MAMMOGRAPHIC IMAGE CLASSIFICATION USING HISTOGRAM INTERSECTION

Erkang Cheng[1], Nianhua Xie[1], Haibin Ling[1], Predrag R. Bakic[3], Andrew D.A. Maidment[3], Vasileios Megalooikonomou[2]

[1]Center for Information Science and Technology, [2]Data Engineering Laboratory (DEnLab), Temple University, 1805 N. Broad St. Philadelphia, PA 19122, USA

[3]Department of Radiology, University of Pennsylvania, 3400 Spruce St., Philadelphia, PA 19104, USA

ABSTRACT

In this paper we propose using histogram intersection for mammographic image classification. First, we use the bag-of-words model for image representation, which captures the texture information by collecting local patch statistics. Then, we propose using normalized histogram intersection (HI) as a similarity measure with the K-nearest neighbor (KNN) classifier. Furthermore, by taking advantage of the fact that HI forms a Mercer kernel, we combine HI with support vector machines (SVM), which further improves the classification performance. The proposed methods are evaluated on a galactographic dataset and are compared with several previously used methods. In a thorough evaluation containing about 288 different experimental configurations, the proposed methods demonstrate promising results.

Index Terms**—**Texture descriptors, bag-of-words, Vector quantization, histogram intersection, classification, x-ray galactograms

1. INTRODUCTION

Texture analysis has been widely used in medical image tasks as well as related fields such as computer vision and pattern recognition. Recently, bag-of-words models, originally used in document analysis, have been successfully extended to image based classification tasks [13]. Despite their success in image retrieval and category classification, bag-of-words models have not been thoroughly studied for image-based diagnosis tasks.

In breast imaging, texture analysis is often desired to assess properties of the underlying breast ductal network. Texture patterns in these images are the radiographic effect of the underlying anatomical arrangement. Analysis of the mammographic texture, the *parenchymal pattern*, using various texture features, has demonstrated difference between the parenchymal pattern in women with low-risk of cancer and women with BRCA1/BRCA2 gene-mutations, which are associated with high risk [10]. On the other hand, analysis of such natural tree-like structures in biomedical images presents special challenges. For example, the surrounding tissue may obscure branching patterns. Galactography can be performed to visualize the breast ductal network by injecting a contrast agent into the lactiferous ducts of the breast. It is useful for visualizing early symptoms of papilloma or ductal ectasia, which cause spontaneous nipple discharge, without identifiable mammographic lesion [1].

Our goal in this paper is to perform texture analysis in particular regions of interests (ROIs) in galactographic images in order to assist diagnosis. We propose using the histogram intersection (HI) in the bag-of-words framework for this task. In this framework, we first represent an image by treating it as a set of local patches. These patches are then vector quantized according to a codebook learned from training images. Then an image is represented by the code histogram of its patches. Such histograms are then compared for classification. For example, previous work in [6] uses a normalized l_1 similarity with K-nearest neighbor (KNN) for galactographic image analysis.

We propose two methods using histogram intersection (HI) for mammographic texture classification. The first method is to use a normalized HI in the KNN classifier. Our second proposal is to combine HI with support vector machines. Both methods are tested on a galactographic dataset containing both normal and pathologic samples. We designed a thorough evaluation containing 288 different configurations and eight different methods (five similarity measures and two classifiers). In the experimental evaluation HI-based methods outperformed other competitors, including previously reported solutions.

In addition to the above contribution, the thorough experiments we performed help understanding the effect of different similarity measures. In particular, we found that normalization plays an important role in designing similarity measures for galactographic texture analysis, especially when used with the KNN classifier. In addition, we observed that SVM based methods perform significantly better than KNN based ones.

The rest of the paper is organized as follows: in Section 2 we provide background information and discuss previous work. In Section 3 we introduce the bag-of-words framework and the histogram intersection based methods. Then, we present the experimental evaluation in Section 4. Finally, Section 5 concludes the paper.

2. BACKGROUND

Texture analysis is an important tool for image analysis and has been used for several decades. Early work for texture analysis often focuses on low level features including color and local gradient statistics. For example, Gabor jets are used in content-based image retrieval systems [5], color distribution is used for image retrieval [14], etc. Recently, patch based methods attract great attention due to their richness in local information and flexibility in object deformation. The bag-of-words (BOW) model, originally used in document analysis, has proved to be very powerful for category classification tasks [13].

In medical imaging, texture analysis has been used particularly for developing computer-aided diagnosis (CAD) systems, which gain increasing popularity due to their ability to improve the precision and accuracy of characterization by radiologists [7]. It has also been used for medical image retrieval, e.g., searching an input image in a database to find images similar to the input. The results are then used for accessing other clinical data and known diagnoses from similar cases [8].

