1

A Novel Relevance Feedback Technique in Image Retrieval

Yong Rui and Thomas S. Huang
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA
E-mail: {yrui, huang}@ifp.uiuc.edu

Abstract

The relevance feedback based approach to image retrieval has been an active research direction in the past few years. Many parameter estimation techniques have been proposed for relevance feedback. However, most of them are either based on ad-hoc heuristics or only partial solutions. In this paper, we introduce the first technique that not only has a solid theoretical framework but also takes into account the multi-level image content model. This technique formulates a vigorous optimization problem. By using Lagrange multipliers, we have derived the explicit optimal solutions for both the query vectors and the weights associated with the two-level image model. Experimental results on real-world image collections have shown the effectiveness and robustness of our proposed algorithm.

1 Introduction

Techniques in content-based image retrieval (CBIR) systems lag far behind their text counterparts (e.g., Inquery [1]), due to the difficulty for human to precisely express their visual queries. Recently, relevance feedback based CBIR techniques have emerged as a promising research direction. These techniques do not require a user to provide accurate initial queries, but rather estimate the user's ideal query by using positive and negative examples (training samples) fedback by the user. The fundamental goal of these techniques is to estimate the ideal query parameters (both the query vectors and the associated weights) accurately and robustly.

MARS [2] introduced both a query vector moving technique and a re-weighting technique to estimate the ideal query parameters. MindReader [3] formulated a minimization problem on the parameter estimation process. These two techniques are among the best known techniques of relevance feedback. They, however, both still have shortcomings. MARS based its parameter estimation on heuristics. Even though MindReader formulated a more vigorous estimation process than MARS, it failed to analyze the necessary conditions for the technique to work. Moreover, neither technique took into account that images contain multiple levels of content.

To address these limitations, this paper proposes a novel framework, which not only has solid mathematical foundations but also takes into account the multi-level image model. Both MARS and MindReader will be special cases of the proposed technique. In Section 2, we will give a detailed description and derivation of the proposed techniques. In Section 3, based on real-world image collections, we will compare the proposed technique against the existing ones and show its effectiveness and robustness in image retrieval. Concluding remarks are given in Section 4.

2 The Proposed Global Optimization Approach

Images contain rich content at multiple levels. At a higher level, human perceives an image's content in terms of its color, texture, shape of the object in the image, or any combination of these features. At a lower level, each of these features (e.g., color) can be characterized by a feature vector. For example, we can use a 6-element color moment feature vector to characterize the color feature of an image [4]. This two-level image content model not only has its root in human visual perception [4] but also leads to less computation and higher accuracy in retrieval performance, which we will show in Section 3. Based on this two-level image model, we define the overall distance between a training sample and the to-be-estimated query as:

$$d_n = \Phi(g_{ni}, U) \quad (1)$$

$$g_{ni} = \Psi_i\,(\vec{x}_{ni}, \vec{q}_i, W_i) \quad (2)$$

where $i = 1, \cdots, I$ is the index of the i^{th} image feature and I the total number of features associated with an image; $n = 1, \cdots, N$ is the index of the n^{th} training

sample and N the total number of the training samples; $\vec{x}_{ni}$ is the n^{th} training sample of the i^{th} feature and $\vec{q}_i$ is the query vector associated with the i^{th} feature; U and W_i are the weights associated with the two levels and $\Phi()$ and $\Psi_i()$ are the distance functions at the two levels. Different specifications of $\Phi()$ and $\Psi_i()$ determine different relevance feedback techniques. For the re-weighting technique used in MARS, $\Psi_i()$ is an Euclidean distance function with W_i being a diagonal matrix. For the MindReader technique, $\Psi_i()$ is a generalized Euclidean distance function and W_i is a full matrix. Note that neither of the two techniques took into account of $\Phi()$ and U (the higher level).

We have shown [4] that if both $\Phi()$ and $\Psi_i()$ take quadratic form, the optimal solutions for $\vec{q}_i$, W_i and U can only be obtained *iteratively* (no explicit solutions). This is not desirable for image retrieval, since fast response time is always one of the most important requirements. We next develop a formulation that guarantees explicit optimal solutions while making $\Phi()$ and $\Psi_i()$ as general as possible.

