CHAPTER

Hashing-based large-scale medical image retrieval for computer-aided diagnosis

8

X. Zhang, S. Zhang
University of North Carolina at Charlotte, Charlotte, NC, United States

CONTENTS

8.1 INTRODUCTION

An important goal in medical imaging informatics is transforming raw images into a quantifiable symbolic form for *indexing, retrieval, and reasoning.* In particular, content-based image retrieval (CBIR) has become important in medical informatics by providing doctors with diagnostic aid in the form of visualizing existing and relevant cases, along with diagnosis information. Therefore computer-aided diagnosis techniques such as case-based reasoning or evidence-based medicine have a strong need to retrieve images that can be valuable for diagnosis. However, this task, of interest to both academia and the healthcare industry, is challenging because medical image content is essentially unstructured. Take tissue images as an example: "accidental characteristics" can be caused by incompletely controlled illumination, variability in tissue texture, image noise introduced in the staining process and normal variations among subjects, impeding the success of existing systems. Most importantly, few systems are able to analyze large-scale (ie, ever-increasing amount and complexity of) medical image databases in real-time. When applied to

Machine Learning and Medical Imaging. http://dx.doi.org/10.1016/B978-0-12-804076-8.00008-6

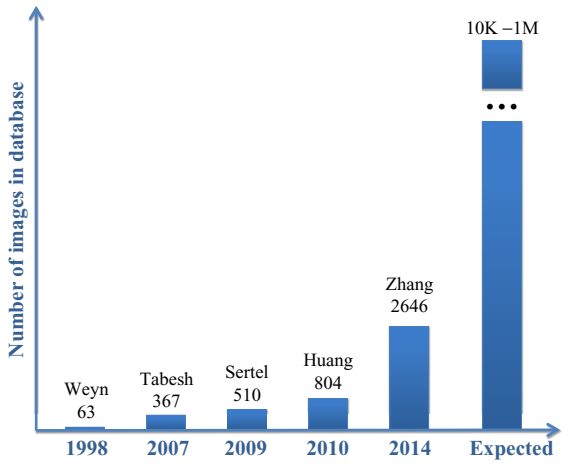

FIG. 8.1

The number of histopathological images analyzed by Weyn et al. (1998), Tabesh et al. (2007), Sertel et al. (2009), Huang and Lai (2010), and in our preliminary work (Zhang et al., 2014a). The expected size is 10K to one million. (The *x*-axis represents the years, and the *y*-axis denotes the number of images.)

full-resolution, real-world data, most methods in this area can only handle relatively small data sets (eg, several hundred images). For the problem of histopathological image analysis, Fig. 8.1 shows the capacity, in terms of number of database images processed, for recent approaches compared to the expected size for a clinically relevant database. The development of large-scale medical image analysis algorithms has lagged greatly behind the increasing quality (and complexity) of medical images and the imaging modalities themselves. These drawbacks limit the effectiveness of current image retrieval systems in research and industrial settings for mining the ever-growing number of medical images stored digitally.

Considering these factors, there is an urgent need to develop an innovative, integrated framework enabling robust and timely measurement, analysis, and characterization of such databases. Efforts in web-scale computer vision and multimedia databases have hinted at the promise of large-scale, data-driven methods for robust tagging, fine-grained object classification. In medical imaging informatics, the ever-increasing amount of medical images also provides a foundation for novel methods of semantic analysis. Transforming these raw images into a quantifiable, symbolic form will facilitate indexing and retrieval, and potentially lead to new avenues of knowledge discovery and decision support. The goal of this chapter is to introduce recent progress in large-scale visual data mining and information retrieval methods for knowledge discovery in potentially massive databases of medical images, particularly in the domain of histopathological image analysis.

8.2 RELATED WORK

Information retrieval in medical images has been widely investigated in this community. For example, Comaniciu et al. (1999) proposed a content-based image-retrieval system that supports decision making in clinical pathology, in which a central module and fast color segmenter are used to extract features such as shape, area, and texture of the nucleus. System performance was assessed through a 10-fold cross-validated classification and compared with that of a human expert on a database containing 261 digitized specimens. Dy et al. (2003) described a new hierarchical approach of CBIR based on multiple feature sets and a two-step approach. The query image is classified into different classes with best discriminative features between the classes. Then similar images are searched in the predicted class with the features customized to distinguish subclasses. El-Naqa et al. (2004) proposed a hierarchical learning approach that consists of a cascade of a binary classifier and a regression module to optimize retrieval effectiveness and efficiency. They applied this to retrieve digital mammograms and evaluated it on 76 mammograms. Greenspan and Pinhas (2007) proposed a CBIR system that consists of a continuous and probabilistic image-representation scheme. It uses information-theoretic image matching via the Kullback-Leibler (KL) measure to match and categorize X-ray images by body region. Song et al. (2011) designed a hierarchical spatial matching-based image-retrieval method using spatial pyramid matching to effectively extract and represent the spatial context of pathological tissues. CBIR has also been employed for histopathological image analysis. For example, Schnorrenberg et al. (2000) extended the biopsy analysis support system to include indexing and content-based retrieval of biopsy slide images. A database containing 57 breast cancer cases was used for evaluation. Zheng et al. (2003) designed a CBIR system to retrieve images and their associated annotations from a networked microscopic pathology image database based on four types of image features. Akakin and Gurcan (2012) proposed a CBIR system using the multitiered approach to classify and retrieve microscopic images. It enables both multi-image query and slide-level image retrieval in order to protect the semantic consistency among the retrieved images. Foran et al. (2011) designed a CBIR system named ImageMiner for comparative analysis of tissue microarrays by harnessing the benefits of high-performance computing and grid technology.