For breast imaging, in particular for galactographic images, many previous studies [2,3,4] have focused on investigating topological descriptors of tree-like structures representing the breast ductal network.

Patch based texture analysis has been applied to x-ray galactographic images [6]. In the work, a similarity derived from normalized l_1 metric is combined with KNN for classification tasks. In comparison, in this paper we use normalized histogram intersection for similarity and also investigate the performance by using support vector machines. Both methods outperform the normalized l_1 on a galactographic dataset.

The use of histogram intersection for image comparison dates back at least to [14], where HI was used to compare color histograms from two images. HI recently became very popular in object/category classification due to its robustness and the fact that it is a Mercer kernel. For example, in [11] and [12] HI was combined with SVM for image matching. To the best of our knowledge, however, it has not been applied to mammographic image analysis. HI also relates to the Mamdani intersection used in fuzzy logic.

3. METHODOLOGY

3.1. Bag-of-words Representation

The basic idea of bag-of-words framework, when applied to images, consists of four steps: (1) build a codebook for local patches; (2) extract local patches for an image; (3) represent an image using the statistics of its quantized local patches; and (4) inference based on the statistics collected in (3). Figure 1 illustrates the process.

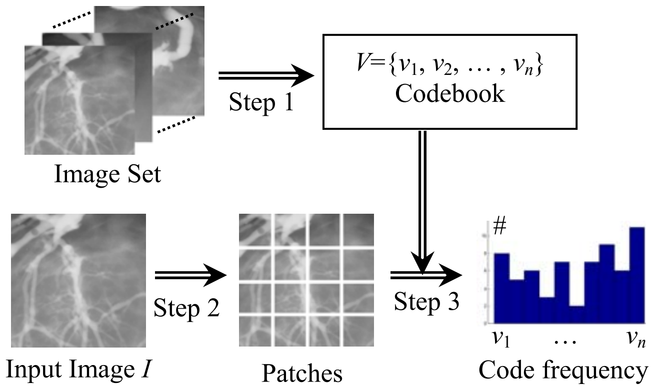

Fig. 1: The bag-of-words framework.

To compare with the previous work [6], we follow the same procedure for codebook generation, patch extraction, and image representation. Specifically, we use the Generalized Lloyd Algorithm (GLA) for codebook clustering, the regular grid decomposition for patch extraction, and code frequency histogram for image representation.

In the following we denote the learned codebook as $V=\{v_1, v_2, \ldots, v_n\}$, containing n vector quantized codes $v_i \in \Re^d$, which correspond to the i^{th} cluster center in the d dimensional feature space (in our case, the intensity values of pixels inside a patch). Then, for an image I, it is first decomposed into m patches $P=\{p_1, p_2, \ldots, p_m\}$, each patch $p_i \in \Re^d$ is then mapped as a word w_i by

$$w_i = \arg\min_{1\le k\le n} \| p_i - v_k \|,$$

where $\|.\|$ denotes the l_2 norm. After that, a code histogram $h \in \Re^n$ is build as

$$h(j)=\#\{w_i\text{: } 1\le i\le n \text{ and } w_i=j\},\ 1\le j\le n$$

where # denotes the cardinality of a set. We use h as a representation of the image I.

3.2. Histogram Similarity

Once the histogram representation of images is ready, the next key step is to find a good histogram similarity measure. Specifically, let h_1 and h_2 be code histograms for images I_1 and I_2 respectively; the task is to find a similarity measure $s(h_1,h_2)$ between them. In previous work [6], the following similarity measure was proposed

$$s_{NL1}(h_1,h_2) = \frac{1}{1+\sum_{i=1..n} |h_1(i)-h_2(i)| / (1+h_1(i)+h_2(i))},$$

which has demonstrated promising classification results for mammography images. We denote this histogram similarity measure as s_{NL1} because it is closely related to the "Normalized" l_1 distance.

In this paper, we propose using histogram intersection instead, which is defined as

$$s_{HI}(h_1, h_2) = \sum_{i=1..n} \min(h_1(i), h_2(i)) \ .$$

This representation, though simple, has been shown to be very useful for image retrieval and classification tasks [14,12,11]. It has the following advantages:

1. It is suitable for partial matching and is therefore robust to occlusion and/or clutters.
2. It is a Mercer kernel (i.e., a symmetric positive definite kernel), which can be combined with kernel based methods such as the Support Vector Machine (SVM).
3. It can be computed very efficiently, which is important when generalizing to large scale tasks.

Although HI can be used directly as a similarity measure, it suffers from the problem where histograms are very noisy. Inspired by [6], we define *Normalized* HI (NHI) as

$$s_{NHI}(h_1, h_2) = \sum_{i=1..n} \frac{\min(h_1(i), h_2(i))}{h_1(i) + h_2(i)} \ ,$$

which generates promising results as shown in Section 4.

