FINDING A SEMANTIC STRUCTURE INTERACTIVELY IN IMAGE DATABASES

Manjeet Rege *Ming Dong* *Farshad Fotouhi*

Machine Vision & Pattern Recognition Lab Database & Multimedia Systems Group
Department of Computer Science, Wayne State University
Detroit, MI 48202, USA
{rege, mdong, fotouhi}@wayne.edu

ABSTRACT

We present a new approach to organize an image database by finding a semantic structure interactively based on multi-user relevance feedback. By treating user relevance feedbacks as weak classifiers and combining them together, we are able to capture the categories in the users' mind and build a semantic structure in the image database. Experiments performed on an image database consisting of general purpose images demonstrate that our system outperforms some of the other conventional methods.

1. INTRODUCTION

Recent development in the field of digital media technology has resulted in the generation of a huge number of images in various applications such as medical image databases, criminal suspect tracking, travel image gallery, personal or family picture collections, etc. It has been shown that grouping these growing number of images into semantically meaningful categories is very helpful in improving the image retrieval accuracy [1].

Machine learning methods have been widely used in semantic image classification to speed up the process while providing a comparable classification accuracy. In [2], both top-down clustering based on K-means algorithm and hierarchical bottom-up clustering have been used to support fast search-by-query and effective browsing on large image databases. Vailaya et al. [3] developed a Bayesian framework to hierarchically classify vacation images. Over 90 percent classification accuracy rate has been reported over a database of 6931 vacation images. Given a few positive and negative natural image examples provided by the users, Guo et al. [4] employed two machine learning techniques, Support Vector Machine (SVM) and Adaboost, to learn the boundary between different categories. Zhang et al. [5] employ the EM algorithm to describe semantic concepts hidden in the region and image distributions of the database. Iteratively, the posterior probabilities of each region in an image to hidden semantic concepts are obtained. Hoi and Lyu [6] use SVMs to learn Web images along with their textual descriptions to search semantic concepts in image databases.

Current technology for content-based image interpretation is limited by the fact that low-level features do not represent the high-level categories accurately. As pointed out in [7, 8], user interaction is essential to accurately capture the semantics between images. The semantics of an image is usually imprecise, and depends on the users' interpretation. In order to support effective search and retrieval, both user's interest and the shift of user's interest over time needs to be reflected in the hierarchical structure of the database. Building a static semantic structure without including the users in the loop can not meet those requirements.

In this paper, we present a new approach to find semantics in an image database, i.e., to find out the meaningful image categories and their relations interactively, based on multi-user relevance feedback. By treating each user as an independent weak classifier, we show that combining multi-user feedback is equivalent to the combinations of weak independent classifiers. Furthermore, by including users in the loop, we also build a semantic structure that reflects the interests of most current users of our system. Our experimental results show that the proposed framework supports effective and efficient search and retrieval in image databases.

2. PROPOSED FRAMEWORK

We are motivated by the following two key observations:

- Each user can be treated as a classifier. The database is partitioned by the user into positive (relevant) and negative (non-relevant) sets. The partition usually has low classification accuracy due to the fact that low-level features do not represent the image content accurately and the users' feedback usually has inevitable noise. Hence, each user can be regarded as a weak classifier.
- Users are independent of each other. A user usually does not communicate with others when he makes a query and provides feedback.

Therefore, combining multi-user feedback is equivalent to the combinations of weak independent classifiers, which as a

1-4244-0367-7/06/$20.00 ©2006 IEEE

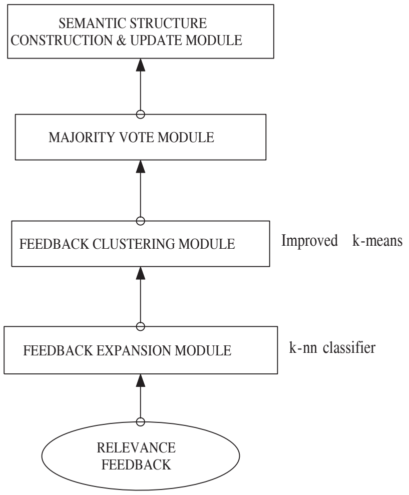
SEMANTIC STRUCTURE CONSTRUCTION & UPDATE MODULE
MAJORITY VOTE MODULE
FEEDBACK CLUSTERING MODULE
Improved k-means
FEEDBACK EXPANSION MODULE
k-nn classifier
RELEVANCE FEEDBACK

Fig. 1. Architectural overview of our system

classification system has been shown to have good generalization performance both theoretically and empirically [9, 10]. Actually, it was shown that the best classifier combination is achieved based on weak classifiers (classifiers having accuracy slightly better than random guessing) [9]. Consequently, the combination makes sense as long as the users agree in their definition of an image category at least 50% of the time.

