Content-Based Color Image Retrieval Using Multi-Variate Feature Vectors

Hideaki Kokubun and Hiroaki Kotera, Graduate School of Science and Technology, Chiba University, Inage-ku, Chiba, Japan

Abstract

This paper introduces a content-based image retrieval system which utilizes color, spatial frequency, and structural features of an object in the image. The major key color areas are extracted by image segmentation applying Bayesian classifier with k-means starter in CIELAB space and the color similarity is measured by summing up the inter-cluster color distances. The mutual correlations in DCT components are used to discriminate the spatial frequency features that carry textural details. In addition, a structural feature of image is simply characterized by down sampled color mosaic pattern. A target color image is retrieved by searching the image with the maximum similarity by cross correlations in the multi-dimensional feature vector space. Color similarity was indispensable to narrow the reliable candidate for almost all the tested images. Although an appropriate combination of spectral or structural features worked effective to retrieve the image with textural similarity, but we have not any definitive solution to select the frequency region yet. Finally, we report a method for evaluating the performance of our system based on psychophysical experiments using z-score.

Introduction

Interest in the potential of digital images has increased enormously over the last few years, as the rapid growth in color imaging applications on the world-wide web sites. Users in many professional fields are exploiting the opportunities offered by the ability to access and manipulate remotely-stored images. So it has required system that desired images can be efficiently retrieved from a huge image database.

In contrast to the text-based image retrieval so far, content-based image retrieval (CBIR) is the system that retrieve similar images from database by comparing features (e.g. color, shape, or structure) automatically extracted from the images themselves. A typical current CBIR system allows users to formulate queries by submitting an example of key image to be retrieved. The system identifies those stored images whose feature values match those of the query most closely, and displays thumbnails of these images on the screen. Fig.1 illustrates the overview of CBIR using multi-dimensional color image features.

Multi-dimensional Feature Vectors

Multi-dimensional image features have been extracted to compare a key image with database images. We utilize the following three major image features.

[1] Mutual distance between image ***Key Colors***
[2] Cross correlation in low dimension Fourier ***Spectra***
[3] Similarity in down sampled rough image ***Structure***

Color Feature Extraction

The major key color areas are extracted by image segmentation applying Bayesian classifier with k-means starter in CIELAB space. Here, we have located K seeds points at the body center in the regularly divided cubes with the higher color population in CIELAB space. Next, the image is segmented into K pieces of key color areas by ***k-means*** method classifying each pixel nearest to these seeds points and renewing the seeds points for their member's statistical deviations to be distributed in uniform. After converging of iterations in k-means process, K sets of gravity center are applied to ***Bayesian*** decision rule and fixed as color feature vector.

The color feature vector $\boldsymbol{C}_i$ of image i is given by

$$\boldsymbol{C}_i = \left[\boldsymbol{C}_{ik}\right],\ \boldsymbol{C}_{ik} = \left[L^*{}_{ik}, a^*{}_{ik}, b^*{}_{ik}\right];\ i = 1 \sim N,\ k = 1 \sim K \qquad (1)$$

where, C_{ik} denotes the gravity center of segmented key color class k for image i. C_i is stored in color feature data base file. Figure 2 shows a stamp image segmented to K = 3 classes.

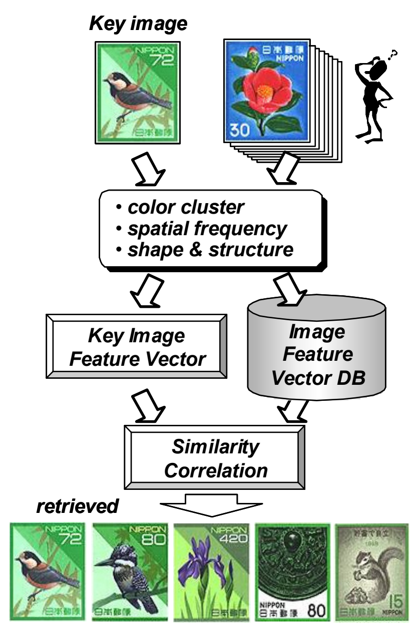

Figure 1. Content-based Image Retrieval System

The mutual color distance between the color feature vectors $\boldsymbol{C}_i$ and $\boldsymbol{C}_j$ is simply measured by the summation of Euclidian distances in gravity color vectors for class k of image i and class l (k, $l=1\sim K$) of image j.

