J. Vis. Commun. Image R. 33 (2015) 104–114

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

# Global Correlation Descriptor: A novel image representation for image retrieval ☆

Lin Feng [a,b], Jun Wu [a,b], Shenglan Liu [c], Hongwei Zhang [d,*]

[a] School of Innovation and Entrepreneurship, Dalian University of Technology, Dalian, Liaoning 116024, China
[b] School of Computer Science and Technology, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China
[c] School of Control Science and Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China
[d] School of Mathematical Sciences, Dalian University of Technology, Dalian, Liaoning 116024, China

ARTICLE INFO

*Article history:*
Received 23 March 2015
Accepted 7 September 2015
Available online 25 September 2015

*Keywords:*
Image retrieval
HSV color space
Texton detection
Global Correlation Descriptor (GCD)
Global Correlation Vector (GCV)
Global Correlation Vector (DGCV)
Feature fusion
Structure element correlation (SEC)

ABSTRACT

The image descriptors based on multi-features fusion have better performance than that based on simple feature in content-based image retrieval (CBIR). However, these methods still have some limitations: (1) the methods that define directly texture in color space put more emphasis on color than texture feature; (2) traditional descriptors based on histogram statistics disregard the spatial correlation between structure elements; (3) the descriptors based on structure element correlation (SEC) disregard the occurring probability of structure elements. To solve these problems, we propose a novel image descriptor, called Global Correlation Descriptor (GCD), to extract color and texture feature respectively so that these features have the same effect in CBIR. In addition, we propose Global Correlation Vector (GCV) and Directional Global Correlation Vector (DGCV) which can integrate the advantages of histogram statistics and SEC to characterize color and texture features respectively. Experimental results demonstrate that GCD is more robust and discriminative than other image descriptors in CBIR.



## 1. Introduction

In recent years, with the rapid development of the Internet and mobile terminals, processing image data effectively and efficiently is still an open issue in image research. Due to the complexity and diversity of images, it is difficult for users to retrieve desired images in the huge image datasets. Generally speaking, there are mainly three categories of image retrieval systems: text-based, content-based and semantic-based [1,2]. However, text-based and semantic-based image retrieval have some obvious limitations. Traditional text-based image retrieval systems need to annotate the images manually in advance. But it's time-consuming due to the amount of images, and people may perceive the same image differently. Besides, semantic-based image retrieval is still an open problem because of the limitations in studying about mechanisms of the primary visual cortex and artificial intelligence. So content-based image retrieval (CBIR) systems [3–6] have attracted more and more attentions in practical applications. These systems can generally be divided into three steps: first, the low-level features of the query image and all the images in dataset need to be extracted; second, choosing appropriate similarity measure, we can obtain the similarities between the query image and the images in dataset; third, sorting images by similarity, these systems return the top images that are most similar to the query image. Thus, the performance of the low-level feature has a great influence on CBIR systems.

Currently, most of CBIR systems are concentrated on extracting substantial features of an image, such as color, texture, shape, and fusion of two or more such features. So far, there have been many great research in extracting a single low-level feature. Due to the different description methods, these features can be classified into global and local features. The image representation based on histogram is one of the most common means of global features extraction, such as color histogram [7], Local Binary Pattern (LBP) [8] and Histogram of Oriented Gradient (HOG) [9], which see the whole image as visual information. However, these features tend to lose the spatial correlation among pixels. To overcome the problem, many researchers proposed their own visual models in succession with different ideas. Sticker and Orengo proposed the concept of color moments [10], which used the first three central moments called mean, standard deviation and skewness. Color correlogram [11] and color coherence vector (CCV) [12] also characterizes the color distributions of pixels and the spatial correlation between pair of colors. In addition, the gray co-occurrence matrix

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
*E-mail address:* hwzhang@dlut.edu.cn (H. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2015.09.002
1047-3203/

[13] and co-occurrence histograms of oriented gradients (CoHOG) [14] characterize the co-occurrence relationship between the values of two pixels. Different from Global feature-based algorithms, local feature-based algorithms focus mainly on key points and salient patches. In [15,16], scale-invariant feature transform (SIFT) was proposed to detect and describe local key points in scale space and has been widely applied to computer vision fields. Gabor wavelets [17], whose kernels are similar to the receptive field of the mammalian cortical simple cells, has been introduced to image analysis models. Based on the biological properties of Gabor wavelets, Liu and Wechsler [18] proposed the Gabor feature representation. And Serre et al. [19] introduced HMAX model based on the hierarchical visual processing in the primary visual cortex (V1). In this model, S1 units also take the form of Gabor filters in different scales and orientations. In [20], the comparison of some local descriptors for image processing can be seen.

Color, texture and shape are significant properties of an image, but simple feature usually has some limitations. To overcome the problem, some researchers proposed image representation based on multi-features fusion, which characters simultaneously two or more low-level features. These methods also can be classified two strategies. One strategy is to extract several features respectively and combine them into an integral vector [21–24], which can express more image information than simple one feature. In [21], Dubey et al. proposed the combination of four feature extraction methods namely color histogram, color moment, Texture and Edge Histogram Descriptor. Jalali et al. [23,24] studied the color processing in the high-level visual area of the primate brain and proposed the Color Quantization Hierarchical Max (CQ-HMAX) model, which use color quantization cores instead of Gabor filters to encode color information. And then combined with HMAX, the model can obtain color and shape information simultaneously. The other strategy is to extract texture and shape in color space directly. In [25], Liu and Yang introduced texton co-occurrence matrix (TCM) in RGB color space. Multi-texton histogram (MTH) [26] and color difference histograms (CDH) [27] integrates the advantages of co-occurrence matrix and histogram and encode color, texture and shape information. Micro-structure descriptor (MSD) [28] extracts micro-structures based on an edge orientation similarity and the underlying colors and structure element correlation statistics characterizes the spatial correlation of micro-structures. In [29], Wang and Wang introduced the concept of structure elements' descriptor (SED) in HSV color space which can describe color and texture information. And then they propose multi-factors correlation (MFC) namely structure element correlation (SEC) [30], gradient value correlation (GVC) and gradient direction correlation (GDC).

However, the strategy that extracts texture and shape in color space directly tends to enhance the characteristics of image color features. Thus, the color information may be dominant in these descriptors and lead to impair the performance of these descriptors in CBIR. In this paper, we propose a novel image representation, called Global Correlation Descriptor (GCD), which can extract color and texture feature of color image respectively.

The rest of this paper is organized as follow. In Section 2, the structure of our GCD model is introduced. In Sections 3 and 4, the color feature and texture feature of GCD model are presented respectively. The experimental results in CBIR are showed in Section 5. Section 6 is the conclusion of the paper.

## 2. Global Correlation Descriptor (GCD)

Color and texture of color image are important features in CBIR. Texture histogram which extracts some regular textons of gray image is a common approach, such as LBP and gray co-occurrence matrix which are introduced based on the different expression of texton. To combine with color information, some researchers defined texton of a color image in color space and extracted texture information based on this color space. However, though merging color and texture simultaneously, these descriptors can get worse performance in CBIR due to enhance the characteristics of image color features. As shown in Fig. 1, there are three flowers with similar texture and shape and different color significantly. The feature vector of these descriptors may be different significantly because the texton of the three flowers focus on their own color layer. To avoid this problem, we propose a new image feature model, called Global Correlation Descriptor (GCD). With this model, color and texture feature are extracted respectively and then are merged into an integral vector.

