The Usage of Color & Edge Histogram Descriptors for Image Mining

Syungog An[†††] · Dong-Won Park[††] · Kulwinder Singh[†] · Ming Ma[†]

ABSTRACT

The MPEG-7 standard defines a set of descriptors that extracts low-level features such as color, texture and object shape from an image and generates metadata in order to represent these extracted information. But the matching performance for image mining ma y not be satisfactory by u sing only on e of these features. Rather than by combining these features we can achieve a better query performance. In this paper we propose a new image retrieval technique for image mining that combines the features extracted from MPEG-7 visual color and texture descriptors. Specifically, we use only some specifications of Scalable Color Descriptor (SCD) and Non-Homogeneous Texture Descriptor also known as Edge Histogram Descriptor (EHD) for the implementation of the color and edge histograms respectively. MPEG-7 standard defines $l_1 - norm$ based matching in EHD and SCD. But in our approach, for distance measurement, we achieve a better result by using cosine similarity coefficient for color histograms and Euclidean distance for edge histograms. Our approach toward this system is more experimental based than hypothetical.

Keywords : MPEG-7 ,Scalable Color Descriptor (SCD), Edge Histogram Descriptor (EHD), distance measurement, cosine similarity coefficient ,Image Mining

칼라와 에지 히스토그램 기술자를 이용한 영상 마이닝 향상 기법

안성옥[†††] · 박동원[††] · Kulwinder Singh[†] · Ming Ma [1)†]

요 약

영상의 칼라, 텍스쳐, 오브젝트의 형체 등과 같은 하위 수준의 특징을 표현할 수 있는 기술자를 MPEG-7 표준에서 규정하고 있다. 하지만, 각각의 기술자를 따로 분석함으로써는 성능 향상에 불충분한 점이 있었다. 본 논문에서는 칼라 기술자와 텍스쳐 기술자를 결합하여 영상검색의 성능을 향상시키는 방법을 제안한다. MPEG-7 표준에서 정의한 $l_1 - norm$ 방법보다, 본 논문에서는 칼라 히스토그램의 경우 코사인 근사도 계수를, 에지 히스토그램의 경우 유클리디언 디스턴스를 적용 실험하여 진일보한 결과를 도출할 수 있었다.

키워드 : 칼라기술자, 텍스쳐기술자, 영상검색, 에지히스토그램, 유클리디언디스턴스

†비 회 원: 배재대학교 IT공학부 박사과정
††비 회 원: 배재대학교 IT공학부 부교수
†††종신회원: 배재대학교 IT공학부교수(교신저자)
논문접수: 2004년 6월16일, 심사완료: 2004년9월14일

1. Introduction

With the rapid increase in computer speed and decline in memory cost, image databases containing thousands or even millions of images are used in many application areas where there is a real need for tools that help navigate such large volumes of data.With the growth in the number of images, manual annotation becomes infeasible both time and cost-wise. Content-Based Image Retrieval (CBIR)[1],[2],[3] describes the process of retrieving desired images from a large collection on the basis of features, such as color, texture and shape that can be automatically extracted from the images. CBIR is a single standard, which can provide a simple, flexible, interoperable solution to the problems of indexing searching and retrieving multimedia data, and will be extremely valuable and widely deployed.

The present paper proposes and evaluates a new retrieval system concept that provides automated features extraction from an image. This concept also enhances indexing and query performance. In addition to this we also discuss about edge and color histograms generation and their implementation for our retrieval system. If we use only color histogram descriptor for image retrieval, there is a probability that many images can have similar color histograms. Our objective was to overcome this weakness of the color histogram descriptor, by combining both the features (color & edge histogram descriptors). We use Euclidean distance and cosine coefficient for similarity measurement between users' query image features and features of images in the database.

This paper is organized as follows. In Section 2 & 2.1 a brief overview of related work and color spaces are given and section 3 is about Color descriptors. Section 4 describes the implementation of the edge histogram, and Section 5 discusses the use of Euclidean distance and cosine similarity coefficient. Section 6 deals with combination of the features. Section 7 analyzes the experimental results. Section 8 of the paper summarizes the conclusion followed by the references.

2. Related Work

In the past few years, several CBIR systems have been developed, notably:

The IBM **QBIC** System - developed at the IBM Almaden Research Center [1].

The **Virage** System - developed by the Virage Inc [2].

