Probabilistic Region Relevance Learning for Content-Based Image Retrieval

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

Douglas R. Heisterkamp
Oklahoma State University
Computer Science Department
219 Mathematical Sciences
Stillwater, OK 74078-1053
doug@cs.okstate.edu

Abstract

Probabilistic feature relevance learning (PFRL) is an effective method for adaptively computing local feature relevance in content-based image retrieval. It computes flexible retrieval metrics for producing neighborhoods that are elongated along less relevant feature dimensions and constricted along most influential ones. Based on the observation that regions in an image have unequal importance for computing image similarity, we propose a probabilistic method inspired by PFRL, probabilistic region relevance learning (PRRL), for automatically estimating 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. Experimental results on general-purpose images show the effectiveness of PRRL in learning the relative importance of regions in an image.

Keywords: *Region-based Image Retrieval, Region Importance, Relevance Feedback*

1. Introduction

In traditional approaches, a content-based image retrieval (CBIR) system extracts some global features (such as color, texture, and shape) from an image. The features are then the components of a feature vector which makes the image correspond to a point in a feature space. In order to determine closeness between two images, a similarity measure is used to calculate the distance between their corresponding feature vectors. Then, the closest images in feature space to a query image are returned to the user as the query results. Because of the semantic gap between high-level concepts and low-level features, the performance of CBIR is not satisfactory. In order to overcome this problem, two major approaches have been suggested: the use of a learning technique, such as relevance feedback to learn the user's high-level concept and region-based image representations that are closer to a user's perception of an image's content.

Relevance feedback (RF) works by gathering semantic information from user interaction. In order to learn a user's query concept, the user labels each image returned in the previous query round as relevant (1) or non-relevant (-1). Based on the feedback, the retrieval scheme is adjusted and the next set of images is presented to the user for labelling. Two main RF strategies have been proposed in CBIR: query shifting [13], and distance reweighting [6, 12, 11]. Query shifting involves moving the query towards the region of the feature space containing relevant images and away from the region containing non-relevant images. Distance reweighting assumes that the relevant images are located along some direction of the feature space. Thus, the task is to determine the features that help the most in retrieving relevant images and increase their importance in determining similarity. In [11], a probabilistic feature relevance learning (PFRL) method that automatically captures the feature relevance based on RF is presented. 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). Retrieved images with RF are used to compute local feature relevance.

In contrast to traditional methods, which compute global features, region-based approaches [1, 9, 14] extract features from segmented regions of an image. Then, images are retrieved according to the similarity between regions. The main objective of using regions is to do a more meaningful retrieval that is closer to a user's perceptions of an image's content. Instead of looking at the image as a whole, we look at the objects in the image and their relationships. The similarity measure that most of these systems [1, 9] use to compare two images is based on individual region-to-region similarity. Both Blobworld [1] and Netra [9] require the

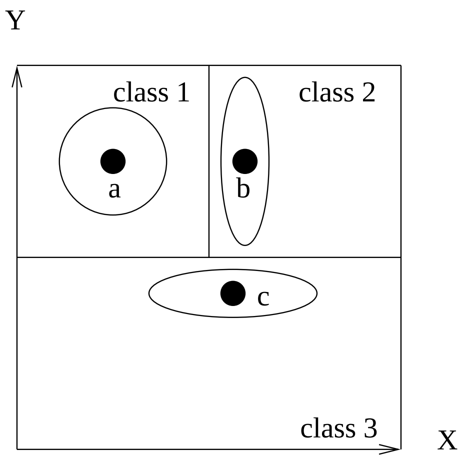


Figure 1. Features are unequal in their differential relevance for computing similarity. The neighborhoods of queries b and c should be elongated along the less relevant Y and X axis respectively. For query a, features X and Y have equal discriminating strength

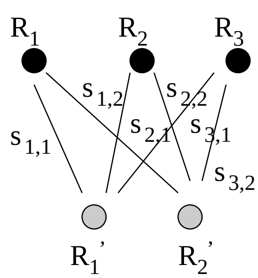


Figure 2. Integrated region matching (IRM)

user to select the region(s) of interest from the segmented image. This information is then used for determining similarity with database images. A major problem with these systems is that the segmented regions they produce usually do not correspond to actual objects in the image. For instance, an object may be partitioned into several regions, with none of them being representative of the object.

