Computer Standards & Interfaces 33 (2011) 59–68

Contents lists available at ScienceDirect

Computer Standards & Interfaces

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

An effective image retrieval scheme using color, texture and shape features

Xiang-Yang Wang [a,b,*], Yong-Jian Yu [a], Hong-Ying Yang [a]

[a] School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China
[b] State Key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications), Beijing 100876, China

ARTICLE INFO

Article history:
Received 19 March 2009
Accepted 9 March 2010
Available online 20 March 2010

Keywords:
Image retrieval
Dynamic dominant color
Steerable filter
Pseudo-Zernike moments

ABSTRACT

In this paper, we present a new and effective color image retrieval scheme for combining all the three i.e. color, texture and shape information, which achieved higher retrieval efficiency. Firstly, the image is predetermined by using fast color quantization algorithm with clusters merging, and then a small number of dominant colors and their percentages can be obtained. Secondly, the spatial texture features are extracted using a steerable filter decomposition, which offers an efficient and flexible approximation of early processing in the human visual system. Thirdly, the pseudo-Zernike moments of an image are used for shape descriptor, which have better features representation capabilities and are more robust to noise than other moment representations. Finally, the combination of the color, texture and shape features provide a robust feature set for image retrieval. Experimental results show that the proposed color image retrieval is more accurate and efficient in retrieving the user-interested images.

1. Introduction

Nowadays, with increased digital images available on Internet, efficient indexing and searching becomes essential for large image archives. Traditional annotation heavily relies on manual labor to label images with keywords, which unfortunately can hardly describe the diversity and ambiguity of image contents. Hence, content-based image retrieval (CBIR) [1] has drawn substantial research attention in the last decade. CBIR usually indexes images by low-level visual features such as color, texture and shape. The visual features cannot completely characterize semantic content, but they are easier to integrate into mathematical formulations [2]. Extraction of good visual features which compactly represent a query image is one of the important tasks in CBIR.

Color is one of the most widely used low-level visual features and is invariant to image size and orientation [1]. As conventional color features used in CBIR, there are color histogram, color correlogram, and dominant color descriptor (DCD). Color histogram is the most commonly used color representation, but it does not include any spatial information. Li et al. [3] 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 retrieve images under the query of special object. Color correlogram describes the probability of finding color pairs at a fixed pixel distance and provides spatial information. Therefore color correlogram yields better retrieval accuracy in comparison to color histogram. Color autocorrelogram is a subset of color correlogram, which captures the spatial correlation between identical colors only. Since it provides significant computational benefits over color correlogram, it is more suitable for image retrieval. DCD is MPEG-7 color descriptors [4]. DCD describes the salient color distributions in an image or a region of interest, and provides an effective, compact, and intuitive representation of colors presented in an image. However, DCD similarity matching does not fit human perception very well, and it will cause incorrect ranks for images with similar color distribution [5,6]. In [7], Yang et al. presented a color quantization method for dominant color extraction, called the linear block algorithm (LBA), and it has been shown that LBA is efficient in color quantization and computation. For the purpose of effectively retrieving more similar images from the digital image databases (DBs), Lu et al. [8] uses the color distributions, the mean value and the standard deviation, to represent the global characteristics of the image, and the image bitmap is used to represent the local characteristics of the image for increasing the accuracy of the retrieval system.

Texture is also an important visual feature that refers to innate surface properties of an object and their relationship to the surrounding environment. Many objects in an image can be distinguished solely by their textures without any other information. There is no universal definition of texture. Texture may consist of some basic primitives, and may also describe the structural arrangement of a region and the relationship of the surrounding regions [9]. In conventional texture features used for CBIR, there are statistic texture features using gray-level co-occurrence matrix (GLCM), Markov random field (MRF) model, simultaneous auto-regressive

* Corresponding author. School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China. Tel.: +86 0411 85992415; fax: +86 0411 85992323.
E-mail address: wxy37@126.com (X.-Y. Wang).

0920-5489/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.csi.2010.03.004

(SAR) model, Wold decomposition model, edge histogram descriptor (EHD), and wavelet moments [2]. Recently, BDIP (block difference of inverse probabilities) and BVLC (block variation of local correlation coefficients) features have been proposed which effectively measure local brightness variations and local texture smoothness, respectively [10]. These features are shown to yield better retrieval accuracy over the compared conventional features. In [11], texture be modeled by the fusion of marginal densities of subband image DCT coefficients. Following this approach, one can extract samples from the texture distribution by utilizing small neighborhoods of scale-to-scale coefficients. Components of the multivariate texture-distributional vectors are formed using the spatially localized coefficients, at different image decomposition levels. Kokare et al. [12] designed a new set of 2D rotated wavelet by using Daubechies' eight tap coefficients to improve the image retrieval accuracy. The 2D rotated wavelet filters that are non-separable and oriented, improves characterization of diagonally oriented textures. Han et al. [13] proposed a rotation-invariant and a scale-invariant Gabor representations, where each representation only requires few summations on the conventional Gabor filter impulse responses, and the texture features are then extracted from these new representations for conducting rotation-invariant or scale-invariant texture image retrieval.

Object shape features can also provide powerful information for image retrieval, because humans can recognize objects solely from their shapes. Usually, the shape carries semantic information [14], and shape features are different from other elementary visual features, such as color or texture features. Basically, shape features can be categorized as boundary-based and region-based. The former extracts features based on the outer boundary of the region while the latter extracts features based on the entire region [2]. Shape matching is a well-explored research area with many shape representation and similarity measurement techniques found in the literature [1]. Shape representation methods include Fourier descriptors, polygonal approximation, invariant moments, B-splines, deformable templates, and curvature scale space (CSS) [2]. Most of these techniques were developed for whole shape matching, i.e., closed planar curve matching. The CSS shape representation method has been selected for moving picture experts group (MPEG)-7 standardization [14]. However, based on the curvature zero-crossing, the CSS method is more suitable for shapes with distinct curvature variations, such as leaf shapes than for smooth shapes with subtle curvature variations. Fourier descriptor has proven to be more efficient and robust than is the CSS in a review of shape representation and description techniques [15]. But, as mentioned in [15], Fourier descriptor was not suitable for partial shape matching. Xu et al. [16] presented an innovative partial shape matching (PSM) technique using dynamic programming (DP) for the retrieval of spine X-ray images. In [17], Wei proposed a novel content-based trademark retrieval system with a feasible set of feature descriptors, which is capable of depicting global shapes and interior/local features of the trademarks.

