Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Scalable histopathological image analysis via supervised hashing with multiple features

Menglin Jiang[a], Shaoting Zhang[b,*], Junzhou Huang[c], Lin Yang[d], Dimitris N. Metaxas[a]

[a] Department of Computer Science, Rutgers University, Piscataway, NJ 08854, USA
[b] Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[c] Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019, USA
[d] Department of Biomedical Engineering, University of Florida, Gainesville, FL 32611, USA

ARTICLE INFO

ABSTRACT

Histopathology is crucial to diagnosis of cancer, yet its interpretation is tedious and challenging. To facilitate this procedure, content-based image retrieval methods have been developed as case-based reasoning tools. Especially, with the rapid growth of digital histopathology, hashing-based retrieval approaches are gaining popularity due to their exceptional efficiency and scalability. Nevertheless, few hashing-based histopathological image analysis methods perform feature fusion, despite the fact that it is a common practice to improve image retrieval performance. In response, we exploit joint kernel-based supervised hashing (JKSH) to integrate complementary features in a hashing framework. Specifically, hashing functions are designed based on linearly combined kernel functions associated with individual features. Supervised information is incorporated to bridge the semantic gap between low-level features and high-level diagnosis. An alternating optimization method is utilized to learn the kernel combination and hashing functions. The obtained hashing functions compress multiple high-dimensional features into tens of binary bits, enabling fast retrieval from a large database. Our approach is extensively validated on 3121 breast-tissue histopathological images by distinguishing between actionable and benign cases. It achieves 88.1% retrieval precision and 91.3% classification accuracy within 16.5 ms query time, comparing favorably with traditional methods.

1. Introduction

For years, histopathology has played a key role in the early diagnosis of breast cancer, which is the second leading cause of cancer-related death among women (American Cancer Society, 2013). Unfortunately, examination of histopathological images is time-consuming and error-prone due to their large size, inter- and intra-observer variability among pathologists, and several other factors (Gurcan et al., 2009; Al-Janabi et al., 2012; Veta et al., 2014). To facilitate this procedure, many content-based image retrieval (CBIR) methods have been proposed as computer-aided diagnosis (CAD) tools (Zheng et al., 2003; Akakin and Gurcan, 2012; Caicedo et al., 2011; Zhang et al., 2015b, 2015c, 2014, 2015; Foran et al., 2011). These methods compare a query image with previously diagnosed cases stored in a database, and return the most similar cases along with the likelihood of abnormality of the query. Pathologists could use the retrieved cases to better diagnose the query case. In addition, CBIR methods also contribute to digital slide archiving, pathologist training, and various other applications (Foran et al., 2011). Especially, the recent development of digital histopathology has been a catalyst for the increase of hashing-based retrieval methods (Zhang et al., 2015b, 2015c, 2014, 2015), which have remarkable computational efficiency and scalability.

In the image retrieval community, it is a common practice to employ multiple features for performance improvement. Nevertheless, except Zhang et al. (2014) and Liu et al. (2014a), few hashing-based approaches in medical image analysis integrate multiple features. Both Zhang et al. (2014) and Liu et al. (2014a) adopted affinity aggregation. They first calculated the aggregated affinity matrix by taking the average of the individual affinity matrices, which were computed using single features. Then, they applied traditional hashing methods, which are designed for one single matrix, to the aggregated affinity matrix. However, affinity aggregation is not suitable for those features that need different kernel functions during the hashing process. Besides, it introduces extra parameters, such as weights of all the matrices, which require to be elaborately

* Corresponding author.
E-mail addresses: szhang16@uncc.edu, rutgers.shaoting@gmail.com (S. Zhang).
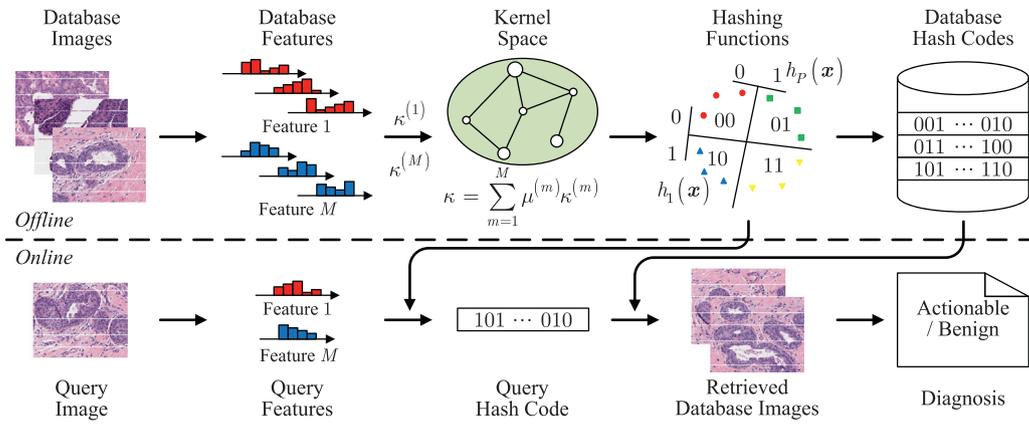
Fig. 1. Overview of the proposed approach.

tuned. Other widely used feature fusion methods in medical image retrieval include feature concatenation (Zheng et al., 2003; Akakin and Gurcan, 2012; Alto et al., 2005; Zheng et al., 2006) and result-level fusion (Zhang et al., 2015, 2015a, Wei et al., 2011). The former approach simply concatenates several features to form a new one. Similar to affinity aggregation, it is not appropriate for features of different formats even with feature normalization, because various features may need different similarity measures during subsequent feature matching. The latter method integrates similarity search results derived using individual features. Obviously, this approach compromises the computational efficiency, since its processing time will be at least the sum of time required by each feature.

To overcome the aforementioned drawbacks, we employ a joint kernel-based supervised hashing (JKSH) approach (Liu et al., 2014b) to incorporate feature fusion into supervised hashing framework, and apply it to scalable histopathological image analysis. The overview of our approach is shown in Fig. 1. Specifically, a joint kernel function is defined as a linear combination of the kernels for individual features, and a series of hashing functions are constructed based on this kernel. During offline process, a supervised optimization algorithm is utilized to learn the kernel weights and hashing functions according to the diagnostic information of database images, which bridges the semantic gap between low-level features and high-level diagnosis. With the learned hashing functions, high-dimensional features extracted from database images are compressed into compact hash codes and stored in hash tables. During online process, the query hash code is computed using the same hashing functions and searched from the hash tables to obtain the most similar database images. The query image is classified (as actionable or benign) according to a weighted majority vote of its retrieved images (Jiang et al., 2015b).

Our approach has many advantages over current methods. First, it improves the retrieval precision by adopting multiple kernel functions as similarity measures for various features, rather than employ the same kernel. Second, parameter tuning is a critical issue to affinity aggregation. We address this problem by automatically learning all the important parameters. The learning algorithm also bridges the semantic gap between low-level features and high-level diagnosis. Finally, utilizing the kernel representation, our approach compresses multiple features into hash codes with little computational overhead than using a single feature. This journal paper has made considerable progress compared with the preliminary results published in a conference paper (Jiang et al., 2015a). First of all, new image feature and kernel function are adopted to achieve even better performance. Second, all the methods are evaluated on a larger dataset containing more challenging

cases, which makes the experimental results more convincing and better demonstrates the scalability of our approach. Finally, this paper is substantially extended to provide more details about our method as well as the techniques at the base of it.

The rest of this paper is organized as follows. Section 2 reviews some relevant work. Section 3 describes the proposed approach. Section 4 presents the experimental results. Finally, Section 5 draws a conclusion.

2. Related work

In this section, we first review the development of hashing techniques, which forms the foundation of our approach. Then we summarize existing CBIR-based CAD methods, with an emphasis on those employing scalable image retrieval.