In order to overcome the problems of innaccurate image segmentation, some approaches have been proposed that consider all the regions in an image for determining similarity [8, 2, 14]. In [8], IRM (Integrated Region Matching) 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. Basically, the "most similar highest priority" principle is used and the smaller the distance between two regions $\mathbf{R_i}$ and $\mathbf{R_j}$ is, the larger their significance credit (weight) $s_{i,j}$ is (See Figure 2). 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, UFM (Unified Feature Matching) [2] was proposed as an alternative to IRM. An image is represented by a set of segmented regions each of which is represented by a fuzzy feature 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.

A key factor in these types of 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 propose a probabilistic method inspired by PFRL[11], probabilistic region relevance learning (PRRL), 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.

1.1 Related Work

Although RF learning has been successfully applied to CBIR systems that use global image representations, not much research has been conducted on RF learning methods for region-based CBIR. Based on the assumption that important regions should appear more often in relevant images than unimportant regions, Jing et al. [7] proposed a $RF * IIF$ (Region Frequency * Inverse Image Frequency) weighting scheme (RFIIF). Let $\mathbf{x} = \{\mathbf{R_i}\}_1^n$ be the variable length representation of a query image, where $\mathbf{R_i}$ represents the features extracted from a region in the image. Let $\mathcal{D} = \{\mathbf{x_i}\}_1^T$ be the set of all images in the database, and $\mathcal{R}^+ = \{\mathbf{x_i}\}_1^k$ be the set of cumulative relevant retrieved images for query image $\mathbf{x}$. For each region $\mathbf{R_i} \in \mathbf{x}$, the region frequency ($RF$) is defined as

$$RF(\mathbf{R_i}) = \sum_{\mathbf{x_j} \in \mathcal{R}^+} s(\mathbf{R_i}, \mathbf{x_j})$$

where $s(\mathbf{R_i}, \mathbf{x_j}) = 1$ if at least one region of $\mathbf{x_j}$ is similar to $\mathbf{R_i}$ and 0 otherwise. Two regions are deemed similar if their $L1$ distance is smaller than a predefined treshold. The

