Adaptive learning for relevance feedback: Application to digital mammography

Jung Hun Oh

Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri 63110

Yongyi Yang

Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, Illinois 60616

Issam El Naqa^{a)}

Department of Radiation Oncology, Washington University School of Medicine, St. Louis, Missouri 63110

(Received 25 April 2010; revised 4 June 2010; accepted for publication 16 June 2010; published 29 July 2010)

Purpose: With the rapid growing volume of images in medical databases, development of efficient image retrieval systems to retrieve relevant or similar images to a query image has become an active research area. Despite many efforts to improve the performance of techniques for accurate image retrieval, its success in biomedicine thus far has been quite limited. This article presents an adaptive content-based image retrieval (CBIR) system for improving the performance of image retrieval in mammographic databases.

Methods: In this work, the authors propose a new relevance feedback approach based on incremental learning with support vector machine (SVM) regression. Also, the authors present a new local perturbation method to further improve the performance of the proposed relevance feedback system. The approaches enable efficient online learning by adapting the current trained model to changes prompted by the user's relevance feedback, avoiding the burden of retraining the CBIR system. To demonstrate the proposed image retrieval system, the authors used two mammogram data sets: A set of 76 mammograms scored based on geometrical similarity and a larger set of 200 mammograms scored by expert radiologists based on pathological findings.

Results: The experimental results show that the proposed relevance feedback strategy improves the retrieval precision for both data sets while achieving high efficiency compared to offline SVM. For the data set of 200 mammograms, the authors obtained an average precision of 0.48 and an area under the precision-recall curve of 0.79. In addition, using the same database, the authors achieved a high pathology matching rate greater than 80% between the query and the top retrieved images after relevance feedback.

Conclusions: Using mammographic databases, the results demonstrate that the proposed approach is more accurate than the model without using relevance feedback not only in image retrieval but also in pathology matching while maintaining its effectiveness for online relevance feedback applications. © 2010 American Association of Physicists in Medicine. [DOI: [10.1118/1.3460839](https://doi.org/10.1118/1.3460839)]

Key words: relevance feedback, incremental learning, mammogram, SVM regression

I. INTRODUCTION

With the growing volume of images used in medicine, given a query image, the capability to retrieve relevant or similar images from large databases is becoming increasingly important.¹⁻³ The key to the successful image retrieval system lies in the development of appropriate similarity metrics for ranking the relevance of images in a database to the query image. A variety of content-based image retrieval (CBIR) techniques have been proposed to overcome the difficulties encountered in textual annotation for large image databases. However, the gap between low-level image features and high-level semantic understanding in CBIR systems still remains a challenging problem.⁴ Recently, many relevance feedback schemes have been developed to improve the performance of offline CBIR systems. Relevance feedback was originally developed in traditional text-retrieval

systems for improving the results of a retrieval strategy.^{5,6} In the image retrieval context, relevance feedback is a postquery process to refine the retrievals by using positive or negative indications from the user's interaction.^{7,8} A fundamental difference of relevance feedback in image retrieval compared to document retrieval is that the latter is based on fixed symbolic representations, with direct mapping to human interpretations, while for images, we can only assume that there exists some sort of mapping between high-level user perception and some extractable low-level image features (e.g., color, texture, shape, etc.).

Despite the progress made in the general area of image retrieval in recent years, its success in biomedicine thus far has been quite limited.⁹ In our previous works,^{6,10,11} we investigated the use of CBIR for digital mammograms, containing clustered microcalcifications (MCs) that are early

signs of breast cancer development. The goal was to provide radiologists with a set of images from past cases that are relevant to the query being evaluated, along with the known pathology of these past cases. That is, we believe that a mammogram retrieval system that presents images with the known pathology that are relevant to the image being evaluated may help radiologists more accurately diagnose breast cancer patients. In this paper, we extend this work and explore new methods to incorporate relevance feedback to refine the image retrieval process by utilizing user's feedback into our proposed retrieval system. The proposed method achieved significant improvement in performance of CBIR in digital mammography over our previous works. Below, we briefly review some of the recent developments in CBIR that were reported in the literature and summarize its application in mammography.

Peng¹² proposed a multiclass form of relevance feedback retrieval instead of the classical two-class approach, in which a chi-squared (χ^2) analysis is used to determine the local relevance of each feature dimension. Zhang and Chen¹³ introduced a general active learning framework for CBIR. For each object in the database, a list of probabilities is maintained, each indicating the probability of the object having one of the attributes. The list of probabilities is used as a feature vector to calculate the distance between the user query and an image in the database. The overall distance between two images is determined by a weighted sum of the semantic distance and a low-level feature distance. In Guo *et al.*,¹⁴ a two-stage retrieval system for natural images was proposed. This system at first employs a classifier such as support vector machine (SVM) or Adaboost that learns the boundary between relevant and irrelevant images to the given query. Then, the relevant images are ranked based on a Euclidean distance metric that uses the color coherence vector coefficients as features. In this experiment with images in Corel photo gallery, the highest precision and recall were approximately 0.41 and 0.47, respectively. A drawback of this approach is that the increased number of support vectors would make the filtering process too slow. Cheng *et al.*¹⁵ proposed a unified relevance feedback framework for Web image retrieval, using both textual features and visual features. To construct an accurate and low-dimensional textual space for the resulting Web images, an effective search result clustering algorithm was employed. Four relevance feedback strategies were compared: Relevance feedback using textual feature only, relevance feedback using visual feature only, linear combination of the relevance feedback in two feature spaces, and the proposed relevance feedback fusion strategy. The average precision was 0.5481, 0.3905, 0.6705, and 0.883, respectively. Yin *et al.*¹⁶ presented a new technique called the virtual feature that digests the cross-session query experiences to estimate the semantic relevance between images compared to the traditional relevance feedback methods that use only within-session query experience. Tao *et al.*¹⁷ designed a novel method called asymmetric bagging random subspace SVM to solve problems arising when the classical SVM-based relevance feedback is used with a small number

of labeled positive feedback samples. In this method, there is an inclination that as the number of feedback samples increases, the precision is better. Therefore, a disadvantage of this model is that many feedback samples are required to achieve satisfactory results. Li and Hsu¹⁸ proposed a graph-theoretic approach that converts the region correspondence estimation problem into an inexact graph matching problem. The relevance feedback step is based on a maximum likelihood method to re-estimate an ideal query and a corresponding image distance measurement. In experimental results using a subset of Corel photo gallery, very low precision values were attained with highest precision of 0.37. The limitation of this work is that the local optimum by the first feedback iteration does not significantly improve the retrieval performance in later iterations. Azimi-Sadjadi *et al.*¹⁹ developed an adaptable CBIR system that attempts to capture high-level semantic user concepts, where learning is implemented in two modes: Model-reference and relevance feedback. The incorporation of user concepts is carried out in an online relevance feedback mode, while the incorporation of the model-reference information is performed in a batch learning mode.