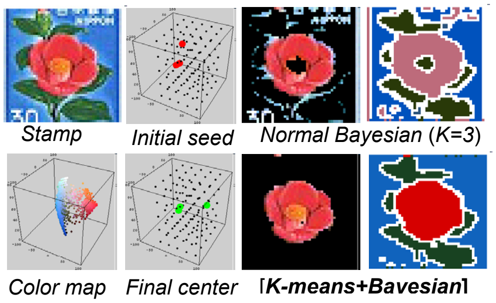

Figure 2. Key color extraction by **[k-means+Bayesian]**

$$\Delta C_{ij} = \sum_{k=1}^{K} \sum_{l=1}^{K} \left\| C_{ik} - C_{jl} \right\| \tag{2}$$

Since ΔC_{ij} takes the minimum value for the image j closest to the searching image i, we introduce the normalized color similarity factor as follows.

$$\rho_{col}(i,j) = 1 - \left[\Delta C_{ij} \Big/ \max_{j=1\sim N} \left\{ \Delta C_{ij} \right\} \right] \tag{3}$$

ρ_{col} takes the maximum value 1 for the same image pair of $i=j$ and the minimum value 0 for the most distant pair.

Figure 3 illustrates the color distance measure between two stamp images segmented to $K=3$ key color areas in each.

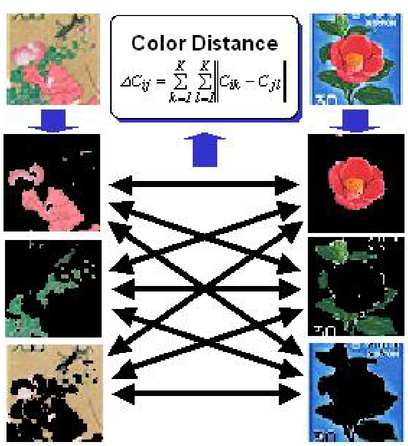

Figure 3. Measuring mutual color distance

Spectral Feature Extraction

The Fourier spectra of image carry the spatial frequency characteristics, which reflect the textual features. Here DCT (Discrete Cosine Transform) was applied to extract the image spectra, because DCT is easy to calculate and gives real values without imaginary part.

RGB image was transformed luminance-chrominance YCC image and DCT was only applied to the luminance Y. The luminance image g is transformed into spatial frequency components G by forward DCT as

$$\boldsymbol{G} = \left[G_{uv} \right] = \boldsymbol{A}^t \boldsymbol{g} \boldsymbol{A} \tag{4}$$

In case of $M \times M$ image, the transform matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \left[a_{uv} \right], \quad a_{uv} = \begin{cases} \dfrac{1}{\sqrt{M}}, & \text{for } v = 1 \\ \dfrac{2}{\sqrt{M}} \cos\left(\dfrac{(2u-1)(v-1)\pi}{2M} \right) & \end{cases} \tag{5}$$
$$\text{for } u = 1,\cdots,M, \; v = 2,3,\cdots,M$$

In general, the power spectra of 2-D natural images are mostly concentrated in the lower spatial frequencies. In order to extract the effective frequency regions and to simplify the computations, 2-D DCT coefficients $\{G_{uv}\}$ were transformed into 1-D spectral vector $\boldsymbol{G}_P$ in the fan-shaped spatial frequency domain as follows.

$$\mathbf{G}_P = \left\lfloor G_p \right\rfloor; \; p = 1,2,\cdots,P$$
$$G_p = \sum_u \sum_v G_{uv} \tag{6}$$
$$\text{for } f_L \le f \le f_H \; ; \; f = \sqrt{u^2 + v^2}$$
$$\frac{2K}{\pi} \tan^{-1}(v/u) \le p \le 1 + \frac{2K}{\pi} \tan^{-1}(v/u)$$

Figure 4 illustrates the overview of $\boldsymbol{G}$ to $\boldsymbol{G}_P$ transformation. First, the effective DCT spectra are extracted in the limited range of radial frequency: $f_L \le f \le f_H$.

Secondly, the circular doughnut region is divided into P sectors by tangential angle $\theta = tan(v/u)$. Finally, the DCT coefficients within the each sector p are summed up to G_p; $p=1\sim P$. Thus, $M \times M$ 2-D DCT coefficient matrix is simply transformed to *P-dimensional* 1-D vector $\boldsymbol{G}_P$.