The structure of our GCD model is shown in Fig. 2. First, we construct image pyramid in color space to obtain color information of multiresolution images. To characterize the color distributions of pixels and the spatial correlation between pair of colors, we propose a new method, called Global Correlation Vector (GCV), to extract color feature in image pyramid. Second, in order to extract texture information unaffected by color, the color image is converted to grayscale. And then we choose Gabor filters in different scales and orientations to process the image. On this basis, we propose a new texture descriptor by defining several directional texton types and introduce Directional Global Correlation Vector (DGCV) to extract texture information. Finally, combining GCV and DGCV, we can get the Global Correlation Descriptor (GCD). The performance of color and texture feature may differ in CBIR in different image datasets, but they are of equal importance for most of the images.

As shown in Fig. 3, in the process of color feature extraction, we can choose appropriate scales for image pyramid. And then combining the GCV of each scale of the pyramid, we can get a more discriminative color feature vector, which is denoted $H_{color}$. Similarly, we choose multiple Gabor filters in different scales and obtain DGCV of each filtered image in the process of texture feature extraction. Then combining these vectors, we can get a more discriminative texture feature vector, which is denoted $H_{texture}$. Finally, GCD can be represented as:

$$H = [H_{color} \quad H_{texture}]^T \tag{1}$$

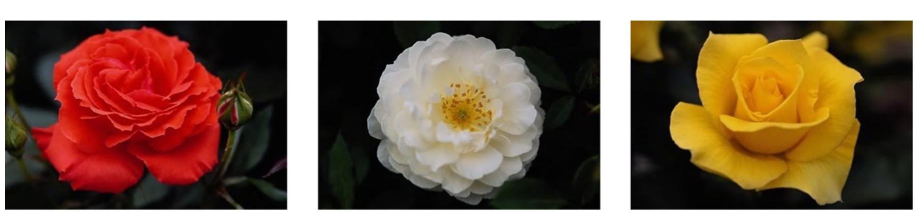


**Fig. 1.** Three flowers with different color and similar texture.

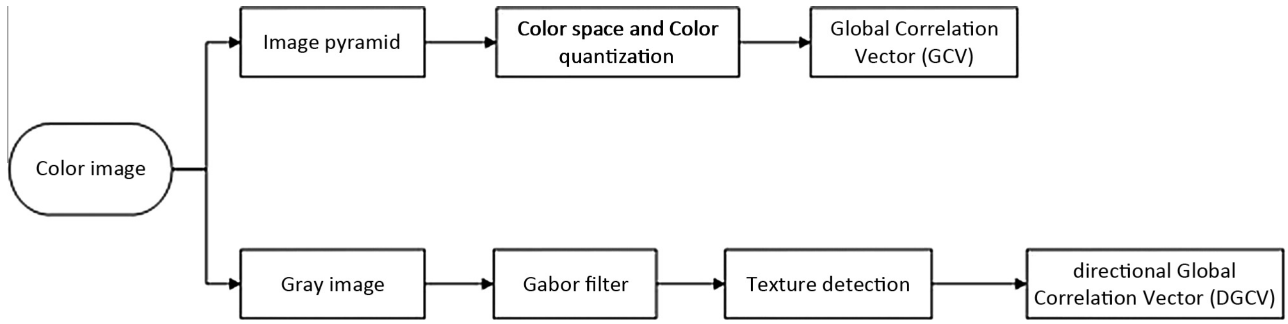



**Fig. 2.** The structure of our GCD model.

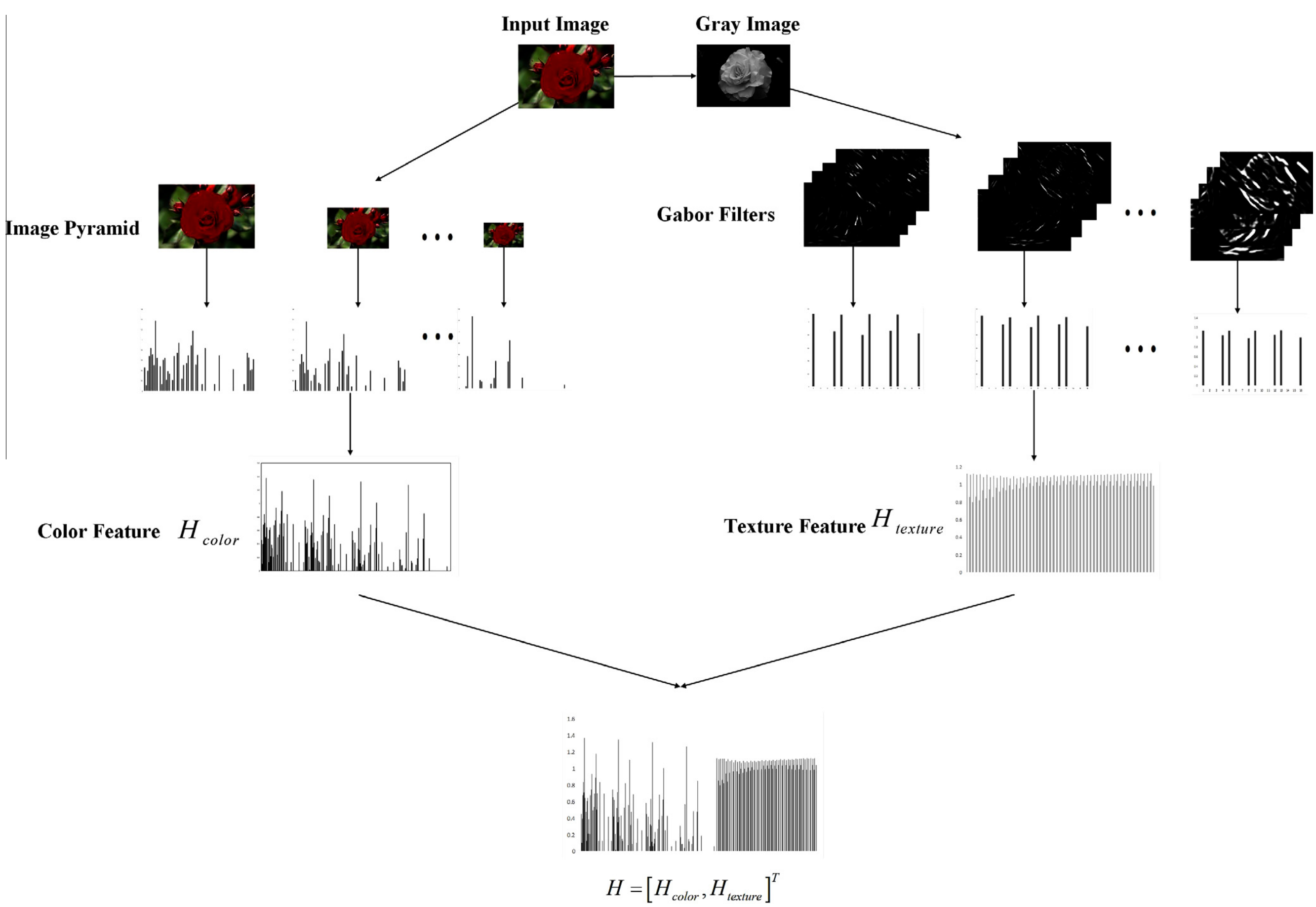



**Fig. 3.** The example of GDC combining GCV and DGCV.

In most images, color and texture information have equal discriminative performance in CBIR, and thus $H_{color}$ and $H_{texture}$ have the same weight in Eq. (1).

