2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Bag of Temporal Co-occurrence Words for Retrieval of Focal Liver Lesions Using 3D Multiphase Contrast-Enhanced CT Images

Yingying Xu, Lanfen Lin
College of Computer Science and Technology
Zhejiang University
Hangzhou, China
cs_ying@zju.edu.cn

Hongjie Hu, Dan Wang, Yitao Liu
Department of Radiology
Sir Run Run Shaw Hospital
Hangzhou, China
hongjiehu@zju.edu.cn

Jian Wang, Xianhua Han, Yen-Wei Chen
Graduate School of Science and Engineering
Ritsumeikan University
Shiga, Japan
chen@is.ritsumei.ac.jp

***Abstract*—Computer-aided diagnosis (CAD) systems have been verified to have the potential to assist radiologists in clinical diagnosis to detect and characterize focal liver lesions (FLLs) based on single- or multiphase contrast-enhanced computed tomography (CT) images. Features extracted from multiphase contrast-enhanced CT images carry more important diagnostic information i.e. enhancement pattern and demonstrate much stronger discriminative ability compared to those of single-phase CT images. In this paper, we propose a new method for multiphase image feature generation called the bag of temporal co-occurrence words (BoTCoW). A temporal co-occurrence image connecting intensity from multiphase images is constructed. Then the bag of visual word (BoVW) model is employed on the temporal co-occurrence images to extract temporal features. The proposed method effectively captures temporal enhancement information and demonstrates the distribution of the evolution patterns. The effectiveness of this method is validated in a retrieval system using 132 FLLs with confirmed pathology type. The preliminary results show that the proposed BoTCoW method outperforms the previously proposed temporal features and multiphase features based on the BoVW model.**

Keywords—Computer-aided diagnosis (CAD) systems; multiphase contrast-enhanced CT images; enhancement pattern; bag of visual words (BoVW); bag of temporal co-occurrence words (BoTCoW);

I. INTRODUCTION

Computer-aided diagnosis (CAD) systems have been verified to have the potential to assist radiologists in clinical diagnosis based on image analysis [1]. Research on the development of CAD systems primarily follows two routes. One is to treat the aided diagnosis as a classification issue, and machine learning methods, such as a support vector machine (SVM), are employed as computer-aided diagnostic classification mechanisms [2,3]. The other is to construct a content-based image retrieval (CBIR) system, wherein, a new case is considered as a query case, cases in the repository with the most similar appearance characteristics are retrieved and rendered to support diagnostic decision making [4,5].

Currently computed tomography (CT) is the most important imaging modality employed to detect and characterize focal liver lesions (FLLs) [6,7]. Several CAD systems based on contrast-enhanced CT images have been proposed to identify different types of liver lesions [1,2,4]. Contrast-enhanced CT scans are divided into four phases before and after the injection of contrast. A non-contrast enhanced (NC) scan is performed before contrast injection. After-injection phases include the arterial (ART) phase (30-40 seconds after contrast injection), portal venous (PV) phase (70-80 seconds after contrast injection) and delay (DL) phase (3-5 minutes after contrast injection). In the most previous work, only one single phase CT scans were used for feature generation [3,4] which neglected the pivotal information conveyed by multiphase scans. Recently, significant research has been conducted to verify that features derived from multiphase contrast-enhanced imaging are more effective than the original features derived from non-enhanced images or single-phase scans [1]. Yang et al. had performed a set of experiments comparing the performance of a retrieval system using single-phase features and multiphase features. Their results showed that triple-phase images generate better results than the use of single-phase ones [8]. It was also shown that features extracted from triple-phase scans outperform those based on single- or dual-phase scans [6].

