A Data Mining Approach to Modeling Relationships Among Categories in Image Collection

Ruofei Zhang, Zhongfei (Mark) Zhang
Department of Computer Science
SUNY at Binghamton
Binghamton, NY 13902
USA
{rzhang, zhongfei}@cs.binghamton.edu

Sandeep Khanzode
Polaris Software Lab
Andheri East, Mumbai 400 096
Maharashtra
India
sandeepkhanzode@hotmail.com

ABSTRACT

This paper proposes a data mining approach to modeling relationships among categories in image collection. In our approach, with 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 α-*Semantics Graph*, to discover the hidden semantic relationships among the semantic categories embodied in the image collection. With the α-*Semantics Graph*, each semantic category is modeled as a unique fuzzy set to explicitly address the semantic uncertainty and semantic overlap among the categories in the feature space. The model is utilized in the semantics-intensive image retrieval application. An algorithm using the classification accuracy measures is developed to combine the built classification tree with the fuzzy set modeling method to deliver semantically relevant image retrieval for a given query image. The experimental evaluations have demonstrated that the proposed approach models the semantic relationships effectively and the image retrieval prototype system utilizing the derived model is promising both in effectiveness and efficiency.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*data mining, image databases*

General Terms

Algorithms, Measurement, Design

Keywords

Relationships, semantic category, fuzzy model, image collection

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
KDD'04, August 22–25, 2004, Seattle, Washington, USA.
Copyright 2004 ACM 1-58113-888-1/04/0008 ...$5.00.

1. INTRODUCTION

Large collections of images have become popular in many applications, from photo collections to Web pages or even video databases. To classify or retrieve them is a challenge which is the focus of many research projects (for instance IBM's QBIC [6]). Almost all of these systems generate low-level image features such as color, texture, shape, and motion, for image index and retrieval. This is partly because low-level features can be computed automatically and efficiently. The semantics of the images, which users are mostly interested in, however, is seldom captured by the low-level features. On the other hand, there is no effective method yet to automatically generate good semantic features of an image. One common compromise is to obtain some semantic information through manual annotation. Since visual data contain rich information and the manual annotation is subjective and ambiguous, it is difficult to capture the semantic content of an image using words precisely and completely, not to mention the tedious and labor-intensive work involved.

It is desirable to organize an image collection in a meaningful manner using image classification. Image classification is the task of classifying images into (semantic) categories based on the available training data. A common approach to image classification involves addressing the following four issues: (i) image features – how to represent the image; (ii) organization of the feature data – how to organize the data; (iii) classifier – how to classify an image; and (iv) semantics modeling – how to address the relationships between the semantic classes.

In this paper, we propose a data mining approach to modeling relationships among categories in image collection. 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 collection and are grouped to create visual dictionaries. Using the visual dictionaries for the training images, a classification tree is constructed. Once the classification tree is obtained, any new image can be classified easily. On the other hand, to model the semantic relationships between the image categories, a representation, called α-Semantics Graph, is generated based on the defined semantics correlations for each semantic category pairs. Based on the α-Semantics Graph each semantic category is modeled as a unique fuzzy set to explicitly address the semantic uncertainty and the semantic overlap between the semantic categories in the feature space. For the image retrieval application, a retrieval algorithm is developed based on the classification tree and the fuzzy semantics model for the semantics-relevant image retrieval.

2. RELATED WORK

Very 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 [2]. Not much work has been done on how to represent imagery (i.e., image features) and how to organize the features. In the following, we review some of the previous work on automatic classification based image retrieval.

Yu and Wolf presented a one-dimensional *Hidden Markov Model* (HMM) for indoor/outdoor scene classification [15]. An image is first divided into horizontal (or vertical) segments and each segment is further divided into blocks. Color histograms of blocks are used to train HMM's for a preset standard set of clusters, such as a cluster of sky, tree, river, a cluster of sky, tree, grass, etc. Maximum likelihood classifiers are then used to classify an image as indoor or outdoor. In general, it is difficult to enumerate a set to cover a general case such as indoor/outdoor. The *configural recognition* scheme proposed by Lipson et al [10] is also a knowledge-based scene classification method. A model template, which encodes the common global scene configuration structure using qualitative measurements, is hand-crafted for each category. An image is then classified to the category whose model template best matches the image by deformable template matching (which requires intensive computation, despite that the images are subsampled to low resolutions) — the nearest neighbor classification.

