A New Semi-Supervised EM Algorithm for Image Retrieval

Anlei Dong and Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, California 92521, USA
{adong, bhanu}@cris.ucr.edu

Abstract

One of the main tasks in content-based image retrieval (CBIR) is to reduce the gap between low-level visual features and high-level human concepts. This paper presents a new semi-supervised EM algorithm (NSS-EM), where the image distribution in feature space is modeled as a mixture of Gaussian densities. Due to the statistical mechanism of accumulating and processing meta knowledge, the NSS-EM algorithm with long-term learning of mixture model parameters can deal with the cases where users may mislabel images during relevance feedback. Our approach that integrates mixture model of the data, relevance feedback and long-term learning helps to improve retrieval performance. The concept learning is incrementally refined with increased retrieval experiences. Experiment results on Corel database show the efficacy of our proposed concept learning approach.

1 Introduction

In recent years, content-based image retrieval (CBIR) has received widespread research interest in the field of computer vision and pattern recognition. Based on the visual features (such as color, texture and shape) extracted from images, CBIR systems attempt to cater to the needs of users who want to retrieve images belonging to their desired concepts in mind.

The relevant feedback mechanism makes it possible for CBIR systems to learn human concepts since users provide some positive and negative image labeling information, which helps systems to dynamically adapt and update the relevance of images to be retrieved. The main techniques in relevance feedback include *query shifting* [1], *relevance estimation* [2] [3], *Bayesian inference* [4], *support vector machine active learning* [5]. All these methods are only adaptations of relevance feedback, instead of systematic concept learning in feature space. Furthermore, Once the user is done with a query and starts a new query, the meta knowledge gained by the systems with previous queries is lost. Meta knowledge is the experience of each query image with various users. This experience consists of the classification of each image into various classes (clusters), relevances (weights) of features and the number of times this image is selected as a query and marked as positive or negative.

Since real image databases experience retrievals from many users, it is possible to exploit previous retrieval experiences (meta knowledge) to learn and refine visual concepts. Some CBIR systems exploiting meta knowledge for concept learning and retrieval improvement have appeared recently [6][7][8]. In all these works, with retrieval experiences in conjunction with relevance feedback, the concept learning is improved, which helps to capture user's desired concept more precisely, and thus, future retrieval performance is improved. This process necessitates a model for the learning process; otherwise, it is only empirical.

We model the image distribution in database feature space as mixture Gaussian densities, which has been adopted for image database analysis by some researchers recently [9] [10]. In [9], the probabilistic image retrieval model based on mixture densities is analyzed, and short term and long term user feedbacks are combined by probabilistic inference. Barnard and Forsyth [10] organize images (with associated words) by a hierarchical model to help browsing and searching. Our work in this paper distinguishes from these approaches by focusing on semi-supervised mixture model fitting by integrating concept learning with multiple users' relevance feedbacks. In [9], only standard EM algorithm is used for model fitting in an unsupervised manner. In [10], some images are used to train the clustering directly (without user intervention); this training stage is unrealistic in real applications because the randomly selected training data set may not represent the image distribution of the entire database well.

The necessity of semi-supervised learning is due to the gap between numeric-oriented feature data and concepts understood by humans. For the pattern recognition tasks in real world, it is usually difficult and sometimes impossible to rely on a pure feature-based clustering (unsupervised clustering) algorithm to

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE

IEEE COMPUTER SOCIETY

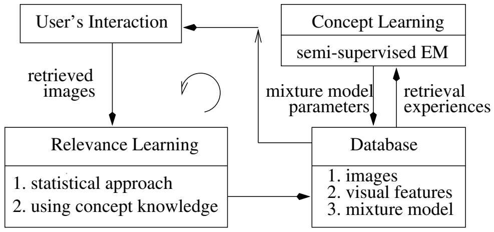


Figure 1: System diagram for concept learning using new semi-supervised EM algorithm (NSS-EM).

obtain satisfactory results.

