CrossMark

# Improving image retrieval by integrating shape and texture features

Yu-Nan Liu[1] · Shan-Shan Zhang[1] · Yu Sang[2] ·
Si-Miao Wang[3]

**Abstract** Content-based image retrieval (CBIR) has been an active research topic in the last decade. Multiple feature extraction and representation is one of the most important issues in the CBIR. In this paper, we propose a new CBIR method based on an efficient integration of texture and shape features. The texture features are extracted on the decomposed images processed by the optimal non-subsampled shearlet transform (NSST), and are represented by the high-frequency sub-band coefficients, which can be modeled by Bessel K Form (BKF) distribution; the shape features are represented by low-order quaternion polar harmonic transforms (QPHTs). The two kinds of features are then integrated by a weighted distance measurement, where Kullback-Leibler distance (KLD) and Euclidean distance (ED) are used for texture and shape features respectively. The integration of shape and texture information provides a robust feature set for image retrieval. Experimental results on standard benchmarks show significant improvements on retrieval performance using the proposed method compared with previous state-of-the-art methods.

**Keywords** Content-based image retrieval · Non-subsampled shearlet transform · BKF distribution · Quaternion polar harmonic transforms

---

✉ Yu-Nan Liu
   lynphd@163.com

1   The Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, People's Republic of China

2   School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China

3   School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China

# 1 Introduction

Digital images are one of the most important media materials, which provide a large amount of information for communication. With advances in information and Internet technology, there is an explosive growth of digital image databases (DBs), which require effective and efficient methods that allow users to search through such large image collections [17]. Depending on the query formats, there are usually three different types of image retrieval methods: text-, content-, and cross-model based methods. Text-based image retrieval (TBIR) is a traditional searching approach which finds the matching between query keywords and image annotations in DBs. Such methods require manual tagging of a large number of images and always fail to retrieve visually similar images. To alleviate the difficulties of text-based methods, an alternative approach called content-based image retrieval (CBIR) [29, 40] has been proposed and has attracted extensive research attention in the last decade. In a typical CBIR system, low-level features related to visual contents such as color, shape, and texture 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 ranked based on their similarities to the query image. In recent years, Cross-model retrieval (CMR) [55, 60] has gained a lot of attentions for retrieval of real world database images (such as images with informative tags or textual descriptions). These methods can effectively solve the problem of dimension disaster and semantic gap by combining text annotation with visual content. However, these methods still require manual annotation of the semantic information of images. In this paper, we focus on CBIR systems, especially on how to improve the performance of image retrieval by extracting of compact and representative visual features [14, 47].

In the early stage of the development of CBIR, most researches only used one kind of features among different low-level visual features. However, it is hard to attain satisfactory retrieval results effectively by using just one feature because, in general, an image contains various visual characteristics. Recently, active researches in image retrieval using a combination of different features have been performed [36, 47, 54, 60]. But it is shown that such a combination of features does not always guarantee better retrieval accuracy [43, 47, 54]. Accordingly, for an advanced CBIR, it is necessary to choose efficient visual 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 novel CBIR method based on an efficient combination of shape and texture features, in which QPHTs is used to extract shape feature and BKF modeling for NSST domain is used to extract texture feature. NSST and QPHTs are two extremely significant technologies that have great advantages in extracting image content. They have been successfully applied in image description and feature extraction. Therefore, they are very useful for CBIR in this paper. The novelty of the proposed method includes: 1) shape information is represented by quaternion polar harmonic transforms (QPHTs) coefficients, which has many desirable properties such as expression efficiency, robustness to noise, and geometric invariance, etc.; 2) image texture is represented by BKF parameters of NSST sub-bands, which are robust to illumination and image blurring, and also reduce computational complexity in the texture retrieval phase; 3) QPHTs coefficients and BKF parameters of NSST are combined effectively for image retrieval.

The rest of this paper is organized as follows. A review of previous related work is presented in Section 2. Some preliminaries on the NSST and BKF are given in Section 3. Section 4 recalls the decomposition about QPHTs. Section 5 describes the proposed CBIR

method by integrating shape and texture features. In Section 6, the effectiveness and efficiency of the proposed method is evaluated. Finally, we conclude this paper in Section 7.

## 2 Related work

Over the past decades, CBIR has been actively investigated by researchers in many applications. Comprehensive surveys exist on the different techniques used in this area [29, 46]. Also, there are some literatures that survey the important CBIR systems. Early systems mostly adopted simple low-level visual features for image retrieval, while more effective features such as SIFT [30, 32], HOG [9] and CNN [1, 34] have been applied recently. Since the work in this paper is related to search using color, texture, and shape features, this section mainly reviews existing works based on these features.

Color is one of the most common and determinant low-level visual feature, which is stable against direction variations and background complexity. As conventional color features used in CBIR, there are color histogram, color moments, and MPEG-7 color descriptor [43]. Li et al. [22] presented a novel algorithm based on running sub-blocks with different similarity weights for object-based image retrieval. By splitting the entire image into certain sub-blocks, the color region information and similarity matrix analysis are used to retrieval images under the query of special object. Chen et al. [8] proposed an adaptive color feature extraction method. Based on the binary quaternion-moment-preserving 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. Wang et al. [53] proposed a CBIR system based on the color histogram of the local feature regions. In their scheme, an RGB color image is converted into YCbCr color space and multi-scale Harris-Laplace detector was applied for extracting feature points. The local feature region construction and local feature regions (LFRs) quantization process were applied. Finally the histogram of the quantized LFRs was used in the retrieval process. Liu et al. [28] presented a novel image feature representation method, namely color difference histogram (CDH) for image retrieval. The features can be considered as a novel visual attribute descriptor combining edge orientation, color and perceptually uniform color difference. Talib et al. [41] came up with a new semantic feature extracted from dominant colors. The newly proposed technique helps reduce the effect of image background on image matching decision where an object's colors receive much more focus. Imran et al. [16] decomposed a color image into sub-images and converted each sub-image into a HSV sub-image. They formed a feature color vector by combining the computed mean, variance and skewness of equalized histograms of each HSV sub image. In general the color histogram matching based CBIR systems is relatively simple and fast but only color information is not sufficient to retrieve the objects having different color features.

