Exploitation of Meta Knowledge for Learning Visual Concepts

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

Abstract

This paper proposes a content-based image retrieval system which can learn visual concepts and refine them incrementally with increased retrieval experiences captured over time. The approach consists of using fuzzy clustering for learning concepts in conjunction with statistical learning for computing "relevance" weights of features used to represent images in the database. As the clusters become relatively stable and correspond to human concept distribution, the system can yield fast retrievals with higher precision. The paper presents discussion on problems such as system mistakenly indentifying a concept, large number of trials to achieve clustering, etc. The experiments on synthetic data and real image database demonstrate the efficacy of this approach.

1. Introduction

Past several years have witnessed the developments of a variety of content-based retrieval methods and systems for image databases. Two topics of interest have been learning concepts from low-level features and relevance feedback from the user.

Tieu and Viola [11] use a boosting technique to learn a classification function when a user selects a few example images at query time. The classifier relies on 20 of the large number of visual features. Cox *et al.* [4] use a Bayesian approach for optimal solutions for multiple visual features. The Multiple Instance Learning problem is formalized in [10] and the Diverse Density algorithm is adopted to learn visual concept. However, this strategy necessitates image segmentation or region selection, which may be brittle and requires extra image processing work. All the above mentioned systems attempt to learn human concepts only with a single user.

Lipson *et al.* [6] use qualitative spatial and photometric relationships to encode class models for classifying scenes by adopting a *configural recognition* scheme. Lim [5] proposes the notion of visual keywords (entities) which can be adapted to visual content domain via learning from examples generated by human during off-line. In both of these two approaches, no relevance feedback is used.

The idea of concepts learning with fuzzy clustering and relevance feedback by exploiting meta knowledge is proposed in [3]. In that paper, the concept is directly given to the system and the user does not seek a concept. As compared to [3], in this paper, the user is given an opportunity to develop his/her own concept and the system identifies the concepts sought by various users.

In relevance feedback, the system attempts to capture the user's concept by dynamically adapting and updating the relevance of the images to be retrieved. The feedback provided by users in the forms of "similar" (positive) images and "dissimilar" (negative) images is an important part of the experience. In these systems, generally once the user is done with a query and starts a new query, the experience (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.

This paper attempts to capture and utilize the previous experiences of the system with various queries to learn visual concepts. The visual concepts are continually learned and refined over time, not necessarily from the interaction with one single user in a single retrieval session. Fuzzy clustering and relevance feedback are the main tools used for this purpose.

The key contribution of the paper is the presentation of an approach that integrates fuzzy clustering with feature relevance learning and exploits meta knowledge

0-7695-1354-9/01 $17.00 © 2001 IEEE

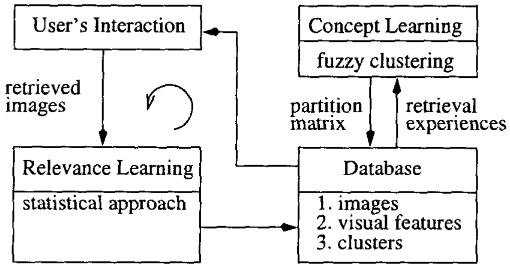


Figure 1. System diagram for concept learning using meta knowledge.

to partition the database into clusters, which can be used for efficient indexing.

2. Technical Approach

Figure 1 illustrates our approach for concept learning by exploiting meta knowledge. Since it is not uncommon that one image can be ascribed into different concepts, we use semi-supervised fuzzy c-means clustering method to learn the concept distribution, and the images' ascriptions to different concepts are represented by the resulting partition matrix. Initially, when the system is presented with a query image, it does not know which concept the user is seeking. It just presents the images to the user using the K-NN search on the entire database. If the user is not satisfied with these retrievals and provides feedback, the system attempts to decide the concept that is sought by the user.

The concept distribution knowledge is derived from semi-supervised fuzzy clustering performed over time. If the desired concept is achieved, the system only needs to search images within the cluster corresponding to this concept; otherwise, it performs statistical relevance learning to estimate feature weights and search images in the entire database. With increased retrieval experiences, the concept learning is improved, which helps to capture user's desired concept more precisely, and thus, future retrieval performance is improved.