One early work for resource selection in distributed visual information systems was reported by Chang et al [3]. The method proposed was based on a metadata base at a query distribution server. The metadata base records a summary of the visual content of the images in each category through image templates and statistical features. The selection of the databases is driven by searching the metadata base using a nearest-neighbor ranking algorithm that uses query similarity to a template and the features of the database associated with the template. Another approach [8] proposes a new scheme for automatic hierarchical image classification. Using banded color correlograms, the approach models the features using singular value decomposition (SVD) [4] and constructs a classification tree. The technique used extracts a certain form of knowledge to classify images. Using a noise-tolerant SVD description, the image is classified in the training data using the nearest neighbor with the first neighbor dropped. Based on the performance of this classification, the categories are partitioned into subcategories, and the interclass dissociation is minimized through using normalized cuts. In this scheme, the content representation is weak (only using color and some kind of spatial information) and the semantic overlap among semantic categories in the feature space is not addressed.

3. IMAGE FEATURES AND VISUAL DICTIONARIES

To capture as much content as possible to describe and distinguish images, we extract multiple semantics-related features as image signatures. Specifically, our framework incorporates color, texture, and shape features to form a feature vector for each image in the collection. Since image features $f \in \mathbb{R}^n$, it is necessary to perform regularization on the feature set such that the visual data can be indexed efficiently. In our approach, we create a visual dictionary for each feature attribute to achieve this objective.

3.1 Image Features

The color feature is represented as a color histogram based on the CIELab space [1] due to its desired property of the perceptual color difference proportional to the numerical difference in the CIELab space. The CIELab space is quantized into 96 bins (6 for L, 4 for a, and 4 for b) to reduce the computational intensity. Thus, a 96-dimensional feature vector C is obtained for each image as a color feature representation.

To extract texture information of an image, we apply a set of Gabor filters [12], which are shown to be effective for CBIR [11], to the image to measure the response. The Gabor filters are one kind of two-dimensional wavelets. The discretization of a two-dimensional wavelet applied to an image is given by

$$W_{mlpq} = \iint I(x,y)\psi_{ml}(x - p\triangle x, y - q\triangle y)dxdy \quad (1)$$

using Gabor function ψ_{ml}, where I denotes the processed image, $\triangle x$, $\triangle y$ denotes the spatial sampling rectangle; p, q are image positions, and m, l specify the scale and orientation of the wavelets, respectively.

Applying the Gabor filter bank to an image results, for every image pixel (p, q), in an M (the number of scales in the filter bank) by L array of responses to the filter bank. We only need to retain the magnitudes of the responses:

$$F_{mlpq} = |W_{mlpq}| \quad m = 0, \ldots, M-1,\ l = 0, \ldots L-1 \quad (2)$$

Hence, a texture feature is represented by 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. In the implementation, a Gabor filter bank of 6 orientations and 4 scales is performed for each image in the collection, resulting in a 48-dimensional feature vector T (24 means and 24 standard deviations for $|W_{ml}|$) for the texture representation.

The edge map is used with water filling algorithm [17] to describe the shape information for each image due to its effectiveness and efficiency for CBIR. A 18-dimensional shape feature vector, S, is obtained by generating edge maps for each image in the collection.

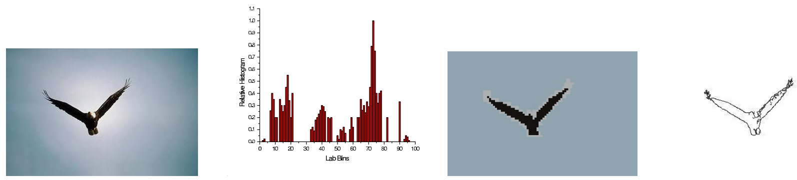
Fig. 1 shows visualized illustrations of the extracted color, texture, and shape features for an example image. These features describe the content of images and are used to index the images.

