CrossMark

Image retrieval based on exponent moments descriptor and localized angular phase histogram

Xiang-Yang Wang[1] · Lin-Lin Liang[1] · Yong-Wei Li[1] ·
Hong-Ying Yang[1]

Abstract Multiple feature extraction and combination is one of the most important issues in the content-based image retrieval (CBIR). In this paper, we propose a new content-based image retrieval method based on an efficient combination of shape and texture features. As its shape features, exponent moments descriptor (EMD), which has many desirable properties such as expression efficiency, robustness to noise, geometric invariance, fast computation etc., is adopted in RGB color space. As its texture features, localized angular phase histogram (LAPH) of the intensity component, which is robust to illumination, scaling, and image blurring, is used in hue saturation intensity (HSI) color space. The combination of above shape and texture information provides a robust feature set for color image retrieval. Experimental results on well known databases show significant improvements in retrieval rates using the proposed method compared with some current state-of-the-art approaches.

1 Introduction

WITH advances in information technology, there is an explosive growth of image databases (DBs), which demands effective and efficient tools that allow users to search through such a large collection. Traditionally, the most straightforward way to implement image database management systems is by means of using the conventional database-management systems

---

✉ Xiang-Yang Wang
    wxy37@126.com

✉ Hong-Ying Yang
    yhy_65@126.com

[1] School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, People's Republic of China

such as relational databases or object-oriented databases. The system of these kinds is usually called keyword-based, in which the images are annotated with keywords. As the databases grow larger, to retrieve a particular image with these methods becomes tedious and inadequate. To solve these problems, content-based image retrieval has emerged as a promising alternative, and has drawn substantial research attention in the last decade [8]. In a typical CBIR, low-level features related to visual content are first extracted from a query image, the similarity between the set of features of the query image and that of each target image in a DB is then computed, and target images are next retrieved which are most similar to the query image. Extraction of good visual features which compactly represent a query image is one of the important tasks in CBIR [13].

Most of the early studies on CBIR have used only a single feature among various low-level visual features. However, it is hard to attain satisfactory retrieval results by using a single feature because, in general, an image contains various visual characteristics. Recently, active researches in image retrieval using a combination of some low-level visual features have been performed [12]. In these methods, some different low-level visual features are extracted and combined, but it is shown that such a combination of features does not always guarantee better retrieval accuracy [12, 40, 41]. It is a challenging work to choose visual features that are complementary to each other and to combine chosen features effectively so as to yield an improved retrieval performance.

In this paper, we propose a new content-based image retrieval method based on an efficient combination of shape and texture features. The novelty of the proposed method includes: 1) Exponent moments (EMs) is introduced to describe image shape, which has many desirable properties such as expression efficiency, robustness to noise, geometric invariance, and fast computation etc.; 2) localized angular phase histogram is used to depict image texture, which is robust to illumination, scaling, and image blurring; 3) EMs descriptor and LAPH are combined effectively for color image retrieval.

The rest of this paper is organized as follows. A review of previous related work is presented in Section II. Section III recalls the decomposition and reconstruction about EMs. Section IV provides the detailed construction of LAPH. Section V discusses the content based color image retrieval using EMD and LAPH. Simulation results in Section VI will show the performance of our CBIR scheme. Finally, Section VII concludes this presentation.

2 Related work

Over the past decades, CBIR has been an active area of research in many applications, and many low-level visual features have been extracted and applied for image retrieval. Color is one of the most common and determinant low-level visual features, which is stable against direction variations, size of image and background complexity. As conventional color features used in CBIR, there are color histogram, color correlogram, color structure descriptor (CSD), and scalable color descriptor (SCD). The latter two are MPEG-7 color descriptors [40]. Chen et al. [6] proposed an adaptive color feature extraction scheme by considering the color distribution of an image. Based on the binary quaternion-moment-preserving (BQMP) thresholding technique, the proposed extraction methods, fixed cardinality (FC) and variable cardinality (VC), are able to extract color features by preserving the color distribution of an image up to the third moment and to substantially reduce the distortion incurred in the extraction process. Li et al. [20] presented a novel algorithm based on running subblocks with

different similarity weights for object-based image retrieval. By splitting the entire image into certain subblocks, the color region information and similarity matrix analysis are used to retrieval images under the query of special object. Liu et al. [26] proposed a novel image feature representation method, namely, the color difference histogram (CDH), which is used to describe image features for image retrieval. For the proposed histogram, orientation and perceptual color information are combined in the unified framework, and both of their spatial layouts are considered. Talib et al. [39] proposed a new semantic feature extracted from dominant colors (weight for each DC). The newly proposed technique helps reduce the effect of image background on image matching decision where an object's colors receive much more focus. Aptoula et al. [4] presented three morphological color descriptors, one making use of granulometries independently computed for each subquantized color and two employing the principle of multiresolution histograms for describing color, using respectively morphological levelings and watersheds.

Texture is an important visual attribute both for human perception and image analysis systems, its role in domain-specific image retrieval is particularly vital due to their close relation to the underlying semantics in these cases. Texture features, such as gray-level co-occurrence matrix (GLCM), Markov random field (MRF) model, simultaneous auto-regressive (SAR) model, Wold decomposition model, edge histogram descriptor (EHD) etc., have long been studied in image processing, computer vision, and computer graphics. He et al. [11] presented a novel method, which uses non-separable wavelet filter banks, to extract the features of texture images for texture image retrieval. Compared to traditional tensor product wavelets (such as DB wavelets), the new method can capture more direction and edge information of texture images. Lasmar et al. [19] introduced two new multivariate models using, respectively, generalized Gaussian and Weibull densities. These models can capture both the subband marginal distributions and the correlation between wavelet coefficients. Aptoula [3] presented the results of applying global morphological texture descriptors to the problem of content-based remote sensing image retrieval. Specifically, they explored the potential of recently developed multiscale texture descriptors, namely, the circular covariance histogram and the rotation-invariant point triplets. Pappas et al. [28] reviewed recently proposed texture similarity metrics and applications that critically depend on such metrics, with emphasis on image and video compression and content-based retrieval. Atto et al. [5] derived a 2-D spectrum estimator from some recent results on the statistical properties of wavelet packet coefficients of random processes, and discussed the performance of this wavelet-based estimator, in comparison with the conventional 2-D Fourier-based spectrum estimator on texture analysis and content-based image retrieval. Rakvongthai et al. [32] investigated the use of complex wavelets for texture retrieval in a noisy environment where the query image is noisy. Based on a statistical framework, the feature vector is formed by modeling an image in the complex wavelet domain and estimating parameters from the image.

Shape is known to play an important role in human recognition and perception. Object shape features provide a powerful clue to object identity. Humans can recognize objects solely from their shapes. The significance of shape as a visual feature can be seen from the fact that every major CBIR system incorporates some shape features in one form or another [44]. By using a mathematical form of analysis, Li et al. [21] compared the amount of visual informa-tion captured by Zernike moments(ZMs) phase and the amount captured by ZM magnitude, and then proposed combining both the magnitude and phase coefficients to form a new shape descriptor for CBIR. Shu et al. [36] suggested a novel shape contour descriptor for shape matching and retrieval. The new descriptor is called contour points distribution histogram

(CPDH) which is based on the distribution of points on object contour under polar coordinates. CPDH not only conforms to the human visual perception but also the computational complexity of it is low. Jian et al. [15] proposed an efficient method based on singular values and potential-field representation for face-image retrieval, in which the rotation-shift-scale-invariant properties of the singular values are exploited to devise a compact, global feature for face-image representation. Liu et al. [25] considered the family of total jBregman divergences (tBDs) as an efficient and robust "distance" measure to quantify the dissimilarity between shapes, and used the tBD-based l1-norm center as the representative of a set of shapes. Anuar et al. [2] proposed a novel technique for trademark retrieval that demonstrates improved performance due to the integration of two shape descriptors. The technique employs the Zernike moments as the global descriptor and the edge-gradient co-occurrence matrix as the local descriptor.

Most of the early studies on CBIR have used only a single feature among various low-level visual features. However, it is hard to attain satisfactory retrieval results by using a single feature because, in general, an image contains various visual characteristics. Recently, active researches in image retrieval using a combination of some low-level visual features have been performed [7]. Lin et al. [24] proposed three image features for image retrieval. The first and second image features are based on color and texture features, respectively called color co-occurrence matrix (CCM) and difference between pixels of scan pattern (DBPSP). The third image feature is based on color distribution, called color histogram for K-mean (CHKM). Yap et al. [43] proposed a content-based image retrieval using Legendre chromaticity distribution moments (LCDM), which can provide a compact, fixed-length and computation effective representation of the color contents of an image. Yu et al. [16] considered multiple features from different views, i.e., color histogram, Hausdorff edge feature, and skeleton feature, to represent cartoon characters with different colors, shapes, and gestures. Each visual feature reflects a unique characteristic of a cartoon character, and they are complementary to each other for retrieval and synthesis. Jacob et al. [14] proposed local oppugnant color texture pattern (LOCTP), which is able to discriminate the information derived from spatial inter-chromatic texture patterns of different spectral channels within a region. It determines the relationship in terms of the intensity and directional information between the referenced pixels and their oppugnant neighbors. The LOCTP strives to use the harmonized link between color and texture, which helps the system to incorporate the human perception. Farsi et al. [9] presented a novel CBIR method based on combination of Hadamard matrix and discrete wavelet transform (HDWT) in hue-min-max-difference colour space. An average normalized rank and combination of precision and recall are considered as metrics to evaluate and compare the proposed method against different methods. Kashif et al. [17] proposed a CBIR scheme based on the three well-known algorithms: colour histogram, texture and moment invariants. Here, colour histogram is used to extract the colour features of an image. Gabor filter is used to extract the texture features, and the moment invariant is used to extract the shape features of an image. Prasad et al. [31] presented a technique to retrieve images by region matching using a combined feature index based on color, shape, and location within the framework of MPEG-7. Dominant regions within each image are indexed using integrated color, shape, and location features. Various combinations of regions are also indexed. Seetharaman et al. [34] proposed a unified scheme for automatic image retrieval based on the multivariate parametric tests. In the proposed technique, mean and covariance are used as representatives of both query and target images, and statistical features such as coefficient of variation, skewness, kurtosis, variance-covariance, spectrum of energy, and number of shapes

