Hidden Semantic Concept Discovery in Region Based Image Retrieval*

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

Abstract

This paper addresses Content Based Image Retrieval (CBIR), focusing on developing a hidden semantic concept discovery methodology to address effective semantics-intensive image retrieval. In our approach, each image in the database is segmented to regions associated with homogenous color, texture, and shape features. By exploiting regional statistical information in each image and employing a vector quantization method, a uniform and sparse region-based representation is achieved. With this representation a probabilistic model based on statistical-hidden-class assumptions of the image database is obtained, to which Expectation-Maximization (EM) technique is applied to analyze semantic concepts hidden in the database. An elaborated retrieval algorithm is designed to support the probabilistic model. The semantic similarity is measured through integrating the posterior probabilities of the transformed query image, as well as a constructed negative example, to the discovered semantic concepts. The proposed approach has a solid statistical foundation and the experimental evaluations on a database of 10,000 general-purposed images demonstrate its promise of the effectiveness.

1. Introduction

Large collection of images are becoming available to the public, from photo collections to Web pages or even video databases. To index or retrieve them is a challenge which is the focus of many research projects (for instance IBM's QBIC [5]). With near one decade research, it is found that content based image retrieval (CBIR) is a practical and satisfactory solution to the challenge. At the same time, it is also well known that the performance of CBIR is mainly limited by the gap between low-level features and high-level semantic concepts [13]. In order to reduce this gap, region based features (describing object level features), instead of raw features of whole image, to represent the visual content of an image is widely used [1, 15, 16, 6, 3].

One important issue significantly affecting success of a region-based CBIR methodology is how to compare two images, i. e., the definition of the image similarity measurement. A straightforward solution adopted by most early systems [1, 8] is to use individual region-to-region similarity as the basis of the comparison. Using such systems, the users are forced to select a limited number of regions from the query image in order to start a query session. As discussed in [15], due to the uncontrolled nature of visual contents in an image, automatically and precisely extracting image objects is still beyond the reach of the state-of-the-art in computer vision. Therefore, the above systems tend to partition one object into several regions with none of them being representative for the object. Consequently, it is often difficult for users to determine which regions should be used for retrieval.

To address this issue, several image-to-image similarity measurements that combine information from all of the regions have been proposed [15, 3]. Such systems only require the users to impose a query image, and therefore relieve the users from puzzling decisions. For example, the SIMPLIcity system [15] uses an integrated region matching as its image similarity measure. By allowing many-to-many relationship of the regions, the approach is robust to inaccurate segmentation.

Ideally what we strive to measure is the *semantic similarity*, which physically is very difficult to define, even to describe. All above methodologies did not explicitly connect the extracted features with the pursued semantics represented by the visual content. Though the defined region-to-region and/or image-to-image similarities attempt to approximate the semantic similarity, the approximation is heuristic and not reliable. Consequently their retrieval accuracies are limited. In this paper, we propose a probabilistic approach to addressing the hidden semantic concept discovery. A new region-based sparse but uniform image representation is developed, which facilitates the indexing scheme based on a region-image-concept probabilistic model with reasonable assumptions. This model has a solid

*This paper was accepted for oral presentation on the 2004 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'04).

statistical foundation and thus is appropriate for the objective of semantics-intensive image retrieval. To describe the semantic concepts hidden in the region and image distributions of a database, the Expectation-Maximization (EM) technique is used. With a derived iterative procedure, the posterior probabilities of each region in an image to hidden semantic concepts are quantitatively obtained, which act as the basis for the *semantic similarity* measure for image retrieval. The retrieval effectiveness is improved because the similarity measure is based on the discovered semantic concepts, which are more reliable than the existing region features proposed in the literature.

The rest of the paper is organized as follows: Section 2 describes the region feature extraction and image representation based on the image segmentation and visual dictionary generated with the feature grouping. In Section 3 the probabilistic region-image-concept model and the hidden semantic concepts discovery through EM technique are described. Section 4 presents the posterior probability based image similarity measure scheme and the supportive relevance feedback based retrieval algorithm as well as discussions on the proposed approach. The experiment results that evaluate the proposed approach against one state-of-the-art CBIR system in several aspects are reported in Section 5. Finally the paper is concluded in Section 6.