Most of the early studies on CBIR have used only a single feature among various color, texture, and shape 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 visual features have been performed [18–20]. In [18], 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 [19], the color histogram is used as a color feature and Haar or Daubechies' wavelet moment is used as a texture feature. In these methods, their feature vector dimension is not considered as an important factor in combining multiple features. It is shown that such a combination of features without increase of feature vector dimension does not always guarantee better retrieval accuracy [20]. Chun et al. [9] proposed a CBIR method which uses the combination of color autocorrelograms of hue and saturation component images and BDIP and BVLC moments of value component image in the wavelet transform domain. In [21,22], Hiremath presents a novel retrieval framework for combining color, texture and shape information. The image is partitioned into non- overlapping tiles of equal size. The color moments and moments on Gabor filter responses of these tiles serve as local descriptors of color and texture respectively. Shape information is captured in terms of edge images computed using gradient vector flow fields. Invariant moments are then used to record the shape features. Accordingly, for an advanced CBIR, it is necessary to choose efficient features that are complementary to each other so as to yield an improved retrieval performance and to combine chosen features effectively without increase of feature vector dimension.

In this paper, we propose 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. The rest of this paper is organized as follows. Section 2 presents dynamic dominant color extraction. Section 3 describes the steerable filter decomposition and texture representation. In Section 4, the pseudo-Zernike moments based shape descriptor is given. Section 5 contains the description of similarity measure for image retrieval. Simulation results in Section 6 will show the performance of our scheme. Finally, Section 7 concludes this presentation.

2. Color feature representation

In general, color is one of the most dominant and distinguishable low-level visual features in describing image. Many CBIR systems employ color to retrieve images, such as QBIC system and VisualSEEK. In theory, it will lead to minimum error by extracting color feature for retrieval using real color image directly, but the problem is that the computation cost and storage required will expand rapidly. So it goes against practical application. In fact, for a given color image, the number of actual colors only occupies a small proportion of the total number of colors in the whole color space, and further observation shows that some dominant colors cover a majority of pixels. Consequently, it won't influence the understanding of image content though reducing the quality of image if we use these dominant colors to represent image.

In the MPEG-7 Final Committee Draft, several color descriptors have been approved including number of histogram descriptors and a dominant color descriptor (DCD) [4]. DCD contains two main components: representative colors and the percentage of each color. DCD can provide an effective, compact, and intuitive salient color representation, and describe the color distribution in an image or a region of interesting. But, for the DCD in MPEG-7, the representative colors depend on the color distribution, and the greater part of representative colors will be located in the higher color distribution range with smaller color distance. It is may be not consistent with human perception because human eyes cannot exactly distinguish the colors with close distance. Moreover, DCD similarity matching does not fit human perception very well, and it will cause incorrect ranks for images with similar color distribution. We will adopt a new and efficient dominant color extraction scheme to address the above problems [7].

According to numerous experiments, the selection of color space is not a critical issue for DCD extraction. Therefore, for simplicity and without loss of generality, the RGB color space is used. Firstly, the RGB color space is uniformly divided into 8 coarse partitions, as shown in Fig. 1. If there are several colors located on the same partitioned block, they are assumed to be similar.

After the above coarse partition, the centroid of each partition ("colorBin" in MPEG-7) is selected as its quantized color. Let $X=(X^R, X^G, X^B)$ represent color components of a pixel with color components Red, Green, and Blue, and C_i be the quantized color for partition i. The

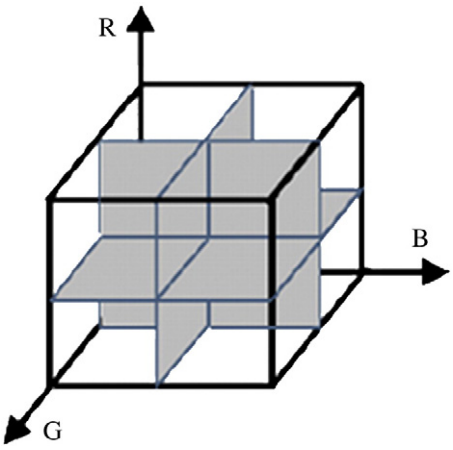


Fig. 1. The coarse division of RGB color space.

average value of color distribution for each partition center can be calculated by

$$\overline{X}_i = \frac{\sum\limits_{X \in C_i} X}{\sum\limits_{X \in C_i} 1} \tag{1}$$

After the average values are obtained, each quantized color

$$C_i = \left(\overline{X}_i^{\mathrm{R}}, \overline{X}_i^{\mathrm{G}}, \overline{X}_i^{\mathrm{B}}\right)(1 \le i \le 8)$$

can be determined. Consequently, we calculate mutual distance of two adjacent C_i, and then merge similar "colorBins" using weighted average agglomerative procedure in the following equation.

$$\begin{cases} X^{\mathrm{R}} = X_1^{\mathrm{R}} \times \left(\dfrac{P_{\mathrm{R},1}}{P_{\mathrm{R},1} + P_{\mathrm{R},2}}\right) + X_2^{\mathrm{R}} \times \left(\dfrac{P_{\mathrm{R},2}}{P_{\mathrm{R},1} + P_{\mathrm{R},2}}\right) \\ X^{\mathrm{G}} = X_1^{\mathrm{G}} \times \left(\dfrac{P_{\mathrm{G},1}}{P_{\mathrm{G},1} + P_{\mathrm{G},2}}\right) + X_2^{\mathrm{G}} \times \left(\dfrac{P_{\mathrm{G},2}}{P_{\mathrm{G},1} + P_{\mathrm{G},2}}\right) \\ X^{\mathrm{B}} = X_1^{\mathrm{B}} \times \left(\dfrac{P_{\mathrm{B},1}}{P_{\mathrm{B},1} + P_{\mathrm{B},2}}\right) + X_2^{\mathrm{B}} \times \left(\dfrac{P_{\mathrm{B},2}}{P_{\mathrm{B},1} + P_{\mathrm{B},2}}\right) \end{cases} \tag{2}$$

In the above Eq. (2), $P_{\mathrm{R}}, P_{\mathrm{G}}, P_{\mathrm{B}}$ represent the percentages in R, G, B components, respectively. The merge processes iterated until the minimum Euclidian distance between the adjacent color cluster centers being larger than the threshold T_{d}. Because the dominant colors should be significant enough, therefore we will merge the insignificant color into its neighboring color. We check each survived color, if its percentage is less than threshold T_{m}, it will be merged into the nearest color. In our paper, we set the T_{d} as 25 and T_{m} as 6%. As a result, we obtain a set of dominant colors, and the final number of dominant colors is constrained to 4–5 on average. The dominant color descriptor is defined as

$$F_C = \{(C_i, P_i), i = 1, \ldots, N\}$$

where N is the total number of dominant colors for an image, C_i is a 3D dominant color vector, P_i is the percentage for each dominant color, and the sum of P_i is equal to 1. As shown in Fig. 2, the final number of dominant colors of the image after color quantized is five, the dominant colors and their percentages are shown in Table 1.

