

# Learning in Region-Based Image Retrieval with Generalized Support Vector Machines

Iker Gondra  
Oklahoma State University  
Computer Science Department  
Stillwater, OK 74078-1053  
gondra@cs.okstate.edu

Douglas R. Heisterkamp  
Oklahoma State University  
Computer Science Department  
Stillwater, OK 74078-1053  
drh@ieee.org

## Abstract

*Relevance feedback approaches based on support vector machine (SVM) learning have been applied to significantly improve retrieval performance in content-based image retrieval (CBIR). Those approaches require the use of fixed-length image representations because SVM kernels represent an inner product in a feature space that is a non-linear transformation of the input space. Many region-based CBIR approaches create a variable length image representation and define a similarity measure between two variable length representations. The standard SVM approach cannot be applied to this approach because it violates the requirements that SVM places on the kernel. Fortunately, a generalized SVM (GSVM) has been developed that allows the use of an arbitrary kernel. In this paper, we present an initial investigation into utilizing a GSVM-based relevance feedback learning algorithm. Since GSVM does not place restrictions on the kernel, any image similarity measure can be used. In particular, the proposed approach uses an image similarity measure developed for region-based, variable length representations. Experimental results over real world images demonstrate the efficacy of the proposed method.*

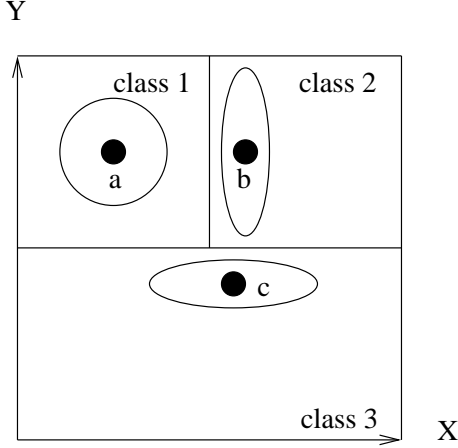
## 1. Introduction

In traditional approaches to content-based image retrieval (CBIR), images are represented by a set of global features and retrieval is performed based on similarity in the feature space. In contrast, region-based approaches [1, 14, 22] extract features from segmented regions of an image. Then, images are retrieved according to similarity among regions. The main objective of using regions is to do a more meaningful retrieval that is closer to a user's perception of an image's content. Instead of looking at an image as a whole, we look at the objects in the image and their

relationships.

Both Blobworld [1] and Netra [14] require the user to select the region(s) of interest from the segmented image. A major problem with this approach is that segmented regions usually do not correspond to actual objects in the image. In order to overcome the problems of inaccurate segmentation, approaches have been proposed that consider all regions in an image for determining similarity [3, 13, 22]. In [13], integrated region matching (IRM) is proposed as a measure that allows a many-to-many region mapping relationship between two images by matching a region of one image to several regions of another image. Thus, by having a similarity measure which is a weighted sum of distances between all regions from different images, IRM is more robust to inaccurate segmentation. Recently, a fuzzy logic approach, unified feature matching (UFM)[3] was proposed as an alternative to IRM. An image is represented by a set of segmented regions each of which is represented by a set of fuzzy features denoting color, texture, and shape characteristics. Because fuzzy features can characterize the gradual transition between regions in an image, segmentation-related inaccuracies are implicitly considered. The similarity between two images is then defined as the overall similarity between two sets of fuzzy features.

Relevance feedback (RF) learning is a common approach that attempts to reduce the semantic gap between high-level concepts and low-level features. It works by gathering semantic information from user interaction. The user labels each image returned in the previous query round as relevant or non-relevant (or a range of values). Based on this feedback, the retrieval scheme is adjusted and the next set of images is presented to the user for labelling. Two main RF learning approaches have been used: query modification, and distance reweighting. Query modification changes the representation of the user's query in a form that is closer (hopefully) to the semantic intent of the user. Distance reweighting changes the calculation of image to image sim-



**Figure 1.** Features are unequal in their differential relevance for computing similarity. The neighborhoods of queries b and c should be elongated along the less relevant Y and X axis respectively. For query a, features X and Y have equal discriminating strength

ilarity to strengthen the contribution of relevant image components in regard to the current query. Probabilistic feature relevance learning (PFRL) [18] is an effective distance reweighting technique that adaptively computes local feature relevance in CBIR systems that use global image representations. It computes flexible retrieval metrics for producing neighborhoods that are elongated along less relevant feature dimensions and constricted along most influential ones (See Figure 1). In [9], we presented an algorithm inspired by PFRL, probabilistic region relevance learning (PRRL). It is based on the observation that regions in an image have unequal importance for computing image similarity. It performs distance reweighting by estimating the relevance of each region in an image based on user’s feedback.

RF schemes based on support vector machine (SVM) [20, 21] learning have been applied to significantly improve retrieval performance in CBIR systems that use global image representations [4, 11, 24]. In [4], relevant images are used to estimate the distribution of target images by fitting a tight hypersphere in the non-linearly transformed feature space. In [24], the problem is regarded as a two-class classification problem and a maximum margin hyperplane in the non-linearly transformed feature space is used to separate relevant images from non-relevant images. Many other approaches, such as [10, 25], have provided improved approaches for utilizing kernel methods and SVMs in CBIR. However, all of these approaches require a valid Mercer kernel. That is, the kernel must satisfy the Mercer conditions [5]. Many region-based CBIR approaches create a variable length image representation and define a similarity measure

between two variable length representations. Thus the standard SVM approach cannot be applied because it violates the requirements that SVM places on the kernel.