2.1. Hashing methods

During the past few years, hashing has become a popular approach to scalable image retrieval. Early work on this topic can be traced back to locality-sensitive hashing (LSH) (Gionis et al., 1999). LSH solves the "curse of dimensionality" by mapping high-dimensional feature vectors to compact binary hash codes. Meanwhile, it ensures that similar features under certain metric are compressed into similar hash codes in Hamming space with high probabilities. While showing a promising direction for large-scale similarity search, LSH has a few limitations, e.g. its projection vectors are randomly generated. In response, a tremendous effort has been made to improve LSH. Representative approaches include, but are not limited to, kernel-based hashing (He et al., 2010; Kulis and Grauman, 2012), unsupervised hashing (Weiss et al., 2008; Liu et al., 2011; Li et al., 2014), (semi-)supervised hashing (Wang et al., 2012; Liu et al., 2012), and multiple-feature hashing (Gönen and Alpaydin, 2011; Zhang et al., 2011; Hassan et al., 2012; Liu et al., 2014b).

The core of hashing is to preserve data similarity in Hamming space. However, many hashing methods, such as LSH, define data similarity in the original feature space, where it might be difficult to find the pattern of data. To solve this problem, kernel functions are introduced to hashing (He et al., 2010; Kulis and Grauman, 2012). For instance, He et al. (2010) proposed an optimal kernel hashing (OKH), which could use any kernel to define hashing functions. As a result, OKH is able to handle any data format (e.g. graph) and utilize various similarity measures (e.g. label consistency).

Another drawback of traditional hashing methods is that they cannot handle the so-called "semantic gap", which means images

with similar low-level features may have different high-level semantic meanings. Supervised information is leveraged to address this issue (Wang et al., 2012; Liu et al., 2012). Perhaps the most prominent one of existing supervised hashing methods is kernel-based supervised hashing (KSH) (Liu et al., 2012). Specifically, when learning the hashing functions, KSH simultaneously minimizes the Hamming distances between semantically similar images and maximizes the distances between dissimilar images.

In addition, researchers improve the performance of hashing methods via the adoption of multiple features, which has been intensively explored in non-hashing image retrieval approaches. For example, Zhang et al. (2011) proposed a composite hashing with multiple information sources (CHMIS), where the hash code of an image is a convex combination of the codes generated by all the hashing functions for individual features. Hassan et al. (2012) incorporated kernels into distance-based hashing (DBH) and presented KernelDBH to embed multiple features into hash codes. KernelDBH defines the hashing functions based on a combination of multiple kernels, and selects the optimal combination using a genetic algorithm.

2.2. CBIR-based CAD approaches

The past decade has witnessed many CBIR-based CAD approaches (Müller et al., 2004; Akgül et al., 2011; Kumar et al., 2013). In the context of histopathology, to name a few, Zheng et al. (2003) developed a CBIR system to retrieve images from an online pathological image database using four features, which are fused via feature concatenation. Akakin and Gurcan (2012) designed a multi-tiered image retrieval system for microscopic images, through which the query image is first classified into a certain disease type and then is searched from database images of the same category. Similarly, Caicedo et al. (2011) combined image annotation and retrieval together. They first annotated the query image with high-level semantic concepts according to its low-level features and then searched it from database images with similar annotations. Besides the above approaches designed for histopathological images, CBIR-based CAD also finds application in many other medical modalities, such as mammographic images (Alto et al., 2005; Zheng et al., 2006; Wei et al., 2011), X-ray images (Avni et al., 2011), endomicroscopic videos (André et al., 2012), and CT images (Yang et al., 2012). These approaches have shown great value of CBIR in medical image analysis. Nevertheless, a vast majority of them fall short of scalability and cannot handle the enormous amount of data in practice.

Recognizing the importance of scalability to medical imaging applications (Langs et al., 2012), researchers have started to develop scalable approaches to medical image retrieval and analysis. For example, Zhang et al. (2015b) employed KSH for efficient retrieval of histopathological images. They also integrated multiple features in anchor graph hashing (AGH) (Liu et al., 2011) through affinity aggregation (Zhang et al., 2014). Subsequently, Zhang et al. (2015) adopted another feature fusion strategy, namely result-level fusion. In particular, KSH and exhaustive search, which employ two different features, are independently conducted for each query image, and their results are consolidated through a query specific re-ranking algorithm (Zhang et al., 2015a). Similar to Zhang et al. (2014), Liu et al. (2014a) applied AGH and affinity aggregation to mammographic image retrieval. Jiang et al. (2015b) exploited vocabulary tree (Nistér and Stewénius, 2006) in large-scale retrieval and analysis of mammographic images. As an alternative to hashing for scalable CBIR, the vocabulary tree framework could retrieve query images from millions of database images in real time (Nistér and Stewénius, 2006). Foran et al. (2011) developed a practical retrieval system for pathological images, which leverages high-performance comput-ers and grid technology. Among the above methods, the work by Zhang et al. (2015b) based on KSH can be regarded as the baseline of our approach. In addition, affinity aggregation (Zhang et al., 2014; Liu et al., 2014a), feature concatenation (Zheng et al., 2003; Akakin and Gurcan, 2012; Alto et al., 2005; Zheng et al., 2006), and result-level fusion (Zhang et al., 2015, 2015a; Wei et al., 2011) are all compared with our approach in the experiments.

3. Methods

In this section, we first formulate the multiple-feature hashing problem as a linear combination of individual kernels, and then explain how to simultaneously learn the kernel weights and hashing functions. After obtaining the hashing functions, features extracted from database histopathological images can be mapped to compact hash codes and stored in hash tables. Given a query image, its hash code is calculated using the same hashing functions and searched from the hash tables to find similar database images, which then vote to determine its diagnosis (Jiang et al., 2015b).

3.1. Joint kernel-based hashing

Suppose we extract M features from N histopathological images. Denote $\boldsymbol{x}_n^{(m)} \in \mathbb{R}^{d^{(m)}}$ as the m-th feature of the n-th image, which is a $d^{(m)}$-dimensional column vector. Then $\boldsymbol{x}_n = \left[\left(\boldsymbol{x}_n^{(1)} \right)^T, \cdots, \left(\boldsymbol{x}_n^{(M)} \right)^T \right]^T \in \mathbb{R}^d$ is the concatenation of all features extracted from the n-th image, where $d = \sum_{m=1}^M d^{(m)}$. A hashing method aims at finding P hashing functions $\{h_1, \cdots, h_P\}$, where P is the desired number of hash bits. Each hashing function, $h_p : \mathbb{R}^d \mapsto \{-1, 1\}$, maps a concatenated feature vector into a binary bit. The nth image is represented as $\boldsymbol{y}_n = [h_1(\boldsymbol{x}_n), \cdots, h_P(\boldsymbol{x}_n)]^T$. Note that here we use "-1" instead of "0"; in the implementation we use "0". In addition, all the vectors in this paper are column vectors.

When designing hashing functions, a classic idea is to preserve "local sensitivity", i.e., similar feature vectors are compressed into similar hash codes (Gionis et al., 1999). Unfortunately, sometimes it is difficult to distinguish between the original features. To solve this problem, kernel functions are introduced to operate the data in an *implicit* higher-dimensional feature space (He et al., 2010; Kulis and Grauman, 2012). Given M features, we can choose M kernel functions $\{\kappa^{(1)}, \cdots, \kappa^{(M)}\}$, where each kernel $\kappa^{(m)}$ is associated with an implicit feature mapping function $\varphi^{(m)}$, i.e. $\kappa^{(m)}\left(\boldsymbol{x}_i^{(m)}, \boldsymbol{x}_j^{(m)} \right) = \varphi^{(m)}\left(\boldsymbol{x}_i^{(m)} \right)^T \varphi^{(m)}\left(\boldsymbol{x}_j^{(m)} \right)$. Without ever computing the mapped features $\varphi^{(m)}\left(\boldsymbol{x}_i^{(m)} \right)$ and $\varphi^{(m)}\left(\boldsymbol{x}_j^{(m)} \right)$, $\kappa^{(m)}$ directly calculates their inner product. Such "kernel trick" improves computational efficiency dramatically. Following Gönen and Alpaydin (2011) and Caicedo et al. (2011), the joint mapping function φ and corresponding kernel κ are defined as:

$$\varphi(\boldsymbol{x}_n) = \left[\sqrt{\mu^{(1)}} \varphi^{(1)}\left(\boldsymbol{x}_n^{(1)} \right)^T, \cdots, \sqrt{\mu^{(M)}} \varphi^{(M)}\left(\boldsymbol{x}_n^{(M)} \right)^T \right]^T, \quad (1)$$

$$\kappa\left(\boldsymbol{x}_i, \boldsymbol{x}_j \right) = \varphi(\boldsymbol{x}_i)^T \varphi\left(\boldsymbol{x}_j \right) = \sum_{m=1}^M \mu^{(m)} \kappa^{(m)}\left(\boldsymbol{x}_i^{(m)}, \boldsymbol{x}_j^{(m)} \right), \quad (2)$$

where $\mu^{(m)}$ is the weight for the m-th feature. Later we will show how to automatically learn the weight vector $\boldsymbol{\mu} = \left[\mu^{(1)}, \cdots, \mu^{(M)} \right]^T$. Eq. (2) demonstrates that $\kappa = \sum_{m=1}^M \mu^{(m)} \kappa^{(m)}$ is actually a linear combination of individual kernels for each feature.

