LEVERAGING COUPLED MULTI-INDEX FOR SCALABLE RETRIEVAL OF MAMMOGRAPHIC MASSES

Menglin Jiang[*] *Shaoting Zhang*[†] *Ruogu Fang*[‡] *Dimitris N. Metaxas*[*]

[*] Department of Computer Science, Rutgers University, Piscataway, NJ, USA
[†] Department of Computer Science, UNC Charlotte, Charlotte, NC, USA
[‡] School of Computing and Information Sciences, Florida International University, Miami, FL, USA

ABSTRACT

Content-based image retrieval techniques have shown great value in computer-aided diagnosis of mammographic masses. Many existing approaches adopt several features to better characterize mammographic regions. However, most of them fuse features through feature concatenation or result-level combination, which cannot fully exert the discriminative power of different features and also sacrifices the overall computational efficiency. To address these drawbacks, we propose to utilize coupled multi-index for index-level feature fusion. Specifically, complementary local features are extracted from the same locations of mammographic regions. Then, they are separately quantized using the "bag of words" (BoW) approach. Finally, quantized features are inserted into a two-dimensional inverted index, with each feature corresponding to one dimension. Experiments are carried out on a large dataset constructed from the digital database for screening mammography (DDSM). Results demonstrate that our approach not only achieves better retrieval precision and diagnostic accuracy than individual features do, but also improves the overall efficiency and scalability.

Index Terms— Mammography, breast masses, image retrieval, feature fusion, coupled multi-index

1. INTRODUCTION

For years, mammography has played a key role in diagnosis of breast cancer, which is the second leading cause of cancer-related death among women [1]. Unfortunately, as a major indicator of breast cancer, mammographic masses are very difficult to diagnose due to their variety in shape, margin, size and obscure boundaries. Consequently, a considerable portion of retrospectively visible masses is missed [2], and biopsies are frequently conducted on normal tissues [3]. To facilitate the diagnosis of mammographic masses, many content-based image retrieval (CBIR) methods have been proposed. Specifically, they first prompt radiologists to label a region of interest (ROI) in the query mammogram, then compare it with database ROIs extracted from previously diagnosed cases, and finally return the most similar cases along with the likelihood of a mass in the query ROI.

In order to describe mammographic regions from variant perspectives, many retrieval methods employ multiple features, such as features related to intensity, texture, and shape properties. While showing a promising direction for better retrieval results, they raise the question of how to consolidate different features. Most of existing methods in the literature solve this problem in two ways. The first category, such as [4–7], simply concatenates several feature vectors to form a new feature. The second category, such as [8–10], first conducts similarity searching using individual features and then integrates their results. Nevertheless, either category has its limitations. In particular, feature concatenation is not suitable for a heterogeneous collection of features even with feature normalization. For instance, variant features may not agree on the adopted distance measure during feature matching, and even if they do, the feature with large distance will dominate the matching process. Result-level fusion, on the other hand, usually introduces parameters such as weight of each feature, which need to be elaborately tuned and may not generalize well to different datasets. Moreover, both categories of methods compromise the computational efficiency and scalability, since the overall processing time will be at least the sum of time required by each feature. This could be a serious problem when many features are exploited to retrieve from a large database.

To overcome the above drawbacks, we propose to fuse complementary local features with coupled multi-index [11]. A coupled multi-index is basically a multi-dimensional inverted index, where each dimension corresponds to one feature. Our method exploits a two-dimensional inverted index to unify two features, namely scale-invariant feature transform (SIFT) [12] and local intensity histogram [13]. The overview of our approach is shown in Fig. 1. Specifically, SIFT and local intensity histogram are extracted from database ROIs and quantized according to the "bag of words" (BoW) method [14]. Then, all the quantized feature couples are inserted into coupled multi-index. Given a query ROI, features are extracted and quantized in the same way, and then searched in the coupled multi-index to find similar database ROIs. These ROIs vote to determine whether the

