Medical Image Retrieval Using Multi-Texton Assignment

Qiling Tang¹ · Jirong Yang² · Xianfu Xia³

Published online: 24 August 2017
© Society for Imaging Informatics in Medicine 2017

Abstract In this paper, we present a multi-texton representation method for medical image retrieval, which utilizes the locality constraint to encode each filter bank response within its local-coordinate system consisting of the k nearest neighbors in texton dictionary and subsequently employs spatial pyramid matching technique to implement feature vector representation. Comparison with the traditional nearest neighbor assignment followed by texton histogram statistics method, our strategies reduce the quantization errors in mapping process and add information about the spatial layout of texton distributions and, thus, increase the descriptive power of the image representation. We investigate the effects of different parameters on system performance in order to choose the appropriate ones for our datasets and carry out experiments on the IRMA-2009 medical collection and the mammographic patch dataset. The extensive experimental results demonstrate that the proposed method has superior performance.

Keywords Image retrieval · Texton · Locality-constrained linear coding · Spatial pyramid pooling

Introduction

Image is one of the important forms of multimedia data, which carries a high amount of information. At present, due to the rapid growth of image repositories, an important challenge is how to effectively deal with increasing data volumes, such as data storage, indexing and retrieval.

Images have been widely applied to the medical field with the development and popularization of medical imaging technology. Hospitals thus produce a large number of medical images (such as X-ray, MRI and CT) for diagnosis [1], treatment planning [2], and assessment of curative effect [3] every day. Diagnostic radiologists have to face the burdensome task of reviewing these images, which occupies much of their time and attention. It is an urgent issue to assist radiologists in the diagnostic process to browse these images efficiently and effectively. Moreover, many hospitals and radiology departments are equipped currently with Picture Archiving and Communication Systems (PACS) [4], which provides a platform for image-based diagnosis, teaching, and research, at the same time also produces a critical need for powerful search engines. Digital image retrieval is playing an increasing role in a broad range of applications within the clinical course.

The technique of content-based image retrieval (CBIR) can help identify relevant image content from a large image archive, which would be valuable to the decision support of medical diagnosis. In the clinical decision-making process, physicians usually integrate patient's data with the prior experiences of similar cases. Therefore, it is beneficial or even important to find the images of similar cases (like same disease, modality, or anatomic region) from the archive to support certain diagnoses [5].

Given a query image, the CBIR system searches for similar images from a reference library through extracting the visual cues of images including low-level features (primitive),

✉ Qiling Tang
qltang@mail.scuec.edu.cn

¹ South Central University for Nationalities, College of Biomedical Engineering, Wuhan 430074, People's Republic of China

² Huibei Key Laboratory for Medical Information Analysis and Tumor Treatment, Wuhan 430074, People's Republic of China

³ Key Laboratory of Cognitive Science, State Ethnic Affairs Commission, Wuhan 430074, People's Republic of China

middle level features (logical), and high-level features (abstract) as the search criteria [6], which is complementary to the text-based retrieval approach that relies on manually assigned annotations and does not consider the visual characteristics of images.

Image representation is a key problem in CBIR system. The bag-of-words (BoW) framework is one of the most popular approaches of the recent years for representing image features in image categorization and retrieval [7, 8], in which an image is first decomposed into a collection of local features, and then these features are mapped to a suitable vocabulary of “visual words” that are obtained by clustering. This image representation method is robust enough to handle spatial translations of features and achieves promising performance for image classification task.

Our work focuses on representing the image contents, that is, using suitable feature representation to characterize medical images. We present a combined texton and locality-constrained linear coding strategy for the retrieval task based on the BoW framework. In our method, the visual dictionary is built by treating textons as words, and then the texton map of an image is generated by assigning the k nearest neighbors to the corresponding pixel via coding scheme, which can reduce the information loss in comparison to the method of the single nearest assignment. Subsequently, spatial pyramid matching method is utilized to capture the spatial layout information of texton distributions and to obtain feature vectors, which further increases the descriptive power of the image representation.

The rest of the paper is organized as follows. “[Related Work](#)” section reviews the related work in the literature. In “[Texton Assignment by Locality-Constrained Linear Coding](#)” section, we introduce the details of multi-texton assignment, including extracting, encoding texton descriptors, and so on. “[Similarity Measure](#)” section describes spatial pooling for texton distributions and similarity measure for texton distributions. “[Experimental Results](#)” section validates the effectiveness of this proposed method on the Image Retrieval in Medical Applications (IRMA) database [9] and on the mammographic patch dataset. Finally, we conclude with a discussion in “[Conclusion](#)” section.

Related Work

CBIR techniques have been extensively explored in the computer vision field over the last 2 decades [10, 11]. In the medical domain, the CBIR system is also becoming an important tool in the clinical decision-making process. A comprehensive review on the development and applications of medical image retrieval was presented in [12].