2.1. Related Work on SS-FCM

The traditional fuzzy c-means (FCM) clustering method [2] is often frustrated by the fact that the lower values of objective function do not necessarily lead to better partitions. This actually reflects the gap between numeric-oriented feature data and concepts understood by humans. The semi-supervised fuzzy c-means (SS-FCM) clustering method [1][7][8] attempts to overcome this limitation when the labels of some of the data are already known.

Pedrycz [8] proposed a semi-supervised fuzzy clustering method with the objective function

$$J = \sum_{i=1}^{c}\sum_{k=1}^{N} u_{ik}^2 d_{ik}^2 + \alpha \sum_{i=1}^{c}\sum_{k=1}^{N} (u_{ik} - f_{ik} b_k)^2 d_{ik}^2 \quad (1)$$

where the notations are specified in the following:
c: the number of clusters,
N: the number of patterns,
$U = [u_{ik}]_{c\times N}$: partition matrix as clustering results,
d_{ik}^2 ($i = 1, 2, \cdots, c$ and $k = 1, 2, \cdots N$): the Mahalanobis distance defined as

$$d_{ik}{}^2 = ||x_k - v_i||^T W_s ||x_k - v_i|| \quad (2)$$

where W_s is a symmetrical positive definite matrix in $\mathbf{R}^n \times \mathbf{R}^n$,
x_k ($k = 1, 2, \cdots, N$): the pattern in $\mathbf{R}^n$,
v_i ($i = 1, 2, \cdots, c$): the prototype of Cluster i.

The first term on the right of (1) is the objective function in traditional fuzzy clustering. In the second term, b_k is defined as

$$b_k = \begin{cases} 1 & \text{if } x_k \text{ is labeled;} \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The matrix $F = [f_{ik}]_{c\times N}$ contains the given label vectors in appropriate columns and zero vectors elsewhere. α ($\alpha \geq 0$) denotes a scaling factor whose role is to maintain a balance between the supervised and unsupervised components within the optimization process. Usually, $\alpha = N/M$, where M is the number of labeled patterns.

The estimations of cluster prototypes and the fuzzy covariance matrices are

$$v_s = \frac{\sum_{k=1}^{N} u_{sk}^2 x_k}{\sum_{k=1}^{N} u_{sk}^2} \quad (4)$$

and

$$W_s{}^{-1} = \left[\frac{1}{\rho_s det(P_s)}\right]^{1/n} P_s \quad (5)$$

respectively, where $s = 1, 2, \cdots, c$, $\rho_s = 1$ (implying all clusters have the similar sizes), and

$$P_s = \frac{\sum_{k=1}^{N} u_{sk}^2 (x_k - v_s)(x_k - v_s)^T}{\sum_{k=1}^{N} u_{sk}^2} \quad (6)$$

The task is to minimize J with respect to the partition matrix U and the prototypes of clusters, with U satisfying two conditions:

(i) $\sum_{i=1}^{c} u_{ik} = 1, \; k = 1, 2, \cdots N,$

(ii) $u_{ik} \geq 0, i = 1, 2, \cdots, c, \; k = 1, 2, \cdots N.$

The Lagrange multiplier technique yields an expression for partition matrix

$$u_{st} = \frac{1}{1+\alpha}\left[\frac{1+\alpha(1-b_j\sum_{j=1}^{c} f_{jt})}{\sum_{j=1}^{c}\frac{d_{st}^2}{d_{jt}^2}} + \alpha f_{st} b_t\right] \quad (7)$$

where $s = 1, 2, \cdots, c$ and $t = 1, 2, \cdots, N$.

2.2. Concept Learning

Let the number of concepts (clusters) and the number of images be c and N, respectively. After a user's retrieval experience, let there be N^+ positive labeled images and N^- negative labeled images, and they are represented by $\mathbf{I}^+ = \{I_1^+, I_2^+, \cdots, I_{N^+}^+\}$ and $\mathbf{I}^- = \{I_1^-, I_2^-, \cdots, I_{N^-}^-\}$ respectively.

