Color Image Retrieval Based on Weighted DCT Spatial Combination Histogram

Yonghua Xie[1,2], Lokesh Setia[2], and Hans Burkhardt[2]

[1] Institute of Computer and Software, Nanjing University of Information Science and Technlolgy, P.R. China
xyh_76@nuist.edu.cn

[2] Institute of Computer Scinece, University of Freiburg, Germany
{Xie, Setia, Burkhardt}@informatik.uni-freiburg.de

Abstract

Color histograms are the most commonly used features for image retrieval due to the good attributes of image translation and rotation invariance. However the traditional color histogram features are usually short of spatial structural information which may cause error retrieval results. This paper presents a color image retrieval method based on Weighted DCT Spatial Combination Histogram (WDSCH). The Equalized DCT coefficients histograms for each divided image block are firstly constructed with the DCT transformation and histogram equilibrium method. The Weighted DCT Spatial Combination Histogram features for the entire image are then extracted in the HSV color space with the weights of the spatial information and energy distribution. Experimental results on the three image datasets with different conditions show that the compact histogram features containing spatial structural relations are robust to image translation, scaling and rotation, and can bring about good retrieval precision and speed simultaneously.
Keyword: HSV color space, block DCT transformation, Weighted DCT Spatial Combination Histogram (WDSCH), color image retrieval

I. Introduction

In recent years, many great efforts have been made to research on the content-based image retrieval (CBIR) methods due to the dramatically increasing amount of available multimedia data. The goal of content-based image retrieval systems is to extract a set of similarity features, such as color, shape and texture features which can effectively characterize the visual content of images, and then use them for retrieval purpose, which is called query by image example. The features should be robust enough for image intrinsic content representation and simple enough for practical retrieval application (Aigrain and Zhang et al, 1996).
Histogram is the most commonly used scheme to represent the features composition of an image. It is invariant to image translation and rotation and can also be normalized to have invariance of image scaling. Generally speaking, the existing histogram features can be divided into two categories: spatial-domain histogram and frequency-domain histogram (Datta and Li, 2005). The most widely used spatial-domain histogram is the global color histogram which represents the color composition and distribution of the whole image (Swain and Ballard, 1991). The main disadvantages of the conventional color histograms are the ignorance of any spatial structural relations, which will cause

the completely disrelated images with the same histogram to be matched (Park and Jeon et al, 2000). Recently many contributions tried to alleviate these problems. For example, Jain et al. combined color histogram with shape information for image retrieval (Jain and Vailaya, 1996). Timothy et al. proposed an intelligent content-based image retrieval system based on color, shape and spatial relations (Timothy and Huang, 2001). Siggelkow et al. presented the principle of integral invariants to construct a global invariant feature histogram (Siggelkow and Schael et al, 2001). The frequency-domain histograms are also called compression-domain histograms, which are normally created directly by the frequency coefficients obtained with the frequency-domain transformation methods, such as DCT coefficients. For example, Lay Joes et al. proposed an image retrieval method based on energy histograms of low frequency DCT coefficients (Lay Joes and Guan, 1999). Wu et al. used subband energy histograms of reordered DCT coefficients for image retrieval (Wu et al, 2002). Lu et al. introduced a DCT-domain vector quantization index histogram based on four codebooks for color image retrieval (Lu and Burkhardt, 2005). However these histogram features are constructed directly by the DCT coefficients of different image blocks, which makes them lack spatial structural information and vulnerable to objects translation. Furthermore, the entire high-dimensional histogram features will enhance the computational complexity greatly, which makes them inapplicable for practical retrieval fields.

To extract the more effective and robust histogram features for retrieval purpose, we take both frequency-domain compressed histogram and spatial structural information into consideration, and then present an image retrieval method based on weighted DCT spatial combination histogram (WDSCH). Experimental results on three image datasets with different conditions show that the compact histogram containing spatial structural relations is robust to image translation, scaling and rotation, and can bring about good retrieval precision and speed simultaneously.

This paper is organized as follows: Section II introduces the HSV color space used for representing color images. The process of constructing weighted DCT spatial combination histogram is explained in section III. Section IV presents the similarity measure and performance evaluation methods. The experimental results and conclusion are provided in section V and section VI respectively.

II. HSV Color Space and Image Representation

The models of human perception of color differences are described in the form of color spaces, so the research on color image application must be done in a given color space. YIQ, LSH, HSV, RGB, HIS etc. are the most frequently used color spaces (Thorell and Smith, 1990).

