# Semantic Repository Modeling in Image Database

Ruofei Zhang[1], Zhongfei (Mark) Zhang[1], Zhongyuan Qin[2]
[1]Sate University of New York, Binghamton, NY 13902, USA
{rzhang, zhongfei}@cs.binghamton.edu
[2]Xi'an Jiaotong Univeristy, Xi'an, Shaanxi 710049, China
zhongyuanqin@yahoo.com.cn

## Abstract

*This work is about content based image database retrieval, focusing on developing a classification based methodology to address semantics-intensive image retrieval. With Self Organization Map based image feature grouping, a visual dictionary is created for color, texture, and shape feature attributes, respectively. Labeling each training image with the keywords in the visual dictionary, a classification tree is built. Based on the statistical properties of the feature space we define a structure, called $\alpha$-Semantics Graph, to discover the hidden semantic relationships among the semantic repositories embodied in the image database. With the $\alpha$-Semantics Graph, each semantic repository is modeled as a unique fuzzy set to explicitly address the semantic uncertainty and the semantic overlap existing among the repositories in the feature space. A retrieval algorithm combining the built classification tree with the developed fuzzy set models to deliver semantically relevant image retrieval is provided. The experimental evaluations have demonstrated that the proposed approach models the semantic relationships effectively and outperforms a state-of-the-art content based image retrieval system in the literature both in effectiveness and efficiency.*

## 1. Introduction

Few studies have considered data classification on the basis of image features in the context of image indexing and retrieval. In the general context of information retrieval, the majority of the related work has been concerned with handling textual information [10]. Not much work has been done on how to represent imagery (i.e., image features) and how to organize the features. With the high popularity and increasing volume of images in centralized and distributed environments, it is evident that the repository selection methods based on textual description is not suitable for visual queries, where the user's queries may refer to not-yet-textually-described image content [13]. One early work of image retrieval through content-based classification was reported by Huang et al [8]. Using banded color correlograms, the approach models the features using singular value decomposition (SVD) [4] and constructs a classification tree. More recently, Djeraba [5] proposed a method for the classification based image retrieval, exploiting the associations among color and texture features and using such associations to discriminate image repositories.

In this paper, we propose a new classification based methodology to content based image retrieval (CBIR). We assume that a set of training images with known class labels is available. Multiple features (color, texture, and shape) are extracted for each image in the set and are grouped to create visual dictionaries. Using the visual dictionaries for the training images, a classification tree is constructed, and any new image can be classified. To model the semantic relationships between the image repositories, a structure, called $\alpha$-Semantics Graph, is generated based on the defined semantics correlations for each semantic repository pairs. Based on the $\alpha$-Semantics Graph each semantic repository is modeled as a unique fuzzy set to explicitly address the semantic uncertainty and the semantic overlap between the semantic repositories in the feature space. A retrieval algorithm is developed based on the classification tree and the fuzzy semantics model for the semantics-relevant image retrieval.

## 2. $\alpha$-Semantics Graph and Classification Based Retrieval

To capture as much content as possible to describe and distinguish images, we extract color, texture, and shape features to form a feature vector for each image in the database. The color feature is represented as a color histogram based on the CIELab space; the texture feature is represented as a vector with each element of the vector corresponding to the energy in a specified scale and orientation sub-band w.r.t. a Gabor filter; and the edge map is used with the water filling algorithm [14] to describe the shape information for each image due to its effectiveness and efficiency for CBIR.

For each feature attribute (i.e., color, texture, and shape) we create a visual dictionary, respectively, using Self Organization Map (SOM) [9] approach as follows. (i) Performing Batch SOM learning [9] algorithm on the feature set to obtain the visualized model (node status) displayed in a 2-dimensional plane map. (ii) Regarding each node as a "pixel" in the 2-dimensional plane such that the map be-