In practice, we don't need to explicitly calculate $\varphi^{(m)}$ or φ. Instead, only $\kappa^{(m)}$ and κ should be computed. To further improve computational efficiency, $R(R \ll N)$ landmark points, denoted as $\{\boldsymbol{z}_1, \cdots, \boldsymbol{z}_R\}$, are randomly selected from all the database feature
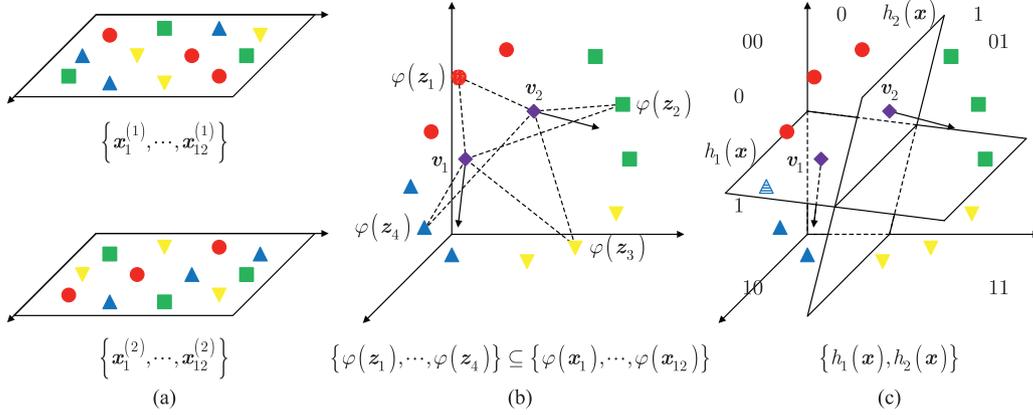
Fig. 2. Illustration of joint kernel-based hashing. (a) Two features are extracted from twelve histopathological images. The mth feature of the nth image is denoted as $\boldsymbol{x}_n^{(m)}$, $m = 1, 2, n = 1, \cdots, 12$. These images can be divided into four groups, which are marked in red, green, blue and yellow, respectively. However, it's difficult to distinguish between the four groups in the original feature space. (b) All the concatenated feature vectors, $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{12}$, are mapped to a joint kernel space, and their projections are represented as $\varphi(\boldsymbol{x}_1), \cdots, \varphi(\boldsymbol{x}_{12})$. Now, it's much easier to find the pattern of features. Four landmark points, $\boldsymbol{z}_1, \cdots, \boldsymbol{z}_4$, are randomly selected from the concatenated feature vectors. Two hyperplanes' normal vectors, which are denoted as $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, are calculated as linear combinations of the projections of landmarks. (c) Two hashing functions h_1 and h_2 are defined based on $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. h_1 and h_2 compress the feature vectors into 2-bit hash codes with similar features mapped to the same code. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

vectors $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N\}$. Define $K_{R \times N}^{(m)} = \left[\kappa^{(m)}\left(\boldsymbol{z}_r^{(m)}, \boldsymbol{x}_n^{(m)}\right)\right]_{R \times N}$ as the kernel matrix for the m-th feature between R landmarks and N database images, and $K_{R \times R}^{(m)} = \left[\kappa^{(m)}\left(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_j^{(m)}\right)\right]_{R \times R}$ as the kernel matrix for the m-th feature between R landmarks. Then, the kernel matrix between landmarks and database images, $K_{R \times N}$, and the kernel matrix between all the landmarks, $K_{R \times R}$, can be easily obtained by calculating the linear combination of $K_{R \times N}^{(m)}$ and $K_{R \times R}^{(m)}$:

$$K_{R \times N} = [\kappa(\boldsymbol{z}_r, \boldsymbol{x}_n)]_{R \times N} = \sum_{m=1}^{M} \mu^{(m)} K_{R \times N}^{(m)}, \qquad (3)$$

$$K_{R \times R} = [\kappa(\boldsymbol{z}_i, \boldsymbol{z}_j)]_{R \times R} = \sum_{m=1}^{M} \mu^{(m)} K_{R \times R}^{(m)}. \qquad (4)$$

For the pth hashing function h_p $(p = 1, \cdots, P)$, its hyperplane's normal vector $\boldsymbol{v}_p$ is defined as a linear combination of the projections of landmarks in the kernel space:

$$\boldsymbol{v}_p = \sum_{r=1}^{R} W(r, p) \varphi(\boldsymbol{z}_r), \qquad (5)$$

where W is a $R \times P$-dimensional matrix, its element $W(r, p)$ denotes the weight of $\varphi(\boldsymbol{z}_r)$ for $\boldsymbol{v}_p$. The hyperplane is represented as:

$$\boldsymbol{v}_p^T \bar{\boldsymbol{x}} + b_p = 0, \qquad (6)$$

where $\bar{\boldsymbol{x}}$ denotes a point in the kernel space, b_p is a parameter related to the distance from the space origin to the hyperplane. $h_p(\boldsymbol{x}_n)$ reflects on which side of the p-th hyperplane $\varphi(\boldsymbol{x}_n)$ is located:

$$h_p(\boldsymbol{x}_n) = \text{sgn}\left(\boldsymbol{v}_p^T \varphi(\boldsymbol{x}_n) + b_p\right). \qquad (7)$$

Define $\boldsymbol{b} = [b_1, \cdots, b_P]^T$ as the vector containing all the hyperplanes' distance parameters. W and $\boldsymbol{b}$ determine the hashing functions, and they will be automatically learned along with $\boldsymbol{\mu}$ using supervised information. The construction of hashing functions is demonstrated in Fig. 2.

Let $K_{R \times N}(:, n)$ represent the nth column of K. Utilizing the definitions of $\boldsymbol{y}_n$, h_p (Eq. (7)), $\boldsymbol{v}_p$ (Eq. (5)), and κ (Eq. (2)) successively, we can represent the hash code of the nth image in a kernel form:

$$\boldsymbol{y}_n = \text{sgn}\left(W^T K_{R \times N}(:, n) + \boldsymbol{b}\right). \qquad (8)$$

3.2. Supervised optimization

In the image retrieval field, "semantic gap", which refers to the difference between low-level features and high-level concepts, is a long-standing problem (Caicedo et al., 2011; André et al., 2012; Yang et al., 2012). Supervised methods, such as KSH (Liu et al., 2012), offer a promise to address this issue. Instead of only considering low-level features, supervised hashing approaches also take high-level training information into consideration and map semantically similar images to similar hash codes. To this end, we incorporate diagnostic information into affinity matrix S. $S(i, j)$, which represents the similarity score between the ith and the jth images, is defined as:

$$S(i, j) = \begin{cases} \exp\left(-\dfrac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{\sigma^2}\right), & \text{if } \boldsymbol{x}_i \in \mathcal{N}_s^k(\boldsymbol{x}_j) \text{ or } \boldsymbol{x}_j \in \mathcal{N}_s^k(\boldsymbol{x}_i) \\ 0, & \text{otherwise} \end{cases}, \qquad (9)$$

where σ is a scaling parameter estimated from the data, and $\mathcal{N}_s^k(\boldsymbol{x})$ represents the set of k-nearest neighbors of $\boldsymbol{x}$ with the same semantic label. Note that S is a sparse matrix, i.e., most of its elements are 0. An example of S is provided in Fig. 3.