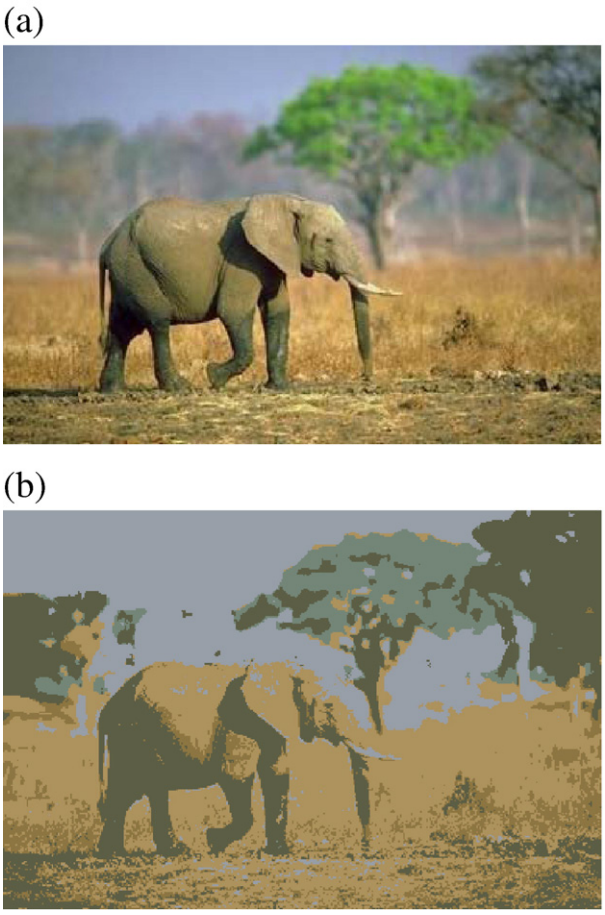


Fig. 2. The original image and its quantized image. (a) The original image; (b) the corresponding quantized image.

3. Steerable filter decomposition and texture representation

Most natural surfaces exhibit texture, which is an important low-level visual feature. Texture recognition will therefore be a natural part of many computer vision systems. In this paper, we propose a new rotation-invariant and scale-invariant texture representation for image retrieval based on steerable filter decomposition.

3.1. Steerable Filter

Oriented filters are used in many vision and image processing tasks, such as texture analysis, edge detection, image data compression, motion analysis, and image enhancement. In many of these tasks, it is useful to apply filters of arbitrary orientation under adaptive control and to examine the filter output as a function of both orientation and phase. One approach to finding the response of a filter at many orientations is to apply many versions of the same filter, each of which is different from the others by some small rotation in angle.

Steerable filter is a class of filters in which a filter of arbitrary orientation is synthesized as a linear combination of a set of "basis

Table 1
The dominant color feature of Fig. 2(a).

N	C_i (R, G, B)	P_i
1	92, 94, 69	0.2642
2	115, 134, 119	0.1041
3	142, 118, 66	0.1035
4	173, 146, 94	0.3179
5	152, 158, 168	0.2103

filters" [23]. The edge located at different orientations in an image can be detected by splitting the image into orientation sub-bands obtained by the basis filters having these orientations. It allows one to adaptively "steer" a filter to any orientation, and to determine analytically the filter output as a function of orientation.

The steering constraint is

$$F_\theta(m,n) = \sum_{k=1}^{N} b_k(\theta) A_k(m,n) \tag{3}$$

where $b_k(\theta)$ is the interpolation function based on the arbitrary orientation θ which controls the filter orientations. And the basis filters $A_k(m,n)$ are rotated version of impulse response at θ.

Because convolution is a linear operation, we can synthesize an image filtered at an arbitrary orientation by taking linear combination of the image filtered with the basis filters. Let * represent convolution with image $I(m,n)$.

$$I(m,n) * F_\theta(m,n) = \sum_{k=1}^{N} b_k(\theta)(I(m,n)) * A_k(m,n)) \tag{4}$$

The structure of steerable filter is illustrated in Fig. 3.

In general, we usually adopt 4–6 oriented sub-bands to extract image texture information when applying steerable filter to analyze digital images. In this paper, we extract the texture feature from four oriented sub-bands, as shown in Fig. 4. Here, Fig. 4(a) represents the original image (Zoneplate), and Fig. 4(b)–(e) are the filtered image in different orientations.

Roughly speaking, the texture image can be seen as a set of basic repetitive primitives characterized by their spatial homogeneity. By applying statistical measures, this information is extracted, and used to capture the relevant image content into feature vectors. More precisely, we use the mean μ and standard deviation σ of the energy distribution of the filtered images $S_i(x,y)(i=1,2,3,4$ represent horizontal orientation, rotation of 45°, vertical orientation, and rotation of −45°, respectively), by considering the presence of homogeneous regions in texture images. Given an image $I(x,y)$, its steerable filter decomposition is defined as:

$$S_i(x,y) = \sum_{x_1} \sum_{y_1} I\left(x_1,y_1\right) B_i(x-x_1, y-y_1) \tag{5}$$

where B_i denotes the directional bandpass filters at orientation $i=1,2,3,4$.

The energy distribution $E_i(x,y)$ of the filtered images $S_i(x,y)$ is defined as

$$E_i = \sum_x \sum_y |S_i(x,y)| \tag{6}$$

Additionally, the mean (μ_i) and standard deviation (σ_i) are found as follows:

$$\mu_i = \frac{1}{MN} E_i(x,y) \tag{7}$$

$$\sigma_i = \sqrt{\frac{1}{MN} \sum_x \sum_y (S_i(x,y) - \mu_i)^2} \tag{8}$$

where M, and N is the width and height of image $I(x,y)$ respectively.

So, the corresponding texture feature vector of the original image $I(x,y)$ should be defined as:

$$F_T = (\mu_1, \sigma_1, \mu_2, \sigma_2, \mu_3, \sigma_3, \mu_4, \sigma_4) \tag{9}$$

4. The pseudo-Zernike moments based shape descriptor

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 feature for content-based image retrieval can be seen from the fact that every major content-based image retrieval (CBIR) system [1] incorporates some shape features in one form or another.

Image shape descriptors, as used in existing CBIR systems, can be broadly categorized into two groups, namely, contour- and region-based descriptors [2]. Contour-based shape descriptors use the boundary information of object shapes. Early work implemented object shapes via Fourier descriptors [1,2]. Exploiting only information from the shape boundaries, contour-based shape descriptors thereby ignore potentially important information in the shape interior. In region-based methods, shape descriptors utilize information from both boundaries and interior regions of the shape. As the most commonly used approaches for region-based shape descriptors, moments and function of moments have been utilized as pattern features in a number of applications [1,2]. The theory of moments, including Hu moments, Legendre moments, wavelet moments, Zernike moments, and Krawtchouk moments, provides useful series expansions for the representation of object shapes.