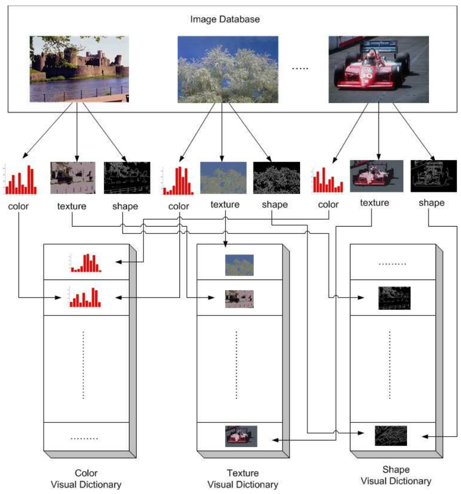


Figure 1: Generation of the visual dictionary

comes a binary image with the value of each pixel $i$ defined as 0 if $count(i) \geq t$ ($count(i)$ is the number of features mapped to the node $i$; the constant $t$ is a preset threshold) and 255 otherwise. (iii) Performing the morphological erosion operation [2] on the resulting binary image to make sparsely connected objects disjointed; the size of the erosion mask is determined to be the minimum to make two sparsely connected objects separate. (iv) With the connected component labeling [2] we assign each separated object a unique ID, a "keyword". For each "keyword", the mean of all the features associated to it is determined and stored. All "keywords" constitute the visual dictionary for the corresponding feature attribute.

Fig. 1 shows the generation of the visual dictionaries. Each entry in a dictionary is one "keyword" representing the similar features. The experiments show that the visual dictionary created captures the clustering characteristics in the feature set very well.

Two issues need to be addressed for semantics-intensive image retrieval. One is the semantic overlap between the semantic image repositories. For example, one repository named "river" has some affinities with the category named "lake". Another is the semantic uncertainty. For instance, an image containing peoples in a "beach" repository is also relevant to users inquiring the retrieval of "people" images.

To address the two issues, we propose a metric to measure the scale of semantic relationships between repositories. The metric is based on statistical measures on the shape of the repository distributions.

**Perplexity**. The perplexity of feature distributions of a repository reflects the uncertainty of the repository. As an analogy to the concept of *perplexity* [12] for a text corpus, we define the *perplexity* of a semantic repository $r_i$ in the image database on the basis of the entropy measurement [11] as

$$\wp(r_i) = 2^{-\frac{1}{N_i}\sum_{j=1}^{N_i} P(C_j)P(T_j)P(S_j)\log\{P(C_j)P(T_j)P(S_j)\}}$$

where $N_i$ is the number of images in $r_i$; $P(C_j)$, $P(T_j)$, and $P(S_j)$ are the occurrence probabilities of the single feature attribute (i.e., color, texture, and shape, respectively) in the repository, respectively. The defined *perplexity* is an approximate measure of the inhomogeneity of the feature distribution in the repository $r_i$. The more perplex in the repository, the bigger $\wp$, and vice versa.

**Distortion**. The distortion is a statistical measure to estimate the compactness degree of the repository. For each repository, $r_i$, it is defined as $D(r_i) = \frac{1}{N_i}\sqrt{\sum_{j=1}^{N_i}\|f_j - c_i\|^2}$ where $f_j$ is the feature point in the repository and $c_i$ is the centroid of the repository. The distortion describes the distribution shape of the repository in the feature space, i.e., the looser the repository, the larger $D$ defined.

Based on these statistical measures on the repositories, we propose a metric to describe the relationship between any two different repositories $r_i$ and $r_j$, $i \neq j$, in the repository set $Re$. The metric, called *semantics correlation*, is a mapping $corr : Re \times Re \longrightarrow \mathbb{R}$. For any repository pair $\{r_i, r_j\}, i \neq j$, it is defined as $corr_{i,j} = L_{i,j}/L_{max}$, where $L_{i,j} = \frac{\sqrt{(D^2(r_i)+D^2(r_j))\wp(r_i)\wp(r_j)}}{\|c_i - c_j\|}$ and $L_{max}$ is the maximal $L_{i,j}$ between any two different semantic repositories in the database, i.e., $L_{max} = \max_{r_k, r_t \in Re,\, k \neq t}(L_{k,t})$. This definition of semantics correlation has following properties: (1) If the perplexity of a repository is large, which means that the homogeneity degree of the repository is weak, it has a larger correlation to other repositories. (2) If the distortion of a repository is large, which means that the repository is looser, it has a larger correlation to other repositories. (3) If the inter-repository distance between two repositories is larger, the repository-pair has a smaller correlation. (4) The range of the semantics correlation is [0,1].