The semi-supervised fuzzy c-means (SS-FCM) clustering methods [11][12] attempt to overcome this limitation when the labels of some of the data are already known; however, these approaches tend to be heuristic since they do not employ explicit models. Recently, some papers on semi-supervised learning based on mixture models have been published. Wu and Huang [13] integrate *multiple discriminant analysis* (MDA) with EM framework so that weak classifiers are boosted by exploring discriminant features in a self-supervised fashion. Another approach dealing with labeled and unlabeled data for Gaussian mixture models [14] is to modify the mixture log-likelihood function as the combination of two terms: the one for unlabeled data and the other for labeled data. These semi-supervised learning approaches assume that the labeled data belong to some specified classes (clusters). In reality, another kind of labeling information that "some data do NOT belong to some classes (clusters)" is also available in many applications, and it may also help the semi-supervised learning. Also these approaches assume that the labeling information is correct.

Figure 1 illustrates our system framework for concept learning using the NSS-EM algorithm on mixture model. Our contributions include: (a) We propose a new semi-supervised EM algorithm for fitting the mixture model, so that the concept learning is incrementally refined with increased retrieval experiences. Our approach can deal with the case of users' mislabeling during relevance feedback. (b) We also propose a new measurement for image dissimilarity by integrating concept learning with relevance feedback to improve retrieval performance.

2 Mixture model

We assume that the database image distribution in feature space is a c-component Gaussian mixture $\mathcal{C} = \{C_1, \ldots, C_c\}$ [9], whose probability density function is

$$f(x;\Psi) = \sum_{i=1}^{c} \pi_i f_i(x;\mu_i,\Sigma_i) \qquad (1)$$

where x is d-dimensional feature, $f_i(x)$ are component densities and π_i $(i = 1, 2, \ldots, c)$ are component proportions ($0 \le \pi_i \le 1$ and $\sum_{i=1}^{c} \pi_i = 1$). The component densities are specified by means μ_i and covariances Σ_i. Ψ is the vector containing all the unknown parameters i.e., $\Psi = \bigcup_{i=1}^{c}\{\pi_i,\mu_i,\Sigma_i\}$. For our image database system with N images, there are c concepts each of which is corresponding to one component. If c is known, the task of concept learning is accomplished by estimating the mixture model parameters Ψ.

Given a set of N independent and identical distribution (i.i.d.) samples $\mathcal{X}=\{x_1, x_2, \ldots, x_N\}$ (corresponding to the visual feature vectors of N database images) from model (1), the *maximum likelihood* (ML) estimation of the unknown parameter vectors $\{\pi_i,\mu_i,\Sigma_i\}$ can be obtained by the *Expectation-Maximization* (EM) approach. Let the associated binary component-indicator vectors for $\mathcal{X}$ be $\mathcal{Z}=\{z_1, z_2, \ldots, z_N\}$, where $z_j = (z_{j1}, \ldots, z_{jc})$ with $z_{ji} = \begin{cases} 1 & \text{if } x_j \text{ is from } i\text{th component} \\ 0 & \text{otherwise} \end{cases}$, for $j = 1, 2, \ldots, N$; $i = 1, 2, \ldots, c$. The complete data log-likelihood function is given by

$$\log L(\mathcal{X},\mathcal{Z};\Psi) = \prod_{j=1}^{N}\prod_{i=1}^{c} z_{ji}\log[\pi_i f_i(x_j;\mu_i,\Sigma_i)] \qquad (2)$$

The EM algorithm produces a sequence of estimations $\{\hat{\Psi}(k),\ k = 0, 1, 2, \ldots\}$ by proceeding iteratively in two steps (E-step and M-step) until some termination criterion is met.

- **E-step:** The conditional expectation of $\mathcal{Z}$ defined as $\tau_{ji} = \text{prob}_{\hat{\Psi}(k)}\{z_{ji} = 1|\mathcal{X}\}$ is derived as
$$\tau_{ji} = \frac{\pi_i f_i(x_j;\hat{\mu}_i,\hat{\Sigma}_i)}{\sum_{h=1}^{c}\pi_h f_h(x_j;\hat{\mu}_h,\hat{\Sigma}_h)} \qquad (3)$$
for $j = 1, 2, \ldots, N$; $i = 1, 2, \ldots, c$.