Recently, several studies were performed to develop CBIR systems for mammography.²⁰⁻³¹ This was initiated by the pioneering work of Swett *et al.*,²³ who developed a computer-based expert system called MAMMO/ICON for automated mammographic image retrieval based on speech recognition technology using findings in the textual report or in the dictation. Mazurowski *et al.*²⁸ proposed an optimization framework for improving a case-based computer-aided decision (CAD) system that was developed for the classification of regions of interests (ROIs) in mammograms. The proposed method is based on the hypothesis that images in the knowledge database vary in their diagnostic importance. Therefore, a different weight was assigned for each stored image. Tourassi *et al.*²⁹ presented several filtering techniques as preprocessing steps for improving the performance of an information-theoretic CAD (IT-CAD) system. In this approach, a ROI database was used, which included true masses and false-positive regions from digitized mammograms and the filters were selected to complement the similarity metric in the IT-CAD system. Park *et al.*³⁰ proposed a simple strategy to remove regionally misclassified ROIs and studied its effect on improving the performance of an interactive computer-aided detection and diagnosis (I-CAD) system. Then, they tested the relationship between the size of the database and the I-CAD performance for reducing false positives in breast masses analysis. The interested reader could find a more detailed description about these methods and others in our review chapter.³²

In this paper, we propose a new unified relevance feedback system incorporating an incremental learning strategy into SVM regression with application to mammogram databases. Particularly, in order to further improve the performance of the proposed relevance feedback strategy, we introduce a new concept called local perturbation that controls neighborhood perturbation by changing certain parameter values in SVM regression of the image samples in the prox-

FIG. 1. The proposed image retrieval framework with relevance feedback.

imity of the current feedback sample. The proposed method enables adaptive online learning by human-machine interaction with objectives to improve the effectiveness of the retrieval system while maintaining high efficiency.

II. OVERVIEW OF THE PROPOSED IMAGE-RETRIEVAL FRAMEWORK

We assume that the user's notion of similarity between a pair of images is expressed as a function of the relevant features of the images. To model this notion of similarity, we then use machine learning methods for the purpose of the image retrieval system. Our goal is to find those images among the many images in the database that are most clinically similar to the query image as judged by the user. Figure 1 illustrates the proposed framework in a functional diagram. For a given query image, we first extract the key features of the image that are represented by an M -dimensional vector $\mathbf{u}$ and quantify them, which characterizes the image. This feature vector is then compared to the corresponding feature vector $\mathbf{v}$ of each image in the database by way of a nonlinear mapping function denoted as $f(\mathbf{u}, \mathbf{v})$. As a result, a similarity coefficient (SC) for a pair of images (the query image and each image in the database) is produced. The images with the highest SCs that are larger than a prescribed threshold value T are then retrieved from the database.

Clearly, the key to this framework lies in the nonlinear mapping function $f(\mathbf{u}, \mathbf{v})$. Ideally, this mapping should have the following properties: (1) $f(\mathbf{u}, \mathbf{v})$ must closely reflect the user's notion of similarity; (2) the mapping should have reasonable computational complexity in order to be efficiently applied in a large-scale database; and (3) $f(\mathbf{u}, \mathbf{v})$ should enable the user to refine the search through using relevance feedback schemes. We adopt a supervised learning approach for the determination of $f(\mathbf{u}, \mathbf{v})$. For this purpose, we first collect labeled similarity scores (e.g., obtained from observer studies) for a set of sample image pairs. We then train a learning machine to capture $f(\mathbf{u}, \mathbf{v})$ with these samples. Suppose that $SC(\mathbf{u}, \mathbf{v})$ denotes the similarity coefficient between an image pair that is characterized by $\mathbf{u}$ and $\mathbf{v}$. Therefore, we can model $SC(\mathbf{u}, \mathbf{v})$ as

$$SC(\mathbf{u}, \mathbf{v}) = f(\mathbf{u}, \mathbf{v}) + \xi, \quad (1)$$

where ξ is the modeling error. Then, the problem of learning similarity between images can be viewed as a regression problem.^{33–35} Now our aim becomes to determine a regres-

FIG. 2. Illustration of ϵ -insensitive support vector machine (ϵ -SVM) for regression. The support vectors are indicated by filled squares.

sion function $f(\cdot, \cdot)$ that generalizes well to unseen images in the testing set.

For simplicity, we view the similarity metric as a function of a single argument $\mathbf{x} = [\mathbf{u}^T \ \mathbf{v}^T]^T$ that is a concatenation of the feature vectors $\mathbf{u}$ and $\mathbf{v}$ of two images to be compared, accordingly redefining the similarity function $f(\mathbf{u}, \mathbf{v})$ as $f(\mathbf{x})$. In this study, we consider a SVM (Ref. 33) for learning the similarity function $f(\mathbf{x})$ though other learning machine approaches could be used.³⁴ The advantage of SVM over other learning methods lies in its robustness and mathematically tractable formulation.

Although SVM was originally designed to solve a binary classification problem, it can also be applied for regression. A SVM formulation in such a case maintains many of the characteristics of the classification case. For nonlinear regression, a SVM in concept first maps the input data vector $\mathbf{x}$ into a higher dimensional space $\mathbf{H}$ through an underlying nonlinear mapping $\Phi(\cdot)$, and then applies a linear regression in this mapped space. That is, a nonlinear SVM regression function can be written in the following form:

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b. \quad (2)$$

Let $\{(\mathbf{x}_i, y_i), i=1, 2, \dots, l\}$ denote a set of training samples, where y_i is the human-observer similarity score for the image pair denoted by $\mathbf{x}_i$. The parameters $\mathbf{w}$ and b in the regression function of Eq. (2) are determined through minimization of the following structured risk:

$$R(\mathbf{w}, b) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l L_\epsilon(\mathbf{x}_i), \quad (3)$$

where $L_\epsilon(\cdot)$ is the so-called ϵ -insensitive loss function which is defined as

$$L_\epsilon(\mathbf{x}) = \begin{cases} |y - f(\mathbf{x})| - \epsilon, & \text{if } |y - f(\mathbf{x})| \geq \epsilon \\ 0, & \text{otherwise} \end{cases}. \quad (4)$$

The function $L_\epsilon(\cdot)$ has the property that it does not penalize errors below the parameter ϵ , as illustrated in Fig. 2. The constant C in Eq. (3) determines the trade-off between the model complexity and the training error. The regression function $f(\mathbf{x})$ in Eq. (2) is also characterized by a subset of the training data known as the support vectors. It can be written as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{l_s} (\alpha_i - \alpha_i^*) K(\mathbf{x}, \mathbf{s}_i) + b, \quad (5)$$

where $\mathbf{s}_i$, $i=1, 2, \dots, l_s$ denote the support vectors; α_i, α_i^* are the Lagrange multipliers associated with the support vectors; and $K(\mathbf{x}, \mathbf{s}_i) = \Phi(\mathbf{x})^T \Phi(\mathbf{s}_i)$ which is called a kernel function. A training sample $(\mathbf{x}_i, y_i)$ is a *margin support vector* when $|f(\mathbf{x}_i) - y_i| = \varepsilon$ and an *error support vector* when $|f(\mathbf{x}_i) - y_i| > \varepsilon$.