The **Photobook** System - developedby the MIT Media Lab [3].

The basic ground is to extract the image contents as feature vectors from each image in the database and to index them for query. The similarity between a query image and images in the database is dependent on computation of the distance function. The Euclidean distance function is widely used to measure such similarities. Finally, the images are arranged in an ascending order of distance and send back to users as a query result.

2.1 Supporting tools for color descriptors

The color space and quantization are the important parts of the color descriptors. These descriptors are utilized in conjunction with other descriptors such as scalable color descriptor (SCD), dominant color descriptor (DCD) etc. The color quantization is a lossy process, which specifies the partitioning of the given color space into discrete bins [4]. The MPEG-7 standard supports different color spaces (i.e. HSV, RGB, YC_bC_r, HMMD etc.) under the specific description.

For example: The perceptually uniform HSV (Hue, Saturation and Value) color space and HMMD (Hue, Min, Max & Diff) color space are used in SCD (Scalable Color Descriptor) and CSD (color structure descriptor) respectively. HSV is a well-known color space, which is widely used in digital image processing applications. HMMD is a new color space introduced by MPEG.Among several existing color spaces, the most common is RGB (red/green/blue) color space (due to availability of the images in RGB format from the scanners) where each pixel is represented by a linear combination of three components red (wavelength 650nm), green (wavelength 510nm) and blue (wavelength 475nm). But RGB is not necessarily the most efficient representation of color. Since the human perception is more sensitive to luminance than color, RGB color space does not provide an easy way to take advantage of this because RGB color space is not perceptually uniform [5], as all the three components (R, G, B) have equal importance and the luminance is present in all the three color components.To represent the color image more efficiently for image mining, luminance should be separated from the color information.

We employ some specifications of SCD for the color histograms implementation where HSV color space is a nonlinear transformation of the RGB color space. Human eyes can distinguish about 128 different hues and 130 saturations and 23 shades. The algorithm for transforming an RGB image into the HSV color space is given below:

```
v=max(R,G,B) ;
s=(v-min(R,G,B))/v ;
if ( R==max(R,G,B) && G==min(R,G,B )) h=5+(R-B)/(R-G) ;
else if ( R==max(R,G,B) && B==min(R,G,B )) h=1-(R-G)/(R-B) ;
else if ( G==max(R,G,B) && B==min(R,G,B )) h=1+(G-R)/(G-B) ;
else if ( G==max(R,G,B) && R==min(R,G,B )) h=3-(G-B)/(G-R) ;
else if ( B==max(R,G,B) && R==min(R,G,B )) h=3+(B-G)/(B-R) ;
else if ( B==max(R,G,B) && G==min(R,G,B )) h=5-(B-R)/(B-G) ;
h*=π/3 ;
```

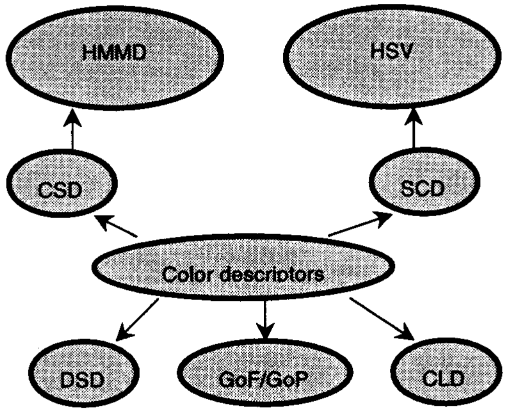
The HSV color space can be interpreted as a cylinder, where h approximately represents an angle (in radians). S is the radius of the circle ($0 \le s \le 1$) and v is the height of the cylinder ($0 \le v \le 1$). We use non-homogenous texture descriptor for the implementation of the edge histogram in YC_bC_r color space.

