



Focal Liver Lesion Classification Based on Tensor Sparse Representations of Multi-phase CT Images

Jian Wang^{1,2}, Xian-Hua Han³, Jiande Sun¹, Lanfen Lin⁴, Hongjie Hu⁵,
Yingying Xu⁴, Qingqing Chen⁵, and Yen-Wei Chen^{2,4}(✉)

¹ School of Information Science and Engineering,
Shandong Normal University, Jinan, China

² College of Information Science and Engineering,
Ritsumeikan University, Kyoto, Japan
chen@is.ritsumei.ac.jp

³ Faculty of Science, Yamaguchi University, Yamaguchi, Japan

⁴ College of Computer Science and Technology,
Zhejiang University, Hangzhou, China

⁵ Department of Radiology, Sir Run Run Shaw Hospital, Hangzhou, China

Abstract. The bag-of-visual-words (BoVW) method has been proved to be an effective method for classification tasks in both natural imaging and medical imaging. In this paper, we propose a multilinear extension of the traditional BoVW method for classification of focal liver lesions using multi-phase CT images. In our approach, we form new volumes from the corresponding slices of multi-phase CT images and extract cubes from the volumes as local structures. Regard the high dimensional local structures as tensors, we propose a K-CP (CANDECOMP/PARAFAC) algorithm to learn a tensor dictionary in an iterative way. With the learned tensor dictionary, we can calculate sparse representations of each group of multi-phase CT images. The proposed tensor was evaluated in classification of focal liver lesions and achieved better results than conventional BoVW method.

Keywords: Multi-phase CT · Tensor analysis · Sparse coding
Image classification · Focal liver lesion

1 Introduction

Liver cancer is one of the leading causes of death worldwide. Early detection of liver cancers by analysis of medical images is a helpful way to reduce death due to liver cancer. High-definition medical images produced by modern medical imaging devices provide more detailed descriptions of tissue structures and thus facilitate more accurate diagnoses. High-definition medical images and large unorganized medical datasets, however, post challenges to doctors from the viewpoint

of analysis and review. Computer-aided diagnosis (CAD) systems will assist doctors by characterizing the focal liver lesion (FLL) images.

Based on clinical observations, different types of liver lesions exhibit different visual characteristics at various time points after intravenous contrast injection. To capture the visual feature transitions of liver tumors over time, multi-phase contrast-enhanced computer-tomography (CT) scanning is generally employed on patients who are thought to have liver problems. In the multi-phase contrast-enhanced CT scan procedure, four phases of images are obtained: noncontrast-enhanced (NC) phase images are obtained from scans before contrast injection, arterial (ART) phase scanned 25–40 s after contrast injection, portal venous (PV) phase 60–75 s after contrast injection, and delayed (DL) phase scanned 3–5 min after contrast injection.

Characterization of FLLs, including classification and retrieval, has attracted considerable research interest recently. Mir *et al.* [1] first presented texture analysis in liver characterization, which illustrated the importance of gray-level distribution for distinguishing normal and malignant tissue. Yu *et al.* in [2] developed a content-based image retrieval system to differentiate among three types of hepatic lesions by using global features derived from a nontensor product wavelet filter and local features based on image density and texture. Roy *et al.* [4] used four types of features, that is, density, temporal density, texture, and temporal texture, which are derived from four-phase medical images, to retrieve the most similar images of five types of liver lesions. Shape feature was adopted in [5] in combination with density and texture features for retrieving five types of FLLs. Comparing to low-level features introduced above, the middle-level feature bag-of-visual-words (BoVW) has been proved to be considerably more effective for classifying and retrieving natural images. Diamant *et al.* [8] learned BoVW representation of the interior and boundary regions of FLLs for classifying three types of FLLs from single-phase CT images. A variant of BoVW called bag of temporal co-occurrence words (BoTCoW) was proposed by Xu *et al.* [9]. In BoTCoW, BoVW was applied to temporal co-occurrence images, which were constructed by connecting the intensities of multi-phase images, to extract temporal features for retrieving five types of FLLs from triple-phase CT images. After a common codebook learning procedure, Diamant *et al.* [11] proposed a visual word selection method based on mutual information to select more meaningful visual words for each specific classification task. In addition to these variants and enhanced versions of BoVW based on the hard-assignment mechanism, Wang *et al.* [12] learned sparse representations of local structures, which is a soft-assignment BoVW method, of multi-phase CT scans for FLL retrieval. Research on learning high-level features by deep learning methods, in particular using the convolutional neural networks (CNNs), is growing rapidly. [13] surveyed the use of deep learning methods in medical image analysis tasks, such as image classification, object detection, segmentation, registration. Due to the difficulties in collecting professional marked medical images, current medical image databases are always too small for deep learning methods. Most of current approaches use pre-trained CNNs to extract feature descriptors from medical images. We have not

