Region Based $\alpha$-Semantics Graph Driven Image Retrieval

Ruofei Zhang, Sandeep Khanzode, Zhongfei (Mark) Zhang
Department of Computer Science
State University of New York, Binghamton, NY 13902, U. S. A.

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 be unanticipated and referring to unextracted image content. 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 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, and the edge map is used through water filling algorithm [13] 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 region 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 becomes a binary image with the value of each pixel $i$ defined as $0$ if $\text{count}(i) \geq t$ ($\text{count}(i)$ 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 sparse connected objects in it disjointed. The size of the erosion mask is determined to be the minimum that makes two sparse connected objects separated. (iv) With 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.

**Distortion**. The distortion is a statistical measure to estimate the compactness degree of the repository. For each repository, \( r_i \), it is defined as

\[
D(r_i) = \frac{1}{N_i} \sum_{j=1}^{N_i} \sqrt{\frac{\sum_{l=1}^{n} \| f_j - c_i \|^2}{N_i}}
\]

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 homogeneity of the feature distributions in the repository \( r_i \). The more perplex in the repository, the bigger \( \varphi \) and vice versa.

**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 discrete value type in a limited domain. To build a classification tree, 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 with the \( \alpha \)-semantics graph follows.
3. Experiment Results

We have implemented the methodology in a prototype system. The evaluation had two parts. Both the classification and image retrieval performance were evaluated.

To provide quantitative evaluations on the performance of image classification, we ran the prototype on a controlled image database. This controlled 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 classification performance reliably with categorization accuracy because the repositories are semantically non-ambiguous and share no semantic overlaps.

The classification performance of the constructed classification tree was compared with the nearest-neighbor classification method (NN method) [1]. For both methods, 40 randomly chosen images for each repository were used to train the classifiers; the classification methods were then tested using the rest 600 images outside the training set. The classification results of our proposed method and raw feature based NN method are shown in Table 1. In the table, for each repository pair, the classification accuracy for every repository is isolated (with no edges connected to other nodes); for the remaining 50% of the COREL collection as the query set. The participation consisted 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.

The image retrieval evaluations were 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 were randomly selected from all repositories of the remaining 50% of the COREL collection as the query set. We invited a group of 5 users to participate the evaluations. The participants consisted 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 α must be decided for the α-semantics graph. For the extreme case α = 0, each node is connected to all other nodes in the 0-Semantics Graph (all repositories are treated as semantics-related to each other); for α = 1, each node is isolated (with no edges connected to other nodes), the 1-Semantics Graph degraded to a repository set. In the experiment we have calculated 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, was used as the α in the prototype.

To evaluate the effectiveness of the semantics correlation measure and the fuzzy model for repositories, we have compared the retrieval precision with and without α-Semantics Graph. Fig. 2 shows the results, in which it is evident that the α-Semantics Graph and derived fuzzy model for repos-
The average computation complexity is the number of images in the database. In our method, the average retrieval complexity is linearly (i.e., the retrieval complexity is $O(m)$, where $m$ is the number of image repositories and $n$ is the number of images in the database). Since $w = \frac{n}{m}$, the overall complexity is $O(\log m + \frac{w}{m})$. In general, $\frac{w}{m} << n$, 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.

To evaluate the influence of the classification accuracy of the classification tree on the training set to the image retrieval performance, we have recorded the statistics of the classifications and the corresponding retrieval precisions for the query image set. The results are shown in Table 2. It indicates that (i) for repositories with higher classification accuracy, a query image belonged to that repository is more likely to be correctly classified and (ii) correct classification of the query image gives a high retrieval precision, however, the incorrect classification of the query image still can return some relevant images due to the $\alpha$-Semantics Graph and the fuzzy model based retrieval algorithm.

Consider that it is difficult to design a fair comparison with existing very few classification-based image retrieval methods, we have computed 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 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 $O(\log n + \frac{n}{m})$ for image similarity calculation, 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, $\frac{n}{m} << n$, 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.

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


