FUSING HETEROGENEOUS FEATURES FOR THE IMAGE-GUIDED DIAGNOSIS OF INTRADUCTAL BREAST LESIONS

Xiaofan Zhang[1], *Hang Dou*[2], *Tao Ju*[2], *Shaoting Zhang*[1]

[1]Department of Computer Science, University of North Carolina at Charlotte, NC, USA
[2]Department of Computer Science and Engineering, Washington University in St. Louis, MO, USA

ABSTRACT

In the analysis of histopathological images, both holistic (e.g., architecture features) and local appearance features demonstrate excellent performance, while their accuracy may vary dramatically among different inputs. This motivates us to investigate how to fuse results from these features to further enhance the accuracy. Particularly, we employ content-based image retrieval approaches to discover morphologically relevant images for image-guided diagnosis, using both holistic and local features. However, because of the dramatically different characteristics and representations of these heterogenous features, their resulting ranks may have no intersection among the top candidates, causing difficulties for traditional fusion methods. In this paper, we employ graph-based query-specific fusion approach where multiple retrieval ranks are integrated and reordered by conducting link analysis on a fused graph. The proposed method is capable of adaptively combining the strengths of local or holistic features for different queries, and does not need any supervision. We evaluate our method on a challenging clinical problem, i.e., histopathological image-guided diagnosis of intraductal breast lesions, and it achieves 91.67% classification accuracy on 120 breast tissue images from 40 patients.

Index Terms— histopathological image analysis, breast lesion, image retrieval, hashing, fusion

1. INTRODUCTION

Recently, digitized tissue histopathology for microscopic examination and automatic disease grading has become amenable to the application of computerized image analysis and computer-aided diagnosis [5]. Many methods have been proposed to tackle this important and challenging use case, by investigating object level and spatially related features and employing learning-based classifiers [2] or content-based image retrieval (CBIR) [14, 6]. In general, accurate analysis of histopathological images requires to examine cell-level information for accurate diagnosis, including individual cells (e.g., appearance and shapes [3]) and architecture of tissue (e.g., topology and layout of all cells [1]). These features cover both local and holistic information, all benefiting the diagnosis accuracy of histopathological images. Therefore, the complementary capability of local and holistic features naturally raises the question of how 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 [1] are represented as a low-dimensional vector of statistics, while local feature can be represented as high-dimensional bag-of-words (BoW) [13] and compressed as binary codes to improve the efficiency [16, 7].

Generally, fusion can be carried out on the feature or rank-levels. In our context (i.e., differentiation of cancers), this means to combine different types of features in a histogram for learning-based classification, or to fuse the ordered results from CBIR methods [4, 15] and then classify via majority voting, both of which are fundamental problems. Unfortunately, most existing fusion methods still have limitations for medical image analysis, especially in terms of the robustness, scalability and generality. For example, feature-level fusion usually produces a new feature vector that has a higher dimensionality, e.g., concatenation of feature vectors. However, when these features are heterogeneous (e.g., having significantly different dimensions and characteristics such as low-dimensional architecture feature [1] and high-dimensional appearance feature [16] in histopathological image analysis), feature-level fusion may not be able to effectively integrate their strengths. On the other hand, rank-level fusion usually needs to decide which features should play a major role in the retrieval, which is quite difficult to determine online for a specific input and a large database. Therefore, it is necessary to employ relatively principled ways to fuse multiple (heterogeneous) features extracted from histopathological images.

In this paper, we focus on the rank-level fusion of local and holistic features for the image-guided diagnosis of breast cancer, i.e., differentiation of the benign (the usual ductal hyperplasia) and actionable (the atypical ductal hyperplasia and ductal carcinoma in situ) cases. Particularly, we use content-based 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 (i.e., retrieval results) obtained from different features, a data-driven and graph-based method is employed for accurate, robust and efficien-