yet seen many applications of deep learning methods in medical image feature extraction, especially in classification of focal liver lesions. To our knowledge, Bag-of-Visual-Words is still the state-of-the-art method in this field.

However, the conventional vector-based BoVW methods, as mentioned above, analyze the multi-phase images separately, in which the temporal co-occurrence information is neglected. In this study, we explore a multilinear generalization of the soft-assignment BoVW, that is, the tensor sparse representation approach, for joint analysis of multi-phase CT images and apply the proposed method for classification of four classes of focal liver lesions.

2 Tensor Sparse Representation of Multi-phase Medical Images

2.1 Tensor Codebook Learning by the Proposed K-CP Algorithm

First, we introduce the notations used throughout this paper. A vector is denoted by a lowercase boldface letter, for example, \mathbf{x} . A matrix is denoted by an uppercase boldface letter, for example, \mathbf{X} . A tensor is denoted by a Lucida Calligraphy letter, for example, \mathbf{X} . We define tensor multiplication in a way similar to that in [14].

Given a set of tensor training samples \mathbf{Y} , we proposed a K-CP method to learn tensor codebook \mathbf{D} . Implementation of the proposed K-CP method comprises two iterated stages: calculation of sparse coefficients, assuming that the codebook is fixed, and codeword update based on the calculated sparse coefficients.

The first stage can be solved easily by using the tensor generalization of Orthogonal Matching Pursuit (OMP) algorithm. The OMP algorithm is a greedy algorithm that finds sparse coefficients of vector-based signals using a given codebook, whose codewords (atoms) are also vectors. In tensor OMP, given a collection of samples $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N]$, where $\mathbf{Y}_i \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_M}$, $i = 1, 2, \dots, N$, is an M^{th} -order tensor and $\mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_M \times N}$ is an $(M + 1)^{\text{th}}$ -order tensor. Suppose a codebook \mathbf{D} comprises of K tensor codewords $\mathbf{D}_k \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_M}$. Then, \mathbf{D} is a $(M + 1)^{\text{th}}$ -order tensor. The tensor OMP can be formulated as follows:

$$i = 1, 2, \dots, N \quad \min_{\mathbf{x}_i} \|\mathbf{Y}_i - \mathbf{D} \bar{\times}_{(M+1)} \mathbf{x}_i\|_2^2, \quad (1)$$

$$s.t. \quad \|\mathbf{x}_i\|_0 \leq T, \forall i$$

where a column vector \mathbf{x}_i in \mathbf{X} represents a combination of the codewords that approximates a sample \mathbf{Y}_i , and T is a sparsity measure.