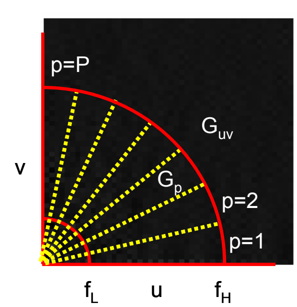

Figure 4. Reduced 1-D spectral vector in segmented fan-shaped domain of 2-D DCT

The similarity of spatial frequency distribution between the two images $\boldsymbol{g}$ and $\boldsymbol{h}$ is measured by the following cross correlation coefficient for 1-D spectral vectors $\boldsymbol{G}$ and $\boldsymbol{H}$ as follows.

$$\rho_{DCT}\left(G,H\right)=\frac{\sum_{p=1}^{P}\left(G_p-\overline{G}\right)\left(H_p-\overline{H}\right)}{\sqrt{\sum_{p=1}^{P}\left(G_p-\overline{G}\right)^2}\sqrt{\sum_{p=1}^{P}\left(H_p-\overline{H}\right)^2}} \qquad (7)$$

Structural Feature Extraction

Besides the color and spectral features, the structural sketch is often useful to retrieve the images intuitively. The simplest way to catch the rough image structure is to use down sampling. The color image with $M \times N$ pixels is resized to $m \times n$ $(=Q)$ small mosaic color pattern by down sampling and its RGB values are sequentially arranged into Q color pixel arrays and used as structural feature. Letting the mosaic pattern for image $\boldsymbol{g}$ be

$$\Omega_g=\left[{}_g\omega_1,\ {}_g\omega_2,\cdots,\ {}_g\omega_Q\right] \qquad (8)$$

The similarity between the mosaic patterns Ω_g and Ω_h for image $\boldsymbol{g}$ and $\boldsymbol{h}$ is given by the cross correlation coefficient

$$\rho_{mos}\left(\Omega_g,\Omega_h\right)=\frac{\sum_{j=1}^{Q}\left({}_g\omega_j-{}_g\overline{\omega}\right)\left({}_h\omega_j-{}_h\overline{\omega}\right)}{\sqrt{\sum_{j=1}^{Q}\left({}_g\omega_j-{}_g\overline{\omega}\right)^2}\sqrt{\sum_{p=1}^{P}\left({}_h\omega_j-{}_h\overline{\omega}\right)^2}} \qquad (9)$$

Figure 5 shows $Q=9\times 12$ mosaic patterns applied for Mask.

Figure 5. Structural feature by down-sampled mosaic

Experimental Result

Retrieval by Single Feature and Mixed-Features

We conducted the image retrieval experiments on the test image sets in Fig. 6, using the following combinations of similarity in multi-dimensional feature vectors.

[a] Color	ρ_{col}
[b] Spectrum	ρ_{DCT}
[c] Structure	ρ_{mos}
[d] Color + Spectrum:	$\rho_{mix}=1/2(\rho_{col}+\rho_{DCT})$
[e] Color + Structure:	$\rho_{mix}=1/2(\rho_{col}+\rho_{mos})$
[f] Color +Spectrum +Structure:	$\rho_{sum}=1/3(\rho_{col}+\rho_{DCT}+\rho_{mos})$

Figure 6. Test image sets

Retrieved Results

Figure 7 shows an examples of “Stamp” and ”Mask” image retrieval.

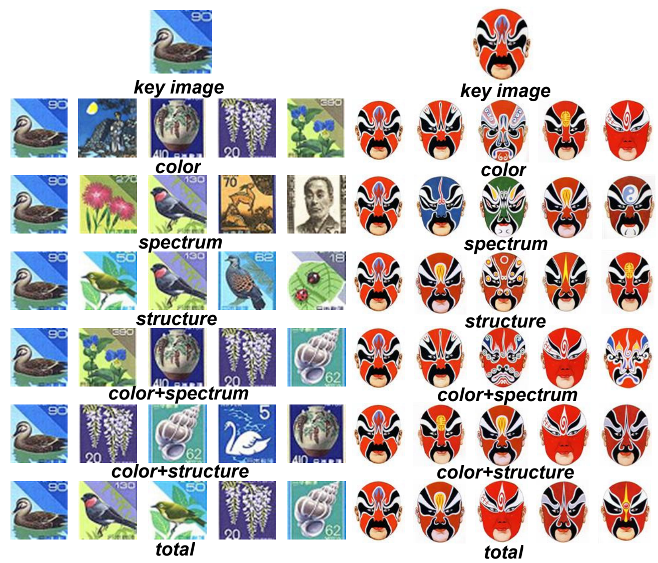

Figure 7. Retrieved example

Each parameter is as follows.