- **M-step:** Update the estimation of Ψ by
$$\hat{\Psi}(k+1) = \arg\max \Phi(\Psi;\hat{\Psi}(k))$$
The updated expression for component proportions is
$$\pi_i(t+1) = \frac{\sum_{j=1}^{N}\tau_{ji}}{N} \quad (i = 1, \ldots, c) \qquad (4)$$
For Gaussian mixture models, the expressions for means and covariances are
$$\hat{\mu}_i(t+1) = (\sum_{j=1}^{N}\tau_{ji})^{-1}\sum_{j=1}^{N}x_j\tau_{ji} \qquad (5)$$
and
$$\hat{\Sigma}_i(t+1) = (\sum_{j=1}^{N}\tau_{ji})^{-1}\sum_{j=1}^{N}(x_j-\hat{\mu}_i)(x_j-\hat{\mu}_i)^T\tau_{ji} \qquad (6)$$

3 New SS-EM algorithm

3.1 Problem

Since different users make a variety of queries and perceive visual content differently in their feedbacks,

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE

they may provide different sets of positive and negative labeling information, each of which is defined as a *retrieval experience* $\mathcal{E} = \{\mathcal{X}^+, \mathcal{X}^-\}$, where $\mathcal{X}^+ = \{x_1^+, x_2^+, \ldots, x_{N^+}^+\}$ are labeled as belonging to (positive for) a certain but unknown component (concept) while another portion of samples $\mathcal{X}^- = \{x_1^-, x_2^-, \ldots, x_{N^-}^-\}$ are labeled as NOT belonging to (negative for) that unknown component (concept). Note that x_i^+ ($i = 1, 2, \ldots, N^+$) and x_j^- ($j = 1, 2, \ldots, N^-$) are image visual feature vectors. We assume that query images in different query sessions are all from the image database with equal probability of being selected.

With more users' executions of their queries, the fitting of the mixture model corresponding to the human concepts may be continually refined over time, hence the concept learning is improved in the *long term*. Each time a retrieval experience is obtained by the system, the fitting of the mixture model should be updated based on the current clustering result.

3.2 Shor-term SS-EM algorithm

Now we consider the learning with only a single retrieval experience $\mathcal{E}$. We first assume that the component with regard to $\mathcal{E}$ is already known, and let it be the hth component with $h \in \{1, 2, \ldots, c\}$.

From the positive and negative labeling information, we already known some binary component-indicator vectors values such that $z_{ji} = \begin{cases} 1 & \text{if } i = h \\ 0 & \text{otherwise} \end{cases}$ for $i = 1, 2, \ldots, c$ and $j \in \mathcal{J}^+$, and $z_{jh} = 0$ for $j \in \mathcal{J}^-$. Thus we modify the complete data log-likelihood function (2) to be

$$\begin{aligned}\log L(\mathcal{X}, \mathcal{Z}; \Psi) = & \prod_{j \in \mathcal{J}^u} \prod_{i=1}^{c} z_{ji} \log\{\pi_i f_i(x_j; \theta_i)\} \\ & + \prod_{j \in \mathcal{J}^+} \log\{\pi_h f_h(x_j; \theta_h)\} \\ & + \prod_{j \in \mathcal{J}^-} \prod_{i=1, i \neq h}^{c} z_{ji} \log\{\pi_i f_i(x_j; \theta_i)\}\end{aligned}$$

In the above expression of the log-likelihood function, the first term is with regard to those unlabeled data $\mathcal{X}^u$, and it is in the same form as that in (2). The second term is for the positive labeled data $\mathcal{X}^+$ whose component-indicator vectors are already known so that there is no need to estimate them. For the negative labeled data $\mathcal{X}^-$, their component-indicator vectors are not totally available and only one of the elements in each vector is pre-determined to be zero. Thus the indicator vectors for $\mathcal{X}^-$ have to be estimated, as demonstrated by the third term.

Based on the modified log-likelihood function, we can implement EM algorithm to estimate parameters in a similar manner as introduced in Section 2. In E-step, for those pre-determined binary component-indicator vector elements, there is no need to estimate i.e., their estimation values are their "real" values, i.e., $\tau_{ji} = \hat{z}_{ji} = \begin{cases} 1 & \text{if } i = h \\ 0 & \text{otherwise} \end{cases}$ for $i = 1, 2, \ldots, c$ and $j \in \mathcal{J}^+$, and $\tau_{jh} = z_{jh} = 0$ for $j \in \mathcal{J}^-$. For other unknown component-indicator vector elements, we have to estimate them, and their estimation expression is given in (3). In M-step, the component proportion estimation is the same as that in (4). For Gaussian mixture components, the estimations for means and covariances are given in (5) and (6), respectively. The result can be derived by using Lagrangian multiplers method to optimize the modified likelihood function, and we do not include it in this paper due to space limitation.