3.2 Visual Dictionary

The creation of the visual dictionary is a fundamental preprocessing step necessary to index features. It is not possible to build a valid classification tree without the preprocessing step in which similar features are grouped. The centers of the feature groups constitute the visual dictionary. Without the visual dictionary, we would have to consider all feature values of all images, resulting in a situation where very few feature values shared by images, which makes it impossible to discriminate categories.

For each feature attribute(color, texture, and shape) we create a visual dictionary, respectively, using Self Organization Map (SOM) [9] approach. SOM is ideal for our problem, as it can project high-dimensional feature vectors to 2-dimensional plane with mapping similar features together while separating different features apart at the same time.

The procedure to create "keywords" in the dictionary is similar to the one developed in [16]. By the procedure, the number of "keywords" is adaptively determined and the similarity-based feature grouping is achieved. Applying this procedure to each feature attribute, a visual dictionary is created for each one.



(a) (b) (c) (d)

Figure 1: An example image and its corresponding color, texture and shape feature maps. (a)The original image. (b)The CIELab color histogram. (c)The texture map. (d) The edge map.

4. α-SEMANTICS GRAPH AND FUZZY MODEL FOR CATEGORIES

Although we can take advantage of the semantics-oriented classification information from the training set, there are still some issues not addressed yet. One is the semantic overlap between the classes. For example, one category named "river" has some affinities with the category named "lake". For some users, the images in the category "lake" are also interesting although they pose a query image of "river". Another issue is the semantic uncertainty, which means that an image in one category may also contain semantic objects inquired by the user although the category is not for the semantics in which the user is interested. For instance, an image containing peoples in an "beach" category is also relevant to users inquiring the retrieval of "people" images. To address these issues, we need to construct a model to explicitly describe the semantic relationships among images and the semantics representation for each category.

4.1 α-Semantics Graph

To describe the uncertainty and overlap of semantic categories quantitatively, we propose a metric to measure the scale, called *semantics correlation*, which reflects the relationships between two semantic categories in the feature space. The semantics correlation is based on statistical measures on the shape of the category distributions.

Perplexity. The perplexity of feature distributions of a category reflects the uncertainty of the category; it can be represented based on the entropy measurement [13]. Suppose there are k elements $s_1, s_2, \ldots, s_k$ in a set with probability distribution $P = \{p(s_1), p(s_2), \ldots, p(s_k)\}$. The entropy of the set is defined as

$$En(P) = -\sum_{i=1}^{k} p(s_i) \log p(s_i)$$

By Shannon's theorem [13], this is the lower bound on the average number of *bits per element (bpe)* required to encode a state of the set. For a particular semantics represented in the images, it is difficult to precisely determine the probability of an image feature $p(s_i)$. Consequently we use the statistics in the training semantic category to estimate the probabilities. Since each image is represented as a 3-component vector $[C, T, S]$, the entropy of each category, r_i, is defined as

$$H(r_i) = -\frac{1}{N_i} \sum_{j=1}^{N_j} P(C_j, T_j, S_j) \log P(C_j, T_j, S_j) \qquad (3)$$

where $P(C_i, T_i, S_i)$ is the joint occurrence probability of an image feature in the category and N_i is the number of images in the category. Assuming that color, texture and shape properties are independent in image representation, i.e., $P(C_j, T_j, S_j) = P(C_j)P(T_j)P(C_j)$ where $P(C_j)$, $P(T_j)$, and $P(S_j)$ are the occurrence probabilities of the single feature attribute in the category, respectively, it follows that

$$H(r_i) = -\frac{1}{N_i} \sum_{j=1}^{N_i} P(C_j)P(T_j)P(S_j) \log\{P(C_j)P(T_j)P(S_j)\} \qquad (4)$$

As an analogy to the concept of *perplexity* [14] for a text corpus, we define the *perplexity* of a semantic category r_i in the image collection as

$$\wp(r_i) = 2^{H(r_i)} \qquad (5)$$

which is an approximate measure of the inhomogeneity of the feature distributions in the category r_i. The more perplex in the category, the bigger $\wp$; and vice versa.