The equations for transforming an RGB image into the YC_bC_r color space [4] are given in equation 1.

$$
\begin{aligned}
Y &= 0.299R + 0.587G + 0.114B \\
C_b &= -0.169R - 0.331G + 0.500B \\
C_r &= 0.500R - 0.419G - 0.081B
\end{aligned} \quad (1)
$$

3. Visual Color descriptors

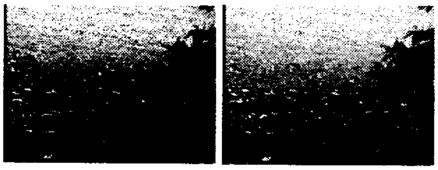
Color is an important visual attribute for both human perception and computer vision and most widely used in Image/video retrieval. But an appropriate color space and color quantization must be specified along with a histogram representation of an image for retrieval purpose. Histogram describes the global distribution of pixels of an image. The main advantage of a color histogram is its small sensitivity to variations in scale, rotation and translation of an image whereas it has drawbacks that similar images can have different color histograms (as shown in figure 2) and different images can have similar color histograms. MPEG-7 standard supports following color descriptors [7]:



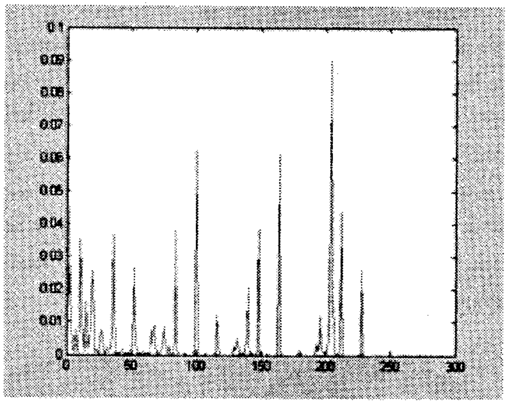
(Figure 1) Classifications of color descriptor

CSD-Color Structure Descriptor
CLD-Color Layout Descriptor
GoF/GoP-Group of Frames or Group of Pictures Descriptor

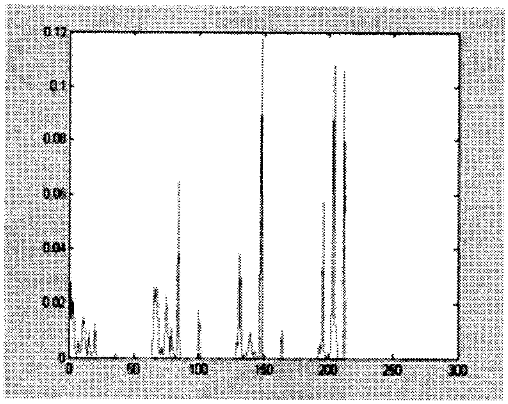
We employ some specifications of SCD for our work as we have mentioned in section 2.We have tried different kinds of quantization schemes for the implementation of the color histograms in HSV color space and we observed that HSV color model is better than RGB color model in our approach using following quantization scheme. Each color component is uniformly quantized, H: 16 bins; S: 4 bins; V: 4 bins. Finally, we concatenate this 16x4x4 histogram and we get a 256-dimensional vector. We do not use Haar transformation.



(a) (b)



256-bin HSV histogram for (a)

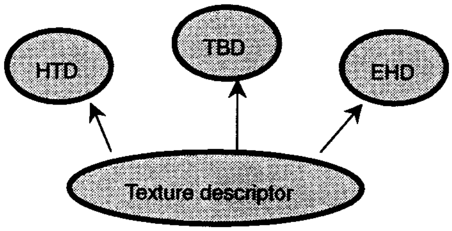


256-bin HSV histogram for (b)

(Figure 2)

4. Non-homogenous texture descriptor (EHD)

Texture defines the visual patterns of an image that are obtained from the presence of multiple colors or intensities. The use of appropriate texture descriptors provides a powerful means for similarity matching and retrieval. Figure 3 shows the classifications of texture descriptor [7].



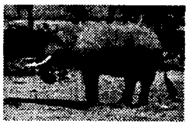
(Figure 3) Classifications of texture descriptor

TBD- Texture Browsing Descriptor

HTD- Homogeneous Texture Descriptor

EHD- Edge Histogram Descriptor (Non-Homogeneous Texture Descriptor)

Edges in images constitute the important roles. An edge histogram represents the frequency and the directionality of the brightness that change in an image. It is a unique feature of an image, which cannot be altered by color histogram or by homogeneous texture attribute. This feature can be used to retrieve images with similar semantic meaning. As we have mentioned in section 2, we use some specifications of the edge histogram descriptor EHD for the implementation of the edge histograms in YC_bC_r color space and we compute edge histogram only from the luminance component of an image (see Fig. 4). This descriptor captures the spatial distribution of edges in an image, which are classified into five categories: vertical, horizontal, 45^o diagonal, -45^o diagonal and isotropic as shown in the figure 5.

