Relevance Feedback Content-Based Image Retrieval Using Query Distribution Estimation Based on Maximum Entropy Principle

Irwin King[1] and Zhong Jin[1,2]

[1] Department of Computer Science and Engineering,
The Chinese University of Hong Kong,
Shatin, New Territories, Hong Kong,
email: {king,zjin}@cse.cuhk.edu.hk

[2] Department of Computer Science,
Nanjing University of Science and Technology,
Nanjing, People's Republic of China

Abstract

In the last few years, we have seen an upsurge of interest in content-based image retrieval (CBIR)–the selection of images from a collection via features extracted from images themselves. Typically the nearest-neighbor rule is used to retrieve images from a query image. However, the underlying query distribution may not be isotropic in nature. Hence, a more sophisticated estimation for the query distribution is required. We propose a novel relevance feedback framework for image retrieval which contains two stages: (1) to estimate the query distribution based on relevance feedback information and (2) to generate a set of inquiries for relevance selection based on the Maximum Entropy Principle. We demonstrate these two stages in detail. Moreover, experiments have been performed on a trademark image database. The results show our proposed framework is effective in image retrieval with a few relevant samples.

1 Introduction

Content-based image retrieval (CBIR) has become one of the most active research areas in the past few years [1, 2]. Generally speaking, primitive visual features representing color, shape, and texture are extracted from an image to represent its content. Similar images can be retrieved from a collection of images on the basis of primitive features–either singly or combined. Successful content-based image retrieval systems require the integration of various techniques in the fields of pattern recognition, image processing, and information retrieval.

Relevance feedback (RF) in information retrieval (IR) is an iterative and interactive process for query reformulation based on user's feedback. Both a query vector moving technique and a re-weighting technique to estimate the ideal query parameter are introduced in MARS [3]. Rui et al. [4] present a relevance feedback based interactive retrieval approach, which effectively takes into account the following two distinct characteristics of CBIR systems: (1) the subjectivity of human perception of visual content and (2) the gap between high-level concepts and low-level features. A novel global optimization framework for relevance feedback is proposed in [5]. Most relevance feedback techniques in CBIR have only utilized information of the relevant retrieved images, and have not made great use of information of the irrelevant retrieved images. Recently, the authors have proposed a re-weighting technique by using both the relevant information and irrelevant information [6].

The key of relevance feedback techniques is in estimating the query distribution function. However, the retrievals under the commonly used nearest-neighbor rule cannot reflect the query distribution function properly since it is isotropic. Thus, most of relevance feedback techniques may fail under the following assumption:

1. The number of relevant retrievals is small
2. The number of iterations is required to be small

In order to improve the retrievals, one novel idea is to retrieve some representative samples of the query distribution function through relevance feedback.

In this paper, we proposed a novel relevance feedback framework for content-based image retrieval based on the Maximum Entropy Principle (MEP). The paper is organized as follows. In Section 2, background information is introduced. A novel framework for content-based image retrieval is proposed in Section 3. In Section 4, experiments are conducted and discussed. Lastly, we draw our conclusion in Section 5.

2 Background Review

Relevance Feedback Relevance feedback technique can be regarded as a form of automatic learning for the unknown query distribution function. It includes two stages in general:

Stage 1. Estimate the query distribution function. For example, we may estimate it by using the Expectation-Maximization (EM) algorithm [7] or estimate it by the classical statistics [8].

Stage 2. Generate the inquiries to be returned to the user. The nearest-neighbor rule is commonly used here. However, in our proposed approach the Maximum Entropy Principle (MEP) is used instead.

EM algorithm has been successfully used in solving the learning problem with mixture models. It can be used to deal with multi-class learning problems on a large of labeled data. Because of a small number of labeled data in RF, EM has its limitations in content-based image retrieval. Nonetheless, a technique has been designed to enlarge the labeled data to circumvent this problem some what [9]. On the other hand, the classical statistical approach can also be used to estimate the query distribution function. However, it cannot deal with two-class problems, and only relevant retrieval information can be utilized.