The task is to first determine which concept the user was seeking so that we can derive correct knowledge from this retrieval and then improve our concept learning by semi-supervised fuzzy clustering later. The index $\mathbf{k}$ of the cluster corresponding to the concept sought is computed as

$$\mathbf{k} = \arg \max_{k=1,2,\cdots,c} P(k) \quad (8)$$

where $P(k)$ is equal to

$$Prob(I_1^+ \in C_k, \cdots, I_{N^+}^+ \in C_k, I_1^- \notin C_k, \cdots, I_{N^-}^- \notin C_k)$$
$$= \prod_{i=1}^{N^+} Prob(I_i^+ \in C_k) \prod_{j=1}^{N^-} Prob(I_j^- \notin C_k)$$
$$= \prod_{i=1}^{N^+} u_{k,I_i^+} \prod_{j=1}^{N^-}(1 - u_{k,I_j^-})$$

with u_{kj} ($k = 1, 2, \cdots, c$ and $j = 1, 2, \cdots, N$) being the element of partition matrix $U_{c\times N}$ and C_k ($k = 1, 2, \cdots, c$) being concept k. This probability maximization method uses the current partition matrix information to decide the sought concept, which necessitates the assumption that current partitioning is not too bad.

Now the images in $\mathbf{I}^+$ are in cluster $\mathbf{k}$ and those in $\mathbf{I}^-$ are not in cluster $\mathbf{k}$. We designate a *positive matrix* $P_{c\times N}$ and a *negative matrix* $Q_{c\times N}$ to represent this kind of knowledge. At the very beginning, when no retrieval has ever been executed on the system, P and Q are initialized to be zero matrices. After a retrieval experience, the elements $\left\{p_{\mathbf{k},I_1^+}, \cdots, p_{\mathbf{k},I_{N^+}^+}\right\}$ in P and the elements $\left\{q_{\mathbf{k},I_1^-}, \cdots, q_{\mathbf{k},I_{N^-}^-}\right\}$ in Q are increased by 1. So the values of $p_{\mathbf{k}j}$ and $q_{\mathbf{k}j}$ represent to what extent people agree and disagree to ascribe an image j into cluster $\mathbf{k}$, respectively.

The motivation for having matrices P and Q is to capture and update previous users' retrieval experiences. In the following, P and Q are processed in the sense of statistics by estimating users' voting whether a certain image contains a specific concept or not.

Define $E = P - Q$, and let

$$b_j = \begin{cases} 0 & \text{the } j\text{th column in } E \text{ is a zero vector;} \\ 1 & \text{otherwise.} \end{cases} \quad (9)$$

for $j = 1, 2, \cdots, c$. Let M be the number of non-zero columns in E, we define

$$\alpha = N/M \quad (10)$$

We then let F be the matrix that has normalized columns of E, i.e., for the elements of F,

$$f_{kj} = \frac{e_{kj} - \min_{i=1,2,\cdots,c} e_{ij}}{\max_{i=1,2,\cdots,c} e_{ij} - \min_{i=1,2,\cdots,c} e_{ij}} \quad (11)$$

for $k = 1, 2, \cdots, c$, $j = 1, 2, \cdots, N$ and kth column in E is a non-zero vector.

If the element e_{kj} of E is negative, $k = 1, 2, \cdots, c$, $j = 1, 2, \cdots, N$, it implies that there are fewer people ascribing image j to cluster k than those opposing to this association, we conclude that image j does not contain concept k and directly predefine the element u_{kj} of partition matrix to zero. If for the jth column of $E_{c\times N}$, there are l_j negative elements whose row indices are $J(j) = \{r_{1,j}, r_{2,j}, \cdots, r_{l_j,j}\}$, we set $e_{kj} = 0$, $j = 1, 2, \cdots, N$, $k \in J(j)$.