Distortion. The distortion is a statistical measure to estimate the compactness degree of the category. For each category, r_i, it is defined as

$$D(r_i) = \frac{1}{N_i} \sqrt{\sum_{j=1}^{N_i} \|f_j - c_i\|^2} \qquad (6)$$

where f_j is the feature point j in this category and c_i is the centroid of the category. The distortion describes the distribution shape of categories, i.e., the looser the category, the bigger D defined.

Based on these statistical measures on the categories, we propose a metric to describe the relationship between any two different categories r_i and r_j, $i \neq j$, in the category set Re. The metric, called *semantics correlation*, is a mapping $corr : Re \times Re \longrightarrow \mathbb{R}$. For any category pair $\{r_i, r_j\}, i \neq j$, it is defined as

$$L_{i,j} = \frac{\sqrt{(D^2(r_i) + D^2(r_j))\wp(r_i)\wp(r_j)}}{\|c_i - c_j\|} \qquad (7)$$

$$corr_{i,j} = L_{i,j}/L_{max} \qquad (8)$$

where L_{max} is the maximal $L_{i,j}$ between any two different semantic categories, and $L_{max} = \max_{r_k, r_t \in Re,\, k \neq t}(L_{k,t})$. This definition of semantics correlation has following properties:

- If the perplexity of a category is large, which means that the homogeneity degree of the category is weak, it has a larger correlation with other categories.
- If the distortion of a category is large, which means that the category is looser, it has a larger correlation with other categories.
- If the inter-category distance between two categories is larger, the category-pair has a smaller correlation.
- The range of the semantics correlation is [0,1].

For convenience, the supplement of the semantics correlation for each semantic category pair is defined as

$$disc_{i,j} = 1 - corr_{i,j} \tag{9}$$

and is called *semantics discrepancy* between the two different semantic categories, resulting in an quantitative measure of the relationship between any two different semantic categories based on their distributions in the feature space.

With semantics correlations defined above, a graph is constructed in the category space. We call the graph α-Semantics Graph. It is defined as follows:

DEFINITION 4.1. *Given a semantic category 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 α-Semantics Graph if it is constructed abiding to the following rules:*

- *The* node *set of the graph is the symbolic category set.*
- *There is an* edge *between any nodes $i, j \in D$ if and only if $corr_{i,j} \geq \alpha$.*
- *The weight of the edge (i, j) is $corr_{i,j}$.*

The α-Semantics Graph uniquely describes the relationships between semantic categories for an arbitrary α. With a tuned α, we can model a semantic category based on its connected neighbors and corresponding edge weights in the α-Semantics Graph.

4.2 Fuzzy Model for Categories

To address the semantic uncertainty and the semantic overlap problems, we propose a fuzzy model for each category based on the constructed α-Semantics Graph. In this model , each semantics category is defined as a fuzzy set while one particular image may belong to several semantic categories.

A fuzzy set F on the feature space $\mathbb{R}^n$ is defined by a mapping $\mu_F : \mathbb{R}^n \rightarrow [0, 1]$ named the *membership function*. For any feature vector $f \in \mathbb{R}^n$, the value of $\mu_F(f)$ is called the degree of membership of f to the fuzzy set F (or, in short, the degree of membership to F). For a fuzzy set F, there is a smooth transition for the degree of membership to F besides the hard cases $f \in F$ ($\mu_F(f) = 1$) and $f \notin F$ ($\mu_F(f) = 0$).

The most commonly used prototype membership functions are cone, trapezoidal, B-splines, exponential, Cauchy, and paired sigmoid functions [7]. Since we could not think of any intrinsic reason why a particular one should be preferred to any other, we tested the cone, trapezoidal, exponential, and Cauchy functions in our system. In general, the performance of the exponential and the Cauchy functions is better than that of the cone and trapezoidal functions. Considering the computational complexity, we use the Cauchy function because it requires much less computation. The Cauchy function is defined as

$$\mathcal{F}(x) = \frac{1}{1 + (\frac{\|x-v\|}{d})^\beta}$$

where d and $\beta \in \mathbb{R}$, $d > 0$, $\beta > 0$, v is the center location (point) of the fuzzy set, d represents the width of the function, and determines the shape (or smoothness) of the function. Collectively, d and β portray the grade of fuzziness of the corresponding fuzzy set. For fixed d, the grade of fuzziness increases as β decreases. If β is fixed, the grade of fuzziness increases with the increase of d.

