감성과학, Vol. 7, No. 4, pp. 57-65. December 2004

Integrating Color, Texture and Edge Features for Content-Based Image Retrieval

내용기반 이미지 검색을 위한 색상, 텍스처, 에지 기능의 통합

Ming Ma* · Dong-Won Park*†

마명* · 박동원*†

Dept. of Information and Communications, PaiChai University*
배재대학교 정보통신공학과

Abstract : In this paper, we present a hybrid approach which incorporates color, texture and shape in content-based image retrieval. Colors in each image are clustered into a small number of representative colors. The feature descriptor consists of the representative colors and their percentages in the image. A similarity measure similar to the cumulative color histogram distance measure is defined for this descriptor. The co-occurrence matrix as a statistical method is used for texture analysis. An optimal set of five statistical functions are extracted from the co-occurrence matrix of each image, in order to render the feature vector for eachimage maximally informative. The edge information captured within edge histograms is extracted after a pre-processing phase that performs color transformation, quantization, and filtering. The features where thus extracted and stored within feature vectors and were later compared with an intersection-based method. The content-based retrieval system is tested to be effective in terms of retrieval and scalability through experimental results and precision-recall analysis.

Key words : Content-based Retrieval, Similarity Measure, Color Clustering, Edge Histogram

요약 : 본 논문에서는 color, texture, shape의 정보를 통합 이용하여 내용기반 영상검색 시스템의 성능을 향상시키는 기법을 고찰하였다. 먼저 영상에 내재되어 있는 color를 분석 추출하여 몇 개의 대표색으로 요약 표현한 다음, 이를 활용한 근사치 측정도를 고안하였다. Texture 정보 분석에 있어서는 영상의 주축 행렬 데이터를 통계적 접근 방법으로 추출하였다. Edge 분석의 방법으로는 Edge 막대그래프에서 색상변환, 양자화, 필터링에 관련된 정보를 선행처리 후 Edge 정보를 추출하였다. 마지막으로, 본 연구의 결과인 내용기반 영상검색 시스템의 효율성을 precision-recall 분석과 실험적 결과를 통하여 입증하였다.

주제어 : 내용기반 검색, 근사치 측정, 칼라 통합, 에지 히스토그램

† 교신저자 : 박동원(배재대학교 정보통신공학과)
E-mail : dwpark@pcu.ac.kr
TEL : 042-520-5639, 016-467-5639
FAX : 042-520-5752

1. Introduction

In recent years, digital images are playing an important role in depicting and disseminating pictorial information, because it is a convenient media for displaying and storing spatial, temporal, spectral and physical information contained in a variety of domains (e.g., satellite images, biomedical images). As a result, large image databases are being created and used in a number of applications, including criminal identification, multimedia encyclopedia, geographic information systems, online applications for art and art history, medical image archives, and trademarks database. These databases typically consist of thousands of images, taking up gigabytes of memory space. While advances in image compression algorithms have alleviated the storage requirement to some extent, the large volume of these images makes it difficult for a user to browse through the entire database. Thus there is a requirement for an efficient and automatic method for indexing and retrieving image from databases [4]. The traditional method of retrieving images is by manually annotated keywords (text based). But it has two main disadvantages. Firstly, it is labor intensive and therefore time consuming and expensive. Secondly, the rich semantics of an image is difficult to be precisely described and different people may describe the same image in different ways. To overcome the drawbacks of the textbased approach, the content based image retrieval (CBIR) approach proposes a technique for retrieving images directly and automatically based on their visual contents such as color, texture, and shape. In a typical content based image retrieval system, the query pattern is query by example, which searches the top N images similar to an example image. Before the retrieval, the visual features are extracted from all images in an image database offline. During the retrieval, the visual features of the example image are compared to those of all images in the image database and the top N images are returned as the query result [9].

Most of the recent work in image database retrieval has concentrated on developing a single concise feature like color, shape, or texture. Although color seems to be a highly reliable attribute for image retrieval, situations where color information is not present in the images require the use of shape and/or texture attributes for image retrieval. Retrieval based on a single image attribute might lack sufficient discriminatory information and might not be able to accommodate large scale and orientation changes. Recently, a number of studies have been carried out which combine the various features for efficient and effective querying by image content [4]. An attempt has been made in this paper to integrate the image representation on the basis of color, texture and edge to improve the retrieval performance.

In section 2 of this paper, the content based retrieval system that we developed is presented. The implemented features and a general framework for proper extensions of the system were also discussed. The performance experiments of the content based retrieval system are given in section 3 and finally, section 4 concludes the paper.

