International DSI / Asia and Pacific DSI 2007 Full Paper (July, 2007)

Visual Information Retrieval by Efficient Feature Extraction

Sang-Sung Park, Young-Geun Shin, Hyun-Chi Jeon, Dong-Sik Jang

Division of Information Management Engineering, Korea University
(hanyul, toctop, chichi, jang@korea.ac.kr)

Abstract

A new visual information retrieval method is proposed in this paper. The proposed method is also used to retrieve similar images from databases. The basic idea is to extract color, the global and local information that images have. The feature of global information is that the color is extracted by RGB model. While, the feature of the local information is that the color is extracted by HSV model. In the case of local feature, unless it has representation, the wrong retrieval result would be caused. GA(Genetic Algorithm) is used to extract the local feature that has representation. The local feature extracted by GA is an optimal representative feature. In the experiment, the recall and precision rate of images are measured using the proposed algorithm. Also, we compared the previous algorithm with the proposed algorithm for the performance of image retrieval. As a result, the proposed algorithm showed higher performance in terms of recall and precision rate.

1. Introduction

The development of computer and communication technologies makes us easily access to the information of outside world. Especially, technology, which is related to production, transmission, and retrieval of multimedia data, becomes one of the main research areas. We focus here images among multimedia data and propose a new content-based image retrieval method. In earlier retrieval systems, a user presents a query in a keyword format and a system retrieves images indexed with the corresponding keyword. However, for a large database the text indexing is cumbersome and time consuming. In addition, the keyword is inherently subjective and unique. A content mismatch occurs when information that indexer ascertains from an image differs from the information that the user is interested in. Another problem is that textual descriptions can never capture the visual content such as color, texture, and shape sufficiently. Some visual properties, such as certain textures and shape, are difficult or nearly impossible to describe with text. Therefore, in order to overcome these faults, image content as well as text keywords are used for image retrieval over the last decade. This method automatically analyzes the image content such as color, texture, and shape and uses those for retrieval purpose.

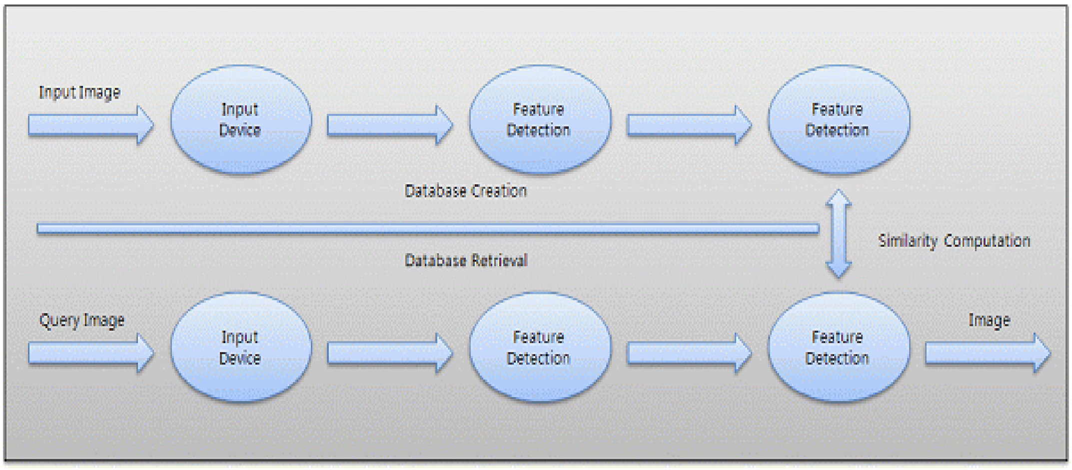

Fig. 1 An image retrieval system

Fig. 1 shows a general block diagram of the image retrieval system in this approach [1]. In this paper, we deal with