Methods to extract effective features and incorporate multiphase information into feature descriptors have attracted significant attention. The bag of visual words (BoVW) model is a popular strategy to represent images applied in image classification and CBIR. BoVW model has been turned out to

978-1-5090-4846-5/16/$31.00 ©2016 IEEE

achieve promising results in image analysis [8]. In [6], Yu et al. divided the lesion into distinct regions and employed BoVW model to represent the regions. The quantized features were averaged over triple phases. The method proposed by Diamant et al. also improved BoVW model for automatic classification of liver lesions in four-phase images [3]. They generated dual dictionaries based on interior and boundary regions of the lesions. Then two histograms were built and were concatenated to represent a lesion. Yang et al. constructed a visual vocabulary for each category by class-wise clustering and aggregated them as one overall vocabulary that is called a category-specific vocabulary for a single phase [8]. To represent multiphase information, the BoVW histograms of each single phase were merged into a single vector. This preliminary study has demonstrated that combining the BoVW representation of multi phases could improve the retrieval performance to a certain extent. Multiphase CT images convey that different types of liver tumors appear to have discriminating evolution patterns after intravenous contrast injection [7] and these evolution patterns play an important role in identification of hepatic lesions in clinical diagnosis [9]. Since the existing BoVW methods are based on averaging features over all phases or linear combinations of multiple histograms, they ignore the temporal enhancement information and relationship among phases.

Some published studies have reported characterization of FLLs using multiphase images to capture the discriminative enhancement patterns. Chi *et al.* [1] proposed to use average density feature, density derivative feature, texture feature and texture derivative feature and linearly combine the features of multiple phases to represent a lesion's heterogeneity and enhancement pattern. In [7], Roy *et al.* working with the same research group as Chi's, extracted the spatially partitioned mean density and texture features of FFLs from four phases to measure lesion enhancement with respect to the surrounding liver tissues, temporal density to measure the temporal enhancement of the lesion in contrast-enhanced phases with respect to the NC phase and temporal texture features to capture the evolution trend of a lesions' texture appearance. Six textural coefficients based on a 3D gray level co-occurrence matrix (GLCM) are obtained to generate texture features. However, similar to average density, features derived from monolithic lesions can only indicate the global property and are insufficient to reveal local and concrete mutation of density.

In this paper, we propose a new method for multiphase image feature generation called the bag of temporal co-occurrence words (BoTCoW). The BoTCoW approach integrates the intensity of each voxel in homologous position among the region of interest (ROI) in NC, ART, and PV phase. A temporal intensity co-occurrence image for the three phases of each case is constructed. A BoVW technique is employed to obtain the vocabulary of temporal co-occurrence words. The BoTCoW regards the evolution pattern as a unity and merges the intensity of multi phases to display the transmutation. The proposed method reveals the distribution of the evolution pattern for one case effectively. To the best of our knowledge, expressing temporal features among multiphase images has not been investigated previously. In this study, BoTCoW features are extracted from 3D multiphase abdominal CT images and the discriminative performance of the features is verified using a retrieval system based on contrast-enhanced CT images.

II. Proposed Work

The multiphase CT image based retrieval system is proposed in this work. The flowchart of the retrieval system is shown in Fig. 1. Data is first preprocessed prior to feature extraction. The lesion of each case is detected manually and outlined by experienced radiologists. A random walk based interactive segmentation algorithm is applied to segment both the lesion and liver tissue from 3D CT volumes [10]. During a clinical CT study, spatial placement of tissues formed in multiple phases has some aberration due to difference in patient body position, respiratory movements and heartbeat. Therefore, to obtain factual variation of density over phases, a non-rigid registration technique based on B-spline is employed to align images of three phases in order to localize the reference lesion in other phases [11]. Then the proposed BoTCoW method is applied to extract temporal features from lesion regions. A database of feature vectors of cases with their corresponding pathology type label is constructed for similar lesion searching. Histogram intersection distance are used for similarity computation.

A. Material

A total of 132 contrast-enhanced CT liver scans containing five types of lesions with confirmed pathology, i.e., cyst, focal nodular hyperplasia (FNH), hepatocellular carcinoma (HCC), hemangioma (HEM) and metastasis (METS), were obtained. The quantity and pathology type of the different lesions considered in this study are given in Table I. Three-phase scans, i.e., NC, ART and PV phases are collected in our database. Based on radiologists' experience, images of these three phases are sufficient to identify the type of lesions, and the delay (DL) phase is typically not used in order to reduce the radiation dose. One lesion per patient was analyzed, which was outlined by experienced radiologists. CT scans are 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.