2. Feature Extraction in Our CBIR System

2.1 Overview of Our CBIR System

Our CBIR system offers retrieval by any combination of color, texture or edge features and image query is specifying an example query image. As it is illustrated in Figure 2.1, the system extracts and stores color, texture and edge features from each image added to the database. During the search period, the system matches appropriate features from query and stored images, calculates a similarity score between the query and each stored image examined, and displays the most similar images on the screen as thumbnails.

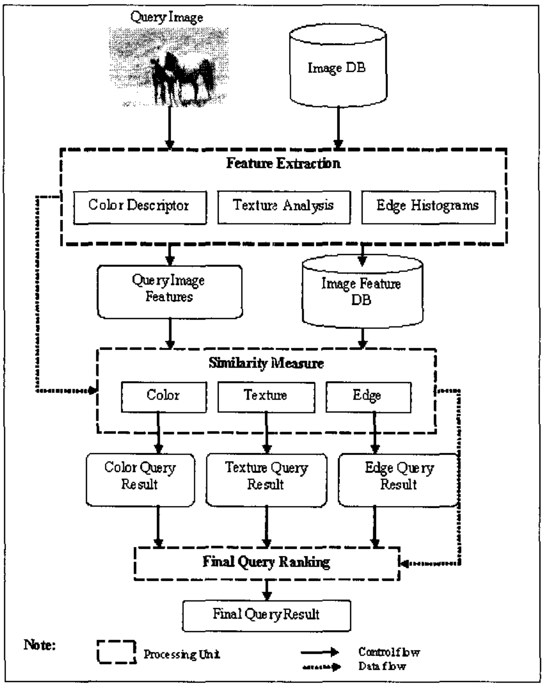


Figure 2.1 The architecture of our CBIR system

2.2 Feature Extraction

2.2.1 Color Descriptor

The color feature extraction first starts with color model transformation. Each pixel in an image has a three dimensional color vector and different color space approaches exist to represent color information. One of these color space is the hardware oriented Red Green Blue Model (RGB), where the color vector of a pixel p is the compound of red, green and blue channels $V_p = (r, g, b)$. Another color space model is the Hue Saturation Intensity Model (HSI) that is based on color descriptions rather than individual color components $V_p = (h, s, i)$. The RGB model has a major drawback: it is not perceptually uniform. Therefore, most of the systems use color space models (such as HIS) other than RGB. Color clustering is performed on each image to obtain its representative colors.

After clustering, only a small number of colors remain and the percentages of these colors are calculated. Each representative color and its corresponding percentage form a pair of attributes that describes the color characteristics in an image. The dominant color descriptor is defined to be

$$F = \{\{c_i, p_i\}, \ i = 1, \ldots, N\} \qquad (2.1)$$

Where N is the total number of color clusters in the image region, c_i is a color vector, p_i is its percentage, and $\sum_i p_i = 1$.

The procedure for color clustering is shown in Figure 2.2. Firstly, we convert the RGB color model into HIS model using the MTM transformer formulas.

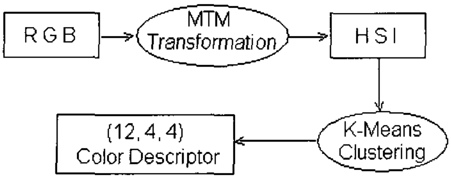


Figure 2.2 Procedure diagram for color clustering

Secondly, an improved K Means clustering algorithm proposed by Alsabti et al.[6] is used on these HSI values to get representative colors and their percentages in the image[7].

The efficient K Means algorithm is proved able to enhance the computational speed of the direct K Means algorithm by an order to two orders of magnitude in the total number of distance calculations and the overall time of computation. This has been shown to be effective in producing good clustering results for many practical applications. The efficient K Means algorithm consists of the following steps:

1. A k-d tree is generated to organize the pattern vectors, by which one can find all the patterns that are closest to a given prototype efficiently. There are some information kept by each node of the tree:

   a. The number of points (m);

   b. The linear sum of the points ($\overline{LS}$), i.e. $\Sigma_{i=1}^{m} \overline{P_i}$

   c. The square sum of the points (SS), i.e. $\Sigma_{i=1}^{m} \overline{P_i}^2$

2. The initial prototypes are derived as in the direct K Means algorithm.
3. A number of iterations are performed as in the direct K Means algorithm until the termination condition is met. And the number of points $C_n{}^i$, the linear sum of the points $\overline{C_{LS}^i}$ and the square sum of the point $\overline{C_{SS}^i}$ are maintained for each cluster I.