Following Weiss et al. (2008) and He et al. (2010), the objective function of JKSH is formulated as:

$$\min_{W, \boldsymbol{b}, \boldsymbol{\mu}} \frac{1}{2} \sum_{i,j=1}^{N} S(i, j) \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2 + \lambda \|V\|_F^2 = \text{Tr}\left(YLY^T\right) + \lambda \|V\|_F^2,$$

$$\text{s.t.} \sum_{n=1}^{N} \boldsymbol{y}_n = \boldsymbol{0}, \ \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}_n \boldsymbol{y}_n^T = I, \ \boldsymbol{1}^T \boldsymbol{\mu} = 1, \ \boldsymbol{\mu} \succcurlyeq \boldsymbol{0}. \qquad (10)$$

Here $V = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_P]$ contains all the hyperplanes' normal vectors, $Y = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N]$ includes all the database hash codes, $L = \text{diag}(S\boldsymbol{1}) - S$ is the graph Laplacian matrix, λ is the parameter that controls how smooth the hashing functions are. In this objective function, $\sum_{i,j=1}^{N} S(i, j) \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2$ guarantees that histopathological images with the same label and similar features are compressed into similar hash codes, $\|V\|_F^2$ is a regularized term that promotes the smoothness of hashing functions, the constraints $\sum_{n=1}^{N} \boldsymbol{y}_n = \boldsymbol{0}$ and $(1/N) \sum_{n=1}^{N} \boldsymbol{y}_n \boldsymbol{y}_n^T = I$ ensure that the generated hash codes are

Fig. 3. A toy example of affinity matrix. (a) Twelve histopathological images are divided into four groups, which are circled in red, green, blue, and yellow, respectively. Images in the same group have identical diagnosis (actionable or benign) and similar feature vectors (2-nearest neighbors). (b) Affinity matrix of these histopathological images. Only images with the same diagnosis and similar features have positive similarity scores. All the other elements in the matrix are 0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

balanced and uncorrelated, $\mathbf{1}^T \boldsymbol{\mu} = 1$ and $\boldsymbol{\mu} \succeq \mathbf{0}$ ensure that $\boldsymbol{\mu}$ is a valid weight vector.

To solve this NP-hard problem, we employ spectral relaxation (Weiss et al., 2008), which ignores the discrete constraint for $\mathbf{y}_n$ in Eq. (8) and allows:

$$\mathbf{y}_n = W^T K_{R \times N}(:, n) + \boldsymbol{b} \in \mathbb{R}^P. \tag{11}$$

The above problem, when either $(W, \boldsymbol{b})$ or $\boldsymbol{\mu}$ is fixed, is convex with respect to the other. Therefore we perform an alternating optimization algorithm, which mainly consists of the following two steps. Derivation of the algorithm is provided in Appendix A.

Step 1. Optimize $(W, \boldsymbol{b})$ **for given** $\boldsymbol{\mu}$. Similar to He et al. (2010), we can find the optimal W by solving the following problem using eigen-decomposition:

$$\min_{W} \operatorname{Tr}\left(W^T C W\right), \ \text{s.t.} \ W^T G W = I, \tag{12}$$

where

$$C = K_{R \times N} L K_{R \times N}^T + \lambda K_{R \times R}, \tag{13}$$

$$G = \frac{1}{N} K_{R \times N}\left(I - \frac{1}{N}\mathbf{1}\mathbf{1}^T\right)K_{R \times N}^T. \tag{14}$$

$\boldsymbol{b}$ has a close-form solution depending on W:

$$\boldsymbol{b} = -\frac{1}{N}W^T K_{R \times N}\mathbf{1}. \tag{15}$$

Step 2. Optimize $\boldsymbol{\mu}$ **for given** $(W, \boldsymbol{b})$. The optimal $\boldsymbol{\mu}$ can be obtained by solving the following quadratic programming problem:

$$\min_{\boldsymbol{\mu}} \frac{1}{2}\boldsymbol{\mu}^T E \boldsymbol{\mu} + \boldsymbol{f}^T \boldsymbol{\mu}, \ \text{s.t.} \ \mathbf{1}^T \boldsymbol{\mu} = 1, \ \boldsymbol{\mu} \succeq \mathbf{0}. \tag{16}$$

Here, E is a $M \times M$-dimensional matrix with its (i, j)th element defined as

$$E(i, j) = 2 \operatorname{Tr}\left(W^T K_{R \times N}^{(i)} L \left(K_{R \times N}^{(j)}\right)^T W\right). \tag{17}$$

$\boldsymbol{f} = \left[f^{(1)}, \cdots, f^{(M)}\right]^T$ is a M-dimensional vector and its mth element is

$$f^{(m)} = \lambda \operatorname{Tr}\left(W^T K_{R \times R}^{(m)} W\right). \tag{18}$$

In summary, the optimization approach works as follows. First, it initializes $\boldsymbol{\mu} = [1/M, \cdots, 1/M]^T$. Then, it iteratively updates $(W, \boldsymbol{b})$ according to step 1 and updates $\boldsymbol{\mu}$ according to step 2 until they converge. In practice, our method usually finds the optimal $(W, \boldsymbol{b})$ and $\boldsymbol{\mu}$ within a few iterations.

4. Experiments

This section validates the JKSH approach on retrieval and classification of breast cancer histopathological images. First, the employed dataset is introduced. Then, the image features and kernel functions adopted in our approach as well as the baseline methods are described. Finally, experimental results are presented and analyzed.

4.1. Dataset

Our experiments are carried out on the breast-tissue microscopic image dataset built in Dundar et al. (2011) and Zhang et al. (2015b). 116 patient cases are collected on a retrospective basis from the IU Health Pathology Lab (IUHPL) according to the approved Institutional Review Board protocol. Based on the diagnosis of nine board-certified pathologists, 57 cases are identified as actionable (atypical ductal hyperplasia, ADH, and ductal carcinoma in situ, DCIS), and the remaining 59 cases are determined as benign (usual ductal hyperplasia, UDH). The H&E-stained tissue specimens are scanned using a ScanScope digitizer (Aperio, Vista, CA, USA) at $40 \times$ magnification. Once the whole-slide images are obtained, a research associate and a pathologist manually identify regions of interest (ROIs) on each slide, which generally show proliferation of cells. 347 and 579 ROIs are extracted from the actionable and benign cases, respectively. Each ROI reaches thousands of pixels along each direction, e.g., 5K × 7K pixels. 3121 images are sampled from these ROIs. 1013 of them are actionable and the other 2108 are benign. Each image has about 2.25M pixels.

Four-fold cross-validation is performed to achieve reliable results—that is, both actionable and benign cases are divided into four parts, and the proposed approach is evaluated four times. During each time, images in one part are used as queries, and the other images form a database. Note that the query and database images are selected from different cases to avoid positive bias. Be-

low we report the average of the evaluation results during four tests.

4.2. Implementation details

The proposed approach is evaluated twice based on three features, namely SIFT (Lowe, 2004), HOG (Dalal and Triggs, 2005), and GIST (Oliva and Torralba, 2001). SIFT has been widely used in medical image retrieval and analysis (Caicedo et al., 2009; Zhang et al., 2014, 2015, 2015b; Caicedo et al., 2011; Liu et al., 2014a; Jiang et al., 2015b; Avni et al., 2011; André et al., 2012; Tamaki et al., 2013), owing to its excellent robustness and discriminative power. Briefly speaking, SIFT features are extracted from a histopathological image as follows. First, scale-invariant keypoints are detected by finding local extrema in the difference-of-Gaussian (DoG) space. Then, for each keypoint, a surrounding region of size 16×16 is identified in a certain scale, and a 128-dimensional gradient orientation histogram is calculated as feature vector. All the SIFT features are quantized using bag-of-words (BoW) method (Sivic and Zisserman, 2003), which is the de facto standard of compressing local features like SIFT (Caicedo et al., 2009; Cruz-Roa et al., 2012; Zhang et al., 2014, 2015, 2015b; Caicedo et al., 2011; Liu et al., 2014a; Jiang et al., 2015b; Avni et al., 2011; André et al., 2012; Yang et al., 2012; Tamaki et al., 2013). In particular, our approach utilizes a SIFT vocabulary of size 1000, i.e. each image is represented as a 1000-dimensional BoW histogram. Histogram intersection (HI) kernel is chosen for SIFT feature, which is a meaningful similarity measure for histograms (Caicedo et al., 2009, 2011; Avni et al., 2011; Tamaki et al., 2013). HOG has been successfully applied to histopathological image retrieval and classification (Zhang et al., 2014). In our paper, HOG features are calculated around the SIFT keypoints. They are also quantized using BoW method and represented as 1000-dimensional histograms. Radial basis function (RBF) kernel is selected for HOG, which is very popular in kernelized learning methods for medical image analysis (Caicedo et al., 2009, 2011; Avni et al., 2011; Tamaki et al., 2013). GIST has demonstrated good performance in medical image retrieval (Liu et al., 2014a). In our paper, the traditional 512-dimensional GIST feature (Oliva and Torralba, 2001) is employed. As for kernel function, GIST also utilizes the RBF kernel. In our conference paper (Jiang et al., 2015a), the proposed approach adopts SIFT and HOG features. In this paper, we employ SIFT and GIST for even better performance. The two configurations are hereafter denoted as "Ours1" and "Ours2". In both Ours1 and Ours2, R (number of landmark points) is set to 200, λ (weight of the smoothness term in the energy function) is set to 0.1, and σ (the scaling parameter used in affinity matrix) is set to 0.01.