Among the above color spaces, HSV color space is an intuitive color description method which quite accords with human visual perception. In general, HSV color space has two distinct characteristics: Firstly, the hue is invariant to the changes in illumination and camera direction and hence more invariant to object retrieval. Secondly, RGB coordinates can be easily and quickly translated to the HSV coordinates by a simple formula. So we choose HSV color space to represent a color image. The transformation from RGB to HSV is defined as follows:

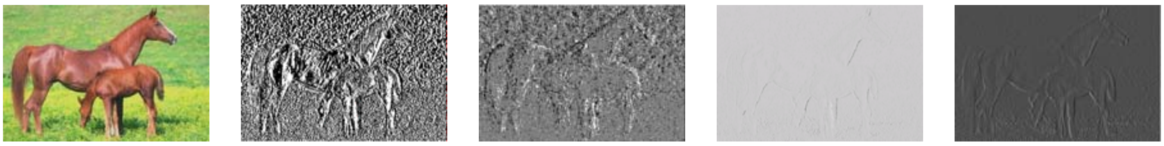
$$S = \frac{V - \min(R,G,B)}{V} \tag{1}$$

$$V = \max(R,G,B) \tag{2}$$

$$H = \begin{cases} 5+\overline{B}, & if\ \ R=\max(R,G,B)\ and\ G=\min(R,G,B) \\ 1-\overline{G}, & if\ \ R=\max(R,G,B)\ and\ G\neq\min(R,G,B) \\ 1+\overline{R}, & if\ \ G=\max(R,G,B)\ and\ B=\min(R,G,B) \\ 3-\overline{B}, & if\ \ G=\max(R,G,B)\ and\ B\neq\min(R,G,B) \\ 3+\overline{G}, & if\ \ B=\max(R,G,B)\ and\ R=\min(R,G,B) \\ 5-\overline{R}, & others \end{cases} \tag{3}$$

where $\overline{R} = \dfrac{V-R}{V-\min(R,G,B)}, \overline{G} = \dfrac{V-G}{V-\min(R,G,B)}, \overline{B} = \dfrac{V-B}{V-\min(R,G,B)}$

Let $f = \{f(x,y)\}, 1 \le x \le M, 1 \le y \le N$ be the original color image. Considering that the difference image $f^{'}(x,y) = f(x,y) - f(x,y+1)$ usually follows the generalized laplacian distribution and can reduce the redundant correlation of the two closer pixels, we choose it to be the experimental image for retrieval purpose. Then we can use f_H, f_S and f_V to represent the hue, saturation and value components of the original difference color image. Figure 1 gives the difference image and its three corresponding HSV components of one original image from MPEG-7 image database.



(a) Original image (b) Difference image (c) Hue (d) Saturation (e) Value

Fig.1 Difference image representation in HSV color space

III. Weighted DCT Spatial Combination Histogram (WDSCH)

A. *Equalized Block DCT Histogram*

The Discrete Cosine Transform (DCT) has been proved successful at separating and concentrating the energy of image data. Though several descriptors for image retrieval have been proposed in MPEG-7 (Koskela and Laaksonen et al, 2002), the requirements for techniques that can directly use DCT coefficients to perform querying and retrieving become important as to reduce the computational complexity. In order to extract the intrinsic and localized features, the image should firstly be divided into several square blocks of 64 pixels before operated by DCT according to the JPEG and MPEG standard. But features extraction for image retrieval application should not be considered as a simple compression task. Too many blocks will increase the vectors dimensionality greatly, which makes them not applicable for practical retrieval on the large image database. Considering the energy distribution of the color image, we usually divide the images with the division methods of 4×4, 6×6, and 8×8 and so on. Note that we use the HSV color space to denote each color image component. In the following description, we only give one example for vectors construction on one of the three components. Let $C = \{C_1, C_2, \ldots, C_K\}$ be the divided blocks set of one image component, where K denotes the number of blocks. Assumed that the size of each block is $m \times n$, then DCT is performed on each image block in set C to obtain the transformed DCT coefficients block set $D = \{D_1, D_2, \cdots, D_K\}$. In order to reduce the computational complexity, we rearrange the transformed DCT block coefficients matrix from the two-dimensional array to the one-dimensional array in the zig-zag sequence, thus we can obtain the corresponding one-dimensional vectors set $E = \{E_1, E_2, \cdots, E_K\}$. Due to that the DCT coefficients probability distribution of one given image, i.e. DCT coefficients histogram, will keep invariant to image rotation and translation, we use it to represent the image features for retrieval application.

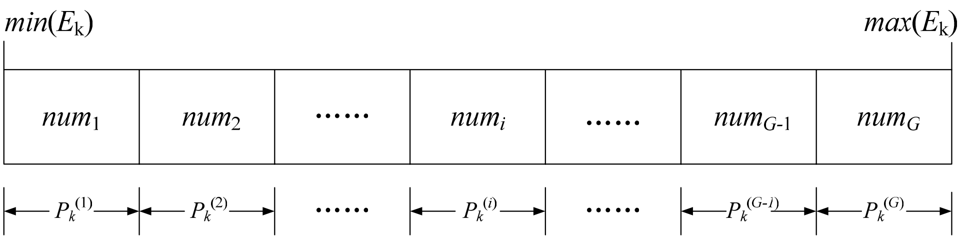


Fig.2 Bins partition of block DCT coefficients

As is shown in Fig.2, let E_k be the one-dimensional DCT coefficients vector of the image block $k, 1 \le k \le K$, then the DCT coefficients histogram can be defined as follows. Firstly, we find the maximum and minimum DCT coefficients of the E_k vector which are described as $\max(E_k)$ and $\min(E_k)$ respectively. Then the interval $[\min(E_k), \max(E_k)]$ can be partitioned into G subset bins according to the coefficients energy distribution. Secondly, let num_i denote the number of the coefficients located in the i subset bin, the corresponding DCT coefficients probability distribution can be defined as