As emphasized in Zhou et al. (2008), scalability is the key factor in CBIR for medical image analysis. In fact, with the ever-increasing amount of annotated medical data, large-scale, data-driven methods provide the promise of bridging the semantic gap between images and diagnoses. However, the development of large-scale medical image analysis algorithms has lagged greatly behind the increasing quality and complexity of medical images. Specifically, owing to the difficulties in developing scalable CBIR systems for large-scale data sets, most previous systems have been tested on a relatively small number of cases. With the goal of comparing CBIR methods on a larger scale, ImageCLEF and VISCERAL provide benchmarks for medical image retrieval tasks (Müller et al., 2005; Langs et al., 2013; Hanbury et al., 2013). Recently, hashing methods have been intensively investigated in the machine

learning and computer vision community for large-scale image retrieval (Wang et al., 2015). They enable fast approximated nearest neighbors (ANN) search to deal with the scalability issue. For example, locality sensitive hashing (LSH) (Andoni and Indyk, 2006) uses random projections to map data to binary codes, resulting in highly compact binary codes and enabling efficient comparison within a large database using the Hamming distance. Anchor graph hashing (AGH) (Liu et al., 2011) has been proposed to use neighborhood graphs which reveal the underlying manifold of features, leading to a high search accuracy. Recent research has focused on data-dependent hash functions, such as spectral graph partitioning and hashing (Weiss et al., 2009) and supervised hashing with kernels (Liu et al., 2012) incorporating the pairwise semantic similarity and dissimilarity constraints from labeled data. These hashing methods have also been employed to solve the dimensionality problem in medical image analysis. In particular, Zhang et al. (2014a, 2015c) built a scalable image-retrieval framework based on the supervised hashing technique and validated its performance on several thousand histopathological images acquired from breast microscopic tissues. It leverages a small amount of supervised information in learning to compress high-dimensional image feature vectors into only tens of binary bits with the informative signatures preserved. The supervised information is employed to bridge the semantic gap between low-level image features and high-level diagnostic information, which is critical to medical image analysis. Instead of hashing and searching the whole image, another approach is to segment all cells from histopathological images and conduct large-scale retrieval among cell images (Zhang et al., 2015d,e,f). This enables cell-level and fine-grained analysis, achieving high accuracy. In addition to using a single feature, it is also possible to fuse multiple types of features in a hashing framework to improve the accuracy of medical image retrieval. Specifically, a composite AGH algorithm (Liu et al., 2011) has been employed for retrieving medical images (Zhang et al., 2014b; Liu et al., 2014), for example, retrieving lung microscopic tissue images for the differentiation of adenocarcinoma and squamous carcinoma. Besides hashing-based methods, vocabulary trees have also been intensively investigated (Nister and Stewenius, 2006) and employed for medical image analysis (Jiang et al., 2015b, 2014).

In this chapter, we introduce technical details of hashing-based large-scale image retrieval for computer-aided diagnosis, with the use case of histopathological image analysis.

8.3 SUPERVISED HASHING FOR LARGE-SCALE RETRIEVAL

8.3.1 OVERVIEW OF SCALABLE IMAGE RETRIEVAL FRAMEWORK

Fig. 8.2 shows a framework for the scalable image retrieval-based diagnosis system. It includes offline learning and run-time search. During the offline learning, we first extract high-dimensional visual features from digitized histopathological images.

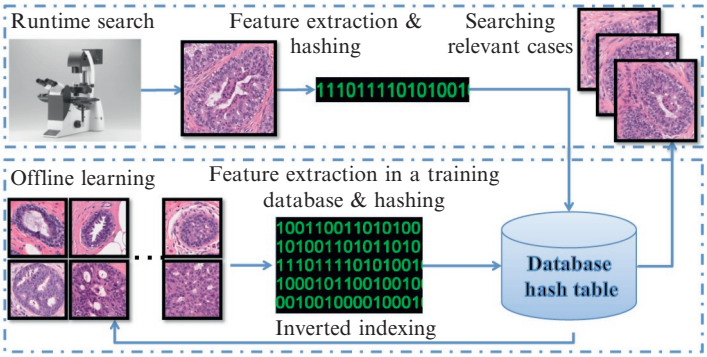

FIG. 8.2

Framework of our large-scale image retrieval system (Zhang et al., 2014a).

These features model texture and appearance information based on SIFT (Lowe, 2004) and are quantized with a bag-of-words (Sivic and Zisserman, 2003). The SIFT descriptor is an effective local texture feature that uses the difference of Gaussian (DoG) detection result and considers the gradient of pixels around the detected region. It can provide an informative description of cell appearance and is robust to subtle changes in staining color. It has been used in both general computer vision tasks and histopathological image analysis.

Although these features can be used directly to measure the similarity among images, computational efficiency is an issue, especially when searching in a large database (eg, exhaustively searching k-nearest neighbors (kNN)). Therefore we employ a hashing method to compress these features into binary codes with tens of bits. Such short binary features allow easy mapping into a hash table for real-time search. Each feature is then linked to the corresponding training images using an inverted index. During a run-time query, high-dimensional features are extracted from the query image and then projected to the binary codes. With a hash table, searching for nearest neighbors can be achieved in a constant time, irrespective of the number of images. The retrieved images (via inverted indices of nearest neighbors) can be used to interpret this new case or for decision support based on majority voting.

8.3.2 KERNELIZED AND SUPERVISED HASHING

In this section, we introduce the key module for histopathological image retrieval, a kernelized and supervised hashing method.

8.3.2.1 Hashing method

Given a set of image feature vectors $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \subset \mathbb{R}^d$ (in our case, $\boldsymbol{x}_i$ is the high-dimensional texture feature extracted from the ith histopathological image), a hashing method aims to find a group of proper hash functions h: $\mathbb{R}^d \mapsto \{1, -1\}^1$, each

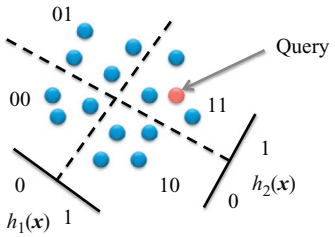

FIG. 8.3

Visualization of desirable hash functions as a hyperplane.

of which generates a single hash bit to preserve the similarity of original features. Searching kNN using tens of bits is significantly faster than traditional methods (eg, Euclidean distance-based brute-force search), owing to constant-time hash-table lookups and/or efficient Hamming distance computation. Note that hashing methods are different from dimensionality-reduction techniques, since a fundamental requirement of hashing is to map similar feature vectors into the same bucket with high probability. Fig. 8.3 visualizes desirable hash functions as a hyperplane to separate higher-dimensional features. Therefore hashing methods need to ensure that the generated hash bits have balanced and uncorrelated bit distributions, which leads to maximum information at each single bit and minimum redundancy among all bits.