B. BoVW model

The bag of visual words (BoVW) model is a widely adopted technique for image representation that is analogous to the bag of words representation for text documents. Patch extraction is the first step in the BoVW approach. In this phase,

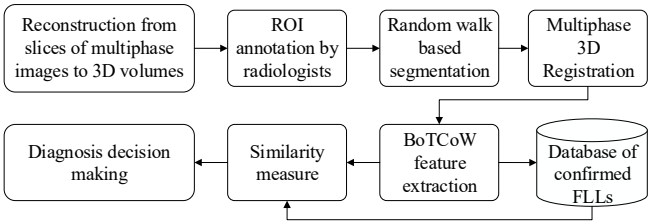

Fig. 1 Flowchart of the retrieval system

TABLE I. Quality and Diagnosis Type of Focal Liver Lesions

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

uniformly sized patches are extracted from an ROI. In medical image, intensity is an important characteristic during diagnosis; thus, to carry more effective information, BoVW representations for medical images always construct patch descriptors based on raw intensity. The key step for BoVW representations is the construction of a visual vocabulary. Clustering algorithms, such as *k*-means, are commonly used to generate clusters of visual words that compose the visual vocabulary for image representation. Given a learned vocabulary, an image is represented as a histogram of visual words in the vocabulary.

C. Bag of Temporal Co-occurrence Word

The proposed bag of temporal co-occurrence words (BoTCoW) is implemented based on conventional BoVW model. The BoTCoW method captures the temporal evolution patterns which is an important issue in clinical diagnosis for radiologists and demonstrates their distribution to discriminate different types of tumors in hepatic tissue. Fig. 2 displays the enhancement pattern of different lesions over three phases. The detailed enhancement characteristics are described as follows. For cysts, FLLs generally appear as a clear round-edged region with a shadow of uniformly low density in the NC phase and show no obvious enhancement after contrast injection. FNH lesions appear to be isodense to the liver parenchyma without any contrast and demonstrates bright ART contrast enhancement in uniformity coefficient. The enhancement continues in the PV phase. In the ART phase, HCC with rich blood supply tend to present hyper-enhancement but shows washout in the PV phase. "Fast in and fast out" is the most prominent characteristic of HCC distinguished from other lesions. HEM exhibits homogeneous enhancement or nodal enhancement of edges occurring in a large lesion in ART. Then, further filling of contrast media inside FLLs gives further enhancement. METs always spread from other organs, and one or more lesions can be detected among the liver tissue. In the ART phase, the peripheral density of METs is slightly greater than that of normal hepatic parenchyma and becomes enhanced in an ulterior manner.

Fig. 3 shows a flowchart of the proposed BoTCoW method. Raw intensity extraction is first performed on the basis of

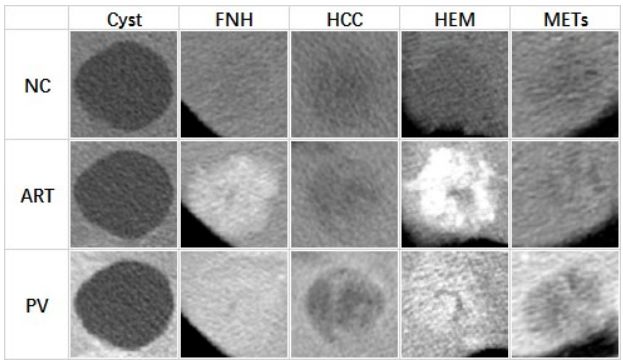

Fig. 2. Evolution patterns of five lesions over three phases.

multiphase images after registration. To reduce the noise level and the undesirable impact caused by registration aberration, the average intensity of a 3×3 square around one voxel is calculated to replace its absolute intensity value for smoothing. Then, we use the triple phase mean filtered images to construct a temporal intensity co-occurrence image as a color image by assigning the intensities of NC, ART, and PV phases as the intensities of R, G and B channels. Five temporal co-occurrence images of different types of FLLs are shown in Fig. 4. Liver is represented by dark-blue, while different FLL will have different color and texture. Temporal co-occurrence image for each case in dataset is constructed.