During each iteration, we traverse the k d tree from the root node with all k candidate prototypes. A pruning function is in turn applied on the candidate's prototypes. The traversal stopped when the number of candidate prototypes is equal to one. The cluster statistics are updated based on the information about the number of points, linear sum and square sum stored for the internal node. A direct K Means algorithm is applied on the leaf node if there is more than one candidate prototype.

In addition, because the human visual system is more sensitive to hues than to saturation or intensity, the H axis is quantized more finely than the S axis and the I axis. In our experiments, we cluster the HSI color space into 12 bins for hue, 4 bins for saturation, and 4 bins for intensity.

2.2.2 Texture Analysis

From the statistical point of view, an image is a complicated pattern on which statistics can be obtained to characterize these patterns. A most popularly use of statistical methods is the co occurrence matrix. The method roughly consists of constructing matrices by counting the number of occurrences of pixel pairs of given intensities at a given displacement.

As presented by Haralick [8], 14 statistical functions can be used for the textural content of images. However, in many applications, appropriate subsets of these functions are enough. The compact use of these statistical functions provides efficiency in terms of computational complexity. Thus, the following five statistical functions are extracted in the system because of their biggest discriminatory power [1][2]:

- **Energy** : the energy of a texture describes the uniformity of the texture. In a homogeneous image there are very few dominant grey tone transitions, so the co-occurrence matrix of this image will have fewer entries of large magnitude. Hence the energy of an image is high when the image is homogeneous(equation 2.2).

- **Entropy** : measures the randomness of the elements of the matrix, when all elements of the matrix are maximally random entropy has its highest value. So, a homogeneous image has a lower entropy than an inhomogeneous image. In fact, when energy gets higher, entropy should get lower(equation 2.3).

$$\sum_{i,j} p(i,\ j)^2 \qquad (2.2)$$

$$-\sum_{i,j} p(i,\ j)\ \log(p(i,\ j)) \qquad (2.3)$$

- **Inverse Difference Moment** : the inverse difference moment, has a relatively high value when the high values of the matrix are near the main diagonal. This is because the squared difference $(i-j)^2$ is then smaller, which increases the value of $\frac{1}{1+(i-j)^2}$ (equation 2.4)

- **Contrast** : It is also called inertia giving the opposite effect with inverse difference moment, when the high values of the matrix are further away from the main diagonal; the value of inertia becomes higher. So inertia and the inverse difference moment are measures for the distribution of grey scales in the image(Figure 5.4). (equation 2.5)

$$\sum_{i,j} \frac{1}{1+(i-j)^2}\, p(i,\ j) \qquad (2.4)$$

$$\sum_{n} n^2 \sum_{i,j|i-j|=n} p(i,\cdot j) \qquad (2.5)$$

- **Correlation** : the correlation feature measures the correlation between the elements of the matrix. When correlation is high the image will be more complex than when correlation is low(equation 2.6).

$$\frac{\sum_{i,j} ijp(i,\ j) - \mu_x\mu_y}{\sigma_x\sigma_y} \qquad (2.6)$$

Where μ_x, μ_y, σ_x, σ_y are the means and standard deviations of $\sum_k p(i,\ k)$ and $p_x(i) = \sum_k p(k,\ j)$.

In this study, the co occurrence matrix and the five statistical functions were implemented in our system. An asymmetric method was used implementing the co-occurrence matrix. The co-occurrence matrix were calculated for 0, 45, 90, and 135. Next, the results were averaged to get the final features of an image. The matrices were only calculated with distance 1.

2.2.3 Edge Histogram

As mentioned above, the image was transformed to HSI color space for generating the color histograms, and the value of intensity which corresponds to the gray scale representation of the image, was processed for edge detection. Thus the intensity channel is convolved separately with the two Sobel gradient operators to produce separate measurements of the gradient component in each orientation(call these Gx and Gy). These can then be combined together to find the absolute magnitude of the gradient at each point and the orientation of that gradient. The orientations are in turn grouped into 8 directions(0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°). Finally the edge histograms are calculated by summing up the number in each direction, i.e.

$$H(i) = |\sum a(x, y) \mid a(x, y) = i|, i = 1,2,\ldots,8 \quad (2.7)$$

where $a(x, y)$ is the direction value of a pixel in the image.