Representing image features has been an active research topic in the development of the CBIR system. Texture features

are highly representative of visual cues, which have received considerable attention and have successfully applied in medical image retrieval tasks. Various methods have been introduced to analyze texture attributes of images, such as gray-level co-occurrence matrix (GLCM) [13], Gabor filters [14], local binary pattern (LBP) [6], and local phase quantization (LPQ) [15].

The texton and patch exemplar methods have proven to be more powerful for texture classification [16]. The texton and patch exemplar representations are closely related to the BoW framework [17]. Different from GLCM or Gabor, the texton-based BoW approach can design task-specific feature representation. The texton method provides a powerful local descriptor for quantitatively capturing texture appearance, which can be utilized to yield visual words for the BoW in many medical imaging applications. Galaro et al. [18] utilized the filter responses obtained with the Haar wavelet to construct the texton vocabulary for identifying anaplastic medulloblastomas. Riaz et al. [19] integrated autocorrelation Gabor features (AGF) in a texton framework to classify and retrieve gastroenterology images. A scheme of combining texton analysis with majority voting over multiple subsampling strategies for mammogram classification was presented in [20]. Haralick statistical descriptors were introduced to texton maps for automatic classification of breast tissue microarray images [21].

In these texton-based CBIR methods, each pixel is mapped to its nearest neighbor in the texton dictionary, and then an image is represented as a histogram of texton occurrence frequencies. However, one inherent component of the transitional nearest neighbor model is the assignment of discrete visual words to continuous image features, which shows a clear mismatch of this hard assignment with the nature of continuous features [22].

To make up for the information loss caused by a single nearest neighbor assignment, we propose a solution of multi-texton assignment, which tends to work better than the hard assignment typically used in visual words models. The main differences between our proposed method and the above texton-based methods are (1) we map a pixel through multi-texton coding instead of assigning to a single nearest visual word in the texton vocabulary, reducing the quantization errors; (2) we take into account the spatial layout among texton features by using multi-scale spatial pyramids, which improve performance over the orderless texton framework in image retrieval tasks.

Texton Assignment by Locality-Constrained Linear Coding

Julesz [23] introduced the concept of textons in order to characterize the preattentive human texture perception, and Malik et al. [24] proposed a computational method of textons based

on the dense description of local texture features, in which the textons (that is, class-specific textural primitives) are viewed as assemblies of filter responses.

The procedure of the texton computation [24] consists of two stages—clustering and assignment. First, in the texton learning stage, the training images are convolved with a filter bank and the resulting filter responses are clustered unsupervisedly using the K -means algorithm. Each cluster builds a Voronoi cell in the space of filter response vectors, and the cluster center is defined as texton; then, these textons are used as a vocabulary to represent images. Second, in the testing stage, given an input image, its corresponding pattern is produced by first convolving it with the filter bank, and then each pixel in this image is assigned to the nearest texton in terms of the Euclidean distance, generating the texton map.

The notion of texton is an effective tool for describing human textural perception, which is successfully applied to different visual processing tasks, including contour detection [25], texture classification [26], and object recognition [27]. It also finds widespread applications in the medical domain, such as breast cancer risk estimation [28], breast density characterization [29], and retinal vessel segmentation [30].

The texton-based methods mentioned above achieve good performance in various medical image tasks. However, it is worth noting that the texton assignment in these works all adopts the traditional method, i.e., mapping each pixel to the texton which lies closest to it in filter response space. Consequently, the filter response of each pixel is approximated by a single texton in the codebook, without considering the relationships between different textons. Such a rough approximation would yield the large quantization errors, which inevitably contributes to reduced performance in the tasks of detection, identification, segmentation, and so on. In order to compensate for the loss of information caused by a single texton assignment, a natural idea is to express the filter response by means of combining multiple textons.

The locality-constrained linear coding (LLC) algorithm proposed in [31] is a very good choice for feature coding that can provide a set of proper weights to the codebook. An extension of this work was presented in [32], using a generalized LLC algorithm together with an approximate geodesic distance for clinical diagnosis of Alzheimer's disease. The essence of LLC is to utilize the locality constraint to encode each descriptor within its local-coordinate system, that is, only a small subset of codebook that are locally similar to the given data would be used for coding rather than the entire set of codebook. In practice, for a given data, it first performs the k -nearest-neighbor search to find local-coordinates of the data, and then solves a constrained least square fitting problem on the local-coordinates.