## 3. Color feature of Global Correlation Descriptor

### 3.1. Color space and color quantization

Color is an intuitive feature and plays an important role in CBIR systems. The HSV color space has been applied to extract color information because it is very close to the visual perception of human eye. In HSV color space [31], the color is defined in terms of three components: Hue (H), Saturation (S) and Value (V). However, it is time-consuming and unnecessary to represent all types of color directly. So in order to reduce the computation, the HSV color space is quantized to 72 bins as the method introduced in [32]. And H, S and V components are divided into 8, 3 and 3 bins respectively.

To obtain color information of multiresolution images, an image pyramid for color image is constructed [15,16,33]. The common method is to smooth the images using Gaussian filters and then reduce image size. So the images is produced from the convolution of a variable-scale Gaussian function. Considering that the image quality is very sensitive to noise and some details of image lose in the process of reduction, the smoothing preprocessing is necessary. In this model, the pyramid has $m$ scales with each neighboring scale different by a ratio of 1/2. And the subimages shouldn't be too small because they lose too much color information to be distinguished.

### 3.2. Global Correlation Vector (GCV)

Color histogram is one of the most common methods to extract color feature of an image and has great performance when applied to image recognition. The essence of this method is to calculate the frequency of occurrence of each color. It is a direct and effective method to describe the global features of the image. However, this method has some limitations due to not considering the spatial correlation between pair of colors.

In order to characterize the spatial correlation, a new method, called structure element correlation (SEC) has been proposed in [28,30]. For an image $f(x,y)$, SEC can be used to describe the structure element $w_0$ of the image where $w_0$ denotes the low-level feature of pixels, such as color value, texton and gradient direction of the pixels. In each block of the image, the center position is denoted by $P_0 = (x_0, y_0)$ and the neighbors of $P_0$ are denoted by $P_i = (x_i, y_i)$. The SEC is defined as the probability of the two neighboring structure elements and can be written as follows:

$$H(w_0) = \begin{cases} Pr(f(P_0) = w_0 \wedge f(P_i) = w_0) \\ where \ \max(|x_0 - x_i|, |y_0 - y_i|) = D \end{cases} \tag{2}$$

where $Pr(\cdot)$ is the probability when the two neighboring structure elements is $w_0$ and the distance is not greater than $D$, namely the ratio of the number of $P_i$ whose structure element is same to $P_0$ to the amount of neighbors of $P_0$. In practice, the distance $D$ between the two neighboring structure elements is taken as 1 and thus Eq. (2) can be simplified:

$$H(w_0) = \frac{N\{f(P_0) = w_0 \wedge f(P_i) = w_0\}}{8\overline{N}\{f(P_0) = w_0\}} \tag{3}$$

where $N$ denotes the co-occurring number of the structure element $w_0$ and $\overline{N}$ denotes the occurring number of $w_0$. However, this method also has some drawbacks. As shown in Fig. 4, the distribution of the black structure element in (a) and (b) are significantly different but have the same SEC which are equal to 0.125. This method disregards the occurring probability of structure elements though it characterizes the spatial correlation.

To solve these problems, our method (GCV) is propose to extract color feature in image pyramid. The distributions of colors and the spatial correlation between pair of colors can be extracted simultaneously. So our GCV can be written as:

$$CF(c_0) = \frac{N\{f(P_0) = c_0 \wedge f(P_i) = c_0\}}{8\overline{N}\{f(P_0) = c_0\}} \cdot \left(\frac{\overline{N}\{f(P_0) = c_0\}}{\overline{N}_{sum}} + 1\right) \tag{4}$$

where $N$ denotes the co-occurring number of the color value $c_0$ and $\overline{N}$ denotes the occurring number of $c_0$. $\overline{N}_{sum}$ denotes the total number of color value of image. In this paper, HSV color space of image pyramid has been quantized to 72 main colors, so the image $f(x,y)$ can be denoted as $f(x,y) = c_0, c_0 \in \{0, 1, \ldots, 71\}$. We can see that the first term in Eq. (4) is SEC and the second term can be seen as color histogram. GCV integrates the advantages of color histogram and SEC, and characterize not only the distributions of colors, but also the spatial correlation of pixels.

With GCV, the different images shown in Fig. 4 can be distinguished. Actually, the Eq. (4) also can be rewritten as:

$$CF(c_0) = \frac{N\{f(P_0) = c_0 \wedge f(P_i) = c_0\}}{8\overline{N}\{f(P_0) = c_0\}} + \frac{N\{f(P_0) = c_0 \wedge f(P_i) = c_0\}}{8\overline{N}_{sum}} \tag{5}$$

The first term in Eq. (5) is still SEC and the second term can be seen as the co-occurring number of the color because $\overline{N}_{sum}$ is constant for images of the same size. With the second term, the two images in Fig. 4 can be distinguished.

So GCV is presented to characterize the color feature of each image in the image pyramid in this paper. And then combining these vectors, we can get the color feature descriptor:

$$H_{color} = [CF_1(c_0)\ CF_1(c_1)\ \cdots\ CF_1(c_{71})\ CF_m(c_0)\ CF_m(c_1)\ \cdots\ CF_m(c_{71})]^T \tag{6}$$

where $m$ denotes the number of scales in the image pyramid.

## 4. Texture feature of Global Correlation Descriptor

### 4.1. Gabor filter for gray image

To rule out the possibility that extracted texture information is dominated by color feature of the images, the color image is converted to gray image. Due to the similarity with the receptive field of the mammalian cortical simple cells [17], Gabor wavelets have been widely used in image analysis. We choose Gabor filters in different scales and orientations to process the image to get the

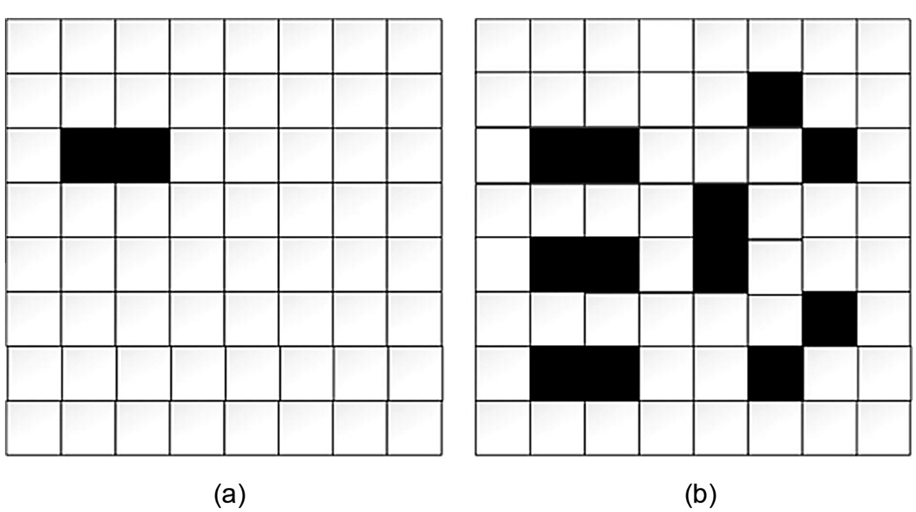



**Fig. 4.** The example of two images which have identical SEC.

filtered image in different orientations. The Gabor filters can be defined as follows:

$$F(x,y) = \exp\left(-\frac{x_0^2+\gamma^2 y_0^2}{2\sigma^2}\right)\times\cos\left(\frac{2\pi x_0}{\lambda}\right) \quad (7)$$
$$s.t.\quad x_0 = x\cos\theta + y\sin\theta,\quad y_0 = -x\sin\theta + y\cos\theta$$

In this paper, all filter parameters can be set as S1 units of HMAX model. We also choose four orientations ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) and a pyramid of scales, spanning a range sizes from $7\times 7$ to $37\times 37$. In addition, the aspect ratio is set to 0.3. The effective width and the wavelength are shown in Table 1. With these filtered images in four orientations, we can define corresponding textons to extract texture information.