Eight baseline methods are implemented for comparison. The first three methods investigate the performance of traditional KSH (Liu et al., 2012) on individual features. They adopt SIFT feature and HI kernel, HOG feature and RBF kernel, GIST feature and RBF kernel, respectively. The fourth to sixth methods carry out widely used feature fusion strategies in medical image retrieval, which are affinity aggregation (Zhang et al., 2014; Liu et al., 2014a), feature concatenation (Zheng et al., 2003; Akakin and Gurcan, 2012; Alto et al., 2005; Zheng et al., 2006), and result-level fusion (Zhang et al., 2015, 2015a; Wei et al., 2011). In particular, affinity aggregation first computes the affinity matrices using SIFT and GIST features, then calculates the average of these two matrices, and finally applies KSH to the aggregated matrix. Note that affinity aggregation could use only one kernel function during the hashing process. Here we choose RBF kernel. Feature concatenation applies KSH to the concatenation of SIFT and GIST features. RBF kernel is used for the concatenated feature. Result-level fusion integrates the retrieval results of the first and third comparison methods via a re-ranking algorithm (Zhang et al., 2015a). The seventh and eighth comparison methods are based on two non-hashing algorithms, namely kNN and SVM, which are very popular in histopathological image analysis. kNN (Tabesh et al., 2007; Yang et al., 2009) calculates the distance between the query feature and each database feature, and returns the database images with smallest distance values. The query image is assigned to the class that is most common among the retrieved database images. SVM with a nonlinear kernel is often employed in histopathological image analysis (Caicedo et al., 2009; Huang and Lai, 2010; Tuzel et al., 2007; Nguyen et al., 2010), owing to its efficiency and the ability to handle linearly inseparable data. Note that SVM could only generate a classification result. Both kNN and SVM use concatenated SIFT and GIST as image feature. The comparison methods will hereafter be referred to as "SIFT", "HOG", "GIST", "Affinity", "Concatenation", "ResultFusion", "kNN", and "SVM".

All the methods are implemented in MATLAB and evaluated on a computer with Intel Core i7 processor (6M cache, 2.40 GHz), 16GB memory, and Windows 7 operating system.

4.3. Results

We first evaluate the *retrieval precision* of all the methods. Precision is the most widely used evaluation metric in image retrieval. In our context, it is defined as the percentage of retrieved histopathological images which have the same pathology (actionable/benign) as that of the query image. Overall the precision scores remain stable as the number of retrieved images increases from 1 to 20. We take the top 20 retrieved images into consideration in the following evaluation. To demonstrate parameter sensitivity, all the hashing methods use a series of hash code lengths, ranging from 8 to 64 bits. The precision scores are summarized in Fig. 4(a). Note that SVM is not included since it cannot generate retrieval result. Four retrieval examples obtained by Ours2 are provided in Fig. 5 for qualitative evaluation.

The above results lead to several conclusions. First, the hashing methods outperform kNN. The reason lies in the adoption of kernel functions and supervised optimization in these hashing methods. Kernel functions model linearly inseparable data, and supervised optimization alleviates the semantic gap issue. Second, for all the hashing methods, as the hash code length increases to 64 bits, the precision scores first increase and then remain relatively stable. This phenomenon is consistent with the conclusion that the first hashing bits are more discriminative than the subsequent bits in eigen-decomposition-based hashing methods (Wang et al., 2012). Third, among all the hashing methods, the three methods utilizing single features perform worst. This is very intuitive, as multiple features provide more information of the image than a single feature does. Fourth, as for the fusion methods, feature concatenation gains marginal improvement over SIFT, while affinity aggregation and result-level fusion obtain considerable improvement. Ours1 and Ours2 substantially surpass all these baseline methods and achieve precision scores of 87.2% and 88.1% respectively. As noted earlier, the proposed approach could select appropriate kernels for different features, whereas affinity aggregation and feature concatenation could employ only one kernel during the hashing process. In addition, our method automatically selects the optimal weights for individual features, which is not guaranteed in result-level fusion. Finally, Our2, which exploits SIFT and GIST, is superior to Ours1, which utilizes SIFT and HOG. The reason is that, compared with HOG, GIST is more complementary to SIFT. Both HOG and SIFT describe the gradient orientation of local image patches, whereas GIST characterizes the gradient information of the entire image.

Then, *classification accuracy* is measured, which refers to the percentage of query images that are correctly classified. Remember that a query image is classified as actionable or benign tissue
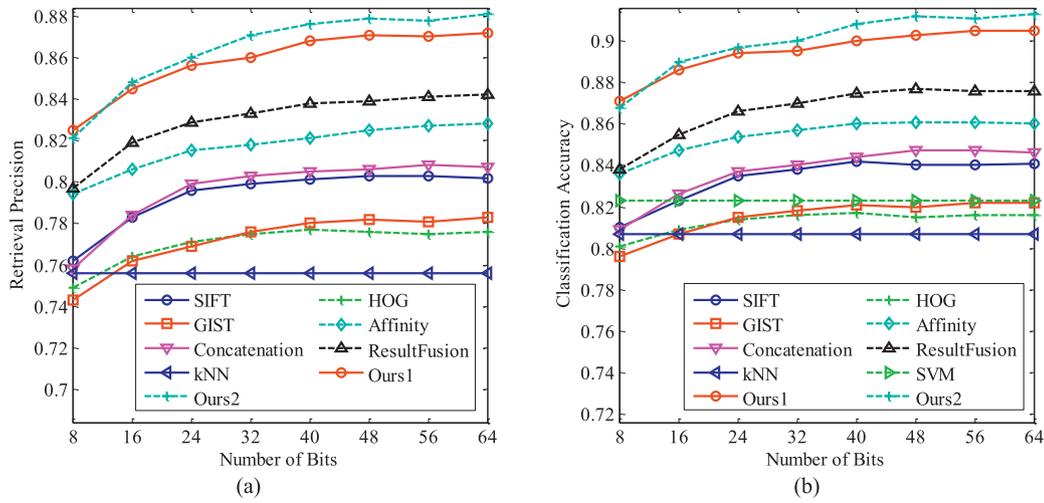
Fig. 4. (a) Retrieval precision and (b) classification accuracy of all the evaluated methods.
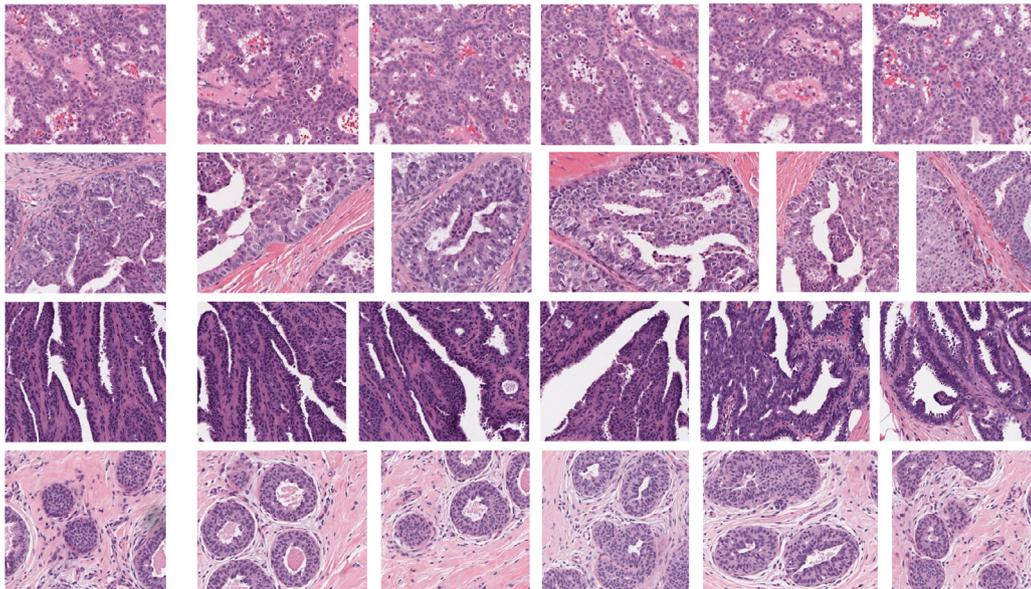
Fig. 5. Four query histopathological images (left) and their retrieved database images obtained by our approach (right). Images in the top two rows are all actionable, and images in the bottom two rows are all benign.

according to a weighted majority vote of its retrieved database images (Jiang et al., 2015b). The accuracy scores are reported in Fig. 4(b).