Maximum Entropy Principle In 1948, Shannon established the foundations of Information Theory [10]. He showed that for a probability distribution, $\mathbf{p} = (p_1, \cdots, p_k)$, over k possible elementary events $\{a_1, \cdots, a_k\}$, the quantity

$$H(\mathbf{p}) = -\sum_{i=1}^{k} p_i \log p_i \qquad (1)$$

is, within a constant factor, the unique quantity in accord with his assumptions.

Jaynes [11] converted Shannon's measure to a powerful instrument for the generation of statistical hypotheses, which is based on what has come to be known as the Maximum Entropy Principle (MEP). The maximum entropy estimate is obtained by determining that probability distribution associated with a random variable over a discrete space which has the greatest entropy subject to constraints on the expectations of a given set of functions of the variable. The Maximum Entropy (MAXENT) solution with no bias (or constraints) is

$$p_i = \frac{1}{k} \quad (i = 1, \cdots, k). \qquad (2)$$

The MEP has been applied to practical problems in many diverse areas. In the early 80's, Cooper et al. [12] made a strong case for applying the maximum entropy approach to the problems of information retrieval. Kantor [13] extended the analysis of the MEP in the context of information retrieval. Recently, Greiff and Ponte [14] took a fresh look at modeling approaches to information retrieval and analyzed classical probabilistic IR models in light of the MEP. They showed that the probabilistically motivated IR models can result from the MEP.

3 Proposed Framework

Suppose an image database S is composed of N distinct images $\{I_1, I_2, \cdots, I_N\}$. It is assumed that for a query image I_q, there are some similar images in the database. The task of image retrieval is to retrieve as many similar images in the database S for a given query image, I_q.

3.1 Problems

In content-based image retrieval, the similarity between two images is determined by features extracted from themselves. In this paper, only a single feature vector is used to represent an image. Suppose an M-dimensional feature vector, $X(I)$, is extracted from each image I. Because of the similarity, it can be assumed that the query image I_q and the corresponding similar images are clustered in the M-dimensional feature space $\Re^M$ with a kind of statistical distribution function $P(x)$. We define $P(x)$ as a **query distribution function** (QDF) in this paper.

Typically, the QDF, $P(x)$, cannot be easily determined directly based on a single query. Therefore, it is sometimes preferable to adopt a less complete, but more computable, characterization. The most important parameters are the mean, μ, and the covariance matrix, Σ, which are defined respectively as:

$$\mu = \int x P(x) dx, \quad (3)$$

$$\Sigma = \int (x-\mu)(x-\mu)' P(x) dx. \quad (4)$$

Commonly, both the query mean, μ, and the query covariance matrix, Σ, are unknown. At the beginning, what we have is only a query image I_q. Fortunately, an estimation for the query mean μ can be given easily from the only one image I_q as

$$\hat{\mu} = X(I_q). \quad (5)$$

Therefore, the similarity function between the query image I_q and any image $I \in S$ can be measured according to the Euler distance as

$$D_1(I) = ||X(I) - \hat{\mu}||^2. \quad (6)$$

According to this similarity function, $D_1(I)$, image retrieval can be conducted as follows: All the N similarity values $\{D_1(I), (i \in S)\}$ are computed and ranked in non-decreasing order and the K most similar ones, $\{I_{j1}, I_{j2}, \cdots, I_{jK}\}$, are returned to the user, where K is the number of images the user wants to retrieve.

In order to improve the retrievals, relevance feedback techniques are used in content-based image retrieval. In order to update the estimation of the query mean, μ, a query vector moving technique by relevance feedback is commonly utilized in content-based image retrieval. It is assumed that for any retrieved image, a degree of relevance can be given by user according to his information need and perception subjectivity. The idea is to use this relevance information to improve the retrievals.

In this paper, suppose that for any retrieved image, the user simply marks the image as `relevant` or `irrelevant`. Moreover, we further assume that T accumulative relevant retrievals $\{I_{r1}, I_{r2}, \cdots, I_{rT}\}$ be given by the user. Since the query image I_q can be regarded as a relevant retrieval, we can denote it as I_{r1}, i.e., we have $I_{r1} = I_q$. Thus, the number of relevant retrievals is no less than 1, i.e., we have $T \geq 1$. In order to improve the retrieval performance, a more accurate estimation for the query mean μ can be proposed on the relevance information:

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} X(I_{rt}), \quad (7)$$

Eqs. (5), (6), and (7) are used as a query moving technique in our proposed framework for content-based image retrieval.