	“Stamp”	“Mask”
Color:	K=3	K=3
Spectrum:	$f_L=2, f_H=24, P=16$	$f_L n=2, f_H=24, P=16$
Structure:	$Q=m\times n=8\times 8=64$	$Q=m\times n=9\times 12=108$

Sometimes, these three key features may work in effective when used independently, but may be more robust when well combined. For example, the structural feature by mosaic worked well but spectral feature didn’t always available when used in single.

Evaluation Method

We suggest the following evaluation methods. We have five images considered to resemble in key image voted by subject from image database, and take out five images ranked higher. To evaluate the selected images, we applied paired comparison method for all combinations in order to calculate Z-score, and decided the order of images. Z-score is shifted so that the minimum value takes 1.0 and the value is defined as a ***point*** gained by the image. Next, we do image retrieval by our CBIR system and keep the following ***weights*** for an image at the five places: 1st: × 5, 2nd: × 4,…, 5th: × 1. Five images provided with Z-score are compared with five images retrieved from our CBIR system. If there is any image to match, we define a ***total point*** as follows: ***total point*** = ***point*** × ***weight***. Total point is normalized and becomes ***evaluation value***.

Evaluation Experiment

The evaluation experiment was done in the next condition.

Figure 8. Key image and voted images

Z-score provided by pair comparison is shown for Fig. 9. Figure 10 shows evaluation value. A color feature was most important in this sample because the evaluation value of total is lower than a feature alone.

Conclusions

A content-based image retrieval system using multi- dimensional image feature vectors is proposed. We introduced a simple quantitative evaluation method for the performance of CBIR system by coupling the rank ***points*** with z-score based on psychophysical experiments.

The future works should be continued to evaluate the validity of proposed system for a large set of image database.

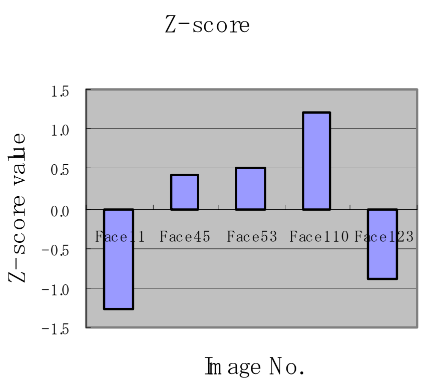

Figure 9. Z-score

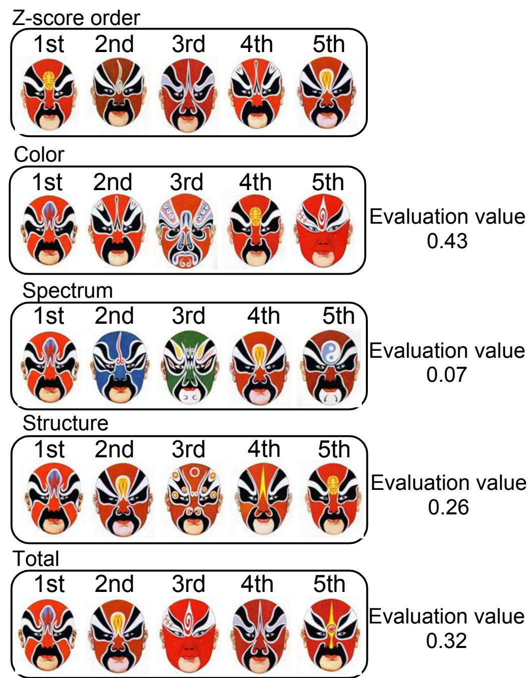

Figure 10. Evaluation result

References

1. S. Asami and H. Kotera, "Hierarchical Image Retrieval System Using Multi-dimension Characteristic Vectors", Proc. Color Forum Japan'02, (2002)
2. S. Asami and H. Kotera, "Content-Based Image Retrieval System By Multi-dimension Feature Vectors", Proc. IS&T's NIP 19, 841- 845 (2003)

Author Biography

Hideaki Kokubun received his B.E. degree in Information and Image Sciences from Chiba University, Japan in 2004. Since 2004, he has been a Master student in Graduate School of Science and Technology of the same university. His current research interests include color image processing and its application to image retrieval.