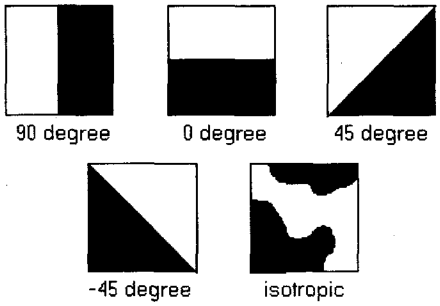


Original RGB Image

Luminance component Cr component Cb component

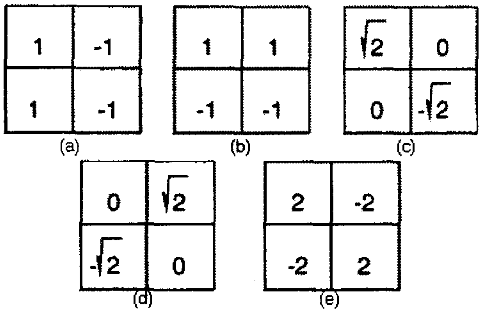
(Figure 4) Depicts original RGB image into YC_bC_r color space

We do not segment the image into 4x4 sub images and compute the edge histogram for each of them because the size of the images in our database is small (200X130 pixels).



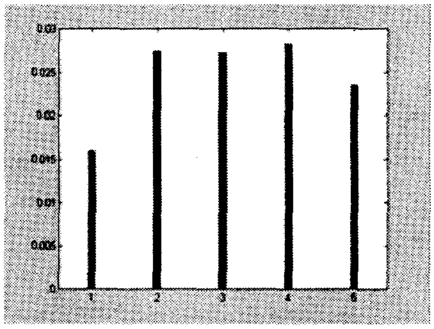
(Figure 5) Shows five edge directions

We use 5 kinds of 2x2 operators in order to compute gradient in 5 directions: 0, +/-45, 90, and isotropic as shown in the figure 6. Thresholding is a simple, non-contextual, efficient segmentation technique. Thresholding [16] can employ either a fixed or an adaptive threshold value. A variety of techniques have been devised to automatically choose a threshold, but no one is robust. We employ Otsu's method [18] for thresholding for all the mentioned operators. Edge histogram is normalized by the size of an image, i.e. each bin value represents the percentage of a certain edge direction in an image.



(Figure 6) MEPEG-7 Edge Filters, left to right (a) 0, (b) 90, (c) -45, (d) 45 degree, and (e) isotropic

(a)



(Figure 7) (Luminance) 5-bin edge histogram of (a)

5. Similarity measurement

In this section we discuss about *cosine* similarity coefficient andEuclidean distance measurement for comparing feature vectors of a query image and feature vectors of images in a database. This is the main component of our retrieval engine, which connects the user to image database and sends back the query result to the user (see figure 8.). Here Euclidean distance and cosine coefficient work as comparators. We have two different kinds of distance measurements, namely Color histogram distance $f_c(q,d)$ and edge histogram distance $f_e(q,d)$ and both lies in the range of {0-1}. Color histogram of each image consists of 256 feature vectors. In order to compare these vectors we have used the following equation:

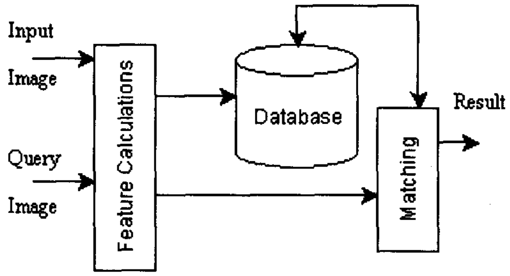
$$f_c(q,d) = \frac{\sum_{i=0}^{255} q_i d_i}{\sqrt{\sum_{i=0}^{255} d_i^2 * \sum_{i=0}^{255} q_i^2}}$$

The above equation is called *cosine* similarity coefficient [15]. The similarity function $f_c(q,d)$ is computed between the color feature vectors of database image defined by $d = (d_0, d_1, \ldots, d_{255})$ and the color feature vectors of the query image, also defined as q =

($q_0, q_1, \ldots, q_{255}$).We have compared the normalized 5-bin edge histograms using following equation.

$$f_e(q,d) = \sqrt{\sum_{i=1}^{n} \{H_q(i) - H_d(i)\}^2}$$

The above equation is known as Euclidean distance [6].