### 4.2. Texture detection

Texture is a significant characteristic of an image. Due to the different understandings of texture, many descriptors have been proposed to analyze texture information. In general, texture are seen as some regular "texton", but defining texton remains a challenge. In this paper, in order to extract texture features of filtered images in four orientations, we propose a new method to represent textons.

We define four directional texton types for filtered image in four different orientations by assuming that the textons of central pixel depend on two neighbors in the corresponding direction. With this idea, gray-levels of neighbors in a zone of width around the gray-level of central pixel are quantized to 1, and ones above or below this are quantized to 0. So given a central pixel in a fixed pattern, the texton can be represented as follows:

$$D_\theta(g_0) = \sum_{i=1}^{P} d_\theta(g_0, g_i)\times 2^i \quad (8)$$

$$d_\theta(g_0, g_i) = \begin{cases} 1, & |g_i - g_0| \leqslant t \\ 0, & |g_i - g_0| > t \end{cases} \quad (9)$$

where $g_0$ denotes the gray-level of central pixel and $g_i$ denotes the gray-level of neighbors in the orientations $\theta$ ($\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$). $P$ denotes the number of neighbors of central pixel in the orientations $\theta$. The central pixels of filtered images have different neighbors because of the different directions of corresponding Gabor filters. In our paper, we choose the neighbors of the central pixels in the $3\times 3$ pattern (in general, the edge pixels have little discriminative information, so their textons can be ignored). Based on this method, there are only four texton types (00, 01, 10, and 11) in a filtered image.

As shown in Fig. 5, considering four Gabor filters in different orientations (highlighted numbers mean the four directions in the local image regions), it can be seen that the four filtered images contain significant directional information. As the filtered orientation $\theta$ is equal to $0^\circ$, we can define textons of the central pixels by comparing its gray-level with its two neighbors in the horizontal direction. For example, when the gray-level of the central pixels is 62, the gray-level of its two neighbors in the horizontal direction are in a zone of width $\pm t$ ($t$ is set as 8 in our experiments). Thus, its texton can be marked as 11 whose corresponding binary value is 3. In addition, there are a large number of pixels whose gray-level are 0 in the filtered images. In the process of extracting texture feature, these pixels need not be considered (they are set as $-1$ in Fig. 5) because they have little discriminative texture information.

### 4.3. Directional Global Correlation Vector (DGCV)

As mentioned above, Global Correlation Vector (GCV) can characterize both the distributions of colors and the spatial correlation between pair of pixels. And this method integrates the advantages of color histogram and structure element correlation (SEC). However, some directional information of texture images can be lost if their texture feature are represented by GCV directly. So in this paper, the method called Directional Global Correlation Vector (DGCV) is proposed to extract texture feature of filtered images. With this method, only the textons of two neighbors in the filtered orientation are considered to compare with the central pixels. For the filtered image, DGCV can be written as:

$$TF_\theta(t_0) = \frac{N_\theta\{f(P_0)=t_0 \wedge f(P_i)=t_0\}}{2\overline{N}\{f(P_0)=t_0\}}\cdot\left(\frac{\overline{N}\{f(P_0)=t_0\}}{\overline{N}_{sum}}+1\right) \quad (10)$$

where $N_\theta$ denotes the co-occurring number of the texton $t_0(t_0 \in \{0,1,2,3\})$ and $P_i$ are the neighbors of $P_0$ in the orientations $\theta$. Similarly, the Eq. (10) also can be rewritten as:

$$TF_\theta(t_0) = \frac{N_\theta\{f(P_0)=t_0 \wedge f(P_i)=t_0\}}{2\overline{N}\{f(P_0)=t_0\}} + \frac{N_\theta\{f(P_0)=t_0 \wedge f(P_i)=t_0\}}{2\overline{N}_{sum}} \quad (11)$$

The second term can be seen as the co-occurring number of the texton because $\overline{N}_{sum}$ is constant. With this term, DGCV has better discriminative performance than SEC and histogram statistics in CBIR. For four filtered images of the same scale, the feature vector is

$$TF_{n\times n} = [TF_{0^\circ}(0)\ TF_{0^\circ}(1)\ TF_{0^\circ}(2)\ TF_{0^\circ}(3)\ \cdots TF_{135^\circ}(0)\ TF_{135^\circ}(1)\ TF_{135^\circ}(2)\ TF_{135^\circ}(3)]^T \quad (12)$$

where $n\times n$ denotes the size of Gabor filter. In this paper, we choose Gabor filters of 16 scales to process the gray image. So combining these features of filtered images, we can get the texture feature vector of the image:

$$H_{texture} = [TF_{7\times 7}\ TF_{9\times 9}\ \cdots\ TF_{37\times 37}]^T \quad (13)$$

Particularly, we have defined four texton types (00, 01, 10, and 11) for filtered images. But considering that it is impossible for them to be adjacent, $TF_\theta$ of texton types 10 and 01 are nonexistent. As a result, we only extract the features of texton types 00 and 11 where the texton type 11 means that the gray-level of the central pixels changes slowly and the texton type 00 means that its gray-level changes greatly.

**Table 1**
The parameters setting of Gabor filters.

| Filter size | $7\times 7$ | $9\times 9$ | $11\times 11$ | $13\times 13$ | $15\times 15$ | $17\times 17$ | $19\times 19$ | $21\times 21$ |
|---|---|---|---|---|---|---|---|---|
| $\theta$ | 2.8 | 3.6 | 4.5 | 5.4 | 6.3 | 7.3 | 8.2 | 9.2 |
| $\lambda$ | 3.5 | 4.6 | 5.6 | 6.8 | 7.9 | 9.1 | 10.3 | 11.5 |
| | $23\times 23$ | $25\times 25$ | $27\times 27$ | $29\times 29$ | $31\times 31$ | $33\times 33$ | $35\times 35$ | $37\times 37$ |
| $\theta$ | 10.2 | 11.3 | 12.3 | 13.4 | 14.6 | 15.8 | 17.0 | 18.2 |
| $\lambda$ | 12.7 | 14.1 | 15.4 | 16.8 | 18.2 | 19.7 | 21.2 | 22.8 |

