

# Active Concept Learning for Image Retrieval in Dynamic Databases

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

## Abstract

*Concept learning in content-based image retrieval (CBIR) systems is a challenging task. This paper presents an active concept learning approach based on mixture model to deal with the two basic aspects of a database system: changing (image insertion or removal) nature of a database and user queries. To achieve concept learning, we develop a novel model selection method based on Bayesian analysis that evaluates the consistency of hypothesized models with the available information. The analysis of exploitation vs. exploration in the search space helps to find optimal model efficiently. Experimental results on Corel database show the efficacy of our approach.*

## 1 Introduction

Visual concept learning is challenging in content-based image retrieval (CBIR) because there exists a big gap between low-level visual features and high-level human concepts. The mechanism of relevance feedback allows a user to label retrieved images as positive or negative depending upon whether or not the images contain the concept that the user is seeking. The system can dynamically update the relevance of images to be retrieved by using the techniques such as *query shifting* [1], *relevance estimation* [2] [3] and *Bayesian inference* [4]. However, these methods are not systematic concept learning methods, but results in only the adaptations of relevance feedback. Once the user is done with a query and starts a new query, the knowledge gained by the system with this iteration is lost.

Recently, some CBIR systems exploiting meta knowledge for retrieval performance improvement have appeared [5] [6]. These systems can improve retrieval performance with retrieval experiences in conjunction with relevance feedback. However, none of them presents a system model that can be used for systematic concept learning.

In this paper, we model the database image distribution in feature space as Gaussian mixture [7], and our task is to estimate this model (called *model fitting*) to achieve concept learning. Model fitting task for mixture model generally includes two steps: (1) model selection, i.e., estimate the number of components; (2) parameter estimation. Model

selection is the prerequisite for parameter estimation, and it is a more challenging and unsolved task.

In unsupervised learning literature, if the true number of components in mixture model  $c$  is known to be in the range  $\{c_{min}, \dots, c_{max}\}$ , it can be generally selected according to some criterion function by

$$\hat{c} = \arg \max_c \{\mathcal{F}(\hat{\Psi}_c, c), c \in \{c_{min}, \dots, c_{max}\}\} \quad (1)$$

where  $\hat{\Psi}_c$  is the mixture parameter estimation when the model is assumed to contain  $c$  components, and the criterion function  $\mathcal{F}(\hat{\Psi}_c, c)$  usually consists of two terms as

$$\mathcal{F}(\hat{\Psi}_c, c) = -\log p(x | \hat{\Psi}_c) + \mathcal{P}(c) \quad (2)$$

The first term is the log-likelihood of the data for the model, and the second term is to penalize higher values of  $c$ . Many researchers attempted to select model by using the criterion functions with this form such as *minimum message length criterion* (MML) [8], *Bayesian inference criterion* (BIC) [9] and *Akaike's information criterion* (AIC) [10]. These methods are proved to be effective for some data in unsupervised manner; however, they are often frustrated by the fact that optimal values of criterion function do not necessarily lead to better model selection or data partition.

Some semi-supervised learning papers [11][12] have appeared in an attempt to improve the model fitting. The most lethal weakness of these approaches is that they are only for parameter estimation and do not deal with model selection. Thus, their practical applications are limited.

The concept learning of our system with regard to the mixture model necessitates model selection, i.e., accessing the number of components. Model selection is not considered in Vasconcelos' approach [7], which is one of the most famous approaches adopting mixture model into multimedia retrieval research. The integration with relevance feedback mechanism of our system makes it possible to carry out model selection in a semi-supervised manner, which is still an unexplored research topic to date.