$$P_k^{(i)} = \frac{num_i}{m \times n} \tag{4}$$

where $1 \le i \le G$. Based on the above steps, the DCT coefficients histogram vector of the block k can be defined as

$$H_k = \{P_k^{(1)}, P_k^{(2)}, \cdots, P_k^{(G)}\} \tag{5}$$

However, DCT transform has the performance of separating and concentrating the principal energy of the image to some few special coefficients, which will bring forth high dimensionality and large-scale dynamic variation of the original DCT coefficients histogram vector. Some traditional methods usually choose the bins of primary DCT coefficients probability distribution, such as DC coefficients, to construct histogram vector (Borko and Saksobhavivat, 1998). It has been proved that these methods can reduce the vector dimensionality and image yawp disturbance, but will lose some useful and discriminative features at the same time. To obtain more efficient features with low dynamic variation, this paper uses the histogram equilibrium method to equalize the original DCT coefficients histogram. With the following transformation function

$$\overline{P}_k^{(i)} = \sum_{j=1}^{i} P_k^{(j)} = \sum_{j=1}^{i} \frac{num_j}{m \times n} \tag{6}$$

the transformed histogram of block k can be constructed as $L_k = \{\overline{P}_k^{(1)}, \overline{P}_k^{(2)}, \cdots, \overline{P}_k^{(G)}\}$. Due to that the dimensionality of original DCT coefficients histogram is G, we proceed to equalize the transformed histogram L_k with G-scale equilibrium method defined as follows

$$Q_k^{(t)} = \frac{1}{G} \sum_{r,s} (\overline{P}_k^{(r)} + \overline{P}_k^{(s)}) \qquad \text{if } \left|\overline{P}_k^{(r)} - \overline{P}_k^{(s)}\right| \big/ G \le T_0 \tag{7}$$

Here T_0 denotes a preset small threshold which restricts the random two elements $\overline{P}_k^{(r)}$ and $\overline{P}_k^{(s)}$ in L_k with the same probability distribution, where $1 \le r, s \le G$, $1 \le t < G$. Then we can obtain the equalized block DCT coefficients histogram vector, which can be described as

$$U_k = \left\{Q_k^{(1)}, Q_k^{(2)}, \cdots, Q_k^{(Z)}\right\} \tag{8}$$

where Q_k denotes the equalized DCT coefficients probability distribution of block k. Z is the scale of the equalized probability distribution which satisfies $1 \le Z < G$. Furthermore, in order to make the equalized DCT coefficient histogram invariant to different image scaling, it should continue to be scaling normalized with the following equation

$$Y_k = U_k / \delta_k \tag{9}$$

Thus we can obtain the final normalized histogram vector $Y_k = \left\{Y_k^{(1)}, Y_k^{(2)}, \cdots, Y_k^{(Z)}\right\}$, where δ_k is the mean of the equalized histogram vector, which can be written as $\delta_k = \frac{1}{Z} \sum_{l=1}^{Z} Q_k^{(l)}$.

Fig.3 gives the construction process of the equalized block DCT coefficients histogram vector for an image block. It can be seen that the proposed histogram features can effectively describe the energy distribution of the image block with highly compact dimensionality and low dynamic variation.

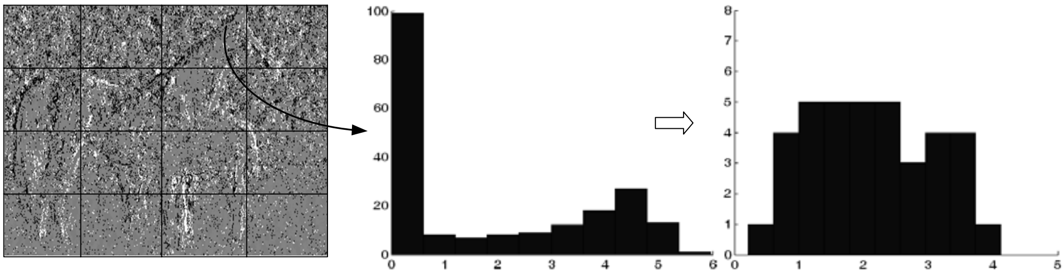


Fig.3 Diagram of constructing equalized block DCT coefficients histogram

B. Weighted DCT Spatial Combination Histogram

Although the dimensionality and scaling equalized block DCT histogram vector is compact, the dimensionality of one component of the entire image will still reach $K \times Z$ if the DCT histogram vectors of each block are combined together directly, where K denotes the number of the image blocks and Z is the dimensionality of the equalized block DCT histogram vector. Furthermore, this simply combined DCT histogram vector is also short of spatial structural relations which will lead to the mismatch errors as the traditional color histogram methods. Due to that the average difference of the most closer two components of the original DCT histogram vector can reflect the energy distribution and gray value variance of the image, we use it to be the spatial information weights for constructing the weighted DCT spatial combination histogram vector.