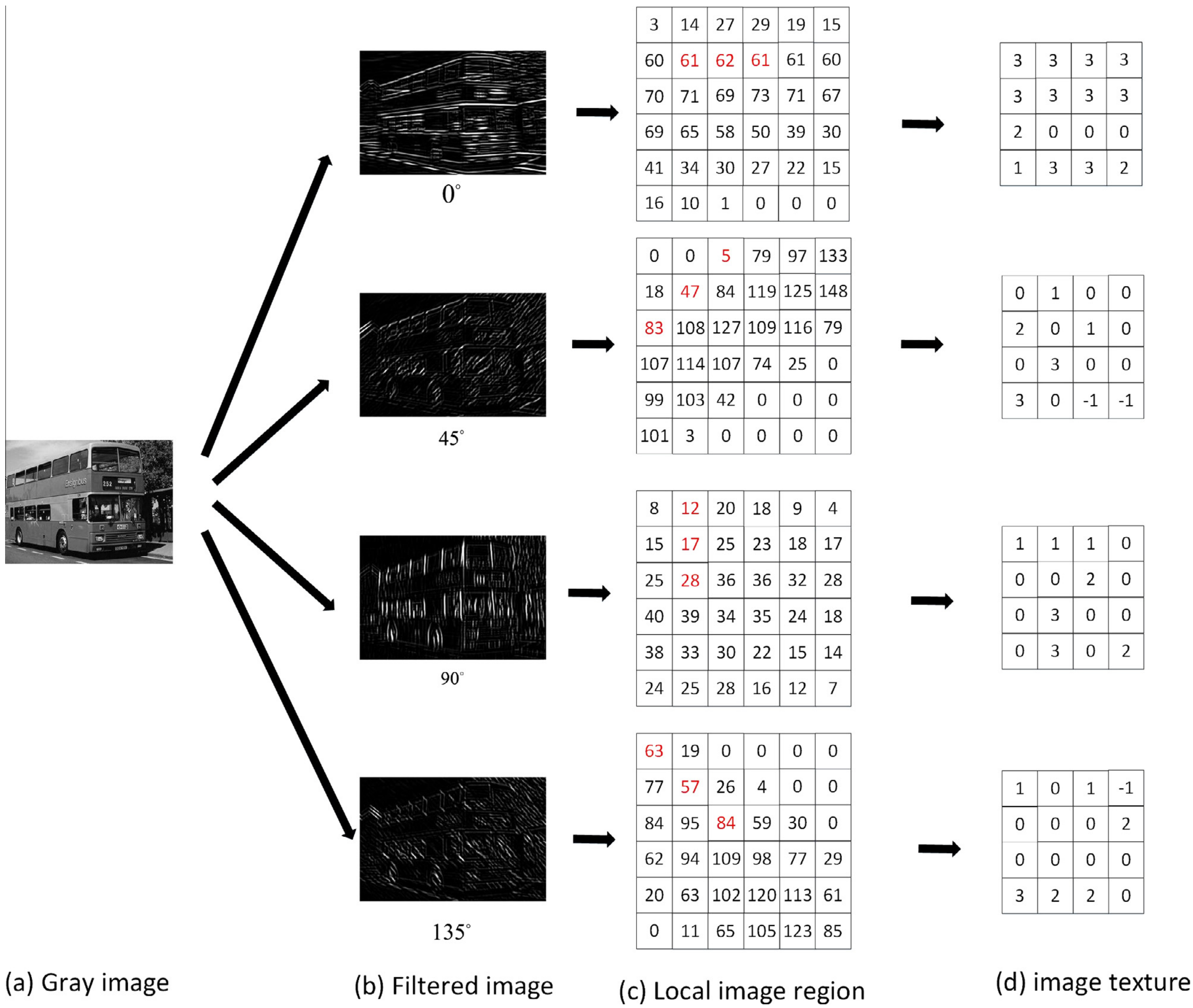



**Fig. 5.** The extraction process of textons for four filtered images.

## 5. Experimental results

In this section, we use three Corel datasets to evaluate the performance of Global Correlation Descriptor (GCD). All Corel images are obtained from Corel Gallery Magic 200,000 (8 CDs) which is widely used in image retrieval field. Three common Corel datasets are used in our experiments (As shown in Fig. 6). The first one is Corel-1000 dataset (Fig. 6(a)), which is divided into 10 categories including landscapes, horses, elephants, human beings, bus, flowers, buildings, mountains, food and dragons. The second one is Corel-5 K dataset (Fig. 6(b)), which contains 50 categories including diverse content. Another one is Corel-10 K dataset (Fig. 6(c)) which contains 100 categories. For the three datasets, every category contains 100 images in JPG or JPEG format.

In these experiment, we randomly choose 10 images from each category in the three datasets, which means that there are 100, 500 and 1000 query images in the three datasets respectively. Then the mean precision-recall pair percentage of these query images is computed. The precision $P$ and recall $R$ can be defined as the following equations:

$$P = I_N/N \tag{14}$$

$$R = I_N/M \tag{15}$$

where $I_N$ denotes the number of the retrieved relevant images and $N$ denotes the total number of the retrieved images. The total number of the relevant images $M$ is a constant ($M = 100$ in our CBIR system).

### 5.1. Implementation details

In the experiments, different color spaces and quantization strategies of color are applied to evaluate the performance of Global Correlation Descriptor (GCD). And the experimental results demonstrate the reason why we choose HSV color space and non-uniform quantization strategy for the proposed framework.

The average retrieval precision and recall of GCD in HSV, RGB and Lab color spaces are listed in Table 2. The total number of the retrieved images are set from 10 to 30 in these experiments. It can be seen from those data that the GCD has best performance in HSV color space and the average retrieval precision ranges from 76.27% to 79.58% for Corel-1000 dataset when the total number of the retrieved images is 10. However, the precisions are not more than 78% when uniform quantization strategies are adopted. In addition, the precisions of GCD in RGB and Lab color space are at most 62.73% and 63.95%, respectively. In the RGB color space, the total color uniform quantization number is increased from 16

(a) Corel-1000 dataset

(b) Corel-5K dataset


(c) Corel-10K dataset

**Fig. 6.** Some images of three Corel datasets.

**Table 2**
Retrieval results of GCD in different color spaces for Corel-1000 dataset.

| Color space | Color quantization strategies | Precision (%) | | | | | Recall (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 10 | 15 | 20 | 25 | 30 | 10 | 15 | 20 | 25 | 30 |
| HSV | 72 (non-uniform) | **79.58** | **77.24** | **74.63** | **73.35** | **71.23** | **7.96** | **11.68** | **14.91** | **18.34** | **21.47** |
| | 72 | 76.85 | 75.14 | 72.84 | 71.46 | 69.78 | 7.69 | 11.32 | 14.56 | 17.87 | 20.90 |
| | 108 | 77.43 | 74.47 | 73.18 | 71.55 | 70.17 | 7.74 | 11.25 | 14.62 | 17.96 | 21.04 |
| | 128 | 77.54 | 73.93 | 71.28 | 69.65 | 68.02 | 7.78 | 11.18 | 14.30 | 17.31 | 20.43 |
| | 192 | 76.27 | 73.12 | 71.49 | 69.78 | 68.31 | 7.63 | 11.06 | 14.38 | 17.44 | 20.56 |
| RGB | 16 | 61.68 | 57.94 | 54.87 | 53.22 | 51.57 | 6.17 | 8.72 | 10.96 | 13.28 | 15.46 |
| | 32 | 62.73 | 59.14 | 55.38 | 52.76 | 51.23 | 6.27 | 8.86 | 11.05 | 13.22 | 15.41 |
| | 64 | 61.68 | 57.94 | 54.87 | 53.22 | 51.57 | 6.17 | 8.72 | 10.96 | 13.28 | 15.46 |
| | 128 | 62.73 | 59.14 | 55.38 | 52.76 | 51.23 | 6.27 | 8.86 | 11.05 | 13.22 | 15.41 |
| Lab | 45 | 60.73 | 56.06 | 52.78 | 50.20 | 48.43 | 6.08 | 8.37 | 10.61 | 12.59 | 14.55 |
| | 90 | 62.56 | 57.91 | 54.97 | 53.15 | 51.30 | 6.26 | 8.73 | 10.95 | 13.34 | 15.36 |
| | 180 | 63.95 | 60.58 | 57.74 | 55.01 | 52.43 | 6.40 | 9.14 | 11.55 | 13.68 | 15.72 |
| | 225 | 67.08 | 62.13 | 58.14 | 56.07 | 53.81 | 6.71 | 9.26 | 11.64 | 14.02 | 16.24 |

The best retrieval results are shown in blod, which means that GCD has the best performance on this condition.