8.3.2.2 Kernelized hashing

Kernel methods can handle practical data that are mostly linearly inseparable. For histopathological images, linear inseparability is an important constraint that needs to be taken into account when building hashing methods. Therefore kernel functions should be considered in hashing methods, $h = sgn(f(x))$ (Kulis and Grauman, 2012), to map the feature vectors into higher-dimensional space. A kernel function is denoted as $\kappa\colon \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}$. The prediction function $f\colon \mathbb{R}^d \mapsto \mathbb{R}$ with kernel κ plugged in is defined as

$$f(\boldsymbol{x}) = \sum_{j=1}^{m} \kappa(\boldsymbol{x}_{(j)}, \boldsymbol{x}) a_j - b, \tag{8.1}$$

where $\boldsymbol{x}_{(1)}, \ldots, \boldsymbol{x}_{(m)}$ are m ($m \ll n$) feature vectors randomly selected from $\mathcal{X}$, $a_j \in \mathbb{R}$ is the coefficient, and $b \in \mathbb{R}$ is the bias.

The bits generated from hash functions h using f aim to keep as much information as possible, so the hash functions should produce a balanced distribution of bits, that is, $\sum_{i=1}^{n} h(\boldsymbol{x}_i) = 0$. Therefore b is set as the median of $\{\sum_{j=1}^{m} \kappa(\boldsymbol{x}_{(j)}, \boldsymbol{x}_i) a_j\}_{i=1}^{n}$, which is usually approximated by the mean. Adding this constraint into Eq. 8.1, we obtain

$$f(\boldsymbol{x}) = \sum_{j=1}^{m} \left(\kappa(\boldsymbol{x}_{(j)}, \boldsymbol{x}) - \frac{1}{n} \sum_{i=1}^{n} \kappa(\boldsymbol{x}_{(j)}, \boldsymbol{x}_i) \right) a_j = \boldsymbol{a}^{\mathrm{T}} \bar{\boldsymbol{k}}(\boldsymbol{x}), \tag{8.2}$$

where $\boldsymbol{a} = [a_1, a_2, \ldots, a_m]^{\mathrm{T}}$. $\bar{\boldsymbol{k}} : \mathbb{R}^d \mapsto \mathbb{R}^m$ is $\bar{\boldsymbol{k}}(\boldsymbol{x}) = \left[\kappa(\boldsymbol{x}_{(1)}, \boldsymbol{x}) - \mu_1, \ldots, \kappa(\boldsymbol{x}_{(m)}, \boldsymbol{x}) - \mu_m\right]^{\mathrm{T}}$, in which $\mu_j = \sum_{i=1}^{n} \kappa(\boldsymbol{x}_{(j)}, \boldsymbol{x}_i)/n$.

The vector $\boldsymbol{a}$ is the most important factor that determines hash functions. In traditional kernelized hashing methods, $\boldsymbol{a}$ is defined as a random direction drawn from a Gaussian distribution (Kulis and Grauman, 2012), without using any other prior knowledge (ie, no semantic information). This scheme works well for natural images, especially scenes, because of large differences in their appearance. However, such differences are very subtle in histopathological images. For example, identifying subtle differences between benign and actionable categories may require characterizing cytoplasmic texture or nuclear appearance. This subtlety motivates us to leverage supervised information to design discriminative hash functions that are suitable for histopathological image retrieval.

8.3.2.3 Supervised hashing

Intuitively, hashing methods minimize the Hamming distance of "neighboring" image pairs (eg, close in terms of the Euclidean distance in the raw feature space). "Neighboring" in our case is defined by its semantic meaning, that is, whether the two images belong to the same category or not. Therefore supervised information can be naturally encoded as similar and dissimilar pairs. Specifically, we assign the label 1 to image pairs when both are benign or actionable, and -1 to pairs when one is benign and the other is actionable (as shown in Fig. 8.4). Then, l ($l \ll n$) feature vectors are randomly selected from $\mathcal{X}$ to build the label matrix S. Note that we need to provide labels for only a small number of image pairs. Therefore labeled data are explicitly constrained by both semantic information and visual similarities, whereas unlabeled data are mainly constrained by visual similarities and implicitly affected by labeled data.

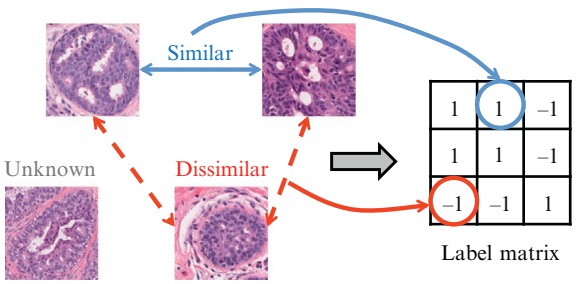

FIG. 8.4

Supervised information is encoded in the label matrix S.

Using this supervision scheme to bridge the semantic gap, r hash functions $h_k(\boldsymbol{x})_{k=1}^{r}$ are then designed to generate r discriminative hash bits based on Hamming distances. However, direct optimization of the following Hamming distances $\mathcal{D}_h(\boldsymbol{x}_i, \boldsymbol{x}_j) = |\{k|h_k(\boldsymbol{x}_i) \neq h_k(\boldsymbol{x}_j), 1 \leq k \leq r\}|$ is nontrivial. Therefore code inner products can be used to simplify the optimization process. As shown in Liu et al. (2012), a Hamming distance and a code inner product are actually equivalent:

$$code_r(\boldsymbol{x}_i) \circ code_r(\boldsymbol{x}_j) = r - 2\mathcal{D}_h(\boldsymbol{x}_i, \boldsymbol{x}_j), \tag{8.3}$$

where $code_r(\boldsymbol{x})$ are r-bit hash codes and the symbol $\circ$ is the code inner product.