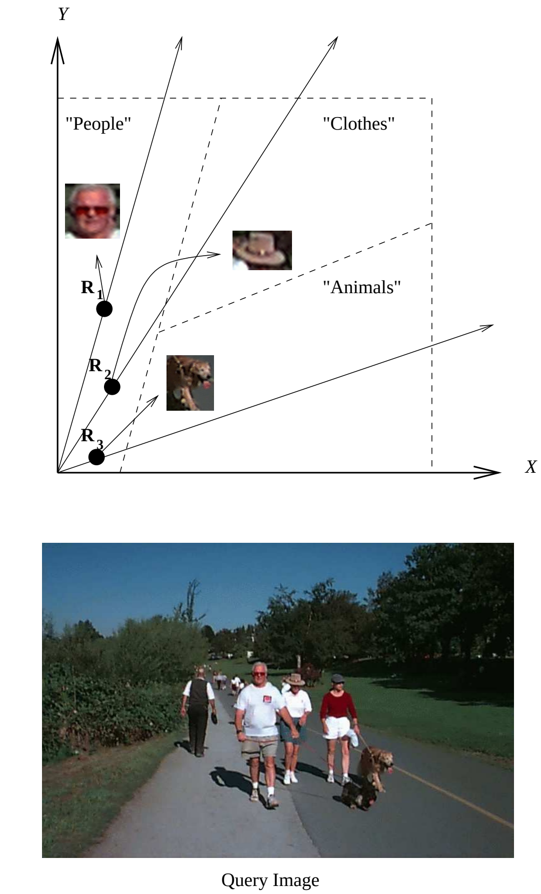


Figure 3. Regions are unequal in their differential relevance for computing similarity. Given that the user is looking for images of people, region $\mathbf{R_1}$ is the most important, followed by $\mathbf{R_2}$ and $\mathbf{R_3}$. Thus, the neighborhood of the similarity metric should be elongated along the direction of $\mathbf{R_1}$ and constricted along the direction of $\mathbf{R_3}$

inverse frequency (IIF) is defined as

$$IIF(\mathbf{R_i}) = log\left(\frac{T}{\sum_{\mathbf{x_j}\ in\mathcal{D}} s(\mathbf{R_i}, \mathbf{x_j})}\right)$$

The region importance (weight) RI is then

$$RI(\mathbf{R_i}) = \frac{RF(\mathbf{R_i}) * IIF(\mathbf{R_i})}{\sum_{j=1}^{n}(RF(\mathbf{R_j}) * IIF(\mathbf{R_j}))}$$

1.2 Paper Outline

The rest of this paper is organized as follows. In Section 2, we describe the probabilistic approach for measuring the importance of each region in a query image. Section 3 describes how user's feedback on the retrieval results is used for estimating the measure of region relevance. A brief description of UFM [2], which is used as the particular region-based image retrieval measure with which PRRL is tested, is given in Section 4. In Section 5, we compare the retrieval performance of UFM against that of UFM with PRRL and UFM with RFIIF for setting region weights. Finally, we give some concluding remarks in Section 6.

2. Region Relevance Measure

Inspired by PFRL[11], we learn the differential region relevance by estimating the strength of each region in predicting the class of a given query. 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 $y \in \{1, -1\}$ 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

$$r_i(\mathbf{x}) = E[f|\mathbf{R_i}] \quad (1)$$

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

$$w_i = \frac{e^{Tr_i(\mathbf{x})}}{\sum_{l=1}^{n} e^{Tr_l(\mathbf{x})}} \quad (2)$$

where T is a parameter that can be chosen to maximize (minimize) the influence of r_i on w_i [11].

1. Use a segmentation method to extract regions and represent current query by $\mathbf{x} = \{\mathbf{R_i}\}_1^n$; initialize region weight vector $\mathbf{w}$ to $\{\frac{1}{n}\}_1^n$.
2. Compute the K most similar images to $\mathbf{x}$ with an overall image-to-image similarity measure using $\mathbf{w}$ for the weighting of regions in $\mathbf{x}$.
3. User marks the K images as relevant or non-relevant.
4. While more RF iterations Do
   (a) $\mathcal{R} \leftarrow \mathcal{R} \bigcup \{marked\ K\ images\}$.
   (b) Update $\mathbf{w}$ from Eqs. (3) and (2) using $\mathcal{R}$.
   (c) Compute the K most similar images to $\mathbf{x}$ with an overall image-to-image similarity measure using $\mathbf{w}$ for the weighting of regions in $\mathbf{x}$.
   (d) User marks the K images as relevant or non-relevant.

Figure 4. The probabilistic region relevance learning (PRRL) algorithm

3. Estimation of Region Relevance

Similarly to PFRL[11] for estimating feature relevance, we use the retrieved images with relevance feedback 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 $0 \leq s(\mathbf{R_i}, \mathbf{R'_j}) \leq 1$ 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 $\hat{s}(\mathbf{R_i}, \mathbf{x_j}) = \max_{j \in \{1,2,\cdots,z\}}(s(\mathbf{R_i}, \mathbf{R'_j}))$. We can use $\mathcal{R}$ to estimate (1), hence (2). 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 $s(\mathbf{R_i}, \mathbf{R'_j}) = 1$), we follow an strategy suggested in [4] and 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, (2) is estimated by

$$\hat{E}[y|\mathbf{R_i}] = \frac{\sum_{j=1}^m y_j 1(S(\mathbf{R_i}, \mathbf{x_j}) > \varepsilon)}{\sum_{j=1}^m 1(S(\mathbf{R_i}, \mathbf{x_j}) > \varepsilon)} \quad (3)$$