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

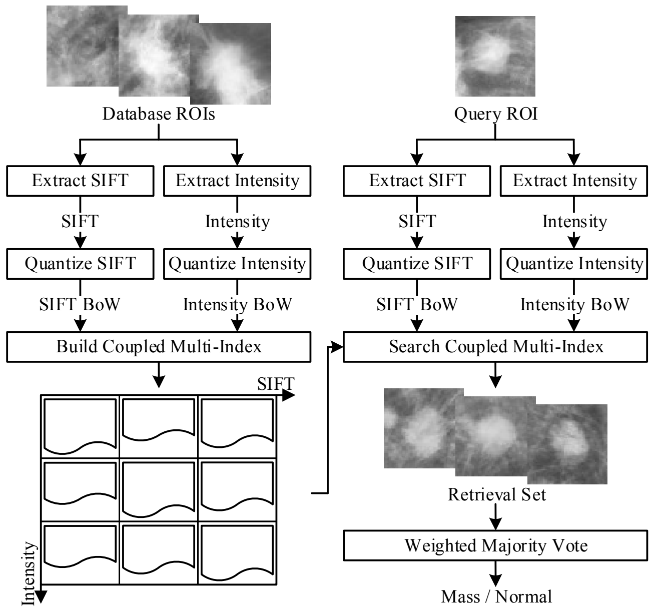

Fig. 1. Overview of the proposed approach.

query contains a mass or not [15].

Our approach has many advantages over current methods. First, it could handle intrinsically different local features since they are quantized separately. Second, no parameters are added and the tuning issue is bypassed. Third, by decreasing the portion of database features that are considered during feature searching, the presented approach improves the overall efficiency and scalability. Last but not least, the index-level fusion also serves as an index, saving the trouble of implementing an extra indexing scheme.

2. PROPOSED APPROACH

In this section, the two features employed in our approach are first described. Then, we revisit the traditional BoW model and inverted index, taking SIFT as an example. Finally, we present coupled multi-index, which incorporates two features at index level. After obtaining the most similar database ROIs along with their similarity scores to query ROI, weighted majority vote [15] is carried out to classify the query ROI as a mass or normal tissue.

Complementary Features: The primary feature chosen in our method is SIFT [12]. SIFT has been successfully applied to medical image retrieval and analysis [15–18], owing to its excellent robustness and discriminative power. Briefly speaking, a series of SIFT features are extracted from a mammographic ROI as follows. Scale-invariant keypoints are first 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.

As for the secondary feature, we adopt local intensity histogram [13]. Intensity-based features, such as mean and standard deviation of intensity values, have been widely used in mammographic mass retrieval methods [5, 8]. Specifically, our approach calculates a 16-dimensional intensity histogram for each SIFT surrounding region, coupling the SIFT feature.

BoW and Inverted Index: Since it is computational infeasible to match vectorial SIFT features in large-scale image retrieval, BoW model [14] is adopted to quantize these features. To this end, a large set of SIFT features extracted from a separate database are used to train a visual vocabulary through k-means clustering. Each cluster center is referred to as a "visual word", which is an analogue of "word" in text retrieval. All the visual words form a "visual vocabulary" $\{\boldsymbol{u}_i\}_{i=1}^{k_S}$, where k_S is the size of the vocabulary.

The SIFT features extracted from database ROIs are quantized using $\{\boldsymbol{u}_i\}_{i=1}^{k_S}$. In particular, a SIFT feature vector, $\boldsymbol{x}^S$, is mapped to the ID of its nearest visual word using a quantization function f^S:

$$f^S\left(\boldsymbol{x}^S\right) = \arg\min_i \left\|\boldsymbol{x}^S - \boldsymbol{u}_i\right\|_2. \quad (1)$$

Thus each ROI is described by a series of visual words, from which comes the name "bag of words".

The quantized database features naturally generate a forward index, which lists the words per ROI. To further accelerate the searching process, the forward index is then "inverted" to develop an inverted index, which lists the ROIs per word. Specifically, the inverted index consists of k_S files, and the i-th file records the IDs of database ROIs containing word i.