To resolve the issue of common SVM kernels not allowing variable-length representations, the following generalization of the Gaussian kernel was introduced in [12]

$$K_{GGaussian}(\mathbf{x}, \mathbf{y}) = e^{-\frac{d(\mathbf{x}, \mathbf{y})}{2\sigma^2}} \quad (1)$$

where  $d$  is a distance measure in the input space between two variable-length image representations  $\mathbf{x} = \{\mathbf{R}_i\}_1^n$  and  $\mathbf{y} = \{\mathbf{R}_i\}_1^m$ , where  $\mathbf{R}_i$  represents the features extracted from a region in the image. Then, using a particular form of (1) with  $d$  being the Earth Mover’s distance (EMD) [19] is proposed. The EMD computes the distance between two distributions represented by sets of weighted features. It is the minimal cost of changing one distribution into the other. The cost is defined in terms of a user-defined ground distance that measures the distance between two features. A distribution can have any number of features. Therefore, EMD can operate on variable-length representations of distributions. Thus an image can be seen as a distribution with a variable number of regions. Then, the kernel proposed in [12] is

$$K_{GEMD}(\mathbf{x}, \mathbf{y}) = e^{-\frac{EMD(\mathbf{x}, \mathbf{y})}{2\sigma^2}}$$

where  $EMD(\mathbf{x}, \mathbf{y})$  is the EMD distance. The ground distance between two regions  $d(\mathbf{R}_i, \mathbf{R}_j)$  is set to the Euclidean distance. In order for EMD to be a true metric, the ground distance must be a metric [19]. Therefore, this approach does not allow for arbitrary image similarity measures.

A generalized SVM (GSVM) [16] allows the use of an arbitrary kernel. In this paper, we propose using a GSVM-based RF learning algorithm that can be applied to region-based CBIR systems that use arbitrary similarity measures. The rest of the paper is organized as follows. In Section 2 we give a brief overview of GSVM. The proposed learning algorithm is presented in Section 3. A brief description of UFM, which is used as the particular region-based image similarity measure used in our learning algorithm, is given in Section 4. In Section 5 we summarize PRRL, whose performance is compared against that of the proposed method. Experimental results are given in Section 6. Finally, we give some concluding remarks in Section 7.

## 2. Generalized Support Vector Machine

Let  $\mathbf{X} \in \mathbb{R}^{m \times n}$  and  $\mathbf{B} \in \mathbb{R}^{n \times l}$ . The kernel  $k(\mathbf{X}, \mathbf{B})$  implements an arbitrary function mapping  $\mathbb{R}^{m \times n} \times \mathbb{R}^{n \times l}$  into  $\mathbb{R}^{m \times l}$ . In particular, given two column vectors  $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$ ,  $k(\mathbf{x}^T, \mathbf{X}^T)$  is a row vector in  $\mathbb{R}^m$ ,  $k(\mathbf{x}^T, \mathbf{b}) \in \mathbb{R}$ , and  $k(\mathbf{X}, \mathbf{X}^T)$  is an  $m \times m$  matrix [16].

Given training data  $\{(\mathbf{x}_i, y_i)\}_1^m$ , where  $\mathbf{x}_i \in \mathbb{R}^n$  and  $y_i \in \{1, 0\}$ , represent it by matrix  $\mathbf{X} \in \mathbb{R}^{m \times n}$  and diagonal matrix of plus or minus ones  $\mathbf{Y} \in \mathbb{R}^{m \times m}$ . Suppose we have a separating hyperplane induced by  $k(\mathbf{X}, \mathbf{X}^T)$  defined as

$$k(\mathbf{x}^T, \mathbf{X}^T)\mathbf{Y} \cdot \mathbf{u} = b \quad (2)$$

where  $\mathbf{u} \in \mathbb{R}^m$  and  $b \in \mathbb{R}$ . In the particular case that  $k$  is an inner product kernel under Mercer's condition, the separating surface becomes

$$\phi(\mathbf{x})^T \phi(\mathbf{X})^T \mathbf{Y} \cdot \mathbf{u} = b$$

where  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^z$ , with  $z \geq n$ . The parameters  $\mathbf{u}$  and  $b$  in (2) can be obtained by solving the following optimization problem

$$\arg \min_{\mathbf{u}, b, \xi} C\mathbf{e} \cdot \xi + \theta(\mathbf{u}) \quad (3)$$

$$\begin{aligned} \text{s.t. } \mathbf{Y}(k(\mathbf{X}, \mathbf{X}^T))\mathbf{Y}\mathbf{u} - \mathbf{e}b + \xi &\geq \mathbf{e} \\ \xi &\geq 0. \end{aligned}$$

where  $\mathbf{e} \in \mathbb{R}^m$  is a column vector of ones,  $\theta$  is some convex function,  $C$  is a positive parameter that weights the separation error  $\mathbf{e} \cdot \xi$  versus suppression of the separating surface parameter  $\mathbf{u}$ . Suppression of  $\mathbf{u}$  can be interpreted as minimizing the number of constraints of (3) with positive multipliers (i.e., number of support vectors). In the particular case that  $\theta$  is a quadratic function induced by a positive definite kernel, we have the standard interpretation of a maximal margin hyperplane [16]. A solution to (3) with corresponding decision function is referred to as a GSVM in [16].