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 (1). The value of ε is chosen so that $\sum_{j=1}^m 1(S(\mathbf{R_i}, \mathbf{x_j}) > \varepsilon) = G$, where $G \leq l$. The probabilistic region relevance learning algorithm is summarized in Figure 4.

4. Unified Feature Matching (UFM)

Chen and Wang [2] proposed unified feature matching (UFM) as an improved alternative to IRM. In UFM, an image is characterized by a fuzzy feature denoting color, texture, and shape characteristics. The similarity between two images is then defined as the overall similarity between two sets of fuzzy features. Because fuzzy features can characterize the gradual transition between regions in an image, segmentation-related inaccuracies are implicitly considered.

The image segmentation algorithm that is used first partitions an image into blocks of 4x4 pixels. Then, a feature vector $\mathbf{f_i} \in \Re^6$ 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 [3, 10]. 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 $\mathbf{h_j} \in \Re^3$ is then extracted for each region $\mathbf{R_j}$ to describe its shape characteristics. The shape features are normalized inertia of order 1 to 3 [5].

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

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

where $\hat{\mathbf{f}}_\mathbf{j}$ is the average of all feature vectors in $\mathbf{R_j}$ and

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

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 $\mu_{\mathbf{R_j},h} : \Re^3 \to [0,1]$ defined as

$$\mu_{\mathbf{R_j},h}(\mathbf{h}) = \frac{1}{1 + \left(\frac{||\mathbf{h} - \hat{\mathbf{h}}_\mathbf{j}||}{d_h}\right)^\alpha}$$

where

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

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

$$\mathbf{F} = [\mathbf{l}_1^t, \mathbf{l}_2^t, \cdots, \mathbf{l}_{Ct}^t, \mathbf{l}_1^q, \mathbf{l}_2^q, \cdots, \mathbf{l}_{Cq}^q]^T$$

where

$$\mathbf{l}_i^t = S(\mu_{\mathbf{R_i},f}, \bigcup_{j=1}^{Ct} \mu_{\mathbf{R'_j},f}) = \frac{d_f + d'_f}{d_f + d'_f + \min_{j=1,\cdots,Ct} ||\hat{\mathbf{f}}_\mathbf{i} - \hat{\mathbf{f}}'_\mathbf{j}||}$$

$$\mathbf{l}_i^q = S(\mu_{\mathbf{R'_i},f}, \bigcup_{j=1}^{Cq} \mu_{\mathbf{R_j},f}) = \frac{d_f + d'_f}{d_f + d'_f + \min_{j=1,\cdots,Cq} ||\hat{\mathbf{f}}'_\mathbf{i} - \hat{\mathbf{f}}_\mathbf{j}||}$$

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 [2].

5. Experimental Results

A subset of 2000 labelled images from the general-purpose COREL image database was used as the data set. There are 20 image categories, each containing 100 pictures. The region-based feature vectors of those images are obtained with the segmentation algorithm described in Section 4 [1].

We tested the performance of UFM, UFM with PRRL (UFM+PRRL), and UFM with RFIIF (UFM+RFIIF). The retrieval performance is measured by *precision* and *recall*, defined as

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

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

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. For UFM+PRRL, and UFM+RFIIF, user's feedback was simulated 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.

The average precision of the 2000 queries with respect to different number of RF iterations is shown in Figure 5.