In addition, the importance of each visual word is measured using term frequency-inverse document frequency (TF-IDF) [19] or its variations. TF of the i-th word in a database ROI d, denoted as $tf^S(i,d)$, represents the number of features extracted from d that are quantized to word i. IDF of the i-th word in the entire database, denoted as $idf^S(i)$, is normally defined using entropy:

$$idf^S(i) = \log \frac{N}{N_i^S}, \quad (2)$$

where N is the total number of database ROIs, and N_i^S is the number of ROIs with at least one SIFT feature quantized to word i. The definition of IDF reflects the observation that visual words occurring frequently in the database are less informative. TF-IDF can also be employed to define the l_2 norm of d:

$$\|d\|_2^S = \sqrt{\sum\nolimits_i \left(tf^S(i,d)\right)^2}, \quad (3)$$

which reflects the number of visual words in d. The TF scores are stored in the inverted index.

For a given query ROI q, its SIFT features are quantized and TF scores are calculated in the above manner. The quantized features are then searched from the inverted index to get the most similar database ROIs. It is worth mentioning that, during the searching process, only those inverted files corresponding to the query visual words are considered. What's more, the portion of involved inverted files can be further decreased by increasing vocabulary size. Compared with forward index, which searches from all the database features, inverted index promotes the searching efficiency dramatically. The similarity score between q and any database ROI d is calculated based on how similar their visual words are:

$$s^S(q,d) = \frac{\sum_i tf^S(i,q) \cdot tf^S(i,d) \cdot \left(idf^S(i)\right)^2}{\|q\|_2^S \cdot \|d\|_2^S}. \quad (4)$$

Note that the visual words are weighted using TF-IDF, and the normalization factor $1\Big/\left(\|q\|_2^S \cdot \|d\|_2^S\right)$ is used to achieve fairness between database ROIs with few and many features.

Coupled Multi-Index: Similar to SIFT, local intensity histogram can also be quantized following the BoW approach. The visual vocabulary, feature vector, and quantization function are denoted as $\{\boldsymbol{v}_j\}_{j=1}^{k_I}$, $\boldsymbol{x}^I$, and f^I respectively, with k_I representing the vocabulary size.

After obtaining f^S and f^I, a couple of SIFT and local intensity histogram extracted from the same SIFT keypoint, $\boldsymbol{x} = \left(\boldsymbol{x}^S, \boldsymbol{x}^I\right)$, is quantized to a pair of visual word IDs:

$$f(\boldsymbol{x}) = \left(f^S\left(\boldsymbol{x}^S\right), f^I\left(\boldsymbol{x}^I\right)\right). \quad (5)$$

Quantized database features are inserted to a coupled multi-index, which is illustrated in Fig. 2. The coupled multi-index is basically a matrix of inverted files, with two dimensions corresponding to SIFT and intensity visual words. The (i,j)-th inverted file stores the IDs of ROIs whose SIFT and intensity features are quantized to $\boldsymbol{u}_i$ and $\boldsymbol{v}_j$ respectively. Apparently, a traditional inverted index can be regarded as one dimension of a coupled multi-index.

TF-IDF and l_2 norm could be easily generalized to coupled multi-index. More specifically, $tf(i,j,d)$ is defined as the number of feature couples in d that are quantized word (i,j). $idf(i,j)$ and $\|d\|_2$ are defined similar to Eq. (2) and Eq. (3):

$$idf(i,j) = \log \frac{N}{N_{i,j}}, \quad (6)$$

$$\|d\|_2 = \sqrt{\sum\nolimits_{i,j} tf^2(i,j,d)}, \quad (7)$$

where $N_{i,j}$ is the number of ROIs with at least one feature couple quantized to word (i,j).

