Relevance Feedback Based on Parameter Estimation of Target Distribution

K.C. Sia and Irwin King
{kcsia, king}@cse.cuhk.edu.hk

Department of Computer Science and Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong

Abstract - Relevance Feedback formulations have been proposed to refine query result in Content-Based Image Retrieval (CBIR) in the past few years. Many of them focus on a learning approach to solve the feedback problem. In this paper, we present an Expectation Maximization (EM) approach to estimate the user's target distribution through user's feedback. Furthermore, we describe how to use Maximum Entropy display to fully utilize user's feedback information. We detail the process and also demonstrate the result through experiments.

Keywords: CBIR, parameter estimation, expectation maximization, relevance feedback

I. Introduction

In Content-Based Image Retrieval Systems, low level features vectors, which may include texture, color and shape, are extracted from the image for storage, manipulation, and retrieval purpose. Many systems use the one-shot approach during retrieval where a query is given in the form of a feature vector and the result is calculated based on the distance between feature vectors of query and images in database, whil the similarity of images is based on this distance measure.

In order to understand the user's need in a search, a relevance feedback approach is used. Every time, the retrieval system presents the user a set of images, the user then selects those, he thinks, which are relevant and gives feedback to the retrieval system. Based on the feedback, the system can capture the user's need more accurately.

In this paper, we describe some current approaches in Section 2. In Section 3, we point out some problems that exist in current approaches. We then propose a relevance feedback architecture that use EM to estimate the user's target and maximum entropy display to utilizes the user's distinguishing power in order to narrow down the retrieval process. In Section 4, we verify the correctness of our algorithm and compare our approach with Rui's approach using the precision versus recall measure. We then give some final remarks and conclude.

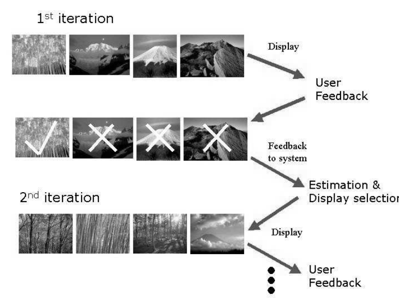

Fig. 1. Relevance feedback architecture

II. Background

Relevance feedback is used to capture users' searching criteria. In each pass, the retrieval system presents user a set of images, user then gives feedback to the system, indicating which one is relevant. The system makes some update to the parameter and presents user another set of images. This process is illustrated in Fig. 1.

Base on this architecture, Rui et. al. [5] formulated a weight updating method to capture user's preference on different features, such as color or texture. Cox et. al. [1] formulated a Bayesian Learning approach to learn which image is more likely to be user's target based on the feedback. In the following section, we review these two approaches in detail.

A. Rui's Weight Updating Method

In [5], objects in image database are modelled as

$$O = O(D, F, R), \tag{1}$$

where D is the raw image data, $F = \{f_i\}$ is a set of low level visual features, such as color, texture, and shape,

0-7803-7278-6/02/$10.00 ©2002 IEEE

and $R = \{r_{ij}\}$ is the set of representations for f_i, which is defined as

$$r_{ij} = [r_{ij1}, ..., r_{ijk}, ..., r_{ijK}]. \quad (2)$$

Moreover, the feature vector is organized in a hierarchical manner. The overall similarity of two images O^a and O^b is defined as

$$\begin{aligned} S(O^a, O^b) &= \sum_i W_i S(f_i^a, f_i^b), & (3) \\ S(f_i^a, f_i^b) &= \sum_j W_{ij} S(r_{ij}^a, r_{ij}^b), & (4) \\ S(r_{ij}^a, r_{ij}^b) &= m(r_{ij}^a, r_{ij}^b, W_{ijk}), & (5) \end{aligned}$$

where m is the distance measure function, while W_i, W_{ij} and W_{ijk} are the weights associated with each features, its representation and each dimension respectively. For each feedback, they will follow the two procedures described below to update the weight in order to capture user's interest in different features.

B. Interweight Updating Procedure

Suppose RT is the set of most similar images according to the overall similarity function Eq. 3, for each feature representation ij, the system retrieves a set of similar images, RT^{ij}, according to that particular feature representation. The weight, W_{ij}, is updated according to

$$W_{ij} = \begin{cases} W_{ij} + R, & RT_l^{ij} \in RT \text{ and } l = 1, ..., N, \\ W_{ij}, & \text{otherwise} \end{cases}$$

where N is the number of most similar images and R is the degree of relevance indicated by the user.