2. Region Based Image Representation

The query image and images in a database are first segmented into homogeneous regions. Then representative properties are extracted for every region by incorporating multiple semantics-related features, specifically, color, texture, and shape properties. The image segmentation and corresponding feature extraction method are similar to those employed in [15], which are shown to be effective. Finally the normalized feature vectors corresponding to color, texture, and shape properties, respectively, are stored to represent each region of all the images in the database.

Since region features $f \in \mathbb{R}^n$, it is necessary to perform regularization on the region property set so they can be indexed efficiently. Noting that many regions from different images are similar in terms of the features, vector quantization (VQ) techniques are used to group similar regions together. Consequently, we create a visual dictionary for region properties to represent the visual content of regions. There are three advantages of the visual dictionary. First, it improves retrieval robustness by tolerating minor variations among visual properties. Without the visual dictionary, because that very few feature values are shared by regions, we have to consider feature vectors of all regions. This manner makes it not effective to compare similarity among regions. However, based on the visual dictionary created, low-level features of regions are quantized such that images can be represented in a way resistant to perception uncertainties [3]. Second, the region-comparison efficiency is improved significantly by preventing from calculating distances between region features. Third, the utilization of visual dictionary reduces the storage space without sacrificing the accuracy.

We create the visual dictionary for region properties by applying Self-Organization Map (SOM)[7] learning. SOM is ideal for our problem as it projects high-dimensional feature vectors to a 2-dimensional plane through mapping similar features together while separating different features apart at the same time.

A procedure is designed to create “code words” in the dictionary. Each “code word” represents a set of visually similar regions. The procedure follows 4 steps:

1. Performing Batch SOM learning [7] algorithm on the region feature set to obtain the visualized model (node status) displayed on the 2-dimensional plane map.

2. 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 following:

   $$p(i) = \begin{cases} 0 & \text{if } count(i) \geq t \\ 255 & \text{else} \end{cases}$$

   where $count(i)$ is the number of features mapped to the node i and the constant t is a preset threshold. The pixel value 0 denotes objects while pixel value 255 denotes background.

3. Performing the morphological erosion operation [2] on the resulting image to make sparse connected objects in the image disjointed.

4. With connected component labeling [2] we assign each separated object a unique ID, a “code word”. For each “code word”, the mean of all the features associated to it is determined and stored. All “code words” constitute the visual dictionary for regional visual properties.

Each labeled component represents a region feature set in which the intra-distance is low. The extent of similarity in each “code word” is controlled by the parameters in the SOM algorithm and the threshold t. With the above procedure, the number of “code words” is adaptively determined and the similarity-based feature grouping is achieved. The experiments show that the visual dictionary created do capture the clustering characteristics existing in the feature set well. We note that the threshold t is highly correlated to the number of “code words” generated; it is determined empirically by balancing the efficiency and accuracy. We discuss

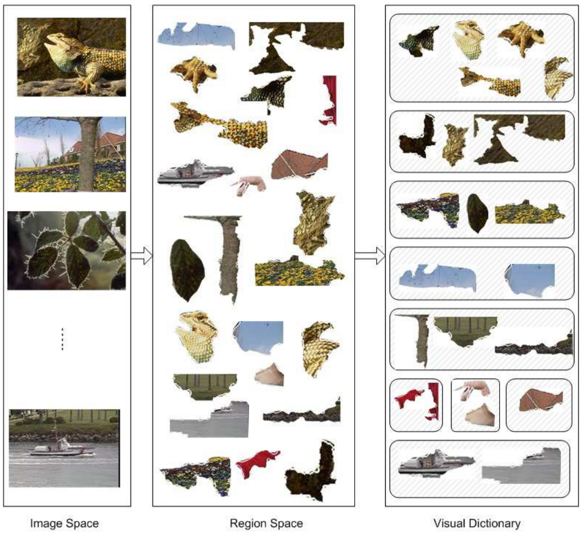

Figure 1: Generation of the visual dictionary.

the choosing of the number of "code words" in the visual dictionary in Section 5. Fig. 1 shows the generation of the visual dictionary. Each round-rectangle in the third column of the figure is one "code word" in the dictionary. In the rest of this paper, we use the terminologies region and "code word" interchangeable; they both denote an entry in the visual dictionary equally.