From Fig. 4, we can see that the classification accuracy scores exhibit similar patterns to those of the retrieval precision scores. In other words, the aforementioned conclusions regarding the precision scores also apply to the accuracy scores. This is expected because the classification results rely on the retrieval results. Second, Fig. 4 demonstrates that the accuracy scores are systematically higher than the precision scores. The reason is because irrelevant retrieved images would not cause a misclassification as long as their weighted votes are less than half of the total votes. Third, SVM is superior to kNN, since it utilizes the supervised information and kernel function. SVM also surpasses HOG and GIST, because it employs concatenated SIFT and GIST as image feature. However, SVM performs worse than other hashing-based comparison methods. Especially, its accuracy score is lower than that of Concatenation, which adopts the same feature as SVM does. This observation indicates that KSH-based methods surpass SVM in terms of classification of histopathological images. Finally, it is worth mention-

ing that our methods considerably outperform all the comparison methods. Ours1 and Ours2 achieve accuracy scores of 90.5% and 91.3%, respectively.

Finally, *query time*, which means the time needed to retrieve and classify a query image, is investigated. Here, the time cost of feature extraction and quantization is not taken into account, since it remains fixed as the database expands and therefore is not the bottleneck for large-scale image analysis. Besides, all the hashing methods are tested when they use 64-bit hash codes. The results are summarized in Table 1.

Table 1 shows that kNN is inefficient. In fact, its computational complexity is linear in the product of database size and feature dimensionality. Obviously, kNN is not suitable for applications that involve high-dimensional features and/or large databases. SVM is much faster than kNN. However, its query time increases with the feature dimensionality (Zhang et al., 2015b). The hashing-based methods exhibit outstanding computational efficiency owing to the compactness of hash codes and the adoption of "kernel trick". Moreover, their query time is sublinear in the database size when they employ hash tables to store the database hash codes. Our

Table 1

Query time (unit is millisecond) of all the evaluated methods. The hashing methods are measured when they use 64-bit hash codes.

Method	SIFT	HOG	GIST	Affinity	Concatenation	ResultFusion	kNN	SVM	Ours1	Ours2
Time	14.5	14.4	13.1	16.2	16.0	30.4	1932.2	58.3	17.4	16.5

methods, along with affinity aggregation and feature concatenation, have only a small computational overhead compared with the hashing methods using single features. As expected, the time cost for ResultFusion is the sum of those for SIFT and GIST.

5. Conclusions

In this paper, we adopt joint kernel-based supervised hashing (JKSH) for fusion of complementary features. Multiple-feature hashing is transformed to a similarity preserving problem with linearly combined kernel functions, which are associated with individual features. An alternating optimization algorithm is performed to learn both the kernel combination and hashing functions efficiently. Superior to traditional multiple-feature hashing methods, the proposed approach can adopt several kernels, therefore it is more suitable for heterogeneous features. In addition, it doesn't introduce new parameters, and has little computational overhead compared with methods employing one feature. Extensive experiments on breast cancer histopathological images demonstrate the efficacy of our approach. Future endeavors will be devoted to extending our approach with online learning so that it could efficiently update hashing functions as new images are added into the database. In addition, we plan to investigate the proposed approach on other data, such as liver whole slide image (Liang et al., 2015b, 2015a, 2016), lung-tissue microscopic images and mammographic images.

Appendix A. Proof of the optimization algorithm

Derivation in this section follows He et al. (2010). From the "balance" constraint $\sum_{n=1}^{N} \boldsymbol{y}_n = \boldsymbol{0}$ and the spectral relaxation (Eq. (11)), we have:

$$\sum_{n=1}^{N} \boldsymbol{y}_n = \sum_{n=1}^{N} \left(W^T K_{R \times N}(:, n) + \boldsymbol{b} \right) = W^T K_{R \times N} \boldsymbol{1} + N\boldsymbol{b} = \boldsymbol{0}. \quad \text{(A.1)}$$

Then, we can easily get the solution to $\boldsymbol{b}$ in Eq. (15).

Using the above solution, we can rewrite the hash code matrix as:

$$\begin{aligned} Y &= [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_N] \\ &= \left[W^T K_{R \times N}(:, 1) - \frac{1}{N} W^T K_{R \times N} \boldsymbol{1}, \cdots, \right. \\ &\quad \left. W^T K_{R \times N}(:, N) - \frac{1}{N} W^T K_{R \times N} \boldsymbol{1} \right] \\ &= W^T K_{R \times N} \left(I - \frac{1}{N} \boldsymbol{1} \boldsymbol{1}^T \right). \end{aligned} \quad \text{(A.2)}$$

The first item in the objective function is transformed to:

$$\begin{aligned} \text{Tr}\left(Y L Y^T \right) &= \text{Tr}\left(W^T K_{R \times N} \left(I - \frac{1}{N} \boldsymbol{1} \boldsymbol{1}^T \right) L \left(I - \frac{1}{N} \boldsymbol{1} \boldsymbol{1}^T \right)^T K_{R \times N}^T W \right) \\ &= \text{Tr}\left(W^T K_{R \times N} L K_{R \times N}^T W \right). \end{aligned} \quad \text{(A.3)}$$

Here we use the fact $\left(I - (1/N) \boldsymbol{1} \boldsymbol{1}^T \right) L \left(I - (1/N) \boldsymbol{1} \boldsymbol{1}^T \right)^T = L$.

According to the definition of $\boldsymbol{v}_p$ (Eq. (5)), the second item in the objective function becomes:

$$\begin{aligned} \lambda \|V\|_F^2 &= \lambda \sum_{p=1}^{P} \|\boldsymbol{v}_p\|^2 = \lambda \sum_{p=1}^{P} \boldsymbol{v}_p^T \boldsymbol{v}_p \\ &= \lambda \sum_{p=1}^{P} \left(\left(\sum_{i=1}^{R} W(i, p) \varphi(\boldsymbol{z}_i) \right)^T \left(\sum_{j=1}^{R} W(j, p) \varphi(\boldsymbol{z}_j) \right) \right) \\ &= \lambda \sum_{p=1}^{P} \sum_{i,j=1}^{R} W(i, p) W(j, p) K_{R \times R}(i, j) \\ &= \lambda \, \text{Tr}\left(W^T K_{R \times R} W \right). \end{aligned} \quad \text{(A.4)}$$

We plug in Eqs. (A.3) and (A.4), and rewrite the objective function as:

$$\begin{aligned} &\min_{W, \boldsymbol{b}, \boldsymbol{\mu}} \, \text{Tr}\left(Y L Y^T \right) + \lambda \|V\|_F^2 \\ &= \min_{W} \, \text{Tr}\left(W^T K_{R \times N} L K_{R \times N}^T W \right) + \lambda \, \text{Tr}\left(W^T K_{R \times R} W \right) \\ &= \min_{W} \, \text{Tr}\left(W^T C W \right), \end{aligned} \quad \text{(A.5)}$$

where $C = K_{R \times N} L K_{R \times N}^T + \lambda K_{R \times R}$ (Eq. (13)). Note that the parameters $\boldsymbol{b}$ and $\boldsymbol{\mu}$ are omitted in the above equation, because $\boldsymbol{b}$ has a close-form solution expressed in terms of W, and $\boldsymbol{\mu}$ is fixed when optimizing W.