C. Intraweight Updating Procedure

For the set of relevant images indicated by user's feedback, the system computes the standard deviation, σ_{ijk}, in each dimension, and the weight for each dimension is updated as Eq. (6). We can see that if σ_{ijk} is large, then the dimension is not ideal for discriminating relevant and irrelevant images, so its weight is updated as follows,

$$W_{ijk} = \frac{1}{\sigma_{ijk}}. \quad (6)$$

D. Cox's Bayesian Formulation Method

In [1], each image is associated with a probability of being the user's target. The retrieval process consists of two steps. In each pass, the system selects a set of images and presents to user. Through the feedback, the system updates the likelihood measure to the query of each image accordingly. The probability is updated using the Bayes' rule as follows,

$$P(T = T_i | H_t) = \frac{P(A_t | T = T_i, D_t, S_{t-1}) P(T = T_i | H_{t-1})}{\sum_{j=1}^{n} P(A_t | T_j, D_t, S_{t-1}) P(T_j | H_{t-1})}. \quad (7)$$

The meaning of Eq. (7) is that the probability of T_i being the target image at iteration t is equal to product of the probability of T_i being the target at iteration $t-1$ and the probability of user give such feedback at iteration t provided that T_i is the target, over the summation of probability of other images.

Moreover, as each image is associated with a probability of being the target, [1] proposed a maximum entropy display strategy to select image presenting to user. As a result, the system is expected to get most information gain from user's feedback. Besides, [3] also details the procedure to apply this strategy. This method inspires the display model of our proposal.

E. Expectation Maximization

Expectation Maximization (EM) algorithm was first proposed in 1977 [2] . It was used to solve the maximum-likelihood from incomplete data. Given a mixture of Gaussians, the EM algorithm estimates the parameter of each mixture, say, mean and variance. Our approach is based on EM algorithm to estimate the parameter of user's target distribution.

Nigam et. al. [4] and Wu et. al. [8] also used the EM algorithm to classify documents and images respectively. The EM approaches utilize the information contained in unlabeled data (irrelevant data) to help estimating user's target distribution more efficiently and accurately. More recently, Wang et. al. [7] proposed a Wiener filter approach to learn user's feedback in an optimal fashion. Tian et. al. [6] proposed the Support Vector Machine (SVM) to classify out the user's target. All these newly suggested methods tends to model the feedback process as a learning process and apply algorithms in computer learning to help.

III. Problem Definition and Proposed Solution

A. Problem Definition

Estimation

In [5], the weight updating method is a distance based similarity measure. The weight is a measure of how important a particular feature, or dimension is in the query process. It makes use of the weight in calculating the distance measure. While in [1], a global update of probability to all images in database is used. It is not parametric based, and the global updating processes seems to be the computational bound when the size of image database grows. In our proposal, we estimate the parameter of user's target distribution. With these parameters, we can capture user's need more accurately, and the process of selecting images to display becomes easier.

Display Selection

Many other approaches always overlook the process of selecting images to display. If we keep displaying the most similar images to user, we have no way to capture user's need in a broader sense. In [1], the most-informative display updating scheme try to achieve this, since each images is associated with a probability value, a maximum entropy display is used to select images. However, when the size of image database is large, the number of permutation is huge, so a sampling approach is used to choose images that maximize the entropy. In our proposal, we have estimated the user's target distribution parameters, thus we can select images located in the boundary to display, as illustrated in Fig. 2. This is analogous to the maximum entropy selection. Through this way, we can have the most information gain from user's feedback. Moreover, our estimation strategy is related to our display selection strategy greatly. Since we select images located around the boundary region, our estimation is not simply calculating the variance of relevant images directly. Instead, we use a new approach, which will be discussed in section III-D.2.

B. Proposed Solution

We propose to use a statistical learning method, expectation maximization (EM) to estimate the distribution of user's search target, which will fully utilize the distinguish power by user in classifying relevant and irrelevant images.