Given a query ROI q, its quantized feature couples are searched from the coupled multi-index, considering only the inverted files corresponding to q's visual words. The similarity score between q and any database ROI d is computed as follows:

$$s(q,d) = \frac{\sum_{i,j} tf(i,j,q) \cdot tf(i,j,d) \cdot idf^2(i,j)}{\|q\|_2 \cdot \|d\|_2}. \quad (8)$$

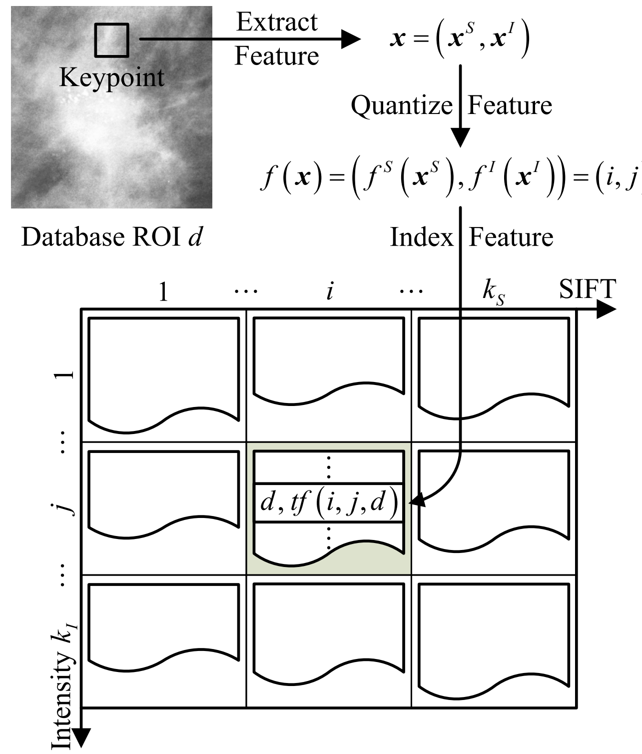

Fig. 2. Illustration of coupled multi-index.

It is noteworthy that coupled multi-index substantially outperforms traditional inverted index with regard to retrieval precision and efficiency. Generally speaking, as the vocabulary size increases, visual words become more discriminative, leading to better retrieval precision. Besides, larger number of inverted files means that less database features need to be considered during feature searching. Thus, the searching speed is accelerated. Unfortunately, inverted index could not afford a large vocabulary, since its computational complexity for feature quantization is linear in vocabulary size, i.e. $O(k_S)$ for SIFT and $O(k_I)$ for local intensity histogram. Coupled multi-index solves this dilemma: it enjoys the benefits of vocabulary size $k_S \cdot k_I$ with a small computational overhead of $O(k_S + k_I)$ for feature quantization.

3. EXPERIMENTS

The experiments are carried out on the dataset constructed in [15] based on the digital database for screening mammography (DDSM) [20]. To the best of our knowledge, this is the largest dataset in mammographic mass retrieval. In particular, 1,840 mass ROIs and 8,713 CAD-generated false positive ROIs (10,553 ROIs in total) form a database, another 500 mass ROIs and 500 false positive ROIs are used as queries.

Three baseline systems are implemented for comparison. The first two methods exploit traditional inverted index on SIFT and local intensity histogram respectively. SIFT [12,15–18] and intensity-based features [5, 8, 13] have demonstrat-

Table 1. Retrieval precision

	SIFT	Intensity	ResultFusion	MultiIndex
Mass	78.3%	74.5%	83.9%	**86.9%**
Normal	81.3%	76.7%	83.5%	**89.3%**
Total	79.8%	75.6%	83.7%	**88.1%**

Table 2. Classification accuracy

	SIFT	Intensity	ResultFusion	MultiIndex
Mass	80.4%	75.3%	85.2%	**89.0%**
Normal	82.6%	78.5%	86.0%	**91.6%**
Total	81.5%	76.9%	85.6%	**90.3%**

ed good performance in medical image retrieval and analysis. The third approach combines the results of the first two methods. Similar to [8,9], the similarity score $s(q, d)$ is a weighted sum of $s^S(q, d)$ and $s^I(q, d)$. These compared methods and the proposed approach are referred to as "SIFT", "Intensity", "ResultFusion" and "MultiIndex" in the following analysis.