**Definition 2.1** *Given a semantic repository set* $D = \{r_1, r_2, \ldots, r_m\}$, *the semantics correlation function* $corr_{i,j}$ *defined on the set* $D$, *and a constant* $\alpha \in \mathbb{R}$, *a weighted undirected graph is called* $\alpha$*-Semantics Graph if it is constructed abiding to the following rules: (1) The* node *set of the graph is the symbolic repository set. (2) There is an* edge *between any nodes* $i, j \in D$ *if and only if* $corr_{i,j} \geq \alpha$. *(3) The weight of the edge* $(i, j)$ *is* $corr_{i,j}$.

To address the semantic uncertainty and the semantic overlap problems, we propose a fuzzy model for each repository based on the constructed $\alpha$-Semantics Graph, where each semantic repository is defined as a fuzzy set using the Cauchy function [7] as the fuzzy membership function such that one particular image may belong to several semantic repositories.

With the three visual dictionaries ready, each image in the training set is represented by a tuple $Img[Color, Texture, Shape]$ while each attribute has a discrete value type in a finite domain. To build a classification tree, the C4.5 algorithm [6] is applied on the training tuple sets obtained. We assume that each image in the training set belongs to only one semantic repository. The splitting attribute selection for each branch is based on information gain ratio [6]. Associated with each leaf node of the classification tree is a ratio $m/n$, where $m$ is the number of images classified to this node and $n$ is the number of incorrectly classified images. This ratio is a measure of the classification accuracy of the classification tree for each repository in the training image set. The image retrieval algorithm based on the classification tree and the fuzzy set model of repositories connected in the $\alpha$-semantics graph follows.

**input** : $rf$, "keyword" tuple of the query image
**output** : $Result$, images retrieved for the query image $rf$
**begin**
    Initialization: Returned image set $Result = \{\}$;
    $Q =$ the repository $q$ is classified by the classification tree;
    $acc_Q =$ the accuracy of the classification associated with $Q$;
    $c_Q =$ the center of the repository $Q$;
    $d_Q =$ the width of the repository $Q$;
    determine the distance between the reference feature $rf$ and the center of the repository $Q$ to be $dist_Q$ with the Cauchy fuzzy set membership function;
    $SetS_Q =$ the images randomly sampled form the repository $Q$ with percentage of $acc_Q$;
    $Result = Result \cup SetS_Q$;
    **for** *each node connected to the node $Q$ in the $\alpha$-Semantics Graph, $V$* **do**
        **if** $\|c_V - c_Q\| >= dist_Q$ **then**
            $dist_V = \|c_V - c_Q\| - dist_Q$;
        **else**
            $dist_V = dist_Q - \|c_V - c_Q\|$;
        **end**
        determine the membership values of the $rf$, using the Cauchy fuzzy set model;
        the percentage sampling in the repository $V$, $PR_V$ = the fuzzy membership value of $rf$ in $V$;
        $SetS_V$ =the images randomly sampled from the repository $V$ with percentage of $PR_V$;
        $Result = Result \cup SetS_V$;
    **end**
    return the set $Result$ in a $L2$ distance based rank to the query image;
**end**

**Algorithm 1:** Retrieval Algorithm

# 3. Experiment Results

We have implemented the methodology in a prototype system. The evaluation consists of two parts: the classification performance and the image retrieval performance.

To provide quantitative evaluations on the performance of image classification, we run the prototype on a controled image database. This controled database consists of 10 image repositories (African people(a1), beach(a2), buildings(a3), buses(a4), dinosaurs(a5), elephants(a6), flowers(a7), horses(a8), mountains and glaciers(a9), and food(a10)), each containing 100 images. Within this controlled database, we can assess the classification performance reliably with the categorization accuracy because the repositories are semantically non-ambiguous and share no semantic overlaps.

Table 1: Results of the classification tree based (upper) and the nearest-neighbor based (lower) image classification experiments for the controlled database

| % | a1 | a2 | a3 | a4 | a5 | a6 | a7 | a8 | a9 | a10 |
|---|---|---|---|---|---|---|---|---|---|---|
| a1 | **52** | 2 | 4 | 0 | 8 | 16 | 10 | 0 | 6 | 2 |
| a2 | 0 | **32** | 6 | 0 | 0 | 0 | 2 | 2 | 58 | 0 |
| a3 | 8 | 4 | **64** | 0 | 8 | 6 | 0 | 0 | 6 | 6 |
| a4 | 0 | 18 | 6 | **46** | 2 | 8 | 0 | 0 | 16 | 4 |
| a5 | 0 | 0 | 0 | 0 | **100** | 0 | 0 | 0 | 0 | 0 |
| a6 | 8 | 0 | 2 | 0 | 8 | **40** | 0 | 8 | 34 | 0 |
| a7 | 0 | 0 | 2 | 0 | 0 | 0 | **90** | 0 | 2 | 6 |
| a8 | 0 | 2 | 0 | 4 | 0 | 6 | 24 | **50** | 6 | 8 |
| a9 | 0 | 6 | 6 | 0 | 2 | 2 | 0 | 0 | **84** | 0 |
| a10 | 6 | 4 | 0 | 2 | 6 | 0 | 8 | 0 | 6 | **68** |