C. Model

Let $DB = \{I_i\}_{i=1}^{n}$ be a set of database objects, and a set of feature parameters be $\theta = \{\theta_i\}_{i=1}^{m}$. For each image in the database, we perform low level feature extraction to map it to a high dimensional data point by function f, which extracts a real-valued d-dimensional vector as,

$$f : I \times \theta \rightarrow R^d, \tag{8}$$

where θ_i means a specific feature, for example, the color histogram, the co-occurrence matrix based texture feature or the Fourier descriptor. Then an image will be mapped to a high dimensional vector, R^d. We assume the user's searching target distribution is a cluster in the high dimensional space. Our goal is to estimate this distribution as accurately as possible, based on the user's

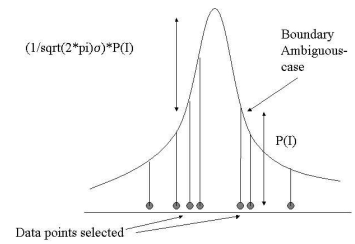

Fig. 2. Fitting μ and σ for data points selected to display

feedback. We focus on one feature at this moment first. For each dimension under this feature, we estimate the mean, μ and variance, δ of the user's target distribution.

D. Relevance Feedback

D.1 Resolving conflicts

In each pass, we will give user a set of images to choose. The user then indicates whether a specific image is relevant or not. Let R^+ and R^- be the set of relevant images and the set of irrelevant images in each pass respectively. Let $Rel(\vec{I_i})$ be the measure of how relevant an image $\vec{I_i}$ and be defined as.

$$Rel_{t+1}(\vec{I_i}) = Rel_t(\vec{I_i}) + 1 \quad \vec{I_i} \in R^+ \tag{9}$$

$$Rel_{t+1}(\vec{I_i}) = Rel_t(\vec{I_i}) - 1 \quad \vec{I_i} \in R^- \tag{10}$$

We only consider images of $Rel(\vec{I}) > 0$ and $Rel(\vec{I}) < 0$. The images are divided into two classes, the relevant one and the irrelevant one. we indicate it as I^+ and I^- respectively. Using equation Eq. (9) and Eq. (10), we can resolve the conflict between successive feedbacks given by user.

D.2 Expectation Maximization Approach

After the user indicated some relevant and irrelevant images, we estimate the mean and variance of user's target distribution in each dimension by the EM algorithm. In Fig. 2, we try to fit the Gaussian distribution having the data points selected by user located in the boundary region; in other words, we are going to maximize the expression: Eq. 11. The reason why we use this expression is that most of the images we give to user to distinguish will fall in the area of **medium** likelihood (boundary case), so we fit our maximum likelihood function in order to make images appear in the **medium** likelihood region.

$$E = \sum_{I_i \in I^+} \sum_{j=1} P(I_i|\theta_j) \times (\frac{1}{\sqrt{2\pi}\delta_j} - P(I_i|\theta j)) \tag{11}$$

$$P(I_i|\theta j) = \frac{1}{\sqrt{2\pi}\delta_j} \exp^{-\frac{(I_{ij}-\mu_j)^2}{2\delta_j^2}} \tag{12}$$

where j is the subscript for dimension and θ_j is the combination of μ and δ for a particular dimension j. I^+ is the set of relevant images and I_{ij} is the value of feature vector of image I_i in dimension j.

For finding the mean, the obvious way is to find the average of all relevant data, i.e.,

$$\mu_j = \frac{\sum_{i \in I^+} I_{ij}}{|I^+|}. \tag{13}$$

In order to find the best fitting δ_j, we differentiate E with respect to δ_j as follows.

$$\frac{dE}{d\delta_j} = 0 \tag{14}$$

$$\sum_{i \in I^+} \left(\frac{-1}{\pi\delta^3} \exp^{-\frac{(I_{ij}-\mu_j)^2}{2\delta_j^2}} + \frac{(I_{ij}-\mu_j)^2}{2\pi\delta^5} \exp^{-\frac{(I_{ij}-\mu_j)^2}{2\delta_j^2}} \right.$$
$$\left. + \frac{1}{\pi\delta^3} \exp^{-\frac{(I_{ij}-\mu_j)^2}{\delta_j^2}} - \frac{(P_{ij}-\mu_j)^2}{\pi\delta^5} \exp^{-\frac{(I_{ij}-\mu_j)^2}{\delta_j^2}} \right) = 0 \tag{15}$$