In the particular case that  $\theta$  in (3) is a convex quadratic function (i.e.,  $\theta(\mathbf{u}) = \frac{1}{2}\mathbf{u} \cdot \mathbf{H}\mathbf{u}$ , where  $\mathbf{H} \in \mathbb{R}^{m \times m}$  is a symmetric positive definite matrix), the Wolfe dual [15, 23] of (3) is

$$\min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \cdot \mathbf{Y}k(\mathbf{X}, \mathbf{X}^T)\mathbf{Y}\mathbf{H}^{-1}\mathbf{Y}k(\mathbf{X}, \mathbf{X}^T)^T\mathbf{Y}\alpha - \mathbf{e} \cdot \alpha$$

$$\begin{aligned} \text{s.t. } \mathbf{e} \cdot \mathbf{Y}\alpha &= 0 \\ 0 &\leq \alpha \leq C\mathbf{e}. \end{aligned}$$

and  $\mathbf{u} = \mathbf{H}^{-1}\mathbf{Y}k(\mathbf{X}, \mathbf{X}^T)^T\mathbf{Y}\alpha$ . If  $k(\mathbf{X}, \mathbf{X}^T)$  is assumed to be symmetric positive definite and  $\mathbf{H} = \mathbf{Y}k(\mathbf{X}, \mathbf{X}^T)\mathbf{Y}$ , then we obtain the dual problem for a standard SVM with  $\mathbf{u} = \alpha$  [16]. The basic idea in [16] is to choose other values for the matrix  $\mathbf{H}$  that will also suppress  $\mathbf{u}$ . In the simplest case, choosing  $\mathbf{H} = \mathbf{I}$  (i.e., the identity matrix) with  $\mathbf{u} = \mathbf{Y}k(\mathbf{X}, \mathbf{X}^T)^T\alpha$  results in the following dual problem

$$\min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \cdot \mathbf{Y}\mathbf{A}\mathbf{Y}\alpha - \mathbf{e} \cdot \alpha \quad (4)$$

$$\begin{aligned} \text{s.t. } \mathbf{e} \cdot \mathbf{Y}\alpha &= 0 \\ 0 &\leq \alpha \leq C\mathbf{e}. \end{aligned}$$

where  $\mathbf{A} = k(\mathbf{X}, \mathbf{X}^T)k(\mathbf{X}, \mathbf{X}^T)^T$  is a positive semidefinite matrix. Thus, this is an always solvable convex quadratic problem for any kernel  $k$  [16]. For more details, see [16].

### 3. Proposed Method

The standard SVM approach cannot be applied in the case of region-based CBIR methods that define a similarity measure between two variable length image representations. This is because the requirements that SVM places on the kernel are violated. Fortunately, a GSVM allows us to use arbitrary kernels. We present a GSVM-based learning approach that allows us to use arbitrary image similarity measures.

Let an image be represented by  $\mathbf{x} = \{\mathbf{R}_i\}_1^n$ , where  $\mathbf{R}_i$  represents the features extracted from a region in the image. Let  $S(\mathbf{x}_i, \mathbf{x}_j)$  be an arbitrary similarity measure between two images. During the RF process for a particular query image, the user marks each retrieved image  $\mathbf{x}_i$  as relevant ( $y_i = 1$ ) or non-relevant ( $y_i = 0$ ). We use the set of cumulative retrievals  $\mathcal{R} = \{(\mathbf{x}_i, y_i)\}_1^m$  as training data in (4). Set  $k(\mathbf{x}_i, \mathbf{x}_j) = S(\mathbf{x}_i, \mathbf{x}_j)$  and let  $s_{\mathbf{x}_i} = [S(\mathbf{x}_i, \mathbf{x}_1) \ S(\mathbf{x}_i, \mathbf{x}_2) \ \dots \ S(\mathbf{x}_i, \mathbf{x}_m)]^T$  (i.e., vector of similarities of  $\mathbf{x}_i$  to all training images). Then, the  $(i, j)^{th}$  entry of matrix  $\mathbf{A}$  in (4) is  $\langle s_{\mathbf{x}_i} \cdot s_{\mathbf{x}_j} \rangle$  (i.e., the dot product of  $s_{\mathbf{x}_i}$  and  $s_{\mathbf{x}_j}$ ). Let  $K_{\mathcal{R}}(\mathbf{x}_i, \mathbf{x}_j) = \langle s_{\mathbf{x}_i} \cdot s_{\mathbf{x}_j} \rangle$ . The equivalent (non-matrix) notation for (4) is then as follows

$$\begin{aligned} \min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K_{\mathcal{R}}(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^m \alpha_i \quad (5) \\ \text{s.t. } \sum_{i=1}^m \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C \end{aligned}$$

Note that (5) is just a standard SVM with an identity kernel. Thus, by representing each image as a vector of its similarity (as given by the arbitrary region-based similarity measure  $S$ ) to all training images, we can use an ordinary SVM.

We regard the problem as a two-class classification problem and use a GSVM to separate relevant images from non-relevant images in the non-linearly transformed feature space. The proposed learning algorithm is summarized in Figure 2.