In the proposed BoTCoW method, we employ the BoVW technique on temporal intensity co-occurrence images to obtain the vocabulary of temporal co-occurrence words and quantize the image based on the vocabulary. Patches are extracted densely from the temporal co-occurrence images to generate descriptors to learn the vocabulary. We employ the *k*-means clustering algorithm to obtain the K clusters as temporal co-occurrence words. The vocabulary $V = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K\}$ is determined by

$$\arg\min_{V=\{w_1,...,w_K\}} \left\{ \sum_i \min_j \left\| \mathbf{x}_i - \mathbf{w}_j \right\|^2 \right\} \quad (1)$$

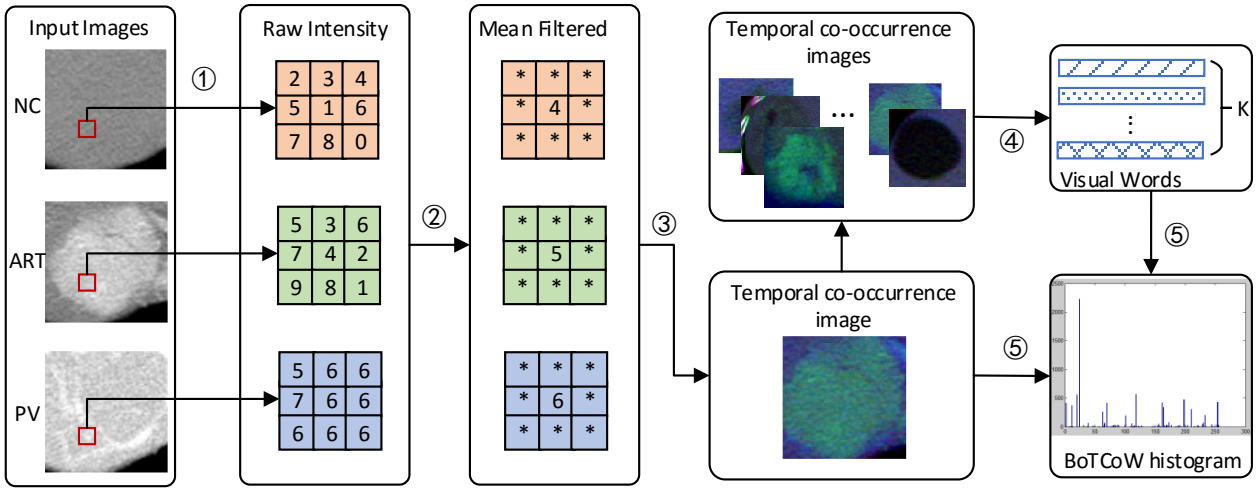

Fig. 3. Flowchart of proposed BoTCoW. Step 1, intensity extraction. Step 2, smoothing. Step 3, temporal co-occurrence image construction. Step 4, dictionary generation. Step 5, quantization

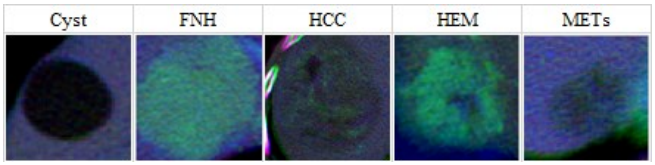

Fig. 4. Temporal co-occurrence images of five different type of FLLs.

where $\mathbf{x}_i$ is the i-th feature vector (patch), $\mathbf{w}_j$ is the j-th center vector (visual word). K is the number of visual words and the dimension of histogram is K. After the vocabulary generation, an image patch is assigned to one visual word $\mathbf{w}$ and the image is delineated by a unique distribution over the temporal co-occurrence words as $(h_1, h_2, ..., h_K)$ which is estimated as follows:

$$h_k = \frac{1}{n}\sum_{i=1}^{n}\begin{cases}1, & if\ \mathbf{w}_k = \arg\min_{\mathbf{w}\in V}(D(\mathbf{w},\mathbf{p}_i))\\ 0\end{cases} \quad (2)$$

where n is the number of patches in the image, $\mathbf{p}_i$ is i-th image patch, $V = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_K\}$ is the generated vocabulary, and $D(\mathbf{w},\mathbf{p}_i)$ denotes the Euclidean distance between a visual word $\mathbf{w}$ ($\mathbf{w} \in V$) and a patch $\mathbf{p}_i$.