content-based image retrieval(CBIR), which is a technique to retrieve images based on their visual properties such as color [2], texture [3], and shape [4]. Systems [5, 6] are well known for supporting this content-based image retrieval. Fast retrieval in databases has been one of the active research areas. In that process, without any clustering schemes and adequate indexing structures, retrievals of similar images are time-consuming because the database system must compare the query image to each image in the database. This cost can be particularly prohibitive if the database images are very large and their features tend to have high-dimensionality. This high-dimensional indexing structure increases the retrieval time and memory space exponentially, as the member of feature dimension increases. Thus, frequently, it does not have any advantages against the simple sequential search. So, fast search algorithms, which can deal with high-dimensional feature data, are often an essential component of the image database. There have been a number of indexing data structures suggested to handle high-dimensional data [7, 8]. The problems of previous image retrieval are as below:
First, the reduction of retrieval efficiency due to the wrong extraction of representative features.
Second, the reduction of speediness due to the extraction of mass image features.
To guarantee the speediness and accuracy for image retrieval, we propose a new method which extracts the representative feature of images. The feature of images is divided into global and local feature. In the case of global feature, the representative feature is extracted by the RGB color model. But the local feature applied the feature extracted by the HSV color model to GA algorithm for guaranteeing representation. The representative feature consists of optimal structure. Thus, it has representation with the small amount of data.

2. Image Features

Color is the dominant component of human perceptions. We used color features to represent images. We used RGB (Red, Green, Blue) color model and HSV(Hue, Saturation, Value) color model. The RGB color model is used for many image retrieval systems [9, 10]. For global image representation and fast search, RGB color histograms of the image, which are quantized into 16 bins per R, G, and B coordinates, are extracted. For local information, the image is divided into rectangular regions. The HSV model correlates well with human color perception. A set of HSV joint histograms in each rectangular region are extracted, and dominant hue, saturation, and value are used as features in that region.

2.1 RGB Color Model

A histogram is a widely used technique to represent image and it has the properties of rotation and translation invariance. A suitable normalization can also provide scale invariance. In the study, we used an RGB color histogram to represent global information of the image.

2.2 HSV Color Model

For representing color, we used HSV (Hue, Saturation, Value) color model because this model is closely related to human visual perception. For color quantization, we uniformly quantized HSV space into 18 bins for hue (each bin consisting of a range of 20 degree), 3 bins for saturation and 3 bins for value for lower resolution. In order to represent the local color histogram, we divided image into equal-sized rectangular regions and extract HSV joint histogram that has quantized 162 bins for each region. And to obtain compact representation, we extract from each joint histogram the bin that has the maximum peak. The HSV representation of an image from RGB is obtained using the following relationships:

$$H = \begin{cases} \theta, & G \ge B, \\ 2\pi - \theta & G \le B, \end{cases}$$
$$where \quad \theta = \cos^{-1}\left[\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2+(R-B)(G-B)]^{\frac{1}{2}}} \right], \tag{1}$$

$$S = 1 - \frac{3}{R+G+B}\left[\min(R,G,B)\right], \tag{2}$$

$$V = \frac{1}{3}(R+G+B). \tag{3}$$

Take hue, saturation, and value associated to the bin as representing features in that rectangular region and normalize to be within the same range of [0,1]. Thus, each image has the dimensional color vector.

3. Representative Feature Extraction Image Retrieval

In this paper, the global and local features are used for image retrieval. But the local feature can not only reduce the retrieval speed due to the mass data but reduce the retrieval accuracy due to the wrong extraction of representative feature. To solve this problem, we propose the algorithm which extracts the optimal representative feature from the local feature, HSV using GA algorithm. Fig. 2 shows the procedure of color feature extraction.