From Eq. (5), we can directly evaluate the regression function through the kernel function $K(|, |)$ without need to specifically address the underlying mapping function $\Phi(|)$. In SVM, the two commonly used kernel types are polynomial kernels and radial basis functions (RBFs) that are known to satisfy Mercer's condition.³³ They are defined as follows:

- Polynomial kernel:

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^p, \quad (6)$$

where $p > 0$ is a constant that defines the kernel order.

- RBF kernel:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \quad (7)$$

where σ is a constant that defines the kernel width. In this work, the RBF kernel was used. Here, the selection of C and σ is important since these parameters can determine the tradeoff between the model overfitting and underfitting.³⁶ For instance, a large value of C would result in a high penalization of training error, increased number of support vectors, and potential lack of generalizability.

III. RELEVANCE FEEDBACK

In this section, we explore strategies to incorporate relevance feedback into our proposed learning-based retrieval approach. We consider the following scenario: For a query image $\mathbf{q}$, a user selects a relevant image $\mathbf{r}$ among the retrieved images to confirm that the retrieved $\mathbf{r}$ is indeed similar to the query $\mathbf{q}$; we want to incorporate this information to further refine the search, hoping that more relevant images could be found for the same query $\mathbf{q}$.

Several strategies could be used to incorporate the feedback information. In one strategy, the SC function could be regarded as a weighted sum of the original query and the relevant image from the user. However, the determination of the appropriate weight would rely on trial and error. Another strategy would be to refine the training around the feedback samples by changing the regularization parameter C , for instance, again, in the case one would rely on heuristics to determine the necessary modification and the answer would be suboptimal. A more robust approach that aims to retain the solution optimality could be based on an incremental learning scenario.^{37,38}

In this scenario, suppose that the online user via relevance feedback introduces a new sample, denoted by $(\mathbf{x}_c, y_c)$. Our objective is to somehow modify the existing SVM in Eq. (5)

to incorporate this newly added piece of information. A straightforward approach to achieve this would be to retrain the SVM using the old samples $Z = \{(\mathbf{x}_i, y_i)\}_{i=1}^l$ and the newly added sample. However, this process is excessively expensive for real-time online applications. An alternative to achieve this in SVM is by retaining only the support vectors with the new sample. However, this technique may yield only an approximate solution.³⁹ Cauwenberghs and Poggio^{37,38} developed a recursive procedure for SVM classification applications, where the SVM solution of $(l+1)$ -sample training set is found in terms of the previous l -sample training set and the new sample. The key to this procedure is to retain the Karush–Kuhn–Tucker (KKT) conditions of the SVM solution from the previous data, while “adiabatically” (i.e., perturbation without loss or gain) adding the new sample. In this work, we extend the idea by Cauwenberghs and Poggio to SVM regression and apply it for relevance feedback in image retrieval using clinical databases.

The proposed approach consists of the following two steps: (1) *Addition of a new sample*. In other words, incorporate the user response $(\mathbf{x}_c, y_c)$ into the existing SVM machine using incremental learning; and (2) *Local perturbation*. In other words, refine the similarity function in the vicinity of $(\mathbf{x}_c, y_c)$. These two steps can be performed adiabatically by applying the same principles as in Ref. 37.

III.A. Addition of a new sample

For the ε -SVM in Eq. (5) trained with $Z = \{(\mathbf{x}_i, y_i)\}_{i=1}^l$, let S denote the set of strict support vectors (i.e., those samples falling precisely on the ε margin), and let M (or E) denote the samples falling inside (or outside) the margin. The KKT conditions corresponding to the SVM solution can then be expressed as

$$\begin{cases} g_i^{(*)} = 0, 0 < \alpha_i^{(*)} < C & (\mathbf{x}_i, y_i) \in S \\ g_i^{(*)} > 0, \alpha_i^{(*)} = 0 & (\mathbf{x}_i, y_i) \in M \\ g_i^{(*)} < 0, \alpha_i^{(*)} = C & (\mathbf{x}_i, y_i) \in E \end{cases}, \quad (8)$$

where $\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0$ and $\alpha_i, \alpha_i^* \in [0, C]$; $g_i = f(\mathbf{x}_i) - y_i + \varepsilon$ and $g_i^* = y_i - f(\mathbf{x}_i) + \varepsilon$ in the upper and lower bounds of $f(\mathbf{x}_i)$, respectively; $g_i^{(*)}$ (denoting both g_i and g_i^*) is the gradient of the SVM objective function with respect to α_i or α_i^* ; and C is the regularization parameter controlling the trade-off between the training error and the model complexity.³³ With the newly added sample $(\mathbf{x}_c, y_c)$, the SVM solution in Eq. (5) is likely to be perturbed (i.e., the membership of the sets $\{S, M, E\}$ would change). This could be represented by differentials of the KKT conditions as in Ref. 37,

$$\begin{cases} \Delta g_i = K(\mathbf{x}_i, \mathbf{x}_c) \Delta \gamma_c + \sum_{j \in S} K(\mathbf{x}_i, \mathbf{x}_j) \Delta \gamma_j + \Delta b \\ \Delta g_i^* = -K(\mathbf{x}_i, \mathbf{x}_c) \Delta \gamma_c - \sum_{j \in S} K(\mathbf{x}_i, \mathbf{x}_j) \Delta \gamma_j - \Delta b \\ \Delta \gamma_c + \sum_{j \in S} \Delta \gamma_j = 0, \end{cases} \quad (9)$$

where $\Delta \gamma_j = \Delta(\alpha_j - \alpha_j^*)$ and α_c and α_c^* are the coefficients corresponding to $(\mathbf{x}_c, y_c)$. Note that Δg_i and Δg_i^* are different only in sign, so it suffices to compute either one of them. Let $\gamma_j = \alpha_j - \alpha_j^*$. Note that after the perturbation, the new support vectors satisfy the following condition:

$$\Delta g_i^{(*)} = 0, \quad \text{for } (\mathbf{x}_i, y_i) \in S. \quad (10)$$

Consequently, Eq. (9) can be rewritten in a matrix form as

$$\mathbf{Q} \cdot \begin{bmatrix} \Delta b \\ \Delta \gamma_{S_1} \\ \vdots \\ \Delta \gamma_{S_{l_s}} \end{bmatrix} = - \begin{bmatrix} 1 \\ K(\mathbf{x}_{S_1}, \mathbf{x}_{S_c}) \\ \vdots \\ K(\mathbf{x}_{S_{l_s}}, \mathbf{x}_{S_c}) \end{bmatrix} \cdot \Delta \gamma_c, \quad (11)$$

where $\mathbf{x}_{S_j}$ denotes the j th support vector, $j=1, 2, \dots, l_s$ and

$$\mathbf{Q} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & K(\mathbf{x}_{S_1}, \mathbf{x}_{S_1}) & \cdots & K(\mathbf{x}_{S_1}, \mathbf{x}_{S_{l_s}}) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(\mathbf{x}_{S_{l_s}}, \mathbf{x}_{S_1}) & \cdots & K(\mathbf{x}_{S_{l_s}}, \mathbf{x}_{S_{l_s}}) \end{bmatrix}. \quad (12)$$