(Figure 8) shows proposed model

6. Combination of the features

This section formulates the problem of combining different features. An example is finding a set of weights of differentfeature distances. This section also presents a **Min-Max** algorithm for finding the best matching image. Suppose we have a query image **q** and images in a database **d.** We have **R** features, thus we have **R** kinds of different distances $f_r(q,d)$, where r = 1 to R and d = 1 to n. We have combined them as a weighted sum of all the distances, i.e.

$$F(q,d) = \sum_{r=1}^{R} w_r f_r(q,d) ------(2)$$

$$and \sum_{r=1}^{R} w_r = 1, w_r \geq 0, \forall r = 1,2,3,\ldots,R ----(3)$$

Further we search a vector **w** that satisfies the equation (3) and the resulting distance measure provides the closest match. In our approach we choose the image that minimizes the maximum distance over all valid set of weights as the best match.

For every database image d, searching for the maximum distance over the weight space turns over to be a linear program. Thus we have the solution as follows: -

Maximize: equation (2), Subject to: equation (3)

where all the distances are constant and $w_r, r = 1,\ldots,R$ are unknown.

The image with minimum "max-distance" is declared as the best match w.r.t. query image.

Color histogram distance $f_c(q,d)$ and edge histogram distance $f_e(q,d)$ bothlies in the range of {0-1}.

The-max-distance $F(q,d) = wf_c(q,d) + (1-w)f_e(q,d), 0 \leq w \leq 1$ of every database image d, is a linear function of w over {0,1}. Thus the maximum either lies at $w = 0$ or $w = 1$

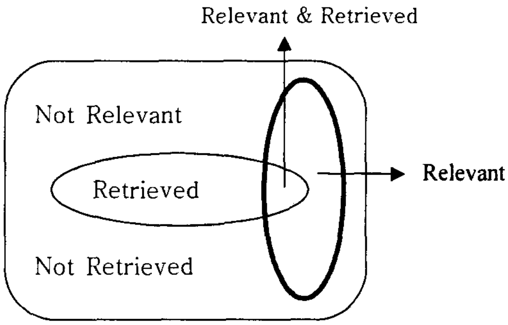
Then we rank the maximum of $f_c(q,d)$ and $f_e(q,d)$ for all database images and take n images with the least distance as our return result.

7. Experiments and results

The performance of the retrieval system can be measured in terms of its recall (or sensitivity) and precision (or specificity). Recall measures the ability of the system to retrieve all models that are relevant. Precision measures that the ability of the system to retrieve only models that are relevant [12]. They are defined as:

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}}$$

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

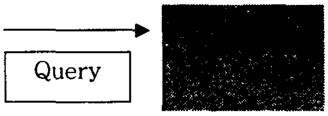


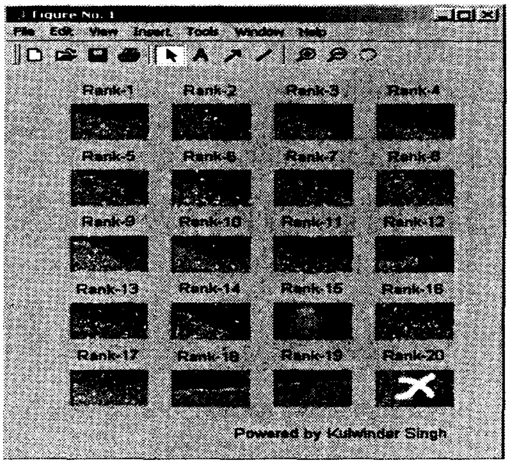
(Figure 9) Set representation for Precision & Recall Calculations

These are the standard measures in IR (Information Retrieval), which give a good indication of the system performance. Also, we can measure the error rate by using the following method [17].