In the texton learning stage, training images from the IRMA-2009 database [9] are convolved with a filter bank to produce feature vectors of each pixel. These vectors are then

clustered using K -means to yield image-specific textons and different textons are aggregated to build a texton dictionary. In this work, the filter bank consists of first and second derivatives of Gaussians at 6 orientations and 3 scales, 8 Laplacian of Gaussian filters, and 4 Gaussian filters. $\mathbb{B} = \{\mathbf{b}_1, \mathbf{b}_1, \dots, \mathbf{b}_m\} \in \mathbb{R}^{d \times m}$ is the obtained texton dictionary in which d is the dimension of the features and m is the number of cluster centers. The process of building texton dictionary is illustrated in Fig. 1.

Here, the emphasis is on how to implement multi-texton assignment for the filter response of each pixel in the testing stage, which is different from the nearest texton assignment approaches in [18–21, 24]. We adopt the LLC algorithm to code the filter responses that are represented by introducing a set of suitable weights to the texton dictionary. The coding strategy of combining multiple textons should reduce the information loss and the quantization error of the single texton assignment.

Given a set of texton features obtained from an image $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$. The first step of the LLC is to compute the distance between feature vector $\mathbf{x}_i$ and each texton in the dictionary $\mathbb{B}$. The k nearest neighboring textons of $\mathbf{x}_i$ based on this distance are selected as local bases to construct a local coordinate system, denoted as $\mathbf{B} = \{\mathbf{b}_1, \mathbf{b}_1, \dots, \mathbf{b}_k\} \subseteq \mathbb{B}$. Then, $\mathbf{x}_i$ can be coded by linear combination of local bases $\mathbf{B}$, and the weight vectors α_i are obtained by solving the following minimization problem:

$$\min_{\mathbf{A}} \|\mathbf{x}_i - \mathbf{B}\alpha_i\|^2 + \lambda \|\mathbf{d}_i \odot \alpha_i\|^2, \quad \text{s.t. } \mathbf{1}^T \alpha_i = 1, \quad \forall i \quad (1)$$

where $[\alpha_1, \alpha_2, \dots, \alpha_n]$ is the set of codes for $\mathbf{X}$. $\mathbf{d}_i \in \mathbb{R}^k$ is the distance function between feature vector $\mathbf{x}_i$ and each basis in $\mathbf{B}$, and $\odot$ denotes the element-wise multiplication. The coefficient λ is a predetermined constant that controls the tradeoff between the two terms in the above expression (here, we used $\lambda = 0.0001$). The values of α_i indicate the contribution of each local basis to the reconstruction of $\mathbf{x}_i$. The purpose of the LLC is to estimate a set of weights that can minimize the reconstruction error of $\mathbf{x}_i$ with the local bases under the locality constraint.

Table 1 presents four medical image examples in which we compute the errors between the filter responses and the texton assignment values obtained with the traditional nearest method and with the LLC method respectively. Comparison to the reconstruction results of the single-texton assignment and the multi-texton assignment shows that both the mean error rate and standard deviation of our proposal for each test image is remarkably better than that of the traditional approach, indicating that each filter bank response can be more accurately represented by the weighted combination of multiple textons. Furthermore, it can be seen from the Table 1 that

Fig. 1 Schematic illustration of building texton dictionary. The filter bank is used for computing feature responses. The corresponding pattern is generated by convolving train images with the filter bank, and

thus each pixel obtains a vector of responses to the filter bank. The resultant filter responses are clustered into textons using the *K*-Means algorithm and then textons are aggregated to form the texton dictionary

the reconstruction error decreases with the increase of the number of combined neighbors.

Similarity Measure

The purpose of image retrieval is to select the similar images to a given query image in a database. For the nearest texton assignment, the similarity between two images is usually computed by representing images as histograms of texton features and then comparing the χ^2 difference between the two histograms. The mean statistics in histogram discards the spatial layout information of texton features, which is incapable of locating an object and severely limits the ability to describe the image. As an extension of the BoW model, spatial pyramid matching (SPM) [33] method adds information about the spatial layout of local features which is successfully applied in the computer vision field [34, 35]. The method partitions an image into multiple spatial subregions from coarser to finer levels, and computes histograms of local features in each

subregion. Finally, a feature vector is formed by concatenating all the histograms to describe the image.

We compute a spatial-pyramid image representation based on locality-constrained linear coding of texton features. Each image is represented by combining multi-texton assignment and spatial pyramid matching. First, an image is divided into 1×1 , 2×2 and 4×4 subregions by using a three-level spatial pyramid. Spatial pooling can effectively merge region features, common approaches such as average pooling and max pooling (i.e., the average and maximum of the codes within each subregion, respectively). Similar to the construction of histograms in SPM, we do average pooling within each of the 21 subregions. For each subregion S , average pooling works as follows:

$$f = \left[\frac{\sum_{i \in S} |\alpha_{i1}|}{\sum_{i \in S} \delta(|\alpha_{i1}| > 0)}, \frac{\sum_{i \in S} |\alpha_{i2}|}{\sum_{i \in S} \delta(|\alpha_{i2}| > 0)}, \dots, \frac{\sum_{i \in S} |\alpha_{im}|}{\sum_{i \in S} \delta(|\alpha_{im}| > 0)} \right] \quad (2)$$

where $|\cdot|$ is an element-wise absolute value operation. α_{ij} is the j th entry of the local-constrained linear coding α_i . $\delta(\cdot)$ is a