Texture is an important visual attribute both for human perception and image analysis systems [61], its role in domain-specific image retrieval is particularly vital due to their close relation to the underlying semantics in these cases. As conventional texture features used in CBIR, there are gray-level co-occurrence matrix (GLCM), Markov random field (MRF) model, Gabor filters, and edge histogram descriptor (EHD) etc. He et al. [13] 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. [21] introduced two new multivariate models using, respectively, generalized Gaussian and Weibull densities. These models can capture both the sub-band 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. Atto et al. [6] 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. [35] presented a novel CBIR scheme based on statistical texture features using the complex wavelets. 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 features also play an important role in human recognition and perception. Many shape descriptors have been proposed for different applications. They are generally categorized into two groups, namely, contour-based descriptors and region-based descriptors [20]. Contour-based methods utilize boundary information which is crucial to human perception in judging shape similarity. It is difficult to extract boundary points from natural images which are rich in texture contents. Region-based methods exploit shape interior information, therefore, they can be applied to more general shapes. Among the region based techniques, rotation invariants [15], gradient features, and curvature scale space are the popular region-based descriptors [45]. By using a mathematical form of analysis, Li et al. [23] compared the amount of visual information 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. [38] 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. [19] 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. [26] considered the family of total Bregman 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. Yap et al. [57] 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. [58] 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. Wang et al. [54] proposed an effective color image retrieval scheme for combining all the three i.e. color, texture and shape information, which achieved higher retrieval efficiency. Farsi et al. [12] presented a novel CBIR method based on combination of Hadamard matrix and discrete wavelet transform in hue-min-max-difference color 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. Seetharaman et al. [36] 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, and spectrum of energy are used. Zhao et al. [59] discussed a novel approach of CBIR, which combines color, texture and shape descriptors to represent the features of the image. This scheme is based on three noticeable algorithms: (1) color distribution entropy takes the correlation of the color spatial distribution in an image into consideration, (2) color level co-occurrence is served as the texture feature, which is a new descriptor that is grounded on co-occurrence matrix to seize the alteration of the texture, and (3) Hu invariant moments are frequently used owing to its invariance under translation, changes in scale, and also rotation. Varish et al. [44] proposed a hierarchical approach for designing a CIBR scheme based on the color and texture features of an image. Singha et al. [39] proposed a CBIR approach based on the combination of Haar wavelet transformation using lifting scheme and the color histogram. The color feature is described by the color histogram, 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. Khokher et al. [20] proposed a CBIR scheme for retrieval images via color, texture, and shape features. Using three specialized histograms (i.e. color, wavelet, and edge histograms), the authors showed that a more accurate representation of the underlying distribution of the image features improves the retrieval quality. Wang et al. [47] proposed a new CBIR method based on an efficient combination of shape and texture features. As its shape features, exponent moments descriptor, which has many desirable properties, is adopted in RGB color space. As its texture feature, localized angular phase histogram of the intensity component is used in hue saturation intensity. This paper combines local binary pattern (LBP) with Legendre moments at multiple resolutions of wavelet decomposition of image. 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 [46, 47]. 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 Texture feature extraction

In recent years, the shearlet transform has been introduced, which can yield nearly optimal approximation properties [37, 48]. The shearlet transform has the following main properties: parabolic scaling, high directional sensitivity, spatially localizing, and optimally sparse. The NSST, which combined the non-subsampled Laplacian pyramid transform with several different combinations of the shearing filters, is the shift-invariant version of the shearlet transform. The NSST differs from the shearlet transform in that the NSST eliminates the

down-samplers and up-samplers. It not only can exactly compute the shearlet coefficients, but can also provide nearly optimal approximation for 2D images. Consequently, introduction of NSST into image retrieval could make use of the good characters of NSST in effectively extracting texture feature from original images.

### 3.1 Non-subsampled shearlet transform (NSST)

Consider the two-dimensional affine system for a continuous wavelet $\psi \in L^2(R^2)$,

$$\left\{ \psi_{ast}(x) = |\det M_{as}|^{\frac{1}{2}} \psi\left(M_{as}^{-1}x - t\right) : t \in R^2, M_{as} \in \Gamma \right\} \tag{1}$$

where $\Gamma$ is the 2 parameter dilation group,

$$\Gamma = \left\{ M_{as} = \begin{pmatrix} a & \sqrt{as} \\ 0 & \sqrt{a} \end{pmatrix} : (a,s) \in R^+ \times R \right\} \tag{2}$$

For any row vector $\xi = (\xi_1, \xi_2) \in R^2$, $a \in R^+$, $s \in R$ and $a \in R^2$ according to the Eq. (2). The continuous shearlet transform of $f \in L^2(R)$ is defined as:

$$SH(a,s,t) = \langle f, \psi_{ast} \rangle \tag{3}$$

The discrete shearlet transform $\hat{\psi}_{jlk}(j \geq 0, -2^j \leq l \leq 2^j - 1)$, which can deal with distributed discontinuities, is obtained by sampling continuous shearlet transform $SH(a,s,t)$ [49]. Each element of $\hat{\psi}_{jlk}$ is supported on a pair of trapezoids of approximate size $2^{2j} \times 2^j$.

A primary advantage of the shearlet transform is that there are no constraints on the size of the supports for the shearing and no restrictions on the number of directions: unlike the construction of the directional filter banks in [49]. Hence, the NSST consists of two phases, which are the non-subsampled Laplacian pyramid and several different combinations of the shearing filters [24]. NSLP can be analyzed through iterative processing as follows:

$$NSLP_{j+1} = A_j f = \left( Ah_j^1 \prod_{k=1}^{j-1} Ah_k^0 \right) f \tag{4}$$

where $f$ is an image, $NSLP_{j+1}$ is the detail coefficients at scale $j+1$, and $Ah_k^0$ and $Ah_j^1$ are low pass and high pass filters of NSLP at scale $j$ and $k$ respectively. Given $N \times N$ image $f_a^0$ and the number of direction $D_j$, the process of the NSST analysis described above at fixed resolution scale $j$ can be summarized below.

Step 1: Apply the NSLP to decompose $f_a^{j-1}$ into a low-frequency image $f_a^j$ of size $N \times N$ and a high-frequency image $f_d^j$;
Step 2: Compute $\hat{f}_d^j$ in pseudo polar grid, then get $Pf_d^j$;
Step 3: Apply a Band-Pass filtering to $Pf_d^j$ to obtain $\left\{ \hat{f}_{d,k}^j \right\}_{k=1}^{D_j}$;
Step 4: Apply inverse FFT to obtain NSST coefficients $\left\{ f_{d,k}^j \right\}_{k=1}^{D_j}$ in pseudo polar grid.