The Histogram Intersection method (denotes in Equation 2.8, 2.9) is employed for similarity calculations as a result of texture vector and edge histogram comparisons between database images and query image.

intersection

$$(h(I), h(M)) = \sum_{j=1}^{K} \min\{h(I)[j], h(M)[j]\} \quad (2.8)$$

match

$$(h(I), h(M)) = \frac{\sum_{j=1}^{K} \min\{h(I)[j], h(M)[j]\}}{\sum_{j=1}^{K} h(M)[j]} \quad (2.9)$$

3. Performance Experiments

The performance of our content based image retrieval system was tested on the database having 1000 images which used in SIMPLIcity[3].

3.1 Evaluating Effectiveness

In order to evaluate effectiveness of our retrieval systems, two well known metrics, precision and recall [5], are used. Since different queries may lead to different precision and recall values, the computation of average effectiveness is needed. To ease this computation, an interpolation procedure is performed in order to facilitate the computation of average of precision and recall values [5], as shown below:

- Individual precision values are interpolated to a set of 11 standard recall levels: 0%, 10%, 20%, 100%.
- Let $j \in \{0,1,2,\ldots,10\}$ be a reference to the j-th standard recall level. Then,

$$P(r_j) = \max_{r_j \le r \le r_{j+1}} P(r) \quad (3.1)$$

- So, the interpolated precision at the j-th standard recall level is the maximum known precision at any recall level between the j-th recall level and the (j+1)-th recall level.

In this section, the interpolated precision-recall graphs for an image are given(Figure 3.1, 3.2). The experiment is to show the different effectiveness of the system by using three features separately and the combination of them. There is no specific reason for presenting the query images, but they provide a comprehensive way to evaluate the retrieval effectiveness of the content based image retrieval system. To conclude that the system is effective, the basic expectation from the interpolated precision-recall graphs has to be non increasing. In all of these graphs, the precision value is 1 for the first few standards recall levels, and while standard recall levels are increasing, the precision values continue in a non increasing manner.

3.2 Query Examples

Figure 3.2 shows the result sets for four example queries using color, texture, edge and the integration of them. Figure 3.1 is a Precision-Recall Graph of searching result by using texture feature, which is replaced by Interpolated

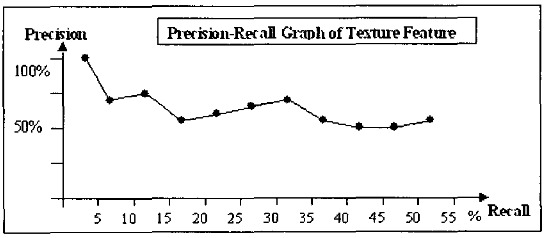
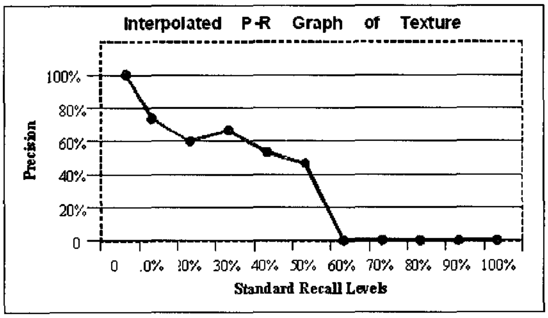
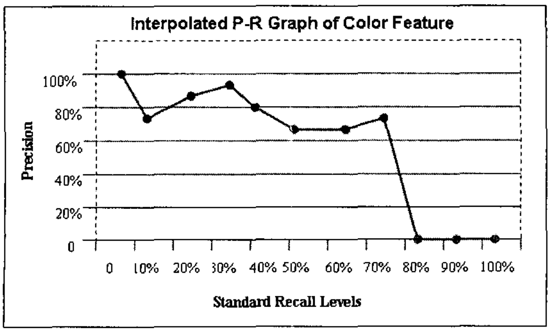


Figure 3.1 Precision Recall Graph of Searching Result

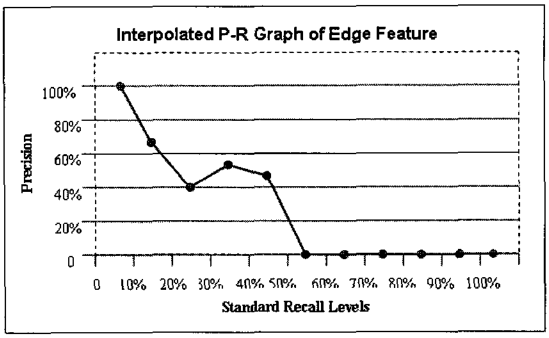
by Using Texture Feature. (Image 435.jpg)