We can now deal with the semi-supervised fuzzy clustering, which is also an optimization problem with the objective function (1). Besides the two constraints (i) and (ii) appearing in 2.1, a new constraint is added as we have discussed above:

1. Given the number of clusters c, the number of images N, positive matrix $P_{c\times N}$ and negative matrix $Q_{c\times N}$. Compute matrix $F_{c\times N}$ and α from P and Q.
2. Compute cluster centers and the fuzzy covariance matrices by (4) and (5).
3. Update partition matrix: If not predefined as 0, the elements are computed by (12).
4. If $\|U - U'\| \leq \delta$ (with δ being a tolerance limit), stop; else, go to 2 with $U = U'$.

Figure 2. Semi-supervised fuzzy clustering for concept learning

(iii) $u_{kj} = 0$, $j = 1, 2, \cdots, N$, $k \in J(j)$.

The estimations of cluster prototypes and the fuzzy covariance matrices are also (2) and (3) respectively. And we derive the expression for partition matrix elements as

$$u_{st} = \frac{1}{1+\alpha}\left[\frac{1+\alpha(1-b_j\sum_{j=1,j\notin J(t)}^{c} f_{jt})}{\sum_{j=1,j\notin J(t)}^{c}\frac{d_{st}^2}{d_{jt}^2}} + \alpha f_{st} b_t\right] \tag{12}$$

where $s = 1, 2, \cdots, c$ and $t = 1, 2, \cdots, N$. Our semi-supervised fuzzy clustering algorithm for concept learning is outlined in Figure 2.

2.3. Improving Retrieval Performance

With the partition matrix $U_{c\times N}$, we define defuzzied partition matrix $Z_{c\times N}$ whose elements are

$$z_{ik} = \begin{cases} 1 & \text{if } u_{ik} \geq \beta(\mathbf{max}_{j=1,2,\cdots,c} u_{jk}); \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

where $i = 1, 2, \cdots, c$ and $k = 1, 2, \cdots, N$. The value of $\beta \in (0, 1]$ represents to what extent we can say that the elements u_{ik} is large enough so that image k can be ascribed to cluster i.

With user's feedback after iteration 0, if L^+ images $\{I_1^+, I_2^+, \cdots, I_{L^+}^+\}$ are labeled positive and L^- images $\{I_1^-, I_2^-, \cdots, I_{L^-}^-\}$ are labeled negative by user, we check if these positive images can be ascribed into one common cluster while negative images are not in this cluster. If $\exists s \in \{1, 2, \cdots, c\}$, the following two conditions are satisfied:

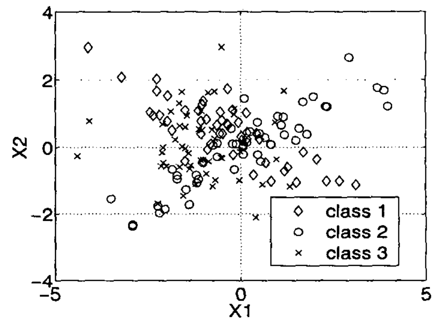
(a) $\forall j \in \{I_1^+, I_2^+, \cdots, I_{L^+}^+\}$, $u_{sj} = 1$,

(b) $\forall i \in \{I_1^-, I_2^-, \cdots, I_{L^-}^-\}$, $u_{si} = 0$,

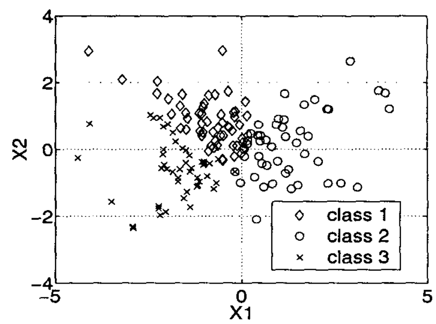
then the current user seems to be seeking the concept corresponding to cluster s. So the system saves tremendous amount of computation for feature relevance learning and searching K images over the entire database; instead, only searching K images within cluster s is needed, i.e., searching among the images whose sth element of the corresponding U column vectors are 1. When above conditions are not satisfied, we use statistical feature relevance approach presented in [9] to perform the retrievals and update clustering.