An example of frequency partition of the NSST is shown in Fig. 1. This type of frequency partition leads to the sparsity of the NSST coefficients, i.e., only the coefficients with both direction and location on the original image edges has significant values. This can be clearly
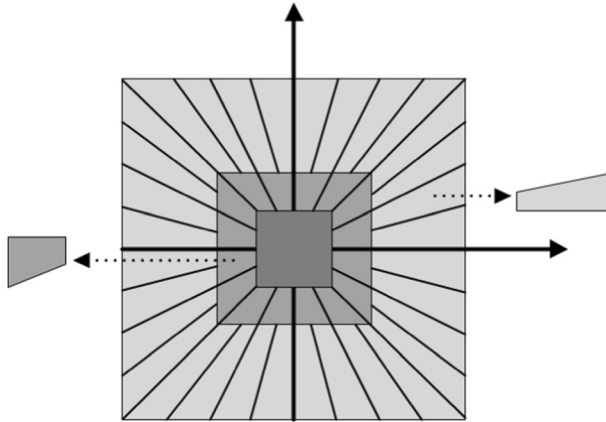
**Fig. 1** An example of frequency partition by the NSST

seen in Fig. 2, where the 2-level NSST is applied on the luminance channel of Barbara image. Here, the numbers of shearing directions are chosen to be 8 and 4 from finer to coarser scale.
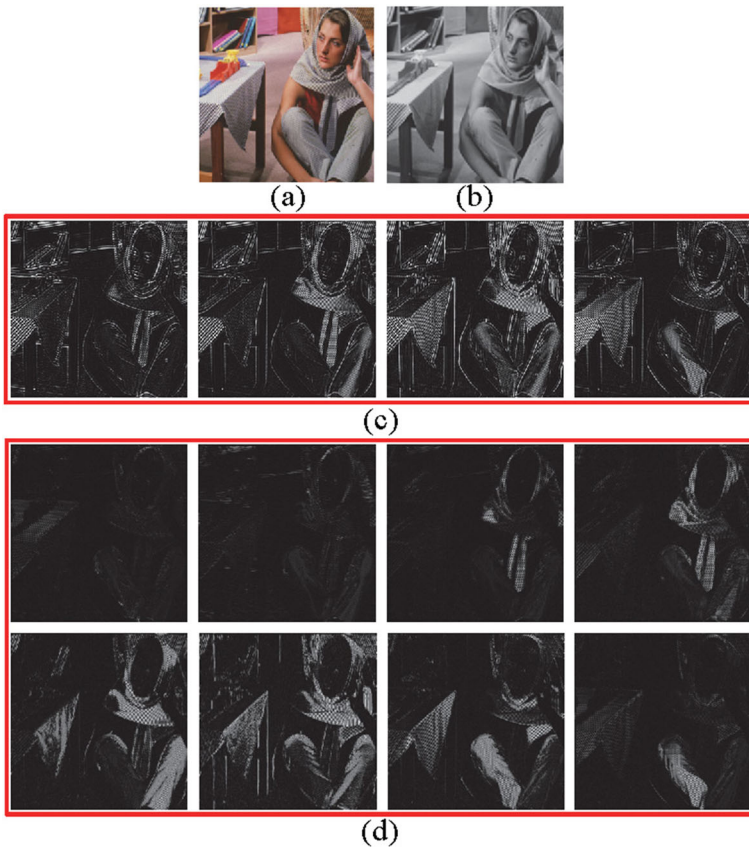


**Fig. 2** The NSST on the luminance channel of color image Barbara: **a** Original image, **b** The approximate NSST coefficients, **c** The detail NSST coefficients at scale 2, **d** The detail NSST coefficients at scale 1

### 3.2 Marginal statistics of NSST coefficients

In the subsection, we will discuss the marginal statistics of the NSST coefficients of images. A standard grayscale dataset image database, namely USC-SIPI image dataset [42], is used to study the marginal statistics, and Fig. 3 plots the histograms of the second scale sub-bands of the images. We apply two-level NSST decomposition, with directional sub-bands being 4 and 8, respectively from coarse to fine, as shown in Fig. 2. Figure 3 demonstrates that these distributions exhibit a sharp peak at zero around and heavy tails on both sides of the peak. This implies that the NSST is sparse, because the majority of coefficients are close to zero. The kurtoses of the four shown distributions shown are 22.74, 22.01, 27.29, and 26.47, which are much higher than the kurtosis of 3 for Gaussian distributions. Therefore, we need to model the NSST coefficients by a non-Gaussian distribution.

### 3.3 BKF Modeling of NSST sub-band coefficients

Using a physical model for image formation, a family of two-parameter probability densities, called Bessel K form (BKF), have been proposed in [11, 52] to model the distribution of arbitrary images that have been filtered by a variety of band-pass filters (e.g., derivative, Gabor, interpolation, steerable filters, etc). It is obvious that NSST decompositions of an image are members of such class of filters. Therefore, the BKF is a suitable model to capture the heavy tail behavior of NSST coefficients densities.
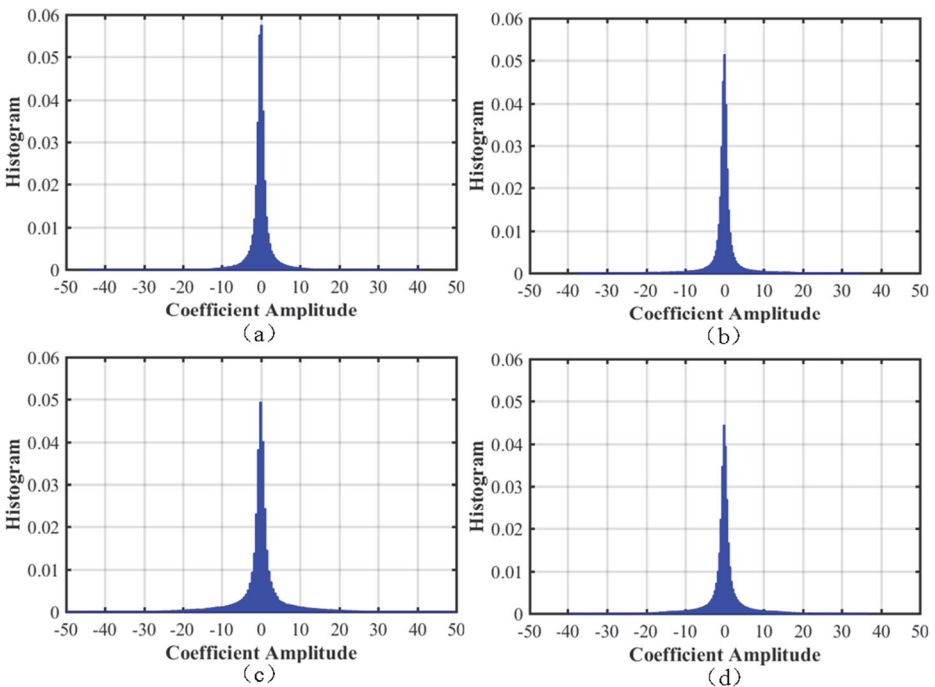


Fig. 3 Marginal statistics of four second scale sub-bands of the image Barbara: The kurtosis of the distributions is measured at **a** 22.74, **b** 22.01 **c** 27.29, and **d** 26.47

Let $g_F$ be a filtered version of an image $g$ by through the bandpass filter $F$. The Bessel K form PDF of $g_F$ has been shown to be [11] for $p>0$, $c>0$

$$f(x;p,c) = \frac{\left(\frac{c}{2}\right)^{-\frac{p}{2}-\frac{1}{4}}\left|\frac{x}{2}\right|^{p-\frac{1}{2}}K_{p-\frac{1}{2}}\left(\sqrt{\frac{2}{c}}|x|\right)}{\sqrt{\pi}\Gamma(p)} \tag{5}$$

where $K_v$ indicates the modified Bessel function defined as

$$K_v(zx) = \frac{\Gamma\left(v+\frac{1}{2}\right)(2x)^v}{\sqrt{\pi}z^v}\int\limits_0^{+\infty}\frac{\cos(zt)dt}{(t^2+x^2)^{v+1/2}}, \tag{6}$$
$$\left(\mathrm{Re}(v) > -\frac{1}{2}, z>0, |\arg x| < \frac{\pi}{2}\right)$$

where $p$ and $c$ are the shape and scale parameters respectively.