Let $\mathbf{R} = \mathbf{Q}^{-1}$ and define the following so-called sensitivity coefficients

$$\beta_0 = \frac{\Delta b}{\Delta \gamma_c}, \quad \beta_j = \frac{\Delta \gamma_j}{\Delta \gamma_c}, \quad j = 1, 2, \dots, l_s. \quad (13)$$

Then, Eq. (11) can be rewritten as

$$\begin{bmatrix} \beta_0 \\ \beta_{S_1} \\ \vdots \\ \beta_{S_{l_s}} \end{bmatrix} = -\mathbf{R} \begin{bmatrix} 1 \\ K(\mathbf{x}_{S_1}, \mathbf{x}_{S_c}) \\ \vdots \\ K(\mathbf{x}_{S_{l_s}}, \mathbf{x}_{S_c}) \end{bmatrix}. \quad (14)$$

The differential changes can then be expressed for each sample as

$$\Delta g_i^{(*)} = \pm \phi_i \Delta \gamma_c, \quad i = 1, \dots, l+1, \quad (15)$$

where ϕ_i , called the margin sensitivity, is given by

$$\phi_i = \begin{cases} 0, & \forall i \in S \\ K(\mathbf{x}_i, \mathbf{x}_c) + \sum_{j \in S} K(\mathbf{x}_i, \mathbf{x}_j) \beta_j + \beta_0, & \text{otherwise} \end{cases}. \quad (16)$$

When the feedback sample $(\mathbf{x}_c, y_c)$ or an existing sample by perturbation is added to S , the matrix $\mathbf{R}$ in Eq. (14) can be computed using the Woodbury's identity as in Ref. 37,

TABLE I. Incremental relevance feedback algorithm.

For each feedback sample $(\mathbf{x}_c, y_c)$, do the following:	
• Initialize $\gamma_c = 0$.	
• If $g_c^{(*)} > 0$, then $(\mathbf{x}_c, y_c) \in M$, and terminate; Otherwise, find the largest increment γ_c such that one of the following conditions first occurs:	
1.	If $g_c^{(*)} = 0$, $(\mathbf{x}_c, y_c) \in S$, update $\mathbf{R}$.
2.	If $\gamma_c = C$, $(\mathbf{x}_c, y_c) \in E$, terminate.
3.	Migrate samples across adjacent sets $\{S, M, E\}$ by checking bounds on γ_i for all samples in Z , and if S changes, update $\mathbf{R}$ accordingly.
	(a) If $\gamma_i = 0$, transfer $(\mathbf{x}_i, y_i)$ from S to M .
	(b) If $\gamma_i = C$, transfer $(\mathbf{x}_i, y_i)$ from S to E .
	(c) If $\gamma_i < C$, transfer $(\mathbf{x}_i, y_i)$ from E to S .
	(d) If $\gamma_i > 0$, transfer $(\mathbf{x}_i, y_i)$ from M to S .
• Repeat this procedure until convergence is achieved.	

$$\mathbf{R} \leftarrow \begin{bmatrix} & 0 \\ & \vdots \\ \mathbf{R} & 0 \\ 0 & \cdots & 0 & 0 \end{bmatrix} + \frac{1}{\phi_c} \beta^T \beta, \quad (17)$$

where $\beta = [\beta_0 \ \beta_{S_1} \ \cdots \ \beta_{S_{l_s}} \ 1]$. Likewise, when a support vector moves from S into M or E , the matrix $\mathbf{R}$ is updated as

$$\mathbf{R}_{ij} \leftarrow \mathbf{R}_{ij} - \frac{\mathbf{R}_{ik} \mathbf{R}_{kj}}{\mathbf{R}_{kk}}, \quad (18)$$

where $\mathbf{R}_{ij}$ denotes the (i, j) th entry of $\mathbf{R}$. The resulting incremental learning algorithm is summarized in Table I.

III.B. Local perturbation by ε parameter

After the feedback sample $(\mathbf{x}_c, y_c)$ is incorporated, the SVM function in Eq. (5) is further modified to control the perturbation in the vicinity of $(\mathbf{x}_c, y_c)$ in the feature space. In other words, we adapt the SVM to the response of the user by refining the learning in the vicinity of $(\mathbf{x}_c, y_c)$ in the feature space. For this purpose, we refine the regression tolerance ε in the SVM function for those samples close to $(\mathbf{x}_c, y_c)$. Let N denote a set of samples in a close neighborhood of the feedback sample $\mathbf{x}_c$. Here, we employ a RBF kernel $K(\mathbf{x}_i, \mathbf{x}_c)$ to identify those samples that are close to $\mathbf{x}_c$. That is, $\mathbf{x}_i \in N$ if $K(\mathbf{x}_i, \mathbf{x}_c) > A$, where A is a prescribed threshold. For each sample in the set N , we modify the regression tolerance ε_i as follows:

$$\varepsilon_i = \begin{cases} \varepsilon_n, & \text{if } \mathbf{x}_i \in N \\ \varepsilon_0, & \text{otherwise} \end{cases}, \quad (19)$$

where ε_0 is the initial tolerance value. In this case, the corresponding differentials for the new KKT conditions are expressed as

$$\Delta g_i^{(*)} = \pm \left[\sum_{j \in S} \Delta(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) + \Delta b \right] + \Delta \varepsilon_i, \quad (20)$$

where $\Delta \varepsilon_i = \varepsilon_n - \varepsilon_0$. As in the case of adding $(\mathbf{x}_c, y_c)$, the current SVM solution is further perturbed by gradually adjusting the tolerance $\Delta \varepsilon_i$ in Eq. (20) so that in each step the migration across the sets $\{S, M, E\}$ is updated, and this pro-

TABLE II. Local perturbation in relevance feedback.

For each sample $\mathbf{x}_j \in N$ in a close neighborhood of a feedback sample $(\mathbf{x}_c, y_c)$, do the following:
• Migrate samples across adjacent sets $\{S, M, E\}$ by checking bounds on γ_i for all samples in Z , and if S changes, update $\mathbf{R}$ accordingly.
(a) If $\gamma_i = 0$, transfer $(\mathbf{x}_i, y_i)$ from S to M .
(b) If $\gamma_i = C_i$, transfer $(\mathbf{x}_i, y_i)$ from S to E .
(c) If $\gamma_i < C_i$, transfer $(\mathbf{x}_i, y_i)$ from E to S .
(d) If $\gamma_i > 0$, transfer $(\mathbf{x}_i, y_i)$ from M to S .
• Repeat this procedure until convergence is achieved.

cedure is repeated until convergence is achieved. Also, the matrix $\mathbf{R}$ is updated accordingly.