3. Experiments

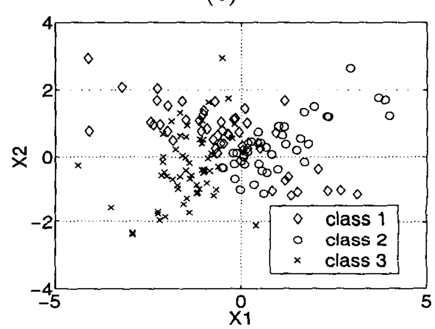
For the experiments on both synthetic data and real image database, we simulate the process of a retrieval system for which queries are selected randomly



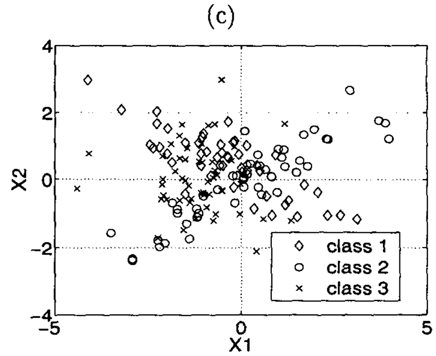
(a)



(b)



(c)



(d)

Figure 3. Fuzzy clustering results. (a) groundtruth labels, (b) 0 experience (47 errors), (c) 10 retrievals (30 errors) and (d) 30 retrievals (5 errors).

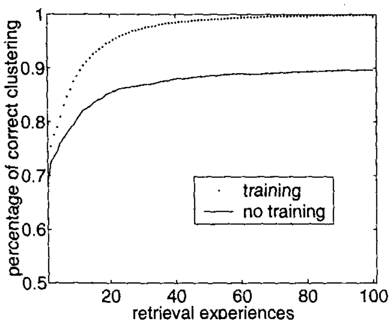


Figure 4. Synthetic data: improved clustering with increased number of retrieval experiences.

among the patterns in the database. For each retrieval, the user's interaction is monitored by a groundtruth matrix $G_{c\times N}$, whose element g_{ij} ($i = 1, 2, \cdots, c$ and $j = 1, 2, \cdots, N$) is defined as

$$g_{ij} = \begin{cases} 1 & \text{if the } j\text{th pattern has concept } i; \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

An important measure for the fuzzy clustering result is *the percentage of correct clustering*, which is defined as

$$\textbf{percentage} = \frac{\sum_i \sum_j g_{ij} \cdot \textbf{xor} \cdot z_{ij}}{cN} \tag{15}$$

where z_{ij} is the element of defuzzied partition matrix Z.

3.1. Synthetic Data

To help the reader to understand the theory of semi-supervised fuzzy clustering, we present an experiment on synthetic data. Figure 3 shows three synthetically created overlapping clusters (two-dimensional, Gaussian distribution). Each cluster contains 50 patterns. Cluster 1 and Cluster 2 are ellipsis with the same mean of $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and they have covariance matrices $\begin{pmatrix} 3.0625 & -1.6238 \\ -1.6238 & 1.1875 \end{pmatrix}$ and $\begin{pmatrix} 3.0625 & 1.6238 \\ 1.6238 & 1.1875 \end{pmatrix}$ respectively. Cluster 3 is a circle with the mean of $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$ and covariance matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Figure 3 (a) shows the cluster distribution.

We implement our clustering algorithm on this synthetic data with $c = 3$, $N = 150$, $K = 8$, and $\beta = 1$. Simulating the system with increased experiences, we randomly select a pattern as the query for each retrieval, and decide the concept (cluster) that is sought by positive and negative images. We then update the fuzzy clustering and derive the defuzzied partition matrix. An example of this process is shown in Figure 3 (b-d), in which the clustering result is improved with increased experiences.