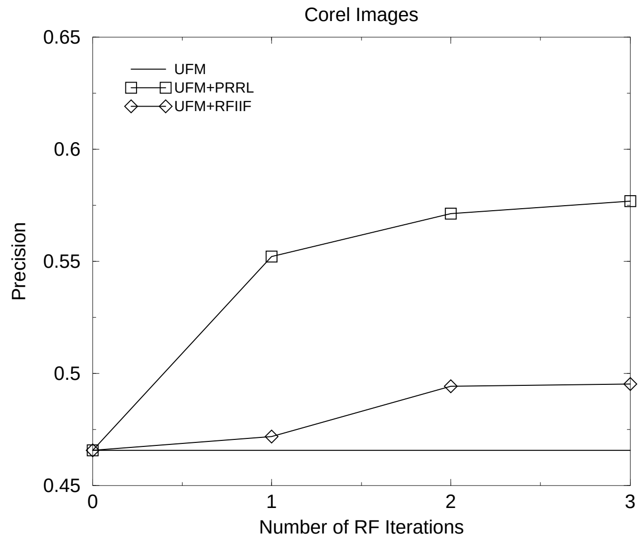


Figure 5. Precision at different number of RF iterations. The size of the retrieval set is 20

Figures 6 through 9 show the precision recall curves after each RF iteration. We can observe that UFM+PRRL has the best performance. It can be seen that, even after only 1 RF iteration, the region weights learned by PRRL result in a very significant performance improvement. Figure 10 shows the retrieval results obtained on a random query image. It is difficult to make objective comparisons with other region-based image retrieval systems such as Netra [9] or Blobworld [1] which require additional information from the user (i.e., important regions and/or features) during the retrieval process.

6. Conclusions and Future Work

Region-based image retrieval frameworks that use an overall image-to-image similarity measure usually set region weights based on some heuristic that is often inconsistent with human perception about the importance of regions in an image. In this paper, we presented a novel probabilistic method for automatically estimating the relative relevance of the regions in an image. The experimental results on general-purpose images show convincingly that learning region relevance based on user's feedback can significantly improve retrieval performance.

Currently, our method only performs intra-query learning. That is, for each given query, the user's feedback is used to learn the relevance of the regions in the query and the learning process starts from ground up for each new query. However, it is also possible to exploit inter-