Therefore the objective function $\mathcal{Q}$ to the binary codes H_l is defined as

$$\min_{H_l \in \{1,-1\}^{l \times r}} \mathcal{Q} = \left\| \frac{1}{r} H_l H_l^{\mathrm{T}} - S \right\|_F^2, \tag{8.4}$$

where $H_l = \begin{bmatrix} h_1(\boldsymbol{x}_1), \ldots, h_r(\boldsymbol{x}_1) \\ \cdots\cdots \\ h_1(\boldsymbol{x}_l), \ldots, h_r(\boldsymbol{x}_l) \end{bmatrix}$ is the code matrix of the labeled data $\mathcal{X}_l$ and S is a label matrix with 1 for similar pairs and -1 for dissimilar pairs. $\|.\|_F$ denotes the Frobenius norm. Define $\bar{K}_l$ as $[\bar{\boldsymbol{k}}(\boldsymbol{x}_1), \ldots, \bar{\boldsymbol{k}}(\boldsymbol{x}_l)]^{\mathrm{T}} \in \mathbb{R}^{l \times m}$, $\bar{\boldsymbol{k}}(\boldsymbol{x}_i)$. The inner product of code matrix H_l can be represented as $H_l H_l^{\mathrm{T}} = \sum_{k=1}^{r} sgn(\bar{K}_l \boldsymbol{a}_k)(sgn(\bar{K}_l \boldsymbol{a}_k))^{\mathrm{T}}$ for binarization. Therefore the new objective function $\mathcal{Q}$ that offers a clearer connection and easier access to the model parameter $\boldsymbol{a}_k$ is

$$\min_{\boldsymbol{a}_k} \quad \mathcal{Q}(\boldsymbol{a}_k) = \left\| \sum_{k=1}^{r} sgn(\bar{K}_l \boldsymbol{a}_k)(sgn(\bar{K}_l \boldsymbol{a}_k))^{\mathrm{T}} - rS \right\|_F^2. \tag{8.5}$$

This can be optimized using (1) spectral relaxation (Weiss et al., 2009) to drop the sign functions and hence convexify the object function, or (2) sigmoid smoothing to replace $sgn()$ with the sigmoid-shaped function. In our implementation, we employ the first strategy to efficiently obtain a solution as the initialization, and use the second strategy to produce an accurate solution.

8.4 RESULTS

Breast-tissue specimens available for this study were collected on a retrospective basis from the IU Health Pathology Lab (IUHPL) according to the protocol approved by the Institutional Review Board (IRB) for this study. All the slides were imaged using a ScanScope digitizer (Aperio, Vista, CA) available in the tissue archival service at IUHPL. 3121 images (around 2250K pixels) were sampled from 657 larger region-of-interest images (eg, 5K×7K) of microscopic breast tissue, which were gathered from 116 patients. Fifty-three of these patients were labeled as benign (usual ductal hyperplasia (UDH)) and 63 as actionable (atypical ductal hyperplasia

(ADH) and ductal carcinoma in situ (DCIS)), based on the majority diagnosis of nine board-certified pathologists. To demonstrate the efficiency of our method, one-fourth of all patients in each category were randomly selected as the test set and the remainder used for training. Note that each patient may have a different number of images. Therefore the number of testing images is not fixed. The approximate number is about 700–900 in each testing process. All the experiments were conducted on a 3.40-GHz CPU with four cores and 16GB RAM, in a MATLAB implementation.

In each image, 1500–2000 SIFT descriptors were extracted from key points detected by DoG (Lowe, 2004). These descriptors were quantized into sets of cluster centers using bag-of-words, in which the feature dimension equals the number of clusters. Specifically, we quantize them into high-dimensional feature vectors of length 10,000, to maximally utilize these millions of cell-level texture features. We provide both qualitative and quantitative evaluations for our proposed framework on two tasks, image classification (ie, benign vs. actionable category) and image retrieval, in terms of accuracy and computational efficiency.

In our system, classification is achieved using the majority vote of the top images retrieved by hashing. We compare our approach with various classifiers that have been widely used in systems for histopathological image analysis. Specifically, kNN has often been used as the baseline in analyzing histopathological images (Tabesh et al., 2007; Yang et al., 2009), owing to its simplicity and proved lower bound, despite the inefficiency in large-scale databases. The Bayesian method is another solution to ensemble statistics of all extracted features and minimize the classification metric, which shows its efficacy in classifying histopathological images (Comaniciu et al., 1999). Boosting methods are always employed to combine multiple weak classifiers for higher accuracy (Yang et al., 2009; Doyle et al., 2012). A support vector machine (SVM) with a nonlinear kernel is commonly used in histopathological images because of its efficiency and the ability to handle linearly inseparable cases (Tuzel et al., 2007; Caicedo et al., 2009; Nguyen et al., 2010; Huang and Lai, 2010). For fair comparison, all parameters of these compared methods were optimized by cross-validation.

In addition, we also compared our proposed method with several dimensionality-reduction algorithms in terms of classification accuracy. Principal component analysis (PCA) has been widely used in this area to preserve variance of original features (Sertel et al., 2009). Graph embedding is a nonlinear dimensionality-reduction algorithm that performs well in grading of lymphocytic infiltration in HER2+ breast cancer histopathology (Basavanhally et al., 2010). Since we use supervised information in generating hash functions, a supervised dimensionality-reduction algorithm, neighborhood components analysis (NCA) (Goldberger et al., 2004), was also chosen for our experimental comparisons.