Figure 1 illustrates our system framework with active concept learning and concept knowledge transplantation. The *contributions* of this paper is: a new active learning approach for mixture model fitting, including a model selection method and a user directed semi-supervised EM (SS-EM) algorithm. Besides, the retrieval experiences derived

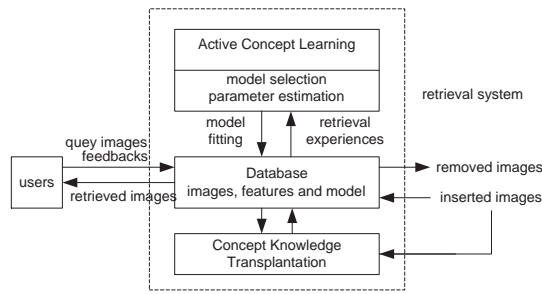


Figure 1: System diagram with active concept learning and concept knowledge transplantation for dynamic image databases.

from previous users' feedback are used to achieve concept learning, which may help to improve future retrieval performance [13]. A concept knowledge transplantation approach can deal with the cases of image insertion and query images being outside the database efficiently [14].

## 2 Motivation

We assume that the database image distribution in feature space is a  $c$ -component Gaussian mixture  $\mathcal{C} = \{C_1, \dots, C_c\}$  [7], whose probability density function is

$$f(x; \Psi_c) = \sum_{i=1}^c \pi_i f_i(x; \mu_i, \Sigma_i) \quad (3)$$

where  $x$  is  $d$ -dimensional feature,  $f_i(x)$  are component densities and  $\pi_i$  ( $i = 1, 2, \dots, c$ ) are component proportions ( $0 \leq \pi_i \leq 1$  and  $\sum_{i=1}^c \pi_i = 1$ ). The component densities are specified by means  $\mu_i$  and covariances  $\Sigma_i$ .  $\Psi_c$  is the vector containing all the unknown parameters i.e.,  $\Psi_c = \bigcup_{i=1}^c \{\pi_i, \mu_i, \Sigma_i\}$ . 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  $\Psi_c$ .

Given a set of  $N$  i.i.d samples  $\mathcal{X} = \{x_1, x_2, \dots, x_N\}$  (corresponding to the visual feature vectors of  $N$  database images) from model (3), let the associated binary component-indicator vectors for  $\mathcal{X}$  be  $\mathcal{Z} = \{z_1, z_2, \dots, z_N\}$ , where  $z_j = (z_{j1}, \dots, z_{jc})$  with  $z_{ji} = \begin{cases} 1 & \text{if } x_j \text{ is from } i\text{th component} \\ 0 & \text{otherwise} \end{cases}$ , for  $j = 1, 2, \dots, N$ ;  $i = 1, 2, \dots, c$ . The EM algorithm can be used to estimate all these model parameters (see [15]).

### 2.1 System events

An image retrieval system with relevance feedback mechanism may encounter two kinds of events at any time during the long-term running: users' queries and database changes (i.e., image insertion or removal). We model the occurrences of these two events as *Poisson* random processes, whose distributions are  $P[\mathcal{N}(t) = k] = \frac{(\lambda_i t)^k}{k!} e^{-\lambda_i t}$

( $k = 0, 1, \dots$ , and  $t$  is the system running time) with  $i = 1$  and 2 respectively. The ratio of the two distribution parameters  $r = \frac{\lambda_1}{\lambda_2}$  specifies the relative occurrence rate of these two events.

Since different users make a variety of queries and perceive visual content differently, 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  $\mathcal{X}^+ = \{x_1^+, x_2^+, \dots, x_{N^+}^+\}$  are labeled as belonging to (positive for) a certain but unknown component (concept) while another portion of samples  $\mathcal{X}^- = \{x_1^-, x_2^-, \dots, x_{N^-}^-\}$  are labeled as NOT belonging to (negative for) that unknown component (concept).

### 2.2 Active model fitting

To accomplish the task of concept learning for the image database, we should do model fitting over the data whose population is continually changing due to dynamic nature of the database. Since retrieval experiences contains positive and negative labeling information, they can help the model fitting; thus, our model fitting is in a semi-supervised manner. Moreover, the fitting should be updated actively with the new retrieval experiences being obtained and images being added to or removed from the database.