D. Similarity and quantitative performance evaluation

The similarity between the query case and database cases is calculated by the histogram intersection which is formulated as follows:

$$S(H_{query}, H_D) = \sum_{i=1}^{K}\min[H_{query}(i), H_D(i)] \quad (3)$$

where H_{query} and H_D are BoTCoW histograms of the query case and the database case, respectively.

The performance is evaluated by a common measure of the precision and recall curve. Precision represents the ratio of retrieved images that are relevant to the class of the query case with respect to the total number of retrieved images. Recall is the number of retrieved images that are relevant to the class of the query case divided by the total number of relevant images in the database. Precision at the top M retrieved FLLs (Prec@M) is defined to represent the proportion of relevant FLLs among the top M results.

III. Experiments

The effectiveness of the proposed BoTCoW method was verified in a CBIR system based on 3D multiphase contrast-enhanced CT images. The dataset was partitioned into training and test set. The training set was used to construct a vocabulary. A leave-one-out validation method was employed to evaluate the system performance. Here one FLL in the test set was selected as a query case, and the remaining 131 FLLs in the dataset were used as labeled cases for retrieval of similar lesions. The experiments were conducted on an Intel Core(TM) i7 (4.00 GHz) with 32 GB RAM, with a MATLAB implement. Details of the run-time of our method are shown in Table II.

TABLE II. Time Measurement of the Proposed Method

Process	Vocabulary generation	Feature calculation	Retrieval
Run-time(sec)	169	0.075	0.0004

Three comparison experiments were performed. The expatiation and comparison results are described in the following sub-sections. The developed medical CBIR system is described in sub-section D. The settings for the corresponding parameters in the BoTCoW algorithm are also discussed in this section.

A. Comparison of BoTCoW and conventional BoVW

We first fixed the sizes of the image patch and visual vocabulary and compared the retrieval performances of BoTCoW and conventional BoVW. Two common BoVW methods proposed by Yang et al. [8], category-specific BoVW (BoVW1) and global BoVW (BoVW2) were compared to the proposed BoTCoW algorithm. Fig. 5 shows the Prec@6 values of the BoVWs and BoTCoW for retrieval of CT images at each single phase and multiphase. The category-specific BoVW is used for retrieval of single phase CT images. Fig. 6 shows the precision-recall curves of the BoTCoW and the conventional BoVW1 and BoVW2 for CBIR based on multiphase CT images. The vocabulary size of the category-specific BoVW was set to be 256 for each category with a step size of 1 and patch size of 11 × 11, which are considered effective for contrast-enhanced CT images when characterizing liver lesions according to previously reported experiments [8]. For a fair

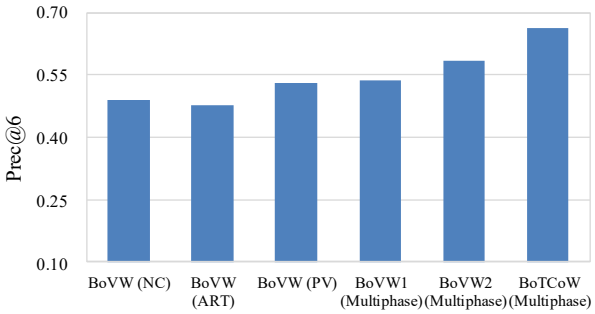

Fig. 5. Prec@6 of the BoVWs and BoTCoW for retrieval at each single phase and multiphase