We restrict ourselves to only two-parameter BKF throughout this paper. For $p=1$, $f$ simply reduces to the double exponential PDF. If $p>1$, we get closer to the Gaussian case (especially when $p>>1$, which is intuitively acceptable using a central limit theorem argument). If $p<1$, the PDF becomes more sharply peaked and the tails are heavier.

BKF distribution has proved useful in the modeling of heavy-tailed data, especially NSST coefficients. To justify the selection of the BKF statistical model, we use the Kolmogorov-Smirnov (KS) metric to compare the empirical PDFs [49] (including Weibull distribution, General Gaussian distribution, Rayleigh distribution, Exponential distribution, Laplacian distributions, Cauchy distributions, and BKF distribution) with the prior PDF (i.e., the histogram) of the NSST coefficients. The KS metric is

$$d_{ks} = \max_{w\in R}|F_h(w)-F_e(w)| \tag{7}$$

where $F_h(w)$ and $F_e(w)$ denote the cumulative density function (CDF) of the prior PDF and the empirical CDF, respectively, and a smaller $d_{ks}$ value indicates a better performance.

Experiments are conducted using four widely used test images Lena, Barbara, and Couple, each of size $512 \times 512$. Each test image is applied by a two-level NSST, where the number of shearing directions is chosen to be 8 and 4 from finer to coarser scale. Then, the distributions of NSST coefficients of the second-level detail sub-bands are fitted with the seven statistical models, where the parameters involved are estimated using moment based estimation technique. The fitted models are further compared with the histogram of the NSST coefficients in the sense of KS metric. Table 1 shows the results concerning the KS metric for various empirical PDFs of the image NSST coefficients in the second finest scale.

It is evident from Table 1 that the BKF distribution fits the empirical data much more accurately than do other distributions.

# 4 Shape feature extraction

Shape is known to play an important role in human recognition and perception [20]. 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 CBIR can be seen from the fact

**Table 1** The KS performance comparison of various statistical models for NSST coefficients

| Test Images | Direction (the second finest scale) | BKF distribution | Weibull distribution | GG distribution | Rayleigh distribution | Exponential distribution | Laplacian distributions | Cauchy distributions |
|---|---|---|---|---|---|---|---|---|
| Lena | 1 | 0.022 | 0.066 | 0.235 | 0.533 | 0.121 | 0.275 | 0.036 |
|  | 2 | 0.009 | 0.078 | 0.217 | 0.582 | 0.178 | 0.356 | 0.027 |
|  | 3 | 0.009 | 0.074 | 0.161 | 0.571 | 0.192 | 0.433 | 0.019 |
|  | 4 | 0.007 | 0.083 | 0.201 | 0.602 | 0.194 | 0.397 | 0.029 |
| Peppers | 1 | 0.015 | 0.029 | 0.075 | 0.399 | 0.082 | 0.0489 | 0.027 |
|  | 2 | 0.015 | 0.037 | 0.101 | 0.435 | 0.095 | 0.361 | 0.020 |
|  | 3 | 0.018 | 0.022 | 0.029 | 0.360 | 0.039 | 0.401 | 0.033 |
|  | 4 | 0.017 | 0.052 | 0.104 | 0.478 | 0.116 | 0.391 | 0.019 |
| Barbara | 1 | 0.015 | 0.048 | 0.172 | 0.461 | 0.106 | 0.301 | 0.018 |
|  | 2 | 0.007 | 0.074 | 0.169 | 0.591 | 0.183 | 0.418 | 0.019 |
|  | 3 | 0.017 | 0.062 | 0.102 | 0.528 | 0.197 | 0.499 | 0.023 |
|  | 4 | 0.012 | 0.061 | 0.132 | 0.521 | 0.170 | 0.433 | 0.015 |
| Average |  | 0.014 | 0.050 | 0.121 | 0.477 | 0.124 | 0.3819 | 0.024 |

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 moment invariants have been utilized as pattern features in a number of applications [2, 23, 47]. The theory of moments provides useful series expansions for the representation of object shapes. In this section, we introduce a robust and effective shape feature based on quaternion polar harmonic transform (QPHT).

## 4.1 Polar harmonic transforms

In 2010, Yap et al. [56] introduced a set of 2D transforms named PHT based on a set of orthogonal projection bases. Compared with other orthogonal moment, PHT has a better image reconstruction, lower noise sensitivity, and lower computational complexity. Besides, the PHT is free of numerical instability issues so that high order moments can be obtained accurately.

The PHT coefficients $M_{n,\,m}$ of order $n$ with repetition $m$, $|n| = |m| = 0, 1, \ldots, \infty$, is defined as

$$M_{n,m} = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 f(r, \theta) \left[ H_{n,m}(r, \theta) \right]^* r dr d\theta \qquad (8)$$

where $[\cdot]^*$ denotes the complex conjugate and the basis $H_{n,\,m}$ can be decomposed into radial and circular components

$$H_{n,m}(r, \theta) = R_n(r) e^{im\theta} \qquad (9)$$

with the radial kernel being a complex exponential in the radial direction

$$R_n(r) = e^{i2\pi n r^2} \qquad (10)$$

And satisfying orthogonality condition

$$\int_0^1 R_n(r) \left[ R_{n'}(r) \right]^* r dr = \frac{1}{2} \delta_{n,n'} \qquad (11)$$

And also

$$\int_0^{2\pi} \int_0^1 H_{n,m}(r, \theta) \left[ H_{n',m'}(r, \theta) \right]^* r dr d\theta = \pi \delta_{n,n'} \delta_{m,m'} \qquad (12)$$

where $\pi$ is the normalization factor, $\delta_{n,n'}$ and $\delta_{m,m'}$ are the Kronecker symbols, and $\left[ H_{n',m'}(r, \theta) \right]^*$ is the conjugate of $H_{n',m'}(r, \theta)$.