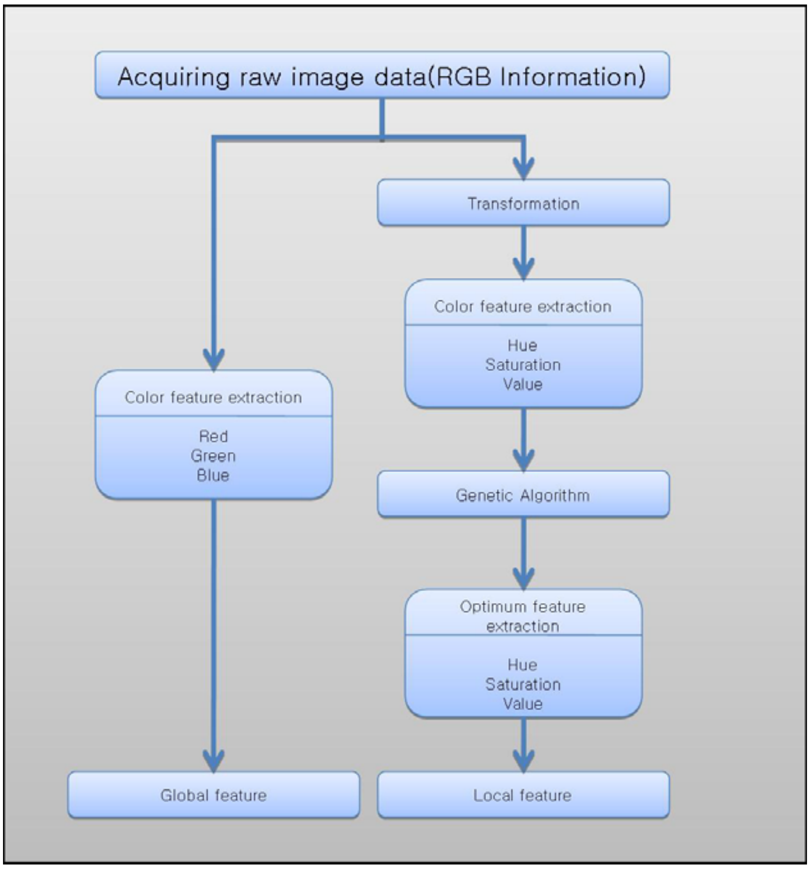

Fig. 2 Color feature extraction procedures

3.1 Genetic Algorithm (GA)

GAs are computational models of evolution. They work on the basis of a set of candidate solutions. Each candidate solution is called a 'chromosome', and the whole set of solutions is called a 'population'. The algorithm allows movement from one population of chromosomes to a new population in an iterative fashion. Each iteration is called a 'generation'. There are various forms of GAs, a simple version, which is called static population model was used in all the experiments [11, 12]. In the static population model, the population is ranked according to the fitness value of each chromosome. At each generation, two (and only two) chromosomes are selected as parents for reproduction. GAs operate iteratively on a population of structures, each one of which represents a candidate solution to the problem at hand, properly encoded as a string of symbols (e.g.,binary). A randomly generated set of such strings forms the initial population from which the GA starts its search. Three basic genetic operators guide this search: selection, crossover, and mutation. The genetic search process is iterative: evaluating, selecting, and recombining strings in the population during each iteration (generation) until reaching some termination condition. The basic algorithm, where P(t) is the population of strings at generation t, is given below:

t = 0
initialize P(t)
evaluate P(t)
while (termination condition is not satisfied) do

begin
select P(t + 1) from P(t)
recombine P(t + 1)
evaluate P(t + 1)
t = t + 1
end

Evaluation of each string is based on a 1tness function that is problem-dependent. It determines which of the candidate solutions are better. This corresponds to the environmental determination of survivability in natural selection. Selection of a string, which represents a point in the search space, depends on the string's 1tness relative to those of other strings in the population. It probabilistically removes, from the population, those points that have relatively low fitness. Mutation, as in natural systems, is a very low probability operator and just flips a specific bit. Mutation plays the role of restoring lost genetic material. Crossover in contrast is applied with high probability. It is a randomized yet structured operator that allows information exchange between points. Its goal is to preserve the fittest individuals without introducing any new value.

3.2 Representative Feature Extractions

GA is used as a tool to get an optimal solution. Previous retrieval algorithms can not use the information obtained in the process of retrieval. But GA algorithm uses information in the process of retrieval for referring to the new retrieval. In addition, it has an adaptive function which retrieves the solution according to the standard of judgment. In other words, GA algorithm calculates the result with evolving.

3.2.1 Feature Vector Encoding

The optimal representative feature is obtained by encoding the color features, h, s and v. Fig. 3 shows the Chromosome Structure.