For each category, the parameters v and d are determined based on the constructed α-Semantics Graph. For the center point of each semantic category r_i, it can be conveniently estimated by the mean vector, c_i, of the feature vectors in the category. For the width d_i, it is determined as follows:

$$d_i = \sum_{r_k \in Nb(r_i)} \|c_i - c_k\| corr_{i,k} \tag{10}$$

where $Nb(r_i)$ is the set of the centroids of all connected nodes to the node r_i in the α-Semantics Graph and $\| \cdot \|$ is the Euclidean distance in $\mathbb{R}^n$. In other words, the width of the membership function for each category is a semantics correlation weighted combination of the distance to its connected nodes in the α-Semantics Graph. Consequently, each category r_i in the training set is modeled as a unique fuzzy set

$$\mathcal{F}_i(f) = \frac{1}{1 + (\frac{\|f-c_i\|}{d_i})^\beta} \tag{11}$$

Denoting the distance between a feature f and c_i as $dist$, the above equation can be equally presented as

$$\mathcal{F}_i(dist) = \frac{1}{1 + (\frac{dist}{d_i})^\beta} \tag{12}$$

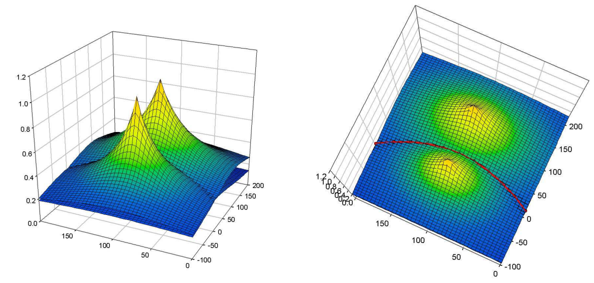
The experiments show that the performance changes insignificantly when β is in the interval $[0.7, 1.5]$, but degrades rapidly outside the interval. Thus, we set $\beta = 1$ in Eq. 11 to simplify the computation.

5. CLASSIFICATION BASED RETRIEVAL ALGORITHM

With the three visual dictionaries ready, an order for the "keywords" in the visual dictionaries is determined and an index to each "keyword" is assigned. Given an image, for each feature attribute, replace it by the index of the "keyword" to which it is assigned in the corresponding visual dictionary. Hence each image in the training set is represented by a tuple $Img[Color, Texture, Shape]$ while each attribute has discrete value type in a limited domain.

To build a classification tree, C4.5 algorithm [5] is applied on the training tuple sets transformed. We assume that each image in the training set belongs to only one semantic category. The splitting attribute selection for each branch is based on information gain ratio [5]. 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 class in the training image set.

A retrieval algorithm is developed based on the classification and its accuracy with the fuzzy model for the category to which the query image is classified. In this algorithm, the category is predicted by the classification tree for the query image. At the same time, a reference feature is determined by inverse analysis from the classification accuracy; the reference feature's membership values of the semantic categories of the neighborhood to the predicted category in the α-Semantics Graph are determined. The intuition is illustrated in Fig. 2. In this figure, two categories modeled with the fuzzy set are shown. Every vector in the feature space is associated with the two categories by the obtained membership values. These membership values are used to guide the sampling percentage in the corresponding semantic categories. In addition, since the algorithm is orthogonal with the distance metric $\mathcal{DM}$, different distance metric $\mathcal{DM}$ can be used for different applications. In our evaluation experiment, we use Euclidian distance as $\mathcal{DM}$ for its simplicity and effectiveness.



(a) (b)

Figure 2: Illustration of two semantic category models in the feature space. (a) Sideview. (b) Topview; the dark (red) curve represents part of the intersection curve.

With this algorithm the images are retrieved not only based on the category the query image is classified to (which is called *primary category*) but also based on the semantics correlations between this primary category and other neighboring categories in the α-Semantics Graph constructed. The percentage of images sampled in each potential relevant categories is determined by the corresponding classification accuracy and its fuzzy model. Intuitively, we give most share to the *primary category*; the shares to the connected categories of the *primary category* in the α-Semantics Graph are based on their semantics correlations with the *primary category*. In other words, more shares are given to the highly semantics-correlated categories while fewer shares to the lowly semantics-correlated categories. Consequently, we have solved the semantic uncertainty and semantic overlap problems explicitly.