978-1-4799-2374-8/15/$31.00 ©2015 IEEE

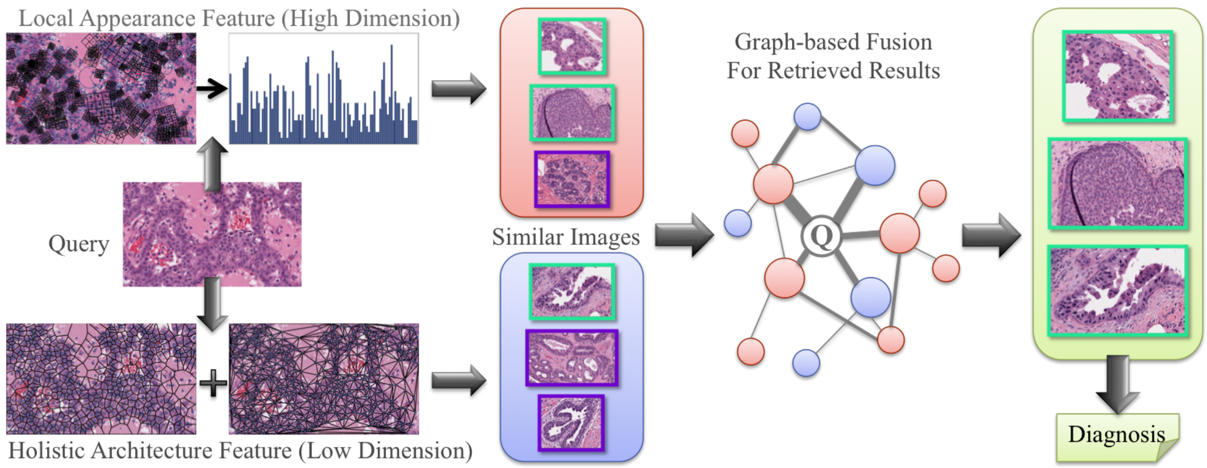

Fig. 1. Overview of the proposed framework. Both holistic architecture feature and local appearance feature are extracted and employed for image retrieval. The retrieval results are fused via the graph-based framework to improve the accuracy. Note that majority voting does not work in this example, since two ranks have no intersection.

t fusion, by evaluating the quality of each rank online [15]. This method provides a unique angle for the fusion of multiple types of information extracted from histopathological images, and is very different from traditional fusion approaches based on voting or combining similarity scores. We validate this proposed method on 120 breast tissue images from 40 patients. The experimental results demonstrate the accuracy and efficiency of our framework.

2. METHODOLOGY

Overview: Fig. 1 shows the overview of our framework. From detected cells, we extract both holistic architecture features [1] and high-dimensional local appearance features [16] (i.e., 10,000 dimensions), both of which are used for image retrieval. To ensure the computational efficiency and scalability, the high-dimensional feature is compressed as tens of hash bits [16, 7]. Combining these complementary features is an intuitive approach to improve the accuracy. However, directly combining them at the feature-level may not be effective due to dramatically different representations. An alternative is to fuse them at the rank-level, i.e., retrieved images. The critical issue is how to measure and compare the quality of ranks on the fly, since fusion process should favor the rank with higher quality. As the similarity scores of retrieved results may vary largely among queries and are not comparable between different ranks, a reasonable approach is to measure the *consistency* among the top candidates. Therefore, for each query image, we construct a weighted undirected graph from the retrieval results of one rank, where the retrieval quality or the relevance is modeled by the weights on the edges. These weights are determined by the overlap ratio (i.e., Jaccard similarity coefficient) of two neighborhood image sets. Then we fuse multiple graphs to one and perform a localized PageRank algorithm [10] to rerank the retrieval results according to their probability distribution. As a result, the fused retrieval results tend to be consistent among different feature representations.

Fusion of Heterogeneous Features: The rank-level fusion of heterogeneous features includes graph construction, graph consolidation, and sub-graph selection [15].

Graph Construction: Given a list of ranked results or retrieval results by one type of features, i.e., the architecture or appearance feature, we assume that the consensus degree among the top candidates reveals the retrieval quality. Therefore, we first build a weighted graph using the constraints derived from the consensus degree. Setting the query as the graph centroid, we use its k-nearest neighbors (kNN) as the first layer of nodes in the graph, and kNN of kNN as the second layer. Note that this setting is different from traditional methods using reciprocal kNN, since such information is usually not available for medical image analysis, i.e., query is not included in the database. Neighboring nodes are connected by edges, whose weight can be defined as the ratio of their common neighbors, i.e., Jaccard similarity, which reflects the confidence of including the connected nodes into the retrieval results. The weight between node i and i' is defined as:

$$w(i,i') = J(i,i') = \frac{|N_k(i) \cap N_k(i')|}{|N_k(i) \cup N_k(i')|} \quad (1)$$

where $|\cdot|$ denotes the cardinality, $N_k(i)$ and $N_k(i')$ include the images that are the top-k retrieved candidates using i and i' as the query, respectively. The range of edge weights is from 0 to 1, with $J(i,i') = 1$ implying that these two histopathological images share exactly the same set of neighbors, in which case we assume that they are highly likely to be similar.