There are three observations with this existing framework that should be noted.

1. The irrelevant retrievals, which users mark as "`irrelevant`", contain relevant information also. This kind of relevant information has not been utilized effectively in the past.

2. The query covariance matrix, Σ, contains much discriminant information of feature components. This discriminant information has not been utilized fully.

3. The retrievals under the nearest-neighbor rule typically cannot efficiently represent the query distribution function $P(x)$ since the underlying distribution is often non-isotropic.

3.2 Estimation Approach

Now, we are required to estimate the query covariance matrix Σ. Suppose that there are T accumulative relevant retrievals $\{I_{r1}, I_{r2}, \cdots, I_{rT}\}$, where $T > 1$. According to the classical statistical theory, Σ can be estimated from the T relevant retrievals and the query image I_q as follows.

$$\hat{\Sigma} = \frac{1}{T-1} \sum_{t=1}^{T} (X(I_{rt}) - \hat{\mu})(X(I_{rt}) - \hat{\mu})', \quad (8)$$

where $\hat{\mu}$ is defined in Eq. (7).

Obviously, the rank of Σ is no more than $T-1$. Only when T is much larger than M, the query covariance matrix Σ can be estimated to be somewhat accurate and can be used in image retrieval. On the other hand, the dimension of the feature vector is needed to be large enough to provide sufficient discriminant information in content-based image retrieval. However, since there are a small number of relevant retrievals, the query covariance matrix Σ will be singular because $T-1$ is often much less than the dimension of the feature vector.

In order to manage with only a small number of relevant retrievals, it is now assumed that the feature vector used is composed of independent components. For an independent feature vector, the query covariance matrix, Σ, is an M-dimensional diagonal matrix expressed as

$$\Sigma = diag\{\sigma_1^2, \cdots, \sigma_M^2\}, \quad (9)$$

where $\sigma_i (i = 1, \cdots, M)$ is the standard deviation.

Under the assumption of Gaussian probability distribution function, the i-th component y_i is a 1-dimensional Gaussian probability distribution function having a mean of μ_i and a standard deviation of σ_i.

Let d_i be the minimum Euler distance for i-th component among the query mean $\hat{\mu}$ and the feature vectors of all the irrelevant retrievals the user marks. According to the following equal-probability condition

$$P\{|y_i - \mu_i| < d_i\} = P\{|y_i - \mu_i| \geq d_i\} = 0.5, \tag{10}$$

An estimation of σ_i can be given as

$$\hat{\sigma}_i = \frac{d_i}{0.6745}. \tag{11}$$

Obviously, Eq. (11) can be utilized by using irrelevant retrieval information in case of only one relevant retrieval, the query image, I_q, i.e., $T = 1$.

3.3 MAXENT Generation

Generally speaking, the retrievals under the nearest neighbor rule are very near to the query means, μ. It is easy to understand that these retrievals cannot effectively reflect the covariance property of the query distribution function. Therefore, the existing framework of retrieval systems may fail in practical use.

In order to improve the final retrievals efficiently, one novel idea is to retrieve some representative retrievals firstly. Now, we propose a retrieval step based on the MEP which we call it as the MAXENT retrieval.

Suppose $Y = [y_1, \cdots, y_M]$ is a random variable having the same distribution function as the query distribution function, $P(x)$. Suppose that $P(x)$ is an M-dimensional Gaussian probability distribution function also. Then, we have

$$Z(Y) = (Y - \mu)'\Sigma^{-1}(Y - \mu). \tag{12}$$

It was proved that this 1-dimensional random variable, $Z(Y)$, has a χ^2 distribution function with a freedom degree of M [15]. For K number of retrievals, $K + 1$ points $\{z_i,\ (i = 1, \cdots, K + 1)\}$ can be determined according to the following equal-probability conditions:

$$\int_{z_i}^{z_{i+1}} \frac{1}{2^{M/2}\Gamma(M/2)} z^{\frac{M}{2}-1} e^{-\frac{z}{2}} dz = \frac{1}{K}. \tag{13}$$

Thus, all similar images in the database S can be divided in the following K subsets:

$$S_i = \{\ I \mid z_i \leq Z(X(I)) < z_{i+1}, I \in S\}. \tag{14}$$

If S_i is nonempty, an element can be randomly selected as a retrieval from the subset S_i. In this way, K retrievals can be retrieved at most. These retrievals can be called MAXENT retrievals because of the similarity of Eq. (2) and Eq. (13).