Assume that the true number of components in the mixture model is known to be in the range of  $\{c_{min}, \dots, c_{max}\}$ . At time  $t$  of the system running, let the system has obtained  $\iota(t)$  retrieval experiences and the current image data is  $\mathcal{X}(t)$ , our task is to find an optimal mixture model

$$\hat{\mathcal{M}} = \arg \max_{\mathcal{M} \in \Omega} \text{prob}\{\mathcal{M} | \mathcal{X}(t), \mathcal{E}_1, \dots, \mathcal{E}_{\iota(t)}\} \quad (4)$$

where  $\Omega$  is the search space containing all possible models in the range of  $\{c_{min}, \dots, c_{max}\}$ . The model  $\mathcal{M}$  is specified by the number of components  $c$  and the parameter estimation  $\Psi_c$ . Note that when  $\iota(t) = 0$ , the problem becomes the traditional unsupervised mixture model fitting task in [8] [9] [10]. For convenience, we omit "(t)" in the notations for  $\mathcal{X}(t)$  and  $\iota(t)$  in the following texts.

## 3 Technical approach

In this section, we first introduce our previous work [13] [14], and then present the model fitting algorithm.

### 3.1 Concept learning and transplantation

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. With the accumulated knowledge contained in  $P$  and  $Q$ , the component-indicator vector elements can be modified. This modification step is inserted between E-step and M-step so that the concept learning result is closer to human understanding. To avoid clustering that lags behind retrieval experience derivation, we implement user directed SS-EM algorithm after every  $s$  ( $s \geq 1$ ) retrieval experiences, where  $s$  is the update step.

The knowledge of mixture model estimation derived from concept learning of the image database can help to improve retrieval performance. For the initial  $K$  nearest neighbor ( $K$ -NN) search, the Euclidean distance in the feature space from one database image  $x_j$  ( $j = 1, 2, \dots, N$ ) to the query  $x_q$  is proposed to contain two terms: feature-based term and concept-knowledge-based term. As the concept learning is improved with the retrieval experiences increased, the second term should be given more credit.

When a new image is inserted or an image is removed from the database, the matrices  $P_{N \times c}$  and  $Q_{N \times c}$  are modified based on the model estimation. If the query image does not belong to the database, the system extracts its visual features, computes component-indicators, and implements  $K$ -NN search using the distance measurement mentioned in the last paragraph. Compared with the traditional  $K$ -NN search that is solely based on visual feature Euclidean distance measurement, this approach yields better retrieval performance since concept knowledge is adopted.

### 3.2 Model selection

If we initially select the model with the true number of components, the fitting (only parameter estimation in this case) with increased retrieval experiences will lead to the model estimation which is very close to groundtruth model. This can be achieved by our user directed SS-EM algorithm in Section 3.1. On the other hand, if the initial assumption on the number of components is not correct, the fitting over time will never yield a good model estimation (at least it will be worse than the case based on the true number of components after enough time).

The previous retrieval experiences obtained by the system can be used in selecting the optimal model. An obvious way to achieve this idea is: the system keep all the candidate model fittings with  $\{\Psi_{c_{min}}, \dots, \Psi_{c_{max}}\}$ . Whenever a new retrieval experience is obtained, the system update all of these fittings by our SS-EM algorithm. At any time, the optimal model is the one which is most consistent with the feature data and the positive and negative information contained in the previous retrieval experiences.

However, when the range of  $\{c_{min}, \dots, c_{max}\}$  is large, and the computational load of EM algorithm is heavy, the updating for all the candidate models will be very slow; thus, the model fitting may lag far behind the available retrieval experience, i.e., the system cannot digest the retrieval experiences on time.