Graph Consolidation: Multiple graphs, denoted as $G^m = (V^m, E^m, w^m)$, are constructed from the retrieved results of holistic and local features. They can be fused together in a natural way, by appending new nodes or consolidating edge weights of existing nodes in the resulting graph:

$$G = (V, E, w), \text{with } V = \cup_m V^m, E = \cup_m E^m,$$
$$\text{and } w(i, i') = \sum_m w^m(i, i') \qquad (2)$$

where $w^m(i, i') = 0$ for $(i, i') \notin E^m$. The rational of this fusion process is that though the rank lists or the similarity scores in different methods or features are not directly comparable, their Jaccard coefficients are comparable as they reflect the consistency of two nearest neighborhoods. In other words, this measure of consensus degree does not rely on the similarity scores or any supervision, so it can be used and compared for different retrieval results from holistic and local features, ensuring the generality.

Sub-Graph Selection: After the candidates from both holistic and local features are fused via the graph consolidation, we need to rank them as per the relevance and select the most similar ones. This can be achieved by conducting a link analysis on the resulting graph to search for the PageRank vector [10], in which the resulting graph is treated as a network. Since this network is built by considering the retrieval relevance, naturally a node is more important or relevant if it has a higher probability to be visited. To compute the equilibrium state of the graph, we define the $|V| \times |V|$ transition matrix $\mathbf{P}$ as $P_{ii'} = w(i, i')/\deg(i)$ for $(i, i') \in E$, and 0 otherwise. This matrix is row-stochastic, with the summation of each row as one. In the *intelligent surfer model* [11], a "surfer" probabilistically hops along the edges of G to different nodes, according to the transition matrix $\mathbf{P}$. Occasionally, with a small probability $1 - \beta$, the surfer jumps according to a fixed distribution over nodes π, which we set as $\pi_q = 0.99$ and uniform otherwise, where q is the index of the query node. We denote p_i^t as the probability for the surfer to be at node i at a time t and $p^t = (p_i^t)$. The equilibrium state of p, where a higher probability reflects a higher relevance to the query, is obtained by the query-dependent PageRank vector as a stationary point using the power method:

$$p^{t+1} = (1 - \beta)\pi + \beta \mathbf{P}^T p^t. \qquad (3)$$

Once p has converged, the histopathological images are ranked according to their probabilities in p, where a higher probability reflects a higher relevance to the query in this equilibrium state of the graph. Using fused results, i.e., a new list of histopathological images from both features, majority voting can be employed for cancer differentiation. To summarize, fusing heterogeneous features via graphs can significantly improve the performance of each individual feature, without sacrificing the scalability and generality.

3. EXPERIMENTS

Histopathological images of breast-tissue 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). All the slides were imaged using a ScanScope digitizer (Aperio, Vista, CA) available in the tissue archival service at IUHPL. 120 images (around 2250K pixels for each image) were gathered from 40 patients, 3 images per patient. 20 of these patients were labeled as benign and others are actionable, based on the majority diagnosis of nine board-certified pathologists. Leave-one-patient-out validation is used to evaluate the accuracy of classification.

We employ two features (holistic and local) as the baseline for fusion. For holistic feature [1], the Voronoi diagram, Delaunay triangulation, minimum spanning tree are constructed and the nuclear density features are computed to model "architecture" of breast tissue, resulting in a 48-dimensional feature vector for each image. For local feature, 1500 to 2000 SIFT descriptors [8] are extracted from each image to describe the cell appearance. These descriptors are quantized into sets of cluster centers using bag-of-words [13], in which the feature dimension equals the number of clusters. Specifically, we quantize them into high-dimensional feature vectors with length 10,000. For efficiency and scalability we compress the high dimensional feature into 48 binary bits with kernelized supervised hashing (KSH) algorithm [7]. Note that this binary representation is not compatible with the holistic feature. We first evaluate the performance of image retrieval using single feature such as the holistic feature, high-dimensional local feature and compressed binary feature. kNN and Support Vector Machine (SVM) are used as the baselines that have been widely employed for histopathological image analysis [1, 12], where $k = 9$ and an RBF kernel with optimized gamma value for SVM.

As shown in Fig. 2, both holistic and local features are able to generate reasonable results, i.e., around 80% accuracy. The only exception is that kNN fails in handling high-dimensional local feature, achieving only 74.17% accuracy. After compression with KSH, the binary codes improve the accuracy to 81.67%. In addition, using hashing representation also significantly improves the computational efficiency, i.e., thousands times faster than using original high-dimensional features, ensuring the scalability. Since both features are fairly effective but not perfect, and they should be complementary as they model different scales of information, it is natural to combine them for higher accuracy.