Table 1 Four medical image examples on the reconstruction error (mean + std.).

nearest neighbor number k
1	15.34 + 0.08	67.47 + 0.38	36.78 + 0.21	97.53 + 2.68
2	13.58 + 0.09	64.24 + 0.36	34.04 + 0.20	95.88 + 2.68
4	11.36 + 0.08	58.13 + 0.35	30.38 + 0.19	88.23 + 2.67
8	8.65 + 0.06	50.36 + 0.33	26.60 + 0.19	79.40 + 2.67

The mean and standard deviation of the reconstruction error are computed by using a single nearest neighbor (i.e., $k = 1$) and multiple nearest neighbors (i.e., $k \geq 2$), respectively

Fig. 2 The flow chart describes the process of our approach on the image retrieval task. The images in the database and the query image are convolved with the filter bank to obtain their feature descriptors. A texton dictionary with M entries is used to code the feature vectors by the LLC scheme and each feature vector is represented by multi-texton

assignment. Pooling with spatial pyramid method is utilized to integrate the codes inside each subregion. All the pooled features are concatenated together to produce the image representation. The similarity between two images is computed by the L_1 norm distance between the two vectors

zero-one indicator function whether α_{ij} is nonzero. The pooled features from different subregions are concatenated to generate a vector representation F of the image.

The returned results of image retrieval systems largely depend on the feature similarity measure. The L_1 norm distance is used to measure the similarity between the feature vectors of query image q and retrieved image r and defined as,

$$dis(q, r) = \|F_q - F_r\|_1 \quad (3)$$

where $\|\cdot\|_1$ denotes the L_1 norm, a popular choice used in the image-to-image distance [33]. Due to local variations in illumination, shadows and reflections, the magnitude of LLC varies over a large range. Thus the distance is divided by their norms $\|F_q\|_1 + \|F_r\|_1$, which makes the similarity measure robust to such variations and increases their discriminative ability.

The retrieved images from the database are sorted according to their similarity with the presented query image. The flow chart of our approach on the image retrieval task is illustrated in Fig. 2.

Experimental Results

In this section, we first test the performance of this proposed scheme on medical image retrieval. The experimental results demonstrate that multi-texton assignment in our method offers advantages over traditional single-texton assignment. For these experiments, the values of the system parameters are determined empirically by analyzing the performance over a range of possible parameters and then selecting the appropriate ones for our datasets. We also present experimental results on classification of mammographic patches and provide a quantitative comparison of our method to existing methods based on the ground truth.

Retrieval Performance Evaluation

The objective of the medical retrieval task is to return a ranked image set from a database, sorted by their relevance to the

query. To evaluate the CBIR system, both the mean average precision (MAP) [36] and the error score proposed by the ImageCLEF campaign [37] are used as the evaluation criteria to measure the retrieval performance of in our work. Average precision (AP) for each query image q is the average of precisions computed at every position in the ranked sequence of the relevant samples,

$$AP(q) = \frac{1}{N_r} \sum_j rel(j) \times P@j \quad (4)$$

where N_r is the number of relevant samples, and j is the rank in the sequence of retrieved samples. $rel(j)$ is an indicator function on the relevance of a rank j (1 for relevant, and 0 for irrelevant), and $P@j$ denotes the precision at the top j retrieved samples. MAP is the mean of AP over all queries,

$$MAP = \frac{1}{Q} \sum_{q \in Q} AP(q) \quad (5)$$

Since in IRMA dataset the four IRMA code axes (the technical, directional, anatomical and biological axes) are separate, the error score E is computed by adopting a hierarchical approach according to the IRMA axes, which is expressed as follows:

$$E = \sum_{i=1}^I \frac{1}{b_i} \frac{1}{i} \delta(l_i, \gamma_i) \quad (6)$$

$$\text{with } \delta(l_i, \gamma_i) = \begin{cases} 0 & \text{if } l_j = l_j \quad \forall j \leq i \\ 0.5 & \text{if } l_j = * \quad \exists j \leq i \\ 1 & \text{if } l_j \neq l_j \quad \exists j \leq i \end{cases}$$

where b_i is the number of possible labels at position i , and I is the depth of the hierarchy. For every axis, the maximal possible error is calculated and the errors are normalized to a range of 0 to 0.25. So a completely wrong decision for that axis gets an error count of 0.25 and a completely correctly predicted axis has an error of 0.00. If all positions in all axes are wrong, error value is 1. Furthermore, this measure suggests that the decisions at an early stage in the code are more important than those at a later stage.