If we let $\Phi()$ be a linear function (in g_{ni}) and $\Psi_i()$ be a quadratic function (in $\vec{x}_{ni}$ and $\vec{q}_i$), we can then formulate the following optimization problem [4]:

$$
\begin{aligned}
min\ J &= \vec{\pi}^{\,T} \times \vec{d} && (3)\\
d_n &= \vec{u}^T \vec{g}_n && (4)\\
\vec{g}_n &= [g_{n1}, \cdots, g_{ni}, \cdots, g_{nI}] && (5)\\
g_{ni} &= (\vec{x}_{ni} - \vec{q}_i)^{\,T} W_i (\vec{x}_{ni} - \vec{q}_i) && (6)\\
s.t. &\quad \sum_{i=1}^{I} \frac{1}{u_i} = 1 && (7)\\
&\quad det(W_i) = 1 && (8)\\
n &= 1, \cdots, N && (9)\\
i &= 1, \cdots, I && (10)
\end{aligned}
$$

where $\vec{\pi} = [\pi_1, \cdots, \pi_N]$ is the degree-of-relevance vector of the N training samples given by the user; U takes the form of a vector ($\vec{u}$) and W_i takes the form of a $(K_i \times K_i)$ matrix, where K_i is the length of the i^{th} feature vector. It is easy to see that if there is no constraints for U and W_i, this optimization problem will reduce to a trivial solution. We therefore enforce Equations (7) and (8) as the constraints.

This problem formulation is general enough to include both MARS and MindReader. If we disregard the higher level of the image model (d_n) and only concentrate on the lower level (g_{ni}), a diagonal matrix of W_i will reduce this formulation to the MARS algorithm and a full matrix of W_i will reduce this formulation to the MindReader approach.

We next will use Lagrange multipliers to solve this constrained optimization problem:

$$
L = \vec{\pi}^{\,T} \times \vec{d} - \lambda(\sum_{i=1}^{I} \frac{1}{u_i} - 1) - \sum_{i=1}^{I} \lambda_i (det(W_i) - 1) \quad (11)
$$

A greater detailed derivation can be found in [4]. We next will only highlight the essential part of the solutions due to page limitations.

2.1 Optimal solution for $\vec{q}_i$

$$
\begin{aligned}
\frac{\partial L}{\partial \vec{q}_i} &= \vec{\pi}^{\,T} \times \begin{bmatrix} \frac{\partial d_1}{\partial \vec{q}_i} \\ \cdots \\ \frac{\partial d_n}{\partial \vec{q}_i} \\ \cdots \\ \frac{\partial d_N}{\partial \vec{q}_i} \end{bmatrix}\\
&= \vec{\pi}^{\,T} \times \begin{bmatrix} -2\ \vec{u}^T \times \begin{bmatrix} 0 \\ \cdots \\ (\vec{x}_{1i} - \vec{q}_i)^{\,T} W_i \\ \cdots \\ 0 \end{bmatrix} \\ \cdots \\ -2\ \vec{u}^T \times \begin{bmatrix} 0 \\ \cdots \\ (\vec{x}_{ni} - \vec{q}_i)^{\,T} W_i \\ \cdots \\ 0 \end{bmatrix} \\ \cdots \\ -2\ \vec{u}^T \times \begin{bmatrix} 0 \\ \cdots \\ (\vec{x}_{Ni} - \vec{q}_i)^{\,T} W_i \\ \cdots \\ 0 \end{bmatrix} \end{bmatrix}\\
&= \vec{\pi}^{\,T} \times \begin{bmatrix} -2\ u_i\ (\vec{x}_{1i} - \vec{q}_i)^{\,T} W_i \\ \cdots \\ -2\ u_i\ (\vec{x}_{ni} - \vec{q}_i)^{\,T} W_i \\ \cdots \\ -2\ u_i\ (\vec{x}_{Ni} - \vec{q}_i)^{\,T} W_i \end{bmatrix}
\end{aligned}
$$

By setting the above equation to zero, we can obtain the final solution to $\vec{q}_i$:

$$
\vec{q}_i^{\,T^*} = \frac{\vec{\pi}^{\,T} X_i}{\sum_{n=1}^{N} \pi_n} \quad (12)
$$

where X_i is the training sample matrix for feature i, obtained by stacking the N training vectors ($\vec{x}_{ni}$) into a matrix. It is therefore an $(N \times K_i)$ matrix. Equation (12) closely matches our intuition. That is, $\vec{q}_i^{T^*}$ (the ideal query vector for feature i) is nothing but the weighted average of the training samples for feature i.

2.2 Optimal solution for W_i

$$
\begin{aligned}
\frac{\partial L}{\partial w_{i_{rs}}} &= \vec{\pi}^{\,T} \times \begin{bmatrix} \vec{u}^T \frac{\partial \vec{g}_1}{\partial w_{i_{rs}}} \\ \cdots \\ \vec{u}^T \frac{\partial \vec{g}_n}{\partial w_{i_{rs}}} \\ \cdots \\ \vec{u}^T \frac{\partial \vec{g}_N}{\partial w_{i_{rs}}} \end{bmatrix} - \lambda_i\ (-1)^{r+s} det(W_{i_{rs}})\\
&= \sum_{n=1}^{N} \pi_n\ (x_{nir} - q_{ir})(x_{1is} - q_{is})
\end{aligned}
$$

$$- \quad \lambda_i \ (-1)^{r+s} det(W_{i_{rs}})$$

After setting the above equation to zero, we get [4]:

$$W_i^* = (det(C_i))^{\frac{1}{K_i}} \ C_i^{-1} \tag{13}$$

where the term C_i is the $(K_i \times K_i)$ weighted covariance matrix of X_i. That is, $C_{i_{rs}} = \sum_{n=1}^{N} \pi_n (x_{nir} - q_{ir})(x_{1is} - q_{is}) / \sum_{n=1}^{N} \pi_n, r, s = 1, \cdots, K_i$. The physical meaning of this optimal solution is that the optimal weight matrix is inversely proportional to the covariance matrix of the training samples.