For the original DCT histogram vector $H_k = \{P_k^{(1)}, P_k^{(2)}, \cdots, P_k^{(G)}\}$ of block k given in Equation (5), the spatial information weights can be defined as

$$W_k = \sum_{\lambda=1}^{G/2} (P_k^{(2\lambda)} - P_k^{(2\lambda-1)}) \Big/ \sum_{\lambda=1}^{G} P_k^{(\lambda)} \tag{10}$$

Then the weighted DCT spatial combination histogram can be obtained by applying the spatial information weights on each equalized histogram block vector

$$F = \sum_{k=1}^{K} W_k Y_k \tag{11}$$

Due to the fact that there are three color components f_H, f_S and f_V of the original color image $f(x, y)$, where $1 \le x \le M$,$1 \le y \le N$, M, N denote the resolution of the image, thus we can obtain the corresponding three weighted DCT spatial combination histogram vectors F_H, F_S and F_V with the same steps mentioned above. Then the three histogram vectors are combined into the one-dimensional histogram vector, which can be described as

$$Hist = \{\Phi_H F_H, \Phi_S F_S, \Phi_V F_V\} \tag{12}$$

Obviously, the vectors dimensionality of the entire color image is $3 \times Z$. Φ_H, Φ_S and Φ_V are the energy distribution coefficients for the three color components, which can be computed as

$$\Phi_\sigma = \sum_{x=1}^{M} \sum_{y=1}^{N} f_\sigma(x, y) \Big/ \sum_{x}^{M} \sum_{y}^{N} f(x, y)\,;\ \sigma \in \{H, S, V\} \tag{13}$$

IV. Similarity Measure and Performance Evaluation

Instead of exact image matching, content-based image retrieval calculates visual similarities between the query image and the test images in an image database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. In our experiment, we adopt the histogram intersection measure to compute the similarities between the query image and the test images, which is defined as

$$sim(I,T)=\frac{\sum_{\rho=1}^{X}\min(Hist^{(\rho)}(I),Hist^{(\rho)}(T))}{\min(\sum_{\rho=1}^{X}Hist^{(\rho)}(I),\sum_{\rho=1}^{X}Hist^{(\rho)}(T))} \tag{14}$$

Here X denotes the total number of the components in the weighted spatial combination histogram vector. $Hist^{(\rho)}(I)$ and $Hist^{(\rho)}(T)$ are the ρ^{th} component of the query image I and test image T respectively. Then the retrieved images will be ranked in sequence with the computed similarities.
Furthermore, in order to evaluate the performance of our proposed method, we adopt two performance evaluation methods of retrieval precision and recall rate, which are usually defined as

$$\text{Recall rate}=\frac{n_r}{n_{db}} \tag{15}$$

$$\text{Retrieval precision}=\frac{n_r}{n_c} \tag{16}$$

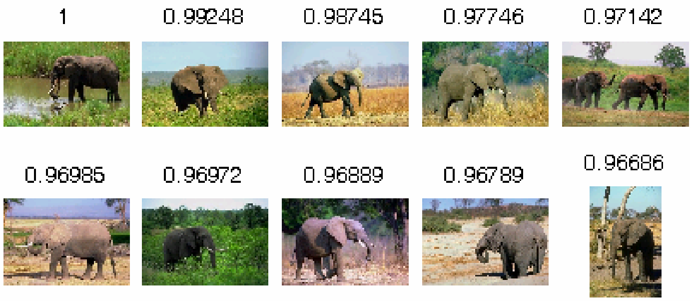
Here n_r is the number of the matched relevant images; n_{db} is the total number of the relevant images in the image database; n_c is the number of the returned images. Then a graph of precision versus recall (PVR) can be created to evaluate the image retrieval performance.

V. Experimental Results

We carry out the experiments on three image datasets of Dataset Ⅰ, Dataset Ⅱ and Dataset Ⅲ. The images in Dataset Ⅰ are directly from a subset of MPEG-7 database with 1000 images. These images are with the size of 384×256 or 256×384 which are categorized into 10 classes including people, buildings, horses, flowers and so on. In order to validate the proposed features robust to image rotation and translation, we rotate each image in Dataset Ⅰ anticlockwise with the angles of 45^0, 90^0, 180^0 and 270^0 to construct the Dataset Ⅱ with 5000 images, in which there are 500 images for each class. Similarly, for testing the proposed features robust to the variety of image scaling, we shrink each image in Dataset Ⅰ with the scaling of 20%, 40% and 60% to construct the Dataset Ⅲ with 4000 images, each class also has 400 images. All the experiments are performed on a Pentium IV computer with 2.4 GHz CPU and 512 MB memory. In view of the computational complexity, the number of original DCT coefficients histogram bins G is set as 50, thus the dimensionality of the final histogram vector used for retrieval experiments can be limited within 150. For testing purpose, the image division methods are chose as 4×4, 8×8 and 12×12, thus we can obtain 16, 64 and 144 divided image blocks respectively for each component of the original color image.
The detailed process for computing the average recall rate (ARR) and average retrieval precision (ARP) are described as follows: Firstly, we successively select each image from different class to be the query image, and all of the images in each corresponding dataset are used for test images. Then to each query image, we divide the relevant images in the first 50 returned images by the total relevant images in the database to obtain the recall rate, and we compute the ratio of the relevant images in the first 50 returned images to obtain the retrieval precision. Finally, after obtaining the recalls and precisions for each query image, we average them to get the average recall and precision. Furthermore, the average time for each query image is computed to obtain the total average retrieval time (ART) in our experiments.

