Concept Learning and Transplantation for Dynamic Image Databases

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

Abstract

The task of a content-based image retrieval (CBIR) system is to cater to users who expect to get relevant images with high precision and efficiency in response to query images. This paper presents a concept learning approach that integrates a mixture model of the data, relevance feedback and long-term continuous learning. The concepts are incrementally refined with increased retrieval experiences. The concept knowledge can be immediately transplanted to deal with the dynamic database situations such as insertion of new images, removal of existing images and query images which are outside the database. Experimental results on Corel database show the efficacy of our approach.

1 Introduction

Meta knowledge in image/video databases can be used to learn and refine visual concepts ([1] [2]). However, this process necessitates a model for the learning process. Vasconcelos [3] analyzed the probabilistic image retrieval model based on mixture densities for the quality of the solution and computational complexity. However, neither relevance feedback nor the exploitation of meta knowledge is considered for the model. Barnard and Forsyth [4] organized images (with associated words) by a hierarchical model for browsing and searching. In [4], some images are used to train the clustering directly; this training stage is unreliable since the training data set may not represent the image distribution of the entire database, especially when some images are added or removed during the database lifetime.

The model estimation by the standard EM method [5] may be far away from ground-truth model due to the gap between numeric-oriented feature data and concepts understood by humans. Recently, some papers on semi-supervised learning based on mixture models have been published [6] [7]. These approaches assume that the labeled data belong to some specified classes. In reality, another kind of labeling information that "some data do NOT belong to some classes" is also available with relevance feedback, and it may improve the effectiveness of the semi-supervised learning. In [8], a new semi-supervised EM (SS-EM) algorithm is proposed that integrates concept learning and relevance feedback. However, the database is assumed to be fixed, and the query images are always from the database itself.

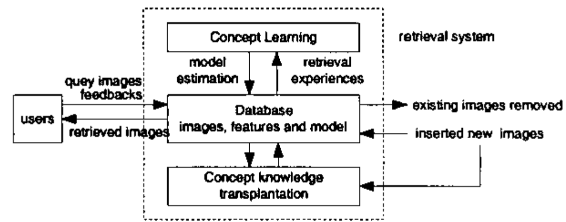


Figure 1: System diagram for concept learning and transplantation for dynamic image databases.

Fig. 1 illustrates our system for concept learning and transplantation. The contributions of this paper are: (a) Unlike most of the research in content-based image retrieval field, in our system, images can be inserted to or removed from the database, and the concept learning knowledge can be immediately transplanted to the new images. We model the dynamic mechanism of image databases, and learn concepts based on the SS-EM algorithm [8]. (b) When the query image does not belong to the database, the system can still efficiently search images using concept knowledge instead of implementing the traditional K nearest neighbor (K-NN) search. The approaches of [1] and [2] are incapable of dealing with this situation.

2 Technical approach

• **Dynamic database model:** An image retrieval system with relevance feedback mechanism may encounter two kinds of events at any time during the long-term operation: (a) users' queries and (b) database changes (i.e., image insertion or removal). We model the occurrences of these two events as *Poisson* random processes, whose distributions are $P[\mathcal{N}(t)=k]=\frac{(\lambda_i t)^k}{k!}e^{-\lambda_i t}$ ($k=0,1,\ldots$) with $i=1$ (query) and 2 (insertion/deletion), 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 $\mathcal{X}^+ = \{x_1^+, x_2^+, \ldots, x_{N^+}^+\}$ 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 estimation of the mixture model for the database corresponding to the human concepts may be continually refined over time, hence the concept learning is improved in the long term.

• **Concept learning:** We assume that the database image distribution in feature space is a c-component Gaussian mixture $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_c\}$, whose pdf is

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

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. $\Psi = \bigcup_{i=1}^{c} \{\pi_i, \mu_i, \Sigma_i\}$. For our image database system with N images, there are c concepts each of which is corresponding to one component. The task of concept learning is accomplished by estimating the mixture model parameters Ψ.

For a set of N i.i.d. samples $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ from the model (1), define the associated binary component-indicator vectors for $z_{ji} = \begin{cases} 1 & \text{if } x_j \text{ is from } i\text{th component} \\ 0 & \text{otherwise} \end{cases}$, for $j = 1, 2, \ldots, N$, $i = 1, 2, \ldots, c$. The *maximum likelihood* (ML) estimate of the unknown parameter vectors can be obtained by the EM approach, which produces a sequence of estimates $\{\hat{\Psi}(t),\ t = 0, 1, 2, \ldots\}$ by proceeding iteratively in two steps (E-step and M-step) until some termination criterion is met: In **E**-step, The conditional expectation of $\mathcal{Z}$ defined as $\tau_{ji} = \mathrm{prob}_{\hat{\Psi}(t)}\{z_{ji} = 1 | \mathcal{X}\}$ is derived as

$$\tau_{ji} = \frac{\pi_i f_i(x_j; \hat{\theta_i})}{\sum_{h=1}^{c} \pi_h f_h(x_j; \hat{\theta_h})} \tag{2}$$

In **M**-step, the component proportions, means and covariances can be estimated [5].