Note that in MARS, W_i is always a diagonal matrix. This limits its ability to model non-linear (quadratic) distance functions. On the other hand, MindReader's W_i is always a full matrix. Even though it can model quadratic functions, itself can not be reliably estimated when the number of training samples (N) is less than the length of the feature vector (K_i). Unlike these two algorithms, the proposed technique dynamically and intelligently switches between a diagonal matrix and a full matrix, depending on the relationship between N and K_i. We have shown that [4] when $N < K_i$, W_i can not be robustly estimated. The proposed algorithm then intelligently forms a diagonal matrix to ensure reliable estimation; and when $N > K_i$, it will form a full matrix to effectively model non-linear distance functions.

2.3 Optimal Solution for $\vec{u}$

To obtain u_i^*, set the partial derivative to zero. We then have

$$\frac{\partial L}{\partial u_i} = \sum_{n=1}^{N} \pi_n \ g_{ni} + \lambda \ u_i^{-2} = 0, \ \forall i \tag{14}$$

Multiply both sides by u_i and summarize over i. We have

$$\sum_{i=1}^{I} u_i \ (\sum_{n=1}^{N} \pi_n \ g_{ni}) + \lambda (\sum_{i=1}^{I} \frac{1}{u_i}) = 0 \tag{15}$$

Since $\sum_{i=1}^{I} \frac{1}{u_i} = 1$, the optimal λ is

$$\lambda^* = -\sum_{i=1}^{I} u_i \ f_i \tag{16}$$

where $f_i = \sum_{n=1}^{N} \pi_n \ g_{ni}$. This will lead to the optimal solution for u_i:

$$u_i^* = \sum_{j=1}^{I} \sqrt{\frac{f_j}{f_i}} \tag{17}$$

This solution tells us, if the total distance (f_i) of feature i is small (meaning it is close to the ideal query), this feature should receive high weight and vice versa.

The solutions for $\vec{q_i}$ and W_i have been partially explored in MARS and MindReader. The solution for u_i, however, has not been investigated by either techniques. In both MARS and MindReader, they used a flat (one-level) image content model, i.e., stack all the feature vectors into a big universal vector. This is not only computationally expensive, but also less effective in retrieval performance. For computation complexity, take MindReader as an example. It needs $O((\sum_i^I K_i)^3 + 2N(\sum_i^I K_i)^2))$ multiplications or divisions while the proposed algorithm only needs $O(\sum_i^I ((K_i)^3 + 2N(K_i)^2))$ operations[4]. Note that the different location of $\sum_i^I$ in the two formular makes *significantly* different computation counts. Similar comparison holds for the MARS algorithm.

3 Experimental Results

We have shown that the proposed algorithm requires much less computation than the existing ones in the previous section. In this section, we will further compare the retrieval performance between these different techniques.

In the experiments reported in this section, the proposed algorithms are tested on the following two data sets. (1) Corel data set: This data set is obtained from Corel Corporation. It contains more than 70,000 images covering a wide range of more than 500 categories. (2) Vistex data set: This data set consists of 832 texture images. Each original 512×512 image is then cut into 16 128×128 nonoverlap small images. The 16 images from the same big image are considered to be relevant images.

We use three visual features in this experiment: color moments, wavelet based texture, and water-fill edge feature. For color moments, we use the HSV color space because of its decorrelated coordinates and its perceptual uniformity. We extract the first two moments (mean and standard deviation) from the three color channels and therefore have a feature vector of length $3 \times 2 = 6$. For wavelet based texture, the original image is fed into a wavelet filter bank and is decomposed into 10 de-correlated sub-bands. For each sub-band, we extract the standard deviation of the wavelet coefficients and therefore have a feature vector of length 10. For water-fill edge feature, there are eighteen (18) elements that are extracted from the edge map of the original image, including *max fill time*, *max fork count*, etc.[5].

We have constructed a CBIR system based on the optimization algorithm developed in Section 2. Figure 1 is its interface.

On the left, are the query image and return results (the top left image is the query image). There is a degree-of-relevance slider associated with each of the images. A user uses these sliders to give his or her relevance feedback to the system. On the right, there are progress controls dynamically displaying the weights (W_i and $\vec{u}$) for a particular query.