From the above analysis, in the case where the cluster index h is already known, the EM algorithm for this semi-supervised learning task is the same as the procedure introduced in Section 2 except that some component-indicator vector elements are predetermined instead of to be estimated.

If the cluster index h is unknown, we can first implement unsupervised EM algorithm on the data, and obtain the clustering result represented by the component-indicator estimations. Based on this initial clustering result, h can be derived from the positive and negative labeling information using a probabilistic method such that

$$h = \arg \max_{i=1,2,\cdots,c} \mathcal{P}(i) \tag{7}$$

where $\mathcal{P}(i)$ is equal to

$$\begin{aligned}& \mathrm{prob}(x_1^+ \in C_i, \cdots, x_{N^+}^+ \in C_i, x_1^- \notin C_i, \cdots, x_{N^-}^- \notin C_i) \\ & = \{\textstyle\prod_{j=1}^{N^+} \mathrm{prob}(x_i^+ \in C_i)\}\{\prod_{j=1}^{N^-} \mathrm{prob}(x_j^- \notin C_i)\} \\ & = \{\textstyle\prod_{j \in \mathcal{J}^+} \tau_{ji}\}\{\prod_{j \in \mathcal{J}^-} (1 - \tau_{ji})\}\end{aligned}$$

for $i = 1, 2, \ldots, c$. It is obvious that the identification of h is dependent on the initial clustering result. One may argue that, if the initial clustering is not good enough, h may be mis-identified so that the NSS-EM algorithm is misled and its clustering result becomes worse. In the following text, we discuss this problem and give an efficient approach to overcome it.

3.3 Long-term SS-EM algorithm

The SS-EM method with a single retrieval experience can be extended to the optimization problem with multiple retrieval experiences.

If some images are randomly selected from a single component, it is possible that their covariance matrix is close to the covariance of the original component [12]. Unfortunately, the labeled images from a single retrieval are not sufficient; these images form a very small agglomeration in feature space compared to the size of a component. This requires that we refine the mixture model fitting continuously until enough experiences are accumulated.

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE

In reality, users making queries on an image database do not always have enough patience to correctly label all the images presented to them by the system. More importantly, different users may ascribe the same image to different concepts. It is obvious that for an image with multiple opinions on its cluster ascription, it should belong to the cluster according to the opinion supported by the majority of users.

In order to capture and accumulate previous users' retrieval experiences in the long-term history, we designate a *positive matrix* $P_{N\times c}$ and a *negative matrix* $Q_{N\times c}$ to represent this kind of knowledge. At the very beginning, when no retrieval has ever been executed on the system, P are Q are initialized to be zero matrices. After a retrieval experience, the elements $\left\{p_{j_1^+,h},\cdots,p_{j_{N+}^+,h}\right\}$ in P and the elements $\left\{q_{j_1^-,h},\cdots,q_{j_{N-}^-,h}\right\}$ in Q are increased by 1. Thus, the values of p_{jh} and q_{jh} represent to what extent people agree and disagree to ascribe an image j into the cluster h, respectively.

With the accumulated knowledge contained in P and Q, the component-indicator vector elements τ_{ji} derived in (3) can be modified as

$$\tilde{\tau}_{ji}=\begin{cases}\tau_{ji}+\frac{p_{ji}-q_{ji}}{\sum_{\iota=1}^{c}(p_{j\iota}+q_{j\iota})} & \text{if } p_{ji}>q_{ji}\\ 0 & \text{if } p_{ji}<q_{ji}\\ \tau_{ji} & \text{if } p_{ji}=q_{ji}\end{cases} \tag{8}$$

for $j=1,2,\ldots,N$ and $i=1,2,\ldots,c$. Then we normalize these modified component-indicator vectors so that $\sum_{i=1}^{c}\tilde{\tau}_{ji}=1$. This means that, based on numeric feature data, the component-indicator estimation is modified with labeling knowledge derived from users' retrieval experience. This modification step is inserted between E-step and M-step so that the concept learning result is closer to human's understanding. The user directed SS-EM algorithm is presented in Figure 2. EM-algorithm for model estimation is computationally intense. To avoid that clustering lags behind retrieval experience derivation in the system, we implement user directed SS-EM algorithm after every s ($s\geq 1$) retrieval experiences, where s is the update step.