3.4 Novel Framework

Our novel framework for image retrieval includes the following stages:

Estimation: Estimate the query mean and the query covariance matrix according to Eq. (7), Eq. (8) for $T > 1$ or Eq. (11) for $T = 1$ by using accumulative relevance retrieval information and irrelevant retrieval information.

Generation Based on MAXENT: Generate subsets S_i $(i = 1, \cdots, K)$ according to Eq. (13) and Eq. (14), and return at most K MAXENT retrievals to the user.

It is noted that these two stages can be performed iteratively.

4 Experiments and Analysis

In this section, we conduct experiments to test the effectiveness of our proposed framework. The experiments are performed on a Sun Ultra 5/270 machine with 128 RAM under Solaris 2.6 using C++. A demo of our system on the web can be found at [16].

4.1 Database

There are 1,400 trademark images with 128×128, ten samples of which are shown in Fig. 1 (a). Based on ten different deformed transformations which include pinch, twirl, ripple, and extension [17], 100 deformed images are generated from the ten samples, and they are shown in Fig. 1 (b).

Some features such as eccentricity, invariant moments, circularity, and Fourier descriptors of approximated boundary, Legendre moments, and Zernike moments have been successful utilized in our previous work on trademark image retrieval [6]. In this paper, a six-dimensional feature vector, which is composed of the first components of six features mentioned above, is used to represent each trademark image. It can be regarded as having six independent components.

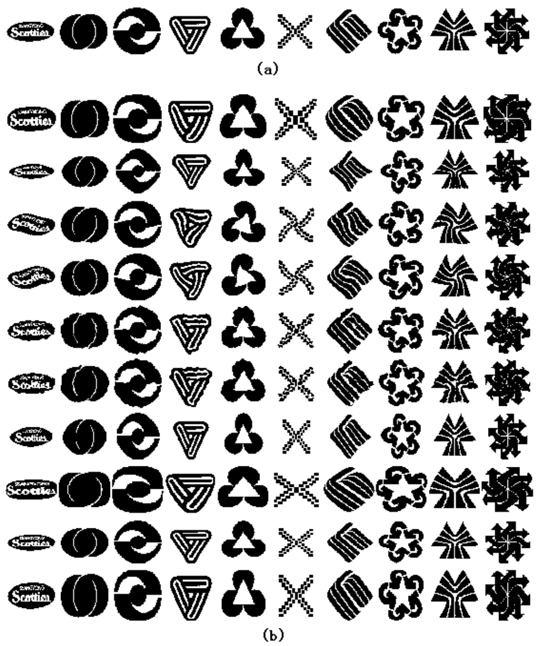

Figure 1: Deformed Images

4.2 Experimental Design

We aim to evaluate the efficiency of the proposed generation stage, Generation MAXENT. The following two commonly used distance functions are used in the generation stage:

Euler Distance: Generate K retrievals under the nearest neighbor rule by the Euler distance in Eq. (6).

Mahalanobis Distance: Generate K retrievals under the nearest neighbor rule by the following Mahalanobis distance:

$$D_2(I) = (X(I) - \hat{\mu})'\hat{\Sigma}^{-1}(X(I) - \hat{\mu}). \quad (15)$$

In order to compare Generation MAXENT with Euler distance and Mahalanobis distance, a set of three-step experiments are designed as follows:

Step 1: For a query image I_q, estimate the query mean according to Eq. (5), and return K retrievals by Euler distance.

Step 2: Perform the estimation stage Estimation MAXENT of the proposed framework, and return retrievals by Euler distance, Mahalanobis distance, and Generation MAXENT respectively.

Step 3: Perform the estimation stage Estimation MAXENT of the proposed framework, and return K retrievals by using the Mahalanobis distance.