Pseudo-Zernike moments consist of a set of orthogonal and complex number moments [24] which have some very important properties. First, the pseudo-Zernike moments' magnitudes are

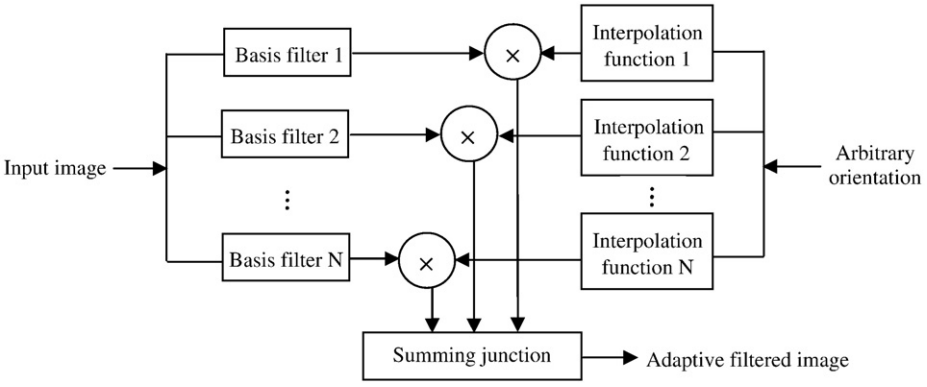


Fig. 3. The structure of steerable filter.

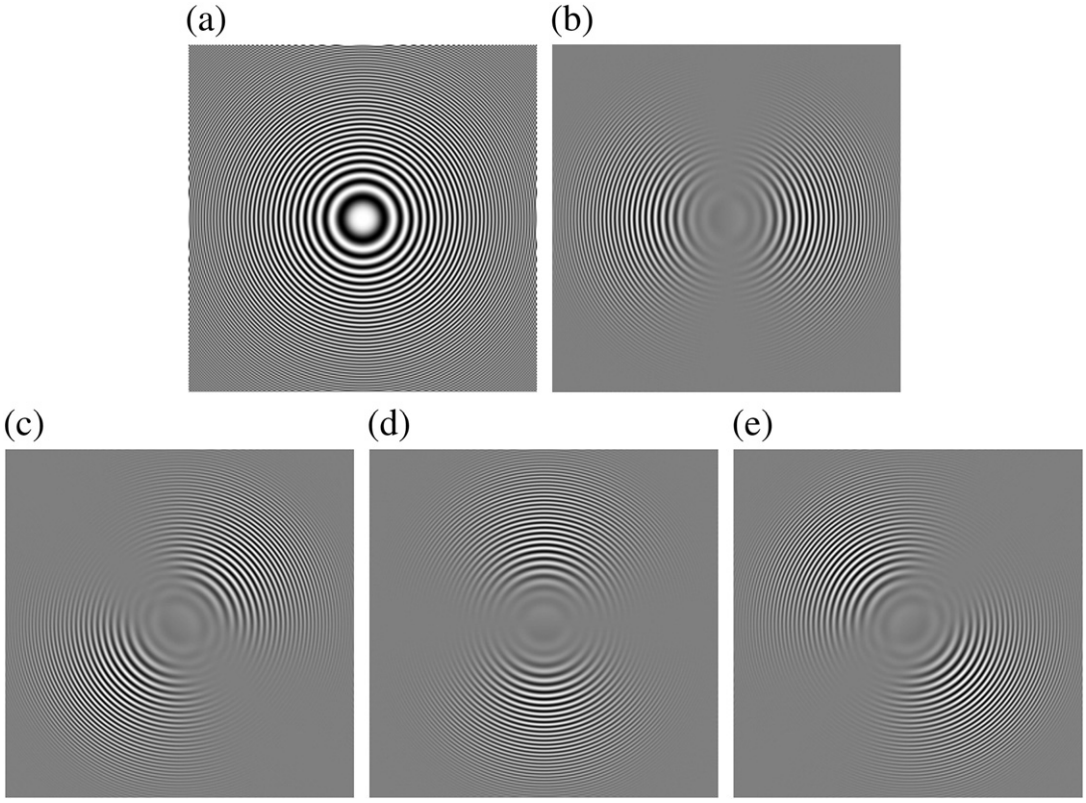


Fig. 4. The filtered images using steerable filter. (a) The original image; (b) the subband image in horizontal orientation; (c) the subband image for rotation of 45°; (d) the subband image in vertical orientation; (e) the subband image for rotation of −45°.

invariant under image rotation. Second, pseudo-Zernike moments have multilevel representation capabilities. Third, pseudo-Zernike moments are less sensitive to image noise. In this paper, the pseudo-Zernike moments of an image are used for shape descriptor, which have better features representation capabilities and are more robust to noise than other moment representations.

4.1. Pseudo-Zernike moments

Pseudo-Zernike moments consist of a set of complex polynomials [24] that form a complete orthogonal set over the interior of the unit circle, $x^2+y^2\le 1$. If the set of these polynomials is denoted by $\{V_{nm}(x,y)\}$, then the form of these polynomials is as follows

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)\exp(jm\theta) \tag{10}$$

where $\rho = \sqrt{x^2+y^2}$, $\theta = \tan^{-1}(y/x)$. Here n is a non-negative integer, m is restricted to be $|m|\le n$ and the radial pseudo-Zernike polynomial $R_{nm}(\rho)$ is defined as the following

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|} \frac{(-1)^s(2n+1-s)!\rho^{n-s}}{s!(n+|m|+1-s)!(n-|m|-s)!} \tag{11}$$

Like any other orthogonal and complete basis, the pseudo-Zernike polynomial can be used to decompose an analog image function $f(x,y)$:

$$f(x,y) = \sum_{n=0}^{\infty} \sum_{\{m:|m|\le n\}} A_{nm}V_{nm}(x,y) \tag{12}$$

where A_{nm} is the pseudo-Zernike moments of order n with repetition m, whose definition is

$$A_{nm} = \frac{n+1}{\pi}\iint_{x^2+y^2\le 1} f(x,y)V^*_{nm}(x,y)dxdy \tag{13}$$