To overcome this computational load problem, we propose an adaptive model selection approach. Assume that the system allows the computation of EM algorithms for  $\mathcal{K}$  ( $1 \leq \mathcal{K} < c_{max} - c_{min} + 1$ ) models at the same time ( $\mathcal{K}$  is determined by the computation capability of the system computer), we select  $\mathcal{K}$  out of  $c_{max} - c_{min} + 1$  models based on their consistencies with the data and the previous retrieval experiences. The consistency of a model  $\mathcal{M}_c$  with

$\iota$  retrieval experiences can be measured by its probability given the data and these retrieval experiences:

$$\begin{aligned} & \text{prob}(\mathcal{M}_c | \mathcal{E}_1, \dots, \mathcal{E}_\iota; \mathcal{X}) \\ &= \frac{\text{prob}(\mathcal{E}_1, \dots, \mathcal{E}_\iota | \mathcal{M}_c; \mathcal{X}) \text{prob}(\mathcal{M}_c | \mathcal{X})}{\text{prob}(\mathcal{E}_1, \dots, \mathcal{E}_\iota | \mathcal{X})} \\ &\propto \text{prob}(\mathcal{E}_1, \dots, \mathcal{E}_\iota | \mathcal{M}_c; \mathcal{X}) \text{prob}(\mathcal{M}_c | \mathcal{X}) \end{aligned} \quad (5)$$

$$= \left\{ \prod_{i=1}^{\iota} \text{prob}(\mathcal{E}_i | \mathcal{M}_c; \mathcal{X}) \right\} \text{prob}(\mathcal{M}_c | \mathcal{X}) \quad (6)$$

The condition for independency required for the deduction from (5) to (6) will be discussed later in this section. For a single retrieval experience  $\mathcal{E}$ , we have

$$\begin{aligned} & \text{prob}(\mathcal{E} | \mathcal{M}_c; \mathcal{X}) \\ &= \sum_{i=1}^c \left\{ \prod_{j=1}^{N^+} \text{prob}(x_j^+ \in C_i) \prod_{j=1}^{N^-} \text{prob}(x_j^- \notin C_i) \right\} \\ &= \sum_{i=1}^c \left\{ \prod_{j \in \mathcal{J}^+} \tau_{ji} \prod_{j \in \mathcal{J}^-} (1 - \tau_{ji}) \right\} \end{aligned} \quad (7)$$

where  $\mathcal{J}^+ = \{j_1^+, j_2^+, \dots, j_{N^+}^+\}$  and  $\mathcal{J}^- = \{j_1^-, j_2^-, \dots, j_{N^-}^-\}$  are the indices for the image in  $\mathcal{X}^+$  and  $\mathcal{X}^-$  respectively, and  $\tau_{ji}$  is conditional expectation that the image  $x_j$  belongs to  $C_i$ .

The term  $\text{prob}(\mathcal{M}_c | \mathcal{X})$  appeared in (6) can be derived by the likelihood function in the unsupervised learning proposed in [16]

$$\mathcal{L}(\Psi_c, \mathcal{X}) = \log p(\mathcal{X} | \Psi_c) + \log \sum_{i=1}^c \sum_{j=1}^N \tau_{ji} \log \tau_{ji} \quad (8)$$

where the second term is the estimated entropy used to penalize the model for its complexity (high value of  $c$ ). Since there are  $c_{max} - c_{min} + 1$  candidate models, we approximate the models' probabilities as

$$\text{prob}(\mathcal{M}_c | \mathcal{X}) \simeq \frac{\exp\{\mathcal{L}(\Psi_c, \mathcal{X}) / \mathcal{L}_{max}\}}{\sum_{k=c_{min}}^{c_{max}} \exp\{\mathcal{L}(\Psi_k, \mathcal{X}) / \mathcal{L}_{max}\}} \quad (9)$$

where  $\mathcal{L}_{max} = \max_{c \in \{c_{min}, \dots, c_{max}\}} |\mathcal{L}(\Psi_c, \mathcal{X})|$ , which is used to normalize the likelihood functions.