in the images are used. Sherin [35] proposed a novel integrated curvelet-based image retrieval scheme (ICTEDCT-CBIR), in which curvelet multiscale ridgelets is integrated with region-based vector codebook subband clustering for enhanced dominant colors extraction and texture analysis. Singh et al. [30] proposed a novel solution to image retrieval system by combining local and global features. Local features are extracted by detecting linear edges of the edge map of the image using Hough transform and then computing the normalized histograms of distances of lines from the centroid of the edge image. The global features are represented by Zernike moments. Susana et al. [38] proposed a color-texture descriptor: the texture component descriptor (TCD), that arise from the decomposition of the image in its textural components, which are groups of blobs with similar attributes either color, shape or orientation. Wang et al. [42] proposed a new and effective color image retrieval scheme which uses the combination of dynamic dominant color, Steerable filter texture feature, and pseudo Zernike moments shape descriptor. Singha et al. [37] proposed a CBIR approach based on the combination of Haar wavelet transformation using lifting scheme and the colour histogram (CH). The colour feature is described by the CH, which is translation and rotation invariant. The Haar wavelet transformation is used to extract the texture features and the local characteristics of an image. The lifting scheme reduces the processing time to retrieve images. In [23], two-dimensional or one-dimensional histograms of the CIELab chromaticity coordinates are chosen as color features, and variances extracted by discrete wavelet frames analysis are chosen as texture features. In these methods, some different low-level visual features are extracted and combined, but it is shown that such a combination of features does not always guarantee better retrieval accuracy [12, 40, 41]. It is a challenging work to choose visual features that are complementary to each other and to combine chosen features effectively so as to yield an improved retrieval performance.

3 Exponent moments descriptor (EMD)

Shape is known to play an important role in human recognition and perception [1]. Object shape features provide a powerful clue to object identity. Humans can recognize objects solely from their shapes. The significance of shape as a feature for content-based image retrieval can be seen from the fact that every major CBIR system incorporates some shape features in one form or another. As the most commonly used approaches for shape descriptors, moments and function of moments have been utilized as pattern features in a number of applications [1]. The theory of moments provides useful series expansions for the representation of object shapes. In this section, we propose a robust and effective shape feature, which is based on a set of new orthogonal moments of images known as Exponent moments.

In 2011, Meng & Ping [27] extended radial harmonic Fourier moments and introduced a new moment named Exponent moments. Compared with other orthogonal moment, EMs has many desirable properties such as better image reconstruction, lower noise sensitivity, geometric invariance, lower computational complexity. Besides, the EMs is free of numerical instability issues so that high order moments can be obtained accurately.

A function set $P_{n,m}(r, \theta)$ defined in a polar coordinate system (r, θ) contains the radial function $A_n(r)$ and Fourier factor in angle direction $\exp(jm\theta)$

$$P_{n,m}(r, \theta) = A_n(r)\exp(jm\theta) \tag{1}$$

where $A_n(r) = \sqrt{2/_r}\exp(j2n\pi r)$, $n, m = -\infty, \cdots, 0, \cdots, +\infty$, $0 \leq r \leq 1$, $0 \leq \theta \leq 2\pi$. According to the characteristic of radial function and Fourier factor in angle direction, the set of $P_{n,m}(r, \theta)$ is orthogonal and sound over the interior of the unit circle

$$\int_0^{2\pi} \int_0^1 P_{n,m}(r, \theta)P_{k,l}^*(r, \theta)rdrd\theta = 4\pi\delta_{n,k}\delta_{m,l} \tag{2}$$

where 4π is the normalization factor, $\delta_{n,k}$ and $\delta_{m,l}$ are the Kronecker symbols, and $P_{k,l}^*(r, \theta)$ is the conjugate of $P_{k,l}(r, \theta)$.

The image $f(r, \theta)$ can be decomposed with the set of $P_{n,m}(r, \theta)$ as

$$f(r, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} E_{n,m}A_n(r)\exp(jm\theta) \tag{3}$$

where $E_{n,m}$ is the EMs of order n with repetition m, whose definition is

$$E_{n,m} = \frac{1}{4\pi} \int_0^{2\pi} \int_0^1 f(r, \theta)A_n^*(r)\exp(-jm\theta)rdrd\theta \tag{4}$$

here, $A_n^*(r)$ is the conjugate of $A_n(r)$.

Following the principle of orthogonal function, the image function $f(r, \theta)$ can be reconstructed approximately by limited orders of EMs ($n \leq n_{max}$, $m \leq m_{max}$). The more orders used, the more accurate the image description

$$f'(r, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} E_{n,m}A_n(r)\exp(jm\theta) \approx \sum_{n=-n_{max}}^{n_{max}} \sum_{m=-m_{max}}^{m_{max}} E_{n,m}A_n(r)\exp(jm\theta) \tag{5}$$

where $\dot{f}(r, \theta)$ is the reconstructed image. The basis functions $A_n(r)\exp(jm\theta)$ of the EMs are orthogonal over the interior of the unit circle, and each order of the EMs makes an independent contribution to the reconstruction of the image.

Below, we will derive and analyze the geometric invariant property of EMs. Let $f^r(r, \theta) = f(r, \theta + \alpha)$ denote the rotation change of an image $f(r, \theta)$ by the angle α, then EMs of $f(r, \theta + \alpha)$ and $f(r, \theta)$ have the following relations

$$E_{n,m}(f^r) = E_{n,m}(f)\exp(\mu m\alpha)$$

where $E_{n,m}(f^r)$ and $E_{n,m}(f)$ are the EMs of $f^r(r, \theta)$ and $f(r, \theta)$, respectively. According to above equation, we know that a rotation of the image by an angle α induces a phase shift $e^{\mu m\alpha}$ of the $E_{n,m}(f)$. Taking the norm on both sides, we have

$$\left|E_{n,m}(f^r)\right| = \left|E_{n,m}(f)\exp(\mu m\alpha)\right| = \left|E_{n,m}(f)\right|\left|\exp(\mu m\alpha)\right| = \left|E_{n,m}(f)\right|$$

So, the rotation invariance can be achieved by taking the norm of the images' EMs. In other words, the EMs modulus $|E_{n,m}(f)|$ are invariant with respect to rotation transform. Besides, the EMs modulus is invariant to scaling if the computation area can be made to cover the same content. In practice, this condition is met because the EMs are defined on the unit disk.

Figure 1 gives some examples of image reconstruction using ZMs and EMs for image "Plane" and "Flower" (moment orders K = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50). As can be seen from the Fig. 1, the reconstructed images using EMs show more visual resemblance to the
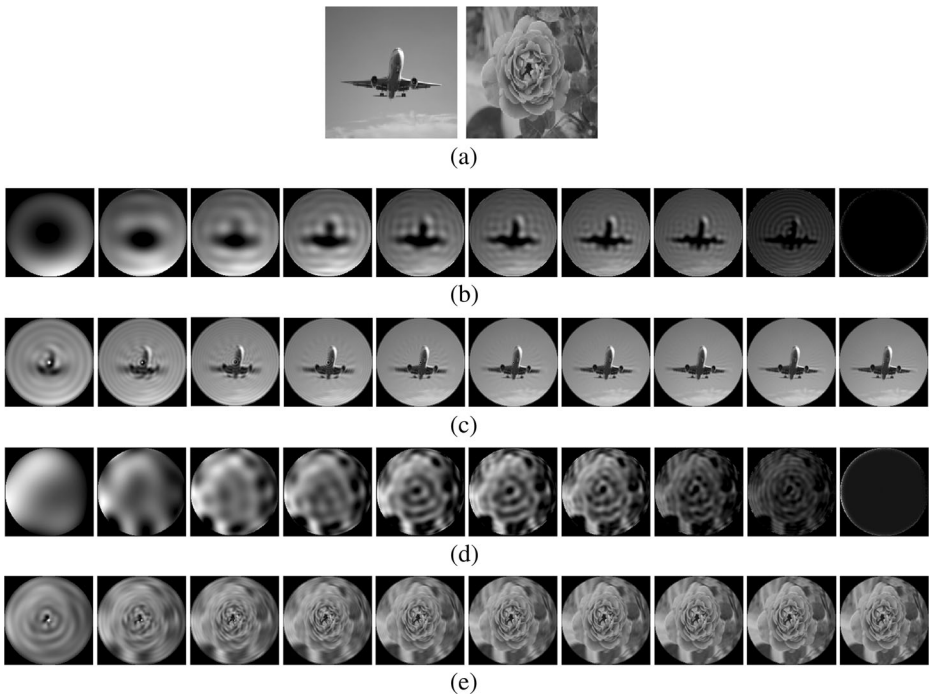
Fig. 1 Some samples of reconstructed images (moment orders K = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50): (**a**) Original images Plane and Flower, (**b**) Reconstructed images Plane for ZMs, (**c**) Reconstructed images Plane for EMs, (**d**) Reconstructed images Flower for ZMs, (**e**) Reconstructed images Flower for EMs