A. *Experimental results on Dataset Ⅰ*

On taking the left-top image as the query image, two examples of the top 10 retrieved images with 12×12 division method on Dataset Ⅰ are ranked by the mutual similarities in Fig.4. It can be seen that the top 10 retrieved images are perceptually similar to the query image without any match errors. The similarities of the obtained relevant images in the first 50 returned images on each class are averagely over 0.83. The corresponding ARR, ARP and ART for each image class with the three division methods are given in Table 1. It can be inferred that the satisfactory retrieval performance can be obtained with our WDSCH features. Firstly, due to that the images in the Dinosaur and Elephant classes are mostly with clear backgrounds and simple spatial structural relations, the average retrieval precision on these two classes are entirely more than 84% with the three division methods. The total average retrieval precision is more than 78% with the three image division methods. Particularly, the best retrieval precision can reach 85.6% with the 12×12 division method. Secondly, although the more image blocks may increase the computational complexity a certain extent, the average retrieval time for each class can still be limited within 6.3s, which validates that our proposed WDSCH vector is highly compact and fairly suitable for practical retrieval. The best retrieval performance with the 12×12 division method further indicates that with the moderate number of image blocks, the extracted histogram vector can compress the redundant information and preserve the spatial structural features simultaneously.



(a)

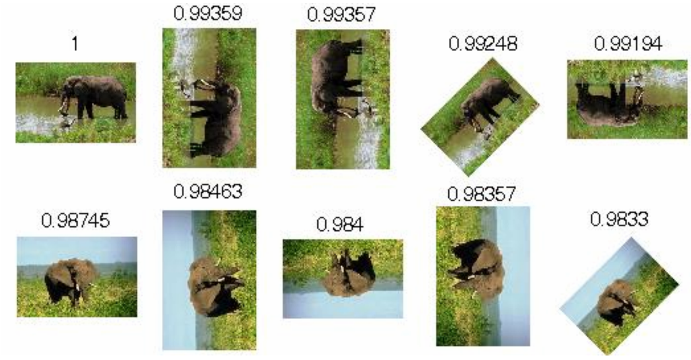
(b)

Fig.4 Two examples of top 10 retrieved images with 12×12 division method on Dataset Ⅰ

Table 1. Average retrieval results on Dataset Ⅰ (ARR, ARP: %; ART: Sec)

Class name	4×4			8×8			12×12		
	ARR	ARP	ART	ARR	ARP	ART	ARR	ARP	ART
People	41.0	82.0	3.2	42.0	84.0	4.8	44.0	88.0	6.1
Beach	33.0	66.0	3.0	35.0	70.0	4.6	38.0	76.0	6.0
Building	42.0	84.0	3.6	43.0	86.0	5.3	44.0	88.0	6.7
Bus	41.0	82.0	3.2	42.0	84.0	4.8	44.0	88.0	6.5
Dinosaur	43.0	86.0	3.0	45.0	90.0	4.2	46.0	92.0	6.2
Elephant	42.0	84.0	3.3	44.0	88.0	4.5	45.0	90.0	6.5
Flower	40.0	80.0	3.6	42.0	84.0	4.9	43.0	86.0	6.7
Horse	41.0	82.0	3.8	42.0	84.0	5.2	44.0	88.0	6.3
Mountain	36.0	72.0	3.4	38.0	76.0	4.3	40.0	80.0	6.1
Food	35.0	70.0	3.5	36.0	72.0	4.4	40.0	80.0	6.2
Total Average	39.4	78.8	3.4	41.0	82.0	4.8	42.8	85.6	6.3

B. *Experimental results on Dataset* *Ⅱ*



(a)

(b)

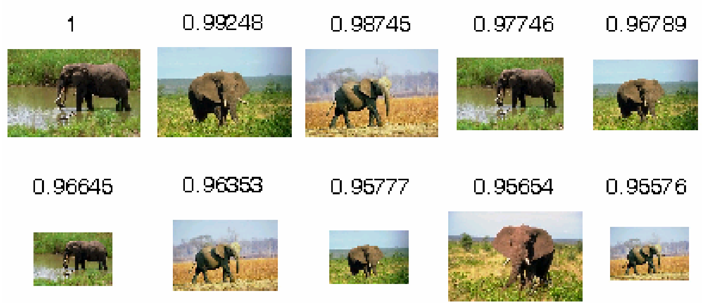
Fig.5 Two examples of top 10 retrieved images with 12×12 division method on Dataset Ⅱ

Table 2. Average retrieval results on Dataset Ⅱ (ARR, ARP: %; ART: Sec)