Retrieval on the IRMA-2009 Medical Collection

We evaluate the proposed approach on the IRMA-2009 medical image collection which contains 15,363 radiographs taken randomly from daily routine work at the Department of Diagnostic Radiology of the RWTH Aachen University [38]. This collection offers 12,677 images for training and 1733 images for testing, and these images were scaled to fit into a 512×512 maximum pixel window keeping the original aspect ratio. All images in the archive are annotated according to the IRMA coding system [38], which consists of four mono-hierarchical axes describing the imaging modality, the body orientation, the body region examined and the biological system examined. The IRMA collection images are subdivided into 193 distinct categories, some of which have large intra-variability and inter-similarity, making the retrieval tasks much difficult [39]. Each image in the IRMA collection can identify its category in terms of the associated code. To evaluate the image retrieval performance, the query images are randomly selected from each category and the remained images in that category are taken as the corresponding ground truth set for the presented queries.

The visual dictionary is built by using the K -means clustering algorithm to find a set of centroids of the textron feature descriptors. An important factor in the construction is the selection of dictionary size, that is, how many words are needed to characterize image contents. For textron assignment, each filter bank response is represented by combining the k -nearest-neighbors in textron dictionary, and thus the number of the nearest neighbors plays an important role in reconstructing the original feature vector.

In order to understand how the parameter choices in our multi-textron representation affect the retrieval performance, we empirically analyze the two main free parameters, the number of the assigned textron, and the size of visual dictionary. Figure 3 shows the effects of different dictionary sizes. As can be seen, the retrieval performance generally improves as the number of dictionary words increases. However, a large vocabulary requires higher computational cost. Since the performance improvement is not significant when dictionary size is more than 1200, this dictionary size is a good choice as the tradeoff between the computational cost and the retrieval performance in our experiments.

To find an appropriate number of the nearest neighbors in textron dictionary for filter bank responses, we assess the impact of the number of assigned textons on the retrieval performance. The experimental results in Table 1 show that using more textons to estimate the response vector can help to reduce the reconstruction error. Similarly, an increase in the number of assigned textons can achieve better retrieval performance in terms of MAP score in Fig. 4. Based on the experimental observation, we set $k = 10$, that is, the 10 nearest neighbors in textron dictionary are selected as local bases to

Fig. 3 Effect of dictionary size on retrieval performance

construct a local coordinate system for each feature vector. Smaller k fails to give satisfactory results and larger k only contributes a slight improvement, but at the expense of higher computational complexity.

Figure 5 presents some retrieval examples from IRMA-2009 medical collection using the proposed method of multi-textron representation, in which the first column represents query images, and the other columns are retrieval results by ranking the images according to the L_1 norm distance between the query image and the image in the database and then selecting the top 10 images as examples.

We quantitatively compare the retrieval precision of our proposed method to related works including the traditional textron histogram method using the nearest neighbor assignment [24], the frequential textron method based on Haralick statistical descriptors [21], the textron framework integrating autocorrelation Gabor features (textron-AGF) [19], and majority voting textron method [20]. Different methods

Fig. 4 Effect of the number of assigned textons on retrieval performance

Fig. 5 Retrieval rankings of medical images. Query images are shown in the left column, and retrieved results in the ascending order of L_1 norm distance by the proposed method shown on the right

are tested on different databases. For comparisons, we repeat these methods on IRMA database and compare their performance in terms of the precision score at the top n retrieved samples ($P@n$), defined as the proportion of relevant samples up to position n [40]. Precision at the top 5, 10, 20 retrieved images is reported in Fig. 6.

The results show that our system generally provides better performance compared to the other retrieval systems based on texton descriptor, which indicates that the multi-texton assignment together with spatial pyramid pooling can better represent the features of medical radiographs. It is also observed that the performance of the proposed framework decreases with the increase in the number of retrieved images, whenever such a decrease is possible.

We further evaluate the classification results with the error score defined in Eq. 6. As is seen in the Table 2, the performance of this proposal is highly competitive with the existing

methods on the IRMA classification and other texton-based classification methods. The best error score is reported as 146.55 in [6]. Nevertheless, our proposed method is able to give satisfactory classification results compared with the other methods. Note that the images in [6] are required a preprocessing procedure of detecting salient regions, and then the data is folded to reduce the effect of irrelevant regions. Thus, the classification performance is highly dependent on the detection results of salient regions, while our method is not affected by the saliency.