The retrieval performance is measured using the following Average Retrieval Precision (ARP):

$$ARP = \frac{relevant\ \ retrieved}{K} \times 100\%, \quad (16)$$

where $K = 10$ is the number of total retrieved images.

4.3 Result Analysis

By taking each of the 100 deformed trademark images as a query respectively, retrieval experiments have been performed. The detailed experimental results can be obtained at [16]. The average ARPs for all the 100 query images are listed in Table 1. The proposed generation stage Generation Based On MAXENT aims to retrieve image samples which can reflect the query distribution function. This is the reason why the ARP of MAXENT in Step 2 in Table 1 is lower than those of Euler distance and Mahalanobis distance. According to the ARP's in Step 3 in Table 1, we can see that the proposed generation stage Generation MAXENT outperforms the commonly used Euler distance and Mahalanobis distance which resulted in a great improvement for retrieval performance.

Table 1: ARP in Percentage

ARP	Generation Stage		
	Euler	Mahalanobis	MAXENT
Step 1	16.5		
Step 2	16.9	20.6	9.8
Step 3	20.6	23.6	48.3

In these experiments, it is shown that our proposed framework can be used effectively and efficiently to estimate the query distribution function resulting in an improved retrieval by using all retrieval information.

5 Conclusion

In this paper, a novel two-stage relevance feedback framework for content-based image retrieval based on query estimation and the Maximum Entropy Principle is proposed. Experiments on a trademark image database show that the framework improves accuracy and speed. Our future work will concentrate on overcoming the difficulty in image retrieval for high-dimensional and non-independent features in large image databases.

Acknowledgment

This paper is supported in part by an Earmarked Grant from the Hong Kong Research Grant Council #CUHK4407/99E.

References

[1] A. K. Jain and A. Vailaya. Shape-based retrieval: A case study with trademark image databases. *Pattern Recognition*, 31(9):1369–1390, 1998.

[2] Alberto Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann Publishers, 1999.

[3] Y. Rui, T. S. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in MARS. In *Proceedings of IEEE International Conference on Image Processing*, pages 815–818, Santa Barbara, California, October 1997.

[4] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. 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.

[5] Y. Rui and T. S. Huang. A novel relevance feedback technique in image retrieval. In *Proceedings ACM Multimedia '99(Part2)*, pages 67–70, Orlando, FL, USA, 1999.

[6] Z. Jin, I. King, and X. Q. Li. Content-based image retrieval by relevance feedback. In *Lecture Notes in Computer Science*, volume 1929, pages 521–529, Lyon, France, Nov. 2-4 2000. Springer.

[7] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum liklihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B 39(1):1–38, July 1977.

[8] A. Papoulis. *Probability, Random Variable, and Stochastic Processes*. p.250, McGraw-Hill, New York, 1965.

[9] Ying Wu, Qi Tian, and Thomas S. Huang. Discriminant-em algorithm with application to image retrieval. In *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 222–227, June 2000.

[10] C. E. Shannon. A mathematical theorey of communication. *The Bell System Technical Journal*, 27:379–423 & 623–656, 1948.

[11] E. T. Jaynes. Information theory and statistical mechanics: Part *I. Physical Review 106*, pages 620–630, 1957.

[12] W. S. Cooper and P. Huizinga. The maximum entropy principle and its application to the design of probabilistic retrieval systems. *Information Technology, Research & Development*, 1:99–112, 1982.

[13] P. B. Kantor. Maximum entropy and the optimal design of automated information retrieval systems. *Information Technology, Research & Development*, 3(2):88–94, 1984.

[14] Warren R. Greiff and Jay M. Ponte. The maximum entropy approach and probabilistic ir models. *ACM Transactions on Information Systems*, 18(3):246–287, July 2000.

[15] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, New York, 2nd edition, 1990.

[16] http://www.cse.cuhk.edu.hk/~miplab/MAXENT/.

[17] D. Y. M. Chan and I. King. Genetic algorithm for weights assignment in dissimilarity function for trademark retrieval. In *Lecture Notes in Computer Science*, volume 1614, pages 557–565, The Netherlands, 1999. Springer.