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. P are Q are initialized to be zero matrices at the beginning. After each 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. The cluster index h is estimated as

$$h = \arg \max_{i=1,2,\cdots,c} \mathcal{P}(i) \tag{3}$$

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 $i = 1, 2, \ldots, c$, where $\mathcal{J}^+$ and $\mathcal{J}^-$ are the indices for the samples in $\mathcal{X}^+$ and $\mathcal{X}^-$ respectively. 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 misidentified by (3).

With the accumulated knowledge contained in P and Q, the component-indicator vector elements τ_{ji} derived in (2) can be modified as

$$\tilde{\tau}_{ji} = \begin{cases} \tau_{ji} + \frac{p_{ji} - q_{ji}}{\sum_{\iota=1}^{c} (p_{j\iota} + q_{j\iota})} & \text{if } p_{ji} > q_{ji} \\ 0 & \text{if } p_{ji} < q_{ji} \\ \tau_{ji} & \text{if } p_{ji} = q_{ji} \end{cases} \tag{4}$$

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 can be inserted between E-step and M-step so that concept the learning result is closer to human's understanding.

• **Improving retrieval performance:** The concept learning result can help to improve retrieval performance. When a query is presented to the system, 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

$$D'(x_q, x_j) = e^{-\frac{\beta n}{N}} D(x_q, x_j) - \sum_{i=1}^{c} \tau_{ji} \tau_{qi} \tag{5}$$

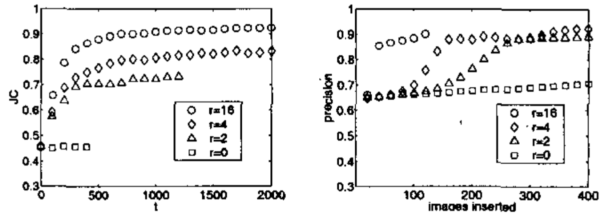
where N is the database size and n is the number of retrieval experiences. The second term on the right side is for concept learning knowledge, which is given more credit as the concept learning is improved with the retrieval experiences increased. The parameter β is to make balance between these two terms.

• **Concept transplantation:** When a new image is inserted, since the database size N is increased by 1, the positive matrix $P_{N \times c}$ and the negative matrix $Q_{N \times c}$ are both modified with one additional row, whose elements are all zeros. Furthermore, the component-indicator estimation of this new image can be computed

- 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
  (a) When new user executes retrieval: **(Learning)**
    1. Derive a retrieval experience, update P and Q.
    2. E-step: Estimate $\mathcal{Z}$ by (2).
    3. Use P and Q to modify $\mathcal{Z}$ estimation by (4).
    4. M-step: Compute component proportions, means and covariances respectively.
    5. Go to 2 until termination criterion is met.

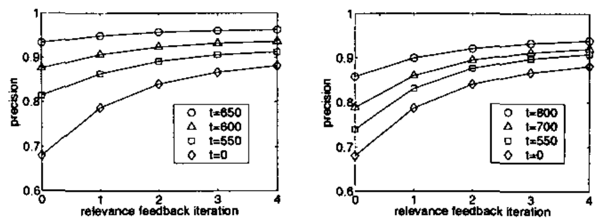
  (b) When images are inserted: **(Transplantation)**
    1. $N \leftarrow N+1$, update P and Q.
    2. Estimate z_N by (2) for new images.

Figure 2: Concept learning and transplantation.



(a) concept learning (b) retrieval precision

Figure 3: Performances with different values of r.



(a) $r = 16$ (b) $r = 2$

Figure 4: Retrieval performance at various times.

by (2) with $j = N+1$ and component proportions, means and covariances being already known. In this way, the database absorbs the new image with concept knowledge transplantation. When some images are removed from the database, the corresponding rows in the matrices $P_{N\times c}$ and $Q_{N\times c}$ are deleted. The relationship between the occurrence rates of user query and database changes influences the speed of concept learning. Obviously, when database changes occur more frequently compared with the event of user query, i.e., when the value of relative occurrence rate r is lower, the concept learning becomes slower.