Fig. 8.5 shows the quantitative results for the classification accuracy. Most methods achieve better accuracy with higher-dimensional features. This is very intuitive, as finer quantization of SIFT features usually provides richer information. In particular, since the SIFT interest points cover most nuclear regions in images, fine quantization (ie, high-dimensional features) indicates analysis on a small scale.

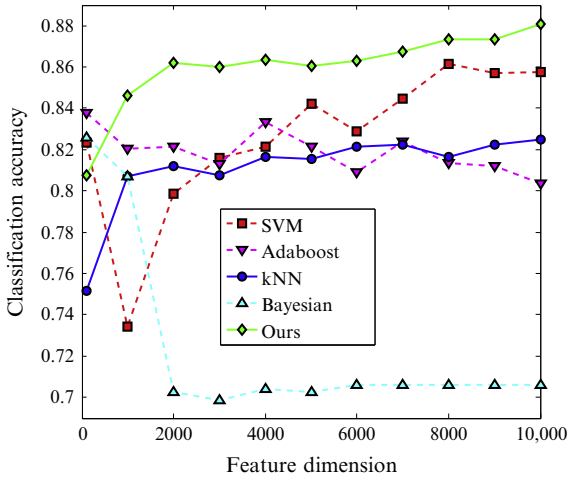

FIG. 8.5

Comparison of classification accuracy with different dimensions of features (from 100 to 10,000).

Exceptions are the Adaboost and Bayesian methods, whose accuracy drops when the feature dimensions increase. This indicates that high-dimensional features do not guarantee the improvement of accuracy. An important factor is the proper utilization of such information. For example, Adaboost is essentially a feature-selection method that chooses only an effective subset of features for the classification. Therefore it may lose important information, especially in high-dimensional space, resulting in accuracy worse than that of our hashing method. Our method is also generally better than kNN and its variations, owing to the semantic information (ie, labels of similar and dissimilar pairs in hashing) that bridges the semantic gap between images and diagnoses. Note that our hashing method needs only a small amount of supervision—in this case, similar or dissimilar pairs of 40% images. This is generally less than the supervised information required by SVM in the training stage. It compares favorably to all other methods when the feature dimension is larger than 1000. The overall classification accuracy is 88.1% for 10,000-dimensional features, 2–18% better than the other methods.

Fig. 8.6 compares the computational efficiency of these methods. With increasing dimensionality the running time of some compared methods increases dramatically. When feature dimensionality reaches 10,000, kNN needs 16 s to classify all query images, and Adaboost needs 5 s. SVM, dimensionality-reduction methods, and the proposed method are much faster. However, the running time for SVM increases with the feature dimensionality, as shown in the expanded view of Fig. 8.6. In contrast, PCA, graph embedding, NCA, and our method achieve constant running time in this data set owing to the fixed size of features after compression. Compared to other

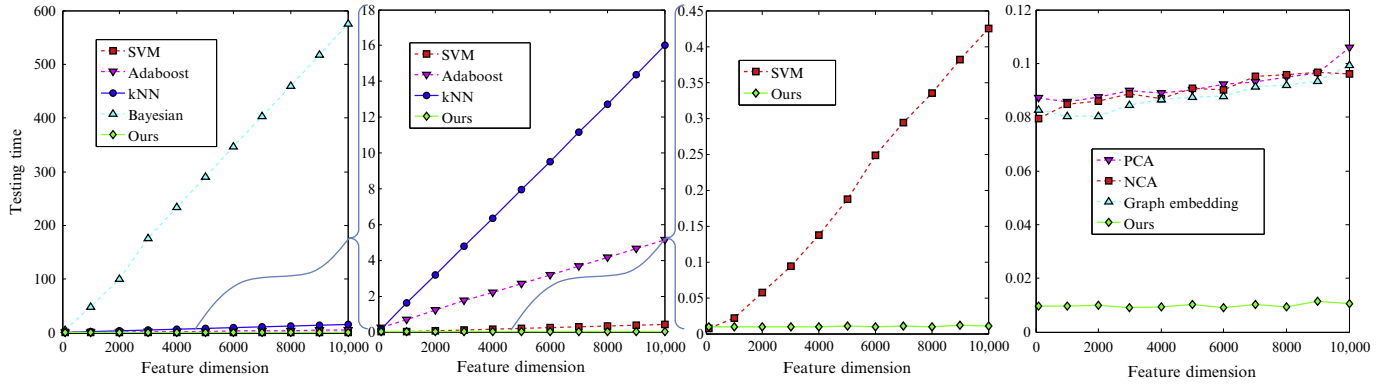

FIG. 8.6

Comparison of the classification running time (seconds) with different dimensions of features, which means the average time of classifying hundreds of test images.

dimensionality-reduction methods, our approach is about 10 times faster because of the efficient comparison among binary codes. In addition, the running time of all kNN-based methods increases with the number of images in a data set, as exhaustive search is needed, while hashing-based methods can achieve $\mathcal{O}(1)$ efficiency using a hash table. To summarize, the average running time of our method is only 0.01 s for all testing images, which is 40 times faster than SVM and 1500 times faster than kNN.

We have also conducted experiments on image retrieval using 10,000-dimensional features. The retrieval precision, evaluated at a given cut-off rank and considering only the topmost results, is reported in Table 8.1, along with the query time and memory cost. The results are quite consistent with the image classification. The mean precision of the hashing method is around 83%, and the standard deviation is 1.1%, which is much better than PCA (Sertel et al., 2009), graph embedding (Basavanhally et al., 2010), and NCA (Goldberger et al., 2004). In most cases, the precision of our method is at least 6% better than the others, except for NCA. Our method is around 3.5% better than NCA on benign cases. To demonstrate statistical significance, we perform t-test for the precision obtained by NCA and by the proposed method on benign cases, under the null hypothesis using a significance level of 0.05. The p-values are found to be 3.6×10^{-6}, 3.2×10^{-6}, and 5.7×10^{-6} at the range of the top 10, 20, and 30 retrievals, respectively, demonstrating that precision values achieved by the proposed technique are indeed significantly better than NCA for the benign cases. In addition, our method is around 14% better than NCA in the actionable cases, resulting in much higher average precision. In fact, most traditional methods produce results as highly unbalanced NCA, that is, the retrieval precision of the benign category is much higher than that of the actionable one. In contrast, our method does not have this problem, owing to the supervised information and the optimization for balanced hash bits. Our framework is also computationally more efficient than traditional methods. The query time of our hashing method is 1000 times faster than kNN and 10 times faster than other dimensionality-reduction methods. Note that our method takes a constant time when using the hash table, independent of the number of feature dimensions and the number of samples. Furthermore, the memory cost is also considerably reduced (10,000 times less than that of kNN). Therefore this method is more applicable to large-scale databases (millions of images) than are other methods.