### 4. Unified Feature Matching

In UFM [4], an image is characterized by a set of segmented regions each of which is represented by a set of

1. Retrieve the  $M$  most similar images to query image by using similarity measure  $S$ .
2. While more RF iterations Do
  - (a) User marks the  $M$  images as relevant or non-relevant
  - (b)  $\mathcal{R} \leftarrow \mathcal{R} \cup \{\text{marked } M \text{ images}\}$ .
  - (c) Compute SVM by solving (5) on training data  $\mathcal{R}$ .
  - (d) Compute the score  $f(\mathbf{x})$  of each database image  $\mathbf{x}$  using SVM decision function  $f(\mathbf{x}) = \sum_{i=1}^{|\mathcal{R}|} \alpha_i y_i K_{\mathcal{R}}(\mathbf{x}, \mathbf{x}_i) + b$ .
  - (e) Retrieve the  $M$  highest-score database images.

**Figure 2.** GSVM-based RF learning algorithm

fuzzy features denoting color, texture, and shape characteristics. The similarity between two images is then defined as the overall similarity between two sets of fuzzy features. To segment an image, it is first partitioned into blocks of 4x4 pixels. Then, a feature vector  $\mathbf{f}_i \in \mathbb{R}^6$  representing color and texture properties is extracted for each block. The first three features are the average color components and the other three represent energy in high frequency bands of the wavelet transforms [6, 17]. The  $C$ -means algorithm is then used to cluster the feature vectors into  $C$  regions  $\{\mathbf{R}_i\}_1^C$ . The number of regions  $C$  is adaptively chosen according to a stopping criteria. A feature vector  $\mathbf{h}_j \in \mathbb{R}^3$  is then extracted for each region  $\mathbf{R}_j$  to describe its shape characteristics. The shape features are normalized inertia [8] of order 1 to 3.

The color and texture properties of each region  $\mathbf{R}_j$  are represented by a fuzzy feature with a Cauchy membership function  $\mu_{\mathbf{R}_j, f} : \mathbb{R}^6 \rightarrow [0, 1]$  defined as

$$\mu_{\mathbf{R}_j, f}(\mathbf{f}) = \frac{1}{1 + \left(\frac{\|\mathbf{f} - \hat{\mathbf{f}}_j\|}{d_f}\right)^\alpha}$$

where  $\hat{\mathbf{f}}_j$  is the average of all feature vectors in  $\mathbf{R}_j$  and

$$d_f = \frac{2}{C(C-1)} \sum_{i=1}^{C-1} \sum_{k=i+1}^C \|\hat{\mathbf{f}}_i - \hat{\mathbf{f}}_k\|$$

is the average distance between cluster centers. The shape characteristics of each region  $\mathbf{R}_j$  are also represented by a fuzzy feature with a Cauchy membership function  $\mu_{\mathbf{R}_j, h} : \mathbb{R}^3 \rightarrow [0, 1]$  defined as

$$\mu_{\mathbf{R}_j, h}(\mathbf{h}) = \frac{1}{1 + \left(\frac{\|\mathbf{h} - \mathbf{h}_j\|}{d_h}\right)^\alpha}$$

where

$$d_h = \frac{2}{C(C-1)} \sum_{i=1}^{C-1} \sum_{k=i+1}^C \|\mathbf{h}_i - \mathbf{h}_k\|$$

is the average distance between shape features.

Let  $\{(\mu_{\mathbf{R}_i, f}, \mu_{\mathbf{R}_i, h})\}_1^{C_q}$  and  $\{(\mu_{\mathbf{R}'_i, f}, \mu_{\mathbf{R}'_i, h})\}_1^{C_t}$  be the fuzzy feature representations for a query and target image respectively. The color and texture similarity between the query and the target image is captured by the similarity vector

$$\mathbf{F} = [\mathbf{I}_1^t, \mathbf{I}_2^t, \dots, \mathbf{I}_{C_t}^t, \mathbf{I}_1^q, \mathbf{I}_2^q, \dots, \mathbf{I}_{C_q}^q]^T$$

where

$$\begin{aligned} \mathbf{I}_i^t &= S(\mu_{\mathbf{R}_i, f}, \bigcup_{j=1}^{C_t} \mu_{\mathbf{R}'_j, f}) \\ &= \frac{d_f + d'_f}{d_f + d'_f + \min_{j=1, \dots, C_t} \|\hat{\mathbf{f}}_i - \hat{\mathbf{f}}'_j\|} \\ \mathbf{I}_i^q &= S(\mu_{\mathbf{R}'_i, f}, \bigcup_{j=1}^{C_q} \mu_{\mathbf{R}_j, f}) \\ &= \frac{d_f + d'_f}{d_f + d'_f + \min_{j=1, \dots, C_q} \|\hat{\mathbf{f}}'_i - \hat{\mathbf{f}}_j\|} \end{aligned}$$

and similarly for the shape similarity, captured by similarity vector  $\mathbf{H}$ . The UFM measure for the query and target image is then defined as