original image in the early orders. The edges of the reconstructed images are also better defined with less jaggedness. In Fig. 2, we show plots of the average reconstruction errors [10] for image "Plane" and "Flower". It can be observed that the reconstructed images using EMs show visual resemblance to the original image in the early orders, and the EMs is free of numerical instability issues. Figure 3 shows the comparison of the moments computing time using ZMs and EMs for image "Plane" and "Flower". Figures 4 and 5 show the EMs magnitudes for image "Plane" and "Flower" under various common image processing
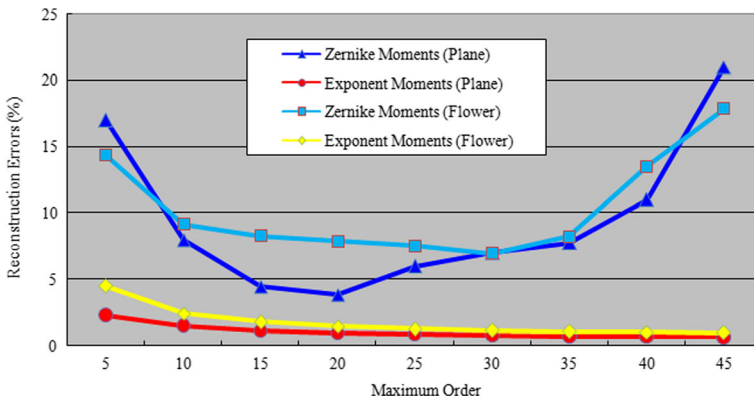
Fig. 2 The reconstruction error yielded by ZMs and EMs in the image reconstruction experiment

Fig. 3 The moments computing time using ZMs and EMs for image "Plane" and "Flower"

operations and geometric transforms. It can be seen that the EMs magnitudes of image have good robustness against common image processing operations and geometric transforms. So, EMs magnitudes (called EMs descriptor, EMD) are suitable for CBIR system.

4 Localized angular phase histogram (LAPH)

Generally, texture features play a very important role in computer vision and pattern recognition, especially in describing the content of images. Importance of the texture feature is due to its presence in many real world images: for example, clouds, trees, bricks, hair, fabric etc., all of which have textural characteristics. Earlier methods for texture representation suffer from two main drawbacks: they are either computationally expensive or retrieval accuracy is poor [18, 22, 33]. In this section, we introduce a new texture feature, the localized angular phase histogram, for CBIR, which are efficient both in terms of accuracy and computational complexity.

4.1 Localized angular phase (LAP) of image pixel

In this paper, LAP is based on the localized Fourier transform that provides information in both the time and frequency domains. In contrast with the 2D short term Fourier transform, LAP applies the 1D Fourier transform over a 1D signal of pixels from local image window. The phase is obtained by computing the arctangent of the division between imaginary and real Fourier coefficients. The phase sign is analyzed to form 8-bit codewords where the distribution of their decimal values is used to describe image pixel texture.

For pixel texture extraction purpose, we use the *HSI* representation of the color image because this color space can control color and intensity information independently. Here, the image pixel texture is extracted from the *I* component, this is because that the *ISS* component closely matches human perception of lightness. For each image pixel, the construction of LAP can be summarized as follows.

1) Construct Local Image Window and Convert Signal

As shown in Fig. 6, the local image window (shadow part) is firstly constructed centered on pixel $s(x, y)$, and 17 pixels from local image window are then converted into a 1D discrete signal along arrow direction. We denote this discrete signal by $p(n), \quad n = 0, 1, \ldots, 16$.
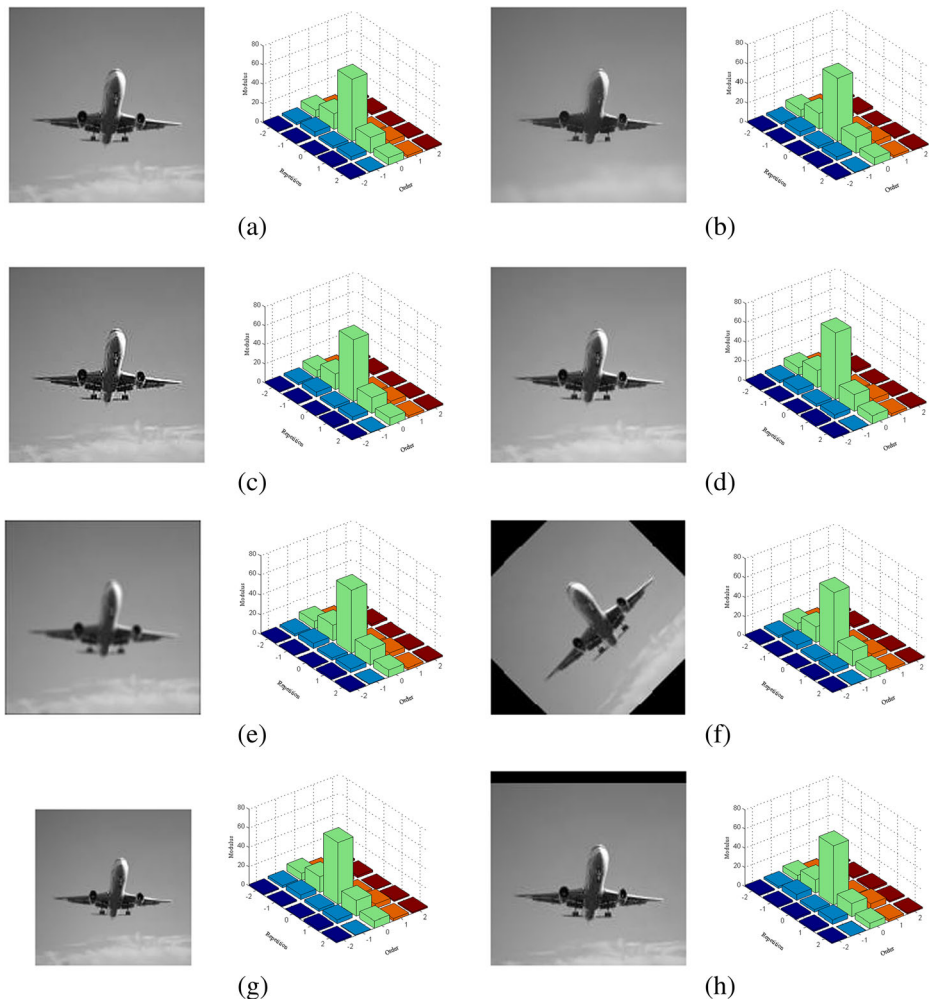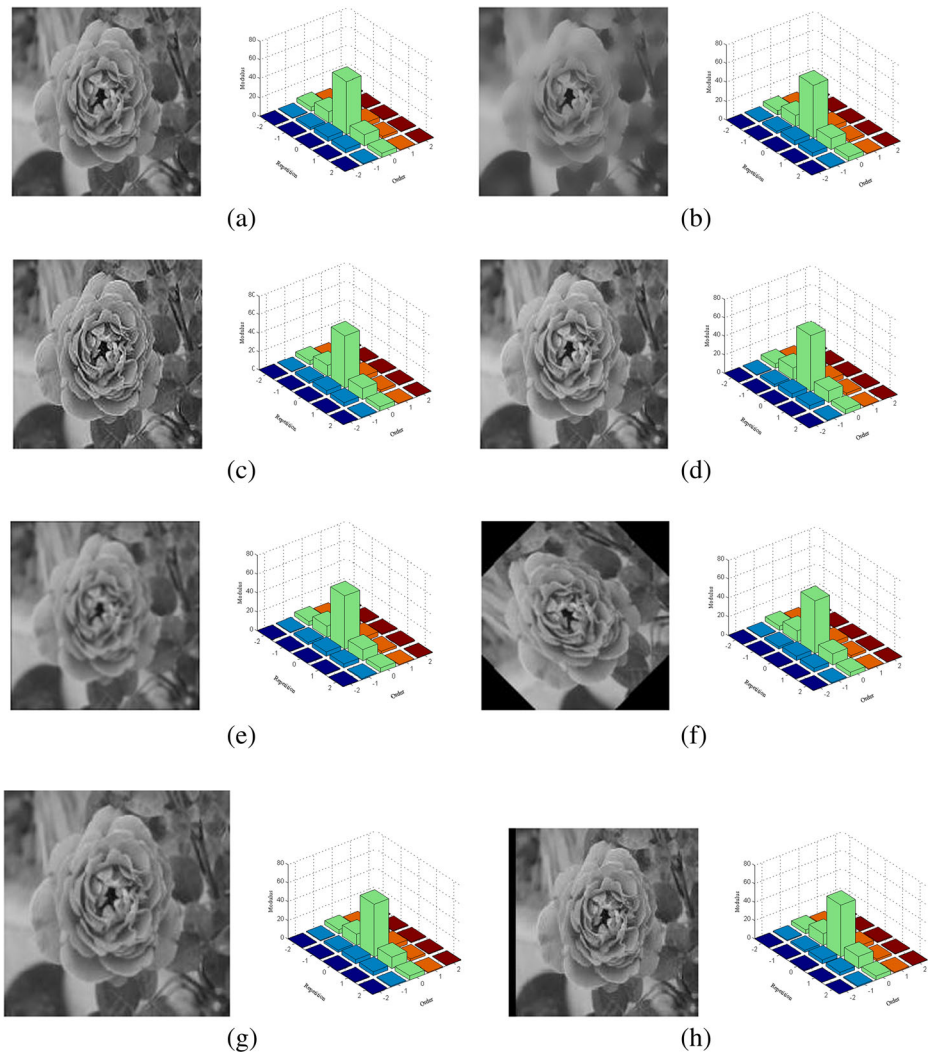
Fig. 4 The EMs magnitudes for image Plane under various common image processing operations and geometric transforms: (**a**) Original image, (**b**) Image blurring, (**c**) Edge Sharpening, (**d**) Light increasing, (**e**) Median filtering, (**f**) Rotation, (**h**) Scaling, (**i**) Translation

2) Perform 1D Fourier Transform

The 1D Fourier transform and inverse transform of $p(n)$ are given by

$$P(k) = \sum_{n=0}^{N-1} p(n)e^{-\frac{2\pi i}{N}kn}, \quad p(n) = \frac{1}{N}\sum_{k=0}^{N-1} P(k)e^{-\frac{2\pi i}{N}kn} \tag{6}$$

where N is the number of samples in $p(n)$, and for local image window, N is 17. Using (6), the discrete signals $p(n)$ are converted to the Fourier coefficients $P(k)$.