This long-term NSS-EM algorithm is the extension of NSS-EM introduced in 3.2: the later only has the extra information from a single retrieval experience while the former has to deal with multiple retrieval experiences. Due to the knowledge accumulation mechanism of matrices P and Q, the learning improvement of the system is guaranteed even though it is possible that the concepts being sought are occasionally mis-identified by (7). For the same reason, in the case that users may mislabel images during relevance feedback, the system can still learn although the learning rate will be slower.

- Given the data $\mathcal{X}$, the number of clusters c, the number of images N. Initialize positive matrix $P_{N\times c}$ and negative matrix $Q_{N\times c}$ to be zero matrices.
- Implement standard EM algorithm on $\mathcal{X}$.
- Repeat
  1. New user executes retrieval.
  2. Derive a retrieval experience, update P and Q.
  3. E-step: Estimate component-indicator vectors $\mathcal{Z}$ by (3).
  4. Use P and Q to modify $\mathcal{Z}$ estimation by (8).
  5. M-step: Compute component proportions, means and covariances by (4), (5) and (6), respectively.
  6. If $\|L-L'\|\leq\delta$ (with δ being a tolerance limit), stop; else, go to 3 with $L'=L$.

Figure 2: Our new semi-supervised EM (NSS-EM) algorithm for long-term concept learning.

3.4 Improving retrieval performance

The knowledge of mixture model estimation derived from concept learning of the image database can help to improve retrieval performance. We use the component-indicator estimation τ_{ji} ($j=1,2,\ldots,N$, $i=1,2,\ldots,c$) to modify the image dissimilarity measurement for the initial search after a query is presented to the system, whose retrieval performance is the most important compared with the subsequent iterations. For the initial K nearest neighbor (K-NN) search, the Euclidean distance in the feature space from one database image x_j ($j=1,2,\ldots,N$) to the query x_q is defined as $D(x_q,x_j)$, which we modify as

$$D'(x_q,x_j)=e^{-\frac{\beta n}{N}}D(x_q,x_j)-\sum_{i=1}^{c}\tau_{ji}\tau_{qi} \tag{9}$$

where N is the database size and n is the number of retrieval experiences. The second term on the right side is with regard to concept learning knowledge, which is derived from

prob{Query q and Image j belong to same class}
$=\sum_{i=1}^{c}$ prob$\{q\in C_i$ and $j\in C_i\}$
$=\sum_{i=1}^{c}$ prob$\{q\in C_i\}$prob$\{j\in C_i\}$

As the concept learning is improved with the retrieval experiences increased, the second term in (9) should be given more credit. The parameter β is to make balance between these two terms. Note the speed of learning improvement depends on the database size N.

4 Experiments

For the experiments the image database, we simulate the process of a retrieval system for which queries are

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE

Figure 3: Sample images of the 12 classes in the database obtained from Corel stock photo library.

selected randomly among the images in the database. Let the number of images the user is presented at each relevance feedback iteration l be 20. To validate the clustering result $\mathcal{R} = \{\mathcal{R}_1, \ldots, \mathcal{R}_c\}$ from an algorithm, we compare $\mathcal{R}$ with the ground-truth mixture model $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_c\}$ by using statistical index. A pair of vectors $\{x_i, x_j\}$ are referred to as (I) BS if both vectors belong to the same component in $\mathcal{C}$ and to the same cluster in $\mathcal{R}$, (II) SD if both vectors belong to the same component in $\mathcal{C}$ and to different clusters in $\mathcal{R}$, (III) DS if both vectors belong to different components in Let λ_1, λ_2 and λ_3 be the number of BS, SD and DS pairs of vectors of $\mathcal{X}$, respectively. The *Jaccard coefficient* [15] is defined as $JC = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}$ and we use it to evaluate clustering result.

We construct an image database with 1200 images, which are selected from Corel stock photo library and divided into 12 classes. These classes are corresponding to the CDs (series number) in the library including *Mayan* & *Aztec Ruins* (33), *Horses* (113), *Owls* (75), *Sunrises* & *Sunsets* (1), *North American Wildflowers* (127), *Ski Scenes* (61, 62), *Coasts* (5), *Auto Racing* (21), *Firework Photography* (73), *Divers* & *Diving* (156), *Land of the Pyramids* (161) and *Lions* (105). We remove some images from these CDs since they do not have good visual features to represent the corresponding concepts, and we add some images from other CDs to some of the 12 classes. Figure 3 shows sample images for all of the 12 concepts. We use texture and color features to represent images. The texture features are derived from 16 Gabor filters We extract means and standard deviations from the three channels in HSV color space. Thus, each image is represented by 22 features.