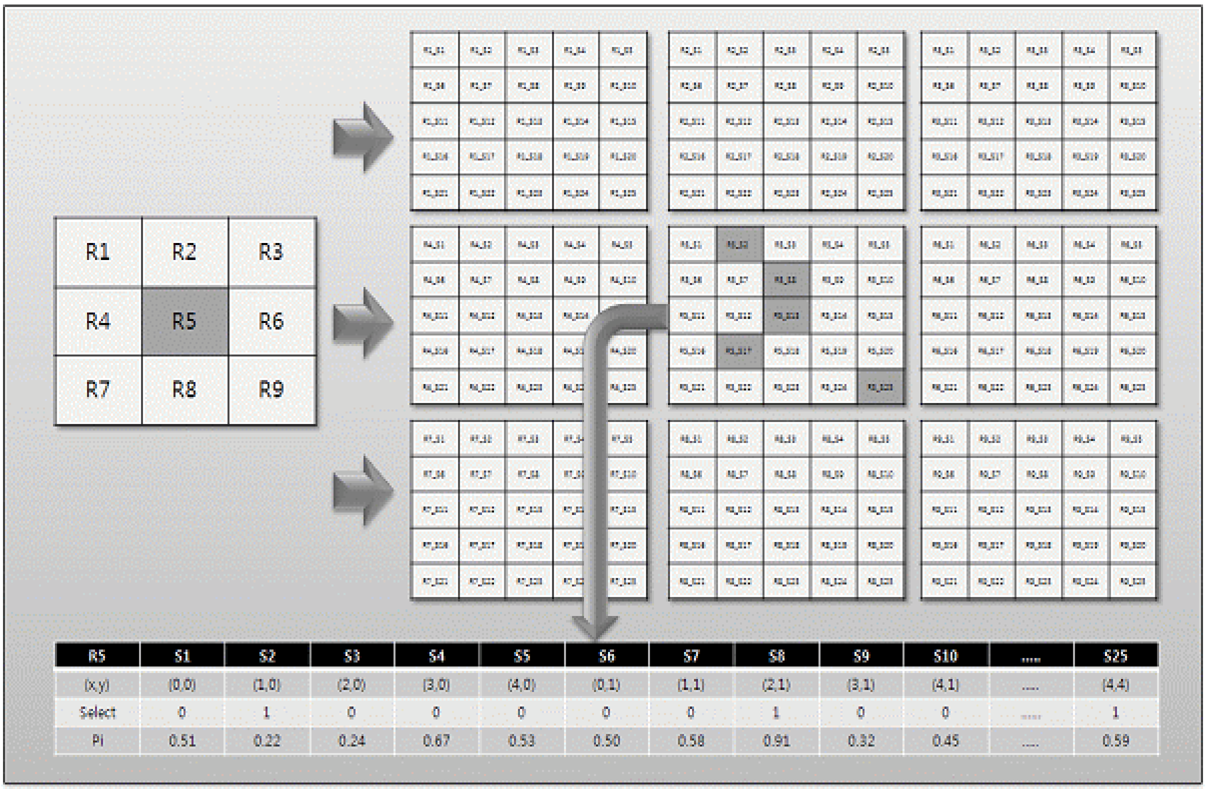

Fig. 3 Chromosome Structure

The selected sub region is indicated as 1, and the deselected region is indicated as 0. In this way, only five regions are randomly selected from the 25 regions. Consequently, the number of regions which have the value 1 from the total of 25 sub regions is five.

3.2.2 Fitness Function

The features for each region generally use the mode of histograms or pixel values. However, the feature using the mode of histograms or pixel values is subject to get trapped in local minima. In the case of using the mode of pixel values, when the distorted information is focused on a part of the region, the representative feature will get trapped in local minima. The histogram method is also subject to get trapped in local minima if the constant noise patterns in an image are converged as a representative feature. Thus, in this paper, we propose three conditions for the fitness function which can not only solve the problem with local minima but guarantee the similarity of each sub region.