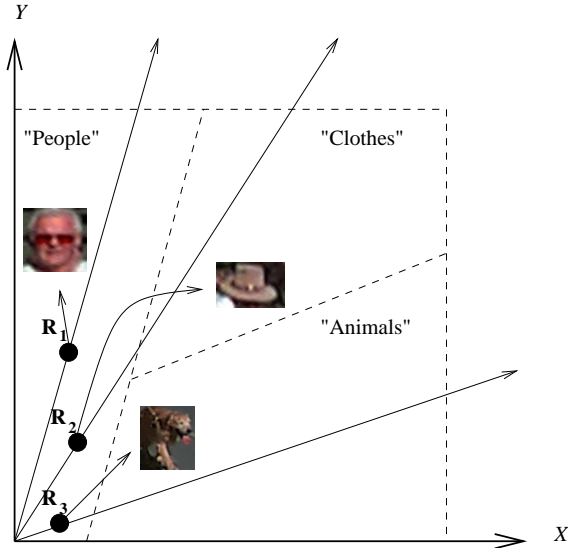
$$m_{(q,t)} = (1 - \rho)[(1 - \lambda)\mathbf{w}_a + \lambda\mathbf{w}_b]^T \mathbf{F} + \rho\mathbf{w}_a^T \mathbf{H}$$

where the normalized weight vectors  $\mathbf{w}_a$  and  $\mathbf{w}_b$  can be set according to some region weighting heuristic,  $0 \leq \lambda \leq 1$  adjusts the importance of  $\mathbf{w}_a$  and  $\mathbf{w}_b$ , and  $0 \leq \rho \leq 1$  determines the significance of  $\mathbf{F}$  (i.e., color and texture similarity) and  $\mathbf{H}$  (i.e., shape similarity). For further details, see [4].

## 5. Probabilistic Region Relevance Learning

A key factor in region-based CBIR systems that consider all the regions to perform an overall image-to-image similarity is the weighting of regions. The weight that is assigned to each region for determining similarity is usually based on prior assumptions such as that larger regions, or regions that are close to the center of the image, should have larger weights. This is often inconsistent with human perception. For instance, a facial region may be the most important when the user is looking for images of people while other larger regions such as the background may be much less relevant.

Based on the observation that regions in an image have unequal importance for computing image similarity (See Figure 3), we presented a probabilistic method inspired by PFRL[18], probabilistic region relevance learning (PRRL) in [9], for automatically capturing region relevance based on user's feedback. PRRL can be used to set region weights in region-based image retrieval frameworks that use an overall image-to-image similarity measure.



Query Image

**Figure 3.** Regions are unequal in their differential relevance for computing similarity. Given that the user is looking for images of people, region  $\mathbf{R}_1$  is the most important, followed by  $\mathbf{R}_2$  and  $\mathbf{R}_3$ . Thus, the neighborhood of the similarity metric should be elongated along the direction of  $\mathbf{R}_1$  and constricted along the direction of  $\mathbf{R}_3$

## 5.1 Region Relevance Measure

Given a query image  $\mathbf{x} = \{\mathbf{R}_i\}_1^n$ , where  $\mathbf{R}_i$  represents the features extracted from a region in the image. Let the class label  $y \in \{1, 0\}$  at  $\mathbf{x}$  be treated as a random variable from a distribution with the probabilities  $\{Pr(1|\mathbf{x}), Pr(0|\mathbf{x})\}$ . Consider the function  $f$  of  $n$  arguments

$$f(\mathbf{x}) \doteq Pr(1|\mathbf{x}) = Pr(y = 1|\mathbf{x}) = E(y|\mathbf{x})$$

In the absence of any argument assignments, the least-squares estimate for  $f(\mathbf{x})$  is simply its expected (average) value

$$E[f] = \int f(\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

where  $p(\mathbf{x})$  is the joint probability density. Now, suppose that we know the value of  $\mathbf{x}$  at a particular region  $\mathbf{R}_i$ . The least-squares estimate becomes

$$E[f|\mathbf{R}_i] = \int f(\mathbf{x})p(\mathbf{x}|\mathbf{R}_i)d\mathbf{x}$$

where  $p(\mathbf{x}|\mathbf{R}_i)$  is the conditional density of the other regions. Because  $f(\mathbf{x}) = 1$  (i.e., the query image is always relevant),  $(f(\mathbf{x}) - 0)$  is the maximum error that can be made when assigning 0 to the probability that  $\mathbf{x}$  is relevant when the probability is in fact 1. On the other hand,  $(f(\mathbf{x}) - E[f|\mathbf{R}_i])$  is the error that is made by predicting  $E[f|\mathbf{R}_i]$  to be the probability that  $\mathbf{x}$  is relevant. Therefore,

$$[(f(\mathbf{x}) - 0) - (f(\mathbf{x}) - E[f|\mathbf{R}_i])] = E[f|\mathbf{R}_i]$$

represents a reduction in error between the two predictions. Therefore, a measure of the relevance of region  $\mathbf{R}_i$  for  $\mathbf{x}$  can be defined as

$$r_i(\mathbf{x}) = E[f|\mathbf{R}_i] \quad (6)$$

The relative relevance can then be used as the weight of region  $\mathbf{R}_i$  in a weighted similarity measure

$$w_i = \frac{e^{Tr_i(\mathbf{x})}}{\sum_{l=1}^n e^{Tr_l(\mathbf{x})}} \quad (7)$$

where  $T$  is a parameter that can be chosen to maximize (minimize) the influence of  $r_i$  on  $w_i$  [9].