We implement our concept learning approach on this database with $c = 12$, $N = 1200$, $s = 50$ and $\beta = 100$. Initially, the Jaccard coefficient by standard EM algorithm is 47.1%. Figure 4 shows that the average Jaccard coefficient is increased with increased retrieval experiences in the long term. Compared with the synthetic data, the concept learning improvement is slower due to the facts that there are more components for real data and components need more data sample

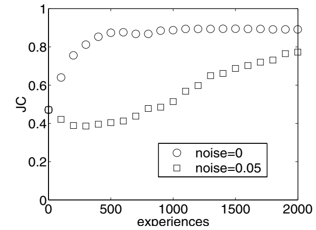
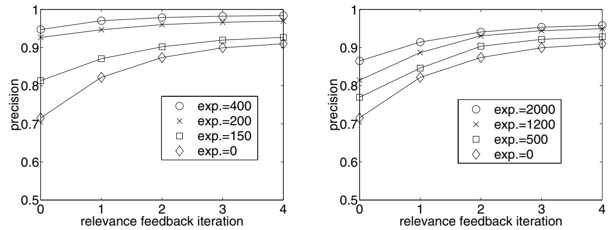


Figure 4: Real data: concept learning is improved with increased retrieval experiences with and without noise.



(a) no noise

(b) noise ratio $\nu = 0.05$

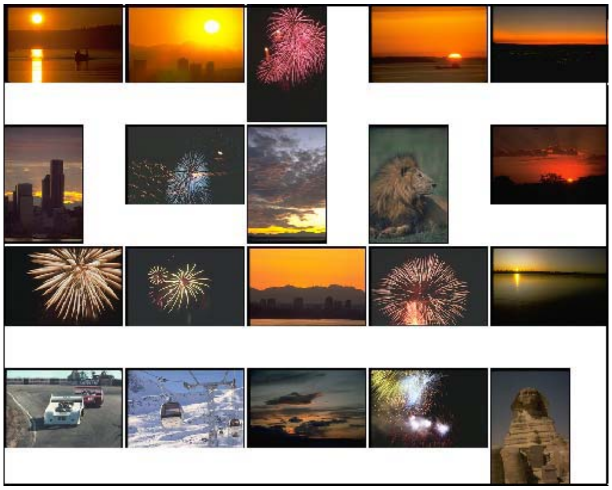
Figure 5: Real data: retrieval performance with various amounts of retrieval experiences.

for higher dimensional features. When the mislabeling noise ratio ν is 0.05, the learning also converges in the long term although the improvement speed is slower. Note the probability that the user mislabels any image at a single relevance feedback iteration is prob(error) $= 1 - (1-\nu)^l$ with l being the number of images presented to the user at each iteration. Thus, when $\nu = 0.05$ and $l = 20$, prob(error) is 0.64, which is quite high.

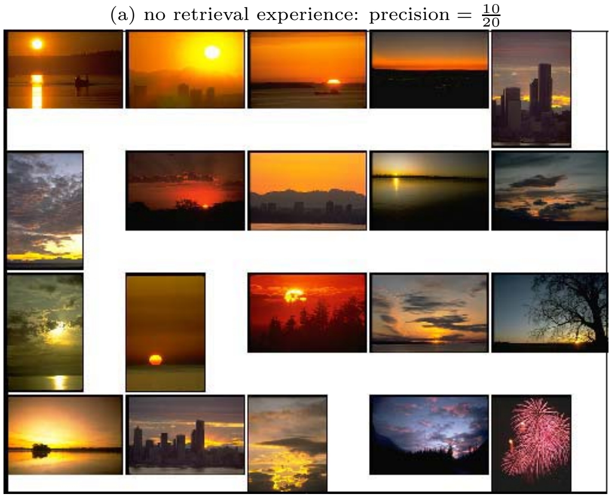
Figure 5 presents the retrieval performances with different amounts of retrieval experiences with and without labeling noise. The *retrieval precision* is defined as the percentage of positive retrievals out of the total retrievals. We select an image in this database as the query, implement our retrieval strategy, and repeat this experiment by changing query until each of the 1200 images has been selected as query. Then we calculate the average precision at each iteration. With increased retrieval experiences, the average precision is improved, especially at initial K-NN search iteration. This has deep significance for retrieval performance in practical applications since users usually don't have enough patience to repeat relevance feedback iterations to search the images.