The architectural overview of our system is shown in Figure 1. Initially, each user is provided with some randomly sampled images from the database and is asked to mark positive images in those samples. Each positive image is then expanded to its k-nearest neighbors in the feature space to obtain the positive set. Since the samples are randomly drawn from the database with equal probabilities, the expansion ratio is given by $\frac{1}{m} \times N$, where m is the number of samples provided to a user and N is the total number of images in the database. Assume a user marks n images out of m samples, the positive set contains $\frac{n}{m} \times N$ images after the expansion.

Next, we combine the expanded positive sets into Q categories using an improved K-means algorithm. One disadvantage of the conventional K-means clustering algorithm is that the exact number of clusters must be decided before clustering. In addition, the clustering results may vary because of the random initialized starting points. The idea behind our improved K-means algorithm is to run K-means algorithm many times and identify those data points that are clustered together frequently based on a two group t-test, such that the clustering results are more stable and consistent. After the completion of the clustering, a majority vote is conducted in each cluster to produce categories $C_1, C_2, \cdots, C_Q$.

We explain the construction of semantic structure from the categories in Section 2.1. The initial semantic structure constructed is not static and is updated as new feedback is collected or new images are added to the database. The update strategy is explained in Section 2.2.

2.1. Initializing the Semantic Structure

We construct a semantic structure in the image database by treating each category as a vertex in a directed graph. Assume C_i and C_j are two categories and C_{ij} is their intersection, $C_{ij} = C_i \cap C_j$. As the meaning of an image is usually imprecise and depends on users' interpretation, the intersection of two categories may or may not be disjoint. For example, mountain images may be classified to both category "Mountain" and category "Nature" by different users such that the intersection of those two categories are not empty. We consider C_i as C_j's ancestor if the following inequality holds,

$$\frac{||C_{ij}||}{||C_i||} < T_1 \quad and \quad \frac{||C_{ij}||}{||C_j||} > T_2 \tag{1}$$

where T_1 and T_2 are two thresholds, and $T_1 < T_2$. A directed edge is then drawn from vertex i to j to link these two categories. We create the semantic structure by constructing a semantic graph from the image database by the algorithm in Figure 2, where C is the set of categories, G is the semantic graph and V is the matrix that saves the edges between any two categories. Finally, the semantic graph is simplified by preserving only the relation between parents and children.

```
Input: $C = \{C_1, C_2, \cdots, C_Q\}$
Output: $G = \{C, V\}$

while($\exists(C_i, C_j)$ not checked)
    $C_{ij} = C_i \bigcap C_j$
    if($\frac{\|C_{ij}\|}{\|C_i\|} < T_1$) and ($\frac{\|C_{ij}\|}{\|C_j\|} > T_2$)
      $V_{ij} = 1$
    elseif($\frac{\|C_{ij}\|}{\|C_i\|} > T_2$) and ($\frac{\|C_{ij}\|}{\|C_j\|} < T_1$)
      $V_{ij} = -1$
    else
      $V_{ij} = 0$
    end if
end while
```

Fig. 2. Algorithm to construct a semantic structure in the image database

2.2. Updating the Semantic Structure

To trace the shift of user's interest, the semantic structure should be updated when new feedback is collected or new images added. First, a simple reject filter is built to add the new images to the existing semantic structure. We calculate the distance L between a new image and the center of each category in the semantic structure. If L is less than a threshold, we add the new image into that category (notice that in our system, one image may belong to more than one category). Otherwise the image will be put into the rejected category. In our experiment, the threshold is set as one standard deviation of the current category. After the filtering process, an image might either belong to some categories or belong to the rejected category. In the following, we refer to the images in the semantic structure as “old images” and the images in the reject category as “new images”. When a user browses the database, sample images from both the “old images” and “new images” will be presented to the user. Generally speaking, the positive images in a new feedback, may contain images purely from “old images”, or purely from “new images”, or from both. Based on the newly collected feedbacks, we generate a set of new categories, which are then compared with all existing categories. Briefly,

- If their relation meets the parent-child relationship defined in Equation 1, we insert the new category as the parent or child node of the category in the semantic graph.
- If the two categories are heavily overlapped (their overlap is greater than threshold T_2), we combine these two categories.
- Otherwise, the new category is considered to be brand new and is inserted as one child of the root node in the semantic structure.