[1]We would like to thank Yixin Chen for providing us with this data

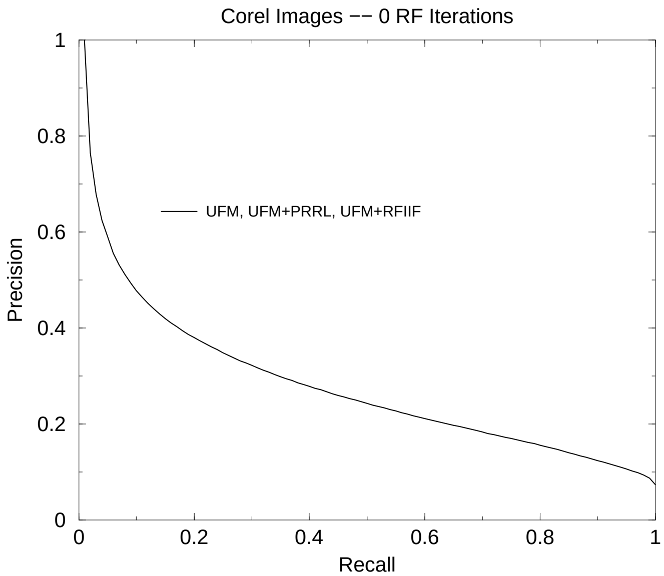


Figure 6. Precision-recall curve with no learning

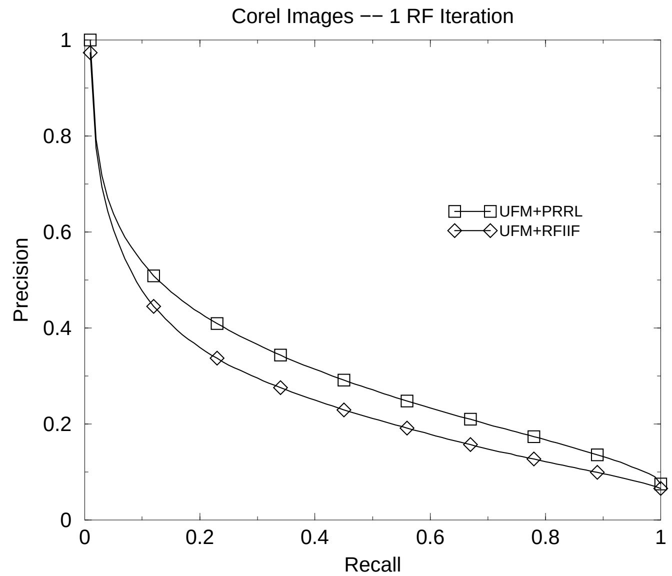


Figure 7. Precision-recall curve after 1 RF iteration

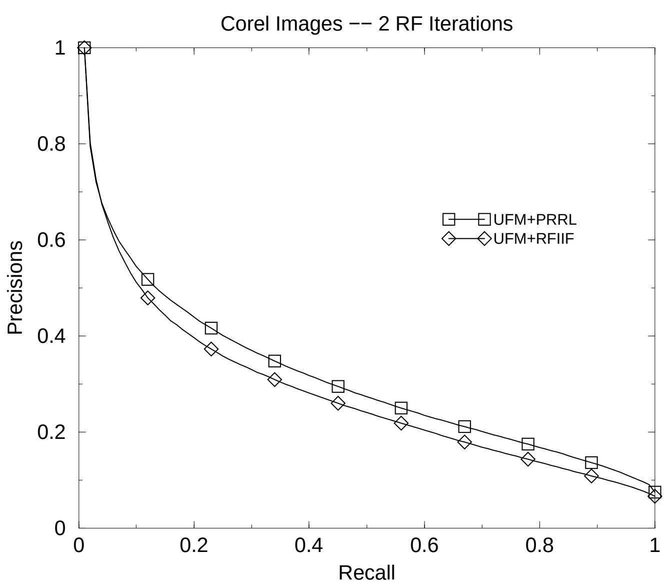


Figure 8. Precision-recall curve after 2 RF iterations

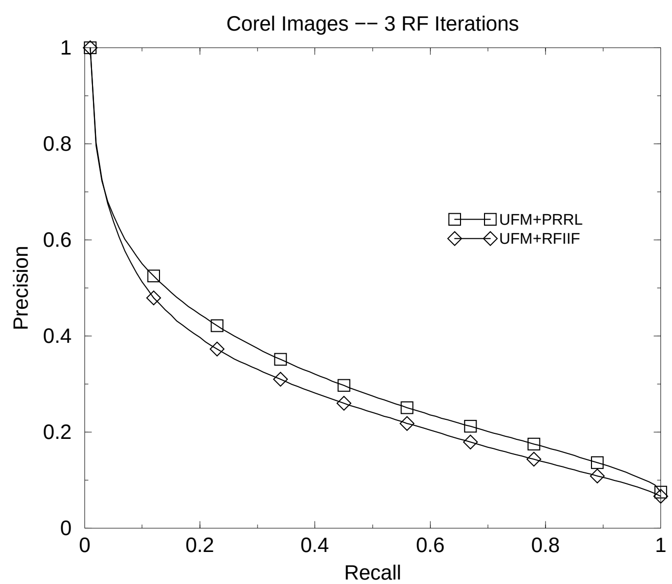
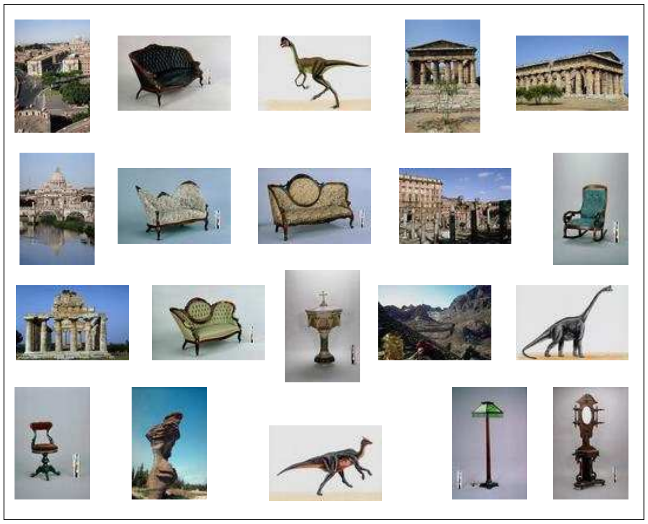
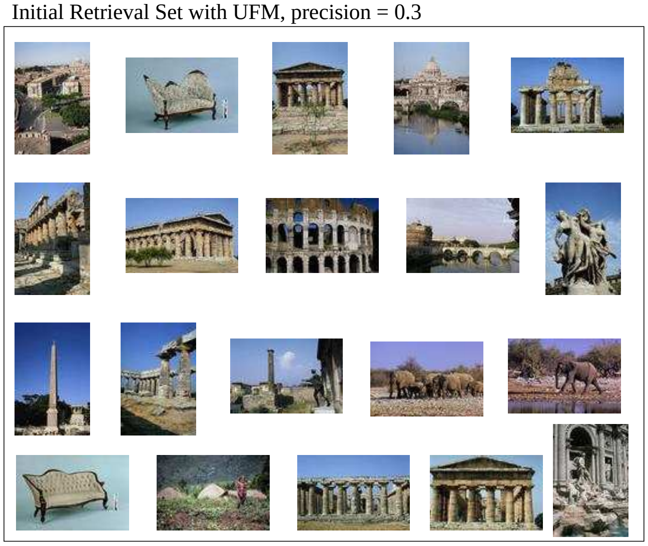


Figure 9. Precision-recall curve of after 3 RF iterations



Initial Retrieval Set with UFM, precision = 0.3



Retrieval Set with UFM+PRRL after 2 RF iterations, precision = 0.75

Figure 10. Retrieval results on a random query image (top leftmost). The images are sorted based on their similarity to the query image. The ranks descend from left to right and from top to bottom.

query learning (i.e., the long-term knowledge accumulated over the course of many query sessions) to enhance the retrieval performance of future queries. Thus, for a new query, instead of starting the learning process from ground up, we could exploit the previously learned region importances of similar queries. This would be very beneficial specially in the initial retrieval set since, instead of using uniform weighting or some other weighting heuristic, we could make a more informed initial estimate of the relevance of regions in the new query. We plan to investigate the possibility of incorporating inter-query learning into the region-based image retrieval framework as part of our future work.

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] 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*, pages 1252–1267, 2002.

[3] I. Daubechies. *Ten Lectures on Wavelets*. Capital City Press, 1992.

[4] J. Friedman. Flexible metric nearest neighbor classifi cation. Technical report, Department of Statistics, Standford University, 1994.

[5] A. Gersho. Asymptotically optimum block quantization. *IEEE Transactions on Information Theory*, IT-25(4):231–262, July 1979.

[6] Y. Ishikawa, R. Subramanys, and C. Faloutsos. Mindreader: Querying databased through multiple examples. In *Proceedings of the 24th VLDB Conference*, pages 433–438, 1998.

[7] 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.

[8] 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.

[9] 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.

[10] Y. Meyer. *Wavelets Algorithms and Applications*. SIAM, Philadelphia, 1993.

[11] 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.

[12] Y. Rui and T. Huang. Relevance feedback: A power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):644–655, 1998.

[13] Y. Rui, T. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. In *Proceedings of the IEEE International Conference on Image Processing*, pages 815–818, 1997.

[14] 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.