III.C. Local perturbation by C parameter

Another approach is to adiabatically perturb the SVM with respect to the regularization parameter C . Following a similar approach to the algorithm presented in Table I, the C value for the feedback sample (instead of ε) and its neighborhood is replaced with a new C_n value. That is, for each sample in the set N , we modify the regularization parameter C_i as follows:

$$C_i = \begin{cases} C_n, & \text{if } \mathbf{x}_i \in N \\ C_0, & \text{otherwise} \end{cases}, \quad (21)$$

where C_0 is the initial regularization parameter value. Similar to ε perturbation, the new C_i of the neighborhood of the feedback sample will cause perturbation of the current SVM solution. If the set S during the migration across the sets $\{S, M, E\}$ is changed, the matrix $\mathbf{R}$ is updated accordingly. As in the previous case, the objective is to determine the changes in the margin support vectors and the bias while preserving the KKT conditions over all the samples. Table II summarizes the C -based local perturbation procedure for relevance feedback.

IV. PERFORMANCE EVALUATION STUDY

The proposed relevance feedback framework was tested on two data sets collected in our previous development of similarity modeling for content-based mammogram retrieval using supervised machine learning.^{10,11} These two data sets consist of clinical mammogram images that contain MC lesions, which were collected by the Department of Radiology at The University of Chicago. Our goal is to automatically retrieve mammogram images that have perceptually similar lesions to that in a query image. In Fig. 3, we show some examples of ROIs extracted from mammograms in the data sets, all of which contain MC clusters (MCCs). For modeling the perceptual similarity between a pair of lesions, human-observer studies were used in these two data sets. Below we describe briefly these two data sets (referred to as A and B, respectively).

FIG. 3. Examples of mammogram regions containing clustered microcalcifications (indicated by circles).

IV.A. Data set A

This feasibility data set consists of a total of 76 mammogram images, all containing multiple MCs, which have a spatial resolution of 0.1 mm/pixel and 10 bit grayscale. For the observer study, a panel of six human observers who have backgrounds in general medical image analysis scored a total of 435 pairs of randomly selected lesion images on a scale from 0 (most dissimilar) to 10 (most similar) based on the spatial geometric distribution pattern of the MCs in a lesion. Detailed information about this data set can be found in Ref. 35.

IV.A.1. Description of MCC features

To characterize the geometric features of MCCs, the following set of shape descriptors were computed for each MC cluster.³⁵

- Compactness of the cluster: A measure of roundness of the region occupied by the cluster.
- Eccentricity of the cluster: The eccentricity of the smallest ellipse of the region (ratio of the distance between the foci and the major axis).
- The number of MCs per unit area.
- The average of the interdistance between neighboring MCs.
- The standard deviation of the interdistance between neighboring MCs.
- Solidity of the cluster region: The ratio between cross-sectional area and the area of the convex hull formed by the MCs.
- The moment signature of the cluster region: Computed based on the distance deviation of the boundary point from the center of the region.
- Cross-sectional area: The area occupied by the cluster.
- Invariant moment: A regional descriptor that is invariant to translation, rotation, or scaling.⁴⁰
- Normalized Fourier descriptor: A frequency-domain characterization of the smoothness of the boundary.⁴¹

All these feature components were then normalized to have the same dynamic range (0,1). Each MCC was then labeled with a feature vector $\mathbf{u}$ formed by these components. Then, two feature vectors corresponding to a given pair of images are concatenated into one vector, which forms a sample together with its observer similarity score. In summary, we have the following data set:

FIG. 4. MDS plot of scores of six observers. No. 7 indicates the average of the six observers; No. 8 is a random observer.

$$Z = \{(\mathbf{x}_i, y_i)\}_{i=1}^l, \quad (22)$$

where $\mathbf{x}_i$ denotes the computed feature vector for the i th MCC pair and y_i is the observer similarity score of the pair. This set was used for the subsequent training and testing of the proposed framework.

IV.B. Data set B

This data set consists of 200 mammogram images from 104 patients with known pathology (46 malignant and 58 benign), all containing multiple MCs and all having a spatial resolution of 0.1 mm/pixel and 10 bit grayscale. For the observer study, ROIs containing MCCs were first extracted from all the mammograms by radiologists. Based on the MCC features, these 200 images were clustered into ten different groups using the k -means algorithm. Based on the clustering results, a total of 300 pairs were randomly selected from the same group and another set of 300 pairs were randomly selected from two different groups. These 600 image pairs were scored by a panel of six expert mammogram readers; different from the feasibility data set A described above, here the image features of individual MCs (in addition to their geometric distribution) were also taken into account by the readers as in their clinical interpretation of mammograms. To examine both intraobserver and interobserver consistencies of the observer ratings, statistical analyses were conducted. Based on these analyses, four observers (Nos. 2, 3, 5, and 6) with the highest intraobserver consistency were selected and their scores were averaged for each of the 600 image pairs. The resulting scores were then used to form training and testing samples. To summarize, in Fig. 4 we show a multidimensional scaling⁴² (MDS) plot of the six observer ratings. In this plot, the original data points are mapped onto a reduced 2D space so that the mapped points indicate their interdistance relationship in their original space. For comparison, the average of all the observers is shown (No. 7), and a random observer is also shown (No. 8) for which random scores were assigned for each of the 600 image pairs. As can be seen, the six observers are more close to each other compared to the random observer; moreover,

the four most consistent observers (Nos. 2, 3, 5, and 6) are also closer to each other than the other two (Nos. 1 and 4).

IV.B.1. Description of MCC features

Besides the geometric distribution features used in data set A above, the following additional features were introduced in order to characterize the image features of individual MCs:^{11,43,44}

- The number of MCs in the cluster.
- The mean effective volume (area times effective thickness) of individual MCs.
- The relative standard deviation of the effective thickness.
- The relative standard deviation of the effective volume.
- The second highest MC-shape-irregularity measure.

These feature components were normalized to have the same dynamic range (0,1). Altogether, there are a total of 12 features for each cluster. The feature vectors for each pair of MCCs were then paired with their similarity score to form a sample.

IV.C. Machine training and performance evaluation

For training and testing of the learning machine, we applied the following leave-one-out cross-validation (LOO-CV) procedure. First, the images were selected in a round-robin fashion so that during each round only a single image was chosen as the query. Thus, the data samples were divided into two sets: One set for training, which consisted of all the samples not involving the current query image, and a second set for testing, which consisted of only those samples corresponding to the current query image. Specifically, let $\mathbf{R}$ denote a set of all samples in our data set and $\mathbf{Q}$ denote a set of all samples including only an image $\mathbf{q}$. In the LOO-CV for a query $\mathbf{q}$, the top scored samples are selected from $\mathbf{Q}$. Let $\mathbf{Q}_1$ denote a top scored sample. In the relevance feedback process with $\mathbf{Q}_1$, parameters that are necessary to calculate SVM regression are updated using $\mathbf{Q}_1$ and $\mathbf{R}-\mathbf{Q}$ samples in the database according to the proposed incremental learning procedure described in Sec. III. Then, the updated SVM regression function is used to calculate the new similarity coefficients for the $\mathbf{Q}$ samples, which are retrieved accordingly. The same recursive procedure is applied in case of multiple feedback samples. Second, this process was repeated in each round for every query image in the data set. Finally, the testing results after LOO-CV were then averaged over all the different rounds to estimate the performance generalization metrics.