Following the principle of orthogonal function [37, 51], the image function $f(r, \theta)$ can be reconstructed approximately by limited orders of PHT coefficients ($n \leq n_{\max}$, $m \leq m_{\max}$). The more orders used, the more accurate the image description

$$f'(r, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} M_{n,m} R_n(r) \exp(im\theta) \approx \sum_{n=-n_{\max}}^{n_{\max}} \sum_{m=-m_{\max}}^{m_{\max}} M_{n,m} R_n(r) \exp(im\theta) \qquad (13)$$

where $f'(r, \theta)$ is the reconstructed image. The basis functions $R_n(r) \exp(im\theta)$ of the PHT are orthogonal over the interior of the unit circle, and each order of the PHT coefficients makes an independent contribution to the reconstruction of the image.

## 4.2 Quaternion polar harmonic transform (QPHT)

A quaternion consists of one real part and three imaginary parts [50] as follows

$$q = a + bi + cj + dk \tag{14}$$

where $a$, $b$, $c$, and $d$ are real numbers, and $i$, $j$, and $k$ are three imaginary units obeying the following rules. $i^2 + j^2 + k^2 = -1 \ ij = -ji = k, jk = -kj = i, ki = -ik = j$

The conjugate and modulus of a quaternion are respectively defined by

$$q^* = a - bi - cj - dk, |q| = \sqrt{a^2 + b^2 + c^2 + d^2} \tag{15}$$

Let $f(r, \theta)$ is the reconstructed color image defined in polar coordinates, we define the right-side QPHT of order $n$ with repetition $m$ as

$$M_{n,m}^R = \frac{1}{\pi} \int_0^{2\pi} \int_0^1 f(r, \theta) \left[ H_{n,m}(r, \theta) \right]^* r dr d\theta$$

$$= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 f(r, \theta) \exp\left(-\mu 2\pi n r^2\right) \exp(-\mu m \theta) r dr d\theta \tag{16}$$

where $\mu$ is an unit pure quaternion chosen as $\mu = (i + j + k)/\sqrt{3}$.

Since the polar complex exponential transform basis functions are orthogonal, the color image $f(r, \theta)$ can be reconstructed approximately from limited orders of QPHT coefficients ($n \leq n_{\max}$, $m \leq m_{\max}$). The more orders used, the more accurate the color image description

$$f'(r, \theta) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} M_{n,m}^R R_n(r) \exp(\mu m \theta)$$

$$\approx \sum_{n=-n_{\max}}^{+n_{\max}} \sum_{m=-m_{\max}}^{+m_{\max}} M_{n,m}^R \exp\left(\mu 2\pi n r^2\right) \exp(\mu m \theta) \tag{17}$$

where $\acute{f}(r, \theta)$ is the reconstructed color image. The basis function $R_n(r) \exp(\mu m \theta)$ of the QPHT is orthogonal over the interior of the unit circle, and each order of the QPHT coefficients makes an independent contribution to the reconstruction of the color image.

Figure 4 gives some examples of image reconstruction using QPHT for standard color image "Lena" and "Barbara" (moment orders $N = 3, 5, 10, 15, 20, 30, 40, 50, 70$, and 100). As more QPHT coefficients are added to the reconstruction process, the reconstructed images get closer to the original images. As can be observed from the reconstructed images, QPHT capture the color image information, especially the edges. Also, it can be observed that the reconstructed color images using QPHT show visual resemblance to the original image in the
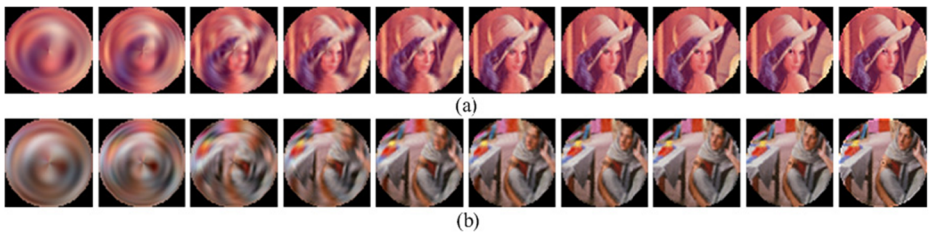


**Fig. 4** Some example of reconstructed images of 128 × 128 (moments orders = 3, 5, 10, 15, 20, 30, 40, 50, 70, and 100): **a** QPHT for image "Lena", **b** QPHT for image "Barbara"

early orders, and the QPHT is free of numerical instability issues. Figure 5 shows the modulus distribution of QPHT coefficients for image Lena under various attacks. It can be seen that the QPHT modulus coefficients have good robustness against various noises, geometric transforms, and color variations. So, QPHT modulus coefficients are suitable for invariant color image description.

From the foregoing, we can obtain rotation, scaling, and translation invariant QPHT modulus coefficients. However, we do not need all the QPHT modulus coefficients in color image retrieval. The number of QPHT modulus coefficients required, however, does not need to be large, since shape features can normally be captured by just a few low-frequency modulus coefficients. Further, the QPHT modulus coefficients $\left|M_{n,-m}^R\right| = \left|M_{n,m}^R\right|$, so only $M_{n,m}^R(n \geq 0, m \geq 0)$ is selected as the shape feature in this paper. Table 2 lists the selected QPHT features for different max orders. From the reconstruction results (see Fig. 4), we can see that QPHT, with the max order up to fifteen, could have a sufficiently good color image representation power.

## 5 The proposed content-based color image retrieval scheme

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
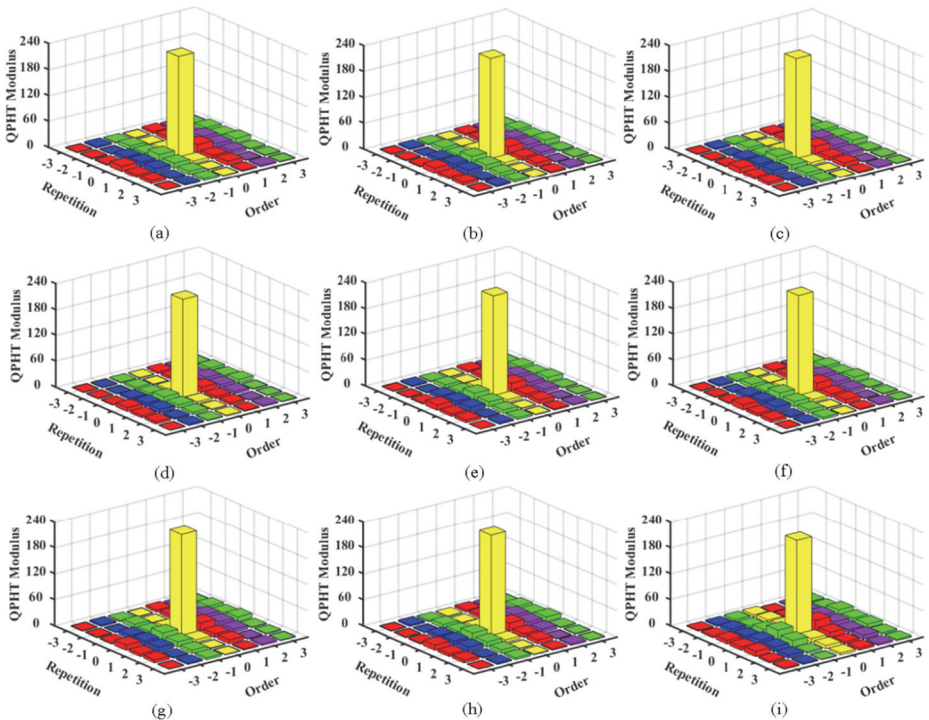