After the Fourier transform, the values of 17 complex coefficients $P(0), P(1), \ldots, P(16)$ are obtained. The next step is to select some complex coefficients to extract the phase information.

Fig. 5 The EMs magnitudes for image Flower under various common image processing operations and geometric transforms: (**a**) Original image, (**b**) Image blurring, (**c**) Edge Sharpening, (**d**) Light increasing, (**e**) Median filtering, (**f**) Rotation, (**h**) Scaling, (**i**) Translation

3) Extract Phase Information

The $P(0)$ is the DC value of the Fourier transform and contains no phase information, thus it is excluded from the selected coefficients. Because the image contains only real values, its Fourier transform becomes centrally symmetric where half of the coefficients are redundant. However, if the number of samples is even, e.g., 16 samples, then the resulting complex coefficients will have another DC value. This extra DC value will reduce the number of useful non redundant complex coefficients. So, to avoid information loss, the LAP method uses 17 samples, instead. This will result in only one DC value. Then, 8 non redundant complex coefficients are selected, whereby half of the complex coefficients are either $P(1), P(2), …,$ $P(8)$ or $P(9), P(10), …, P(16)$. So, the phase information C can be extracted from these 8
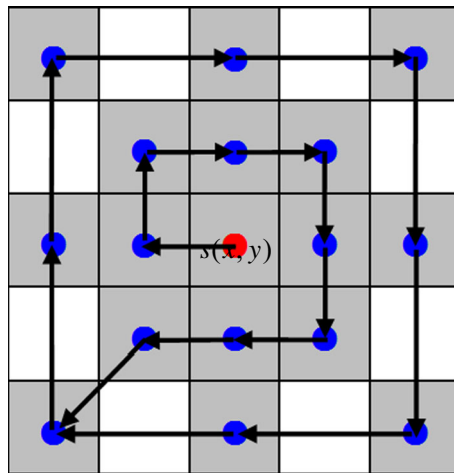
Fig. 6 Local image window for computing LAP of image pixel s(x,y)

complex Fourier coefficients by computing the arctangent of the division between imaginary and real Fourier coefficients.

$$C = [C_1 C_2 C_3 C_3 C_5 C_6 C_7 C_8]$$
$$C_i = \arctan\frac{a_i}{b_i} \cdot (i = 1, 2, \cdots, 8)$$

where a_i and b_i are real part and imaginary part of the selected complex Fourier coefficients.

4) Construct Localized Angular Phase (LAP)

Phase information matrix C is quantized into 8-bit binary code by using the following formula

$$b(k) = \begin{cases} 1, if\, C_k \geq 0 \\ 0, otherwise \end{cases} \tag{7}$$

where $b(k)$ is the sign of each phase information C_k.

By arranging $b(1)$, $B(2)$, …, $b(8)$, the 8 bit binary code can be formulated, and a binomial factor is assigned as 2 for each $b(k)$, hence it is possible to transform (7) into a unique LAP number (LAP element), given by

$$LAP = \sum_{k=1}^{8} b(k)2^{k-1} \tag{8}$$

Based on (8), this LAP is a decimal value between 0 and 255 resulting from the 8-bit binary code. Figure 7 shows the color images and their LAP matrices.

4.2 Localized angular phase histogram (LAPH) of image

Considerable research has been carried out on the basis of image content. The most popular representation of image information is global histogram. Statistically, the histogram denotes the joint probability of intensities of the three image channels, thus describing the global pixel distribution in an image. In general, the histogram provides useful clues for the subsequent expression of similarity between images, due to its robustness to background complications
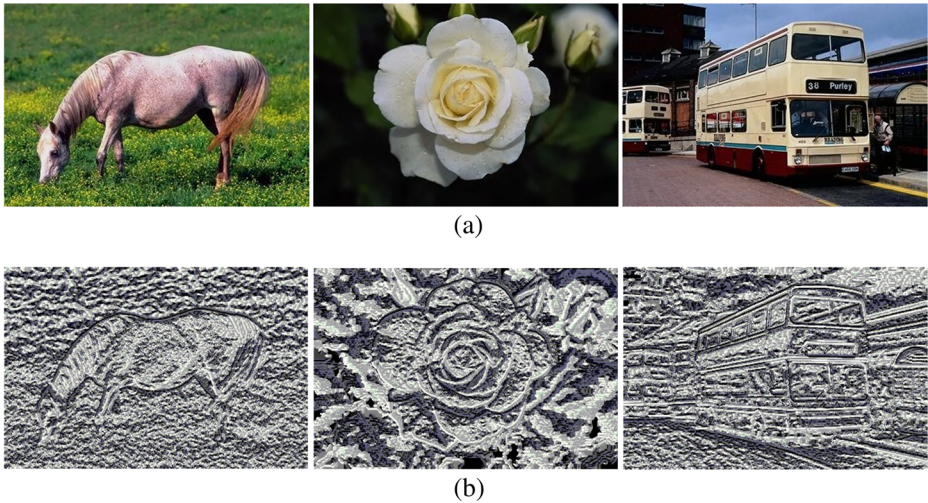
(a)

(b)

Fig. 7 The color images and their LAP matrices: (a) The color image, (b) The LAP matrix

and object distortion. Moreover, it is translation, scale, and rotation invariant, very simple to implement and systems encountering histograms exhibit a fast retrieval response that makes real-time implementation easier. In this paper, we introduce a new texture feature, the localized angular phase histogram (LAPH), for CBIR, which is efficient both in terms of accuracy and computational complexity.

Let L denote the LAP matrix of color image I, which contains N LAP elements, and the corresponding LAPH can be presented as

$$H(k) = \frac{n_k}{N}(k = 0, 1, \cdots, 255) \qquad (9)$$

where n_k is the total number of LAP elements in the k th bin.

Figures 8 and 9 shows the traditional color histogram(TCH) and localized angular phase histogram (LAPH) for two images with similar/different content. From Figs. 8 and 9, we can see that LAPH can reflect effectively the image content, and is superior to TCH.

5 Content-based color image retrieval using EMD and LAPH

For content-based color image retrieval (CBIR), image features in color image database are extracted and stored in an index file that is linked to the original color images. The descriptor of the query color image is represented in vector form and the similarity is calculated between the descriptor vectors of database color images and of the query color image. This section presents a content-based color image retrieval scheme based on EMD and LAPH. Figure 10 describes our image retrieval system framework.

5.1 EMD computing and selection

According to Section II, we can compute rapidly the EMD, i.e. EMs magnitudes. However, we don't need too much EMs magnitudes in color image retrieval, since color image features can normally be captured by just a few low-frequency EMs magnitudes. The choice of the max
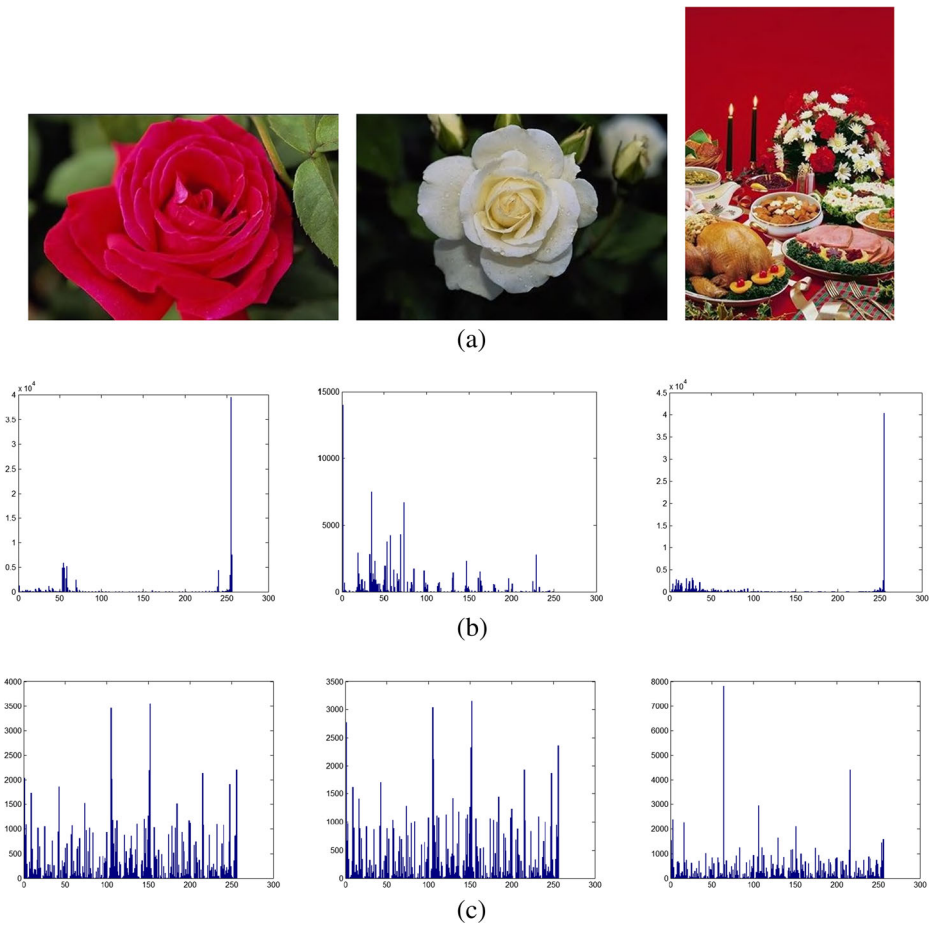
Fig. 8 The TCH and LAPH for different images with similar/different content: (**a**) The color images with similar/different content, (**b**) The TCH for different images with similar/different content, (**c**) The LAPH for different images with similar/different content