Figure 6 shows two different retrieval results with the same query image with different retrieval experiences where there is not labeling noise. In (a), there is no retrieval experience, and K-NN search only yields 10 out of 20 *sunset* images (row 1: image 1, 2, 4, 5; row 2: 1, 3, 5; row 3: 3, 5; row 4: image 3); In (b), after 200 retrieval experiences, 19 *sunset* images are presented (except the last one) by our approach.

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE



(a) no retrieval experience: precision = $\frac{10}{20}$



(b) after 200 retrieval experiences: precision = $\frac{19}{20}$

Figure 6: Retrieval results for the same query (the first image) with different retrieval experiences. The user is looking for *sunset* images.

5 Discussions and conclusions

This paper proposed mixture model as the image distribution in feature space of the image database. Based on this model, a new semi-supervised EM (NSS-EM) algorithm is presented for concept learning, which can deal with the case of users' mislabeling during relevance feedback. We also integrate concept learning result with relevance feedback to improve retrieval performance. The experiments shows the efficacy of our concept learning approach. Note that since the number of visual features used in CBIR is usually large, the computation load of EM algorithm is heavy. Thus, it is desired to reduce the dimensionality in feature space during EM iterations.

Acknowledgements

This research was supported by grants F49620-02-1-0315 and IIS-0114036. The contents of the information do not necessarily reflect the position or the policy of the US Government.

References

[1] J. J. Rocchio and G. Salton, "Information search optimization and iterative retrieval techniques," *Proc. AFIPS*, vol. 27, pp. 293–305, 1965.

[2] Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8 (5), pp. 644–655, Sept. 1998.

[3] J. Peng, B. Bhanu, and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval," *CVIU*, vol. 75, no. 1-2, pp. 150–164, July-August 1999.

[4] I. J. Cox, M. L. Matt, T. P. Minka, T. V. Papathoma, and P. N. Yianilos, "The Bayesian image retrieval system, *PicHunter*: theory, implementation, and psychophysical experiments," *IEEE Trans. on Image Processing*, vol. 9 (1), pp. 20–37, Jan. 2000.

[5] S. Tong and E. Chang, "Supporting vector machine active learning for image retrieval," *Proc. ACM int. conf. on Multimedia*, pp. 107–118, 2001.

[6] P. Yin, B. Bhanu, K. Chang, and A. Dong, "Improving retrieval performance by long-term relevance information," *Proc. ICPR*, vol. III, pp. 533–536, Aug. 2002.

[7] B. Bhanu and A. Dong, "Concept learning with fuzzy clustering and relevance feedback," *Engineering Applications of AI*, vol. 15, pp. 123–138, Sept. 2002.

[8] J. Fournier and M. Cord, "Long-term similarity learning in content-based image retrieval," *Proc. ICIP*, vol. 1, pp. 441–444, Sept. 2002.

[9] N. Vasconcelos, *Bayesian Models for Visual Information Retrieval*, Ph.D thesis, MIT, 2000.

[10] K. Barnard and D. Forsyth, "Learning the semantics of words and pictures," *Proc. IEEE Int. Conference on Computer Vision*, vol. 2, pp. 408–415, 2001.

[11] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Trans. on SMC*, vol. 27 (5), pp. 787–795, Oct. 1997.

[12] A. M. Bensaid, L. O. Hall, J. C. Bezdek, and L. P. Clarke, "Partially supervised clustering for image segmentation," *Pattern Recognition*, vol. 29 (5), pp. 859–871, May 1996.

[13] Y. Wu and T. S. Huang, "Towards self-exploring discriminating features for visual learning," *Engineering Applications of AI*, vol. 15, pp. 139–150, Sept. 2002.

[14] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using EM," *Machine Learning*, vol. 39, pp. 103–134, 1999.

[15] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 1999.

1063-6919/03 $17.00 © 2003 IEEE