

# CONTENT-BASED IMAGE RETRIEVAL FOR DIGITAL MAMMOGRAPHY<sup>1</sup>

Issam El Naqa, Yongyi Yang, Nikolas P. Galatsanos, and Miles N. Wernick

Dept. of Electrical and Computer Engineering, Illinois Institute of Technology  
3301 S. Dearborn Street, Chicago, IL 60616

## ABSTRACT

*In this work, we explore the use of a learning-based framework for retrieval of relevant mammogram images from a database, for purposes of aiding diagnoses. A fundamental issue is how to characterize the notion of similarity between images for use in assessing relevance of images in the database. We investigate the use of several learning algorithms, namely, neural networks and support vector machines, in a two-stage hierarchical learning network for predicting the perceptual similarity from similarity scores collected in human-observer studies. The proposed approach is demonstrated using microcalcification clusters extracted from a database consisting of 76 mammograms. Initial results demonstrate that the proposed two-stage hierarchical learning network outperforms a single-stage learning network.*

## 1. INTRODUCTION

With the growing volume of images used in medicine, the capability to retrieve relevant images from large databases is becoming increasingly important. Despite the progress made in the general area of image retrieval in recent years (e.g., [1-4]), its success in biomedicine thus far has been quite limited [5].

In this work we investigate the use of content-based image retrieval (CBIR) for digital mammograms. The goal is to provide the radiologist with a set of images from past cases that are relevant to the one being evaluated, along with the known pathology of these past cases. Our expectation is that a library of relevant past images would assist radiologists to better diagnose difficult cases. We expect such a facility to be useful in medical education and training as well.

The goal of CBIR is to obtain from a possibly very large image database those images that are similar in content to an image of interest (the *query image*). The key in CBIR is to apply appropriate similarity metrics for ranking the relevance of images in a database to the query image. A popular approach is to employ measured image features to provide a description of the content of the image. The image features may be extracted at either a low level (such as local edges) or at a high level (such as a grey level

histogram), or both [1]. The query image can then be compared to the images in the database on the basis of their features. Those images in the database having the highest similarity to the query image are retrieved.

A challenging task in such an approach is how to define a measure of similarity that corresponds to similarity as judged by the user. The notion of similarity of images is highly dependent on the application domain. In this work, we propose to develop a CBIR system that can retrieve mammograms that have “perceptually similar” lesions to those in a query mammogram, the purpose being to identify mammograms that are medically relevant. Clearly, the similarity metric in this application must conform closely to the notion of similarity used by radiologists when they interpret mammograms. It is reasonable to expect that simple, mathematical distance metrics will not be adequate.

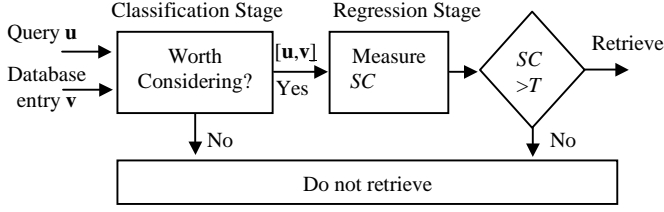
We propose to model the perceptual similarity between two lesion images as a nonlinear function of those images, and use learning algorithms (support vector machines [6] and neural networks) to learn this function from similarity scores collected in human-observer studies. The basic idea of modeling perceptual similarity of images by a learning algorithm for image retrieval was initially developed in our previous work in [7]. This idea is further expanded in this work by using a two-stage hierarchical learning network for retrieval of mammograms.

## 2. METHODOLOGY

The proposed two-stage hierarchical retrieval framework is illustrated in Fig. 1. It consists of a cascade of a binary classifier stage and a regression stage for predicting the similarity coefficients (*SCs*) between a query image and those in the database. The purpose of using such a two-stage network is as follows. First, it is often the case that the majority of the images in the database are not similar at all to a given query image; thus, the first-stage classifier can speed up the retrieval by quickly disqualifying many mammogram images from further consideration. Second, the learning machine in the second-stage can be fine-tuned to predict *SCs* accurately only for image pairs that are at least somewhat similar. Those images surviving the first

---

<sup>1</sup> This research was supported by the NIH/NCI under grant CA89668.



**Figure 1.** The proposed two-stage retrieval framework.

stage are then compared to the query in this stage to obtain a numerical  $SC$ . Images with the highest  $SC$ s are retrieved. During the training phase, the first-stage classifier functions as a coarse, binary learner, of which the purpose is to simply identify whether a database entry is sufficiently similar to the query for further consideration. On the other hand, the second stage functions as a more refined learner, of which the purpose is to decide quantitatively the similarity between a surviving entry from the first stage and the query.