order value $n_{\max}$ will depend on the size of the given color image and also on the resolution needed. Besides, we must consider fully the symmetrical characteristic of EMs magnitudes distribution ($|E_{n,m}| = |E_{-n,-m}|$) when the EMs magnitudes are selected. Table 1 lists the selected EMD for different max orders (In this paper, the max order value is selected as 5). So, the shape feature vector based on EMD in RGB color space is given by

$$\mathbf{F}_1 = \left(\mathbf{E}^R, \mathbf{E}^G, \mathbf{E}^B \right) \tag{10}$$

where $\mathbf{E}^R$, $\mathbf{E}^G$, $\mathbf{E}^B$ denote the EMD of Red, Green, and Blue components, respectively.

5.2 LAPH computing

According to Section III, we can compute rapidly the LAP of each image pixel, and further extract the LAPH of image in HSI color space. So, the new texture feature vector based on LAPH is denoted as $\mathbf{F}_2$.
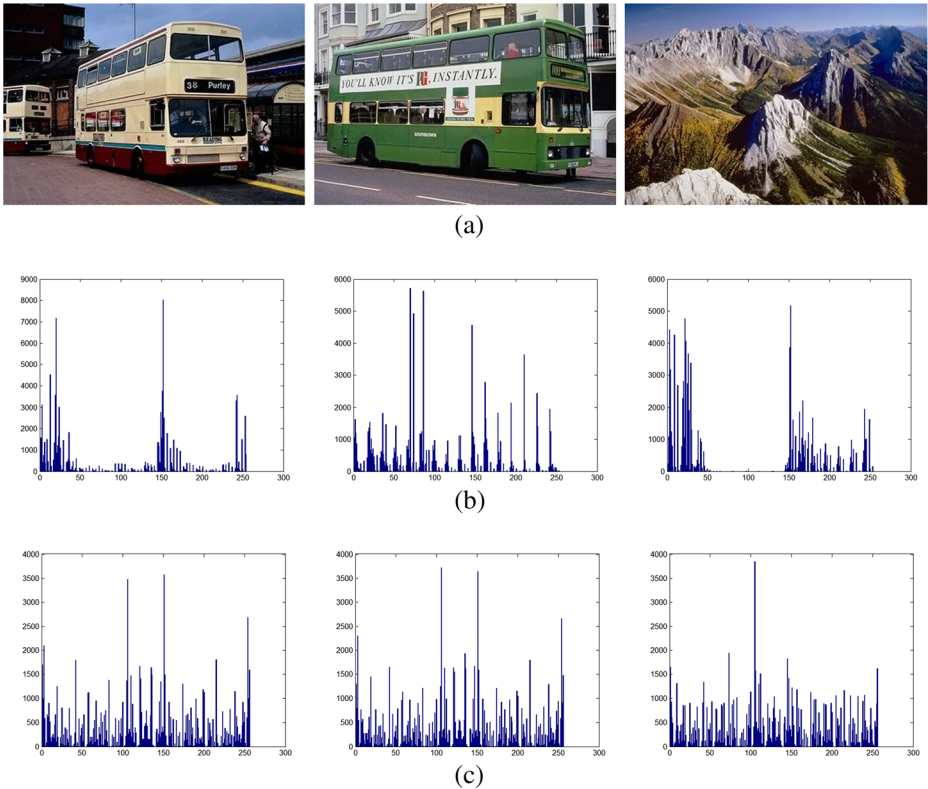
(a)

(b)

(c)

Fig. 9 The TCH and LAPH for different images with similar/different content: (**a**) The different color images with similar/different content, (**b**) The TCH for different images with similar/different content, (**c**) The LAPH for different images with similar/different content

5.3 Similarity measure

After the shape feature vector $\mathbf{F}_1$ and texture feature vector $\mathbf{F}_2$ are extracted, the retrieval system combines these feature vectors, calculates the similarity between the combined feature vector of the query image and that of each target image in an image DB, and retrieves a given number of the most similar target images.
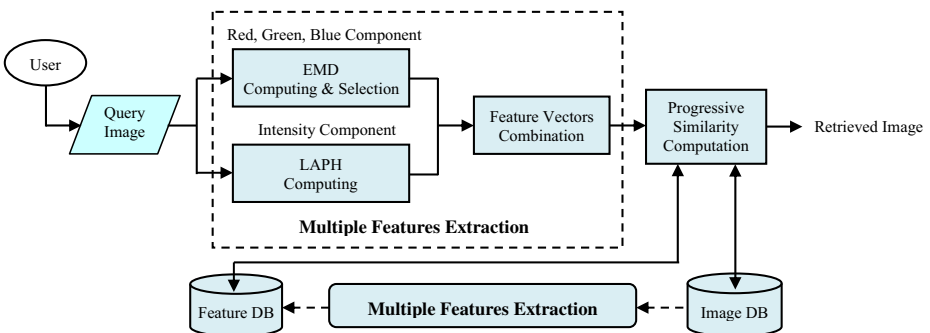
Fig. 10 Block diagram of the proposed color image retrieval system

Table 1 List of the selected EMD for different max orders

Order	EMD	No. of Moments	Accumulative No.
1	$\|E_{0,0}\|$	1	1
2	$\|E_{-1,1}\|; \|E_{0,1}\|; \|E_{1,1}\|; \|E_{1,0}\|$	4	5
3	$\|E_{-2,2}\|; \|E_{-1,2}\|; \|E_{0,2}\|; \|E_{1,2}\|; \|E_{2,2}\|;$ $\|E_{2,1}\|; \|E_{2,0}\|; \|E_{2,-1}\|$	8	13
4	$\|E_{-3,3}\|; \|E_{-2,3}\|; \|E_{-1,3}\|; \|E_{0,3}\|; \|E_{1,3}\|;$ $\|E_{2,3}\|; \|E_{3,3}\|; \|E_{3,2}\|; \|E_{3,1}\|; \|E_{3,0}\|; \|E_{3,-1}\|; \|E_{3,-2}\|$	12	25
5	$\|E_{-4,4}\|; \|E_{-3,4}\|; \|E_{-2,4}\|; \|E_{-1,4}\|; \|E_{0,4}\|; \|E_{1,4}\|; \|E_{2,4}\|;$ $\|E_{3,4}\|; \|E_{4,4}\|; \|E_{4,3}\|; \|E_{4,2}\|; \|E_{4,1}\|; \|E_{4,0}\|; \|E_{4,-1}\|; \|E_{4,-2}\|; \|E_{4,-3}\|$	16	41

(1) Shape Feature Similarity Measure

The shape feature similarity is given by

$$S_1(Q,I) = \left| \mathbf{F}_1^Q - \mathbf{F}_1^I \right| \tag{11}$$

where $\mathbf{F}_1^Q$ denotes the shape feature vector of query image Q, and $\mathbf{F}_1^I$ denotes the shape feature vector of target image I.

(2) Texture Feature Similarity Measure

We give the texture feature similarity as follows

$$S_2(Q,I) = \left| \mathbf{F}_2^Q - \mathbf{F}_2^I \right| \tag{12}$$

where $\mathbf{F}_2^Q$ denotes the texture feature vector of query image Q, and $\mathbf{F}_2^I$ denotes the texture feature vector of target image I.

(3) Feature Similarity Measure

So the distance used for computing the similarity between the query feature vector and the target feature vector is given as

$$S(I,Q) = w_1 S_1(Q,I) + w_2 S_2(Q,I)$$
$$w_1 + w_2 = 1 \tag{13}$$

where w_1 and w_2 are the weights of the shape and texture features respectively.

When retrieving images, we firstly calculate the similarity between the query image and each target image in the image DB, and then sort the retrieval results according to the similarity value.

6 Simulation results

To evaluate the performance of the proposed algorithm, we conduct an extensive set of CBIR experiments by comparing the proposed scheme to several state-of-the-art image retrieval approaches [7, 9, 43].