Based on the visual dictionary, each image can be represented in a uniform vector model. In this representation, an image is a vector with each dimension corresponding to a "code word". More formally, the uniform representation $\vec{I}_u$ of an image I is a vector $\vec{I}_u = \{w_1, w_2, \ldots, w_M\}$ where M is the number of "code words" in the visual dictionary. For a "code word" $C_i, 1 \leq i \leq M$, if there exists a region R_j of I that corresponds to it, then $w_i = W_{Rj}$ for $\vec{I}_u$, where W_{Rj} is the number of occurrence of R_j in the image I; otherwise, $w_i = 0$. This uniform representation is sparse, for an image usually contains few regions comparing with the number of "code words" in the visual dictionary. Based on this representation of every image, the database is modeled as a $M \times N$ "code word"-image matrix which records the occurrence of every "code word" in each image, where N is the number of images in the database.

3. Probabilistic Hidden Semantic Model

To achieve the automatic semantic concept discovery, a region-based probabilistic model is constructed for the image database with the representation by the "code word"-image matrix. The probabilistic model is analyzed by the EM technique to discover the latent semantic concepts, which act as a basis for effective image retrieval through comparing concept similarities among images.

With a uniform "code word" vector representation for each image in the database, we propose a probabilistic model. In this model, we assume the (region, image) are known i.i.d. samples from an unknown distribution. We assume these samples are associated with an unobserved *semantic concept* variable $z \in Z = \{z_1, \ldots, z_K\}$. Each observation of one region ("code word") $r \in R = \{r_1, \ldots, r_M\}$ in an image $g \in G = \{g_1, \ldots, g_N\}$ belongs to one concept class z_k. To simplify the model, we have two more assumptions. First, observation pairs (r_i, g_j) are generated independently. Second, the pairs of random variable (r_i, g_j) are conditionally independent given the respective hidden concept z_k, i. e., $P(r_i, g_j|z_k) = P(r_i|z_k)P(g_j|z_k)$. The region and image distribution are considered as a randomized data generation process described as follows:

- choose a concept with probability $P(z_k)$;
- select a region $r_i \in R$ with probability $P(r_i|z_k)$; and
- select an image $g_j \in G$ with probability $P(g_j|z_k)$.

As a result one obtains an observed pair (r_i, g_j), while the concept variable z_k is discarded.

Based on the theory of the generative model [9], the process is equivalent to the following way:

- select an image g_j with probability $P(g_j)$;
- pick a concept z_k with probability $P(z_k|g_j)$; and
- generate a region r_i with probability $P(r_i|z_k)$.

Translating this process into a joint probability model results in the expression

$$\begin{aligned} P(r_i, g_j) &= P(g_j)P(r_i|g_j) \\ &= P(g_j)\sum_{k=1}^{K} P(r_i|z_k)P(z_k|g_j) \end{aligned} \quad (1)$$

Inverting the conditional probability $P(z_k|g_j)$ in (1) with the application of the Bayes' rule results in

$$P(r_i, g_j) = \sum_{k=1}^{K} P(z_k)P(r_i|z_k)P(g_j|z_k) \quad (2)$$

Following the likelihood principle, one determines $P(z_k)$, $P(r_i|z_k)$, and $P(g_j|z_k)$ by maximization of the log-likelihood function

$$\mathcal{L} = \log P(R, G) = \sum_{i=1}^{M}\sum_{j=1}^{N} n(r_i, g_j)\log P(r_i, g_j) \quad (3)$$

where $n(r_i, g_j)$ denotes the number of regions r_i occurred in image g_j. From (3) and (1) we derive that the model is a statistical mixture model [9], which can be resolved by applying the EM technique [4]. The EM alternates in two steps: (i) an expectation (E) step where the posterior probabilities are computed for the hidden variable z_k, based on the current estimates of the parameters, (ii) an maximization (M) step, where parameters are updated to maximize the expectation of the complete-data likelihood $\log P(R, G, Z)$ given the posterior probabilities computed in the previous E-step.