| % | a1 | a2 | a3 | a4 | a5 | a6 | a7 | a8 | a9 | a10 |
|---|---|---|---|---|---|---|---|---|---|---|
| a1 | **33** | 11 | 10 | 0 | 7 | 12 | 6 | 8 | 10 | 3 |
| a2 | 3 | **35** | 4 | 0 | 0 | 20 | 1 | 13 | 14 | 10 |
| a3 | 7 | 7 | **45** | 3 | 5 | 17 | 0 | 3 | 13 | 0 |
| a4 | 4 | 13 | 7 | **40** | 0 | 8 | 2 | 4 | 18 | 4 |
| a5 | 0 | 0 | 1 | 0 | **88** | 0 | 6 | 5 | 0 | 0 |
| a6 | 3 | 0 | 6 | 0 | 2 | **46** | 0 | 9 | 27 | 7 |
| a7 | 1 | 1 | 2 | 8 | 0 | 0 | **78** | 0 | 2 | 10 |
| a8 | 1 | 3 | 0 | 7 | 0 | 11 | 18 | **34** | 15 | 11 |
| a9 | 4 | 7 | 9 | 0 | 2 | 4 | 0 | 0 | **69** | 5 |
| a10 | 10 | 4 | 5 | 6 | 3 | 6 | 10 | 0 | 23 | **33** |

The classification performance of the constructed classification tree is compared with the nearest-neighbor classification method (NN method) [1]. For both methods, 40 randomly selected images for each repository are used to train the classifiers; the classification methods are then tested using the rest 600 images outside the training set. The classification results of our proposed method and the raw feature based NN method are shown in Table 1. In the table each row lists the percentage of images in one repository classified to each of the 10 repositories. Numbers on the diagonal show the classification accuracy for every repository. The classification performance of our proposed method is clearly better than that of the NN method since (i) the overall number of misclassification between repositories is smaller and (ii) the overall number of correct classification is larger.

The image retrieval evaluations are performed on a general-purpose color image database containing 10,000 images from COREL collection of 96 semantic repositories. Each semantic repository has 85-120 images. We randomly take 50% of them as the training set to train the image classifier. To evaluate the image retrieval performance, 1,500 images are randomly selected from all repositories of the remaining 50% of the COREL collection as the query set. We invite a group of 5 users to participate the evaluations. The participants consists of CS graduate students as well as lay-people outside the CS Department. The relevancy of the retrieved images is subjectively examined by the users and the retrieval accuracy is the average values across all query sessions.

Before we evaluate the prototype system, an appropriate $\alpha$ must be determined for the $\alpha$-semantics graph. For the

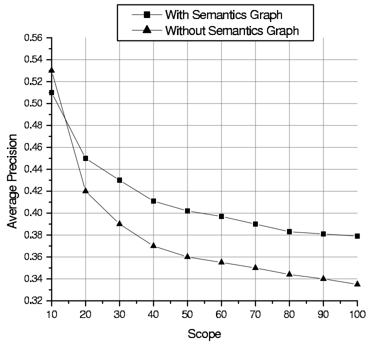


Figure 2: Average precision comparison with/without $\alpha$-Semantics Graph

extreme case $\alpha = 0$, each node is connected to all other nodes in the 0-Semantics Graph (all repositories are treated as semantics-related to each other); on the other hand, for $\alpha = 1$, each node is isolated (with no edges connected to other nodes), and the 1-Semantics Graph is degraded to a repository set. In the experiment we have computed the pair-wise semantics correlation $corr_{i,j}$ for all the repository pairs in the training set; the third quartile, which is obtained as 0.649 for the training set, is used as the $\alpha$ in the prototype.