**Fig. 5** The modulus distribution of QPHT coefficients for image "Lena" under various attacks: **a** No attack, **b** Gaussian filtering, **c** salt and pepper noise, **d** JPEG compression 50, **e** light increasing, **f** contrast lowering, **g** translation, **h** rotation, **i** scaling

**Table 2** List of the selected QPHT features for different max orders

| Order | QPHT modulus coefficients | No. of moments | Accumulative No. |
|---|---|---|---|
| 0 | $\left|M_{0,0}^{R}\right|$ | 1 | 1 |
| 1 | $\left|M_{1,1}^{R}\right|$; $\left|M_{1,0}^{R}\right|$ | 2 | 3 |
| 2 | $\left|M_{2,2}^{R}\right|$; $\left|M_{2,1}^{R}\right|$; $\left|M_{2,0}^{R}\right|$ | 3 | 6 |
| 3 | $\left|M_{3,3}^{R}\right|$; $\left|M_{3,2}^{R}\right|$; $\left|M_{3,1}^{R}\right|$; $\left|M_{3,0}^{R}\right|$ | 4 | 10 |
| 4 | $\left|M_{4,4}^{R}\right|$; $\left|M_{4,3}^{R}\right|$; $\left|M_{4,2}^{R}\right|$; $\left|M_{4,1}^{R}\right|$; $\left|M_{4,0}^{R}\right|$ | 5 | 15 |
| ...... | | | |
| 8 | $\left|M_{8,8}^{R}\right|$; $\left|M_{8,7}^{R}\right|$; $\left|M_{8,6}^{R}\right|$; $\left|M_{8,5}^{R}\right|$; $\left|M_{8,4}^{R}\right|$; $\left|M_{8,3}^{R}\right|$; $\left|M_{8,2}^{R}\right|$; $\left|M_{8,1}^{R}\right|$; $\left|M_{8,0}^{R}\right|$ | 9 | 45 |
| 9 | $\left|M_{9,9}^{R}\right|$; $\left|M_{9,8}^{R}\right|$; $\left|M_{9,7}^{R}\right|$; $\left|M_{9,6}^{R}\right|$; $\left|M_{9,5}^{R}\right|$; $\left|M_{9,4}^{R}\right|$; $\left|M_{9,3}^{R}\right|$; $\left|M_{9,2}^{R}\right|$; $\left|M_{9,1}^{R}\right|$; $\left|M_{9,0}^{R}\right|$; | 10 | 55 |
| ...... | | | |

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 an efficient combination of shape and texture features. Figure 6 describes our image retrieval system framework.

### 5.1 Shape and texture features

According to Section 4.2, we can compute rapidly the QPHT coefficients. However, we don't need too much QPHT coefficients in color image retrieval, since color image features can normally be captured by just a few low-order QPHT coefficients. The shape feature vector based on QPHT is represented as $F = \left[\tilde{M}_{00}, \tilde{M}_{01}, \tilde{M}_{10}, \ldots, \tilde{M}_{nm}\right]$. Normalize the QPHT coefficients:

$$M_{ij} = \frac{\tilde{M}_{ij} - \mu_F}{\sigma_F} \qquad (18)$$
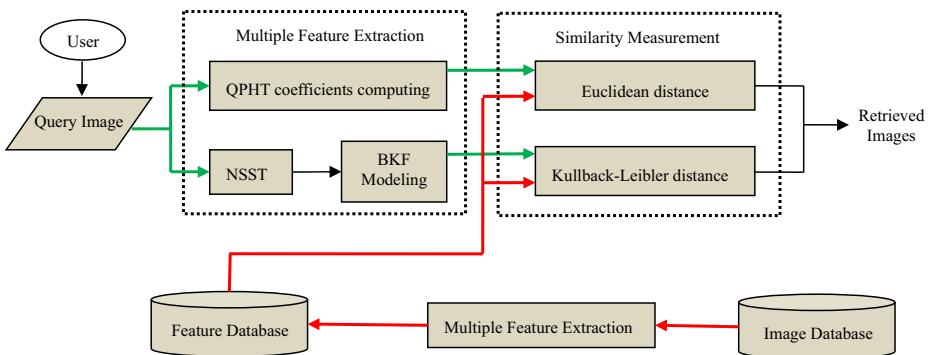


**Fig. 6** Block diagram of the proposed content-based image retrieval system. The green and red lines indicate the path of query image and database image, respectively

where $\mu_F$ and $\sigma_F$ are mean and standard deviation of $F$ respectively. Then the normalized shape feature vector is written as:

$$V_{\text{shape}} = [M_{00}, M_{01}, M_{10}, \ldots, M_{nm}] \tag{19}$$

Effective parameter estimation is necessary for accurate modeling NSST coefficients accurately. Many studies in the literatures [11, 49], describe methods parameter estimation using the statistical model, such as the maximum likelihood (ML) estimator, the moment/Newton-step (MN) estimator, and the moment based method (MM). The MM method is a feasible and effective parameter estimation approach; we will use this method in this paper to estimate the parameters of BKF, as follows [49],

$$\hat{p} = \frac{3(n-2)(n-3)m_2^2}{(n-1)\left[(n+1)m_4 - 3(n-1)m_2^2\right]} \tag{20}$$

$$\hat{c} = \frac{nm_2}{(n-1)\hat{p}} \tag{21}$$

where $n$ indicates the number of samples used in the estimate, and $m_2$ and $m_4$ are the second and fourth order sample central moments, respectively.