It should be pointed out that in case of digital images, Eq. (13) cannot be applied directly, but rather, its approximate version has to be employed. For instance, given a digital image of size $M\times N$, its pseudo-Zernike moments are computed as

$$\hat{A}_{nm} = \frac{n+1}{\pi}\sum_{i=1}^{M}\sum_{j=1}^{N} h_{nm}\left(x_i,y_j\right)f\left(x_i,y_j\right) \tag{14}$$

where the value of i and j are taken such that $x_i^2+y_j^2\le 1$, and

$$h_{nm}\left(x_i,y_j\right) = \int_{x_i-\frac{\Delta x}{2}}^{x_i+\frac{\Delta x}{2}}\int_{y_j-\frac{\Delta y}{2}}^{y_j+\frac{\Delta y}{2}} V^*_{nm}(x,y)dxdy \tag{15}$$

where $\Delta x = \frac{2}{M}, \Delta y = \frac{2}{N}$. $h_{nm}(x_i,y_j)$ can be computed to address the nontrivial issue of accuracy. In this research, we adopt the following formulas which are most commonly used in literature to compute pseudo-Zernike moments of discrete images:

$$\hat{A}_{nm} = \frac{n+1}{\pi}\sum_{i=1}^{M}\sum_{j=1}^{N} V^*_{nm}\left(x_i,y_j\right)f\left(x_i,y_j\right)\Delta x\Delta y \tag{16}$$

4.2. The advantages of pseudo-Zernike moments

The pseudo-Zernike moments have the following advantages.

4.2.1. The invariance properties

The reason we use pseudo-Zernike moments for shape descriptor is that they have some very important properties, i.e., their magnitudes are invariant under image rotation and image flipping. We now elaborate on these invariance properties.

If image $f(r,\theta)$ is rotated α degrees counterclockwise, the image under rotation is $f'(r,\theta)=f(r,\theta-\alpha)$, and the pseudo-Zernike moments of $f'(r,\theta)$ can be expressed by

$$A'_{nm}=\frac{n+1}{\pi}\int_{-\pi}^{\pi}\int_0^1 f(r,\theta-\alpha)\left[R_{nm}(r)e^{-jm\theta}\right]^* r dr d\theta \tag{17}$$

Assume $\theta'=\theta-\alpha$, we can get

$$\begin{aligned}A'_{nm}&=\frac{n+1}{\pi}\int_{-\pi}^{\pi}\int_0^1 f\left(r,\theta'\right)\left[R_{nm}(r)e^{-jm\left(\theta'+\alpha\right)}\right]^* r dr d\theta' \\ &=\left[\frac{n+1}{\pi}\int_{-\pi}^{\pi}\int_0^1 f\left(r,\theta'\right)\right]\left[R_{nm}(r)e^{-jm\theta'}\right]e^{-jm\alpha}=A_{nm}e^{-jm\alpha}\end{aligned} \tag{18}$$

It can be shown that the pseudo-Zernike moments of the resulting image are $A'_{nm}=A_{nm}\exp(-jm\alpha)$, which leads to $|A'_{nm}|=|A_{nm}|$. Therefore, it is robust to rotation.

It can also be shown that if an image $f(r,\theta)$ is flipped horizontally, the pseudo-Zernike moments of the resulting image are

$A_{nm}^{(hf)}=A_{nm}^*$, if m is even;

$A_{nm}^{(hf)}=-A_{nm}^*$, if m is odd.

Similarly, if an image $f(r,\theta)$ is flipped vertically, the pseudo-Zernike moments of the result image $A_{nm}^{(vf)}=A_{nm}^*$ for m even or odd. In either case, the magnitudes of the pseudo-Zernike moments do not change, indicating that the pseudo-Zernike moments has perfect robustness to image flipping attacks.

4.2.2. Good robustness against noise

Pseudo-Zernike moments have been proven to be superior to other moment functions such as Zernike moments in terms of their feature representation capabilities. Pseudo-Zernike moments can offer more feature vectors than Zernike moments. There are $(n+1)^2$ linearly independent pseudo-Zernike polynomials of $order \le n$ as compared to $(n+1)(n+2)/2$ of Zernike polynomials due to additional constraint of $(n-m)=even$. Under the same order, pseudo-Zernike moments have more low-order moments than Zernike moments, therefore pseudo-Zernike moments are less sensitive to image noise than the conventional Zernike moments.

4.2.3. Orthogonal property

Pseudo-Zernike moments are orthogonal. Orthogonal moments have been proven to be more robust in the presence of noise, and they are able to achieve a near zero value of redundancy measure in a set of moment functions.

4.2.4. Multilayer expression

For an image, the low-order moments of the pseudo-Zernike moments can express the outline of the image; and the high-order moments of the pseudo-Zernike moments can express the detail of the image.

Fig. 5 shows the binary image H and its reconstructed image. From left to right, the orders are 5, 10, 15, 20, and 25, respectively. It is not difficult to see that the low-order pseudo-Zernike moments of digital image contain contour, but the high-order ones give the detail.

4.2.5. It is convenient to implement

The higher order pseudo-Zernike moments can be constructed arbitrarily, not like the Hu moments.

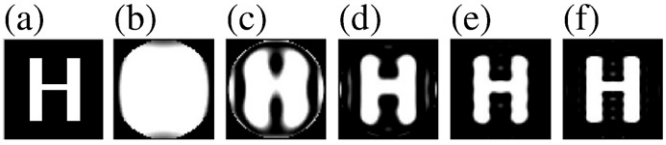


Fig. 5. The binary image H and its reconstructed image. (a) Original image; (b) the reconstructed image (order is 5); (c) the reconstructed image (order is 10); (d) the reconstructed image (order is 15); (e) the reconstructed image (order is 20); (f) the reconstructed image (order is 25).

On the whole, pseudo-Zernike moment is a good figure description operator, though it hasn't scaling and translation invariance, but the scaling and translation invariance can be achieved by preprocessing the image to a standard image via image normalization technology.

4.3. Shape feature representation

Pseudo-Zernike moments are not scale or translation invariant. In our work, the scaling and translation invariance are firstly obtained by normalizing the image, and then $|\hat{A}_{nm}|$ is selected as shape feature set for image retrieval.

The pseudo-Zernike moments based shape feature vector is given by:

$$F_S=\left(|\hat{A}_{00}|,|\hat{A}_{10}|,|\hat{A}_{11}|,|\hat{A}_{20}|,\cdots,|\hat{A}|_{54},|\hat{A}|_{55}\right) \tag{19}$$

5. Similarity measure

After the color, texture and shape feature vectors 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.

5.1. Color feature similarity measure

MPEG-7 DCD similarity matching does not fit human perception very well, and it will cause incorrect ranks for images with similar color distribution [7], so we adopt a modification on distance function to improve the robustness of color feature similarity measure.