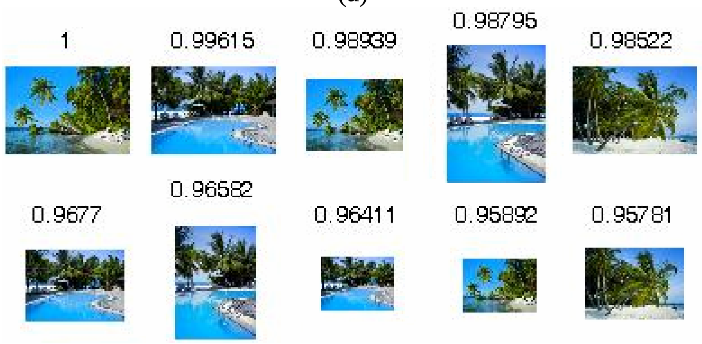
Class name	4×4			8×8			12×12		
	ARR	ARP	ART	ARR	ARP	ART	ARR	ARP	ART
People	8.6	86.0	14.3	9.0	90.0	17.5	9.2	92.0	19.2
Beach	7.6	76.0	14.5	8.0	80.0	17.6	8.4	84.0	19.5
Building	8.8	88.0	14.6	9.0	90.0	16.9	9.2	92.0	19.3
Bus	8.4	84.0	13.8	8.6	86.0	16.2	9.0	90.0	19.4
Dinosaur	9.0	90.0	14.7	9.2	92.0	15.6	9.6	96.0	18.2
Elephant	8.8	88.0	13.3	9.0	90.0	15.8	9.2	92.0	18.8
Flower	8.4	84.0	14.2	8.8	88.0	16.7	9.0	90.0	19.2
Horse	8.4	84.0	14.4	8.6	86.0	16.6	8.8	88.0	19.0
Mountain	7.8	78.0	14.9	8.2	82.0	16.8	8.4	84.0	19.6
Food	7.8	78.0	13.8	8.0	80.0	17.2	8.2	82.0	19.8
Total Average	8.4	83.6	14.3	8.6	86.4	16.7	8.9	89.0	19.2

Fig.5 gives two examples of the top 10 retrieved images with 12×12 division method on Dataset Ⅱ ranked by the mutual similarities with the left-top query image. The results show that rotated images with the angles of 45^0, 90^0, 180^0 and 270^0 are all top retrieved and have almost the same similarities with the original query image. The minute errors are mainly caused by the quantizing error during the process of histogram equalization and spatial information weights combination. In addition, the computed similarities of the relevant images in the first 50 returned images on each class are over 0.86 on average. Table 2 gives the ARR, ARP and ART for the each image class on Dataset Ⅱ with the three division methods. It can be seen that the ARP on Dataset Ⅱ outperforms that on Dataset Ⅰ with the change of 3.8% or so. The best total retrieval precision of 89.0% with the 12×12 division method validates that the proposed WDSCH features is quite robust to image rotation and translation, even for those images with complex and cluttered background.

C. *Experiments on Dataset Ⅲ*



(a)



(b)

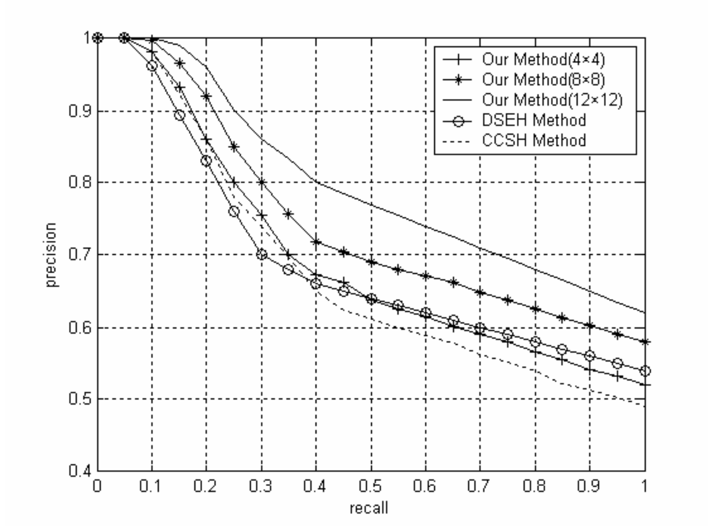
Fig.6 Two examples of top 10 retrieved images with 12×12 division method on DatasetⅢ

Table 3. Average retrieval results on Dataset Ⅲ (ARR, ARP: %; ART: Sec)