Classification on the Mammographic Patch Dataset

The proposed approach is also applied to classify the mammographic images. Mammographic patch dataset (*12er-patches*) from the IRMA project is used for experimental test. According to metadata annotations, the patches centering the

Fig. 6 Precision at $P@5$, $P@10$, and $P@20$ for different texton-based retrieval methods

Table 2 Quantitative comparison of various methods on the irma-2009 benchmark with the error score. A smaller error score indicates a better classification performance

	Method	Error score
Texton-based methods:	Proposed method	158.43
	Majority voting texton [20]	165.38
	Texton-AGF [19]	173.75
	Haralick-based texton [21]	193.46
	Traditional texton [24]	216.81
Other existing methods:	MS LBP [6]	146.55
	TAUbiomed [41]	169.50
	Idiap [37]	178.93
	CNNC + RBC [42]	224.13
	VPASabanci [43]	261.16

The lower part of this table lists some results reported in the literature

lesion are extracted from screening mammographies of different BI-RADS (Breast Imaging Reporting Data System) classes. These patches are divided into four classes in terms of the BI-RADS breast tissue density, corresponding to (I) almost entirely fatty, (II) scattered fibro glandular tissue, (III) heterogeneously dense tissue, and (IV) extremely dense tissue [44]. For each of the four tissue density classes, according to BI-RADS lesion assessment, suspicious regions are labeled as three categories of pathology: (1) no finding (negative), (2) benign, and (5) highly suggested malignant. Figure 7 exemplifies the resulting patches. The dataset of *12er-patches* contains 2796 patches of 12 classes, each with 233 images.

Visually, breast tissues of various densities with or without lesions are distinguished through gray level intensity in mammographies [44]. Texton is an effective tool for representing texture patterns which contain information about the variations of gray levels. It is desirable to choose the texton-based BoW approach for describing breast density and lesion characteristics. The proposed method of the multi-texton assignment with spatial pyramid pooling is employed to represent texture features of mammographic patches, and then the multi-class SVM proposed by Crammer and Singer [45] is used for classification with ten-fold cross validation of all the patches, instead of the L_1 norm distance in the retrieval task. We compute the classification accuracy (*Acc*), sensitivity (*Sens*), specificity (*Spec*), positive predictive value (*PPV*), and *F*-measure as performance indicators. These terms are defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sens = \frac{TP}{TP + FN}$$

$$Spec = \frac{TN}{TN + FP}$$

$$PPV = \frac{TP}{TP + FP}$$

where *TP*, *TN*, *FP*, and *FN* are the number of true positives, true negative, false positives, and false negatives, respectively. The positive predictive value is called precision, and sensitivity is called recall. *F*-measure is the harmonic mean of precision and recall, i.e., $F\text{-measure} = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$.

Fig. 7 Patches from the dataset of *12er-patches* in the IRMA mammography reference database. From left to right, the four columns correspond to the BI-RADS tissue density, classes I to IV. From top to bottom, the three rows depict BI-RADS assessment categories 1, 2, 5, respectively

Table 3 Comparisons of different texton-based methods on the mammographic patch dataset

Methods	Acc (%)	Sens (%)	Spec (%)	PPV (%)	F-measure
Traditional texton [24]	57.8	60.8	63.4	57.5	0.591
Haralick-based texton [21]	60.5	62.2	65.3	59.6	0.609
Texton-AGF [19]	63.9	67.1	68.7	63.0	0.650
Majority voting texton [20]	66.4	69.5	72.1	65.3	0.673
Proposed method	68.2	70.4	72.6	67.2	0.688

Table 3 presents the classification results of different methods on the *12er-patches* dataset. The results indicate that our method of multi-texton assignment with SPM outperforms the competing algorithms in the classification task of mammographic texture patterns.

Conclusion

In this paper, we have presented a multi-texton BoW framework for medical image retrieval and classification, in which feature representation consists of a series of processing steps including feature extraction, feature encoding and feature pooling. This work contribution is twofold: We first utilize the locality constraint to project each filter bank response into the local-coordinate system composed of the k -nearest-neighbor in texton dictionary and find a weighted linear combination of these neighbors for reconstructing each original response. This strategy reduces the quantization errors in mapping process, which is a substantial improvement over previous work of the single texton assignment based on the nearest neighbor; In order to enhance the ability to describe the image, spatial pyramid pooling technique is subsequently employed to capture the spatial layout information of texton distributions, making up for such spatial information loss in the histogram statistics approach.

To verify the effectiveness of our approach, extensive experiments are performed on the IRMA-2009 medical collection and the mammographic patch dataset. We explored the effects of different parameters of the system on the retrieval task and, thus, choose the optimal dictionary size and the optimal number of nearest neighbors to achieve high performance in the retrieval and classification.