We define  $\mathcal{M}_c$ 's consistency with the data  $\mathcal{X}$  and  $\iota$  retrieval experiences as the log-based value of (6)

$$\begin{aligned} & \text{cons}(\mathcal{M}_c; \mathcal{E}_1, \dots, \mathcal{E}_\iota, \mathcal{X}) \\ &= \sum_{i=1}^{\iota} \log \text{prob}(\mathcal{E}_i | \mathcal{M}_c; \mathcal{X}) + \log \text{prob}(\mathcal{M}_c | \mathcal{X}) \end{aligned} \quad (10)$$

which appropriately represents the probability  $\text{prob}(\mathcal{M}_c | \mathcal{E}_1, \dots, \mathcal{E}_\iota; \mathcal{X})$ . By (7), (8) and (9), the value of (10) can be computed. Note that when  $\iota = 0$  (no retrieval experience), the consistency only depends on the second term of (10), which is derived from the model likelihood function for unsupervised learning. This is the case at the very beginning of the system running. With retrieval experiences increased, the unsupervised criterion represented by the second term of (10) exerts less influence while the accumulated retrieval experiences plays a more important role on the consistency measurement.

The current optimal model is

$$\hat{c}_{opt} = \arg \max_{c \in \{c_{min}, \dots, c_{max}\}} \text{cons}(\mathcal{M}_c; \mathcal{E}_1, \dots, \mathcal{E}_t, \mathcal{X}) \quad (11)$$

The selection of  $\mathcal{K}$  out of  $c_{max} - c_{min} + 1$  models for SS-EM updating should be based on the models' current consistencies. However, we do not directly choose the  $\mathcal{K}$  models which have highest consistency values; instead, we give opportunity of being selected to each of the  $c_{max} - c_{min} + 1$  models, whose probability is

$$\text{prob}(\mathcal{M}_c) = \frac{\exp\{\eta t \cdot \text{cons}(\mathcal{M}_c; \mathcal{E}_1, \dots, \mathcal{E}_t, \mathcal{X})\}}{\sum_{i=c_{min}}^{c_{max}} \exp\{\eta t \cdot \text{cons}(\mathcal{M}_i; \mathcal{E}_1, \dots, \mathcal{E}_t, \mathcal{X})\}} \quad (12)$$

where  $\eta$  is a parameter to be discussed later. This equation implies that models with higher consistency values have better chances of being selected. Thus, the search direction for optimal model tends to be towards the models which have good consistencies with the feature data and the accumulated retrieval experiences. In other words, the optimal model search should *exploit* the current model estimation. On the other hand, due to the possibility that the model estimation based on the true number of components may not be good (especially at the early stage), we still give chances to the models with lower consistency values, i.e., we want to *explore* the whole search space. The relationship between *exploitation* and *exploration* changes with time  $t$  increased: at the early stage, *exploration* is more important so that all of the candidate models have chances to be selected for updating with retrieval experiences. With more retrieval experiences improving these model estimations, *exploitation* becomes the main concern since the model with good consistency is very likely to be the optimal model. The term  $\eta t$  in (12) assigns the probabilities for models being selected for SS-EM updating in the way that reflects this *exploitation* & *exploration* relationship.

The independency condition for the deduction from (5) to (6) may not be satisfied if all the obtained retrieval experiences are used to measure model consistency. For example, if two retrieval experiences with the same image sets (they are not independent) are both counted in the computation of (6) to measure model consistency, the positive and negative information contained in these two experiences may be overly used; thus, the derived consistency may not correctly reflect whether or not the model is good. To avoid this problem, we select retrieval experiences for each pair of which have no overlap, i.e., no common images. However, this condition is too strict for the system to obtain enough retrieval experiences for consistency measurement. Thus, we allow each pair of selected retrieval experiences to share a small percentage of common images. We set an independency threshold  $\varrho$  such that the maximal allowed number of common images between two retrieval experiences is  $\varrho l$  ( $l$  is the number of images that are presented to the user at each relevance feedback iteration). Thus, enough retrieval experiences can be accumulated and the condition for in-

dependency is not deteriorated. Another advantage of this experience selection method is that it alleviates the load of computation for consistency; otherwise, the retrieval experiences used for consistency computation may tend towards infinity with time  $t$  increased.