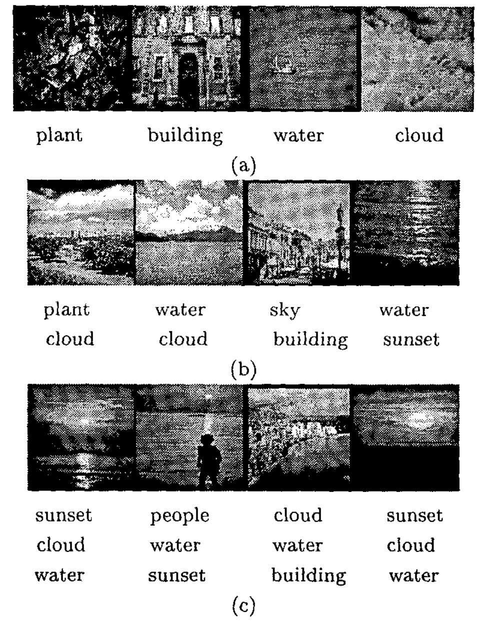


Figure 5. Sample real data with multiple concepts: sample images containing (a) 1 concept, (b) 2 concepts, (c) 3 concepts.

Figure 4 shows the average percentage of correct clustering with increased experiences. Notice that only 89.7% of correct clustering is achieved after 100 experiences. This is because the partition matrix derived from the initial fuzzy c-means clustering without any experience is far away from groundtruth matrix. After a user's experience, the system may mistakenly decide the concept sought. This incorrect knowledge will mislead the fuzzy clustering which may cause the updated partition matrix to be farther away from groundtruth matrix. After a retrieval experience, if the correctly sought concept is directly given instead of deriving it by computation, this is called a training experience. Figure 4 also gives the performance curve with training experiences, which help clustering result to finally reach 100%. The role of training stage will be discussed further in the real data experiment.

3.2. Real Data

Many image databases for retrieval research are derived from Corel Photo Collection. However, since each image in this collection has only one groundtruth label, it is not suitable for our problem where an image may belong to multiple clusters. We constructed an image database which contains 1047 images, some samples are shown in Figure 5. For each image, we ascribe it to a concept only if this concept occupies a significant area in the image. There are 9 concepts (of sizes): *plant* (115), *sky* (128), *animal* (100), *sunset* (199), *building* (249), *texture* (152), *people* (185), *cloud* (204) and *water* (146). On the average, each image contains 1.41 concepts. We use texture and color features to represent images. The texture features are derived from 16 Gabor filters [9]. 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 fuzzy clustering method on this database, with $c = 9$, $N = 1047$, $K = 16$, and $\beta = 0.5$. For the reasons of the big gap between low-level features and a human concept, the initial fuzzy clustering is far away from groundtruth labeling. We can set a training stage at the beginning of the system's running online. Let there be t training experiences, in each of which on the average L images are labeled positive or negative, the amount of concept knowledge derived from training is estimated to be $\frac{tL}{cN}$, which denotes the percentage of elements whose values are given in advance out of all the elements in the groundtruth matrix.

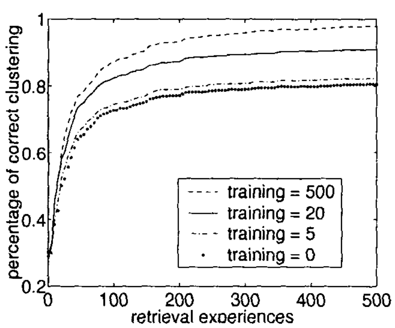


Figure 6. Real data: improved clustering with different amounts of training.

Figure 6 shows the fuzzy clustering performance of the system going through 500 retrieval experiences starting with different amounts of training experiences. With increased number of initial training experiences, fuzzy clustering is improved. Compared with the case