6.1 Image database

The color image retrieval systems have been implemented in MATLAB 7.0 programming environment on a Pentium 4 (2 GHz) PC. To check the retrieval efficiency of proposed method, we perform experiments over 10000 images from 150 categories of the COREL photo gallery, in which each category contains 100 images. Every database image is of size 256×384 or 384×256, which cover a variety of topics, such as "Flowers", "Buses", "Beach", "Elephants", "Sunset", "Buildings", "Horses", etc.

We also perform experiments over 10000 images from 256 object categories of the Caltech image database. The Caltech image database comprises 30607 images, in which each category has a minimum of 80 images. Caltech images are harvested from other popular online image database, and they represent a diverse set of lighting conditions, poses, backgrounds, image sizes, and camera systematics. The categories were hand-picked by the authors to represent a wide variety of natural and artificial objects in various setting. The organization is simple and the images are ready to use, without the need for cropping or other processing. Corel and Caltech images have been widely used by the image processing and CBIR research communities. Figure 11 shows our image retrieval system interface.

6.2 The performance of parameters selection

In our image retrieval, the EMs magnitudes (EMD) are used to capture the shape feature. To evaluate the overall performance of the proposed image feature in retrieval, a number of experiments were performed on our image retrieval.

In order to choose the "good" maximum EMs order value for extracting image feature, we randomly selected 500 images as query images from the above image database, and tested the image retrieval accuracies for different maximum EMs order values. Figure 12 shows the mean of retrieval precisions of 500 times query results for different maximum EMs order values, which reflects the image retrieval efficiency. From the Fig. 12, we can obtain the optimal maximum EMs order value $n_{\max} = 5$.

Figure 13 shows the average retrieval precisions of 500 times query results for different feature weight values w_1, w_2, which reflects the image retrieval efficiency. In Figs. 14 and 15, we demonstrate our retrieval results with EMD only, LAPH only, and both EMD and LAPH, respectively. The image at the top of left-hand corner is the query image; other 20 images are the retrieval results

6.3 Comparative performance evaluation

We report experimental results that show the feasibility and utility of the proposed algorithm and compare its performance with three state-of-the-art image retrieval approaches [7, 9, 43]. To simulate the practical situation of online users, the sequence of query images used in all the experiments is generated at random.

Figures 16, 17 and 18 show the image retrieval results using the scheme [43], scheme [7], scheme [9], and the proposed method. The image at the top of left-hand corner is the query image; other 20 images are the retrieval results.

In order to further confirm the validity of the proposed algorithm, we randomly selected 500 images as query images from the above image database (The tested 10
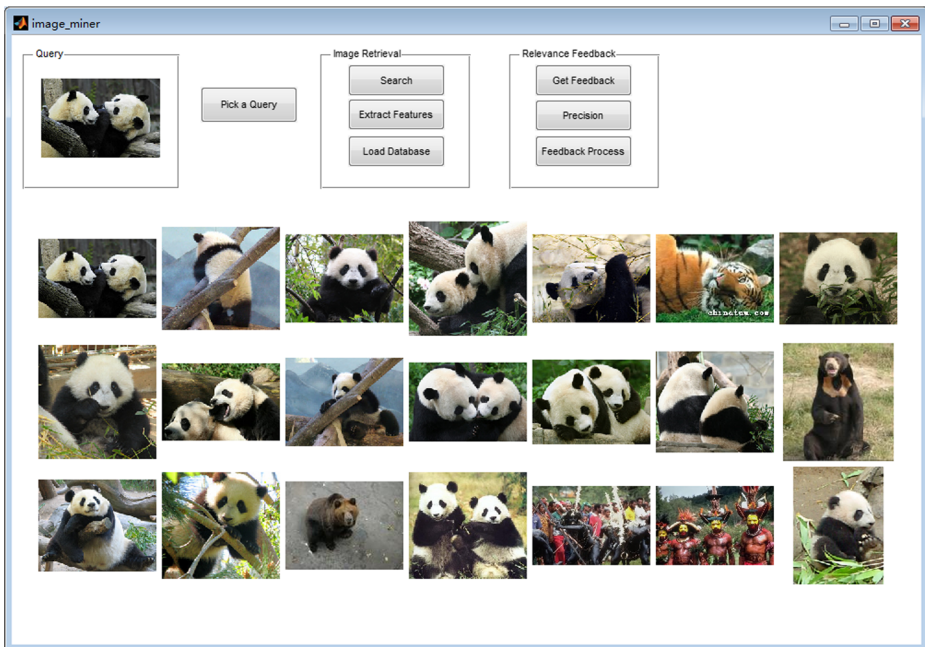
Fig. 11 User interface of our image retrieval system

semantic class includes bus, horse, flower, dinosaur, building, elephant, people, beach, scenery, and dish). Each kind is extracted 50 images, and each time returns the first 20 most similar images as retrieval results. To each kind of image, the average normal precision and average normal recall of 10 times query results are calculated. These values are taken as the retrieval performance standard of the algorithm, as shown in Fig. 19.

We also compared the proposed method with some state-of-the-art image retrieval approaches [9, 29] on another two publicly available retrieval datasets, NUS-WIDE and Oxford Buildings. The NUS-WIDE is a large-scale web image dataset collected from Flickr as a benchmark for evaluating multimedia search techniques, which contains 269648 images and their ground-truth annotations for 81 concepts. The Oxford Buildings contains 5062 high resolution images (1024 × 768) showing either one of the Oxford landmarks (the dataset contains 11 landmarks), or other places in Oxford. The database includes 5 queries for each landmark (55 queries in total), each of them including a bounding box that locates the object of interest. Figure 20 presented the average retrieval performance on NUS-WIDE and Oxford Buildings.

According to the Figs. 16, 17, 18, 19 and 20, we see that the image retrieval accuracy by the proposed method is competitive with the other tested methods. The effectiveness of the proposed image retrieval results from: (1) Exponent moments descriptor (EMD) is adopted to depict the image shape, which has many desirable properties such as expression efficiency, robustness to noise, geometric invariance, fast computation, etc.; (2) Localized angular phase histogram (LAPH) is used to describe texture information, which is robust to illumination, scaling, and image blurring.
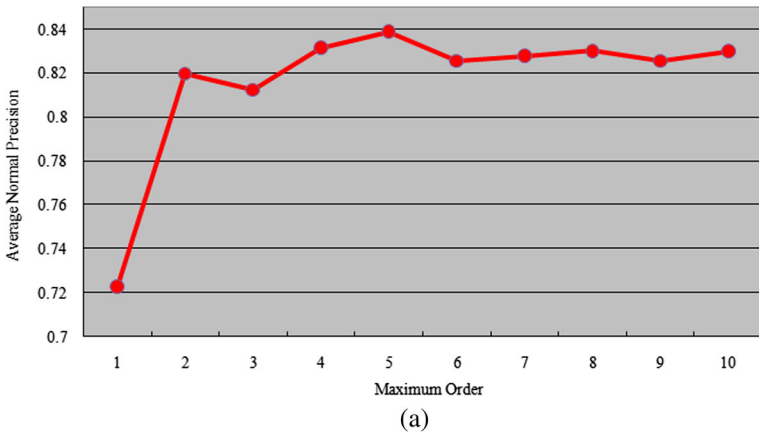
(a)

(b)

Fig. 12 The mean of retrieval precisions of 500 times query results for different maximum EMs order values: (a) Average retrieval performance for different maximum EMs order values ($n_{max} = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$), (b) Average retrieval performance for different maximum EMs order values ($n_{max} = 5, 10, 15, 20, 25$)
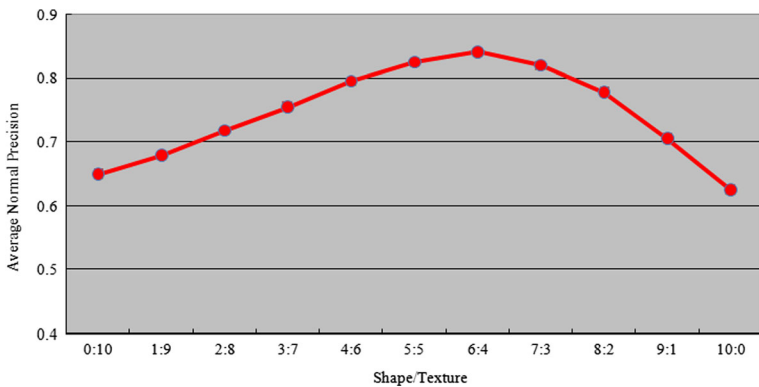