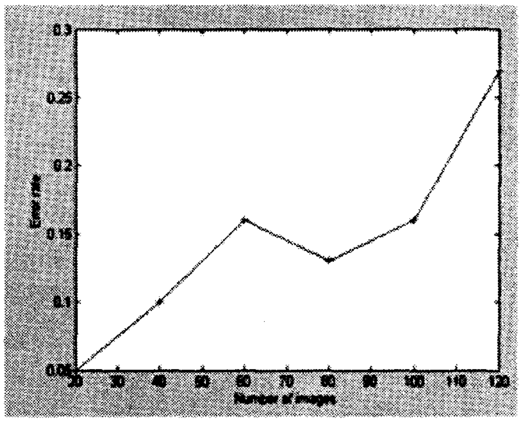
$$\textit{Error rate} = \frac{\textit{Number of non- relevant images retrieved}}{\textit{Total number of images retrieved}}$$

We have performed all the experiments on heterogeneous database [11] (A test database has been used in **SIMPLIcity**paper) containing 1000 images. Also, We have used a total of 448 images from the publicly available **Minerva benchmark**[14]. It should be also noted that our proposed approach does not involve any manipulation of the images. Our main aim was to retrieve those images from the database that are deemed as relevant w.r.t. the query image. Each database image was manually marked as relevant or irrelevant on the basis of our query image. The results are presented in Table 1.The figures 10, 11, 12, 13, 14 and 15 show the statistics and query performance.

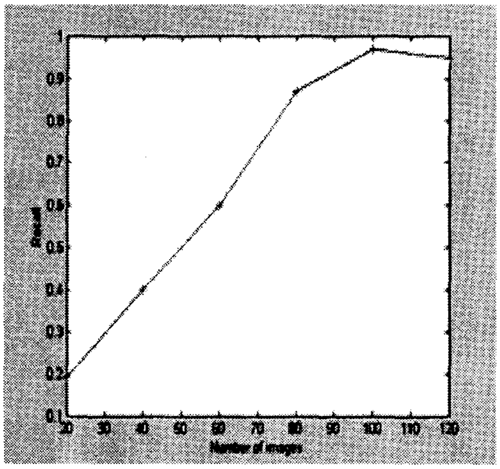




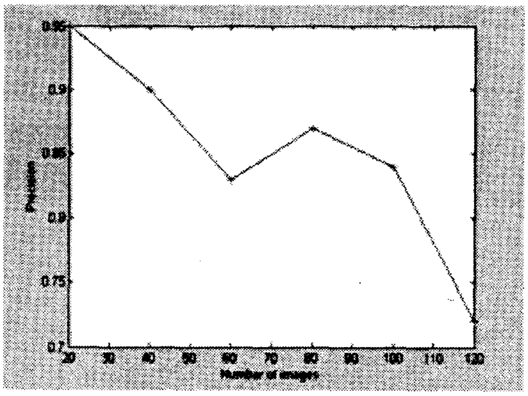
(Figure 13) Shows the query performance (MINERVA database)



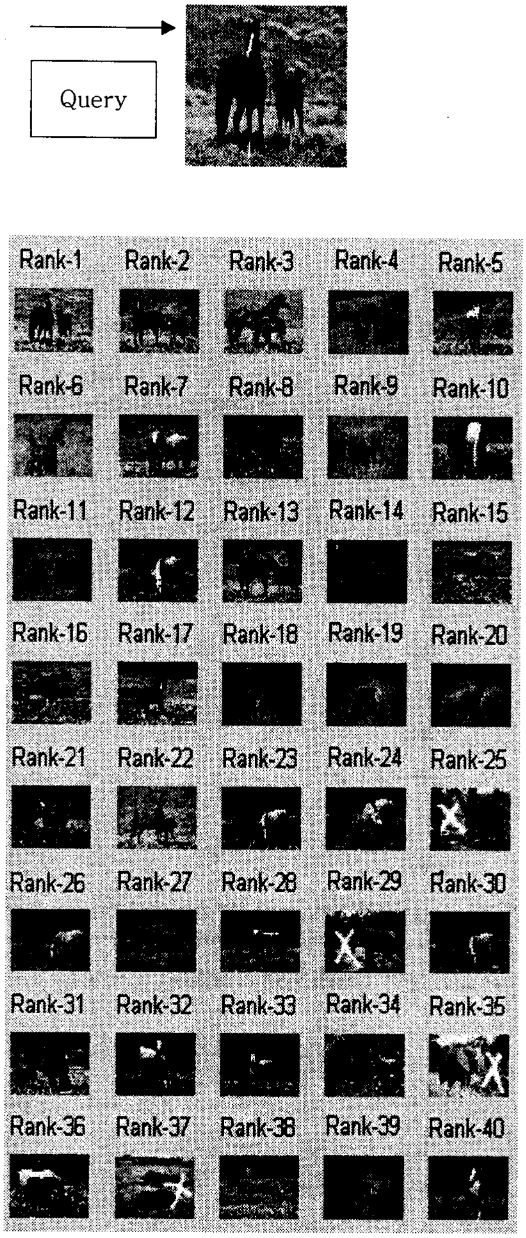
(Figure 10) Error rate versus number of images graph