In the codeword update stage, each tensor codeword is updated individually. To update codeword \mathbf{D}_k , we first find the row vector \mathbf{x}_k^T in \mathbf{X} , in which each entry corresponds to the coefficient of a sample in \mathbf{Y} to \mathbf{D}_k . Then, we define the approximation error without using codeword \mathbf{D}_k as follows:

$$\mathbf{E}_k = \mathbf{Y} - \sum_{j \neq k}^K \mathbf{D}_j \circ \mathbf{x}_j^T \quad (2)$$

The total reconstruction error can be written as follows:

$$\| \mathbf{Y} - \mathbf{D} \times_{(M+1)} \mathbf{X} \|^2 = \| \mathbf{E}_k - \mathbf{D}_k \circ \mathbf{x}_k^T \|^2 \tag{3}$$

Algorithm 1. Tensor codebook learning via TOMP & K-CP

1. Task: Find the optimal dictionary that represent the data samples $\{\mathbf{y}_i\}_{i=1}^N$ as sparse compositions. Objective function is:

$$\min_{\mathbf{D}, \mathbf{X}} \{ \| \mathbf{Y} - \mathbf{D} \times_{(M+1)} \mathbf{X} \|_F^2 \}, \text{ s.t. } \| \mathbf{x}_i \|_0 \leq T, \forall i$$

2. Initialize: Set dictionary matrix $\mathbf{D}^{(0)}$ with l^2 normed columns

3. **while** (stopping rule not met) **do**

4. **Sparse coding stage:** Compute \mathbf{x}_i to approximate example \mathbf{y}_i using any pursuit algorithm, by solving:

$$i = 1, 2, \dots, N \quad \min_{\mathbf{x}_i} \{ \| \mathbf{y}_i - \mathbf{D} \times_{(M+1)} \mathbf{x}_i \|_2^2 \}, \text{ s.t. } \| \mathbf{x}_i \|_0 \leq T$$

5. **Codebook update stage:** Update each dictionary atom \mathbf{D}_k , $k=1, 2, \dots, K$

6. --Define the group of indices corresponding to atom \mathbf{D}_k

$$\omega_k = \{ i \mid 1 \leq i \leq N, \mathbf{x}_k^T(i) \neq 0 \}$$

7. --Compute the overall representation error:

$$\mathbf{E}_k = \mathbf{Y} - \sum_{j \neq k} \mathbf{D}_j \mathbf{x}_k^T$$

8. --Obtain \mathbf{E}_k^R by restricting \mathbf{E}_k using ω_k

9. --Apply CP on \mathbf{E}_k^R . Codeword \mathbf{D}_k is updated by the rank one tensor in the decomposition result.

10. **end while**

Our aim is to find the optimal \mathbf{D}_k that well approximates the reconstruction error \mathbf{E}_k in Eq. (3), which can be solved easily by applying CP decomposition on \mathbf{E}_k .

CP (CANDECOMP/PARAFAC decomposition) decomposes a P^{th} -order tensor \mathbf{D} into a sum of rank-one tensors [14].

$$\mathbf{D} \approx \sum_{r=1}^R \lambda_r (\mathbf{d}_r^1 \circ \mathbf{d}_r^2 \circ \dots \circ \mathbf{d}_r^P) \tag{4}$$

where \circ denotes the outer product. We suppose the vector \mathbf{d}_r^p is normalized to unit length, and the weight of each rank-one tensor is λ_r .

However, applying CP on E_k directly would fill the coefficient vector \mathbf{x}_k^T , which means that the sparsity would be destroyed. Therefore, we construct a constraint vector $\boldsymbol{\omega}_k = (i|1 \leq i \leq N, \mathbf{x}_k^T \neq 0)$ that captures the nonzero entries of \mathbf{x}_k^T . According to $\boldsymbol{\omega}_k$, we must restrict E_k and \mathbf{x}_k^T to E_k^R and \mathbf{x}_k^R , respectively. By applying CP to E_k^R with a rank-one tensor component, D_k can be updated by using the decomposition result and the coefficient vector \mathbf{x}_k^T can be updated by zero-padding the weight λ , as in Eq. (4)

The process of applying the CANDECOMP/PARAFAC (CP) decomposition to the reconstruction residual tensor is executed K times to update each of the K tensor codewords in each iteration. Thus this method is called K-CP method.