To evaluate the performance of the proposed retrieval framework, we used the so-called *precision-recall curves*.¹ The retrieval *precision* is defined as the proportion of the images retrieved that are truly relevant to a given query; the term *recall* is measured by the proportion of the images that are actually retrieved among all the relevant images to a query. Mathematically, they are given by

$$\text{Precision} = \frac{\text{number of relevant images that are retrieved}}{\text{total number of retrieved images}},$$

$$\text{Recall} = \frac{\text{number of relevant images that are retrieved}}{\text{total number of relevant images}}. \quad (23)$$

The precision-recall curve is a plot of the retrieval precision versus the recall over a continuum of the operating threshold. To calculate the precision and recall, we need to decide a threshold that indicates whether or not a retrieved image is relevant to a query image. As the ground truth in these calculations, we considered an image to be truly relevant to a query provided that its corresponding observer similarity score is larger than a preselected threshold T_2 . In our data sets, the degree of similarity was described using a scale from 0 (most dissimilar) to 10 (most similar). Accordingly to this scale, a value of $T_2=7$ was used as the threshold for deciding whether or not a retrieved image is relevant to a query image. This value seems to reflect a high degree of similarity on this scale and it is consistent with our previous studies.^{11,35}

To evaluate the efficacy of the proposed relevance feedback method, we performed the following experiments: For each query, the trained retrieval network (offline learning) was first applied to retrieve images from the database; among the images retrieved, the one with the highest SC (based on the pre-existing observer data) was chosen as the relevant feedback image. This chosen image was then paired up with the query along with their corresponding observer similarity score to form the feedback sample. The proposed relevance feedback procedure was then applied to retrieve a new set of images. The precision-recall curves were then computed based on this new set of images. In our experiments, the SVM was trained with a radial basis function of width $\sigma=1.5$ and $C=100$ using the observer data (offline learning) as in our previous work.⁶

V. RESULTS AND DISCUSSION

We tested the proposed relevance feedback method for improving performance of image retrieval systems in mammography using a feasibility data set A scored by nonexperts³⁵ and clinical data set B scored by expert radiologists. In each case, we used LOO-CV procedure described in Sec. IV C.

V.A. Study on demonstrative data set A

As a demonstration, we evaluated the proposed relevance feedback system using data set A that consists of 76 mammogram images. We compared the performance of the relevance feedback system with one, three, and five feedback samples, to that when no feedback sample was used (offline case). Figure 5 shows the precision-recall curves using our relevance feedback system with $\varepsilon_n=0.5$. As expected, as the number of feedback samples increased, further improvement in the performance was observed. In particular, with only one

FIG. 5. Precision-recall curves using relevance feedback with 76 mammogram images. RFB stands for relevance feedback.

feedback sample, the performance was improved considerably compared to the offline case. However, for the three and five feedback samples, no distinctive difference was observed. These data demonstrated the feasibility of the proposed method and encouraged investigating its application on mammogram images scored by expert radiologists as discussed below.

In order to test the improvement in performance statistically, we applied a bootstrapping procedure. From data set A, we generated 20 000 bootstrap data sets using sampling with replacement, each of which consisted of 30 samples. The area under the precision-recall curve (AUPRC) was computed for each bootstrap data set for the offline SVM and relevance feedback systems. A paired t -test was performed to compare the offline SVM and our relevance feedback systems, which yielded p -value=0.002 after Bonferroni correction for multiple comparisons in all cases.

V.B. Study on clinical data set B

V.B.1. Relevance feedback with local perturbation

To test the proposed relevance feedback system based on incremental learning with SVM regression, we performed experiments with several ε_n and C_n values for the feedback samples and their neighboring samples. Figure 6 shows precision-recall curves for the different number of feedback samples when the averaged scores of observers 2, 3, 5, and 6 were used. As can be seen, in general, as the number of feedback samples increased, better performance was attained. In particular, when $C_n=1$, a slightly better performance was achieved compared to $C_n=150$. Interestingly, for C_n values greater than 150, the performance remained unchanged (results not shown). For ε_n parameter, choosing $\varepsilon_n=0.5$ provided the best performance compared to other ε_n and C_n values. Note that when $\varepsilon_n=2$ was used, the performance degraded dramatically. It is obvious that the parameter ε greatly influences the performance since it directly regulates the width of the margin tube in ε -SVM (cf. Fig. 2).

FIG. 6. Precision-recall curves using relevance feedback with different parameter values.

For our relevance feedback system with $C_n=1$ and $\varepsilon_n=0.5$, tenfold CV was performed. Overall AUPRC after tenfold CV was less than that of LOO-CV. For example, for the relevance feedback system with five feedback samples, AUPRC was 0.73 and 0.79 after tenfold CV and LOO-CV, respectively. The same bootstrapping procedure that was used with data set A was applied using the relevance feedback system with $C_n=1$ and $\varepsilon_n=0.5$. From data set B, we generated 20 000 bootstrap data sets using sampling with replacement, each of which consists of 200 samples. Histo-

grams of AUPRC for these 20 000 bootstrap data sets are shown in Fig. 8. In comparison of the offline SVM and our relevance feedback systems, a paired t -test yielded $p=0.002$ after multiple comparisons correction in all cases.

Figure 7 displays the AUPRC using default parameter values for C_n and ε_n (in this study, $C_n=1$ and $\varepsilon_n=0.5$), and using both $C_n=1$ and $\varepsilon_n=0.5$. In addition, we tested our relevance feedback system with both $C=100$ and $\varepsilon=1$ (parameter values used for training the SVM regression) without local perturbation to the neighborhood of the feedback

FIG. 7. AUPRC histograms using relevance feedback with different parameter values. The “w/o LP” means “without local perturbation.”

sample (allowing only perturbation by the feedback sample). It is quite clear that with local perturbation, the performance is greatly improved. We also observed that the performance with both $C_n=1$ and $\varepsilon_n=0.5$ is better compared to the results obtained when $C_n=1$ or $\varepsilon_n=0.5$ was chosen alone. It suggests that not only the selection of a good parameter value is important, but also it is worthy that we apply proper combination of parameter values. Using a two-way ANOVA test, $p=0.0186$ and $p<0.0001$ were obtained for the parameters and the relevance feedback systems, respectively.