The current work focuses on X-ray medical images, whereas it is common to use different imaging modalities (such as X-ray, CT, PET, MRI, angiography, and sonography) to provide complementary information about a special patient. The multimodality medical image retrieval has potential applications in diagnosis, training, and research [46]. In future work, we intend to extend the multi-texton representation approach on the multimodality medical image repositories containing a diverse collection of data, achieving higher CBIR

performance utilizing complementary information from different modalities.

Acknowledgments The main image dataset used in this study is courtesy of the IRMA Group, Aachen, Germany, <http://irma-project.org>. This work was supported by the Fundamental Research Funds for the Central Universities (CZP17033).

References

- Madabhushi A, Agner S, Basavanahally A, Doyle S, Lee G: Computer-aided prognosis: Predicting patient and disease outcome via quantitative fusion of multi-scale, multi-modal data. *Comput Med Imaging Graph* 35(7–8):506–514, 2011
- Zaidi H, Veas H, Wissmeyer M: Molecular PET/CT imaging-guided radiation therapy treatment planning. *Acad Radiol* 16(9): 1108–1133, 2009
- Marcus C, Ladam-Marcus V, Cucu C, Bouché O, Lucas L, Hoeffel C: Imaging techniques to evaluate the response to treatment in oncology: Current standards and perspectives. *Crit Rev Oncol Hematol* 72(3):217–238, 2009
- Greenspan H, Pinhas AT: Medical image categorization and retrieval for PACS using the GMM-KL framework. *IEEE Trans Inf Technol Biomed* 11(2):190–202, 2007
- Dhara AK, Mukhopadhyay S, Dutta A, Garg M, Khandelwal N: Content-based image retrieval system for pulmonary nodules: assisting radiologists in self-learning and diagnosis of lung cancer. *J Digit Imaging* 30(1):63–77, 2017
- Camlica Z, Tizhoosh HR, Khalvatid F: Medical image classification via SVM using LBP features from saliency-based folded data. In: *Proc IEEE Int Conf Mach Learn Appl*, 2015. p. 128–132.
- Tang Q, Liu Y, Liu H: Medical image classification via multiscale representation learning. *Artif Intell Med* 79:71–78, 2017
- Lin Z, Brandt J: A local bag-of-features model for large-scale object retrieval. In: *Proc Eur Conf Comput Vis*, 2010, p. 294–308
- Lehmann TM, Deselaers T, Schubert H, Guld MO, Thies C, Fischer B, Spitzer K: IRMA—a content based approach to image retrieval in medical applications. In: *IRMA Int Conf*, 5033: 911–912, 2006.
- Puzicha J, Hofmann T, Buhmann JM: Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In: *Proc IEEE Conf Comput Vis Pattern Recognit*, 1997, pp 267–272
- Arandjelovic R, Zisserman A: Three things everyone should know to improve object retrieval. In: *Proc IEEE Conf Comput Vis Pattern Recognit*, 2012, pp 2911–2918
- Kumar A, Kim J, Cai W, Fulham M, Feng D: Content-based medical image retrieval: A survey of applications to multidimensional and multimodality data. *J Digit Imaging* 26(6):1025–1039, 2013
- Gletsos M, Mouggiakakou SG, Matsopoulos GK, Nikita KS, Nikita AS, Kelekis D: A computer-aided diagnostic system to characterize