6. EXPERIMENT RESULTS

We have implemented the approach in a prototype system on a platform of Pentium IV 2.0 GHZ CPU with 256M memory. The image retrieval evaluations were performed on a general-purpose color image collection containing 10,000 images from COREL collection of 96 semantic categories. Each semantic category has 85–120 images. Images in the same category are often not all visually similar. For this image collection we randomly shuffle the images in each category and take 50% of them as the training set to train the image classifier. To evaluate the image retrieval performance, 1,500 images were randomly selected from all categories of the remaining 50% of the COREL collection as the query set. 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 α must be decided. For the extreme case $\alpha = 0$, each node is connected to all other nodes in the 0-Semantics Graph (all categories are treated as semantics-related to each other); for $\alpha = 1$, each node is isolated (with no edges connected to other nodes), the 1-Semantics Graph degraded to a category set. In the experiment we have calculated pair-wise semantics correlation $corr_{i,j}$ for all the category pairs in the training set; the third quartile, which is obtained as 0.649 for the training set, was used as the α in the prototype.

Fig. 3 shows an excerpted α-Semantics Graph example for the categories in the training set. The annotation of each category is labeled on its node. The length of each edge between two nodes in the figure is proportional to the *semantics discrepancy* between the two corresponding categories. It is noticeable that the semantic uncertainty and the semantic overlap among categories described in Section 4.1 are measured explicitly. For example, for the "outdoor scene" category, category "castle" is more semantics-correlated than "beach" category; category "waterfall" has strong semantics correlations with "fishing", "rafting", and "beach" categories; "peasant life" category is connected to "outdoor scene" and "fashion model" categories. These semantics correlations measured in the feature space among categories agree well to the subjective perceptions of the image contents.

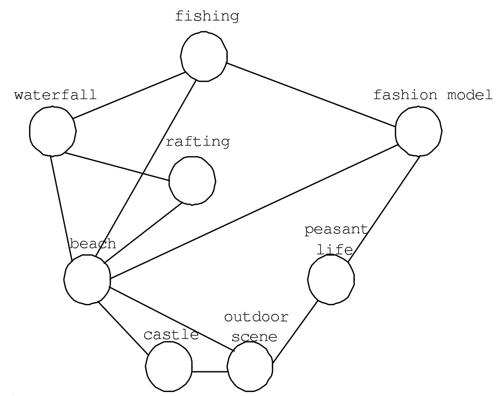


Figure 3: An example of α-semantics graph.

(a) (b) (c)

Figure 4: Three test images. (a)This image is associated with a single category by α-semantics graph. (b)This image is associated with 3 categories. (c)This image is associated with 7 categories.

Fig. 4 shows three test images with 0, 2 and 6 categories connected in the constructed α-semantics graph respectively. The *primary category* assigned to Fig. 4(a) is category "china", which is correct, without any edges connected. Fig. 4(b) is assigned *primary category* as "people" and two categories, "building" and "outdoor scene", are connected to the *primary category* with the corresponding semantics correlations 0.652 and 0.723, respectively. Based on the subjective observation, "building" is not relevant while the *primary category* "people" and the other connected category "outdoor scene" are. The *primary category* of Fig. 4(c) is "winter season" with connections to "building", "beach", "European town ", "mountain", "sea shore", and "vacation resort" categories in the α-semantics graph. Although the *primary category* "winter season" assigned to this image by the classification tree is not semantically relevant, there are 4 semantically relevant categories ("building", "European town", "sea shore", and "vacation resort") connected with the *primary category* ("winter season") this image is classified to. Thus, the retrieval accuracy is significantly improved by incorporating these categories into the fuzzy model described in Section 5.