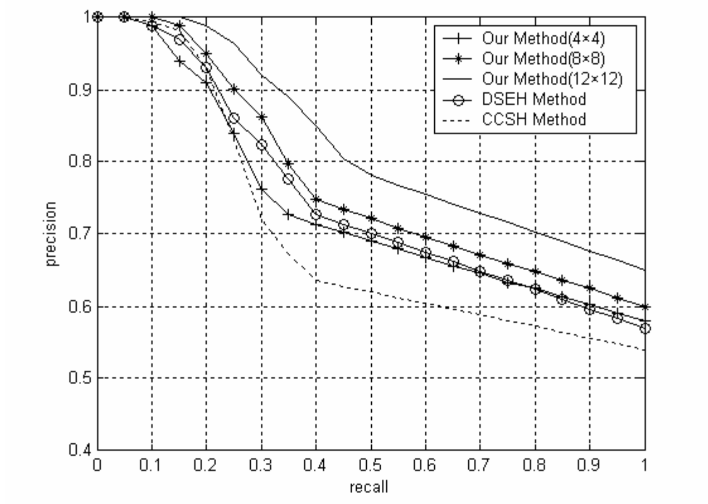
Class name	4×4			8×8			12×12		
	ARR	ARP	ART	ARR	ARP	ART	ARR	ARP	ART
People	10.3	82.0	12.3	10.5	84.0	12.8	10.8	86.0	13.6
Beach	9.3	72.0	12.5	9.8	78.0	13.6	10.0	80.0	13.5
Building	11.0	88.0	12.6	11.3	90.0	12.3	11.3	90.0	13.3
Bus	10.5	84.0	11.8	10.8	86.0	13.7	11.0	88.0	13.5
Dinosaur	10.8	86.0	11.7	11.5	92.0	12.3	11.8	94.0	13.0
Elephant	10.5	84.0	11.3	11.0	88.0	12.4	11.5	92.0	13.2
Flower	10.8	86.0	12.2	11.0	88.0	13.1	11.3	90.0	14.2
Horse	10.0	80.0	12.4	10.5	84.0	12.8	10.8	86.0	14.0
Mountain	10.0	80.0	12.4	10.3	82.0	12.6	10.5	84.0	14.0
Food	9.8	78.0	12.1	10.0	80.0	13.2	10.3	82.0	13.6
Total Average	10.3	82.0	12.1	10.7	85.2	12.9	10.9	87.2	13.6

Fig.6 gives two examples of the top 10 retrieved images with 12×12 division method on DatasetⅢ ranked by the mutual similarities with the left-top query image. It can be seen that the scaling transformation images can be retrieved and the similarities are more close to those of the corresponding query images. The minute errors are also caused by the quantizing error during the process of histogram equalization and spatial information weights combination. Furthermore, the computed similarities of the relevant images in the first 50 returned images on each class are over 0.85 averagely. Table 3 gives the ARR, ARP and ART for all of the image classes on DatasetⅢ with the three image division methods. It can be seen that the good retrieval performance can also be obtained on DatasetⅢ. The best retrieval precision of 87.2% with the 12×12 division method shows that the proposed WDSCH vector is invariant to image scaling transformation.

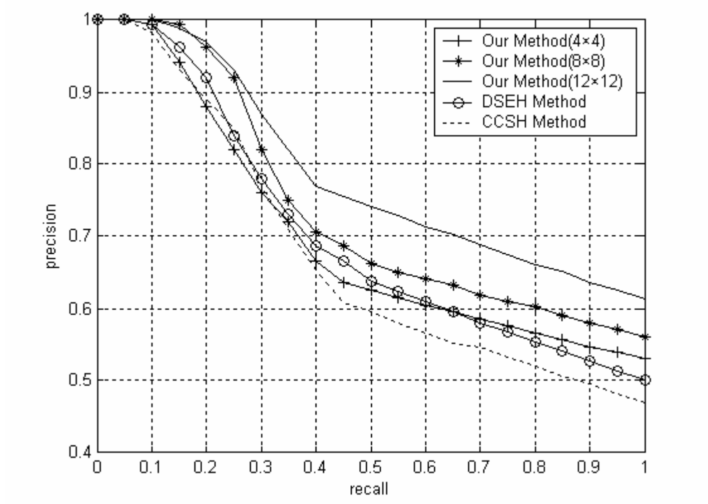
D. Contrastive results



(a) Retrieval PVR graph on Dataset Ⅰ



(b) Retrieval PVR graph on Dataset Ⅱ



(c) Retrieval PVR graph on Dataset Ⅲ

Fig.7 Contrastive retrieval results on thee image datasets

In order to test the efficiency of our proposed features, we compare the retrieval performance of our method with those of the Color Combination Shape Histogram (CCSH) method (Timothy and Huang, 2001) and the DCT Subband Energy Histogram (DSEH) method (Wu et al, 2002) on each image dataset. The graphs of average precision versus recall (PVR) are given in Fig.7. Table 4 gives the overall contrastive average retrieval results on the three image datasets. It is shown that our method can get much better retrieval performance than the other two methods. Firstly, the overall ARP of the 12×12 division WDSCH method can reach 87.3%, which outperforms those of the DSEH method and CCSH method with the change of 10.5% and 13% respectively. Secondly, the maximum ART of the 12×12 WDSCH division method only cost 13.2s, which validates that our method can obtain the fastest retrieval speed with the highly compact vectors. The contrast results further verify that the rotation, translation and scaling robustness of our presented WDSCH features.