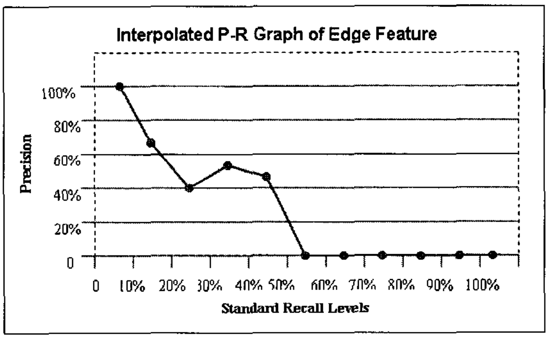
(a)



(b)



(c)



(d)

Figure 3.2 Interpolated Precision Recall Graph of Searching Result

Precision recall Graph in further experiments for lack of the consistent criterion.

by Using (a)Texture, (b) Color, (c) Edge and (d) Integration Feature (Image 435.jpg).

4. Conclusions and Future Work

We have designed and implemented a content-based image retrieval system that evaluates the similarity of each image in its data store to a query image in terms of color, texture and edge characteristics, and returns the images within a desired range of similarity.

For the color content extraction, the color descriptor is used. The expressiveness of this technique is accelerated via color space transformation and quantization, and the color features are extracted by the help of an improved K Means Clustering algorithm, an effective method for determining the optimal number of classes. From among the existing approaches to texture analysis within the domain of image processing, we have adopted the statistical approach to extract texture features from both the query images and the images of the data store. Five distinct functions have been selected as an optimal subset of the set of statistical features that can be extracted from co-occurrence matrix. For the edge feature extraction, a well known and effective technique and histograms were used. The histogram intersection method has been used as the similarity measure between two feature vectors.

Our system has been tested on images used in SIMPLIcity and shown to be an efficient tool for image retrieval. Based on the experimental results, the retrieval based on the color feature is relatively robust to background complication and independent of image size and orientation compared to texture and edge features. The poor performances shown by texture and edge occur because the whole image is processed for feature extraction, which means treating the image as if it is composed of one sole texture. However, texture feature is proved superior to others when searching an image with a huge background and uniformity. And edge feature works as the accessorial component of color feature by virtue of the capability of spatial discrimination. Finally, the retrieval by integrating these features with a proper proportion shows the best performance.

A crucial future work to be done on our system is to further enhance its capability of querying with respect to the certain object regions in images in which better results can be obtained by using texture feature. Thus, the semiautomatic and fully automatic feature extraction processes will be included. For the former, the user is provided a drawing facility that is used for the specification of regions of interest. For the latter, the system tries to capture a rectangular region that represents the interested content of the image.

Reference

[1] Gotlieb, C. C. and Kreyszig, H. E. (1990). Texture descriptors based on co occurrence matrices. Computer Vision, Graphics, and Image Processing 51

[2] Haralick, R. M., Shanmugam, K., and I. Dinstein. (1973). Textural features for image classification. IEEE Transactions on Systems, Man and Cybernetics, 3 : 610621

[3] Jain, K. and Vailaya, A. (1996). Image Retrieval using Color and Shape. Pattern Recognition, Vol. 29, 1233-1244

[4] Jones, K. S. (1981). Information Retrieval Experiment. Butter worth and Co.

[5] Konak, E. S. (2002). A Content Based Image Retrieval System for Texture and Color Queries, 13-70

[6] Ma, M., Singh, K., D W. Park, (2003). Image Retrieval By Contents / Using Color & Edge, Proc. of. The 3rd International Workshop on Content Based Multimedia Indexing Sep.22-24, IRISA, Rennes, France.

[7] Rui, Y. and Huang, T. S. (1999). Image Retrieval: Current Techniques, Promising Directions and Open Issues, Journal of Visual Communication and Image Representation

[8] Singh, K., Ma, M., Park, D W. (2003). A Content- Based Image Retrieval Using FFT & Cosine Similarity Coefficient, Proc. of. The 5th IASTED International Conference Signal and Image processing, (SIP 2003), Honolulu, Hawaii, U.S.A. 13-15

[9] Smith, J. R. and Chang, S. F. (1995). Single color extraction and image query, in Proc. IEEE Int. Conf. on Image Proc.