**Table 3**
Retrieval results of GCD with different similarity/distance measures for Corel-1000 dataset.

| Similarity measure | Precision (%) | | | | | Recall (%) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | 15 | 20 | 25 | 30 | 10 | 15 | 20 | 25 | 30 |
| $L1$ | **79.58** | **77.24** | **74.63** | **73.35** | **71.23** | **7.96** | **11.68** | **14.91** | **18.34** | **21.47** |
| Euclidean | 77.87 | 74.92 | 73.95 | 69.37 | 67.31 | 7.79 | 11.20 | 14.42 | 17.35 | 20.17 |
| Canberra | 74.95 | 71.86 | 69.12 | 66.75 | 64.94 | 7.50 | 10.78 | 13.84 | 16.74 | 19.49 |
| Weighted $L1$ | 78.72 | 75.95 | 73.82 | 71.61 | 69.75 | 7.87 | 11.67 | 14.76 | 17.94 | 20.86 |
| $\chi^2$ statistics | 76.95 | 73.54 | 71.23 | 68.55 | 66.76 | 7.70 | 11.02 | 14.21 | 17.17 | 19.98 |

The best retrieval results are shown in blod, which means that GCD has the best performance on this condition.

**Table 4**
Retrieval results with different methods for each category of Corel-1000 dataset.

| Category | Performance (%) | MTH | MSD | CDH | SED | Our method |
|---|---|---|---|---|---|---|
| African | Precision | 69.17 | 83.33 | 77.50 | 82.50 | **87.50** |
| | Recall | 8.30 | 10.00 | 9.30 | 9.90 | **10.50** |
| Beach | Precision | 61.67 | 43.33 | 56.67 | 28.33 | **68.33** |
| | Recall | 7.40 | 5.20 | 6.80 | 3.40 | **8.20** |
| Building | Precision | 45.83 | 63.33 | 47.50 | 47.50 | **61.67** |
| | Recall | 5.50 | 7.60 | 5.70 | 5.70 | **7.40** |
| Bus | Precision | 68.33 | 76.67 | 71.67 | 73.33 | **80.00** |
| | Recall | 8.20 | 9.20 | 8.60 | 8.80 | **9.60** |
| Dinosaur | Precision | 100.00 | 100.00 | 100.00 | 90.00 | **100.00** |
| | Recall | 12.00 | 12.00 | 12.00 | 10.80 | **12.00** |
| Elephant | Precision | **70.83** | 65.00 | 62.50 | 55.00 | 67.50 |
| | Recall | **8.50** | 7.80 | 7.50 | 6.60 | 8.10 |
| Flower | Precision | 75.00 | 86.67 | 60.83 | 72.50 | **88.33** |
| | Recall | 9.00 | 10.40 | 7.30 | 8.70 | **10.60** |
| Horse | Precision | 100.00 | 97.50 | 91.67 | 62.50 | **100.00** |
| | Recall | 12.00 | 11.70 | 11.00 | 7.50 | **12.00** |
| Mountain | Precision | 39.17 | 29.17 | 44.17 | 40.00 | **55.00** |
| | Recall | 4.70 | 3.50 | 5.30 | 4.80 | **6.60** |
| Food | Precision | 52.50 | **76.67** | 45.00 | 64.17 | 74.17 |
| | Recall | 6.30 | **9.20** | 5.40 | 7.70 | 8.90 |

The best retrieval results are shown in blod, which means that GCD has the best performance on this condition.

$(4 \times 2 \times 2)$ to 128 $(8 \times 4 \times 4)$ bins. And in Lab color space, the total color quantization number is increased from 72 $(8 \times 3 \times 3)$ to 192 $(12 \times 4 \times 4)$ bins. It can be seen that the performance of GCD can be better when the quantization number of color is increased. But GCD may obtain too much noisy features to reduce the performance if the quantization number of color is too large. Thus, we choose HSV color space and non-uniform quantization strategy as introduced in [32] to characterize color feature in this paper.

### 5.2. Similarity measure

In CBIR systems, retrieval accuracy and recall depends on both the performance of feature descriptor and the choice of similarity/distance measures. So it is also a key step to select an appropriate similarity/distance measure for GCD in image retrieval. In the experiments, we compare several common similarity/distance measures, such as the Euclidean distance, the $L1$ distance, $\chi^2$ statis-

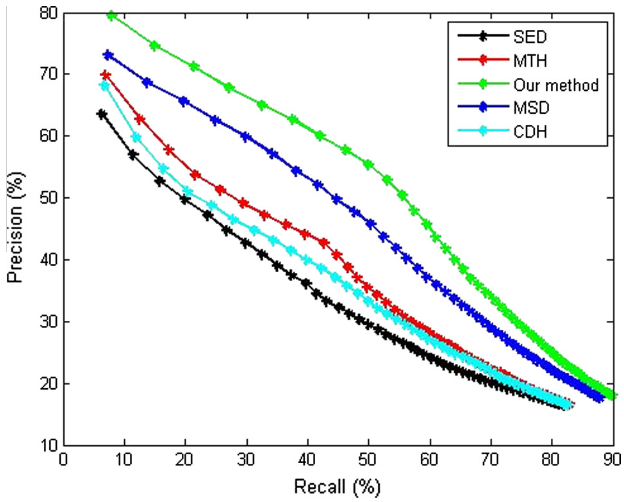



**Fig. 7.** The average retrieval performance comparison for Corel-1000 dataset.

tics, the Canberra distance and the weighted $L1$ distance as mentioned in [26–28].

For two feature vectors $X = (x_1, x_2, \ldots, x_n)^T$ and $Y = (y_1, y_2, \ldots, y_n)^T$ extracted from images, their similarity/distance measures can be represented as:

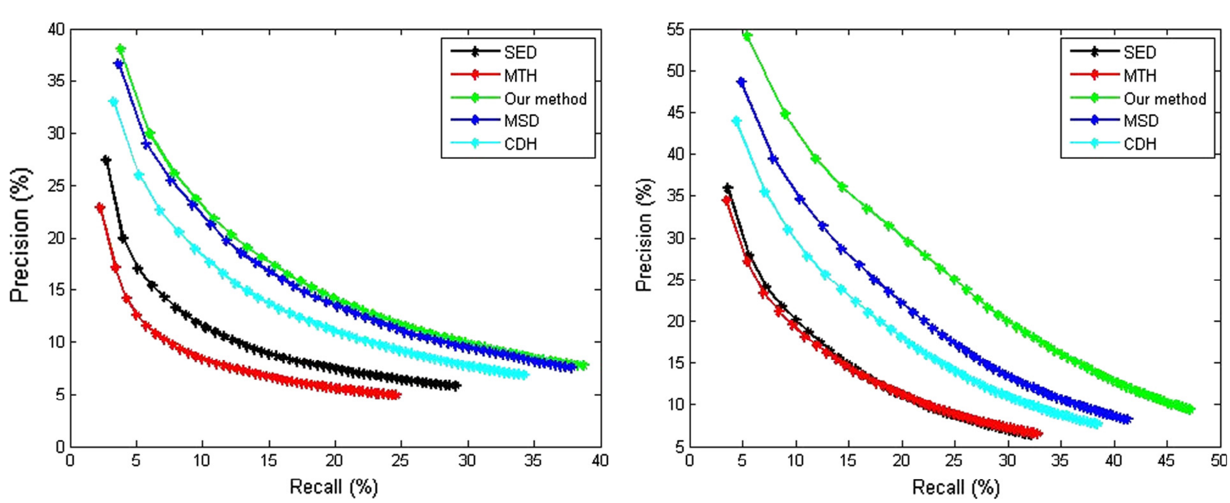
- $L1$ distance: $D(x, y) = \sum_i |x_i - y_i|$.
- Euclidean distance: $D(x, y) = \sum_i (x_i - y_i)^2$.
- $\chi^2$ statistics: $D(x, y) = \sum_i \frac{(x_i - y_i)^2}{x_i + y_i}$.
- Canberra distance: $D(x, y) = \sum_i \frac{|x_i - y_i|}{x_i + y_i}$.
- Weighted $L1$ distance: $D(x, y) = \sum_i \frac{|x_i - y_i|}{1 + x_i + y_i}$.