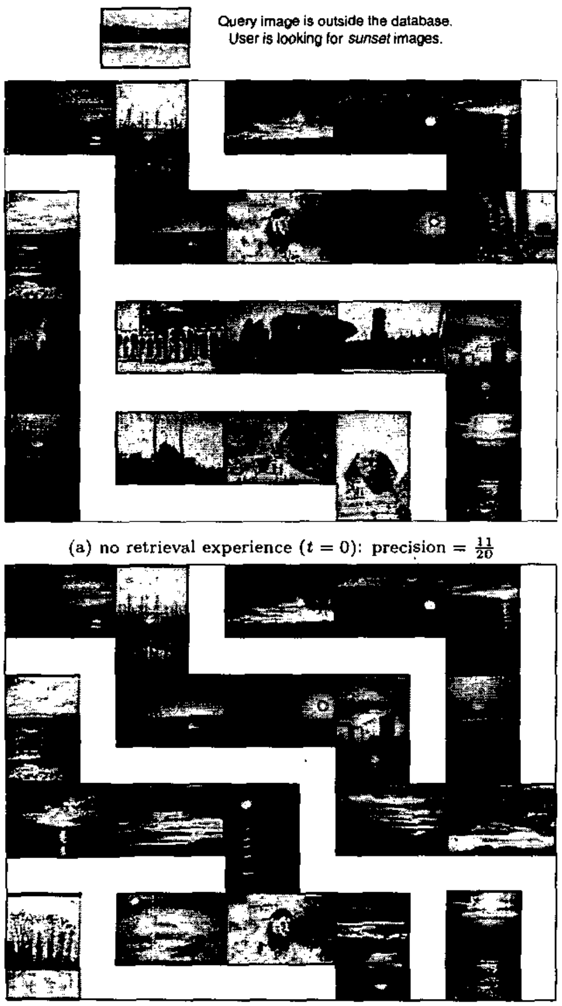
If the query image does not belong to the database, the system extracts its visual features, computes τ_{qi} ($i = 1, 2, \ldots, c$) by (2), and implements K-NN search using the distance measurement given in (5). Compared with the traditional K-NN search that is solely based on visual feature Euclidean distance measurement, this approach yields better retrieval performance since concept knowledge is adopted.

The concept learning and transplantation algorithm is presented in Fig. 2. EM-algorithm for mixture 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.

3 Experiments

We collect 1200 images from Corel stock photo library and divide them into 12 classes. Images are represented by 22 texture and color features [8]. 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 a 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 $\mathcal{C}$ and to the same cluster in $\mathcal{R}$. Let ξ_1, ξ_2 and ξ_3 be the number of BS, SD and DS pairs of vectors of $\mathcal{X}$, respectively. We use *Jaccard coefficient* $JC = \frac{\xi_1}{\xi_1+\xi_2+\xi_3}$ to evaluate clustering result.

We randomly select 800 out of the 1200 images as the initial database images, i.e., $N = 800$, and insert the other 400 images while the system is running. Our concept learning approach on the database is implemented with $c = 12$, $s = 50$ and $\beta = 100$. We set the system running time as $t = 0, 1, 2, \ldots$; at each t, one of the events happens: user query or image insertion. Queries within the database and images to be inserted are randomly selected. Fig. 3(a) shows the concept learning improvement. Initially, the Jaccard coefficient by standard EM algorithm is 45.6%. If there are only image insertions in the random process, i.e., $r = 0$, the concept learning cannot be improved. When the users' queries happen more frequently (r is higher), the concept learning will be faster. Note that when $r = 0$ and $r = 2$, after all the 400 images are inserted into the system, t is around 400 and $(400 + 2 * 400) = 1200$ respectively, and it cannot reach $t = 2000$. After an image is inserted, we use the rest of the images outside the database as queries and implement concept transplantation method to compute the retrieval precision (at relevance feedback iteration 0) by (5) . As shown



(a) no retrieval experience ($t = 0$): precision $= \frac{11}{20}$

(b) $t = 300$: precision $= \frac{19}{20}$

Figure 5: Retrieval precision is improved with retrieval experiences increased.

in Fig. 3(b), the precision increases with more images being inserted to the system due to the reason that concept learning is improved since more retrieval experiences are derived. 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 with regard to the query image can be found increases. This also explains that when $r = 0$, the precision still becomes a slightly higher with more images being inserted although concept learning is not improved. Since we only observe the process with t being from 0 to 2000, when $r = 16$, only around $2000/(1+16) = 117$ images have been inserted. This is reason that the curve for $r = 16$ cannot reach 400 in x-axis.

Fig. 4 presents the retrieval performance improvement with increased running time for $r = 16$ and $r = 2$. 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 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. Fig. 5 shows two different retrieval results with the same query image (outside the database) after different running time. In (a), there is no retrieval experience, and K-NN search only yields 11 out of 20 *sunset* images (row 1: all the 5 images; row 2: 1, 2, 4; row 3: 5; row 4: 1 and 5); In (b), when $t = 300$, 19 *sunset* images are presented (except the 3rd image on the last row) by our concept transplantation approach.

4 Conclusions

This paper proposed a new concept learning approach where the learned concept knowledge can be transplanted to new incoming images or to query images outside the database efficiently.

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] 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.

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

[3] N. Vasconcelos, "On the complexity of probabilistic image retrieval," *Proc. ICCV*, vol. 2, pp. 400–407, 2001.

[4] K. Barnard and D. Forsyth, "Learning the semantics of words and pictures," *Proc. ICCV*, vol. 2, 2001.

[5] G. McLachlan and D. Peel, *Finite Mixture Models*, John Wiley & Sons, 2000.

[6] 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.

[7] 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.

[8] A. Dong and B. Bhanu, "A new semi-supervised EM algorithm for image retrieval," *Proc. IEEE CVPR*, 2003.