In terms of runtime efficiency, the incremental learning relevance feedback system was around 56 times faster than the full SVM training while achieving similar effectiveness, using MATLAB scripting version 7.5 and running under a Windows XP system on a personal computer with dual Intel CPUs of 3 GHz and 3 GB RAM. This improvement in speed while maintaining effectiveness was also demonstrated in our earlier work by learning the regression of a toy sinc function example from both sequentially and randomly selected data points.²⁶

All the results above were obtained by assuming that all the feedback samples were selected and used all at once for

incremental learning. We also considered an alternative strategy for successive learning as follows. For each query, the top image selected by the user among the retrieved images was initially used as a feedback sample for incremental learning. Next, the updated SVM was applied to retrieve the images again for the same query. Then, the top retrieved image was used again as a feedback sample (if it was not already selected in the previous round, otherwise, the second top image was selected). This successive selection procedure was repeated for each query image in the database. In the experiments with this scenario, however, we did not yield distinctive improvement compared to the original proposed scenario. This could be attributed to the fact that the number of similar images to the query is somewhat limited in data sets we used and the performance improvement becomes saturated after the first few similar images are retrieved.

V.B.2. Adaptation to individual observer's scores

Figure 9 illustrates precision-recall curves that result from experiments with $\varepsilon_n=0.5$ when each observer's scores were used instead of the averaged scores as in the previous experiments. Figure 10 shows corresponding AUPRC histograms. When the scores of observer 6 were used, the best AUPRC was achieved, while the worst AUPRC was in case of observer 2. It is interesting to note that AUPRC values of the average observer are relatively close to those of observers 3 and 5 and are relatively distant from those of observers 2 and 6, which consorts with the results of MDS in Fig. 4. In a two-way ANOVA test, $p<0.0001$ and $p<0.0001$ were obtained for the observers and the relevance feedback systems, respectively. Figure 11 displays a sample case to demonstrate how well our proposed image retrieval system works. Given a query image, the top row represents the top three images retrieved after relevance feedback. In contrast, the bottom row shows results by the offline trained SVM. The numbers in brackets are the observer similarity score (left) and the machine response score (right). In the offline training, the highest similarity image with an observer score of 7.50 was ranked second. On the other hand, after relevance feedback, the image was ranked first with a closer machine score to the corresponding observer score. Also, note that overall, the retrieved images were well ranked with respect to their observer scores.

V.B.3. Relationship to clinical pathology

Figure 12 shows the average matching fraction for the top k -retrieved images ($k=1, 2, 3, 4$, and 5) that actually match the disease condition of the query (benign versus malignant) after the relevance feedback with one, three, and five feedback samples, each of which was analyzed using LOO-CV evaluation. For comparison, we also displayed the matching fraction when the observer score (ground truth) and the score by the offline trained SVM were used. It is clear that overall, the average matching fraction evaluated from the image retrieval system trained by relevance feedback is higher than that by the observer score and the offline trained SVM,

FIG. 8. Histograms of AUPRC for 20 000 bootstrap experiments.

FIG. 9. Precision-recall curves showing adaptation to individual observer's scores.

achieving a matching fraction as high as 82.4% that is a significantly improved result compared to 72.5% in our previous work.¹¹ In this case, the improvement was more pronounced for the small number of retrieved images (one, two, and three) due to the limited size of the used database. We believe that this improvement could be attributed to the fact that quantitative image features were used in the similarity

model, which may actually be more consistent and descriptive of the underlying pathology; another factor could be that the learned similarity model has a smoothing effect in which it could remove some of the random variations in the observer scores. In a two-way ANOVA test, $p < 0.0001$ and $p = 0.0058$ were obtained for the number of retrieved images and the relevance feedback systems, respectively.

FIG. 10. AUPRC histograms using $\epsilon_n = 0.5$ and each observer's scores instead of averaged scores.

FIG. 11. An example to show the effectiveness of relevance feedback. Given a query image, the top three images (top row) retrieved using relevance feedback and retrieval results (bottom row) by the offline trained SVM are shown.

FIG. 12. Average matching fraction for the top k retrieved images ($k=1, 2, 3, 4$, and 5) that actually match the disease condition of the query.

V.B.4. Advantages over existing relevance feedback systems

Compared to other existing relevance feedback techniques, the proposed method has the following salient features:

- (1) The proposed method enables efficient online learning by adapting the current trained model to changes prompted by the user's relevance feedback, which improves the performance of the image retrieval system over time.
- (2) In contrast to retraining the whole system each time a feedback sample is provided, the proposed CBIR system is trained efficiently by using a new incremental learning process.
- (3) Our experimental results indicate that even with only one feedback sample, the proposed system can show significant improved performance due to the local perturbation scheme employed.

VI. CONCLUSION

In this paper, we have presented a novel relevance feedback scheme to improve the effectiveness of retrieval of relevant mammogram images using clinical data scored by experienced radiologists. To integrate relevance feedback into mammogram image retrieval in a practical way, incremental learning with SVM regression was proposed. The basic idea of incremental learning is to achieve performance similar to that of full retraining of the retrieval system while reducing computational time substantially. The proposed framework is computationally efficient since it trains the SVM incrementally on the relevance feedback samples. Besides explicit relevance feedback, we also introduced local perturbation by regression parameters to further improve the performance. Our experimental results demonstrated that the proposed approach can achieve higher efficiency while maintaining its effectiveness for online relevance feedback application to mammographic databases. In particular, compared to our previous work, the 10% improved matching fraction (from

72.5% to 82.4%) was achieved. For the clinical use of the proposed image retrieval system, we need to collect more mammogram images, score them by radiologists, and train the system with the proposed relevance feedback approach. We expect that the well-trained system will help radiologists more accurately diagnose breast cancer patients.

ACKNOWLEDGMENTS

This work was supported in part by NIH Grant Nos. EB009905 and CA128809.

^{a)} Author to whom correspondence should be addressed. Electronic mail: ielnaqa@medphys.mcgill.ca; Present address: Medical Physics Unit, McGill University, Montréal, Québec H3G 1A4, Canada.

¹ A. D. Bimbo, *Visual Information Retrieval* (Morgan Kaufman, San Francisco, 1999).

² Y. Rui and T. Huang, "Image retrieval: Current techniques, promising directions and open issues," *J. Visual Commun. Image Represent* **10**, 39–62 (1999).

³ X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive reviews," *Multimedia Syst.* **8**, 536–544 (2003).

⁴ D. Tao, X. Li, and S. J. Maybank, "Negative samples analysis in relevance feedback," *IEEE Trans. Knowl. Data Eng.* **19**(4), 568–580 (2007).

⁵ R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval* (Addison-Wesley Longman Publishing Co., Inc., Boston, 1999).

⁶ I. El Naqa, Y. Yang, N. P. Galatsanos, and M. N. Wernick, "Content-based image retrieval for digital mammography," in *Proceedings of the IEEE International Conference on Image Processing*, 2002, pp. 141–144.

⁷ T. Kurita and T. Kato, "Learning of personal visual impression for image database systems," in *Proceedings of the International Conference on Document Analysis and Recognition*, 1993, pp. 547–552.

⁸ Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Trans. Circuits Syst. Video Technol.* **8**(5), 644–655 (1998).

⁹ S. Wong, "CBIR in medicine: Still along way to go," in *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1998, pp. 115.