The average retrieval precision and recall of GCD with different similarity/distance measures are listed in Table 3. The total number of the retrieved images are also set from 10 to 30 in the experiments. It can be seen that the $L1$ distance performs better than other similarity/distance measures and it is much more computationally efficient. Euclidean distance is one of the most commonly used similarity/distance measures, but not always the best one because the distances put too much emphasis on features that are greatly dissimilar. Both Canberra distance and weighted $L1$ distance can be considered as a weighted $L1$ distance with different weights. In addition, $\chi^2$ statistics is also a common similarity/distance measure. However, they are not suitable for our retrieval system although they have good performance. So we choose $L1$ distance as similarity/distance measure for our CBIR system in this paper.

### *5.3. Retrieval performance*

As mentioned above, multi-feature fusion based feature extraction methods are classified into two categories: extracting several features respectively and combine them into an integral vector and extracting texture and shape in color space directly. Motivated by the idea in 1st category, GCD is proposed to improve the performance of descriptors in 2nd category by using the 1st feature extracting category and the novel statistical methods. Based on structure element correlation (SEC) which is proposed in MSD and MFC, GCV and DGCV are introduced as statistical methods to characterize image features. And color and texture features are represented respectively with these statistical methods. Thus GCD outperforms other methods in 2nd category because of not only the 1st category but also the novel statistical methods in GCD. Moreover, though adopting the first category by combining several proposed features, these descriptors still have no logically sound explanation about how to choose and combine proposed features. So through comparing with the methods in 2nd category, this paper not only disproves these methods but demonstrates that the first category is feasible to avoid the problem of the 2nd category.

So in this subsection, we compare GCD with four image feature descriptors which belong to the 2nd category, namely MTH in [26], MSD in [28], CDH in [27] and SED in [29]. In order to illustrate the robust performance of GCD for most image categories, we test GCD and the other four descriptors for each category of Corel-1000 dataset. We set the total number of the retrieved images N to 12 in this experiment. As shown in Table 4, we can see that GCD performs better significantly than other descriptors in most categories. The precision and recall of MTH is higher than that of GCD (about 3.33% precision and 0.40% recall) in elephant class. And the precision and recall of MSD is higher than that of GCD (about 2.50% precision and 0.30% recall) in food class. Therefore, comparing with these four descriptors, the image retrieval system based on GCD is more robust.




(a) Corel-5K dataset (b) Corel-10K dataset

**Fig. 8.** The average retrieval performance comparison for (a) Corel-5 K dataset; (b) Corel-10 K dataset.

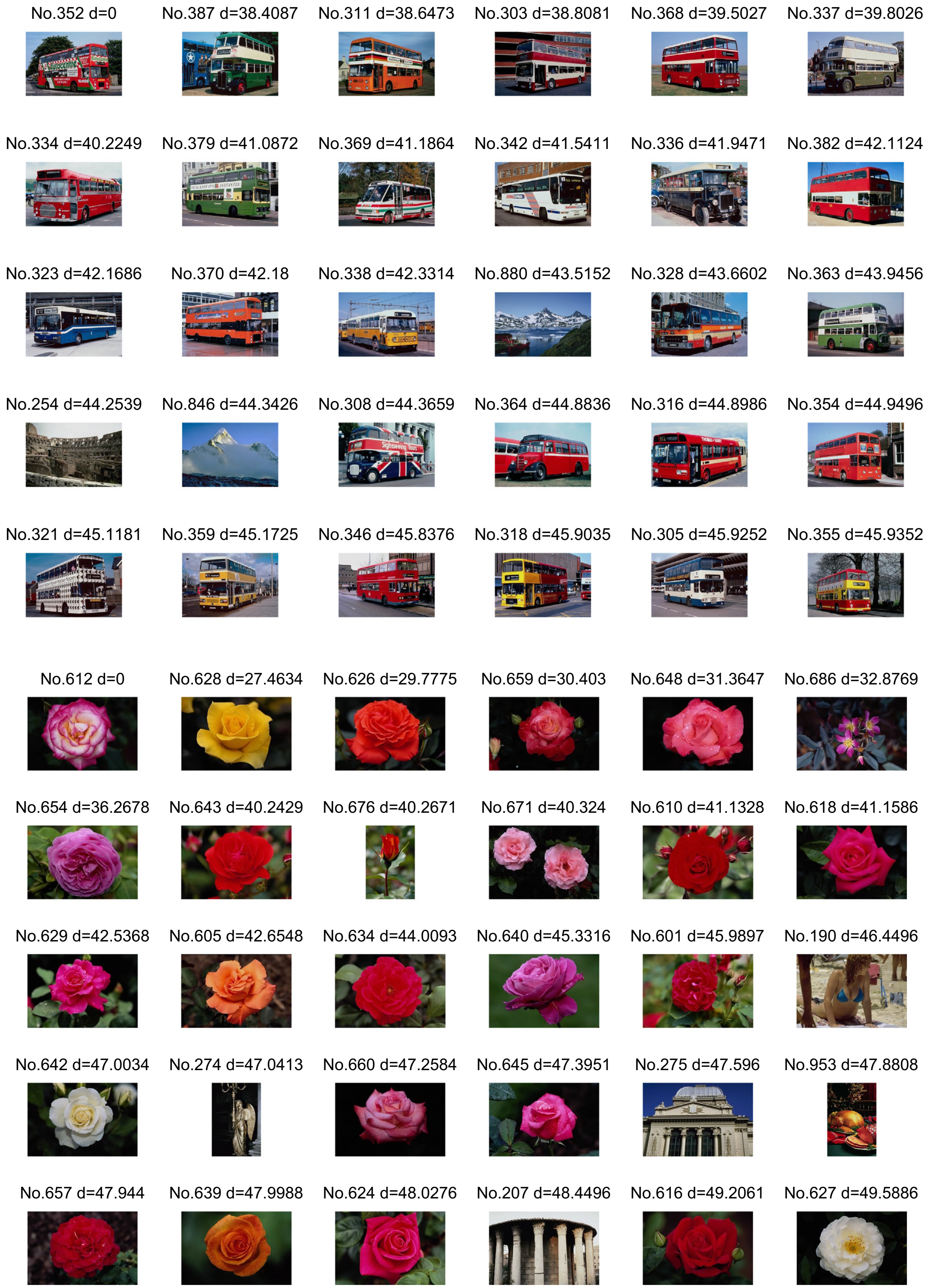



**Fig. 9.** The retrieval results of some images (Nos. 352 and 612 are the query images).

In addition, the overall average retrieval precision and recall of GCD with these descriptors for Corel-1000 dataset are illustrated as Fig. 7. It can be seen from the results that the average precision of GCD can reach at most 79.58% while that of other four descriptors are only 69.83%, 73.15%, 68.23% and 63.67%. And the average precision and recall of GCD are significantly higher than that of others when the total number of the retrieved images is increasing. Thus based on these analysis of retrieval results, our GCD is more discriminative and robust than MTH, MSD, CDH and SED. Fig. 9 shows some examples of image retrieval in this dataset. It can be seen that most of the retrieved images are correct when 30 images are returned.