Fig. 8.7 shows our image-retrieval results. The top five relevant images are listed for each query image. The differences between certain images in different categories are very subtle. Our accurate results demonstrate the efficacy of the proposed method. Specifically, the features capturing local texture and appearance are very robust to various image sizes, cell distributions, and occlusions by the blood. The supervised information also improves the retrieval precision by correlating binary code with diagnosis information. These retrieved images are clinically relevant in potential (ie, retrieved images belong to the same category as the query image) and thus can be useful for decision support.

Table 8.1 Comparison of Retrieval Precision for the Top 10, 20, and 30 Results (Denoted as P@10, P@20, and P@30, Respectively), Along With the Memory Cost of Training Data and Query Time of All Test Images

	kNN		PCA		NCA		Graph Embedding		Ours	
	Benign	**Actionable**	**Benign**	**Actionable**	**Benign**	**Actionable**	**Benign**	**Actionable**	**Benign**	**Actionable**
P@10	0.779	0.687	0.762	0.705	0.799	0.697	0.672	0.487	**0.836**	**0.830**
P@20	0.773	0.653	0.758	0.681	0.800	0.689	0.673	0.486	**0.839**	**0.829**
P@30	0.770	0.631	0.755	0.667	0.800	0.685	0.670	0.480	**0.837**	**0.833**
STD	0.024		0.028		0.020		0.012		0.011	
Time (s)	15.77		10.07		10.04		10.03		<0.01	
Memory	134.58MB		0.65MB		0.65MB		0.65MB		0.01MB	

Both mean values and the standard deviation (STD) of 20 experiments are reported. The best precision in each row for benign and actionable categories are highlighted in bold.

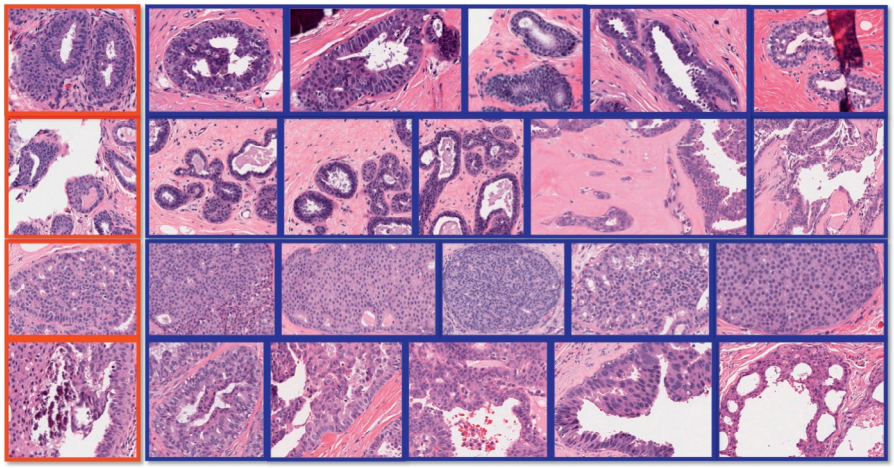

FIG. 8.7

Four examples of our image retrieval (query marked in red and in the first column, and retrieved images marked in blue). The first two rows are benign; the last two rows are actionable. (A) Query. (B) Retrieved Images

8.5 DISCUSSION AND FUTURE WORK

In this chapter, we introduce a *large-scale image-retrieval framework* for medical image analysis. We employ hashing to achieve efficient image retrieval and present an improved kernelized and supervised hashing approach for real-time image retrieval. Note that even though we use histopathological image analysis as a use case, this framework is applicable to other problems in this area. The potential applications of our framework include image-guided diagnosis, decision support, education, and efficient data management. For example, the efficient retrieval of relevant cases from medical databases will provide usable tools to assist clinicians' diagnoses and support efficient medical image data management, such as picture archiving and communication systems (PACS). More specifically, it provides efficient reasoning in large-scale medical image databases using techniques for scalable and accurate medical image retrieval in potentially massive databases to provide real-time querying for the most relevant and consistent instances (eg, similar morphological profiles) for decision support. In addition to the resulting tools for medical image processing, disease detection and information retrieval, their use will allow for the exploration of structured image databases, in medical education and training.

An important extension to the existing framework is to integrate multiple features for accurate retrieval and diagnosis. For example, accurate analysis of histopathological images requires examination of cell-level information for accurate diagnosis, including individual cells (eg, appearance (Caicedo et al., 2009; Zhang et al., 2015c)