Applying Bayes' rule with (1), we determine the posterior probability for z_k under (r_i, g_j):

$$P(z_k|r_i, g_j) = \frac{P(z_k)P(g_j|z_k)P(r_i|z_k)}{\sum_{k'=1}^{K} P(z_{k'})P(g_j|z_{k'})P(r_i|z_{k'})} \quad (4)$$

The expectation of the complete-data likelihood $\log P(R, G, Z)$ for the estimated $P(Z|R, G)$ derived from (4) is

$$\sum_{(i,j)=1}^{K} \sum_{i=1}^{M} \sum_{j=1}^{N} n(r_i, g_j) \log \left[P(z_{i,j})P(g_j|z_{i,j})P(r_i|z_{i,j})\right] P(Z|R, G) \quad (5)$$

where

$$P(Z|R, G) = \prod_{m=1}^{M} \prod_{n=1}^{N} P(z_{m,n}|r_m, g_n)$$

In (5) the notation $z_{i,j}$ is the concept variable that associates with the region-image pair (r_i, g_j). In other words, (r_i, g_j) belongs to concept z_t where $t = (i, j)$.

Maximizing (5) with Lagrange multipliers to $P(z_l)$, $P(r_u|z_l)$, and $P(g_v|z_l)$, respectively, under the following normalization constraints

$$\sum_{k=1}^{K} P(z_k) = 1, \sum_{k=1}^{K} P(z_k|r_i, g_j) = 1, \sum_{i=1}^{M} P(r_i|z_l) = 1 \quad (6)$$

for any r_i, g_j and z_l, the parameters are determined as

$$P(z_k) = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} n(r_i, g_j)P(z_k|r_i, g_j)}{\sum_{i=1}^{M} \sum_{j=1}^{N} n(r_i, g_j)} \quad (7)$$

$$P(r_u|z_l) = \frac{\sum_{j=1}^{N} n(r_u, g_j)P(z_l|r_u, g_j)}{\sum_{i=1}^{M} \sum_{j=1}^{N} n(r_i, g_j)P(z_l|r_i, g_j)} \quad (8)$$

$$P(g_v|z_l) = \frac{\sum_{i=1}^{M} n(r_i, g_v)P(z_l|r_i, g_v)}{\sum_{i=1}^{M} \sum_{j=1}^{N} n(r_i, g_j)P(z_l|r_i, g_j)} \quad (9)$$

Alternating (4) with (7)–(9) defines a convergent procedure that approaches a local maximum of the expectation in (5).

The number of concepts, K, must be determined in advance to proceed the EM model fitting. Ideally, we intend to choose the value of K that best agrees to the number of semantic classes in the database. One readily available notation of the fitting goodness is the log-likelihood. Given this indicator, we can apply the Minimum Description Length (MDL) principle [10] to select among values of K. This can be operationalized as follows [10]: choose K to maximize

$$\log(P(R, G)) - \frac{m_K}{2} \log(MN) \quad (10)$$

where the first term is expressed in (3) and m_K is the number of free parameters needed for a model with K mixture components. In our probabilistic model, we have

$$m_K = (K-1)+K(M-1)+K(N-1) = K(M+N-1)-1$$

As a consequence of this principle, when models with different values of K fit the data equally well, the simpler model is selected. For our experiment database, K is determined through maximizing (10).

4. Posterior Probability based Image Retrieval

Based on the probabilistic model, we derive the posterior probability of each image in the database to every discovered concept by applying Bayes' rule as

$$P(z_k|g_j) = \frac{P(g_j|z_k)P(z_k)}{P(g_j)} \quad (11)$$

which is determined with the estimations of (7)–(9). The posterior probability vector $P(Z|g_j) = [P(z_1|g_j), P(z_2|g_j), \ldots, P(z_K|g_j)]^T$ is used to quantitatively describe the semantic concepts associated with the image g_j. This vector is considered as a representation of g_j (which originally is represented in the M-dimensional "code word" space) in the K-dimensional *concept space* determined by the estimated $P(z_k|r_i, g_j)$ in (4).