¹⁰ I. El Naqa, Y. Yang, N. P. Galatsanos, and M. N. Wernick, "Relevance feedback based on incremental learning for mammogram retrieval," in *Proceedings of the IEEE International Conference on Image Processing*, 2003, pp. 729–732.

¹¹ L. Wei, Y. Yang, M. N. Wernick, and R. M. Nishikawa, "Learning of perceptual similarity from expert readers for mammogram retrieval," *IEEE J. Sel. Top. Signal Process.* **3**(1), 53–61 (2009).

¹² J. Peng, "Multi-class relevance feedback content-based image retrieval," *Comput. Vis. Image Underst.* **90**, 42–67 (2003).

¹³ C. Zhang and T. S. Chen, "An active learning framework for content-based information retrieval," *IEEE Trans. Multimedia* **4**, 260–268 (2002).

¹⁴ G. D. Guo, A. K. Jain, W. Y. Ma, and H. J. Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. Neural Netw.* **13**(4), 811–820 (2002).

¹⁵ E. Cheng, F. Jing, and L. Zhang, "A unified relevance feedback framework for web image retrieval," *IEEE Trans. Image Process.* **18**(6), 1350–1357 (2009).

¹⁶ P. Y. Yin, B. Bhanu, K. C. Chang, and A. Dong, "Long-term cross-session relevance feedback using virtual features," *IEEE Trans. Knowl. Data Eng.* **20**(3), 352–368 (2008).

¹⁷ D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(7), 1088–1099 (2006).

¹⁸ C.-Y. Li and C.-T. Hsu, "Image retrieval with relevance feedback based on graph-theoretic region correspondence estimation," *IEEE Trans. Multimedia* **10**(3), 447–456 (2008).

¹⁹ M. R. Azimi-Sadjadi, J. Salazar, and S. Srinivasan, "An adaptable image retrieval system with relevance feedback using kernel machines and selective sampling," *IEEE Trans. Image Process.* **18**(7), 1645–1659 (2009).

²⁰ G. D. Tourassis, B. Harrawood, S. Singh, J. Y. Lo, and C. E. Floyd, "Evaluation of information-theoretic similarity measures for content-

- based retrieval and detection of masses in mammograms," *Med. Phys.* **34**, 140–150 (2007).
- ²¹B. 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.* **33**, 111–117 (2006).
 - ²²S. K. Kinoshita, P. M. de Azevedo-Marques, R. R. Pereira, J. A. Rodrigues, and R. M. Rangayyan, "Content-based retrieval of mammograms using visual features related to breast density patterns," *J. Digit. Imaging* **20**, 172–190 (2007).
 - ²³H. A. Swett, P. G. Mutalik, V. P. Neklesa, L. Horvath, C. Lee, J. Richter, I. Tocino, and P. R. Fisher, "Voice-activated retrieval of mammography reference images," *J. Digit. Imaging* **11**, 65–73 (1998).
 - ²⁴J. Sklansky, E. Y. Tao, M. Bazargan, C. J. Ornes, R. C. Murchison, and S. Teklehaimanot, "Computer-aided, case-based diagnosis of mammographic regions of interest containing microcalcifications," *Acad. Radiol.* **7**, 395–405 (2000).
 - ²⁵H. Qi and W. E. Snyder, "Content-based image retrieval in picture archiving and communications systems," *J. Digit. Imaging* **12**, 81–83 (1999).
 - ²⁶I. El Naqa, "Content-based image retrieval by similarity learning for digital mammography," Ph.D. thesis, Electrical Engineering, Illinois Institute of Technology, Chicago, IL, 2002.
 - ²⁷I. El Naqa, M. Wernick, Y. Yang, and N. Galatsanos, "Image retrieval based on similarity learning," Proceedings of the IEEE International Conference on Image Processing, 2000, pp. 722–725.
 - ²⁸M. A. Mazurowski, P. A. Habas, J. M. Zurada, and G. D. Tourassi, "Decision optimization of case-based computer-aided decision systems using genetic algorithms with application to mammography," *Phys. Med. Biol.* **53**, 895–908 (2008).
 - ²⁹G. D. Tourassi, R. Ike, S. Singh, and B. Harrawood, "Evaluating the effect of image preprocessing on an information-theoretic CAD system in mammography," *Acad. Radiol.* **15**, 626–634 (2008).
 - ³⁰S. C. Park, R. Sukthankar, L. Mummert, M. Satyanarayanan, and B. Zheng, "Optimization of reference library used in content-based medical image retrieval scheme," *Med. Phys.* **34**, 4331–4339 (2007).
 - ³¹P. Mazzoncini de Azevedo-Marques, N. A. Rosa, A. J. M. Traina, C. Traina, Jr., S. K. Kinoshita, and R. M. Rangayyan, "Reducing the semantic gap in content-based image retrieval in mammography with relevance feedback and inclusion of expert knowledge," *Int. J. CARS* **3**, 123–130 (2008).
 - ³²I. El Naqa, L. Wei, and Y. Yang, *Ubiquitous Health and Medical Informatics: The Ubiquity 2.0 Trend and Beyond* (IGI Global, Hershey, 2010).
 - ³³V. Vapnik, *Statistical Learning Theory* (Wiley, New York, 1998).
 - ³⁴D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Netw.* **2**(16), 568–576 (1991).
 - ³⁵I. El Naqa, Y. Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N. Wernick, "A similarity learning approach to content-based image retrieval: Application to digital mammography," *IEEE Trans. Med. Imaging* **23**(10), 1233–1244 (2004).
 - ³⁶S. Chen, S. Zhou, F. F. Yin, L. B. Marks, and S. K. Das, "Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis," *Med. Phys.* **34**, 3808–3814 (2007).
 - ³⁷G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," *Adv. Neural Inf. Process. Syst.* **13**, 409–415 (2001).
 - ³⁸C. P. Diehl and G. Cauwenberghs, "SVM incremental learning, adaptation and optimization," in Proceedings of the IEEE International Joint Conference Neural Networks, 2003, pp. 2685–2690.
 - ³⁹N. Syed, H. Liu, and K. Sung, "Incremental learning with support vector machines," in Proceedings of the International Joint Conference on Artificial Intelligence, 1999.
 - ⁴⁰R. C. Gonzalez and R. E. Woods, *Digital Image Processing* (Addison-Wesley, Reading, 1992).
 - ⁴¹L. Shen, R. M. Rangayyan, and J. L. Desautels, "Application of shape analysis to mammographic calcifications," *IEEE Trans. Med. Imaging* **13**, 263–274 (1994).
 - ⁴²I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications* (Springer-Verlag, New York, 1997).
 - ⁴³H.-P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt, and D. D. Adler, "Computerized analysis of mammographic microcalcifications in morphological and texture feature space," *Med. Phys.* **25**, 2007–2019 (1998).
 - ⁴⁴Y. Jiang, R. M. Nishikawa, and J. Papaioannou, "Dependence of computer classification of clustered microcalcifications on the correct detection of microcalcifications," *Med. Phys.* **28**, 1949–1957 (2001).