To compare the three relevance feedback techniques, we have performed both subjective tests and object

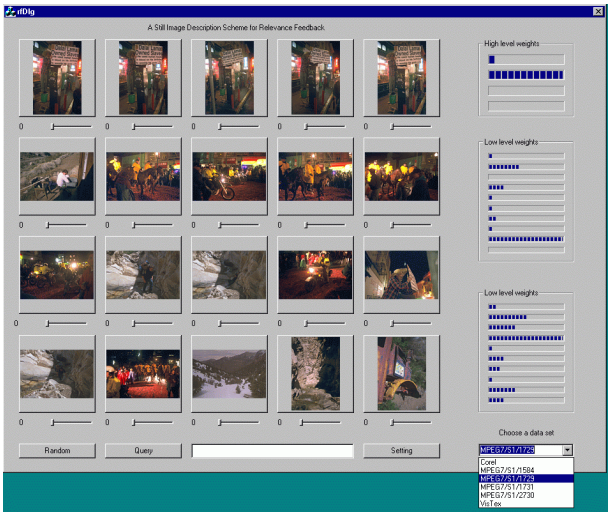

Figure 1: The interface of the demo system

Table 1: Retrieval performance.

	0 rf	1 rf	2 rf
c(MARS)	71.83	84.51	86.94
c(MindReader)	71.83	84.70	87.49
c(New)	71.83	85.75	89.53
e(MARS)	40.11	59.24	62.73
e(MindReader)	40.11	38.35	43.07
e(New)	40.11	59.24	62.73
cte(MARS)	79.80	97.08	98.12
cte(MindReader)	79.80	85.44	87.36
cte(New)	87.88	96.98	98.25

tests. During subjective tests, both the Corel dataset and the Vistex dataset have been used. Users having various academic background (computer science, art, library science, etc.), as well as users from industry, have been invited to compare which technique they like the best. All the users rated the proposed technique the best in terms of both accuracy and fast response.

For objective tests, to avoid *ad-hoc* evaluations, we have decided to use the Vistex dataset as our primary test set. By doing so, we can establish an *unbiased* ground truth by considering the 16 images cut from the same big image as relevant images. During the retrieval process, at each iteration, the top 20 images are returned to the user. The numbers in Table 1 are the average retrieval performance over all the 832 images in the database and the retrieval performance is defined as

$$\frac{relevant\ ones\ retrieved}{16} \times 100\% \qquad (18)$$

The experimental results are summarized in Table 1, where "c", "e", and "cte" denote the experimental results by using color alone, edge alone, and all color, texture and edge, respectively. The symbol "rf" denotes how many iterations of relevance feedback.

The reason that we have chosen the three cases "c", "e", and "cte" is to simulate different retrieval conditions. Recall that the feature vector lengths for color, texture, and edge are 6, 10, and 18, respectively. Then "c" simulates the $N > K_i$ condition; "e" simulates the $N < K_i$ condition, and "cte" simulates a mixed condition. Based on the experimental results, we have the following observations:

- The proposed algorithm is the best in all the retrieval conditions (*robustness*);
- MindReader works well, *only if* the condition $N > K_i$ is satisfied, due to its difficulty of estimating C_i from insufficient data;
- Regardless of which technique used, the relevance feedback technique can significantly increase the retrieval performance.

One note needs to be pointed out: even though in this experiment the MARS technique works reasonably well, there are cases (see MindReader [3]) where it is not effective, due to its limited ability to model non-linear distance functions.

4 Conclusions

In this paper, we have proposed a novel relevance feedback technique which is not only comprehensive enough to include previous proposed approaches but also simple enough to allow us to derive explicit optimal solutions for the query estimation. Because of its two-level image content model, it requires less computation. Furthermore, because of its dynamic and intelligent selection of diagonal or full matrix for W_i, it achieves the best performance in all the retrieval conditions.

5 Acknowledgment

The Corel data set of images were obtained from the Corel collection and used in accordance with their copyright statement.

References

[1] J. P. Callan, W. B. Croft, and S. M. Harding, "The inquery retrieval system," in *Proc. of 3rd Int Conf on Database and Expert System Application*, Sept 1992.

[2] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," in *Proc. IEEE Int. Conf. on Image Proc.*, 1997.

[3] Y. Ishikawa, R. Subramanya, and C. Faloutsos, "Mindreader: Query databases through multiple examples," in *Proc. of the 24th VLDB Conference*, (New York), 1998.

[4] Y. Rui, "Efficient indexing, browsing and retrieval of image/video content," Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1999.

[5] S. X. Zhou, Y. Rui, and T. S. Huang, "Water-filling algorithm: A novel way for image feature extraction based on edge maps," in *Proc. IEEE Int. Conf. on Image Proc.*, 1999.