- CT focal liver lesions: design and optimization of a neural network classifier. *IEEE Trans Info Technol Biomed* 7(3):153–162, 2003
14. Lee C, Chen S, Tsai H, Chung P, Chiang Y: Discrimination of liver diseases from CT images based on Gabor filters. In: *Proc IEEE Int Symp Biomed Comput-based Med Syst*, 2006
15. Nanni L, Brahnam S, Lumini A: A very high performing system to discriminate tissues in mammograms as benign and malignant. *Expert Syst Appl* 39(2):1968–1971, 2012
16. Manik V, Andrew Z: A statistical approach to material classification using image patch exemplars. *IEEE Trans Pattern Anal Mach Intell* 31(11):2032–2047, 2009
17. Jégou H, Douze M, Schmid C: Improving bag-of-features for large scale image search. *Int J Comput Vis* 87(3):316–336, 2010
18. Galaro J, Judkins AR, Ellison D, Baccon J, Madabhushi A: An integrated texton and bag of words Cclassifier for identifying anaplastic medulloblastomas. In: *Conf Proc IEEE Eng Med Biol Soc*, 2011, pp 3443–3446
19. Riaz F, Silva FB, Ribeiro MD, Coimbra MT: Invariant Gabor Texture Descriptors for Classification of Gastroenterology Images. *IEEE Trans Biomed Eng* 59(10):2893–2904, 2012
20. Li Y, Chen H, Rohde GK, Yao C, Cheng L: Texton analysis for mass classification in mammograms. *Pattern Recognit Lett* 52:87–93, 2015
21. Fernández-Carrobles MM, Bueno G, Déniz O, Salido J, García-Rojo M, González-López L: Frequential versus spatial colour textons for breast TMA classification. *Comput Med Imaging Graph* 42:25–37, 2015
22. van Gemert JC, Veenman CJ, Smeulders AWM, Geusebroek J: Visual word ambiguity. *IEEE Trans Pattern Anal Mach Intell* 32(7):1271–1283, 2010
23. Julesz B: Textons, the elements of texture perception, and their interactions. *Nature* 290:91–97, 1981
24. Leung T, Malik J: Representing and recognizing the visual appearance of materials using three-dimensional textons. *Int J Comput Vis* 43(1):29–44, 2001
25. Tang Q, Sang N, Liu H: Contrast-dependent surround suppression models for contour detection. *Pattern Recogn* 60:51–61, 2016
26. Xie J, Zhang L, You J, Shiu S: Effective texture classification by texton encoding induced statistical features. *Pattern Recogn* 48(2):447–457, 2015
27. Shotton J, Winn J, Rother C, Criminisi A: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int J Comput Vis* 81: 2–23, 2009
28. Li X, Williams S, Bottema MJ: Constructing and applying higher order textons: Estimating breast cancer risk. *Pattern Recogn* 47(3): 1375–1382, 2014
29. Petroudi S, Brady M: Breast density characterization using texton distributions. In: *Conf Proc IEEE Eng Med Biol Soc*, 2011, pp 5004–5007
30. Zhang L, Fisher M, Wang W: Retinal vessel segmentation using multi-scale textons derived from keypoints. *Comput Med Imaging Graph* 45:47–56, 2015
31. Wang J, Yang J, Yu K, Lv F, Huang T, Gong Y: Locality-constrained linear coding for image classification. In: *Proc IEEE Conf Comput Vis Pattern Recognit*, 2010, pp 3360–3367
32. Zhang P, Wee C, Nieghammer M, Shen D, Yap P: Large deformation image classification using generalized locality-constrained linear coding. In: *Proc Int Conf Med Image Comput Comput Assist Interv (MICCAI)*, 2013, pp 292–299
33. Yang J, Yu K, Gong Y, Huang T: Linear spatial pyramid matching using sparse coding for image classification. In: *Proc IEEE Conf Comput Vis Pattern Recognit*, 2009, pp 1794–1801
34. Lazebnik S, Schmid C, Ponce J: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: *Proc. IEEE Conf. Comput. Vis. Pattern Recogn* 2: 2169–2178, 2006.
35. He K, Zhang X, Ren S, Sun J: Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 37(9):1904–1916, 2015
36. Meij E, Trieschnigg D, Rijke MD, Kraaij W: Conceptual language models for domain-specific retrieval. *Inform Process Manag* 46(4): 448–469, 2010
37. Tommasi T, Caputo B, Welter P, Guld MO, Deserno TM: Overview of the CLEF 2009 medical image annotation track. In: *CLEF 2009 Workshop. Lect Notes Comput Sci* 6242(85–93), 2010
38. Lehmann TM, Schubert H, Keysers D, Kohnen M, Wein BB: The IRMA code for unique classification of medical images. In: *Proc SPIE-Med Imaging*, 5033:440–451, 2003
39. Rahman MM, Desai BC, Bhattacharya P: Medical image retrieval with probabilistic multiclass support vector machine classifiers and adaptive similarity fusion. *Comput Med Imaging Graph* 32(2):95–108, 2008
40. Jiang W, Er G, Dai Q, Gu J: Similarity based online feature selection in content-based image retrieval. *IEEE Trans Image Process* 15(3):702–712, 2006
41. Avni U, Greenspan H, Konen E, Sharon M, Goldberger J: X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. *IEEE Trans Med Imaging* 30(3):733–746, 2011
42. Liu X, Tizhoosh HR, Kofman J: Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform. In: *Proc IEEE Int Joint Conf Neural Networks*, 2016, pp 2872–2878
43. Unay D, Soldea O, Ozogür-Akyuz S, Cetin M, Ercil A: Medical image retrieval and automatic annotation: VPA-SABANCI at imageCLEF 2009. *Working Notes for CLEF 2009 Workshop*, 2009.
44. de Oliveira JEE, Lopes APB, Camara-Chavez G, de Araújo A, Deserno TM, Mammo SVD: A content-based image retrieval system using a reference database of mammographies. In: *Proc IEEE Int Symp Biomed Comput-based Med Syst*, 2009, pp 1–4
45. Crammer K, Singer Y: On the algorithmic implementation of multiclass kernel-based vector machines. *J Mach Learn Res* 2: 265–292, 2001
46. Pinho E, Godinho T, Valente F, Costa C: A multimodal search engine for medical imaging studies. *J Digit Imaging* 30(1):39–48, 2017