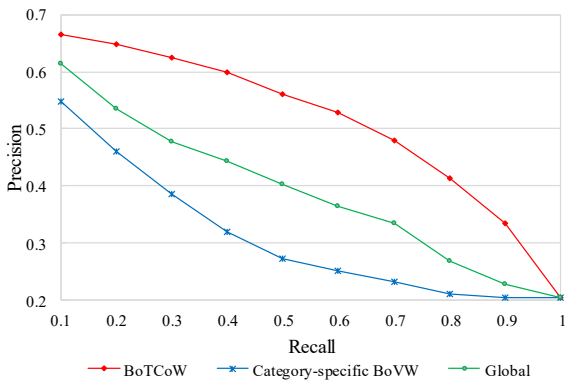

Fig. 6. Comparison results of BoTCoW and conventional BoVW

comparison, the parameters of the BoVW model (e.g. patch size, step size) are the same for the three BoVW-based approaches. For each phase in category-specific BoVW, the total vocabulary size was 256 × 5. Note that generating a 256 × 5 sized vocabulary for the global BoVW algorithm requires a rather high computational cost; thus, the vocabulary size of the global BoVW and the proposed BoTCoW was 256. As shown in Fig. 5, the retrieval performance can be significantly improved by the use of multiphase CT images. As shown in Figs. 5 and 6, the proposed BoTCoW outperforms both global BoVW and category-specific BoVW. In addition, the global BoVW with a relatively small vocabulary size yields a contiguous result compared to the category-specific BoVW.

B. Comparison of BoTCoW and previous temporal features

The temporal features in Roy's work were proposed to demonstrate the different enhancement patterns of different lesions over multiple phases [7]. The method was executed in our experiments for comparison to the proposed BoTCoW. The retrieval performance was evaluated by precision and recall curve. Fig. 7 shows that the proposed BoTCoW generates reasonably better results than the Roy's temporal features.

C. Impact of vocabulary size and patch size

Vocabulary size (K) and patch size are important factors for retrieval performance. We varied the vocabulary size and patch size in our experiments. The patch size was set to 3 × 3 to assess the impact of the vocabulary size on retrieval performance. Table III shows that, generally, a larger vocabulary size leads to higher performance. However, the higher retrieval performance comes at the expense of increased computational cost. When the vocabulary size was set to 256, the system can obtain reasonably desirable performance at lower computational expense. Thus, the vocabulary size is set to 256 to assess the impact of the patch size. The results are shown in Table IV. Note that a 3 × 3 patch size generated better retrieval performance than other patch sizes.

TABLE III. SYSTEM PERFORMANCE WITH VARIOUS VOCABULARY SIZE

Vocabulary Size	Prec@2	Prec@6	Prec@10	Prec@15
16	0.6818	0.6616	0.6417	0.6056
64	0.6856	0.6667	0.6492	0.6253
256	0.6932	0.6679	0.6500	0.6268
512	0.6818	0.6705	0.6500	0.6258
1024	0.6880	0.6717	0.6545	0.6258

TABLE IV. SYSTEM PERFORMANCE WITH VARIOUS PATCH SIZE

Patch Size	Prec@2	Prec@6	Prec@10	Prec@15
3 × 3	0.6932	0.6679	0.6500	0.6268
5 × 5	0.6553	0.6629	0.6515	0.6227
7 × 7	0.6629	0.6629	0.6500	0.6207
9 × 9	0.6515	0.6578	0.6508	0.6217
11 × 11	0.6629	0.6616	0.6439	0.6146