1) A Condition for Consistency

Homogeneity is that the internal natures of each separated region such as brightness and texture are similar and consistent. To satisfy the nature, variance which can be expressed by numerical value is used as a measure of homogeneity. If a region has lots of the same pixel values, the variance would have a small value and the smallest value among the sum of variance for every region is the best segmentation for homogeneity. Q_{r1} , the variance of the region 1 ($r1$) is obtained as below.

$$Q_{r1} = \frac{1}{5}\sum_{i=1}^{5}(p_i - \overline{p}) \tag{4}$$

5 indicate the number of each sub region. $\overline{p}$ is the mean or mode of pixel values at p_i and $\overline{p}$ is the mean of five sub regions. Thus, the evolvement should proceed in order to have a small value of Q_{r1} because the region will have lots of similar pixels as the value of Q_{r1} decreases.

2) A Condition for Robustness from Noise

The distance between regions is obtained by Euclidean distance and it is summed to prevent the randomly selected sub region from being converged by noise. This is to evenly select the sub region from every region. If the mean value is obtained by the only noise regions, the result will get trapped in local minima. It can be interpreted that the sample is evenly selected from all parts of the image as the value of Q_{r2} increases. On the other hand, the sample would be selected from the specific region as the value of Q_{r2} decreases.

$$Q_{r2} = \sum_{i=1}^{5}\sum_{j=1}^{5}\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{5}$$

(x_i, y_i) are the coordinates of the sub region , P_i .

3) A Condition for Robustness from Constant Patterns

In the case of patterns which are not associated with image features, the improper representative features can be extracted because the many patterns converged as mode. For example, there is a black line on the outer frame of an image and the representative feature converges into the pattern. To exclude this vertical and horizontal adjacency, the equation (6) is proposed.

$$Q_{r3} = \sum_{i=1}^{5}\sum_{j=1}^{5}\min(x_i - x_j, y_i - y_j) \tag{6}$$

The adjacency grows less and less as Q_{r3} increases and the adjacency grows more and more as Q_{r3} decreases. The standard to retrieve a solution for a problem in genetic algorithms is the fitness or objective function. In this paper, the fitness is defined as below to extract the optimal representative feature.

$$F = \lambda Q_{r1} + \frac{1}{Q_{r2}} + \frac{1}{Q_{r3}} \tag{7}$$

F is the fitness function and λ is the weight value. The initial value for λ was fixed as 0.95.

3.2.3 Initialization

An initial population size ${}_{25}C_5$ for a genetic algorithm should be generated, which may be randomly 10 selected.

3.2.4 Selection

Based on the fitness values of bit strings in the current population, pairs of “parent” bit strings are selected which undergo the genetic operation “crossover” to produce pairs of “offspring” bit strings forming the next generation. The probability of selection of a particular chromosome is proportional to the fitness function value. In this study, Roulette-Wheel selection method is employed.

3.2.5 Crossover

Crossover exchanges information between two parent bit strings and generates two offspring bit strings for the next population. In this study, one-point crossover is employed.

3.2.6 Mutation

Mutation can extend the scope of the solution space and reduce the possibility of falling into local extremes. In general, the probability of applying mutation is very low.

3.2.7 Parameter and stopping criterion

- Probability of applying crossover = 0.92
- Ptonsnility of applying mutation = 0.05

In this study, the following two stopping criteria are employed:

(1) Average fitness of population is smaller than threshold1(0.44)

(2) The minimum value in the fitness values of each chromosome is smaller than threshold2(0.40)

4. Experimental results

In order to evaluate the proposed algorithm, all experiments were performed on a Pentium IV with 512Mbyyes of main memory and 100Gbytes of storage. All programs have been implemented in Visual C++. We gathered 400 images where most of them have dimensions of 192 x 128 pixels. These were gathered from public sources and represent natural scenes such as animals and plants. In order to measure the retrieval effectiveness, we selected eight query images from the database and performed the experiments based upon the relevance of the retrieved images to the query. The query image is compared to the database images based on the RGB color histogram and HSV joint histogram features. Retrieval effectiveness can be defined in terms of precision and recall rates. A precision rate can be defined as the percent of retrieved images similar to the query among the total number of retrieved images. A recall rate is defined as the percent of retrieved images which are similar to the query among the total number of images similar to the query in the database. There are recall rate and precision function as in (8), (9).

$$recall = \frac{relevant\ correctly\ retrieved}{all\ relevant} \tag{8}$$

$$precision = \frac{relevant\ correctly\ retrieved}{all\ retrieved} \tag{9}$$

Table 1 shows recall and precision rates for eight queries.