For consistency computation, there is an overflow problem that cannot be ignored. For (7) whose computation involves the multiplications of many probabilities, if there exist some probabilities whose values are zero or very small, it may lead that  $\text{prob}(\mathcal{E} | \mathcal{M}_c; \mathcal{X})$  is zero. This will cause overflow problem for computing log-based value in (10). To avoid this overflow problem, we adopt a small value threshold  $\varepsilon$  to prevent the probabilities from being too small in the way that

$$\tau_{ji} = \frac{\tau_{ji} + \varepsilon}{(1 + c\varepsilon)} \quad (13)$$

for  $j = 1, 2, \dots, N$ ;  $i = 1, 2, \dots, c$ . We use  $\varepsilon = 0.01$  in this paper.

From (10), we observe that consistency value becomes smaller with  $t$  increased (more retrieval experiences). Although this does not influence the ranks of the model consistencies at time  $t$ , it influences the probabilities for the models being selected for updating as shown in (12). Set the summation of the image numbers in all the  $t$  retrieval experiences is  $\mathcal{S}$ , the model consistency in (10) can be regarded as the summation of  $\mathcal{S} + 1$  log-based probabilities. Thus, we normalize the consistency by dividing the expression in (10) with  $\mathcal{S} + 1$ . Figure 2 summarizes our algorithm for the active model selection.

## 4 Experiments

We use the database containing 1,200 images and 12 classes [13]. To validate the concept learning result from the algorithm, we compare the clustering result  $\mathcal{R} = \{\mathcal{R}_1, \dots, \mathcal{R}_c\}$  with the groundtruth mixture model  $\mathcal{C} = \{C_1, \dots, C_c\}$  by using *Jaccard coefficient* ( $JC$ ) [13]. One of the advantages of Jaccard coefficient is that it can evaluate a clustering result whose cluster number is not necessarily the true component number.

We randomly select 800 out of the 1,200 images as the initial database images  $\mathcal{X}(0)$ , i.e.,  $N = 800$ , and insert the other 400 images during the system's running. We set the system running time as  $t = 0, 1, 2, \dots$ ; at each  $t$ , one of the events happens: user query or image insertion. This is a random process derived from the two events' poisson random processes with their relative occurrence rate  $r$  ( $r = 16$ ). Our active concept learning approach with model selection on the database is implemented with  $c_{min} = 10$ ,  $c_{max} = 13$ , the number of models selected for SS-EM updating  $\mathcal{K} = 1$ , the independency threshold  $\varrho = 0.25$ , the update step is  $s = 10$ .

At the initial stage of the system running, since the models' consistencies with the limited retrieval experiences may not have convincing statistical significance, we give all the

- $t = 0$ .
- Given the data  $\mathcal{X}(t)$ , the range for the number of components  $\{c_{min}, \dots, c_{max}\}$ .
- Implement standard EM algorithms on  $\mathcal{X}(t)$  with regard to  $c_{min}, \dots, c_{max}$ , and get parameter estimations  $\Psi_{c_{min}}, \dots, \Psi_{c_{max}}$ .
- Repeat ( $t = t + 1$ )

(a) When: new user executes retrieval and get retrieval experience  $\mathcal{E}_t$ :

1.  $\mathcal{X}(t) \leftarrow \mathcal{X}(t - 1)$ .
2. Count  $\mathcal{E}_t$  for consistency measurement if the condition for independence with threshold  $\rho$  is satisfied; otherwise, go to “Repeat”.
3. Compute consistencies for  $\mathcal{M}_{c_{min}}, \dots, \mathcal{M}_{c_{max}}$  by (10), and normalize by dividing them with  $S + 1$ .
4. Select  $\mathcal{M}_{\hat{c}_{opt}}$  as current optimal model by (11).
5. Randomly select  $\mathcal{K}$  models for SS-EM updating by the probabilities given in (12) and get  $\mathcal{K}$  updated parameter estimations. Prevent component-indicator elements to be 0 by (13).

(b) When Image  $I^+$  is added:  $\mathcal{X}(t) \leftarrow \mathcal{X}(t - 1) \cup I^+$ .

(c) When Image  $I^-$  removed:  $\mathcal{X}(t) \leftarrow \mathcal{X}(t - 1) - I^-$ .