We compare our graph fusion method with several classical methods for fusion, including both feature and rank-level approaches. For feature-level fusion, we normalize and concatenate different features into a histogram and classify them with either kNN or SVM. Since the dimensions of features are largely different, it is not likely to obtain reasonable re-

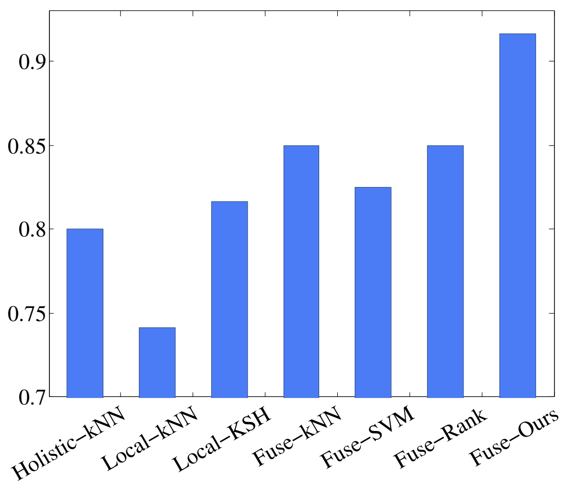

Fig. 2. Quantitative comparison of the classification accuracy. We compare the performance of each single feature, and the fusion of both holistic and local features.

sults without doing normalization. Therefore, normalization ensures that each feature contributes "equally" to the concatenated one [9]. For rank-level fusion, we combine different retrieval results via rank aggregation [4] and classify the query image with majority voting.

As shown in Fig. 2, concatenation of feature vectors marginally improve the classification accuracy, i.e., around 1-3% better than the baseline, due to the dramatically different characteristics of heterogeneous features. kNN is slightly better than SVM, because of the normalization on the distance-level [9]. On the other hand, rank aggregation also merely improves the accuracy by 3%, since there may be no intersection among the top candidates retrieved by the local and holistic features. Our graph fusion method determines online which features should play a major role in the retrieval, in an unsupervised scheme. As a result, our fusion of heterogeneous features significantly improves the accuracy by around 10%, i.e., achieving 91.67% overall accuracy on this challenging problem. In addition, since this fusion process is applied on the retrieved results, i.e., a small subset of the whole dataset, it is very efficient and only takes milliseconds, ensuring sound scalability.

4. CONCLUSIONS

In this paper, we investigate the fusion of heterogenous features for histopathological image analysis. Specifically, we employ a graph-based framework to fuse the holistic architecture feature and the local appearance feature that is represented as hashing code. These features are complementary but have dramatically different characteristics and representations, causing difficulties for traditional fusion methods. Our framework is able to measure online the retrieval quality by the consistency of the neighborhoods of candidate images, without using any supervision. Therefore, the fused results significantly improve the baseline using single feature. In the future, we will test our method on larger dataset (e.g., thousands of images) and employ more features for fusion. We are also interested in the relevance feedback that takes domain experts' feedback to improve the retrieval and fusion methods.

5. REFERENCES

[1] A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, and A. Madabhushi. Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology. *TBME*, 57(3):642–653, 2010.

[2] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski. Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In *ISBI*, pages 496–499, 2008.

[3] M. M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel, and M. N. Gurcan. Computerized classification of intraductal breast lesions using histopathological images. *TBME*, 58(7):1977–1984, 2011.

[4] R. Fagin, R. Kumar, and D. Sivakumar. Efficient similarity search and classification via rank aggregation. In *ACM SIGMOD*, pages 301–312. ACM, 2003.

[5] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener. Histopathological image analysis: A review. *TBME*, 2:147–171, 2009.

[6] G. Langs, H. Müller, B. H. Menze, and A. Hanbury. VISCERAL: Towards large data in medical imaging - challenges and directions. In *MCBR-CDS MICCAI workshop*, volume 7723 of *Springer LNCS*, 2013.

[7] W. Liu, J. Wang, R. Ji, Y.-G. Jiang, and S.-F. Chang. Supervised hashing with kernels. In *CVPR*, pages 2074–2081, 2012.

[8] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, Nov. 2004.

[9] A. Makadia, V. Pavlovic, and S. Kumar. Baselines for image annotation. *IJCV*, 90(1):88–105, 2010.

[10] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. 1999.

[11] M. Richardson and P. Domingos. The intelligent surfer: Probabilistic combination of link and content information in pagerank. In *NIPS*, pages 1441–1448, 2001.

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

[13] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *ICCV*, 2003.

[14] L. Yang, W. Chen, P. Meer, G. Salaru, L. A. Goodell, V. Berstis, and D. J. Foran. Virtual microscopy and grid-enabled decision support for large-scale analysis of imaged pathology specimens. *TITB*, 13(4):636–644, 2009.

[15] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas. Query specific rank fusion for image retrieval. *TPAMI*, 2014.

[16] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang. Towards large-scale histopathological image analysis: Hashing-based image retrieval. *TMI*, 34(2):1–11, 2015.