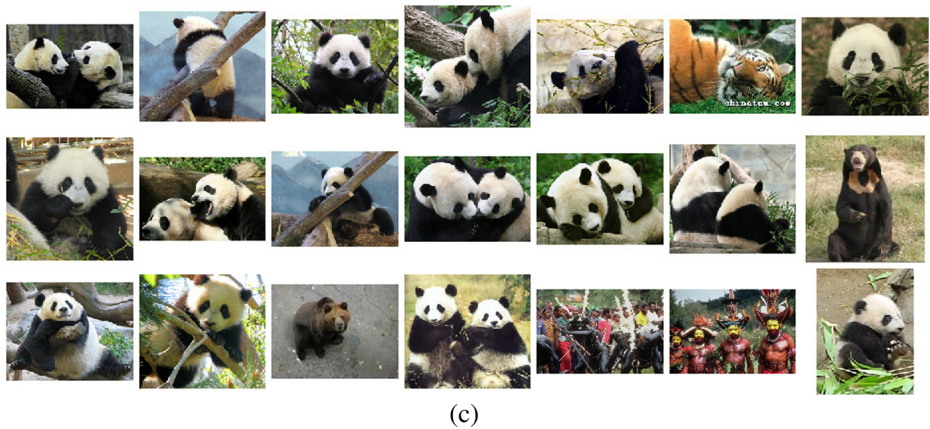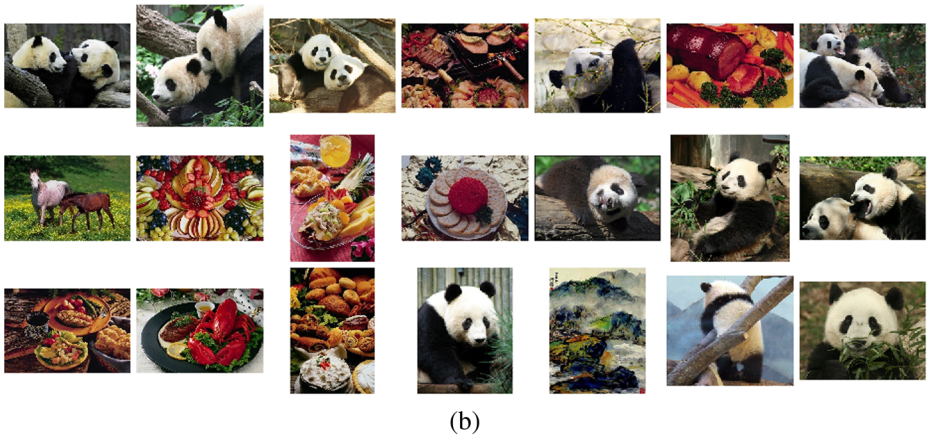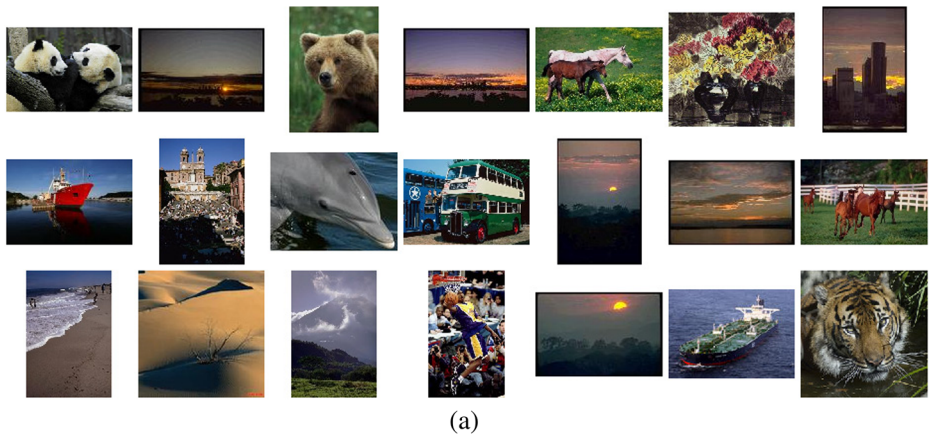Fig. 13 The average retrieval precision for different feature weight values

(a)

(b)

(c)

Fig. 14 Our image retrieval results (Panda): (a) By taking only the EMD, (b) By taking only the LAPH, (c) By taking both EMD and LAPH

In order to improve furtherly the retrieval performance, we can also add relevance feedback to this scheme. The image retrieval with relevance feedback has four main
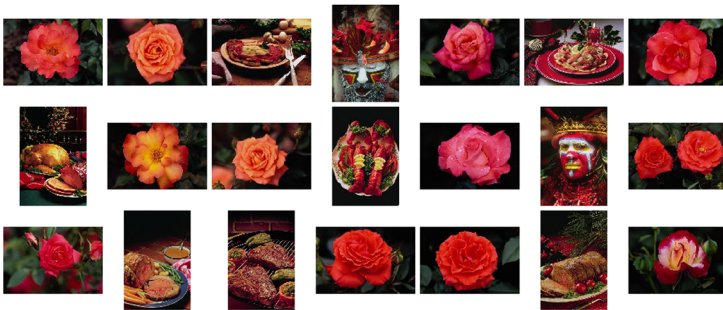
(a)

(b)

(c)

Fig. 15 Our image retrieval results (Horse): (**a**) By taking only the EMD, (**b**) By taking only the LAPH, (**c**) By taking both EMD and LAPH
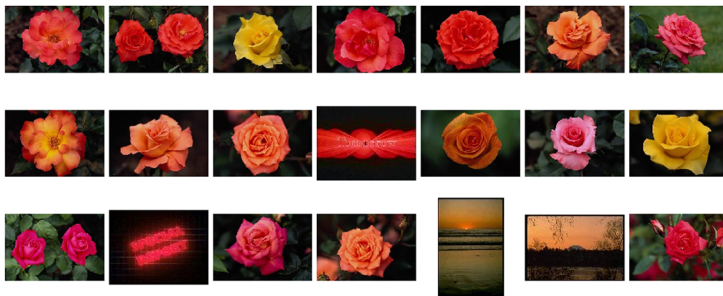
components: query, retrieval, labeling, and learning. When a query is submitted, its low-level visual features (EMD and LAPH) are extracted. Then, all images in the database are
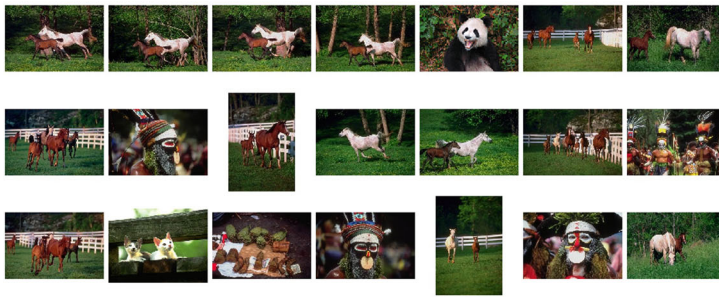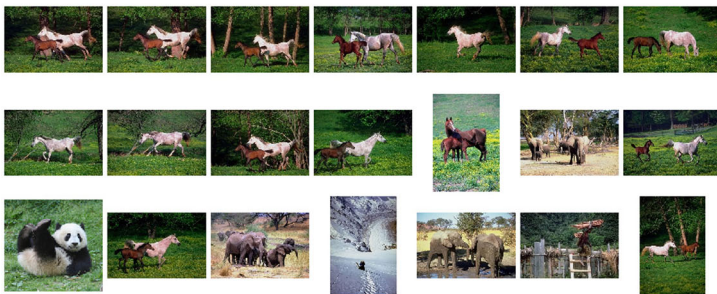
(a)

(b)

(c)

(d)

Fig. 16 The image retrieval results (Flower) using different schemes: (**a**) The retrieval scheme [43], (**b**) The retrieval scheme [7], (**c**) The retrieval scheme [9], (**d**) The proposed retrieval method

(a)

(b)

(c)

(d)

Fig. 17 The image retrieval results (Horse) using different schemes: (a) The retrieval scheme [43], (b) The retrieval scheme [7], (c) The retrieval scheme [9], (d) The proposed retrieval method
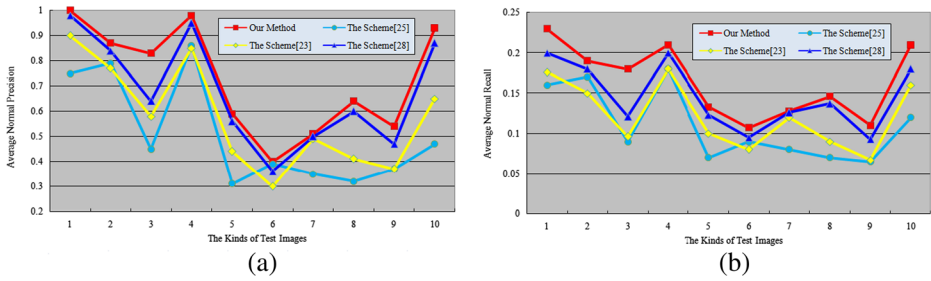
(a)

(b)

(c)

(d)

Fig. 18 The image retrieval results (Bust) using different schemes: (**a**) The retrieval scheme [43], (**b**) The retrieval scheme [7], (**c**) The retrieval scheme [9], (**d**) The proposed retrieval method

Fig. 19 The average retrieval performance of four schemes (COREL dataset): (**a**) The average normal precision, (**b**) The average normal recall

sorted based on a similarity metric. If the user is satisfied with the result, the retrieval process is ended. If the user is not satisfied, he can label some images as positive feedbacks and/or some images as negative feedbacks. Using this feedback process, the system is trained based on machine learning using the embedded relevance feedback algorithm. Then, all the images are re-sorted based on the recalculated similarity metric. If the user is still not content with the result, he repeats the process.