Figure 2: Active model selection algorithm.

models equal probabilities to be selected for SS-EM updating, instead of assigning the probabilities by (12). This conservative strategy refrains the possibility that the model with the true number of components is discarded for future updating because of its low consistency initially. We set this initial stage to be  $t \in [0, 200]$ . After the initial stage, the probabilities are assigned by (12), where we set  $\eta = 0.5$ .

Figure 3 shows the active model selection process. From Figure 3(a), we observe the oscillations of the consistency values for all the models from  $t = 0$  to  $t = 80$  due to the reason that the retrieval experiences are limited initially. Then all the consistency curves become relatively smooth, and the consistency of  $\mathcal{M}_{12}$  gradually increases with a slow speed (Figure 3(b)). This means that the model with true number of components fits the obtained retrieval experiences better. Another observation is that  $\mathcal{M}_{12}$  has the highest consistency after  $t = 140$ ; thus,  $\mathcal{M}_{12}$  is always regarded as the current optimal model by the system (Figure 3(c)). After  $t = 200$ ,  $\mathcal{M}_{12}$  has the dominant probability to be selected for SS-EM updating.

Figure 3(d) shows that the model fitting is improved during the process of the active concept learning as more retrieval experiences are obtained. Note that in spite of the overall increasing trend for the  $JC$  curve, the  $JC$  value occasionally decreases after SS-EM updating, which is caused by the reason that the concept (component) sought by the

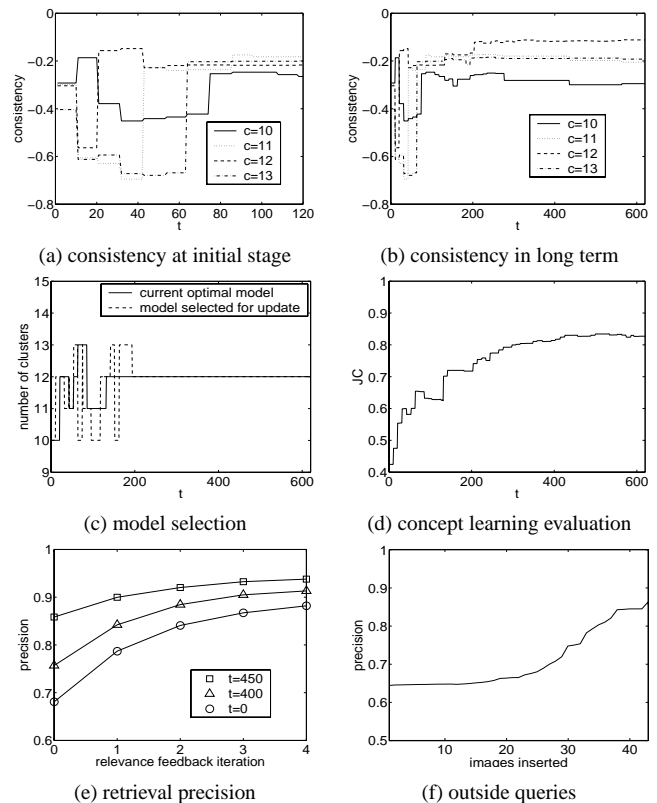


Figure 3: Active concept learning process.

user may not be correctly identified. Figure 3(e) presents the improvement in retrieval performance with increased  $t$ . At time  $t$ , 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 database images has been selected as a query. Then we compute 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.

To demonstrate the efficacy of our concept knowledge transplantation method, at the moment  $t$  when a new image is inserted during active concept learning process, we use the rest of the images outside the database (i.e., the 1,200 images  $-\mathcal{X}(t)$ ) as queries and simulate the relevance feedback iterations by using the knowledge transplantation method. In Figure 3(f), the average retrieval precision at iteration 0 increases with more images being inserted to the system due to the reason that concept learning is improved since more retrieval experiences are obtained in the process. Note that another factor that improves the precision is that with more images being inserted, there are more relevant images within each class; thus, the probability that relevant images are selected increases.

