Semantic-Based Analysis and Retrieval Techniques for Geographic Image Data

Rong Zhao
Department of Computer Science
State University of New York
Stony Brook, NY 11794-4400, USA
rzhao@cs.sunysb.edu

Abstract

In this study we present the results of our work that seeks to find the hidden correlation, and thus, to negotiate the gap, between low-level features and high-level concepts in image contents. We introduce two visual feature indexing techniques, structured color texture and color anglogram, to represent image contents. Our study also involves a dimension reduction technique called latent semantic indexing which has been used in textual information retrieval for years. Experimental results on sample image sets show that these techniques are able to extract the underlying semantic structure of image contents. We intend to apply these techniques to geographic image data, hoping to generate more meaningful metadata models and to improve image annotations.

1. Introduction

Geographic image data is available and has been used in many different application domains such as homeland security, urban planning, water resource management, transportation monitoring, etc. As in the old saying, “a picture is worth a thousand words”. Besides, the sets of images available in databases and on the Internet have been growing dramatically. The sheer volume of image data presents a daunting challenge for those who try to develop techniques to access these data. In many cases, for example, in the case of homeland security, analysis and retrieval of image data has to be both highly accurate and highly efficient. Content-based image retrieval systems normally conduct indexing and searching based on low-level visual features, while users usually have a more abstract and conceptual notion of what they are looking for. Using low-level features to correspond to high-level abstractions is one aspect of the semantic gap between content-based analysis and retrieval methods and the concept-based users. To negotiate this gap, we introduce two visual feature indexing techniques, structured color texture and color anglogram, to represent meaningful image contents. Our study also involves a dimension reduction technique called latent semantic indexing which has been used in textual information retrieval for years. Experimental results on sample image sets show that these techniques are able to extract the underlying semantic structure of image contents. We intend to apply these techniques to geographic image data, hoping to generate more meaningful metadata models and to improve image annotations.

2. Structured Color Texture

We propose a structured color texture indexing technique, in which combinations of color transitions, either random or structured, are used to represent the visual contents of an image. Dominant color transitions are extracted and their properties such as edge directions at the transitions, inter-transition distance, and frequency of occurrence are computed. Studies have shown that basic textural patterns such as coarseness, contrast, and directionality possess good correspondence with human perception. In our approach, color transitions represent contrast; inter-transition distance corresponds to coarseness; and edge direction relates to directionality. A major advantage of this approach is that it makes no assumption that the entire image possesses only one homogeneous texture pattern. This approach also allows for partial image matching, which can be used for achieving object extraction.

Color texture patterns are global statistical features in the image contents, and thus, they cannot provide information about whether or not all the transitions result from the same object. Therefore, misleading results may be generated if two dramatically different texture patterns happen to share the same color transition statistically. To solve this problem, we extend the color texture indexing method by adding structural analysis of both the scale and gradient (i.e., directionality) components of those textural
features. For each color transition, various types of structures and their properties can be extracted and analyzed. Preliminary results show that this approach provides very promising performance with regard to recall and precision.

3. Color Anglogram

Spatial layout of a set of points can be coded through an anglogram that is computed by discretizing and counting the angles produced by the Delaunay triangulation of these points. Anglograms encode the spatial correlation of these feature points and are invariant to translation, scale, and rotation.

To construct color anglograms, each image is decomposed into a number of non-overlapping blocks. Each individual block is abstracted as a unique feature point labeled with its spatial location and feature values. The feature values can be dominant or average hue and saturation values in the corresponding block. For each set of feature points labeled with a particular value, a Delaunay triangulation is constructed and the feature point histogram is computed by discretizing and counting the number of either the two largest angles or the two smallest angles in the Delaunay triangles. Finally, the image will be indexed by using color anglogram, i.e., the concatenated feature point histograms for each feature value.

4. Latent Semantic Indexing

Latent Semantic Indexing (LSI) was introduced to overcome a fundamental problem that plagues existing textual information retrieval techniques. The problem is that users want to retrieve documents on the basis of conceptual content, while individual keywords provide unreliable evidence about the conceptual meaning of a document. There are usually many ways to express a given concept and users tend to have different preference of picking terms. This is called synonymy and the prevalence of synonyms tends to decrease the recall performance. On the other hand, most words have multiple meanings and are used in different contexts. Hence, the terms in a query may literally match the terms in documents that are not of any interest to the user at all. In information retrieval this is addressed as polysemy which may lead to poor precision performance.

LSI makes use of Singular Value Decomposition (SVD) which is a dimension reduction method allowing the arrangement of the space to reflect the major associative patterns in the data, and ignore the smaller, less important influences. As a result, terms that did not actually appear in a document may end up close to the document in the transformed space, if that is consistent with the major patterns of association in the data. Retrieval proceeds by using the terms in a query to identify a point in the semantic space, and documents in its neighborhood are returned as relevant results to the query.

5. Experiments

In our experiment, we compared the performance of our color anglogram approach with that of the color histogram method, due to the fact that color histogram provides one of the best performances among existing techniques. Each image is converted into the HSV color space. For each pixel of the image, hue and saturation are extracted and each quantized into a 10-bin histogram. Then, the two histograms $h$ and $s$ are combined into one $h \times s$ histogram with 100 bins, which is the representing feature vector of each image. This is a vector of 100 elements, $F = [f_1, f_2, f_3, \ldots, f_{100}]^T$.