The complete algorithm is shown in Figure 3. In the algorithm, we have used the C', C, and $\tilde{C}$ to denote the new categories, the categories in the existing semantic structure, and the categories in the updated structure respectively. G and $\tilde{G}$ stand for the semantic graph before update and after update. We also assume there are totally N new categories.

```
Input: $C' = \{C'_1, C'_2, \cdots, C'_N\}, G = \{C, V\}$
Output: $\tilde{G} = \{\tilde{C}, \tilde{V}\}$

while($\exists(C'_i, C_j)$ not checked)
    if($\frac{\|C'_i \bigcap C_j\|}{\|C'_i\|} < T_1$) and ($\frac{\|C'_i \bigcap C_j\|}{\|C_j\|} > T_2$)
      $\tilde{V}_{C'_i, C_j} = 1$
      $\tilde{C} = \{\tilde{C}, C'_i\}$
    elseif($\frac{\|C'_i \bigcap C_j\|}{\|C'_i\|} > T_2$) and ($\frac{\|C'_i \bigcap C_j\|}{\|C_j\|} < T_1$)
        $\tilde{V}_{C'_i, C_j} = -1$
        $\tilde{C} = \{\tilde{C}, C'_i\}$
    elseif($\frac{\|C'_i \bigcap C_j\|}{\|C'_i\|} > T_2$) and ($\frac{\|C'_i \bigcap C_j\|}{\|C_j\|} > T_2$)
        $\tilde{C} = \{\tilde{C}, C'_i \bigcup C_j\}$
    else
        $\tilde{C} = \{\tilde{C}, C'_i\}$
    end if
end while
```

Fig. 3. Algorithm to update the semantic structure when new images are added or new feedback collected

3. EXPERIMENTAL RESULTS

We tested the proposed framework on an image database consisting of 1583 general purpose images comprising of 10 categories (0: Image Database, 1: Nature, 2: Dawn, 3: Flower, 4: Autumn Tree, 5: Building, 6: Mountain, 7: Snow Mountain, 8: Chinese building, 9: Indoor). The low level features extracted and normalized from the images were color histogram, color coherence histogram, edge histogram, and edge coherence histogram.

In order to simulate the real world scenario in our experiment, we use feedback obtained from human subjects instead of simulated ones. This helps us capture the diversity in human perception in the interpretation of semantic meaning of images. The proposed framework is tested on both initialization and updating stages of the semantic structure. Assuming the users know the ground truth, we first collect 80 feedbacks on 595 images to initialize the semantic structure. Then, 988 new images are added and 50 new feedbacks are collected. We observed that in the updated structure, some additional categories are added while some of the existing categories are expanded due to the addition of new images. Space constraints prevent us from displaying the actual hierarchies.

Table 1 compares the classification accuracies of the first 10 positive sets and the final category generated by majority voting for each image category. The accuracies are obtained by comparing the classification results with the ground truth. Table 1 shows that our system is able to learn quickly from a few feedbacks and generate most of the categories in the database with relatively high accuracy.

We compare our approach with the top-down (generated by K-means with $K = 3$ and depth=2) and bottom-up (given by the hierarchical clustering and only nodes within the top 3 levels and with at least 70 images are kept) clustering approaches in terms of query precision and recall. For a query q belonging to a category in the ground truth, we retrieve its k-nearest neighbors in all three hierarchies. The retrieval is done in every node of the three hierarchies. The precision p and recall r are calculated as follows,

$$p = \frac{||R(q) \bigcap G(q)||}{||R(q)||} \qquad r = \frac{||R(q) \bigcap G(q)||}{||G(q)||} \tag{2}$$

where $R(q)$ is the set of retrieved images and $G(q)$ is the set of all images that lies in the same category of the ground truth with the query image q. Only the maximum precision and recall in each semantic structure are recorded. For each category in the ground truth, we make 20 random queries. The mean and variance of query precision and recall for category “Dawn” are shown in Figure 4 as a function of k (the number of retrieved images). It is obvious that query precision and recall based on our semantic structure constantly has greater mean and less variance than that with clustering approaches. We get similar results for all other categories.

Comments on Thresholds T_1 and T_2: In our experiment we set T_1 at 0.3 and T_2 at 0.7. The parent-child relationship between category C_i and category C_j will be established

Table 1. Classification accuracy before and after voting for each image category

	1	2	3	4	5	6	7	8	9	10	average	vote
Dawn	.810	.800	.628	.612	.668	.613	.685	.623	.758	.644	.685	.815
Flower	.580	.444	.657	.489	.508	.443	.477	.500	.467	.536	.510	.678
Autumn	.748	.677	.654	.724	.703	.697	.546	.717	.750	.494	671	.936
Bld.	.632	.662	.631	.720	.515	.661	.757	.633	.556	.833	.660	.702
Mt.	.773	.676	.765	.703	.672	.643	.723	.763	.794	.689	.720	.918
Snow Mt.	.622	.692	.692	.737	.731	.750	.651	.882	.773	.593	.712	.903
Nature	.811	.801	.747	.788	.782	.802	.766	.789	.776	.796	.786	.895
Indoor	.820	.818	.832	.794	.750	.681	.865	.765	.595	.670	.759	.893
Ch. B.	.280	.252	.386	.357	.318	.311	.412	.415	.430	.339	.350	.489