First of all, *retrieval precision* is evaluated, which is defined as the percentage of retrieved database ROIs that are relevant to query ROI. The precision scores of all the methods at top 10 retrievals are summarized in Table 1. An example is provided in Fig. 3 for visual evaluation. The results show that our method remarkably outperforms SIFT and Intensity, and also surpasses ResultFusion.

Second, *classification accuracy* is measured, which refers to the percentage of query ROIs that are correctly classified. Remember that a query ROI is classified as mass or normal tissue according to a weighted majority vote of its retrieved database ROIs [15]. The accuracy scores at top 10 retrievals are reported in Table 2. Once again, our method consistently outperforms all the baseline approaches. Furthermore, the classification accuracy is systematically higher than the retrieval precision, since irrelevant retrievals would not cause a misclassification as long as they remain a minority of the retrieval set. Especially, the proposed approach achieves a satisfactory classification accuracy of 90.3%.

Finally, efficiency and scalability are investigated using the *processing time* needed to retrieve a query ROI. As shown in Table 3, "BoW" means the time cost of feature extraction and quantization, "Search" indicates the time needed to search quantized feature from inverted index or multi-index, and "Total" is the entire processing time, i.e. sum of "BoW" and "Search". Apparently, the time cost for ResultFusion is the sum of those for SIFT and Intensity. As for MultiIndex, since quantization of a feature couple is basically quantization of separate features, the time needed for BoW calculation would be the sum of those for SIFT and Intensity BoW calculation. Nevertheless, since its vocabulary size $k = k_S \cdot k_I$ is much larger than k_S and k_I, the number of database features considered in each search is much smaller, leading to shorter searching time. Altogether, efficiency and scalability are significantly improved using coupled multi-index. Moreover, as the database becomes larger, the ratio of searching time over total time will increase, and the advantage of multi-index will be further enhanced.

Table 3. Processing time (second)

	SIFT	Intensity	ResultFusion	MultiIndex
BoW	0.31	**0.22**	0.53	0.53
Search	0.83	1.05	1.89	**0.21**
Total	1.14	1.27	2.42	**0.74**

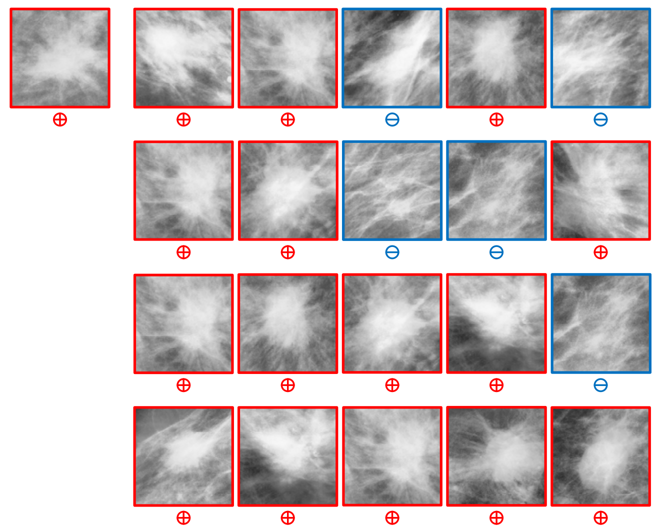

Fig. 3. A query ROI (left) and its top 5 retrieved database ROIs (right) returned by SIFT, Intensity, ResultFusion and MultiIndex (from top to bottom). For each ROI, its class is shown below. $\oplus$ stands for a mass, and $\ominus$ represents normal tissue

4. CONCLUSION

In mammographic mass retrieval, it is widely recognized that complementary features could lead to better performance. However, existing fusion methods may not fully integrate the strength of variant features, and they also sacrifice the computational efficiency and scalability. In this paper, we exploit coupled multi-index for index-level fusion of SIFT and local intensity histogram. The proposed approach is superior to current methods in many aspects. Specifically, it is suitable for heterogeneous local features, introduces no extra parameters, improves the overall efficiency and scalability, and serves as an index structure. Experiments are implemented on a dataset [15] built from DDSM [20], demonstrating the efficacy of our approach.