As this maximum likelihood objective function is hard to solve, we use EM algorithm to estimate the δ_j. We make use of the parameter from the previous step to derived the new parameter value. We substitute $\frac{\delta_j^{\frac{1}{2}}}{\sqrt{2\pi}\delta_j} \exp^{-\frac{(I_{ij}-\mu_j)^2}{2\delta_j^2}}$ by $\delta_{old_j}^{\frac{1}{2}} P(I_{ij}|\theta_{old_j})$, and come out with an update equation for δ_j:

$$\delta_j^2 = \frac{\sum_{i \in I^+}((I_{ij}-\mu_j)^2 2^{\frac{1}{4}} \pi^{-\frac{3}{4}} - \delta_{old_j}^{\frac{1}{2}} P_{old}^{\frac{1}{2}} (I_{ij}-\mu_j)^2 2^{\frac{1}{2}} \pi^{-\frac{1}{2}})}{\sum_{i \in I^+}(2^{\frac{1}{4}} \pi^{-\frac{3}{4}} - \delta_{old_j}^{\frac{1}{2}} P_{old}^{\frac{1}{2}} 2^{\frac{1}{2}} \pi^{-1})} \tag{16}$$

E. Maximum Entropy Display

Since we have captured the μ and δ of user's target distribution, we proceed to select images that lie in the boundary case to display, and let user determine whether they are relevant or irrelevant. We choose images that are $\pm k\delta$ away from the μ such that $P(\mu \pm k\delta) = \frac{1}{\sqrt{2\pi}\delta} - P(\mu \pm k\delta)$. After solving this equation, we found that k is equal to 1.1774. In our experiment, we choose points around $\mu + k\delta$ or $\mu - k\delta$ in each dimension to display. This is analogous to the maximum entropy display as we choose the ambiguous images for user to classify, thus we fully utilize the power of user in distinguishing different image classes.

TABLE I
PARAMETERS OF THE SYNTHETIC DATA

Dimension	Class No.	Class Size	Range of μ	Range of δ
4	50	50	[-1,1]	[0.2,0.6]
6	70	50	[-1.5,1.5]	[0.2,0.6]
8	85	50	[-1.5,1.5]	[0.15,0.45]

IV. Experiments

Here, we propose a set of experiments to verify the correctness and measure the performance of our proposed algorithm. We would like to make sure of the convergence property is met. Moreover, we compare our proposed method with the Rui's method in terms of accuracy.

1. We generate a mixture of Gaussians with class labels and store in a text file.
2. Base on our proposed algorithm, the program selects 18 data points in each iteration, and presents their class label to user.
3. The user choose one class as his target, if he find data points come from his target, he gives feedback to system indicating that they are relevant.
4. After several iteration, we see if the estimated distribution parameters converge towards the parameters used to generate the user's target class
5. We use Root-Mean-Square error to measure the difference between actual and estimated μ and δ.

A. Experimental Setup

In the experiments we focus only on synthetic data sets. These data sets are generated by Matlab as mixture of Gaussians. We specify the mean and variance for each class and use Matlab random function to generate. Our experiments were performed on program written by C++ running on Sun Ultra 5/400 with 256Mb ram.

The synthetic data sets are mixture of Gaussians, and the parameters are listed in Table 1. The μ of each class is uniformly distributed within the range, and the δ of each class is value lies within the range indicated of probability 0.68 (the two values are the ± 1 standard deviation of the gaussian function used to generate δ).

A.1 Convergence Experiment

To test the convergence property, we make sure that the Root-Mean-Square (RMS) error decreases in each iteration. Tables 2-4 demonstrate the RMS error of estimated mean and standard deviation along each iteration. The fields indicated as not applicable are those with fewer than 3 relevant samples given. It is because our algorithm starts to estimate the mean and variance when 2 and 3 relevant data points are accumulated respectively.