To apply the latent semantic indexing technique, a feature-image matrix, $A = [F_1, \ldots, F_n]$, where $n$ is the total number of images in the data set, is constructed using the feature vector of each image. Each row corresponds to one of the feature elements and each column is the entire feature vector of the corresponding image. Singular Value Decomposition is performed on the feature-image matrix. The result comprises three matrices, $U$, $\Sigma$, and $V$, where $A = UV^T$. The dimensions of $U$, $\Sigma$, and $V$ are $100 \times 100$, $100 \times n$, and $n \times n$, respectively. For our data set, the total number of images, $n$, is greater than 100. To reduce the dimensionality of the transformed space, we use a rank-$k$ approximation, $A_k$, of the matrix $A$, where $k = 12$. This is defined by $A_k = U_k \Sigma_k V_k^T$. The dimension of $A_k$ is the same as $A$, 100 by $n$. The dimensions of $U_k$, $\Sigma_k$, and $V_k$ are $100 \times 12$, $12 \times 12$, and $n \times 12$, respectively.

The following normalization process will assign equal emphasis to each image of the feature vector. For the feature-image matrix $A=[V_1, V_2, \ldots, V_n]$, we have $A_i$, which is the $i^{th}$ component in vector $V_i$. Assuming a Gaussian distribution, we can obtain the mean, $\mu$, and standard deviation, $\sigma$, for the $i^{th}$ component of the feature vector. Then we normalize the feature-image matrix into the range of [-1,1] by using,

$$A_i = \frac{A_i - \mu_i}{\sigma_i}$$

It can easily be shown that the probability of an entry falling into the range of [-1, 1] is 68%. In practice, we map all the entries into the range of [-1, 1] by forcing the out-of-range values to be either -1 or 1, whichever is the nearest. Then we shift the entries into the range of [0, 1] by using,

$$A_i = \frac{A_i + 1}{2}.$$
After this normalization process, each component of the feature-image matrix is a value between 0 and 1, and thus will not bias the importance of any component in the computation of similarity.

Applying different weights to different components has been a common practice for improving retrieval performance. Both global weight and local weight are considered. A global weight indicates the overall importance of that component in the feature vector across all the images. Therefore, the same global weighting is applied to an entire row of the matrix. A local weight is applied to each element indicating the relative importance of the component within its vector. The value for any component \( A_{i,j} \) is thus \( L(i,j)G(i) \), where \( L(i,j) \) is the local weighting for feature component \( i \) in image \( j \), and \( G(i) \) is the global weighting for that component.

Common local weighting techniques include term frequency, binary, and log of term frequency, whereas common global weighting methods include normal, gfidf, idf, and entropy. Based on previous research, it has been found that \( \log(1 + \text{term frequency}) \) helps to dampen effects of large differences in frequency and thus has the best performance as a local weight, whereas entropy is the appropriate method for global weighting. The entropy method is defined as,

\[
1 + \sum \frac{p_{ij} \log(p_{ij})}{\log(\text{number of documents})}
\]

where,

\[
p_{ij} = \frac{tf_{ij}}{gf_{i}}
\]

is the probability of that component, \( tf_{ij} \) is the raw frequency of component \( A_{i,j} \), and \( gf_{i} \) is the global frequency, i.e., the total number of times that component \( i \) occurs in all the images.

The global weights give less emphasis to those components that occur frequently or in many images. Theoretically, the entropy method is the most sophisticated weighting scheme, taking the distribution property of feature components over the set of all the images into account.

The measures of recall and precision are used in evaluating the shot detection performance. First, we applied color histogram and evaluated it with and without latent semantic indexing. It can be noticed that better performance is achieved by integrating color histogram with latent semantic indexing. This validates our beliefs that LSI can help discover the correlation between visual features and high-level concepts.

For our experiments with color anglogram, we divide each image into 64 blocks and compute the average hue value and average saturation value of each block. The average hue values are quantized into 10 bins, so are the average saturation values. For each quantized hue (saturation) value, we apply Delaunay triangulation on the point feature map. We count the two largest angles of each triangle in the triangulation, and categorize them into a number of anglogram bins each of which is \( 5^\circ \). Our vector representation of a image thus has 720 elements: 36 bins for each of the 10 hue values and 36 bins for each of the 10 saturation values. In this case, for each video clip the dimension of its feature-image matrix is \( 720 \times n \), where \( n \) is the total number of images. As is discussed above, we reduce the dimensionality of the feature-image matrix to \( k = 12 \). Based on the experimental results of our previous studies, we notice that normalization and weighting has a negative impact on the performance of color anglogram. Therefore, we do not apply normalization and weighting on the elements in the feature-image matrix.

From the results we notice that color anglogram achieves better performance than color histogram. We also notice that the best performance of both recall and precision is obtained by integrating color anglogram with latent semantic indexing.

### 6. Conclusions

In this paper, we presented some of the work that we have conducted to bridge the semantic gap in content-based image retrieval. We proposed a structured color texture indexing approach which integrates texture, color, and structural information into a more meaningful representation of image contents. We introduced a novel technique for spatial color indexing, color anglogram, which is invariant to rotation, scaling, and translation. We also experimented with a dimension reduction technique, latent semantic indexing, to explore semantic correlations between low-level features and high-level concepts.

Structured color texture indexing provides us with some very promising recall and precision results. Experimental results show that color anglogram is fairly accurate in capturing and emphasizing meaningful spatial color features in the image contents. Results show that latent semantic indexing can not only reduce the dimensionality of image data but also correlate visual features to construct higher-level semantic clusters.

Within the next two months we intend to conduct more experiments with the structured color texture technique, due to the fact that texture patterns are widely used in analysis and retrieval of geographic image data. We also plan to combine the three techniques and to apply such an integrated analysis/retrieval framework to several geographic image data sets. How to use these techniques to improve image retrieval will also be analyzed. Detailed comparison of experimental results and more extensive discussions with regard to geographic image data will be included in the full version of this paper and will be presented on the workshop.