5. REFERENCES

[1] American Cancer Society, *Breast Cancer Facts & Figures 2013-2014*, American Cancer Society, Atlanta, 2013.

[2] R. L. Birdwell, D. M. Ikeda, K. F. O'Shaughnessy, and E. A. Sickles, "Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection," *Radiology*, vol. 219, no. 1, pp. 192–202, 2001.

[3] R. A. Hubbard, K. Kerlikowske, C. I. Flowers, B. C. Yankaskas, W. Zhu, and D. L. Miglioretti, "Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: A cohort study," *Ann. Intern. Med.*, vol. 155, no. 8, pp. 481–492, 2011.

[4] H. Alto, R. M. Rangayyan, and J. E. Leo Desautels, "Content-based retrieval and analysis of mammographic masses," *J. Electron. Imaging*, vol. 14, no. 2, pp. 023016–1–17, 2005.

[5] B. Zheng, A. Lu, L. A. Hardesty, J. H. Sumkin, C. M. Hakim, M. A. Ganott, and D. Gur, "A method to improve visual similarity of breast masses for an interactive computer-aided diagnosis environment," *Med. Phys.*, vol. 33, no. 1, pp. 111–117, 2006.

[6] F. Narváez, G. Díaz, and E. Romero, "Multi-view information fusion for automatic BI-RADS description of mammographic masses," in *Medical Imaging 2011: Computer-Aided Diagnosis*, 2011, vol. 7963 of *Proc. SPIE*, pp. 79630A–1–7.

[7] F. Narváez, G. Díaz, F. Gómez, and E. Romero, "A content-based retrieval of mammographic masses using a curvelet descriptor," in *Medical Imaging 2012: Computer-Aided Diagnosis*, 2012, vol. 8315 of *Proc. SPIE*, pp. 83150A–1–7.

[8] Y. Tao, S.-C. B. Lo, M. T. Freedman, and J. Xuan, "A preliminary study of content-based mammographic masses retrieval," in *Medical Imaging 2007: Computer-Aided Diagnosis*, 2007, vol. 6514 of *Proc. SPIE*, pp. 1–12.

[9] C.-H. Wei, Y. Li, and P. J. Huang, "Mammogram retrieval through machine learning within BI-RADS standards," *J. Biomed. Inform.*, vol. 44, no. 4, pp. 607–614, 2011.

[10] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas, "Query specific fusion for image retrieval," in *Proc. ECCV*, 2012, pp. 660–673.

[11] L. Zheng, S. Wang, Z. Liu, and Q. Tian, "Packing and padding: Coupled multi-index for accurate image retrieval," in *Proc. IEEE CVPR*, 2014, pp. 1947–1954.

[12] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[13] D. Shen, "Image registration by local histogram matching," *Pattern Recognit.*, vol. 40, no. 4, pp. 1161–1172, 2007.

[14] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. IEEE ICCV*, 2003, pp. 1470–1477.

[15] M. Jiang, S. Zhang, H. Li, and D. N. Metaxas, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 783–792, 2015.

[16] J. C. Caicedo, A. Cruz, and F. A. Gonzalez, "Histopathology image classification using bag of features and kernel functions," in *Proc. AIME*, 2009, pp. 126–135.

[17] B. André, T. Vercauteren, A. M. Buchner, M. B. Wallace, and N. Ayache, "Learning semantic and visual similarity for endomicroscopy video retrieval," *IEEE Trans. Med. Imaging*, vol. 31, no. 6, pp. 1276–1288, 2012.

[18] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Trans. Med. Imaging*, vol. 34, no. 2, pp. 496–506, 2015.

[19] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, 1988.

[20] M. Heath, K. Bowyer, D. Kopans, W. P. Kegelmeyer Jr, R. Moore, K. Chang, and S. Munishkumaran, "Current status of the digital database for screening mammography," in *Digital Mammography*, pp. 457–460. Springer Netherlands, 1998.