To evaluate the effectiveness of the semantics correlation measurement and the fuzzy model for categories, we have

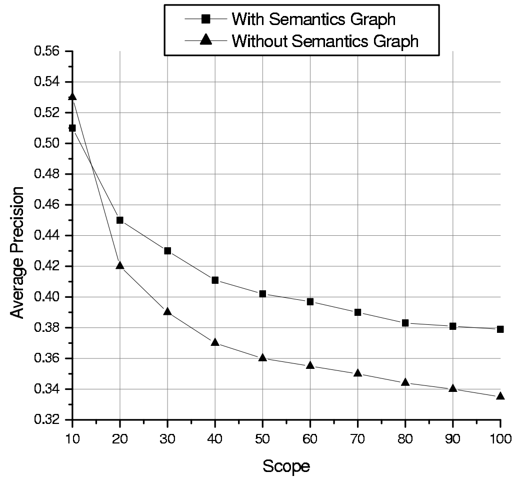


Figure 5: Average precision comparison with / without α-Semantics Graph.

Table 1: The classification and retrieval precision statistics.

	Average classification accuracy	Average retrieval precision
Correctly classified images	90.2%	67.2%
Incorrectly classified images	81.5%	34.3%

compared the retrieval precision with and without α-Semantics Graph. Fig. 5 shows the results. From the figure, it is evident that the α-Semantics Graph and the derived fuzzy model for categories improve the retrieval precision significantly. These results substantiate our motivations: by explicitly addressing the semantic uncertainty and the semantic overlap, the false positives can be substantially reduced for image retrieval.

To evaluate the influence of the performance of the classification tree to the image retrieval results, two evaluation statistics are recorded. They are:

Average classification accuracy: The average value of the classification accuracy for training images in all categories.

Average retrieval precision: The average ratio of relevant images in top 50 retrieved images for every query image.

The results are shown in Table 1.

Another advantage of our method is its high online query efficiency. In most state-of-the-art CBIR systems, the search is performed linearly. In other words, the computation complexity is $O(n)$ for an image collection with n images. In our method, the average computation complexity is $O(\log m)$ for image classification and $O(w)$ for image similarity computation, where m is the number of image categories and w is the average number of images in a category. Thus, the overall complexity is $O(\log m + w)$. 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 validated in the experiment. The observed average query time for returning top 30 images is less than 0.5 second.

7. CONCLUSIONS

In this paper, we have proposed a data mining approach to modeling relationships among categories in image collections. A semantics correlation based structure, called α-Semantics Graph, is proposed to represent the semantic uncertainty and the semantic overlap explicitly. Founded on the α-Semantics Graph, each semantic category 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. The model derived is utilized in the image retrieval application. A unique image retrieval algorithm is developed through integrating the classification results and the fuzzy model for each category. With the effective supervised learning applied to the image collection and the precise modeling of image semantic categories, the proposed methodology inaugurates a new generation of content-based image retrieval approaches, which aims at achieving more semantics-relevant performance.

8. REFERENCES

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

[2] S. F. Chang, J. R. Smith, M. Beigi, and A. Benitez. Visual information retrieval from large distributed online repositories. *Comm. ACM*, 40(2):63–67, 1997.

[3] W. Chang, G. Sheikholeslami, J. Wang, and A. Zhang. Data resource selection in distributed visual information systems. *IEEE Trans. on Knowledge and Data Engineering*, 10(6):926–946, Nov./Dec. 1998.

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

[6] M. F. et al. Query by image and video content: The qbic system. *IEEE Computer*, 28(9):23–32, September 1995.

[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] P. Lipson, E. Grimson, and P. Sinha. Configuration based scene classification and image indexing. In *The 16th IEEE Conf. on Compuer Vision and Pattern Recognition Proceedings*, pages 1007–1013, 1997.

[11] W. Y. Ma and B. S. Manjunath. A comparison of wavelet transform features for texture image annotation. In *Internation Conference on Image Processing*, pages 2256–2259, 1995.

[12] B. S. Manjunath and W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8), August 1996.

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

[14] 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.

[15] H. Yu and W. Wolf. Scenic classification methods for image and video databases. In *SPIE International Conference on Digital Image Storage and Archiving Systems*, volume 2606, pages 363–371, 1995.

[16] R. Zhang and Z. Zhang. Hidden semantic concept discovery in region based image retrieval. In *IEEE International Conference on Computer Vision and Pattern Recogntion (CVPR) 2004*, Washington, DC, June 2004.

[17] 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.