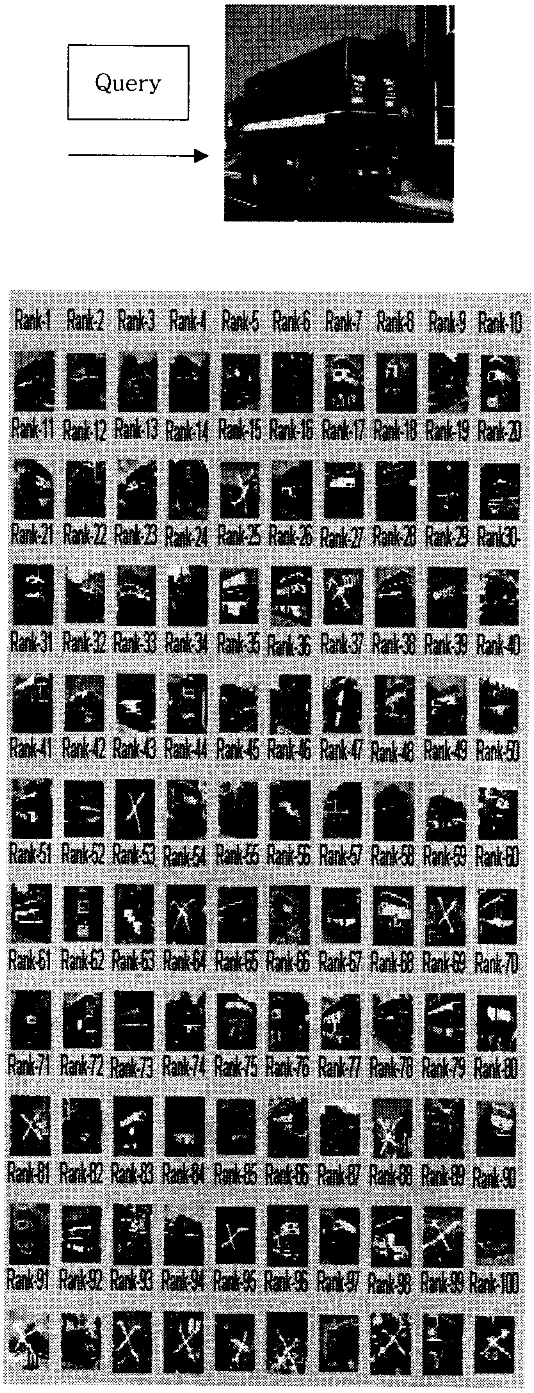
(Figure 11) Recall versus number of images graph



(Figure 12) Precision versus number of images graph



(Figure 14) Shows the query performance (SIMILIcity database)



(Figure 15) Shows the query performance (SIMILIcity database)

8. Conclusion

The working efficiency of our system with theses combined features is much better than other systems developed so far such as color-based indexing. The retrieval performance of the system depends upon color space, quantization scheme and mode of integration of the features, We have seen in our experiment that the precision can be kept high by retrieving only a few images and we can always make the recall 1, simply by retrieving all images. The performance of mpeg-7 edge histogram could improve by employing some adaptive thresholding.

In our near future work we are looking forward to integrating MPEG-7 edge histogram descriptor, color histogram and contour shape descriptor.

References

[1] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, *Query by image and video content: the QBIC system, Computer, vol. 28,* no. 9, pp. 23-32, Sept. 1995.

[2] A. Gupta and R. Jain, *Visual information retrieval, Comm. Assoc. Comp. Mach., vol. 40,* no. 5, pp. 70-79, May 1997.

[3] A. Pentland, R. W. Picard, and S. Sclaroff, Photobook: Content-based manipulation of image databases, *SPIE Storage and Retrieval Image and Video Databases II,* San Jose, 1995

[4] B. S. Manjunath (Editor), Philippe Salembier (Editor), Thomas Sikora (Editor), Introduction to MPEG-7: Multimedia Content Description Interface, (John Wiley & Sons Ltd., England-2002)

[5] Iain E. G. Richardson, *Video Codec Design* (John Wiley & Sons Ltd., England-2002)

[6] Linda G. Shapiro, George C. Stockman, *Computer Vision*(Upper Saddle River, NJ: Prentice-Hall, 2001).