## 5.2 Estimating Region Relevance

Similarly to PFRL[18] for estimating feature relevance, retrieved images with relevance feedback are used to estimate region relevance. Let  $\mathcal{R} = \{(\mathbf{x}_j, y_j)\}_1^m$  be the set of cumulative retrievals for  $\mathbf{x}$ . Let  $\mathbf{x}_j = \{\mathbf{R}_j'\}_1^z$ . Let  $0 \leq$

1. Retrieve the  $M$  most similar images to query image by using similarity measure  $S$ .
2. While more RF iterations Do
  - (a) User marks the  $M$  images as relevant or non-relevant.
  - (b)  $\mathcal{R} \leftarrow \mathcal{R} \cup \{\text{marked } K \text{ images}\}$ .
  - (c) Update weights of regions in query image with (8) and (7) using  $\mathcal{R}$ .
  - (d) Retrieve the  $M$  most similar images to query image by using similarity measure  $S$ .

**Figure 4.** Probabilistic region relevance learning

$s(\mathbf{R}_i, \mathbf{R}'_j) \leq 1$  denote the similarity between region  $\mathbf{R}_i$  in  $\mathbf{x}$  and region  $\mathbf{R}'_j$  in  $\mathbf{x}_j$  in a region-based CBIR system. Also, let  $\hat{s}(\mathbf{R}_i, \mathbf{x}_j) = \max_{j \in \{1, 2, \dots, z\}} (s(\mathbf{R}_i, \mathbf{R}'_j))$ . We can use  $\mathcal{R}$  to estimate (6), hence (7). Note that  $E[f|\mathbf{R}_i] = E[y|\mathbf{R}_i]$ . However, since there may be no  $\mathbf{x}_j \in \mathcal{R}$  for which  $\mathbf{R}'_j = \mathbf{R}_i$  (i.e., no  $\mathbf{R}'_j$  such that  $s(\mathbf{R}_i, \mathbf{R}'_j) = 1$ ), a strategy suggested in [7] is followed and we look for data in the vicinity of  $\mathbf{R}_i$  (i.e., we allow  $s(\mathbf{R}_i, \mathbf{R}'_j)$  to be smaller than 1). Thus, (7) is estimated by

$$\hat{E}[y|\mathbf{R}_i] = \frac{\sum_{j=1}^m y_j 1(S(\mathbf{R}_i, \mathbf{x}_j) > \varepsilon)}{\sum_{j=1}^m 1(S(\mathbf{R}_i, \mathbf{x}_j) > \varepsilon)} \quad (8)$$

where  $1(\cdot)$  returns 1 if its argument is true, and 0 otherwise. Thus,  $0 \leq \varepsilon \leq 1$  is an adaptive similarity threshold that changes so that there is sufficient data for the estimation of (6). The value of  $\varepsilon$  is chosen so that  $\sum_{j=1}^m 1(S(\mathbf{R}_i, \mathbf{x}_j) > \varepsilon) = G$ , where  $G \leq m$ . The probabilistic region relevance learning algorithm is summarized in Figure 4.

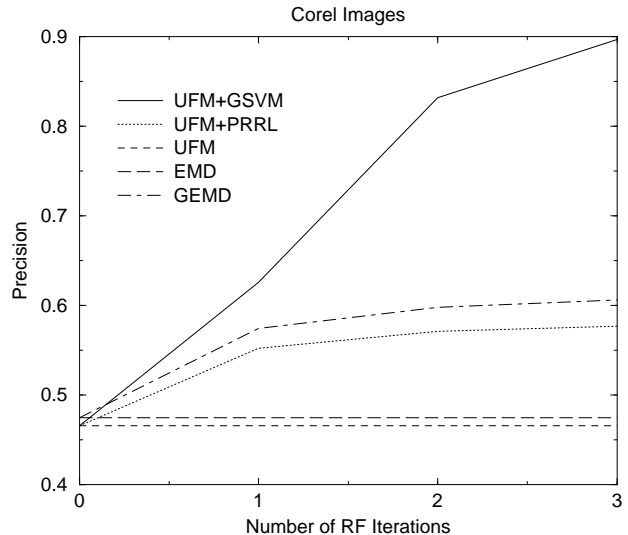
## 6. Experimental Results

We tested the performance of the UFM and EMD similarity measures with not RF learning, UFM with the proposed GSVM-based learning algorithm (UFM+GSVM), UFM with distance reweighting by PRRL (UFM+PRRL), and SVM learning with kernel  $K_{GEMD}$  (GEMD). The retrieval performance is measured by *precision* and *recall*, which are defined as

$$precision = \frac{\text{number of relevant images retrieved}}{\text{number of images retrieved}}$$

$$recall = \frac{\text{number of relevant images retrieved}}{\text{number of relevant images in database}}$$

A subset of 2000 labelled images from the general-purpose COREL image database was used as the data set.