Table 4. Overall contrastive average retrieval results (ARR, ARP: %; ART: Sec)

Overall Performance	CCSH	DSEH	WDSCH		
			4×4	8×8	12×12
ARR	17.9	18.2	19.4	20.1	20.9
ARP	74.3	76.8	81.2	84.5	87.3
ART	15.6	20.8	9.9	11.5	13.2

VI. Conclusion

This paper presents a color image retrieval approach based on Weighted DCT Spatial Combination Histogram (WDSCH) features. The advantages of the histogram features include three aspects. Firstly, the equalized block DCT histogram effectively describes the energy distribution of the image block with highly compact dimensionality and low dynamic variation. Secondly, the normalized Weighted DCT Spatial Combination histogram features include spatial structural relations and energy distribution and gray value variance of the image, which are proved to be more robust to image rotation, translation and scaling transformation. Thirdly, the spatial combination histogram with spatial information weights are highly compacted which make our method quite suitable for practical image retrieval application. In the future, we will continue researching the histogram-based image retrieval theory and applied system development.

Acknowledgements

This work is supported by the NUIST Science Foundation and the National Natural Science Foundation of China (No. 60472061). The authors also gratefully acknowledge the anonymous reviewers, Dr. Qing Wang and Marco Reisert for their useful comments and suggestions.

References

[1] P. Aigrain, H. J. Zhang, D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review", in *Multimedia Tools and Applications*, Vol. 3, No. 3, 1996, pp.179-202.

[2] R. Datta, J. Li, Z. W. James, "Content-based image retrieval-approaches and trends of the new age" in *MIR'05*, November, Singapore, 2005.

[3] M. J. Swain, D. H. Ballard, "Color indexing" in *International Journal of Computer Vision*, Vol. 7, No. 1, 1991, pp.11-32.

[4] D. K. Park, Y. S. Jeon, C. S. Won et al, "A composite histogram for Image Retrieval" in *IEEE International Conference on Multimedia and Expo, ICME 2000*, Vol.1, pp.355 -358.

[5] A. K. Jain, A. Vailaya, "Image retrieval using color and shape" in *Pattern Recognition*, Vol. 29, No. 8, 1996, pp.1233-1244.

[6] K. S. Timothy, J. Y. Huang et al, "An intelligent content-based image retrieval system based on color, shape and spatial relations" in *Proceedings of National Science Council ROC(A)*, Vol. 25, No.4, 2001, pp.232-243.

[7] S. Siggelkow, M. Schael and H. Burkhardt, "SIMBA-Search images by appearance" in B. Radig and S. Florczyk, editors, *Pattern Recognition*, *DAGM*, Munchen, 2001, pp.9-16.

[8] Jose A. Lay and L. Guan, "Image retrieval based on energy histograms of the low frequency DCT coefficients" in *International Conference on Acoustics, Speech and Signal Processing* Vol. 6, 1999, pp.3009–3012.

[9] D. S. Wu and L. N. Wu, "Image retrieval based on subband energy histograms of reordered DCT coefficients" in *International Conference on Signal Processing*, Vol. 1, 2002, pp. 596-599.

[10] Z. M. Lu and H. Burkhardt, "Colour image retrieval based on DCT-domain vector quantisation index histograms" in Electronics Letters, Vol. 41, No. 17, 2005, pp. 956-957.

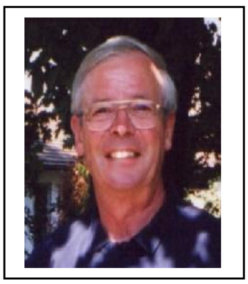
[11] L. G. Thorell and W. J. Smith, Using computer colour effectively, Prentice Hall, New Jersey, 1990.

[12] M. Koskela, J. Laaksonen and E. Oja, "PicSOM-self-organizing image retrieval with MPEG-7 content descriptors" in *IEEE Transactions on Neural Networks*, Vol. 13, No. 4, 2002, pp.841-853.

[13] F. Borko, P. Saksobhavivat, "A fast content-based multimedia retrieval technique using compressed data" in *SPIE 3527: Conference on Multimedia Storage and Archiving Systems III*, Boston, 1998, pp.561-571.

Yonghua Xie received his Master's degree from Nanjing University of Information Science and Technology, China in 2002. In 2006 he received his PhD in Pattern Recognition and Intelligent System from Nanjing University of Science and Technology, China. Now he is guest scholar in Computer Science Department of Freiburg University, Germany. His major research filed includes feature extraction and image retrieval, pattern recognition theory and application.

Lokesh Setia obtained his Bachelor's degree in Electrical Engineering from the Indian Institute of Technology in 1999. In 2001, he received a scholarship from the DAAD to pursue a master study at the University of Applied Science, Offenburg, Germany. Since 2003, he has been a Doctor candidate at the University of Freiburg, Germany. His research interests include image retrieval, relevance feedback methods.



Hans Burkhardt obtained his Dr. Ing. degree in 1974 from the University of Karlsruhe, Germany. Since 1997, he has been a full professor in the Computer Science Department, University of Freiburg, Germany. He is a Member of the Academy of Science and Humanities, Heildelberg and a Fellow of the International Association for Pattern Recognition. His research interests include invariants in pattern recognition, optimal image-restoration methods.