The above two stages are iterated until a pre-specified reconstruction error is achieved or the maximum iteration number is reached. The details of the K-CP method for overcomplete tensor codebook learning are given in Algorithm 1.

2.2 FLL Classification Using Tensor Sparse Representations of Spatiotemporal Structures

For each patient in the dataset, there are triple-phase (NC/ART/PV) CT images, which is explained in detail in Sect. 3.1. Based on the structure of the dataset, spatiotemporal features are extracted by using the BoVW models, in which codebooks are learned by the proposed tensor sparse coding method.

To capture the temporal feature of multi-phase CT images, corresponding slices from triple-phase CT images were center-aligned according to the tumor masks and stacked to form three-layer volumes. By this operation, the temporal co-occurrence information is transformed into spatial information in the third dimension of the constructed volumes. A spatiotemporal codebook can be learned by applying our proposed method on the tensor training samples, which are local descriptors extracted from three-layer volumes. Spatiotemporal feature of each medical case can be then calculated by summarizing the representations of local descriptors using mean pooling method. Spatiotemporal feature of a query case can also be calculated based on the learned spatiotemporal codebook

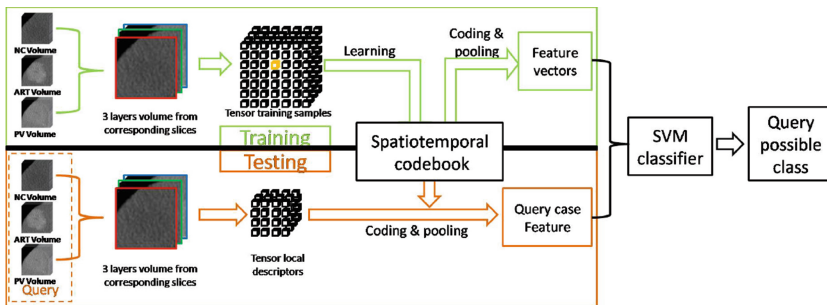


Fig. 1. Learning spatiotemporal features via the proposed tensor sparse coding method from multi-phase images

under the same mechanism. Features of the query and cases in the dataset were fed into a support-vector machine (SVM) classifier with a Radial basis function (RBF) kernel to predict the possible class that the query case may belong to. The workflow is shown in Fig. 1.

3 Experiments and Results

3.1 Multi-phase Medical Dataset

A multi-phase medical dataset was constructed with the help of radiologists to evaluate the performance of the proposed method. The dataset comprises four types of FLLs collected from 111 medical cases. For each medical case, triple-phase (NC/ART/PV) CT images were collected, with spacing of $(0.5 - 0.8) \times (0.5 - 0.8) \times (5/7) \text{ mm}^3$. The size of a CT slice was fixed to 512×512 pixels, while the number of CT slices was set depending on the region scanned (full body or only the abdomen). All tumors in each CT image were manually marked by an experienced medical doctor. In our experiments, however, only the major tumor, that is, the tumor with the largest volume, was considered. As a result, 111 FLLs were selected for use in our experiments, including 38 lesions of the cyst class, 19 cases of focal nodular hyperplasia (FNH), 26 cases of hepatocellular carcinoma (HCC), and 28 cases of hemangioma (HEM). Examples of the four types of FLLs are shown in Fig. 2.