For each query image, after obtaining the corresponding "code words" we attain its representation in the discovered concept space by plugging it in the EM iteration derived in Section 3. The only difference is that the $P(r_i|z_k)$ and $P(z_k)$ are fixed to be the values we have obtained with the whole database modeling (which are obtained in the indexing phase, i. e., determining the concept space representation of every image in the database).

With the proposed probabilistic model, it is able to concurrently obtain $P(z_k|r_i)$ and $P(z_k|g_j)$ such that both regions and images have an interpretation in the concept space simultaneously, while image clustering based approaches, e. g. [6], do not have this flexibility. Every region and/or image can be represented as a weighted sum of the discovered concept axes. In this aspect, the model acts like a factoring analysis [9] but the model offers important advantages, e. g., each weight has a clear probabilistic meaning and the factoring are two folded.

In designing a region-based image retrieval methodology, we note two characteristics of the region representations existing:

1. The number of segmented regions in one image is normally small.

2. Not all regions in one image are semantically relevant, some are unrelated or even non-relevant; which regions are (ir)relevant depends on user's querying subjectivity.

Incorporating the "code words" corresponding to unrelated or non-relevant regions hurts the retrieval accuracy because the occurrence of these regions in an image "fools" the probabilistic model such that erroneous concept representations would be generated. To address the two characteristics in image retrieval, we employ relevance feedback for the similarity measurement in the concept space. Relevance feedback has been demonstrated great potential to capture users' querying subjectivity both in text retrieval and image retrieval [14, 12]. A retrieval algorithm is designed to integrate with the probabilistic model to deliver a better performance.

In the algorithm, we move the query point in the "code word" space toward good example points (relevant images labeled by the user) and away from bad example points (irrelevant images labeled by the user) such that the region representation has more support to the probabilistic model. At the same time, the query point is expanded with "code words" of labeled relevant ones. On the other hand, we construct a negative example "code word" vector by applying the similar moving strategy such that it lies near bad example points and away from good example points. The vector moving strategy uses form of the Rocchio's formula [11]. The Rocchio's formula for relevance feedback and feature expansion has proven to be one of the best iterative optimization technique in the field of information retrieval. With the modified query vector pos and a constructed negative example neg, their representations in the discovered concept space are obtained and their similarities to each image in the database are measured through cosine metric of the corresponding vectors in the concept space, respectively. Then the images are ranked based on the similarity to pos as well as the dissimilarity to neg. The algorithm is shown in Algorithm 1.

The parameters α, β, and γ in the algorithm are assigned a value of 1.0 in our current implementation of the system for the sake of simplicity. However, other values may be given to emphasize the different weights between the good sample points and bad sample points.

5. Experiment Results

We have implemented the approach in a prototype system on a platform of Pentium IV 2.0 GHZ CPU and 256M memory. The following reported evaluations are performed on a general-purpose color image database containing 10,000 images from the COREL collection with 96 semantic categories. These categories include "landscape", "fashion", "historical building", "city life", etc. Each semantic category consists of 85–120 images. To evaluate the image retrieval performance, 1,500 images are randomly selected from all the categories as the query set. The relevancy of the retrieved images are subjectively examined by users. Unless otherwise noted, the default results of the experiments are average of the top 30 returned images for each query using the 1,500 queries.