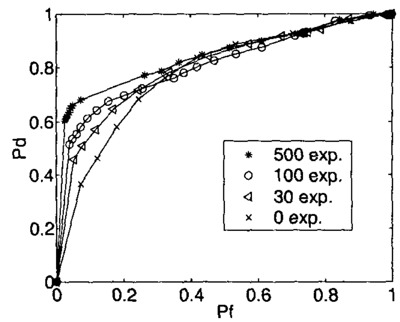


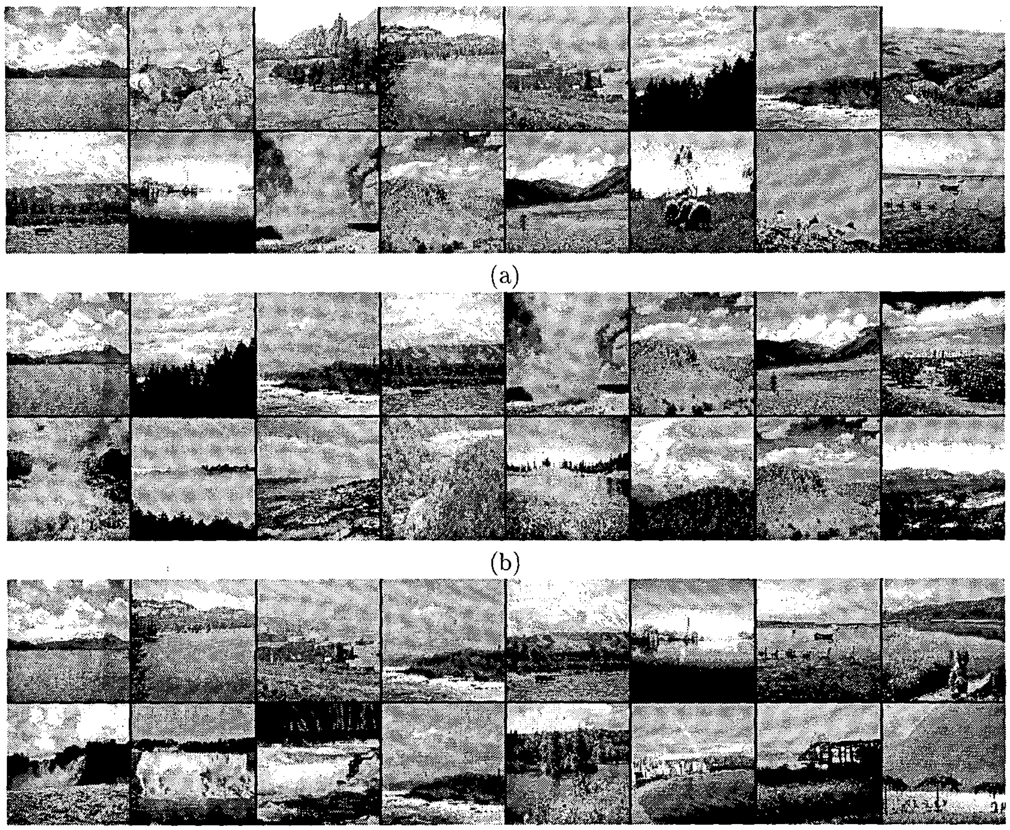
Figure 7. Real data: ROC curves for database classification with different amounts of retrieval experiences.

that has no training, 20 training experiences improve the clustering significantly. In our experiment, $L = 26$, so the amount of concept knowledge derived from the 20 training experiences is 5.5% . We also observe from Figure 6 that even with training experiences, the percentage of correct clustering still cannot converge to 100%, which again reflects the gap between image features and human visual concepts.

Figure 8. Real data: retrieval performance with various amounts of experiences.

For Concept k, $k = 1, 2, \cdots, c$, in the corresponding kth rows in groundtruth matrix G and defuzzied partition matrix Z, for $j = 1, 2, \cdots, N$, let
$N1$ = number of j that give $g_{kj} = 1$,
μ = number of j that give $g_{kj} = 1$ and $z_{kj} = 0$,
$N0$ = number of j that give $g_{kj} = 0$,
ν = number of j that give $g_{kj} = 0$ and $z_{kj} = 1$.

We define the Probability of detection and Probability of false alarms as $Pd = (N1 - \mu)/N1$ and $Pf = \nu/N0$. Calculating the average Pd and Pf over the c concepts, we obtain the ROC curves for detection



(a)

(b)

(c)

Figure 9. Different retrieval results with the same query (the first image) containing the concepts of *cloud* and *water*. The retrievals are shown after 500 experiences. Initially K-NN search yields the images in (a). When the user seeks *cloud*, 7 images having *cloud* are labeled positive (row 1: image 1, 6, 7; row 2: image 1, 3, 4, 5). After searching the *cloud* cluster, the retrieved images are shown in (b) with 12 correct images (except row 2: image 4, 5, 6, 8). When the user seeks *water*, 7 images in (a) are labeled positive (row 1: image 1, 4, 5, 7 and row 2: image 1, 2, 8). After searching the corresponding cluster, the retrieved images are shown in (c) with 15 correct images (except row2: image 7).

performance of partition matrix with different amounts of experiences shown in Figure 7. With the value of defuzzy parameter β decreased, Pd and Pf both becomes larger. Observe that with more retrieval experiences, in the case when β is not very large, the detection ability of partition matrix is improved.