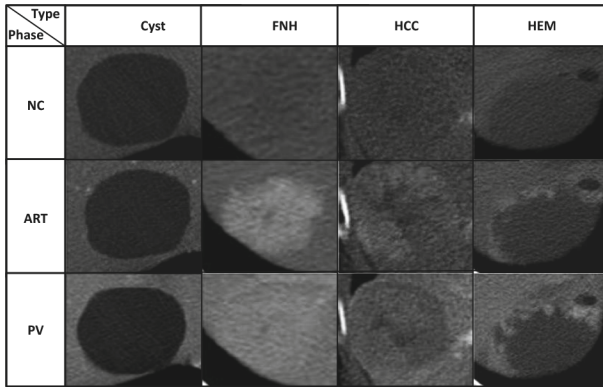


Fig. 2. Examples of each lesion type on 3 phases. Rows are images belong to same contrast phase, while columns are images from same lesion: cyst, FNH, HCC, HEM

3.2 Evaluation Method

Considering the constructed small dataset, the leave-one-out cross-validation method is used in performance evaluation. The classification accuracy are calculated for quantitative measurement, shown as follows:

$$Accuracy = TP / (TP + FP) \quad (5)$$

where, TP is number of correct classified cases, FP represents the number of miss classified cases. $(TP + FP)$ is the total number of cases in the corresponding FLL type.

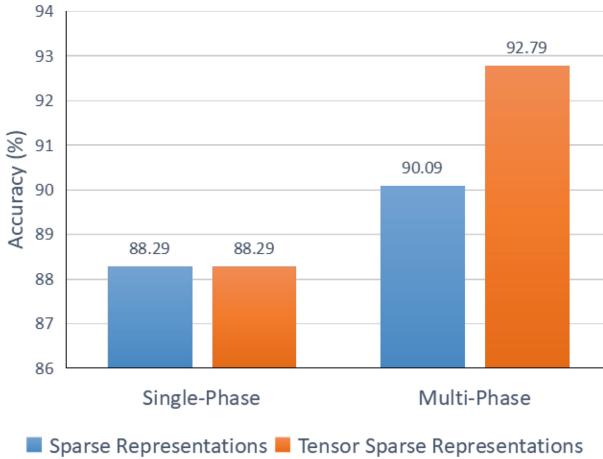


Fig. 3. Comparing classification performance by the proposed tensor sparse representation and the conventional sparse representation method using sing-/multi-phase CT images

3.3 Experimental Results

We compared the classification performance of the proposed tensor sparse representation method with the conventional sparse representation method over both single-/multi-phase medical images, as shown in Fig. 3. We used PV phase images in the single-phase experiments as most of related works do, since most liver lesion types can be visualized clearly on PV phase images. It's interesting that both the two methods got exactly the same results using single-phase images. The accuracy is more significantly improved, however, by the proposed tensor sparse representation method than the conventional one when using multi-phase images, which emphasizes that the proposed method is more effective in capturing the temporal information from multi-phase images. The detailed classification result of the proposed method is shown in Table 1. Due to the clear texture features and temporal enhancement features of Cyst and FNH, they are much easier to be classified from the others when using the temporal co-occurrence information captured by the proposed tensor sparse representation method.

A comparison of the performance of the proposed method with those of the state-of-the-art methods is given in Table 2. As mentioned in previous sections, considerable research effort has been invested to exploring variants and enhanced versions of the BoVW model for FLL characterization. Most of the state-of-the-art methods are based on the BoVW framework. Table 2 shows a comparison of the

proposed method with a few other BoVW models. The proposed tensor sparse coding method outperforms the other methods by preserving spatiotemporal features captured from multi-phase CT images, especially for FNH that shows significant different contrast enhancement features in different phases.

Table 1. The performance of the proposed method

FLL type	Cyst	FNH	HCC	HEM	Accuracy (%)
Cyst	36	0	0	2	94.74
FNH	0	18	1	0	94.74
HCC	0	2	24	0	92.31
HEM	1	1	1	25	89.29

Table 2. Compare the classification accuracy (%) of the proposed method with those of state-of-the-art methods

Methods	Cyst	FNH	HCC	HEM	Averaged
Dual dictionary BoVW [8]	72.09	67.75	96.08	40.00	74.00
BoVW-MI [11]	93.23	90.53	66.67	68.00	79.80
TextureSpecific BoVW [10]	79.16	94.12	80.39	69.24	80.13
BoTCoW [9]	95.83	82.35	84.32	84.61	87.42
The proposed method	94.74	94.74	92.31	89.29	92.79