and shapes (Dundar et al., 2011)) and architecture of tissue (eg, topology and layout of all cells (Basavanhally et al., 2010)). These features cover both local and holistic information, all benefitting the diagnostic accuracy of histopathological images. Therefore the complementary descriptive capability of local and holistic features motivates us to integrate their strengths to yield more satisfactory results. However, their characteristics, algorithmic procedures, and representations can be dramatically different, making them nontrivial to fuse. For example, architecture features (Basavanhally et al., 2010) are represented as a low-dimensional vector of statistics, while local features can be represented as high-dimensional bag-of-words (BoW) (Sivic and Zisserman, 2003) and compressed as binary codes to improve the efficiency (Liu et al., 2012; Zhang et al., 2015c). To tackle this feature fusion problem, we have focused on the rank-level fusion of local and holistic features for the image-guided diagnosis of breast cancer, that is, differentiation of the benign and actionable cases. In particular, we conduct image retrieval to discover clinically relevant instances from an image database, which can be used to infer and classify the new data. Given image ranks (ie, retrieval results) obtained from different features, a data-driven and graph-based method (Zhang et al., 2012, 2015a,b) is employed for accurate, robust, and efficient fusion, by evaluating the quality of each rank online. Fig. 8.8 shows the overview of this graph-based feature fusion for image retrieval. This provides an effective solution for the fusion of heterogeneous information in the domain of histopathological image analysis, and the preliminary experimental results demonstrate the accuracy and efficiency of our framework. In addition to rank-based fusion, it is also possible to integrate multiple features in the hashing framework, either at the distance level or at the kernel level (Jiang et al., 2015a).

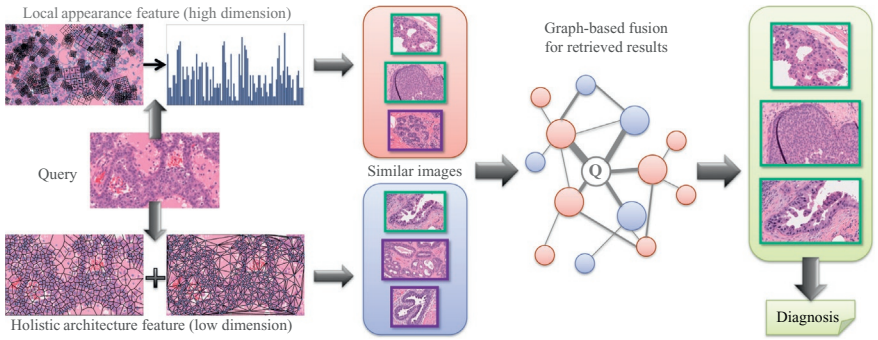

FIG. 8.8

Overview of the graph-based feature fusion for image retrieval (Zhang et al., 2015b). Both holistic architecture feature and local appearance feature are extracted and employed for image retrieval. The retrieval results are fused via the graph-based framework to improve the accuracy. Note that majority voting does not work in this example, since two ranks have no intersection.

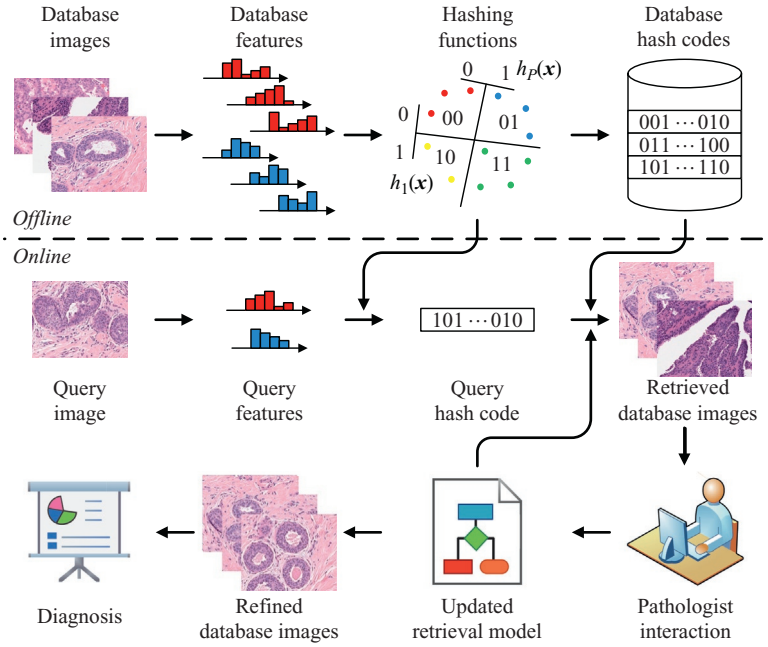

FIG. 8.9

Example of incorporating domain knowledge from pathologists into the loop of hashing model updating.

In the future, we will focus on the intelligent interaction and visualization that integrates expert feedback and automated algorithms for efficient decision making and provides a comprehensive understanding of the query results and supports semantic interaction functions. Interaction and visualization is another important yet challenging tool for effective computer-aided diagnosis and medical data mining. To achieve our ultimate goal of assisting efficient decision making and reasoning using medical image databases, we plan to incorporate users in the loop to incorporate the domain knowledge of experts, as shown in Fig. 8.9. While the automated methods are designed to process millions of images, human users can only reasonably work with much fewer images at a time. The main challenge will be bridging the gap between the large-scale automated algorithms and the knowledge that domain experts can provide, but at much smaller scales. We plan to design a visual analysis system with a set of feature-based query, visualization, comparison, and learning methods for revealing the relevant image features and relationships. This system will support the analysis of the retrieved relevant image sets, extracted image features, and feature similarities among the retrieved image sets, and will provide efficient interaction methods to enhance the query algorithms and obtain finer-tuned results. To summarize, the components of large-scale retrieval and intelligent

interaction will be coordinated for the purpose of scalable and interactive mining to provide a semantic interface between users and data through the language of feature similarities. The overall framework will be designed to address the challenges of both *scalable and interactive* mining within medical imaging informatics, and each aspect of the design and development will be driven by the goals of efficiency, robustness, and effective integration of user input.

REFERENCES

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

Andoni, A., Indyk, P., 2006. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In: IEEE Symposium on Foundations of Computer Science (FOCS), Berkeley, CA.

Basavanhally, A.N., Ganesan, S., Agner, S., Monaco, J.P., Feldman, M.D., Tomaszewski, J.E., Bhanot, G., Madabhushi, A., 2010. Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. IEEE Trans. Biomed. Eng. 57 (3), 642–653.