Considering that the color feature of query image Q is $F_C^Q=\{(C_i,P_i), i=1,...,N_Q\}$ and the color feature of each target image I in an image DB is $F_C^I=\{(C_i,P_i), i=1,...,N_I\}$, we define a new color feature similarity as follows.

$$S_{\text{Color}}(Q,I)=\sum_{i=1}^{N_Q}\sum_{j=1}^{N_I} d_{i,j}S_{i,j} \tag{20}$$

where N_Q and N_I denote the number of the dominant colors of the query image Q and the target image I respectively; $d_{i,j}=\|C_i^Q-C_j^I\|$ denotes the Euclidean distance between the dominant color C_i^Q of query image Q and the dominant color C_i^I of target image I; $S_{i,j}=[1-|p_i^Q-p_j^I|]\times\min(p_i^Q,p_j^I)$ denotes the similarity score between two different dominant colors. Here, p_i^Q and p_j^I are the percentages of the ith and jth dominant color in query image and target image, respectively. The $\min(p_i^Q,p_j^I)$ is the intersection of p_i^Q and p_i^Q, which represents the similarity between two colors in percentage. The term in bracket, $1-|p_i^Q-p_j^I|$, is used to measure the difference of two colors in their percentage. If p_i^Q equals to p_i^Q, then their percentage is same and the color similarity is determined by $\min(p_i^Q,p_j^I)$; otherwise, a large difference between p_i^Q and p_i^Q will decrease the similarity measure.

5.2. Texture feature similarity measure

The texture feature similarity is given by

$$S_{\text{Texture}}(Q,I)=\left(\sum_{i=1}^{4}\left[\left(\mu_i^Q-\mu_i^I\right)^2+\left(\sigma_i^Q-\sigma_i^I\right)^2\right]\right)^{1/2} \tag{21}$$

where μ_i^Q and σ_i^Q denote the texture feature of query image Q, μ_i^I and σ_i^I denote the texture feature of target image I.

5.3. Shape feature similarity measure

We give the shape feature similarity as follows.

$$S_{\text{Shape}}(Q,I) = \left(\sum_{i=0}^{5} \sum_{j}^{i} \left(|\hat{A}_{ij}^{Q}| - |\hat{A}_{ij}^{I}| \right)^2 \right)^{1/2} \tag{22}$$

where $|\hat{A}_{ij}^{Q}|$ and $|\hat{A}_{ij}^{I}|$ denote the shape feature of the query image Q and the target image I respectively.

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

$$S(I,Q) = w_C S_{\text{Color}}(Q,I) + w_T S_{\text{Texture}}(Q,I) + w_S S_{\text{Shape}}(Q,I)\; w_C + w_T + w_S = 1$$

where w_C, w_T, w_S is the weight of the color, texture, and shape features respectively.

It should be pointed out that $S(Q,I)$ is typically not equivalent to $S(I,Q)$ according to the formula above, so we define the final similarity as:

$$S'(Q,I) = \frac{S(I,Q) + S(Q,I)}{2} \tag{23}$$

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. Experimental results

In this paper, we propose a new and effective color image retrieval scheme for combining all the three i.e. color, texture and shape information, which achieve higher retrieval efficiency. Fig. 6. shows the block diagram of the proposed retrieval method.

The color image retrieval systems have been implemented in Visual C++ 6.0 environment on a Pentium 4 (2 GHz) PC. To check the retrieval efficiency of the proposed method, we have tested the performance with a general purpose image database that consists of 11,000 images of 110 categories from the Corel Image Gallery. Corel images have been widely used by the image processing and CBIR research communities. They cover a variety of topics, such as “flower”, “bus”, “eagle”, “gun”, “sunset”, “horse”, etc. To evaluate the overall performance of the proposed image feature in retrieval, a number of experiments were performed on our image retrieval and the state-of-the-arts, and the methods to be compared include traditional color histogram and the color histogram of subblocks [3]. The parameter settings are:

$$T = 25, \qquad \delta = 0.06, \qquad w_C = w_T = w_S = {}^1/_3$$

Figs. 7 and 8 show the image retrieval results using the traditional color histogram, the color histogram of subblocks [3], 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 50 images as query images from the above image database (the tested 10 semantic class includes bus, horse, flower, dinosaur, building, elephant, people, beach, scenery, and dish). Each kind is extracted 5 images, and each time returns the first 20 most similar images as retrieval results. To each kind of image, the average normal precision, and the average normal recall of 5 times query results are calculated. These values are taken as the retrieval performance standard of the algorithm, as shown in Fig. 9.

7. Conclusion

CBIR is an active research topic in image processing, pattern recognition, and computer vision. In this paper, a CBIR method has been proposed which uses the combination of dynamic dominant color, Steerable filter texture feature, and pseudo-Zernike moments shape descriptor. Experimental results showed that the proposed method yielded higher retrieval accuracy than the other conventional methods with no greater feature vector dimension. It was all the more so for multiresolution image DBs. In addition, the proposed method almost always showed performance gain of average normal precision, average normal recall, and average retrieval time 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.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant Nos. 60773031 and 60873222, the

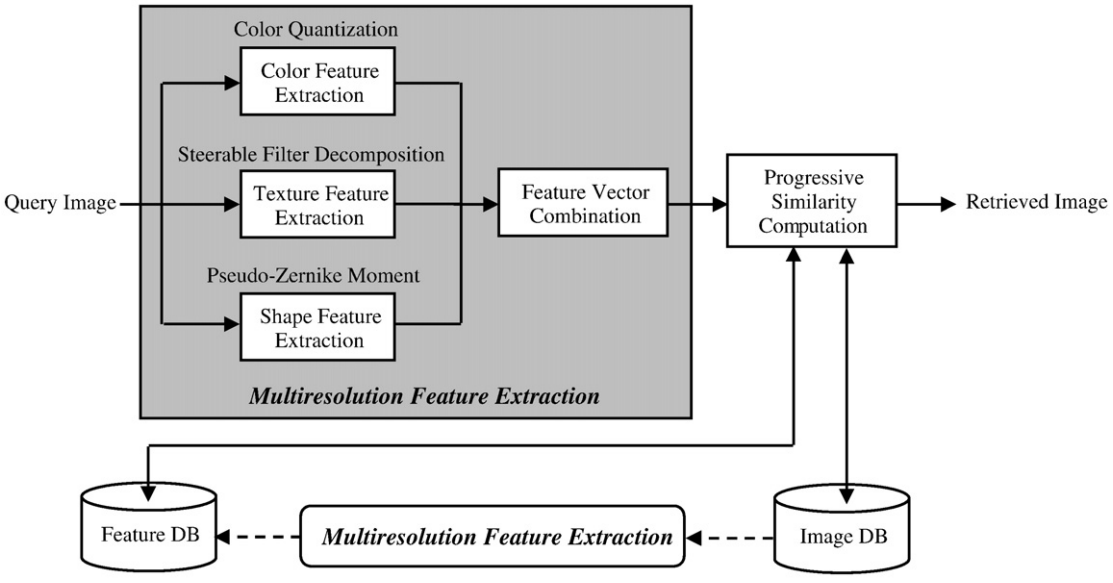


Fig. 6. Block diagram of the proposed method.