TABLE II

FOUR DIMENSIONAL TEST CASE DATA

Iteration	Feedback Given	RMS mean	RMS std
1	1.5	not applicable	not applicable
2	1.5	0.292545	0.20655
3	5.5	0.217373	0.203525
4	6.5	0.19565	0.180268
5	5.75	0.202975	0.16099
6	9.25	0.156245	0.134668
7	7	0.154993	0.1253
8	5.25	0.146323	0.116223
9	7	0.13309	0.111628

TABLE III

SIX DIMENSIONAL TEST CASE DATA

Iteration	Feedback Given	RMS mean	RMS std
1	1.25	not applicable	not applicable
2	4	0.269095	not applicable
3	3.75	0.237395	not applicable
4	4.25	0.23813	0.182255
5	2.75	0.286803	0.172855
6	7.25	0.207565	0.136693
7	9.5	0.1705	0.122663
8	8.5	0.151863	0.122808
9	9.5	0.155308	0.121773
10	8.25	0.143003	0.10449

The data below for each dimension is an average of 4 test cases for that particular dimension.

A.2 Performance Experiment

We have implemented the intraweight updating version (Eq. 6) of Rui's approach to compare with our EM approach to see how we can improve the retrieval accuracy by estimating user's target. We carry out this experiment using synthetic data, which is a mixture of Gaussians with the parameter same as the 4 dimensional case in the previous experiment.

According to the intraweight update equation, we update the weight base on variance of retrieved relevant data points. After several iterations (normally 6 to 7), we compute a K-nn(K nearest neighbors) search incorporating the weight measure to analyze the precision versus recall measure. For our proposed algorithm, as we can estimate the μ and δ of the target distribution, we again perform a K-nn search starting from the mean μ while dropping those data points away from the μ more than two δ. Fig. 6,7,8,9 shows the precision versus recall graph for 4 test cases. We have done 9 cases, in 4 of them, EM outperforms Rui's approach, while in other 5 cases, EM ties or performs a little better than Rui's approach.

TABLE IV

EIGHT DIMENSIONAL TEST CASE DATA

Iteration	Feedback Given	RMS mean	RMS std
1	1.75	0.203065	not applicable
2	5.25	0.22882	0.13232
3	9.25	0.215707	0.087263
4	9.75	0.176893	0.059613
5	12	0.215953	0.059937
6	13.5	0.199765	0.065423
7	15.75	0.16033	0.052733
8	16	0.147903	0.052118
9	15.25	0.111283	0.057955
10	16.25	0.10208	0.05726

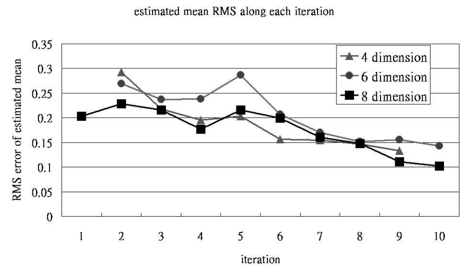

Fig. 3. RMS of estimated mean along each feedback iteration

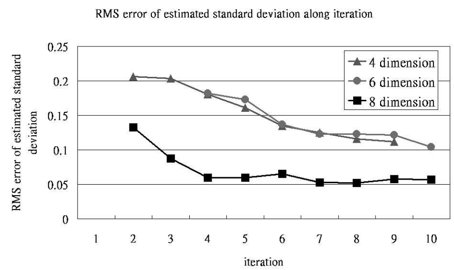

Fig. 4. RMS of estimated standard deviation along each feedback iteration

V. Conclusion

In this paper, we proposed an approach for CBIR to estimate the user's target through learning from user's feedback via EM algorithm. We have demonstrated the correctness and accuracy of our algorithm. We proposed a display selection strategy that utilizes the information