Now, let's look at the other constraint $(1/N) \sum_{n=1}^{N} \boldsymbol{y}_n \boldsymbol{y}_n^T = I$. It can be rewritten as:

$$\begin{aligned} &\frac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}_n \boldsymbol{y}_n^T \\ &= \frac{1}{N} \sum_{n=1}^{N} \left(\left(W^T K_{R \times N}(:, n) - \frac{1}{N} W^T K_{R \times N} \boldsymbol{1} \right) \right. \\ &\quad \left. \left(W^T K_{R \times N}(:, n) - \frac{1}{N} W^T K_{R \times N} \boldsymbol{1} \right)^T \right) \\ &= \frac{1}{N} \left(\sum_{n=1}^{N} W^T K_{R \times N}(:, n) K_{R \times N}(:, n)^T W \right. \\ &\quad - \frac{1}{N} \sum_{n=1}^{N} W^T K_{R \times N}(:, n) \boldsymbol{1}^T K_{R \times N}^T W \\ &\quad - \frac{1}{N} \sum_{n=1}^{N} W^T K_{R \times N} \boldsymbol{1} K_{R \times N}(:, n)^T W \\ &\quad \left. + \frac{1}{N^2} \sum_{n=1}^{N} W^T K_{R \times N} \boldsymbol{1} \boldsymbol{1}^T K_{R \times N}^T W \right) \\ &= \frac{1}{N} \left(W^T K_{R \times N} K_{R \times N}^T W \right. \\ &\quad - \frac{1}{N} W^T K_{R \times N} \boldsymbol{1} \boldsymbol{1}^T K_{R \times N}^T W \\ &\quad - \frac{1}{N} W^T K_{R \times N} \boldsymbol{1} \boldsymbol{1}^T K_{R \times N}^T W \\ &\quad \left. + \frac{1}{N} W^T K_{R \times N} \boldsymbol{1} \boldsymbol{1}^T K_{R \times N}^T W \right) \end{aligned}$$

$$= \frac{1}{N} W^T K_{R \times N} \left(I - \frac{1}{N} \mathbf{1} \mathbf{1}^T \right) K_{R \times N}^T W$$

$$= W^T G W = I, \tag{A.6}$$

where $G = (1/N) K_{R \times N} \left(I - (1/N) \mathbf{1} \mathbf{1}^T \right) K_{R \times N}^T$ (Eq. (14)). In the above derivation, we use the facts $\sum_{n=1}^{N} K_{R \times N}(:, n) K_{R \times N}(:, n)^T = K_{R \times N} K_{R \times N}^T$, $\sum_{n=1}^{N} K_{R \times N}(:, n) = K_{R \times N} \mathbf{1}$, and $\sum_{n=1}^{N} K_{R \times N}(:, n)^T = \mathbf{1}^T K_{R \times N}^T$. With the objective function represented by Eq. (A.5) and the constraint represented by Eq. (A.6), we can get Eq. (12).

Given fixed $(W, \boldsymbol{b})$, using Eqs. (3), (4), (A.3), and (A.4), we can rewrite the objective function as

$$\mathrm{Tr}\left(Y L Y^T\right) + \lambda \|V\|_F^2$$

$$= \mathrm{Tr}\left(W^T \left(\sum_{i=1}^{M} \mu^{(i)} K_{R \times N}^{(i)} \right) L \left(\sum_{j=1}^{M} \mu^{(j)} K_{R \times N}^{(j)} \right)^T W \right)$$

$$+ \lambda \, \mathrm{Tr}\left(W^T \left(\sum_{m=1}^{M} \mu^{(m)} K_{R \times R}^{(m)} \right) W \right) \tag{A.7}$$

$$= \mathrm{Tr}\left(\sum_{i,j=1}^{M} \mu^{(i)} \mu^{(j)} W^T K_{R \times N}^{(i)} L \left(K_{R \times N}^{(j)} \right)^T W \right)$$

$$+ \lambda \, \mathrm{Tr}\left(\sum_{m=1}^{M} \mu^{(m)} W^T K_{R \times R}^{(m)} W \right).$$

According to the definitions of E (Eq. (17)) and $\boldsymbol{f}$ (Eq. (18)), the above equation can be further transformed to

$$\mathrm{Tr}\left(Y L Y^T\right) + \lambda \|V\|_F^2$$

$$= \sum_{i,j=1}^{M} \mu^{(i)} \mu^{(j)} \mathrm{Tr}\left(W^T K_{R \times N}^{(i)} L \left(K_{R \times N}^{(j)} \right)^T W \right)$$

$$+ \sum_{m=1}^{M} \mu^{(m)} \lambda \, \mathrm{Tr}\left(W^T K_{R \times R}^{(m)} W \right) \tag{A.8}$$

$$= \frac{1}{2} \sum_{i,j=1}^{M} \mu^{(i)} \mu^{(j)} E(i, j) + \sum_{m=1}^{M} \mu^{(m)} f^{(m)}$$

$$= \frac{1}{2} \boldsymbol{\mu}^T E \boldsymbol{\mu} + \boldsymbol{f}^T \boldsymbol{\mu}.$$

Using the above equation and the constraints $\mathbf{1}^T \boldsymbol{\mu} = 1, \boldsymbol{\mu} \succcurlyeq \mathbf{0}$, we can obtain Eq. (16). Note that in Eq. (16), the parameter $(W, \boldsymbol{b})$ is omitted since it's fixed when we optimize $\boldsymbol{\mu}$.

References

Akakin, H.C., Gurcan, M.N., 2012. Content-based microscopic image retrieval system for multi-image queries. IEEE Trans. Inf. Technol. Biomed. 16 (4), 758–769.

Akgül, C.B., Rubin, D.L., Napel, S., Beaulieu, C.F., Greenspan, H., Acar, B., 2011. Content-based image retrieval in radiology: current status and future directions. J. Digit. Imaging 24 (2), 208–222.

Al-Janabi, S., Huisman, A., van Diest, P.J., 2012. Digital pathology: current status and future perspectives. Histopathology 61 (1), 1–9.

Alto, H., Rangayyan, R.M., Desautels, J.E.L., 2005. Content-based retrieval and analysis of mammographic masses. J. Electron. Imaging 14 (2), 023016–1–17.

American Cancer Society, 2013. Breast Cancer Facts & Figures 2013–2014. American Cancer Society, Atlanta, GA, USA.

André, B., Vercauteren, T., Buchner, A.M., Wallace, M.B., Ayache, N., 2012. Learning semantic and visual similarity for endomicroscopy video retrieval. IEEE Trans. Med. Imaging 31 (6), 1276–1288.

Avni, U., Greenspan, H., Konen, E., Sharon, M., Goldberger, J., 2011. X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. IEEE Trans. Med. Imaging 30 (3), 733–746.

Caicedo, J.C., Cruz, A., González, F.A., 2009. Histopathology image classification using bag of features and kernel functions. In: Proceedings of Artificial Intelligence and Medication, pp. 126–135.

Caicedo, J.C., González, F.A., Romero, E., 2011. Content-based histopathology image retrieval using a kernel-based semantic annotation framework. J. Biomed. Inform. 44 (4), 519–528.

Cruz-Roa, A., González, F.A., Galaro, J., Judkins, A.R., Ellison, D., Baccon, J., Madabhushi, A., Romero, E., 2012. A visual latent semantic approach for automatic analysis and interpretation of anaplastic medulloblastoma virtual slides. In: Proceedings of Medical Image Computation and Computational Assistance Intervention (Part 1), pp. 157–164.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computational Visualisation and Pattern Recognition, pp. 886–893.

Dundar, M.M., Badve, S., Bilgin, G., Raykar, V.C., Jain, R.K., Sertel, O., Gurcan, M.N., 2011. Computerized classification of intraductal breast lesions using histopathological images. IEEE Trans. Biomed. Eng. 58 (7), 1977–1984.

Foran, D.J., Yang, L., Chen, W., Hu, J., Goodell, L.A., Reiss, M., Wang, F., Kurc, T., Pan, T., Sharma, A., Saltz, J.H., 2011. ImageMiner: a software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology. J. Am. Med. Inform. Assoc. 18 (4), 403–415.