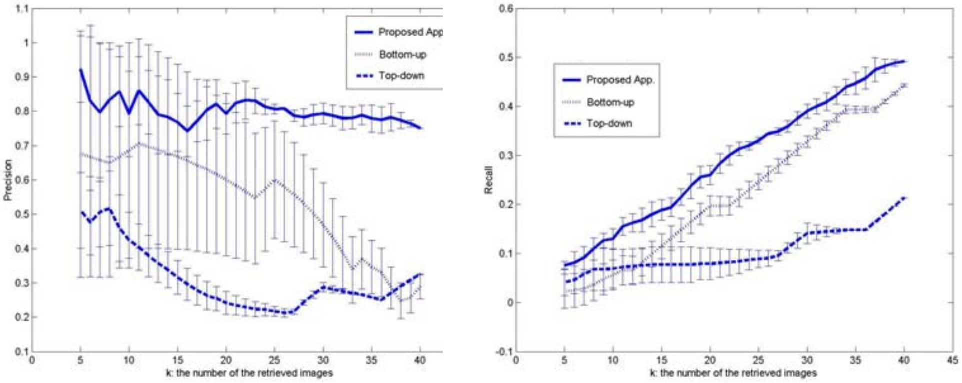

Fig. 4. Comparison of mean and variance of query precision (on the left) and recall (on the right) for category "Dawn" in three different hierarchies built by proposed framework, bottom-up clustering, and top-down clustering respectively.

if the intersection of C_i and C_j have at most 30% overlap with parent category C_i and at least 70% overlap of the child category C_j. This requirement could be weakened by either increasing T_1 or decreasing T_2. For example, if we set T_1 at 0.4 and T_2 at 0.6, we will get the same semantic hierarchy as before. In other words, our semantic hierarchy is not overwhelmingly sensitive to the two thresholds. On the other hand, we could put more strict requirements on the parent-child relationship if we decrease T_1 or increase T_2. The semantic hierarchy will be different in that case. In general, the choice of T_1 and T_2 should obtain a good balance on both noise tolerance and the required accuracy on category relationships. Currently, the best values have to be decided following the trial and error procedure.

4. CONCLUSIONS

In this paper, we present a new approach to organize an image database by finding a semantic structure interactively based on multi-user relevance feedback. Our approach is based on the observation that users can be treated as weak independent classifiers. Experimental results show that the proposed approach significantly outperforms some of the other conventional methods for image database organization.

5. REFERENCES

[1] J.Wang, J.Li, and G.Wiederhold, "Simplicity: Semantics-sensitive integrated matching for picture libraries," *IEEE Trans. on PAMI*, vol. 23, no. 9, 2002.

[2] J. Chen, C.A. Bouman, and J.C. Dalton, "Hierarchical browsing and search of large image databases," *IEEE Trans. on Image Processing*, vol. 9, pp. 442–455, 2000.

[3] A.Vailaya, M. A. T. Figueiredo, A. K. Jain, and H. J. Zhang, "Image classification for content-based image retrieval," *IEEE Trans. on Image Processing*, vol. 10, no. 1, pp. 117–129, 2001.

[4] G.Guo, A.K.Jain, W.Ma, and J.Zhang, "Learning similarity measure for natural image retrieval with relevance feedback," *IEEE Trans. on Neural Networks*, vol. 13, no. 4, 2002.

[5] R. Zhang and Z. Zhang, "Hidden semantic concept discovery in region based image retrieval," in *Proc. of IEEE CVPR*, 2004.

[6] C. Hoi and M. Lyu, "Web image learning for searching semantic concepts in image databases," in *Proc. of WWW*, 2004.

[7] R. Jain ed., ," in *Proc. of US NSF Workshop Visual Information Management Systems*, 1992.

[8] S. Santini, A. Gupta, and R. Jain, "Emergent semantics through interactuion in image databases," *IEEE Trans. on KDE*, vol. 13, no. 3, pp. 337–351, May/June 2001.

[9] C. Ji and S. Ma, "Combinations of weak classifiers," *IEEE Trans. on Neural Networks*, vol. 8, no. 1, 1997.

[10] L. Kuncheva, "A theoretical study on six classifier fusion strategies," *IEEE Trans. on PAMI*, vol. 24, no. 2, pp. 281–286, February 2002.