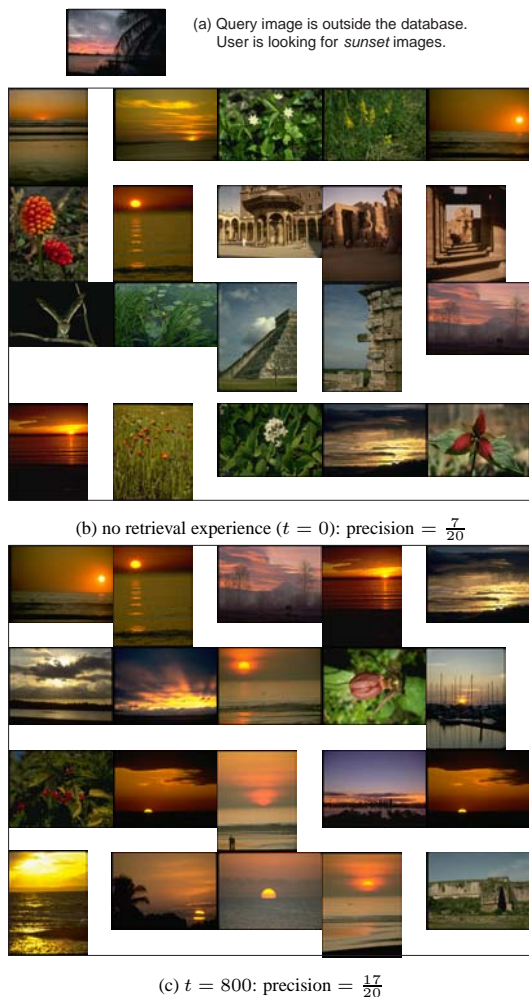


Figure 4: Retrieval precision is improved as retrieval experiences are increased.

Figure 4 shows two different retrieval results with the same query image in (a) (outside the database) at different running time. In (b), there is no retrieval experience, and  $K$ -NN search only yields 7 out of 20 *sunset* images (row 1: 1, 2, 5; row 2: 2; row 3: 5; row 4: 1 and 4); In (c), when  $t = 800$ , 17 *sunset* images are presented (except the 4th image on row 2, the 1st image on row 3 and the last image on row 4) by our concept transplantation approach.

## 5 Conclusions

This paper proposed an active concept learning approach for dynamic image databases. The model selection in a semi-supervised manner, which is still an unexplored topic to date, is achieved with the help of user directed semi-supervised EM algorithm, and the *exploitation* vs. *exploration* analysis makes the model selection efficient. Our experimental results show that the concept learning results can help to improve retrieval performance.

**Acknowledgements:** This research was supported in

part 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. Papatoma, 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] 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.
- [6] B. Bhanu and A. Dong, "Concept learning with fuzzy clustering and relevance feedback," *Engineering Applications of Artificial Intelligence*, vol. 15, pp. 123–138, April 2002.
- [7] N. Vasconcelos, *Bayesian Models for Visual Information Retrieval*, Ph.D thesis, MIT, 2000.
- [8] M. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. PAMI*, vol. 24 (3), pp. 381–396, March 2002.
- [9] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [10] M. Whindham and A. Cutler, "Information ratios for validating mixture analysis," *Journal of Royal Statistical Soc.*, vol. 87, pp. 1188–1192, 1992.
- [11] Y. Wu and T. S. Huang, "Towards self-exploring discriminating features for visual learning," *Engineering Applications of Artificial Intelligence*, vol. 15, pp. 139–150, Sept. 2002.
- [12] 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.
- [13] A. Dong and B. Bhanu, "A new semi-supervised EM algorithm for image retrieval," *Proc. CVPR*, vol. II, pp. 662–667, 2003.
- [14] A. Dong and B. Bhanu, "Concept learning and transplantation for dynamic image databases," *Proc. ICME*, vol. I, pp. 765–768, 2003.
- [15] G. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley & Sons, 2000.
- [16] R. J. Hathaway, "Another interpretation of the EM algorithm for mixture distributions," *Statistics & Probability Letters*, vol. 4, pp. 53–56, 1986.