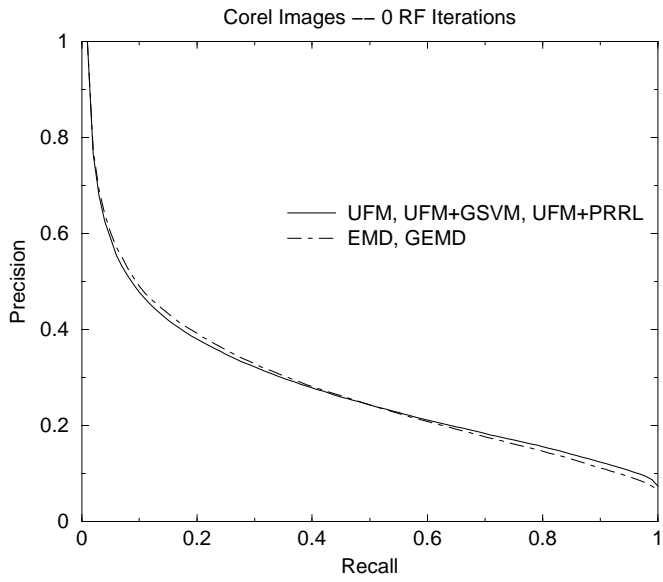
**Figure 5.** Precision at different number of RF iterations. The size of the retrieval set is 20

The region-based feature vectors of those images are obtained with the segmentation algorithm used by UFM and described in Section 4<sup>1</sup>. There are 20 image categories, each containing 100 pictures. Every image is used as a query image. A uniform weighting scheme is used to set the region weights of each query and target images (i.e.,  $\mathbf{w}_a = \mathbf{w}_b = \{\frac{1}{C_q + C_t}\}_1^{C_q + C_t}$ ). For UFM+GSVM, UFM+PRRL, and GEMD, we simulated user's feedback by carrying out 3 RF iterations for each query. Because the images in the data set are labelled according to their category, it is known whether an image in the retrieval set would be labelled as relevant or non-relevant by the user.

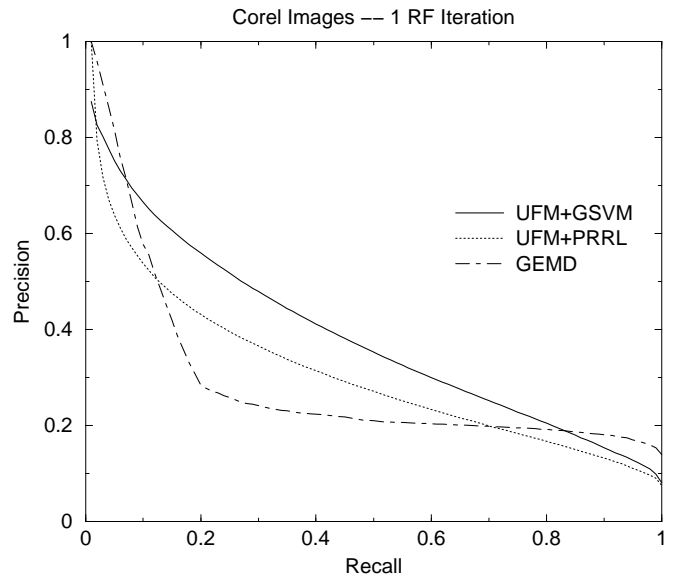
In UFM+GSVM and GEMD, after each RF iteration, the set of labelled cumulative retrieved images is used as training data and the resulting decision function is used to rank database images. We used the LIBSVM [2] package for creating the SVM. In UFM+PRRL, after each RF iteration, PRRL is used to update the region weights of the query image.

The average precision of the 2000 queries with respect to different number of RF iterations is shown in Figure 5. Figures 6 through 9 show the precision-recall curves after each RF iteration. We can observe that, even after only 1 RF iteration the RF learning methods (i.e., UFM+GSVM, UFM+PRRL, and GEMD) result in a very significant performance improvement. Also, UFM+GSVM performs much better than UFM+PRRL and GEMD.

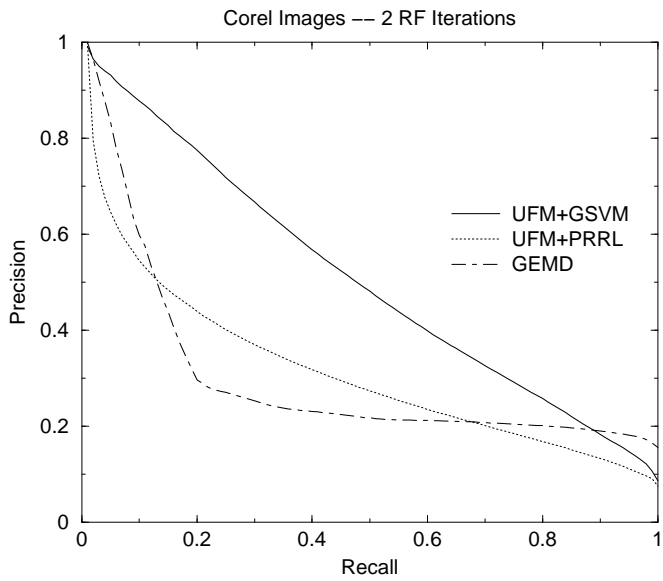
<sup>1</sup>We would like to thank Yixin Chen for providing us with the segmentation results