### 2.1 First-Stage Classifier

Let  $\mathbf{u}$  and  $\mathbf{v}$  denote the feature vectors of the query and a candidate image, respectively. The task of the first-stage classifier is to determine whether the images described by  $\mathbf{u}$  and  $\mathbf{v}$  are sufficiently similar for further consideration. We consider two types of pattern classifiers: (1) a Fisher discriminant, and (2) a support vector machine (SVM).

#### A. Fisher Discriminant Classifier

The Fisher discriminant is based on the principle of projecting the data along the direction along which different classes are well separated [8]. With  $\mathbf{x} = [\mathbf{u}^T \ \mathbf{v}^T]^T$ , the decision function is of the form

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^T \mathbf{x} + w_o \quad (1)$$

where  $\mathbf{w}$  is the discriminant vector, which is determined based on the training data samples [8].

#### B. SVM Classifier

The SVM is a universal constructive learning procedure based on statistical learning theory [6]. For classification, the main idea of the SVM technique is to separate the classes in a mapped high-dimensional feature space with a hyperplane that maximizes the separation margin between them. The discriminant function in such a case is characterized by a subset of the training data known as support vectors:

$$f(\mathbf{x}) = \sum_{i=1}^{l_s} \alpha_i d_i K(\mathbf{x}_i, \mathbf{x}) + \alpha_o \quad (2)$$

where  $l_s$  is the number of support vectors,  $d_i$  are the desired output (+1 or -1) for the support vectors  $\mathbf{x}_i$ , the coefficients  $\alpha_i$  are determined through training, and  $K(\cdot, \cdot)$  is a kernel function (e.g., polynomials or radial basis functions) satisfying the Mercer's conditions [6].

In its original form, an SVM classifier is formulated by minimizing the following cost function (for notational simplicity, we illustrate the concept by the linear case)

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i \quad (3)$$

$$\text{subject to } d_i(\mathbf{w}^T \mathbf{x}_i + w_o) \geq 1 - \xi_i, \text{ and} \\ \xi_i \geq 0 \text{ for all } i = 1, 2, \dots, l,$$

where  $l$  is the total number of training samples,  $C$  is the regularization parameter, and  $\xi_i$  are slack variables.

For our task, we instead use the following modified cost function:

$$\Phi(\mathbf{w}, \xi) = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C^+ \sum_{i:d_i=+1} \xi_i + C^- \sum_{i:d_i=-1} \xi_i \quad (4)$$

where  $C^+ > C^-$ . This allows us to impose a greater penalty ( $C^+$ ) on missed retrievals than on retrieval of non-similar images ( $C^-$ ). The rationale is that the first-stage classifier is for pre-screening only and should be designed to pass marginal cases to the second stage for further consideration.

The dual problem of (5) is maximization of

$$J(\vec{\alpha}) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j d_i d_j \mathbf{x}_i^T \mathbf{x}_j \quad (5)$$

subject to the constraints:

$$(1) \sum_{i=1}^l \alpha_i d_i = 0$$

$$(2) 0 \leq \alpha_i \leq C^+ \text{ for } d_i = +1 \\ 0 \leq \alpha_i \leq C^- \text{ for } d_i = -1.$$

This maximization is accomplished by quadratic programming.

### 2.2 Regression Stage

The user's assessment of similarity between a pair of images is modeled as a function of their feature vectors  $\mathbf{u}$  and  $\mathbf{v}$ . With  $\mathbf{x} = [\mathbf{u}^T \ \mathbf{v}^T]^T$ , we write the following regression model for the similarity coefficient:

$$SC = g(\mathbf{x}) + \varepsilon, \quad (6)$$

where  $\varepsilon$  is the modeling error, and  $g(\mathbf{x})$  is the similarity function to be determined.

In our previous work [7], we applied and evaluated a general regression neural network (GRNN) and an SVM for estimating the similarity function  $g(\mathbf{x})$ . We continue to use these to design the second-stage network.

#### A. GRNN for Regression

The GRNN computes a conditional mean estimator of  $SC$ . It is based on an estimate of the joint probability density of the inputs and the output obtained by the Parzen method [9]. The output of the GRNN can be represented as:

$$g(\mathbf{x}) = \frac{\sum_i d_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right)}{\sum_i \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right)} \quad (7)$$

where  $(\mathbf{x}_i, d_i)$  are input-output pairs of the training examples. The parameter  $\sigma$  is determined empirically during the training phase.