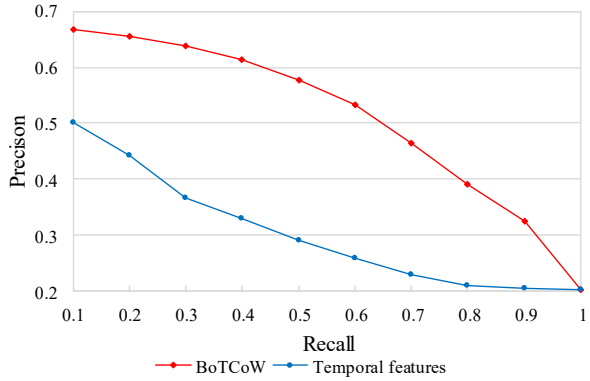

Fig. 7. Comparison of BoTCoW and previous temporal features

D. System Construction

We have developed a medical CBIR system based on our BoTCoW model. In addition to the feature of temporal enhancement by the proposed BoTCoW in this paper, a shape feature, which was proposed in our previous work [9], is also used for medical image retrieval. We applied principle component analysis (PCA) to the 3D lesion mask image and calculated its eigenvectors and eigenvalues. The three eigenvalues are used to represent the sphericity of the lesion as shape features. The performance of retrieval system using BoTCoW and shape features was displayed in Fig. 8 in terms of precision of top M retrieved results (M is from 1 to 8 for the system only concerns the most similar cases). One example of the system retrieval results was exhibited in Fig. 9. Six most similar cases were retrieved and rendered to radiologists along with their CT images, corresponding diagnosis reports, clinical data, *etc*.

IV. CONCLUSION

We have proposed a BoTCoW method to discriminate lesions by representing the distribution of evolution pattern in an ROI for one case based on multiphase contrast-enhanced CT images. The effectiveness of the proposed method was validated in a retrieval system based on 132 FFLs with confirmed pathology type. The proposed BoTCoW algorithm captures temporal enhancement information effectively and demonstrates the distribution of the evolution patterns. We also