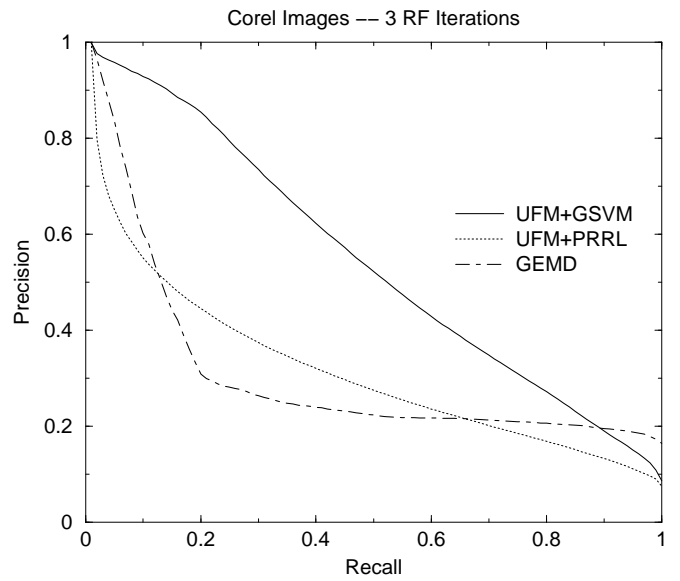
**Figure 6.** Precision-recall curve with no learning



**Figure 7.** Precision-recall curve after 1 RF iteration



**Figure 8.** Precision-recall curve after 2 RF iterations



**Figure 9.** Precision-recall curve after 3 RF iterations

## 7. Conclusions and Future Work

We presented a GSVM-based RF learning algorithm that can be used in region-based CBIR systems that use arbitrary similarity measures. The experimental results on general-purpose images show convincingly the efficacy of the proposed method in improving image retrieval performance. Currently, for each query, the user's RF is used to learn a SVM and the learning process starts from ground up for each new query. However, it is also possible to exploit the long term learning accumulated over the course of many query sessions. This would be very beneficial specially in the initial retrieval set since, instead of ranking images based only on a region-based similarity measure, we could make a more informed initial estimate of the relevance of images to the user's query concept. We plan to investigate the possibility of incorporating long-term learning into the region-based CBIR framework. With GSVM, there is ample opportunity to adapt other SVM-based CBIR approaches to region-based image retrieval. This will also be part of our future research.

## References

- [1] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its applications to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.
- [2] C. Chang and C. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [3] Y. Chen and J. Wang. A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1252–1267, 2002.
- [4] Y. Chen, X. Zhou, and T. Huang. One-class svm for learning in image retrieval. In *Proceedings of IEEE International Conference on Image Processing*, pages 34–37, 2001.
- [5] N. Cristianini and J. Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, 2000.
- [6] I. Daubechies. *Ten Lectures on Wavelets*. Capital City Press, 1992.
- [7] J. Friedman. Flexible metric nearest neighbor classification. Technical report, Department of Statistics, Stanford University, 1994.
- [8] A. Gersho. Asymptotically optimum block quantization. *IEEE Transactions on Information Theory*, IT-25(4):231–262, 1979.
- [9] I. Gondra and D. Heisterkamp. Probabilistic region relevance learning for content-based image retrieval. In *Proceedings of the 2004 International Conference on Imaging Science, Systems, and Technology (to appear)*, 2004.
- [10] D. Heisterkamp, J. Peng, and H. Dai. Adaptive quasiconformal kernel metric for image retrieval. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 388–393, 2001.
- [11] P. Hong, Q. Tian, and T. Huang. Incorporate support vector machines to content-based image retrieval with relevance feedback. In *Proceedings of the IEEE International Conference on Image Processing*, pages 750–753, 2000.
- [12] F. Jing, M. Li, L. Zhang, H. Zhang, and B. Zhang. Learning in region-based image retrieval. In *Proceedings of 2nd International Conference on Image and Video Retrieval*, pages 206–215, 2003.
- [13] J. Li, J. Wang, and G. Wiederhold. Irm: Integrated region matching for image retrieval. In *Proceedings of the 8th ACM International Conference on Multimedia*, pages 147–156, 2000.
- [14] W. Ma and B. Manjunath. Netra: A toolbox for navigating large image databases. In *Proceedings of IEEE International Conference on Image Processing*, pages 568–571, 1997.
- [15] O. Mangasarian. *Nonlinear Programming*. McGraw-Hill, New York, 1969.
- [16] O. Mangasarian. Generalized support vector machines. In A. Smola et al. (eds.), *Advances in Large Margin Classifiers*, pages 135–146. MIT Press, 2000.
- [17] Y. Meyer. *Wavelets Algorithms and Applications*. SIAM, Philadelphia, 1993.
- [18] J. Peng, B. Bhanu, and S. Qing. Probabilistic feature relevance learning for content-based image retrieval. *Computer Vision and Image Understanding*, 75(1/2):150–164, 1999.
- [19] Y. Rubner, C. Tomasi, and L. Guibas. A metric for distributions with applications to image databases. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 59–66, 1998.
- [20] B. Scholkopf. *Support vector learning*. PhD thesis, Technischen Universitat, Berlin, Germany, 1997.
- [21] V. Vapnik. *The nature of statistical learning theory*. Springer, New York, 1995.
- [22] J. Wang, G. Li, and G. Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:947–963, 2001.
- [23] P. Wolfe. A duality theorem for nonlinear programming. *Quarterly of Applied Mathematics*, 19:239–244, 1961.
- [24] L. Zhang, F. Lin, and B. Zhang. Support vector machine learning for image retrieval. In *Proceedings of IEEE Conference on Image Processing*, pages 721–724, 2001.
- [25] X. Zhou and T. Huang. Small sample learning during multimedia retrieval using biasmap. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 11–17, 2001.