In the proposed method, a three-level non-subsampled shearlet transform (NSST) is applied to each color image, and then 20 directions sub-bands can be obtained. The probability density function is utilized to model the 20 high-pass sub-bands, and the scale parameter and shape parameter of each sub-band are estimated. Forty parameters can be obtained to form the BKF statistical model features (BSMFs), which can efficiently represent the texture of remote sensing images. The BSMFs vector is described as follows:

$$V_{\text{texture}} = [p_1, c_1, p_2, c_2, \ldots, p_{20}, c_{20}] \tag{22}$$

## 5.2 Similarity measurement

The texture feature similarity between two NSST sub-bands can be figured out effectively by the BKF parameters. Meanwhile, the NSST coefficients in different sub-bands are independent. Therefore, the overall distance between two images is the sum of all the Kullback-Leibler distance (KLD) across the corresponding high-frequency NSST sub-bands. The texture feature distance between the query image and the database image is represented as follows:

$$D_1\left(V_{texture}^{I_Q}, V_{texture}^{I_T}\right) = \sum_{j=1}^{J} \sum_{d=1}^{D_j} f_Q^{(j,d)} \log\left(\frac{f_Q^{(j,d)}}{f_T^{(j,d)}}\right) \tag{23}$$

where $f_Q^{(j,d)}$ and $f_T^{(j,d)}$ represent the BKF statistical model in the two images $I_Q$ and $I_T$, respectively, for the high-frequency sub-band of the $j$th scale and the $d$th direction. There is no need for normalization on texture feature vectors in this method of similarity measurement.

The similarity measurement between the shape feature vectors is selected to be Euclidean distance, which is the most common distance measurement and is defined as follows:

$$D_2\left(V_{shape}^{I_Q}, V_{shape}^{I_T}\right) = \sum_{i=1}^{K} \left(V_{shape}^{I_Q} - V_{shape}^{I_T}\right)^2 \tag{24}$$

where $V_{shape}^{I_Q}$ is the shape feature vector of the query image, $V_{shape}^{I_T}$ is the shape feature vector of image in the database, and $K$ is the number of vector elements.

The final distance between the image $I_Q$ and $I_T$ is defined by the weighted distance formula as follows:

$$D\left(I_Q, I_T\right) = \omega_1 D_1\left(V_{texture}^{I_Q}, V_{texture}^{I_T}\right) + \omega_2 D_2\left(V_{shape}^{I_Q}, V_{shape}^{I_T}\right) \tag{25}$$

where $\omega_1$ and $\omega_2$ are the weights of the shape and texture features respectively and $\omega_1 + \omega_2 = 1$. We use a minimum distance criterion and sort the database images for each query.

# 6 Simulation results

In this paper, we propose a new and effective CBIR method for combining texture and shape feature, which achieve higher retrieval efficiency. To evaluate the performance of the proposed algorithm, we conduct an extensive set of experiments by comparing the proposed scheme to the several state-of-the-art pipelines including traditional handcraft feature-based methods [5, 10, 18, 20, 25, 27, 44, 47] and CNN-based methods [1, 4, 7, 31, 33, 34]. First, we conduct the parameter selection experiment, and then compare the proposed algorithm with the recent multi-feature fusion retrieval method, and finally compare the proposed algorithm with other excellent methods (including Local-based algorithms and CNN-based algorithms).

## 6.1 Image database and evaluation criteria

The proposed color image retrieval system has been implemented by using MATLAB R2011b on the platform Intel core i5–7500 @ 3.4GHz, 16G RAM, 64 bit, Microsoft Windows 10 OS.

To check the retrieval efficiency of proposed method, we perform experiments on several well-known image benchmark datasets. The first image dataset used in this work is that of Wang et al. [47]. It is a subset of the COREL photo collection and is composed of 10,000 color images from 150 semantic categories, in which each category contains 100 images. Every database image is stored in JPEG format with size 384 × 256 or 256 × 384. This dataset covers a variety of topics, such as 'Flowers', 'Buildings', 'Elephants', 'Buses', 'Planes', and 'Foods', etc., with corresponding category ID's denoted by integers from 1 to 100, respectively. This category information availability is an advantage of this dataset since it makes evaluation of retrieval results easier. Ideally, the goal is to retrieve images belonging to the same category as the query image.

We also perform experiments on other 3 public datasets: UK-bench [47], Holidays [47], and Oxford [47]. **UK-bench dataset** consists of 10,200 images of 2250 different objects. Each object image is taken under four different viewpoints to get four visually similar images. The standard accuracy measure used for the UK-bench is computing the precision at top 4 images then the results averages over all queries. The best accuracy can be achieved is 4, e.g. 1

indicates only one relevant image to the query is retrieved at top 4 images, and 4 indicates all relevant images are successfully retrieved and ranked. **Holidays dataset** contains 500 images groups and in all 1491 personal Holidays photos undergoing various transformations. The number of photos in an image group is variable. The dataset contains a large variety of scene types such as nature, water, and fire effects, etc. The resolution of the images is very high $(2448 \times 3204)$ and for our experiments we scale them to the size $128 \times 128$ using bicubic interpolation of MATLAB. **Oxford dataset** contains 5062 high-resolution images $(1024 \times 768)$ showing either one of 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.

The performance of an image retrieval system is normally measured using precision $P(N)$ and recall $R(N)$ for retrieving top $N$ images defined by

$$P(N) = \frac{I_N}{N} \tag{26}$$

$$R(N) = \frac{I_N}{M} \tag{27}$$

where $I_N$ is the number of relevant retrieved from top $N$ positions and $M$ is the total number of images in the dataset that are similar to the query image. The precision and recall measure the accuracy of image retrieval with relevancy to the query and database image. While the precision provides the accuracy of retrieval out of the top $N$ retrieved images, the recall provides the accuracy with respect to the total number of relevant images in the database which are similar to the query image. Thus, only recall cannot measure the effectiveness of a retrieval system, precision must also be computed. The average normal precision of a single query is the mean of all the precision scores for each of the top $N_R$ retrieval:

$$\overline{P}(q) = \frac{1}{N_R} \sum_{N-1}^{N_R} P(N) \tag{28}$$

The mean average precision (mAP) is the mean of the average precision scores over all queries Q:

$$mAP = \frac{1}{Q} \sum_{q=1}^{Q} \overline{P}(q) \tag{29}$$

The mAP measure contains both the precision and recall information and represents the entire ranking.

## 6.2 The performance of parameters selection

In our image retrieval, shape feature is represented by quaternion polar harmonic transforms (QPHTs) coefficients and texture feature is represented by BKF parameters of NSST sub-bands. To evaluate the overall performance of the proposed image feature in retrieval, a number of experiments were performed on our image retrieval.