input : q, "code word" vector of the query image
output : Images retrieved for the query image q
begin
  $rs = \{rel_1, rel_2, \ldots, rel_a\}$, where rel_i is "code word" vector of each image labeled as relevant;
  $is = \{ire_1, ire_2, \ldots, ire_b\}$, where ire_j is "code word" vector of each image labeled as irrelevant;
  $pos = \alpha q + \beta(\frac{1}{a}\sum_{i=1}^{a} rel_i) - \gamma(\frac{1}{b}\sum_{j=1}^{b} ire_j)$;
  $neg = \alpha\sum_{j=1}^{b} ire_j + \beta(\frac{1}{b}\sum_{j=1}^{b} ire_j) - \gamma(\frac{1}{a}\sum_{i=1}^{a} rel_i)$;
  for $k = 1$ *to* K **do**
    determine $P(z_k|pos)$ and $P(z_k|neg)$ with EM and (11);
  end
  $n = 1$;
  while $n <= N$ **do**
    $sim1(g_n) = \frac{P(Z|pos) \bullet P(Z|g_n)}{\|P(Z|pos)\|\|P(Z|g_n)\|}$;
    $sim2(g_n) = \frac{P(Z|neg) \bullet P(Z|g_n)}{\|P(Z|neg)\|\|P(Z|g_n)\|}$;
    if $(sim1(g_n) > sim2(g_n))$ **then**
      $sim(g_n) = sim1(g_n) - sim2(g_n)$;
    else
      $sim(g_n) = 0$;
    end
    rank the images in the database based on $sim(g_n)$;
  end
end

Algorithm 1: Retrieval Algorithm.

In the experiment, the parameters of the image segmentation algorithm [15] used is adjusted with considering the balance of the depiction detail and the computation intensity such that there are an average 8.3207 regions in each image. To determine the size of the visual dictionary, different numbers of "code words" have been selected and evaluated. The average precision (without query expansion and moving strategy) within the top 20 (30, 50) images, denoted as P(20) (P(30), P(50)), are shown in Fig. 2. It is indicated that the general trend is that the larger the visual dictionary size, the higher the retrieval accuracy. However, a larger

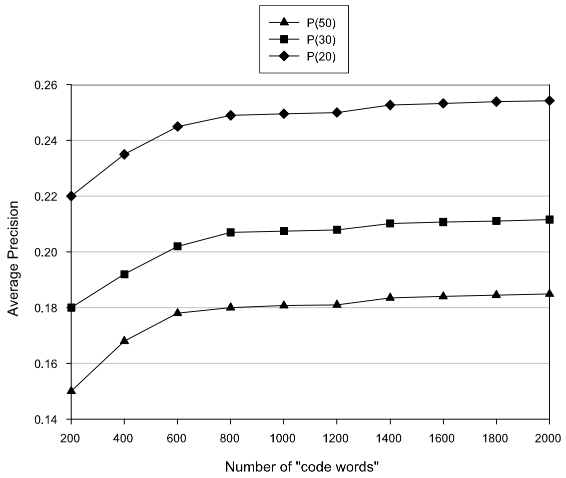

Figure 2: Average precision (without query expansion and movement) for different sizes of the visual dictionary.

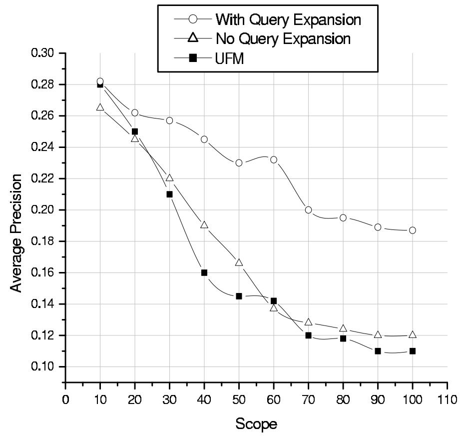

Figure 3: Average precision comparisons between two versions of our approach and UFM.

visual dictionary size means larger number of image feature vectors, which implies higher computation complexity in the hidden semantic concept discovery. Also, a larger visual dictionary leads to a larger storage space. Therefore, we use 800 as the number of the "code words", which corresponds to the first turning point in Fig. 2. Since there are total 83,307 regions in the database, in average each "code word" represents 104.13 regions.

Applying the method of estimating the number of hidden concepts described, the number of the concepts is determined to be 132. Performing the EM model fitting, we obtain the conditional probability of each "code word" to every concept, i. e., $P(r_i|z_k)$. Manually checking the visual contents of the region sets corresponding to the top 10 highest "code words" in every semantic concept, we observe that these discovered concepts indicate semantic interpretations, such as "people", "building", "outdoor scenery", "plant", "automotive race".