Gionis, A., Indyk, P., Motwani, R., 1999. Similarity search in high dimensions via hashing. In: Proceedings of Very Large Data Base, pp. 518–529.

Gönen, M., Alpaydin, E., 2011. Multiple kernel learning algorithms. J. Mach. Learn. Res. 12, 2211–2268.

Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B., 2009. Histopathological image analysis: a review. IEEE Rev. Biomed. Eng. 2, 147–171.

Hassan, E., Chaudhury, S., Gopal, M., 2012. Feature combination in kernel space for distance based image hashing. IEEE Trans. Multimedia 14 (4), 1179–1195.

He, J., Liu, W., Chang, S., 2010. Scalable similarity search with optimized kernel hashing. In: Proceedings of the ACM International Conference on Knowledge and Discovring Data Mining, pp. 1129–1138.

Huang, P.-W., Lai, Y.-H., 2010. Effective segmentation and classification for HCC biopsy images. Pattern Recognit. 43 (4), 1550–1563.

Jiang, M., Zhang, S., Huang, J., Yang, L., Metaxas, D.N., 2015. Joint kernel-based supervised hashing for scalable histopathological image analysis. In: Proceedings of the Medical Image Computing and Computational Assistance Intervention (Part 3), pp. 366–373.

Jiang, M., Zhang, S., Li, H., Metaxas, D.N., 2015. Computer-aided diagnosis of mammographic masses using scalable image retrieval. IEEE Trans. Biomed. Eng. 62 (2), 783–792.

Kulis, B., Grauman, K., 2012. Kernelized locality-sensitive hashing. IEEE Trans. Pattern Anal. Mach. Intell. 34 (6), 1092–1104.

Kumar, A., Kim, J., Cai, W., Fulham, M., Feng, D., 2013. Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. J. Digit. Imaging 26 (6), 1025–1039.

Langs, G., Hanbury, A., Menze, B.H., Müller, H., 2012. VISCERAL: towards large data in medical imaging - challenges and directions. In: Proceedings of the International Workshop of Medical Content-Based Retr. Clin. Support, pp. 92–98.

Li, Y., Chen, C., Liu, W., Huang, J., 2014. Sub-selective quantization for large-scale image search. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 2803–2809.

Liang, Y., Wang, F., Treanor, D., Magee, D., Teodoro, G., Zhu, Y., Kong, J., 2015. A 3D primary vessel reconstruction framework with serial microscopy images. In: Proceedings of Medical Image Computing and Computational Assistance Intervention (Part 3), pp. 251–259.

Liang, Y., Wang, F., Treanor, D., Magee, D., Teodoro, G., Zhu, Y., Kong, J., 2015. Liver whole slide image analysis for 3D vessel reconstruction. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, pp. 182–185.

Liang, Y., Wang, F., Treanor, D., Magee, D.R., Roberts, N., Teodoro, G., Zhu, Y., Kong, J., 2016. A framework for 3D vessel analysis using whole slide images of liver tissue sections. Int. J. Comput. Biol. Drug Des. 9 (1–2), 102–119.

Liu, J., Zhang, S., Liu, W., Zhang, X., Metaxas, D.N., 2014. Scalable mammogram retrieval using anchor graph hashing. In: Proceedings of the IEEE International Symposium on Biomedical Imaging, pp. 898–901.

Liu, W., Wang, J., Ji, R., Jiang, Y., Chang, S., 2012. Supervised hashing with kernels. In: Proceedings of the IEEE on Computational Visualisation and Pattern Recognition, pp. 2074–2081.

Liu, W., Wang, J., Kumar, S., Chang, S., 2011. Hashing with graphs. In: Proceedings of the International Conference on Machine Learning, pp. 1–8.

Liu, X., He, J., Lang, B., 2014. Multiple feature kernel hashing for large-scale visual search. Pattern Recognit. 47 (2), 748–757.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60 (2), 91–110.

Müller, H., Michoux, N., Bandon, D., Geissbühler, A., 2004. A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. Int. J. Med. Inform. 73 (1), 1–23.

Nguyen, K., Jain, A.K., Allen, R.L., 2010. Automated gland segmentation and classification for Gleason grading of prostate tissue images. In: Proceedings of the International Conference on Pattern Recognition, pp. 1497–1500.

Nistér, D., Stewénius, H., 2006. Scalable recognition with a vocabulary tree. In: Proceedings of the IEEE Computation and Visualisation Pattern Recognition, pp. 2161–2168.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: a holistic representation of the spatial envelope. Int. J. Comput. Vis. 42 (3), 145–175.

Sivic, J., Zisserman, A., 2003. Video google: A text retrieval approach to object matching in videos. In: Proceedings of the IEEE International Conference on Computation and Visualisation., pp. 1470–1477.

Tabesh, A., Teverovskiy, M., Pang, H., Kumar, V.P., Verbel, D., Kotsianti, A., Saidi, O., 2007. Multifeature prostate cancer diagnosis and Gleason grading of histological images. IEEE Trans. Med. Imaging 26 (10), 1366–1378.

Tamaki, T., Yoshimuta, J., Kawakami, M., Raytchev, B., Kaneda, K., Yoshida, S., Takemura, Y., Onji, K., Miyaki, R., Tanaka, S., 2013. Computer-aided colorectal tumor classification in NBI endoscopy using local features. Med. Image Anal. 17 (1), 78–100.

Tuzel, O., Yang, L., Meer, P., Foran, D.J., 2007. Classification of hematologic malignancies using texton signatures. Pattern Anal. Appl. 10 (4), 277–290.

Veta, M., Pluim, J.P.W., van Diest, P.J., Viergever, M.A., 2014. Breast cancer histopathology image analysis: a review. IEEE Trans. Biomed. Eng. 61 (5), 1400–1411.

Wang, J., Kumar, S., Chang, S., 2012. Semi-supervised hashing for large-scale search. IEEE Trans. Pattern Anal. Mach. Intell. 34 (12), 2393–2406.

Wei, C.-H., Li, Y., Huang, P.J., 2011. Mammogram retrieval through machine learning within BI-RADS standards. J. Biomed. Inform. 44 (4), 607–614.

Weiss, Y., Torralba, A., Fergus, R., 2008. Spectral hashing. In: Proceedings of Neural Information Processing System, pp. 1753–1760.

Yang, L., Chen, W., Meer, P., Salaru, G., Goodell, L.A., Berstis, V., Foran, D.J., 2009. Virtual microscopy and Grid-enabled decision support for large-scale analysis of imaged pathology specimens. IEEE Trans. Inf. Technol. Biomed. 13 (4), 636–644.

Yang, W., Lu, Z., Yu, M., Huang, M., Feng, Q., Chen, W., 2012. Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single- and multiphase contrast-enhanced CT images. J. Digit. Imaging 25 (6), 708–719.

Zhang, D., Wang, F., Si, L., 2011. Composite hashing with multiple information sources. In: Proceedings of ACM International Conference Research and Development Information Retr., pp. 225–234.

Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D.N., 2015. Query specific rank fusion for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. 37 (4), 803–815.

Zhang, X., Dou, H., Ju, T., Xu, J., Zhang, S., 2015. Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis. IEEE J. Biomed. Health Inform PP (99), 1. doi:10.1109/JBHI.2015.2461671.

Zhang, X., Liu, W., Dundar, M., Badve, S., Zhang, S., 2015. Towards large-scale histopathological image analysis: hashing-based image retrieval. IEEE Trans. Med. Imaging 34 (2), 496–506.

Zhang, X., Xing, F., Su, H., Yang, L., Zhang, S., 2015. High-throughput histopathological image analysis via robust cell segmentation and hashing. Med. Image Anal. 26 (1), 306–315.

Zhang, X., Yang, L., Liu, W., Su, H., Zhang, S., 2014. Mining histopathological images via composite hashing and online learning. In: Proceedings of Medical Image Computing Computational Assistance Intervention(Part 2), pp. 479–486.

Zheng, B., Lu, A., Hardesty, L.A., Sumkin, J.H., Hakim, C.M., Ganott, M.A., Gur, D., 2006. A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment. Med. Phys. 33 (1), 111–117.

Zheng, L., Wetzel, A.W., Gilbertson, J.R., Becich, M.J., 2003. Design and analysis of a content-based pathology image retrieval system. IEEE Trans. Inf. Technol. Biomed. 7 (4), 249–255.