7 Conclusion

CBIR has drawn substantial research attention in the last decade. CBIR usually indexes images by low-level visual features which, though they cannot completely characterize semantic content, are easier to integrate into mathematical formulations. In this paper, we have proposed a content-based image retrieval approach using Exponent moments descriptor and localized angular phase histogram. Experimental results showed that the proposed method yielded higher retrieval accuracy than the other conventional methods with no greater feature vector dimension. In addition, the proposed method almost always showed performance gain in of average normal precision and average normal recall over the other methods. As further studies, the proposed retrieval method is to be evaluated for more various DBs and to be applied to video retrieval.
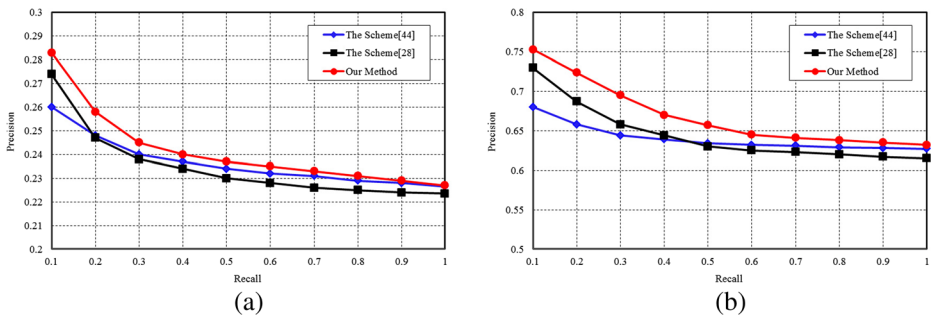

Fig. 20 The average retrieval performance on NUS-WIDE and Oxford Buildings: (**a**) The NUS-WIDE dataset, (**b**) The Oxford Buildings dataset

References

1. Amanatiadis A, Kaburlasos VG, Gasteratos A, Papadakis SE (2011) Evaluation of shape descriptors for shape-based image retrieval. IET Image Process 5(5):493–499
2. Anuar FM, Setchi R, Lai Y (2013) Trademark image retrieval using an integrated shape descriptor. Expert Syst Appl 40(1):105–121
3. Aptoula E (2014) Remote sensing image retrieval with global morphological texture descriptors. IEEE Trans Geoscience Remote Sensing 52(5):3023–3034
4. Aptoula E, Lefèvre S (2009) Morphological description of color images for content-based image retrieval. IEEE Trans Image Process 18(11):2505–2517
5. Atto AM, Berthoumieu Y, Bolon P (2013) 2-D wavelet packet spectrum for texture analysis. IEEE Trans Image Process 22(6):2495–2500
6. Chen WT, Liu WC, Chen MS (2010) Adaptive color feature extraction based on image color distributions. IEEE Trans Image Process 19(8):2005–2016
7. Chun YD, Kim NC, Jang IH (2008) Content-based image retrieval using multiresolution color and texture features. IEEE Trans Multimedia 10(6):1073–1084
8. Datta R, Joshi D, Li J (2008) Image retrieval: ideas, influences, and trends of the new age. ACM Comput Surv 40(2):1–60
9. Farsi H, Mohamadzadeh S (2013) Colour and texture feature-based image retrieval by using hadamard matrix in discrete wavelet transform. IET Image Process 7(3):212–218
10. Gholamreza A, Ali E, George B, Mircea N (2005) Accurate and efficient computation of high order Zernike moments. First Int Symp Visual Comput, Lecture Notes Comput Sci 3804:462–469. doi:10.1007/11595755_56
11. He Z, You X, Yuan Y (2009) Texture image retrieval based on non-tensor product wavelet filter banks. Signal Process 89(8):1501–1510
12. Hong C, Yu J, Tao D (2015) Image-based 3d human pose recovery by multi-view locality sensitive sparse retrieval. EEE Trans Industrial Electronics 62(6):3742–3751
13. Hu W, Xie N, Li L, Zeng X (2011) A survey on visual content-based video indexing and retrieval. IEEE Trans Syst, Man Cybernetics, Part C: Appl Rev 41(6):797–819
14. Jacob IJ, Srinivasagan KG, Jayapriya K (2014) Local oppugnant color texture pattern for image retrieval system. Pattern Recogn Lett 42:72–78
15. Jian M, Lam KM (2014) Face-image retrieval based on singular values and potential-field representation. Signal Process 100:9–15
16. Jun Y, Dongquan L, Dacheng T, Hock Soon S (2012) On combining multiple features for cartoon character retrieval and clip synthesis. IEEE Trans Syst, Man Cybernetics, Part B: Cybernetics 42(5):1413–1427
17. Kashif I, Michael OO, Anne J (2012) Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics. J Comput Syst Sci 78(4):1258–1277
18. Kokare M, Biswas PK, Chatterji BN (2006) Texture image retrieval using new rotated complex wavelet filters. IEEE Trans Syst Man Cybernetics Part B 35(6):1168–78
19. Lasmar NE, Berthoumieu Y (2014) Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms. IEEE Trans Image Process 23(5):2246–2261
20. Li X (2003) Image retrieval based on perceptive weighted color blocks. Pattern Recogn Lett 24(12):1935–1941
21. Li S, Lee MC, Pun CM (2009) Complex Zernike moments features for shape-based image retrieval. IEEE Trans Systems, Man Cybernetics, Part A: Systems Humans 39(1):227–237
22. Li C, Li J, Fu B (2013) Magnitude-phase of quaternion wavelet transform for texture representation using multilevel copula. IEEE Signal Proc Letters 20(8):799–802
23. Liapis S, Tziritas G (2004) Color and texture image retrieval using chromaticity histograms and wavelet frames. IEEE Trans Multimedia 6(5):676–686
24. Lin CH, Chen RT, Chan YK (2009) A smart content-based image retrieval system based on color and texture feature. Image Vis Comput 27(6):658–665
25. Liu M, Vemuri BC, Amari SI, Nielsen F (2012) Shape retrieval using hierarchical total Bregman soft clustering. IEEE Trans Pattern Analysis Machine Intell 34(12):2407–2419
26. Liu GH, Yang JY (2013) Content-based image retrieval using color difference histogram. Pattern Recogn 46(1):188–198
27. Meng M, Ping ZL (2011) Decompose and reconstruct images based on exponential Fourier moments. J Inner Mongolia Normal Univ (Natural Sci Ed) 40(3):258–260

28. Pappas TN, Neuhoff DL, de Ridder H, Zujovic J (2013) Image analysis: focus on texture similarity. Proc IEEE 101(9):2044–2057
29. Park U, Park J, Jain AK (2014) Robust keypoint detection using higher-order scale space derivatives: application to image retrieval. IEEE Signal Proc Letters 21(8):962–965
30. Pooja CS (2011) Improving image retrieval using combined features of Hough transform and Zernike moments. Opt Lasers Eng 49(12):1384–1396
31. Prasad BG, Biswas KK, Gupta SK (2004) Region-based image retrieval using integrated color, shape, and location index. Comput Vis Image Underst 94(1–3):193–233
32. Rakvongthai Y, Oraintara S (2013) Statistical texture retrieval in noise using complex wavelets. Signal Process Image Commun 28(10):1494–1505
33. Saipullah KM, Kim DH (2012) A robust texture feature extraction using the localized angular phase. Multimedia Tools Appl 59(3):717–747
34. Seetharaman K, Jeyakarthic M (2014) Statistical distributional approach for scale and rotation invariant color image retrieval using multivariate parametric tests and orthogonality condition. J Vis Commun Image Represent 25(5):727–739
35. Sherin MY (2012) ICTEDCT-CBIR: Integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval. Comput Electrical Eng 38(5):1358–1376
36. Shu X, Xiao-Jun W (2011) A novel contour descriptor for 2D shape matching and its application to image retrieval. Image Vis Comput 29(4):286–294
37. Singha M, Hemachandran K, Paul A (2012) Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram. IET Image Process 6(9):1221–1229
38. Susana A, Annaa S, Maria V, Xavier O (2012) Low-dimensional and comprehensive color texture description. Comput Vis Image Underst 116(1):54–67
39. Talib A, Mahmuddin M, Husni H (2013) A weighted dominant color descriptor for content-based image retrieval. J Vis Commun Image Represent 24(3):345–360
40. Van De Sande KEA, Gevers T, Snoek CGM (2010) Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Analysis Machine Intell 32(9):1582–1596
41. J. Wan, D. Wang, S. C. Hoi (2014) Deep learning for content-based image retrieval: a comprehensive study. Proceedings of the ACM International Conference on Multimedia. Orlando, FL, USA, 2014: 157–166
42. Wang XY, Yu YJ, Yang HY (2011) An effective image retrieval scheme using color, texture and shape features. Comput Standards Interfaces 33(1):59–68
43. Yap PT, Paramesran R (2006) Content-based image retrieval using Legendre chromaticity distribution moments. IEE Proc-Vision, Image Signal Proc 153(1):17–24
44. Zhang D, Lu G (2004) Review of shape representation and description techniques. Pattern Recogn 37(1):1–19

Xiang-Yang Wang is currently a professor with the Multimedia and Information Security Laboratory, School of Computer and Information Technology, Liaoning Normal University, China. His research interests lie in the areas of information security, image processing, pattern recognition, and computer vision. He is the author of two books. He has published over 80 papers in international journals (including IEEE/ACM Transactions) and 25 papers in international conferences and workshops. Mr. Wang is a Reviewer for many leading international and national journals and conferences, including IEEE/ACM Transactions.

Lin-Lin Liang received the B. E. degree from the School of Computer and Information Technology, Liaoning Normal University, China, in 2015, where she is currently pursuing the M. S. E. degree. Her research interests include image retrieval and signal processing.

Yong-Wei Li received the B. E. degree from the School of Computer and Information Technology, Liaoning Normal University, China, in 2012, where he is currently pursuing the M. S. E. degree. His research interests include image retrieval and signal processing.

Hong-Ying Yang is currently a professor with the School of Computer and Information Technology at the Liaoning Normal University, China. Her research interests include signal processing and communications, digital multimedia data hiding.