In terms of computational complexity, despite of the iterative nature of EM, the computing time for the model fitting at $K = 132$ is acceptable (less than 1 second). The average number of iterations upon convergence for an image is less than 5.

To show the effectiveness of the probabilistic model in image retrieval, we have compared the accuracy of our method with that of UFM [3]. UFM is a method based on fuzzified region representation to build region-to-region similarity measures for image retrieval. We compare our approach with UFM because it is available to us and reflects the performance of the state-of-the-art CBIR systems. In addition, since the same image segmentation and feature extraction methods are used in UFM as in ours, a fair comparison of the performance is expected.

The systematic comparison results on the 1,500 query image set are shown in Fig. 3. Two versions of our approach (one with query expansion and movement and another without) and UFM are compared. It is shown that the performance of our probabilistic model has higher overall precision than UFM and the query expansion and movement with interaction of constructed negative example boost the retrieval accuracy significantly.

6. Conclusions

The main contributions of this work are the identification of the problem existing in most region-based CBIR methods —- unreliable region evidence in semantic content, and the development of a promising hidden semantic concept discovery technique to solve for the problem. Through performing image segmentation with multiple features and developing a SOM based quantization method to generate a visual dictionary, a uniform and sparse region-based representation scheme is obtained. On the basis of this representation a probabilistic model of the image database is defined. The model assumes that the regions, hidden semantic concepts, and images are random variables and the objective is to discover concept distributions with samples from the (region, image) distributions. Based on this model, the EM method is applied to derive an iterative procedure to discover the hidden semantic concepts in the database. An elaborated relevance feedback based retrieval algorithm is designed to support the model and improve the retrieval accuracy. The image querying is performed by integrating the posterior probabilities of the transformed images to discovered semantic concepts. Supported by the solid statistical foundation, this approach enables a retrieval by higher order semantic indicants which are more reliable, hence improves the retrieval accuracy. The experimental evaluations on a database of 10,000 general-purpose images demonstrate the promising effectiveness of the approach in image retrieval.

References

[1] C. Carson, M. Thomas, and e. a. S. Belongie. Blobworld: A system for region-based image indexing and retrieval. In *the 3rd Int'l Conf. on Visual Information System Proceedings*, pages 509–516, Amsterdam, Netherlands, June 1999.

[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] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of Royal Statistical Society*, 39(1):1–38, 1977.

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

[6] F. Jing, M. Li, H. Zhang, and B. Zhang. An efficient region-based image retrieval framework. In *ACM Multimedia 2002*, Juan-les-Pins, France, December 2002.

[7] 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.

[8] W. Y. Ma and B. Manjunath. Netra: A toolbox for navigating large image databases. In *IEEE Int'l Conf. on Image Processing Proceedings*, pages 568–571, Santa Barbara, CA, 1997.

[9] G. Mclachlan and K. E. Basford. *Mixture Models*. Marcel Dekker, Inc., Basel, NY, 1988.

[10] J. Rissanen. *Stochastic Complexity in Statistical Inquiry*. World Scientific, 1989.

[11] J. J. J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retreival System — Experiments in Automatic Document Processing*, pages 313–323. Prentice Hall, Inc, Englewood Cliffs, NJ, 1971.

[12] Y. Rui, T. S. Huang, S. Mehrotra, and M. Ortega. A relevance feedback architecture in content-based multimedia information retrieval systems. In *IEEE Workshop on Content-based Access of Image and Video Libraries, in conjunction with CVPR'97*, pages 82–89, June 1997.

[13] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. on PAMI*, 22:1349–1380, 2000.

[14] N. Vasconcelos and A. Lippman. Bayesian relevance feedback for content-based image retrieval. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, Hilton Head, South Carolina, June 2000.

[15] J. Z. Wang, J. Li, and gio Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Trans. on PAMI*, 23(9), September 2001.

[16] R. Zhang and Z. Zhang. Addressing cbir efficiency, effectiveness, and retrieval subjectivity simultaneously. In *the 5th ACM Int'l Workshop on Multimedia Information Retrieval*, Berkeley, CA, November 2003. in conjunction with ACM Multimedia (ACM MM) 2003.