Table 1 Recall and precision rates for eight queries

Query	N1	N2	Recall rate	Precision rate
Rose	6	50	12%	60%
Polar Bear	10	50	20%	100%
Horse	9	50	18%	90%
Sunset	8	50	16%	80%
Eagle	9	50	18%	90%
Airplane	9	50	18%	90%
tiger	5	50	10%	50%
Valley	7	50	14%	70%
Average			15.78%	78.75%

In the best case, the retrieval performance provides a 20% recall rate for the whale query image and a 100% precision rate for the eagle query image. The overall performance was a 15.78 recall rate and a 78.75% precision rate.

Table 2 Comparison of previous and proposed methods

	Existing method		Proposed method	
Average	Recall rate	Precision rate	Recall rate	Precision rate
	14.75%	73.75%	15.78%	78.75%

Table 2 shows the comparison result of previous and proposed methods. The experimental results proved that the proposed method had a higher recall and precision rate than previous method.

5. Conclusions

In this paper, global and local features are extracted for effective image retrieval. The local feature which is subject to get trapped in local minima extracted the representative feature using Genetic Algorithm. The average performance was a 15.78% recall rate and a 78.75% precision rate. As a result, the proposed method showed more excellent performance than a previous method. In short, we believe that this is a promising technique for content-based image retrieval in large databases. For further improvement, we are investigating a relevance feedback approach in order to capture users' visual perception.

Acknowledgement
This work is supported by the Brain Korea 21 Project in 2007

References

[1] M. Swain, & D. Ballard; Color indexing, International Journal of Computer Vision, Vol. 7, No. 1, pp. 11-32, 1991.

[2] J. R. Smith, & S.-E. Chang; VisualSEEK: A fully automated content-based image query system, In Proc. ACM Multimedia, pp.87-98, 1996.

[3] R. M. Haralick, K. Shanmugam, & I. Dinstein; Texture features for image classification, IEEE Trans. On Sys. Man. And Cyb. SMC-3(6), pp. 610-621, 1973.

[4] A.K Jain, & A. Vailaya; Shape-based retrieval: a case study with trademark image databases, Pattern Recognition, Vol. 31, pp. 1369-1390, 1998.

[5] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, & P. Yanker; Query by image content: The QBIC system, IEEE Computer, Vol. 28, No. 9, pp. 23-31, 1995.

[6] A. Pentland, R. W. Picard, & S. Sclaroff; Photobook: Content-based manipulation of image databases, International Journal of Computer Vision, Vol. 18, No. 3, pp. 233-254, 1996.

[7] D.A. White, & R. Jain; Similarity indexing with the SS-tree, In Proc. 12th IEEE International Conference on Data Engineering, pp. 516-523, 1996.

[8] K.-I. Lin, H. V. Jagadish, & C. Faloutsos; The TV-tree: An index structure for high dimensional data, VLDB Journal, Vol. 3, No. 4, pp. 517-549, 1994.

[9] A. K. Jain, & A. Valilaya; Image retrieval using color and shape, pattern recognition, Vol. 29, pp. 1233-1244, 1996.

[10] A. Berman, & L. G. Shapiro; Efficient image retrieval with multiple distance measures, In Proc. SPIE Vol. 3022: Storage and Retrieval for Image and Video Databases V, pp. 12-21, 1997.Dd

[11] D. Whiney; A genetic algorithm tutorial, Technical Report, Department of computer science, Colorado state university, CS-93-103, 1993.

[12] R. L. Haupt; An introduction to genetic algorithms for electromagnetics, IEEE Magazine, Antennas Propagation, Vol. 37, pp. 7-15, 1995..