Caicedo, J.C., Cruz, A., Gonzalez, F.A., 2009. Histopathology image classification using bag of features and kernel functions. In: Artificial Intelligence in Medicine. Springer, New York, pp. 126–135.

Comaniciu, D., Meer, P., Foran, D.J., 1999. Image-guided decision support system for pathology. Mach. Vis. Appl. 11 (4), 213–224.

Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A., 2012. A boosted Bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE Trans. Biomed. Eng. 59 (5), 1205–1218.

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

Dy, J.G., Brodley, C.E., Kak, A., Broderick, L.S., Aisen, A.M., 2003. Unsupervised feature selection applied to content-based retrieval of lung images. IEEE Trans. Pattern Anal. Mach. Intell. 25 (3), 373–378.

El-Naqa, I., Yang, Y., Galatsanos, N.P., Nishikawa, R.M., Wernick, M.N., 2004. A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Trans. Med. Imaging 23 (10), 1233–1244.

Foran, D.J., Yang, L., et al., 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.

Goldberger, J., Hinton, G.E., Roweis, S.T., Salakhutdinov, R., 2004. Neighbourhood components analysis. In: Advances in Neural Information Processing Systems, pp. 513–520.

Greenspan, H., Pinhas, A.T., 2007. Medical image categorization and retrieval for PACS using the GMM-KL framework. IEEE Trans. Inform. Technol. Biomed. 11 (2), 190–202.

Hanbury, A., Müller, H., Langs, G., Menze, B.H., 2013. Cloud-based evaluation framework for big data. In: FIA Book 2013, LNCS volume. Springer, New York.

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

Jiang, M., Zhang, S., Liu, J., Shen, T., Metaxas, D.N., 2014. Computer-aided diagnosis of mammographic masses using vocabulary tree-based image retrieval. In: IEEE International Symposium on Biomedical Imaging. ISBI, pp. 1123–1126.

Jiang, M., Zhang, S., Huang, J., Yang, L., Metaxas, D.N., 2015a, Joint kernel-based supervised hashing for scalable histopathological image analysis. In: Medical Image Computing and Computer-Assisted Intervention. Springer, New York, pp. 366–373.

Jiang, M., Zhang, S., Li, H., Metaxas, D.N., 2015b. 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.

Langs, G., Müller, H., Menze, B.H., Hanbury, A., 2013. VISCERAL: towards large data in medical imaging—challenges and directions. In: MCBR-CDS MICCAI workshop, LNCS vol. 7723. Springer, New York.

Liu, W., Wang, J., Kumar, S., Chang, S.F., 2011. Hashing with graphs. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 1–8.

Liu, W., Wang, J., Ji, R., Jiang, Y.G., Chang, S.F., 2012. Supervised hashing with kernels. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 2074–2081.

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

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

Müller, H., Geissbühler, A., Ruch, P., 2005. ImageCLEF 2004: combining image and multi-lingual search for medical image retrieval. In: Multilingual Information Access for Text, Speech and Images. Springer, New York, pp. 718–727.

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

Nister, D., Stewenius, H., 2006. Scalable recognition with a vocabulary tree. In: IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2161–2168.

Schnorrenberg, F., Pattichis, C., Schizas, C., Kyriacou, K., 2000. Content-based retrieval of breast cancer biopsy slides. Tech. Health Care 8 (5), 291–297.

Sertel, O., Kong, J., Catalyurek, U.V., Lozanski, G., Saltz, J.H., Gurcan, M.N., 2009. Histopathological image analysis using model-based intermediate representations and color texture: follicular lymphoma grading. J. Signal Process. Syst. 55 (1–3), 169–183.

Sivic, J., Zisserman, A., 2003. Video google: a text retrieval approach to object matching in videos. In: IEEE International Conference on Computer Vision, pp. 1470–1477.

Song, Y., Cai, W., Feng, D., 2011. Hierarchical spatial matching for medical image retrieval. In: ACM International Workshop on Medical Multimedia Analysis and Retrieval, pp. 1–6.

Tabesh, A., Teverovskiy, M., Pang, H.Y., 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.

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

Wang, J., Liu, W., Kumar, S., Chang, S.F., 2015. Learning to hash for indexing big data—a survey. arXiv 1509.05472

Weiss, Y., Torralba, A., Fergus, R., 2009. Spectral hashing. In: Advances in Neural Information Processing Systems, pp. 1753–1760.

Weyn, B., van de Wouwer, G., van Daele, A., Scheunders, P., van Dyck, D., van Marck, E., Jacob, W., 1998. Automated breast tumor diagnosis and grading based on wavelet chromatin texture description. Cytometry 33 (1), 32–40.

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. Inform. Technol. Biomed. 13 (4), 636–644.

Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D., 2012. Query specific fusion for image retrieval. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (Eds.), European Conference on Computer Vision, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 660–673.

Zhang, X., Liu, W., Zhang, S., 2014a, Mining histopathological images via hashing-based scalable image retrieval. In: IEEE International Symposium on Biomedical Imaging, pp. 1111–1114.

Zhang, X., Yang, L., Liu, W., Su, H., Zhang, S., 2014b, Mining histopathological images via composite hashing and online learning. In: International Conference on Medical Image Computing and Computer Assisted Intervention. Springer, New York, pp. 479–486.

Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D.N., 2015a. 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., 2015b. Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis. IEEE J. Biomed. Health Inform. PP (99), 1–1.

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

Zhang, X., Su, H., Yang, L., Zhang, S., 2015d, Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 5361–5368.

Zhang, X., Su, H., Yang, L., Zhang, S., 2015e, Weighted hashing with multiple cues for cell-level analysis of histopathological images. In: Information Processing in Medical Imaging, pp. 303–314.

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

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

Zhou, X.S., Zillner, S., Moeller, M., Sintek, M., Zhan, Y., Krishnan, A., Gupta, A., 2008. Semantics and CBIR: a medical imaging perspective. In: ACM International Conference on Content-Based Image and Video Retrieval, pp. 571–580.