Fig. 7. The image retrieval results (horses) using different schemes. (a) The retrieval based on traditional color histogram; (b) the retrieval based on color histogram of subblocks [3]; (c) the proposed retrieval method.

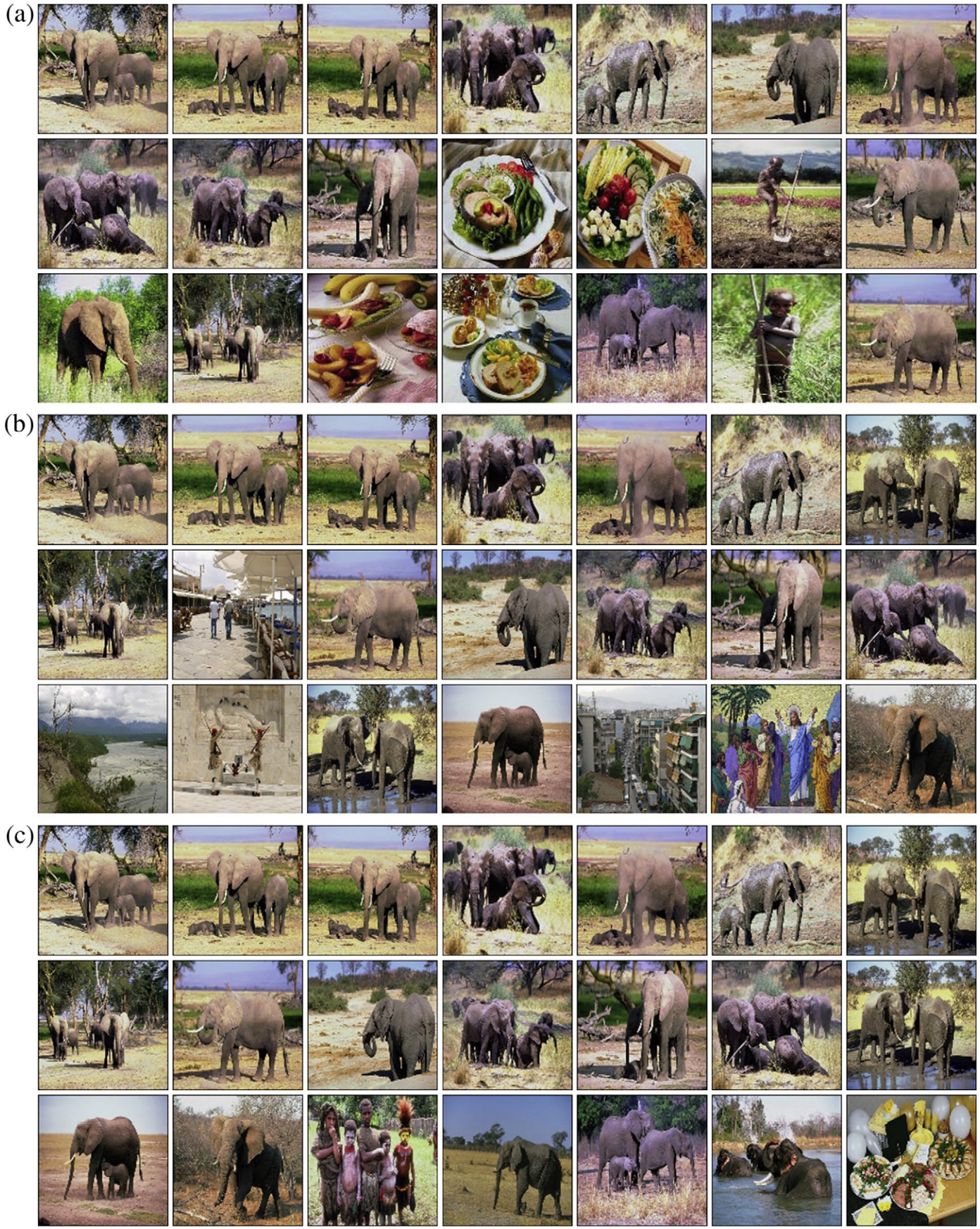


Fig. 8. The image retrieval results (elephants) using different schemes. (a) The retrieval based on traditional color histogram; (b) the retrieval based on color histogram of subblocks [3]; (c) the proposed retrieval method.

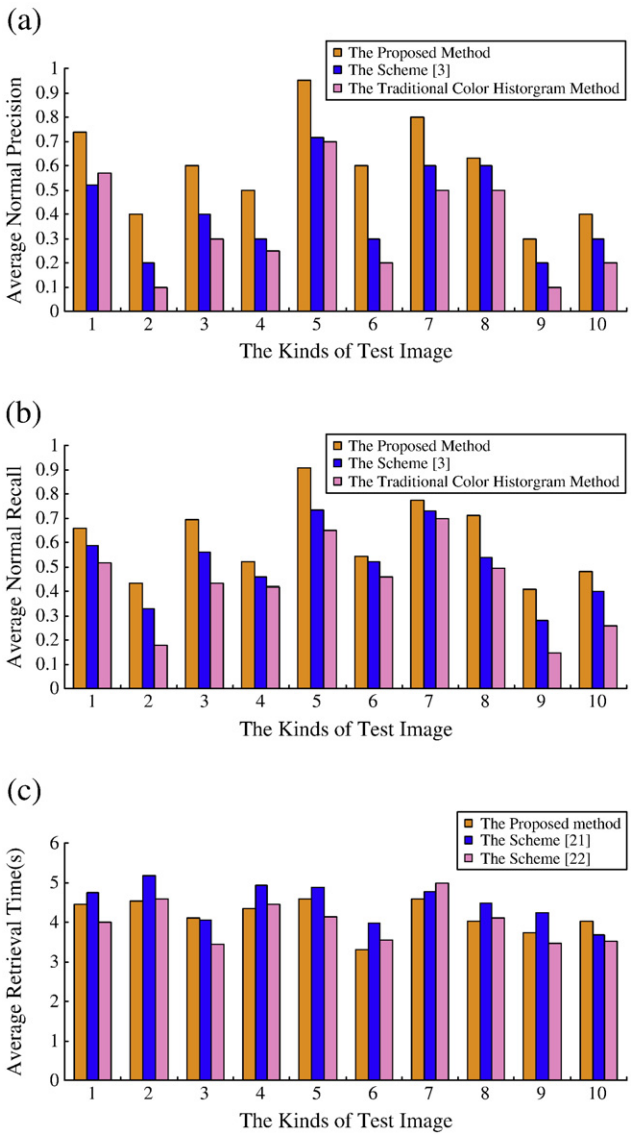


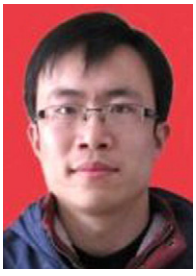
Fig. 9. The average retrieval performance of three schemes. (a) The average normal precision. (b) The average normal recall. (c) The average retrieval time.