4 Conclusion

In this paper, we proposed the K-CP method to learn tensor sparse representations of multi-phase medical images. We learned tensor codebooks by using the proposed method and build BoVW models for extracting spatial features and temporal co-occurrence of multi-phase medical images. Experiments of applying the proposed method on focal liver lesion classification showed that the proposed method achieved more significant improvement from single-phase to multi-phase images than conventional sparse representation method and the proposed method outperforms the state-of-the-art methods in this task.

Acknowledgments. This research was supported 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. 18H03267, and No. 18H04747, in part by the Key Science and Technology Innovation Support Program of Hangzhou under the Grant No.20172011A038, and in part by the National Key Basic Research Program of China under the Grant No. 2015CB352400.

References

1. Mir, A.H., Hanmandlu, M., Tandon, S.N.: Texture analysis of CT-images. *IEEE Eng. Med. Biol.* **5**, 781–786 (1995)
2. Yu, M., Lu, Z., Feng, Q., Chen, W.: Liver CT image retrieval based on non-tensor product wavelet. In: *International Conference on Medical Image Analysis and Clinical Applications (MIACA)*, pp. 67–70 (2010)
3. Duda, D., Kretowski, M., Bezy-Wendling, J.: Texture characterization for hepatic tumor recognition in multiphase CT. *Biocybern. Biomed. Eng.* **26**(4), 15–24 (2006)
4. Roy, S., Chi, Y., Liu, J., Venkatesh, S.K., Brown, M.S.: Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions. *IEEE Trans. Biomed. Eng.* **61**(11), 2768–2778 (2014)
5. Xu, Y., et al.: Combined density, texture and shape features of multi-phase contrast-enhanced CT images for CBIR of focal liver lesions: a preliminary study. In: Chen, Y.-W., Toro, C., Tanaka, S., Howlett, R.J., Jain, L.C. (eds.) *Innovation in Medicine and Healthcare 2015. SIST*, vol. 45, pp. 215–224. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-23024-5_20
6. Yu, M., Feng, Q., Yang, W., Gao, Y., Chen, W.: Extraction of lesion-partitioned features and retrieval of contrast-enhanced liver images. In: *Computational and Mathematical Methods in Medicine* (2012)
7. Yang, W., Lu, Z., Yu, M., Huang, M., Feng, Q., Chen, W.: Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single- and multi-phase contrast-enhanced CT images. *J. Digit Imaging* **25**, 708–719 (2012)
8. Diamant, I., et al.: 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 (2016)
9. Xu, Y., et al.: Bag of temporal co-occurrence words for retrieval of focal liver lesions using 3D multiphase contrast-enhanced CT images. In: *2016 23rd International Conference on Pattern Recognition (ICPR 2016)* (2016)
10. Xu, Y., et al.: 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. *Int. J. Comput. Assist. Radiol. Surg.* **13**(1), 151–164 (2018)
11. Diamant, I., Klang, E., Amitai, M., Konen, E., Goldberger, J., Greenspan, H.: Task-driven dictionary learning based on mutual information for medical image classification. *IEEE Trans. Biomed. Eng.* **64**(6), 1380–1392 (2017)
12. Wang, J., et al.: Sparse codebook model of local structures for retrieval of focal liver lesions using multi-phase medical images. *Int. J. Biomed. Imaging*, vol. 2017, Article ID 1413297, 13 pages (2017)
13. Litjens, G.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**(9), 60–88 (2017)
14. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Rev.* **51**(3), 455–500 (2009)
15. Foruzan, A.H., Chen, Y.-W.: Improved segmentation of low-contrast lesions using sigmoid edge model. *Int. J. CARS* **11**(7), 1267–1283 (2016)