### B. Support Vector Machine (SVM) for Regression

SVM learning can also be applied for nonlinear regression. The SVM formulation in such a case maintains many of the characteristics of the classification case, such as the quadratic form and avoidance of the curse of dimensionality by solving the dual problem. The SVM output function is also characterized by the support vectors as follows:

$$g(\mathbf{x}) = \sum_{i=1}^{l_i} w_i K(\mathbf{x}_i, \mathbf{x}) + w_o, \quad (8)$$

In this case, the following so-called  $e$ -insensitive penalty term is used in place of the slack variables in the cost function in (4):

$$L(\mathbf{x}_i, d_i) = \begin{cases} |d_i - g(\mathbf{x}_i)| - e, & |d_i - g(\mathbf{x}_i)| \geq e \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

### C. Model Training and Parameter Selection

For training of each stage, feature vectors extracted from examples consisting of image pairs are presented to the learning machine along with a  $SC$  for each image pair given by human observers.

To determine the best parameters for each learning machine, we use a statistical re-sampling procedure called *bootstrapping* [10], according to which the learning machine is repeatedly trained and tested systematically using samples randomly drawn from the available training data set. The accuracy of the trained machine is then estimated using this procedure. The model with the best accuracy is adopted.

## 3. EXPERIMENTAL RESULTS

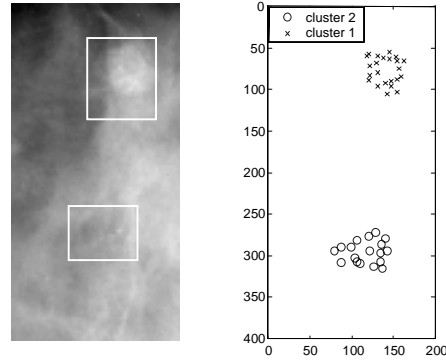
### 3.1 Data Set

The proposed retrieval framework was developed and tested using a database of mammogram images provided by the Department of Radiology at the University of Chicago. It consists of 76 different mammogram images, digitized with a spatial resolution of 0.1 mm/pixel and 10-bit grayscale. All these images contain microcalcification clusters (MCCs). The microcalcifications in this database have been identified by expert radiologists.

Microcalcifications are tiny calcium deposits that appear as small bright spots in a mammogram. MCCs provide information to radiologists for diagnosis of cancer. For example, linearly distributed MCCs are typically

malignant, while round clusters are typically benign [11]. Our objective is to apply the proposed framework to retrieve mammograms containing similar MCCs to that in a query mammogram.

For the training and testing of the two-stage network, MCCs were first extracted from all the mammograms in the database (Fig. 2). These MCCs were then paired up in a random order and scored by a panel of six human observers based on their geometric similarity on a scale from 0 (most dissimilar) to 10 (most similar). A total of 465 such MCC pairs have been scored and recorded. For these 465 pairs those with scores above 5 were used as “similar” samples when training the first-stage classifier, while the rest were used as “not similar” samples. For the regression stage, only the “similar” samples were used.



**Figure 2.** *Left:* a mammogram region, *Right:* extracted microcalcification clusters.

To describe the geometric features of MCCs, the following set of parameters, believed to be relevant to the user’s perception of similarity, were computed for each cluster: (1) area occupied by the cluster; (2) compactness of the cluster; (3) eccentricity of the cluster; (4) the number of microcalcifications per unit area; the average (5) and the standard deviation (6) of the inter-distance among cluster points; (7) solidity of the cluster region; (8) the first invariant moment of the cluster region; (9) the moment signature of the cluster region; and (10) the Fourier descriptor of the cluster boundary. For brevity, the computational details of these parameters are omitted in this proposal. Each of these feature components was normalized to have the same dynamic range (0,1).

### 3.2 Training and Retrieval Results

#### A. Classification Stage

The first-stage classifier was tested using: (1) a Fisher discriminant; (2) an SVM with linear kernel (SVM/linear); (3) an SVM with radial basis function kernel (SVM/rbf); and (4) an SVM using our modified cost function (for both linear and rbf kernels). The best results were achieved by (3) SVM/rbf and (4) MSVM/rbf, which are summarized in Table 1. Both the average and standard deviation of the two-class classification errors (false positive (FP), i.e.,

non-similar pairs for retrieval, and false negative (FN), i.e., missed retrieval) were given in Table 1.

**Table 1.** Misclassification rates by trained classifier

Classifier	Model Parameters	FP		FN	
		Avg.	Std.	Avg.	Std.
SVM/rbf	$\sigma=5, C=1000$	0.072	0.091	0.104	0.079
MSVM/rbf	$\sigma=2.5, C^+/C^-=4$	0.088	0.100	0.087	0.092

### B. Regression Stage

The regression stage was trained using only those image pairs that were classified as similar by the first-stage classifier. The training and testing results are summarized in Table 2, where the *correlation* was computed between the output of the trained network and the *SCs* assigned by the human observers over the test samples. The *precision* of the retrieval [1] was computed as the fraction of the retrieved images that are truly relevant.