Open Foundation of State Key Laboratory of Networking and Switching Technology of China under Grant No. SKLNST-2008-1-01, the Open Foundation of State Key Laboratory of Information Security of China under Grant No. 03-06, the Open Foundation of State Key Laboratory for Novel Software Technology of China under Grant No. A200702, and Liaoning Research Project for Institutions of Higher Education of China under Grant No. 2008351.

References

[1] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Computing Surveys 40 (2) (2008) 1–60.
[2] J. Vogel, B. Schiele, Performance evaluation and optimization for content-based image retrieval, Pattern Recognition 39 (5) (2006) 897–909.
[3] Xuelong Li, Image retrieval based on perceptive weighted color blocks, Pattern Recognition Letters 24 (12) (2003) 1935–1941.
[4] ISO/IEC 15938-3/FDIS Information Technology—Multimedia Content Description Interface—Part 3 Visual Jul. 2001, ISO/IEC/JTC1/SC29/WG11 Doc. N4358.
[5] L.M. Po, K.M. Wong, A new palette histogram similarity measure for MPEG-7 dominant color descriptor, Proceedings of the 2004 IEEE International Conference on Image Processing (ICIP'04), Singapore, October 2004, pp. 1533–1536.
[6] Rui Min, H.D. Cheng, Effective image retrieval using dominant color descriptor and fuzzy support vector machine, Pattern Recognition 42 (1) (2009) 147–157.
[7] Nai-Chung Yang, Wei-Han Chang, Chung-Ming Kuo, Tsia-Hsing Li, A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval, Journal of Visual Communication and Image Representation 19 (2) (2008) 92–105.
[8] Tzu-Chuen Lu, Chin-Chen Chang, Color image retrieval technique based on color features and image bitmap, Information Processing and Management 43 (2) (2007) 461–472.
[9] Young Deok Chun, Nam Chul Kim, Ick Hoon Jang, Content-based image retrieval using multiresolution color and texture features, IEEE Transactions on Multimedia 10 (6) (2008) 1073–1084.
[10] Y.D. Chun, S.Y. Seo, N.C. Kim, Image retrieval using BDIP and BVLC moments, IEEE Transactions on Circuits and Systems for Video Technology 13 (9) (2003) 951–957.
[11] C. Theoharatos, V.K. Pothos, N.A. Laskaris, G. Economou, Multivariate image similarity in the compressed domain using statistical graph matching, Pattern Recognition 39 (10) (2006) 1892–1904.
[12] Manesh Kokare, P.K. Biswas, B.N. Chatterji, Texture image retrieval using rotated wavelet filters, Pattern Recognition Letters 28 (10) (2007) 1240–1249.
[13] Ju. Han, Kai-Kuang Ma, Rotation-invariant and scale-invariant Gabor features for texture image retrieval, Image and Vision Computing 25 (9) (2007) 1474–1481.
[14] B.S. Manjunath, P. Salembier, T. Sikora, Introduction to MPEG-7: Multimedia Content Description Interface, Wiley, Chichester, 2002.
[15] D. Zhang, G. Lu, Review of shape representation and description techniques, Pattern Recognition 37 (1) (2004) 1–19.
[16] Xu. Xiaoqian, Dah-Jye Lee, Sameer Antani, L. Rodney Long, A spine X-ray image retrieval system using partial shape matching, IEEE Transactions on Information Technology in Biomedicine 12 (1) (2008) 100–108.
[17] Chia-Hung Wei, Yue Li, Wing-Yin Chau, Chang-Tsun Li, Trademark image retrieval using synthetic features for describing global shape and interior structure, Pattern Recognition 42 (3) (2009) 386–394.
[18] S. Liapis, G. Tziritas, Color and texture image retrieval using chromaticity histograms and wavelet frames, IEEE Transactions on Multimedia 6 (5) (2004) 676–686.
[19] A. Vadivel, A.K. Majumdar, S. Sural, Characteristics of weighted feature vector in content-based image retrieval applications, Proceedings of IEEE International conference on Intelligent Sensing and Information Processing, Chennai, India, Jan. 2004, pp. 127–132.
[20] Y.D. Chun. Content-based image retrieval using multiresolution color and texture features. Ph.D. thesis, Dept. Elect. Eng., Kyungpook National Univ., Daegu, Korea, Dec. 2005 [Online]. Available: http://vcl.knu.ac.kr/phd_thesis/ydchun.pdf.
[21] P.S. Hiremath, J. Pujari, Content based image retrieval using color, texture and shape features, Proceedings of the 15th International Conference on Advanced Computing and Communications, Guwahati, Inde, December 2007, pp. 780–784.
[22] Y. Deng, B.S. Manjunath, C. Kenney, An efficient color representation for image retrieval, IEEE Transactions on Image Processing 10 (1) (2001) 140–147.
[23] M. Jacob, M. Unser, Design of steerable filters for feature detection using Canny-like criteria, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (8) (2004) 1007–1019.
[24] A. Khotanzad, Y.H. Hong, Invariant image recognition by Zernike moments, IEEE Transactions on Pattern Anal. Mach. Intell 12 (5) (1990) 489–497.

Xiangyang Wang was born in Tieling, China, in 1965. He is currently a professor with the School of Computer and Information Technology at the Liaoning Normal University, China. He obtained his B.S. degree from the Lanzhou University, China and his M.S. degree from the Jilin University, China, in 1988 and 1995, respectively. His research interests include signal processing and communications, digital multimedia data hiding and information assurance, applications of digital image processing, and computer vision. He has published more than 150 journal papers, 20 conference papers, and contributed in 2 books in his areas of interest.



Yongjian Yu received his B.S. degree from the School of Computer and Information Technology, Liaoning Normal University, China, in 2006, where he is currently pursuing the M.S. degree. His research interests include image retrieval and signal processing.

Hongying Yang is currently an assistant professor with the School of Computer and Information Technology at the Liaoning Normal University, China. She received her B.S. degree from the Liaoning Normal University, China in 1989. Her research interests include signal processing and communications, and digital multimedia data hiding.