3.2. Image Classification using Histogram Intersection

K Nearest Neighbor (KNN) is a widely used classifier, due to its simplicity and strong performance. To compare with previous work in [6], we use KNN with the normalized HI instead of the normalized l_1 similarity.

While KNN achieves pretty good results for our task, it is a simple classifier that may not work well for complicated tasks. In addition, as shown in [6], the performance of KNN can be very sensitive to the choice of K.

In addition to KNN, we investigate using Support Vector Machines (SVM), which have demonstrated excellent performance in many classification tasks. Let h be the code histogram representation of an input image I, the SVM classifier has the following form

$$sign\left(\sum_{i=1..n_s} \alpha_i l_i K(h_i, h) + b\right),$$

where n_s is the number of support vectors, h_i, l_i are the labels of corresponding support vectors, and α_i and b are parameters learned from training samples. The kernel $K(.,.)$ is very important since it determines the behavior of the classifier. When $K(.,.)$ is symmetric positive definite, aka a Mercer kernel, SVM implicitly maps the original problem into a (usually) non-linear space. Histogram intersection has been proved to be a Mercer and therefore readily to be used in SVM, that is,

$$K(h_i, h) = s_{NH}(h_i, h) \ .$$

4. EXPERIMENTAL RESULTS

In this section, we describe experiments using the proposed methods for texture classification of regions of interest (ROIs) manually extracted from x-ray galactograms [2]. These ROIs are taken from the galactogram area behind the nipple; such ROI selection has been shown to be highly indicative of a woman's risk to develop pathology [9].

4.1. Dataset

We use the same dataset in [6], which is part of a study in [2]. The dataset contains 23 galactographic images collected retrospectively from 14 patients. The images in the dataset form two groups: 10 cases with no radiological findings (NF) and 13 cases with radiological findings (RF). The images were digitized with a spatial resolution of (100 micron)^2/pixel using a Lumisys digitizer (Sunnyvale, CA).

For each image, a ROI of size 256×256-pixel was manually segmented in the region of the breast behind the nipple. Our task is to classify each ROI based on its texture pattern. Figures 2 and 3 show several example texture patterns in the dataset.

4.2. Methods

We investigate the HI-based methods on the task along with several other approaches. Specifically, our experiments involve five different similarity measures and two classifiers, including: (1) l_1+KNN, (2) Normalized l_1+KNN [6], (3) Chi square+KNN, (4) Chi square+SVM, (5) HI+KNN, (6) HI+SVM, (7) Normalized HI+KNN, and (8) Normalized HI+SVM.

To reduce the randomness in the experiments, we perform the experiments in many different configurations and report the average classification rates. To reduce the effects brought by different patch sizes, three different sizes are tested: 4×4, 8×8, 16×16. Codebooks with different numbers of codes are also used (256 or 512). For KNN, we tested K=1,2,3,4. Finally, to reduce the randomness in vector quantization, 12 random codebooks are generated given a fixed training/testing configuration. To summarize, the above combination generates 3×2×4×12=288 different configurations for KNN-based methods and 3×2×12=72 configurations for SVM-based methods.

4.3. Results

We use leave-one-out scheme for evaluating the performance of different approaches. Specifically, each time one ROI is used for testing and the rest 22 ROIs are used for training. For each training/testing separation, we test all different configurations described above, and report the average classification rate. Finally, we calculate the average classification rate over all different ROIs. The classification rates are summarized in Table 1. In addition, some example ROIs that are frequently classified correctly or frequently misclassified are shown in Figures 2 and 3.

	KNN	SVM
l_1	59.2%	n/a
Normalized l_1 [6]	69.8%	n/a
Chi square	58.2%	76.5%
HI	59.2%	**80.7%**
Normalized HI	**70.5%**	73.3%

Table 1: Average classification rates of different methods.

From the experimental results we have the following observations. First, histogram intersection does help improve the performance, especially when combined with SVM. Second, SVM is helpful as we expected. Third, for both HI and l_1, normalization greatly helps to improve the performance when KNN is used. Fourth, from the example images in Figures 2 and 3, we can see that the problem is very challenging from the perspective of human vision, which may guide us for searching for better representation in the future.

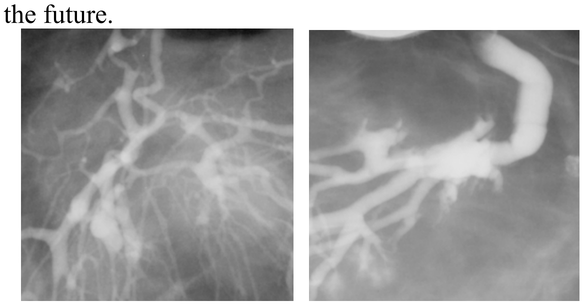

Fig. 2: The NF ROI (left) and the RF ROI (right) with the best classification rates.