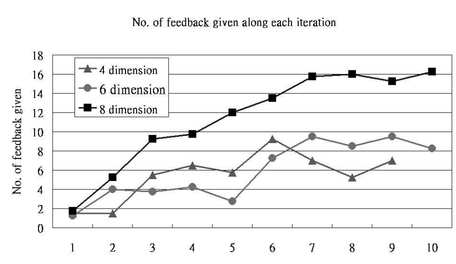

Fig. 5. No. of feedback given along each feedback iteration

0-7803-7278-6/02/$10.00 ©2002 IEEE

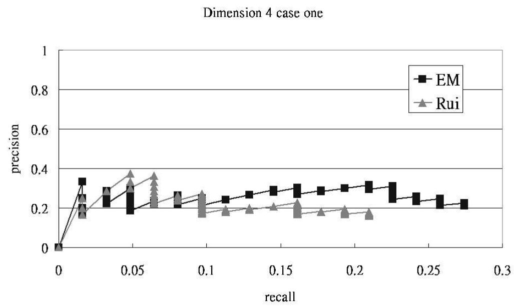

Fig. 6. precision recall graph : tie case

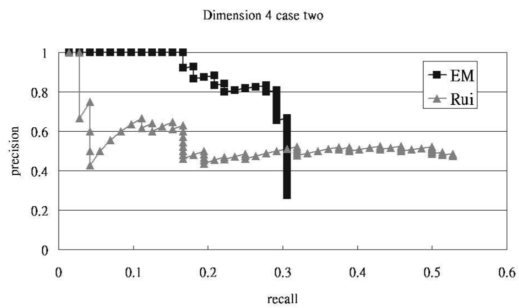

Fig. 7. precision recall graph : win case

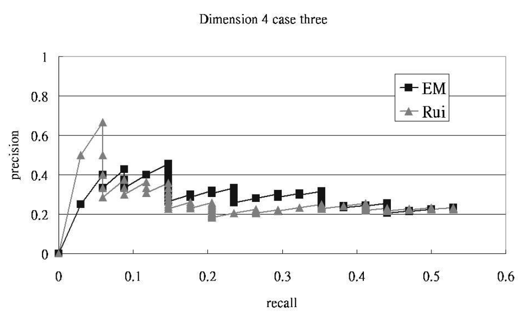

Fig. 8. precision recall graph : tie case

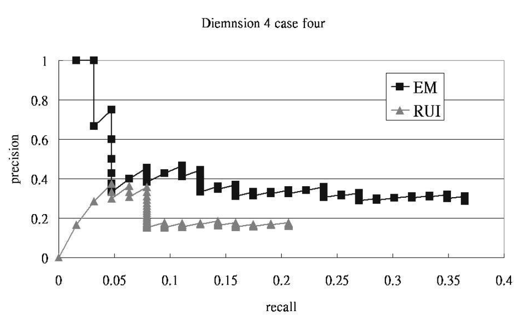

Fig. 9. precision recall graph : win case

given by user in dividing image classes in contrast to the K-nn search approach for result display. Moreover, we try to solve the conflicts between successive feedbacks from user. Also, our method is based on the parameter estimation of target distribution, we do not need to perform a global update each image in the database accordingly.

However, our algorithm also has weaknesses. Since we need to accumulate up to 3 relevant data points before estimating the standard deviation, user might find this too long. Moreover, we use maximum entropy strategy for display, thus, user might feel that the system cannot present the most relevant data in the retrieval process.

Acknowledgement

This research is supported in part by an Earmarked Grant from the Hong Kong's University Grants Committee (UGC), CUHK #4407/99E.

References

[1] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, and P. N. Yianilos. The bayesian image retrieval system, pichunter, theory, implementation, and psychophysical experiments. *IEEE Transactions on Image Processing*, 9(20-37), January 2000.

[2] A. Dempster, N. Laird, and R. D.B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39:185–197, 1977.

[3] I. King and Z. Jin. Relevance feedback content-based image retrieval using query distribution estimation based on maximum entropy principle. In L. Zhang and F. Gu, editors, *Proceedings to the International Conference on Neural Information Processing (ICONIP2001)*, volume 2, pages 699–704, Shanghai, China, November 14-18 2001. Fudan University, Fudan University Press.

[4] K. Nigam, A. K. Mccallum, S. Thrun, and T. Mitchell. Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39(2/3):103–134, 2000.

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

[6] Q. Tian, P. Hong, and T. Huang. Update relevant image weights for content-based image retrieval using support vector machines. In *Proceedings to the IEEE Conference on Multimedia and Expo*, volume 2, pages 1199–1202, June 2000.

[7] T. Wang, Y. Rui, and S.-M. Hu. Optimal adaptive learning for image retrieval. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, December 2001.

[8] Y. Wu, Q. Tian, and T. Huang. Discriminant-em algorithm with application to image retrieval. In *Proceedings to the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 222–227, June 2000.