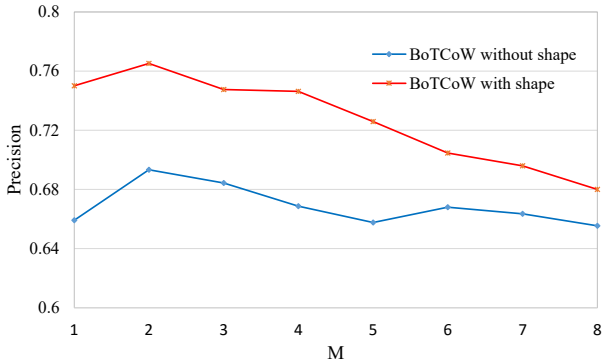

Fig. 8. Precision of top M retrieved results with and without shape feature

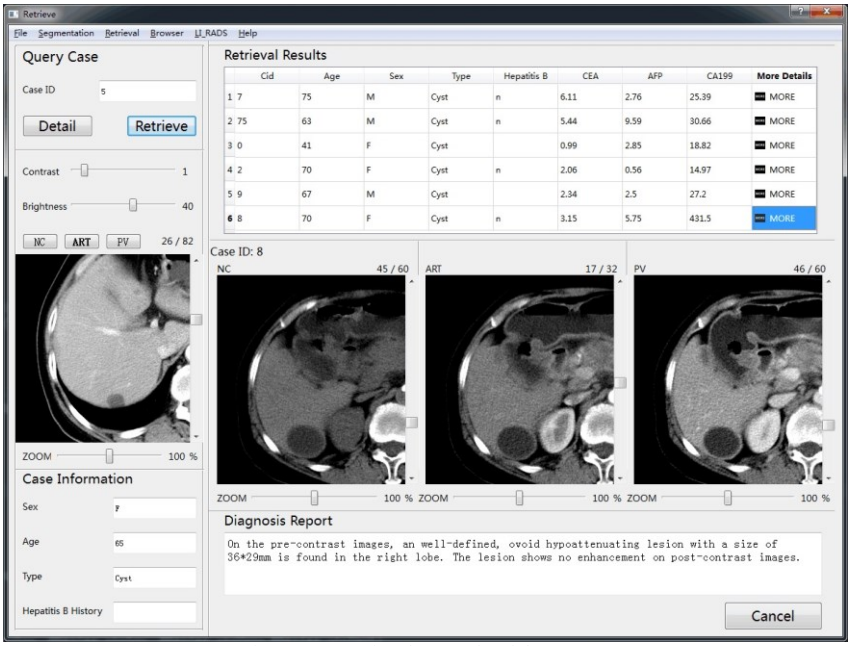

Fig. 9. One retrieval example of the system

developed a medical CBIR system based on our proposed method. The precision at the top 6 is about 0.7, which can be improved by increasing the size of database. Furthermore, in this study, only the intensity of multiphase images was used to represent variation over phases. In future, texture-based temporal co-occurrence patterns may be used to represent the enhancement information. For example, the temporal co-occurrence images can be produced by local binary pattern (LBP) code rather than the intensity value. The BoVW model can then be applied to generate the temporal local binary pattern code co-occurrence words and quantize the multiphase images based on the vocabulary. In addition, a more efficient clustering algorithm for generating a dictionary in this model should be developed according to additional experimental efforts.

ACKNOWLEDGMENT

This research was supported in part by National Science and Technology Support Program of China under the Grant No.2013BAF02B10, 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).

REFERENCES

[1] Y. Chi, J. Zhou, S. K. Venkatesh, Q. Tian, and J. Liu, “Content-based image retrieval of multiphase CT images or focal liver lesion characterization,” Med. Phys., vol. 40, no. 10, art. 103502, 2013.

[2] I. Diamant, J. Goldberger, E. Klang, and M. Amitai. "Multi-phase liver lesions classification using relevant visual words based on mutual information." in IEEE International Symposium on Biomedical Imaging IEEE, 2015.

[3] I. Diamant, A. Hoogi, C. Beaulieu, M. Safdari, E. Klang, M. Amitai, H. Greenspan, and D. Rubin, “Improved patch based automated liver lesion classification by separate analysis of the interior and boundary regions,” IEEE J Biomed Health Inform. 2015 Sep . PMID: 26372661.

[4] S.A. Napel, C.F. Beaulieu, C. Rodriguez, J. Cui, J. Xu, D. Korenblum, H. Greenspan, Y. Ma, and D. L. Rubin, “Automated retrieval of CT images of liver lesions on the basis of image similarity: Method and preliminary results,” Radiology, vol. 256, no. 1, pp. 243–252, 2010.

[5] C.B. Akgül, D.L. Rubin, S. Napel, et al. “Content-Based Image Retrieval in Radiology: Current Status and Future Directions,” Journal of Digital Imaging, pp. 208-22, 2010.

[6] M. Yu, Q. Feng, W. Yang, Y. Gao, and W. Chen, “Extraction of lesion-partitioned features and retrieval of contrast-enhanced liver images,” Comput. Math. Methods Med., vol. 2012, pp. 12, 2012.

[7] S. Roy, Y. Chi, J. Liu, S. Venkatesh, and M. Brown, “Three-dimensional spatio-temporal features for fast content-based retrieval of focal liver lesions,” IEEE Transactions on Bio-Medical Engineering, vol. 92, pp. 1–10, June 2014.

[8] W. Yang, Z. Lu, and M. Yu, et al, “Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single- and multiphase contrast-enhanced CT images,” J. Digit. Imag., vol. 25, pp. 708–719, 2012.

[9] Y. Xu, L. Lin, H. Hu, C. Jin, J. Wang, X. Han and Y.-W. Chen, “Combined Density, Texture and Shape Features of Multi-phase Contrast-Enhanced CT Images for CBIR of Focal Liver Lesions: A Preliminary Study,” in Innovation in Medicine and Healthcare 2015, Springer International Publishing, pp. 215-224, 2015.

[10] C. Dong, Y.-W. Chen, L. Lin, H. Hu, C. Jin, T. Tateyama, X. Han, “Simultaneous Segmentation of Multiple Organs Using Random Walks,” Journal of Information Processing Society of Japan, Vol.24, No.2, pp. 320-329, 2016.

[11] C. Dong, Y.-W. Chen, T. Seki, R. Inoguchi, C. Lin, and X. Han, “Non-rigid image registration with anatomical structure constraint for assessing locoregional therapy of hepatocellular carcinoma,” Computerized Medical Imaging & Graphics, vol.45, pp. 75-83, 2015.