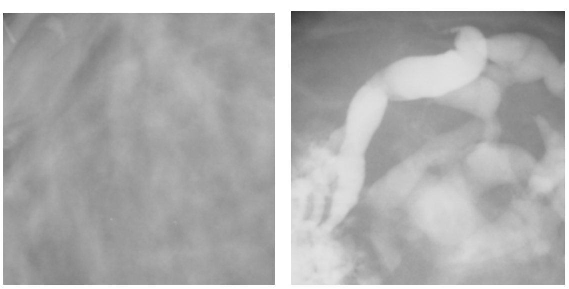

Fig. 3: The NF ROI (left) and the RF ROI (right) with the worst classification rates.

5. CONCLUSION

We investigate using histogram intersection (HI) in the bag-of-words framework for mammography image classification. Specifically, we show that both normalized HI and HI+SVM outperform previous state-of-the-art methods on a real mammography image dataset. We also notice that, when using KNN classifiers, normalization is an important step that helps to improve the accuracy of histogram similarities.

In the future, we plan to advance the study by including more data and thorough study. We will explore mainly along two directions: more effective texture representation and advanced classification techniques. In addition, it is also interesting to extend the study to different anatomic structures and modalities.

6. ACKNOWLEDGEMENTS

This work was supported in part by NSF Research Grant IIS-0916624. The funding agency specifically disclaims responsibility for any analyses, interpretations and conclusions.

7. REFERENCES

[1] B. Pereira and K. Mokbel, "Mammary ductoscopy: past, present, and future," Mammary ductoscopy: past, present, and future, vol. 10, pp. 112-116, 2005.

[2] Bakic, PR., Albert, M., Maidment, AD., Classification of galactograms with ramification matrices: preliminary results, Academic Radiology, Vol. 10 (2) pp. 198-204, 2003.

[3] V. Megalooikonomou, M. Barnathan, D. Kontos, P. R. Bakic, A. D.A. Maidment, "A Representation and Classification Scheme for Tree-like Structures in Medical Images: Analyzing the Branching Pattern of Ductal Trees in X-ray Galactograms," IEEE Transactions on Medical Imaging, Vol. 28, No, 4, pp. 487-493, April 2009.

[4] Kontos, D., Megalooikonomou, V., Javadi, A., Bakic, PR., Maidment, AD., Classification of galactograms using fractal properties of the breast ductal network, Proceedings of the IEEE Int'l Symposium on Biomedical Imaging, 2006.

[5] J. Ashley, M. Flickner, J. Hafner, D. Lee, W. Niblack, D. Petkovic, "The query by image content (QBIC) system", ACM SIGMOD Record, 24(2), pp. 475, November 1995.

[6] V. Megalooikonomou, J. Zhang, D. Kontos, P.R. Bakic, "Analysis of texture patterns in medical images with an application to breast imaging", Proceedings of the SPIE Conference on Medical Imaging, 2007.

[7] Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K. Improving breast cancer diagnosis with computer-aided diagnosis. Academic Radiology 1999; 6:22-33

[8] Tourassi GD, Vargas-Voracek R, Catarious DM, Floyd CE. Computer-assisted detection of mammographic masses: a template matching scheme based on mutual information. Medical Physics 2003; 30:2123-2130

[9] Li, H., Giger, ML., Huo, Z., et al. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location, Medical Physics, Vol. 31, Issue 3, pp. 549-555, 2004.

[10] H. Li, M.L. Giger, O.I. Olopade, A. Margolis, L. Lan, & M.R. Chinander, "Computerized Texture Analysis of Mammographic Parenchymal Patterns of Digitized Mammograms," *Academic Radiology*, vol. 12, 2005, 863-873.

[11] Grauman, K. & Darrell, T. "The Pyramid Match Kernel: Efficient Learning with Sets of Features". Journal of Machine Learning Research, 8 (Apr): 725--760, 2007.

[12] Ling, H. & Soatto, S., "Proximity distribution kernels for geometric context in category recognition. In Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV), 2007.

[13] Zhang, J.; Marszalek, M.; Lazebnik, S. & Schmid, C. Local Features and Kernels for Classification of Texture and Object Categories: A Comprehensive Study International Journal of Computer Vision, 2007, 73, 213-238

[14] M. J. Swain and D. H. Ballard. ``Color Indexing", International Journal of Computer Vision, 7(1):11-32, 1991.