[7] T. Sikora, *The MPEG-7 Visual Standard for Content Description-An Overview, IEEE*Transaction on circuits and video technology, VOL.II, No.6, June 2001

[8] Y. Rui, T. S. Huang, S. Mehrotra, and M. Ortega, A relevance feedback architecture in content-based multimedia information retrieval systems, in *Proc. of IEEE Wrkshop on Content-Based Access of Image and Video Libraries,* in Conjunction with IEEE CVPR'97, 1997.

[9] V. Athitsos, M. J. Swain, and C. Frankel, Distinguishing, photographs and graphics on the world wide web, in *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries,* 1997.

[10] S. Sclaroff, L. Taycher, and M. La Cascia, Imagerover: A content-based image browser for the world wide web, in *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries,* 1997

[11] James Z. Wang, Jia Li, Gio Wiederhold, SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries, *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol 23, no.9, pp. 947-963, 2001.

[12] I. Witten, A.Moat, and T. Bell (1999) Managing Gigabytes. Morgan Kaufmann, Second edition.

[13] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, Blobworld: *A system for region-based image indexing and retrieval, Third Int. Conf. on Visual Information Systems,* June 1999.

[14] M.Singh and S. Singh, Minerva scene analysis benchmark, Proc. 7th Australian and New Zealand Intelligent Information Systems Conference, pp. 231-235, Perth, 2001.

[15] K. Singh, M. Ma and D.W. Park, *A Content-Based Image Retrieval using FFT & Cosine Similarity Coefficient,* 5th IASTED Int. Conf. on Signal and Image processing (SIP-2003), Hawaii, Aug 13-15, 2003.

[16] R.C. Gonzalez, R.E.Woods, *Digital Image Processing* (Indian Branch- Delhi, Addison Wesley Longman (Singapore) Pte Ltd., 2000)

[17] H. Muller, W. Muller, D.M. Squire, S.M. Maillt, T. Pun, *Performance evaluation in content-based image retrieval: overviews and proposals,* Pattern Recognition Letters 22,593-601,2001

[18] N. Otsu. *A threshold selection method from gray-level histograms,* IEEE Transactions on Systems, Man, and Cybernetics (SMC-9), 9(1): 62-66, 1979.

안성옥

1983년 2월 고려대학교수학교육과졸업(이학사).
1985년 2월 고려대학교 대학원 수학과 졸업
(전산학전공 석사).
1989년 8월 고려대학대학원 수학과 졸업
(전산학전공 박사).
1991년~현재 배재대학교 IT공학부 교수.
1993년~1994년 펜실베이니아 주립대학교
컴퓨터학과 방문교수.
1994년~1995년 뉴욕주립대학교
컴퓨터학과 방문교수
관심분야: 데이터베이스, 컴퓨터교육, 영상처리
게임공학
E-Mail: sungohk@pcu.ac.kr

박동원

1983 고려대학교 전기공학과
(공학사)
1985 Florida Institute of
Technology 공학석사
(전산학 전공)
1993 Texas A&M University
공학박사 (전산학 전공)
1994 ~ 현재 배재대학교 IT공학부 부교수

관심분야 : 실시간통신, 영상처리
E-Mail : dwpark@pcu.ac.kr

Ming Ma

1997 Dept. of Computer
Science and Information
Engineering, Inner
Mongolia Agricultural
University, China B.E
2002 Dept. of Information and
Communications Engineering Paichai
University, Korea M.S
2004 ~ Dept. of Information and
Communications Engineering Paichai
University, Korea
Research Interests:
Multimedia information retrieval,
Image processing and computer vision.
E-Mail : ming@pcu.ac.kr

Kulwinder Singh

1996 Dept. of Electronic
Engineering,
Ravishanker Shukla
University, India B.E
2002 Dept. of Information and
Communications Engineering Paichai
University, Korea M.S
2004 ~ Dept. of Information and
Communications Engineering Paichai
University, Korea
Research Interests:
Multimedia information retrieval,
Biometric Authentication.
E-Mail : singh@pcu.ac.kr