Figure 8 presents the retrieval performances with different amounts of experiences starting with 20 training experiences. The *retrieval precision* is defined as

$$\textbf{precision} = \frac{\textbf{number of positive retrievals}}{\textbf{number of total retrievals}} \tag{16}$$

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 1047 images has been selected as query. Then we calculate the average precision at each iteration. Among these 1047 queries, the number of those leading to direct search within a cluster is 174, 289 and 421, respectively corresponding to 200, 300 and 500 experiences. If the percentage of correct clustering is high, the retrieval with direct search within a cluster yields a high precision after iteration 0, so it is not strange that with increased experiences, the average retrieval precision is improved. The more important aspect of direct search within one cluster is that the computational time at iteration 1 is decreased by $1/c$ compared with that of searching the entire database. This has deep significance for retrieval performance in practical applications.

Figure 9 shows two different retrievals with the same query image which is regarded as containing the concepts of both *cloud* and *water* based on the concept learning after 500 experiences.

4. Conclusions

This paper proposes a promising approach for an integrated content-based image retrieval system that uses meta knowledge to improve future image retrieval performance. The visual concepts are continually learned and refined over time. With increased retrieval experiences, concept learning by the modified semi-supervised fuzzy clustering method yields improved concept distribution knowledge, which leads to faster retrievals with higher precision. Implementing experiments on larger real data set with more concepts is our current work of interest. The dynamic concept creation, splitting and merging are the topics of future research.

5. Acknowledgements

This work was supported by DARPA/AFOFR grant F49620-97-1-0184. The contents of the information do not necessarily reflect the position or the policy of the US Government.

References

[1] A. Bensaid, L. Hall, J. Bezdek and L. Clarke, "Partially supervised clustering for image segmentation," *Pattern Recognition*, Vol. 29, No. 5, pp. 859-871, 1996.

[2] J. Bezdek, J. Keller, R. Krisnapuram and N. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing,* Kluwer Academic Publisher, 1999.

[3] B. Bhanu and A. Dong, "Concepts learning with fuzzy clustering and relevance feedback," *Proc. Workshop on Machine Learning and Data Mining in Pattern Recognition*, pp. 102-116, 2001.

[4] I. Cox, *et al.*, "The Bayesian image retrieval system, *PicHunter*: theory, implementation, and psychophysical experiments," *IEEE Trans. on Image Processing*, Vol. 9, No. 1, pp. 20-37, 2000.

[5] J. Lim, "Learning visual keywords for content-based retrieval," *Proc. IEEE Conf. on Multimedia Computing and Systems*, Vol. 2, pp. 169-173, 1999.

[6] P. Lipson, E. Grimson and P. Sinha, "Configuration based scene classification and image indexing," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1007-1013, 1997.

[7] W. Pedrycz, "Algorithm of fuzzy clustering with partial supervision" *Pattern Recognition Letters*, Vol. 31, No. 3, pp. 13-20, 1999.

[8] W. Pedrycz and J. Waletzky, "Fuzzy clustering with partial supervision," *IEEE Trans. on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 27, No. 5, pp. 787-795, 1997.

[9] J. Peng, B. Bhanu and S. Qing, "Probabilistic feature relevance learning for content-based image retrieval," *Computer Vision and Image Understanding*, Vol. 75, pp. 150-164, 1999.

[10] A. Ratan, O. Maron, W. Grimson and T. Lozano-Perez, "A framework for learning query concepts in image classification," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 423-429, 1999.

[11] K. Tieu and P. Viola, "Boosting image retrieval," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 1, pp. 228-235, 2000.