To evaluate the effectiveness of the semantics correlation measure and the fuzzy model for the repositories, we have compared the retrieval precisions with and without $\alpha$-Semantics Graph. Fig. 2 shows the results, in which it is evident that the $\alpha$-Semantics Graph and the derived fuzzy model for the repositories improve the retrieval precision substantially.

Considering that it is difficult to design a fair comparison with existing very few classification-based image retrieval methods, we have compared the average retrieval precision of our method with that of UFM [3], a state-of-the-art CBIR system, as shown in Fig. 3. It is clear that both the absolute precision and potential (attenuation trend) of our method are superior to those of UFM.

Another advantage of our method is its high online query efficiency. In most CBIR systems, the search is performed linearly (i.e., the retrieval complexity is $O(n)$ where $n$ is the number of images in the database). In our method, the average computation complexity is $O(\log m)$ for image classification and $O(w)$ for image similarity comparison, where $m$ is the number of image repositories and $w$ is the average number of images in a repository. Since $w = \frac{n}{m}$, the overall complexity is $O(\log m + \frac{n}{m})$. In general, $m << n$. Hence, with image classification the computation complexity of our method is much more tractable than that of the linear search methods. This conclusion is also supported in the experiment.

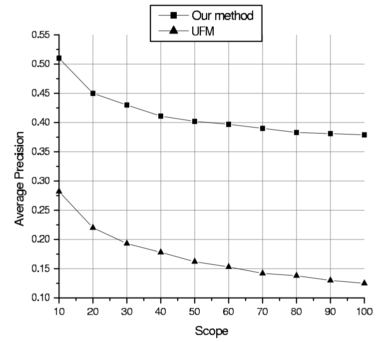


Figure 3: Average precision comparison of our method with that of UFM

# 4. Conclusions

A semantics correlation based structure, called $\alpha$-Semantics Graph, is proposed to explicitly represent the semantic uncertainty and the semantic overlap existing in an image database. Founded on the $\alpha$-Semantics Graph, each semantic repository is modeled as a fuzzy set which captures the statistical distribution in the feature space. With the generation of a multiple feature (color, texture, and shape) supported visual dictionary, a classification tree is trained using a provided training set. A unique image retrieval algorithm is developed and is demonstrated with promising performance for image retrieval.

# References


[1] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Trans. on PAMI*, 24(8):1026–1038, 2002.

[2] K. R. Castleman. *Digital Image Processing*. Prentice Hall, Upper Saddle River, NJ, 1996.

[3] Y. Chen and J. Z. Wang. A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Trans. on PAMI*, 24(9):1252–1267, 2002.

[4] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of American Sociation of Information Science*, 41:391–407, 1990.

[5] C. Djeraba. Association and content-based retrieval. *IEEE Transaction on Knowledge and Data Engineering*, 15(1):118–135, January 2003.

[6] M. H. Dunham. *Data Mining, Introductory and Advanced Topics*. Prentice Hall, Upper Saddle River, NJ, 2002.

[7] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. John Wiley & Sons, New York, 1999.

[8] J. Huang, R. Kumar, and R. Zabih. An automatic hierarchical image classification scheme. In *The Sixth ACM Int'l Conf. Multimedia Proceedings*, 1998.

[9] T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, and A. Saarela. Self organization of a massive document collection. *IEEE Trans. on Neural Networks*, 11(3):1025–1048, May 2000.

[10] M. Koster. Alweb: Archie-like indexing in the web. *Computer Networks and ISDN Systems*, 27(2):175–182, 1994.

[11] C. Shannon. Prediction and entropy of printed english. *Bell Sys. Tech. Journal*, 30:50–64, 1951.

[12] G. Taubin and D. B. Cooper. Recognition and positioning of rigid objects using algerbraic moment invariants. In *SPIE Geometric Methods in Computer Vision Proceedings*, volume 1570, pages 175–186, 1991.

[13] R. Zhang and Z. Zhang. Addressing cbir efficiency, effectiveness, and retrieval subjectivity simultaneously. In *ACM Multimedia 2003 Multimedia Information Retrieval Workshop*, Berkeley, CA, November 2003. in conjunction with ACM Multimedia (ACM MM) 2003.

[14] X. S. Zhou, Y. Rui, and T. S. Huang. Water filling: A novel way for image structural feature. In *IEEE Conf. on Image Processing Proceedings*, 1999.