**Table 2.** Retrieval results for each stage.

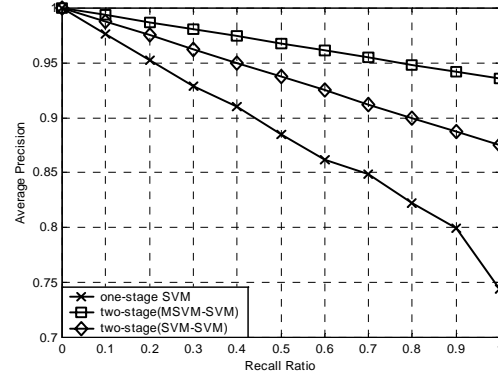
Classification Stage	Regression Stage	Average correlation	Average precision
SVM/rbf $C=1000, \sigma=5$	SVM/rbf $C=20k, \sigma=1.5$	0.802	0.900
MSVM/rbf $\sigma=2.5, C^+/C^-=4$	SVM/rbf $C=20k, \sigma=0.5$	0.776	0.937

### C. Retrieval Performance of Overall Network

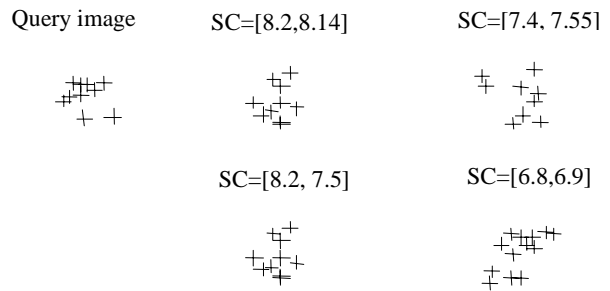
The performance of the resulting overall network is summarized in Fig. 3 using the precision-recall curves for the parametric settings listed in Table 2. A precision-recall curve [1] is a plot of the retrieval precision versus the recall fraction, defined as the fraction of the relevant images that are actually retrieved, over a continuum of the similarity threshold. For comparison, we also show in Fig. 3 the retrieval results when a single-stage SVM was used for the overall learning as in our previous work [7]. Note that the best performance (the top curve) was achieved by the proposed hierarchical two-stage network (modified SVM classifier followed by SVM regression). As an example, we show in Fig. 4 some MCCs retrieved for a given query MCC by both the two-stage and single-stage networks.

## 4. CONCLUSIONS AND FUTURE WORK

In this work, we proposed a hierarchical learning framework for retrieval of relevant mammogram images. Our most current work (not included in this proposal due to space limitation, but will be presented at the conference) demonstrates that the performance of this approach can be further improved by use of relevance feedback. Currently we are studying the practical application of the proposed CBIR framework to a mammogram database based on similarity data to be given by experienced radiologists.



**Figure 3.** Precision-recall comparative study.



**Figure 4.** First row: top two MCCs retrieved by the two-stage network; Second row: top two MCCs retrieved by the single-stage network. Numbers in brackets on top of each cluster are the User SCs (left) and the Machine SCs (right).

## 5. REFERENCES

- [1] A. D. Bimbo, *Visual Information Retrieval*, Morgan Kaufman Publishers, 1999.
- [2] T. Kato *et al.*, "A sketch retrieval method for full color image database," *11<sup>th</sup> Int. Conf. Patt. Recog.*, 1992.
- [3] W. Niblack *et al.*, "The QBIC project: querying images by content using color, texture, and shape," *SPIE Storage and Retrieval Image and Video Databases*, 1994.
- [4] Y. Rui and T. Huang, "Image retrieval: current techniques, promising directions and open issues", *J. Vis. Comm. and Imag. Rep.*, vol. 10, 1999.
- [5] S. Wong, "CBIR in medicine: still a long way to go," *IEEE workshop on Content-Based Access of Image and Video Libraries*, Santa Barbara, CA, June, 1998.
- [6] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.
- [7] I. El Naqa, M. N. Wernick, Y. Yang, and N. P. Galatsanos, "Image retrieval based on similarity learning," *IEEE Int. Conf. Imag. Proc.*, Vancouver, Canada, 2000.
- [8] B. Ripley, *Pattern Recognition and Neural Networks*, Cambridge, 1996.
- [9] D. F. Specht, "A general regression neural network," *IEEE Trans. Neural Network*, vol. 2, no. 16, 1991.
- [10] R. L. Kennedy *et al*, *Solving Data Mining Problems through Pattern Recognition*, Prentice Hall, 1998.
- [11] A American College of Radiology, *Illustrated Breast Imaging Reporting and Data System*, 3<sup>rd</sup> Ed., 1998.