To validate the universality of our GCD algorithm, we choose the other two Corel datasets, namely Corel-5 K and Corel-10 K datasets, to compare GCD with these image feature descriptors. There are 10 images chosen randomly for each category as query images in the Corel-5 K and Corel-10 K datasets. So there are 500 and 1000 query images in the two datasets respectively. Fig. 8 show the experimental results of comparison among GCD and the other four descriptors in these datasets. It can be seen that the GCD algorithm have better performance than MTH, MSD, CDH and SED.

MTH, MSD, CDH and SED are proposed based on the fusion of low-level feature, and have good performance in CBIR. MTH and CDH integrates the advantages of co-occurrence matrix and histogram and encode color, texture and shape information. But it is hard to fully represent the content of images. Micro-structure descriptor (MSD) extracts micro-structures based on an edge orientation similarity and the underlying colors. And structure element correlation statistics is used to characterize the spatial correlation of these micro-structures. SED also describes color and texture features by defining five structure elements in HSV color space. However, these descriptors define texton type directly in color space and then extract texture feature including color information. Although merging color and texture simultaneously, the performance of these descriptors may be limited in CBIR due to enhancing the characteristics of color features. The GCD algorithm extracts color and texture respectively to solve this problem. Then GCV and DGCV are proposed to represent these color and texture, respectively. With these method, the advantages of color histogram and SEC can be integrated. So the final GCD algorithm have better performance in CBIR systems.

## 6. Conclusion

In this paper, we propose a novel image feature descriptor, namely Global Correlation Descriptor (GCD), which characterizes color and texture features. With this method, color and texture information can be expressed equivalently. Image pyramid in HSV color space is used to describe color information of multiresolution images. And we propose a new method to describe the texton types of filtered images. In addition, Global Correlation Vector (GCV) and Directional Global Correlation Vector (DGCV) which can integrate the advantages of color histogram and SEC are presented to extract color and texture features, respectively. The experimental results have demonstrated that the GCD algorithm is much more robust and discriminative than other image descriptors in CBIR.

## Acknowledgment

This work was supported by National Natural Science Foundation of PR China (Grant Nos. 61173163 and 61370200).

## References

[1] Y. Liu, D. Zhang, G. Lu, et al., A survey of content-based image retrieval with high-level semantics, Pattern Recogn. 40 (1) (2007) 262–282.
[2] Y. Liu, D. Zhang, G. Lu, Region-based image retrieval with high-level semantics using decision tree learning, Pattern Recogn. 41 (8) (2008) 2554–2570.
[3] Y. Rui, T.S. Huang, S.F. Chang, Image retrieval: current techniques, promising directions, and open issues, J. Vis. Commun. Image Represent. 10 (1) (1999) 39–62.
[4] R. Datta, J. Li, J.Z. Wang, Content-based image retrieval: approaches and trends of the new age, in: Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, ACM, 2005, pp. 253–262.
[5] O.A.B. Penatti, F.B. Silva, E. Valle, et al., Visual word spatial arrangement for image retrieval and classification, Pattern Recogn. 47 (2) (2014) 705–720.
[6] M.E. ElAlami, A new matching strategy for content based image retrieval system, Appl. Soft Comput. 14 (2014) 407–418.
[7] M.J. Swain, D.H. Ballard, Color indexing, Int. J. Comput. Vis. 7 (1) (1991) 11–32.
[8] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[9] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, vol. 1, IEEE, 2005, pp. 886–893.
[10] M.A. Stricker, M. Orengo, Similarity of color images, in: IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology, International Society for Optics and Photonics, 1995, pp. 381–392.
[11] J. Huang, S.R. Kumar, M. Mitra, et al., Image indexing using color correlograms, in: Proceedings, 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997, IEEE, 1997, pp. 762–768.
[12] G. Pass, R. Zabih, J. Miller, Comparing images using color coherence vectors, in: Proceedings of the Fourth ACM International Conference on Multimedia, ACM, 1997, pp. 65–73.
[13] R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybernet. (6) (1973) 610–621.
[14] T. Watanabe, S. Ito, K. Yokoi, Co-occurrence histograms of oriented gradients for pedestrian detection, Advances in Image and Video Technology, Springer, Berlin, Heidelberg, 2009, pp. 37–47.
[15] D.G. Lowe, Object recognition from local scale-invariant features, The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999, vol. 2, IEEE, 1999, pp. 1150–1157.
[16] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[17] J.P. Jones, L.A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, J. Neurophysiol. 58 (6) (1987) 1233–1258.
[18] C. Liu, H. Wechsler, A Gabor feature classifier for face recognition, Proceedings, Eighth IEEE International Conference on Computer Vision, 2001, ICCV 2001, vol. 2, IEEE, 2001, pp. 270–275.
[19] T. Serre, L. Wolf, S. Bileschi, et al., Robust object recognition with cortex-like mechanisms, IEEE Trans. Pattern Anal. Mach. Intell. 29 (3) (2007) 411–426.
[20] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
[21] R.S. Dubey, R. Choubey, J. Bhattacharjee, Multi feature content based image retrieval, Int. J. Comput. Sci. Eng. 2 (06) (2010) 2145–2149.
[22] Z.C. Huang, P.P.K. Chan, W.W.Y. Ng, et al., Content-based image retrieval using color moment and Gabor texture feature, 2010 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 2, IEEE, 2010, pp. 719–724.
[23] S. Jalali, C. Tan, J.H. Lim et al., Visual Recognition using a Combination of Shape and Color Features, 2013.
[24] S. Jalali, P.J. Seekings, C. Tan, et al., Classification of marine organisms in underwater images using CQ-HMAX biologically inspired color approach, in: The 2013 International Joint Conference on Neural Networks (IJCNN), IEEE, 2013, pp. 1–8.
[25] G.H. Liu, J.Y. Yang, Image retrieval based on the texton co-occurrence matrix, Pattern Recogn. 41 (12) (2008) 3521–3527.
[26] G.H. Liu, L. Zhang, Y.K. Hou, et al., Image retrieval based on multi-texton histogram, Pattern Recogn. 43 (7) (2010) 2380–2389.
[27] G.H. Liu, J.Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recogn. 46 (1) (2013) 188–198.
[28] G.H. Liu, Z.Y. Li, L. Zhang, et al., Image retrieval based on micro-structure descriptor, Pattern Recogn. 44 (9) (2011) 2123–2133.
[29] X. Wang, Z. Wang, A novel method for image retrieval based on structure elements' descriptor, J. Vis. Commun. Image Represent. 24 (1) (2013) 63–74.
[30] X. Wang, Z. Wang, The method for image retrieval based on multi-factors correlation utilizing block truncation coding, Pattern Recogn. 47 (10) (2014) 3293–3303.
[31] M. Singha, K. Hemachandran, Performance analysis of color spaces in image retrieval, Assam Univ. J. Sci. Technol. 7 (2) (2011) 94–104.
[32] L. Junling, K. Degang, Z.H. Wei, et al., Image retrieval based on weighted blocks and color feature, in: 2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC), IEEE, 2011, pp. 921–924.
[33] T. Lindeberg, Scale-space theory: a basic tool for analyzing structures at different scales, J. Appl. Stat. 21 (1–2) (1994) 225–270.