To find the optimal maximum order of QPHTs and the number of NSST sub-bands, we randomly selected 500 different images as query images from the COREL dataset to test the performance of the

proposed algorithm. Figure 7a shows the average retrieval precision and average feature extraction time for different maximum orders of QPHTs. Figure 7b plots the average retrieval precision and average feature extraction time for different scales and directions of NSST decomposition. With more shape features and texture features, the performance of image retrieval tends to become better because more information can be represented by the indexing feature space. However, the number of features will increase accordingly, which will inevitably reduce the computational efficiency of image retrieval. In order to achieve better trade-off between the average retrieval precision and average feature extraction time, we choose the maximum order of QPHTs is $n_{max}$=9 and NSST decomposition [29, 40] in the rest of experiments. Figure 8 shows the average retrieval precisions of 500 times query results for different feature weight values $\omega_1$ and $\omega_2$, which reflects the image retrieval efficiency. In Figs. 9 and 10, we demonstrate our retrieval results with shape feature only, texture feature only, and both shape feature and texture feature, respectively. It clearly shows that integrating the results of shape- and texture-based queries provides better retrieval effectiveness than either of the individual feature based queries.

## 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 [20, 44, 47], which retrieval using a combination of several low-level visual features. To simulate the practical situation of online users, we randomly selected 1000 images as query images from the COREL dataset (The tested 10 semantic class includes people, beaches, buildings, buses,
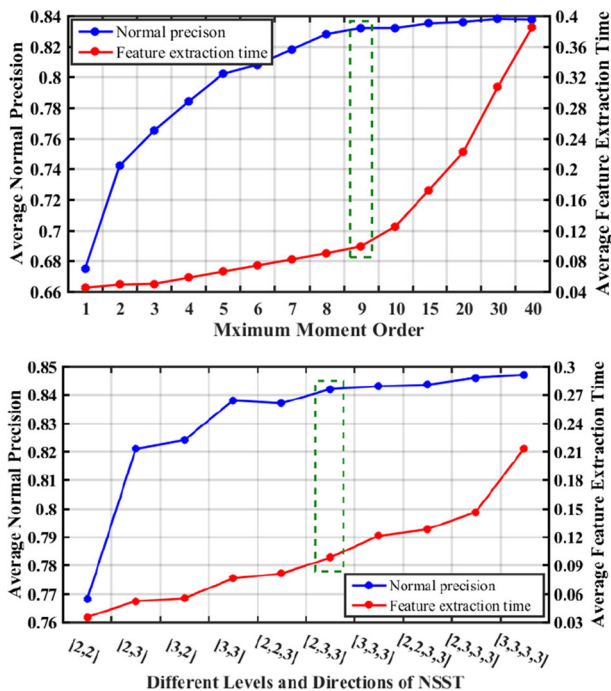


**Fig. 7** The mean of retrieval precisions and average feature extraction time of 500 times query results for **a** different maximum QPHTs order values, **b** different NSST decomposition
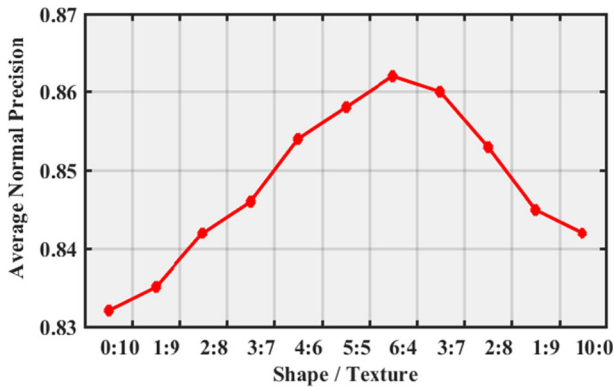
**Fig. 8** The average retrieval precision for different feature weight values

dinosaurs, elephants, flowers, horses, mountains, and foods). Each kind is extracted 20 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 20 times query results are
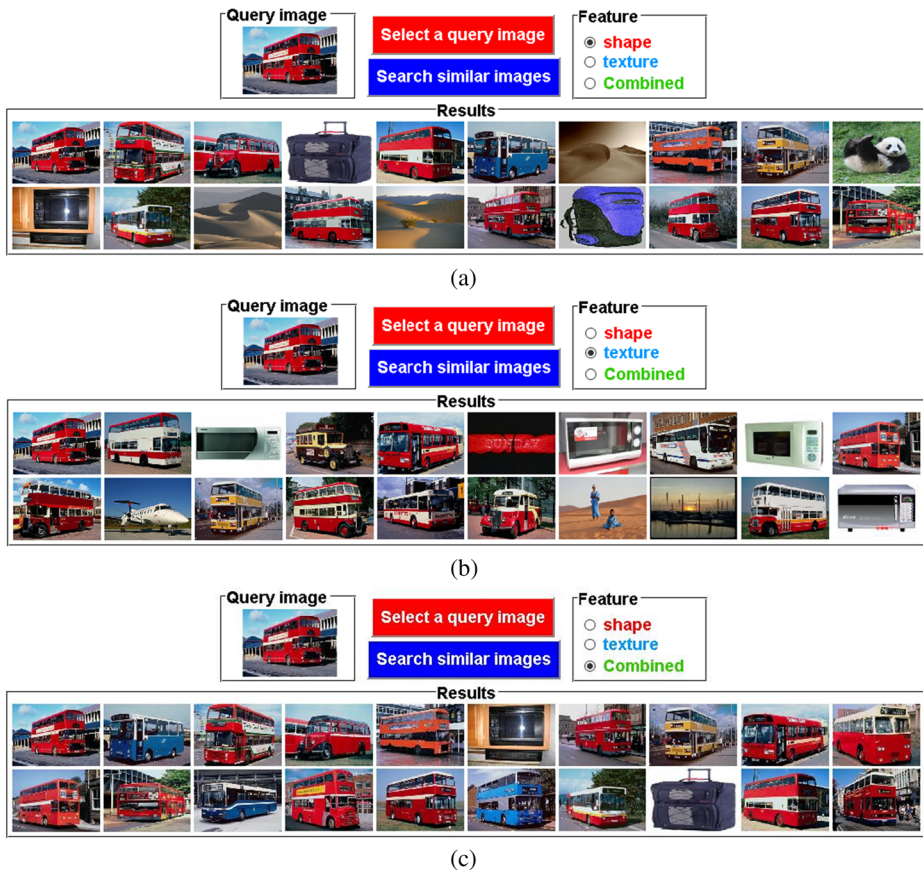


(a)



(b)



(c)

**Fig. 9** Our image retrieval results (Bus): **a** By taking only the shape feature, **b** By taking only the texture feature, **c** By taking both shape feature and texture feature

Fig. 10 Our image retrieval results (Plane): **a** By taking only the shape feature, **b** By taking only the texture feature, **c** By taking both shape feature and texture feature

calculated. These values are taken as the retrieval performance standard of the algorithm, as shown in Fig. 11. According to the Fig. 11, we see that the image retrieval accuracy by the proposed method is competitive with the other methods.

As stated earlier, retrieval efficiency is another parameter to measure the performance of the CBIR system. Efficiency is closely related with the storage requirements and the responsiveness of the system. We examine retrieval efficiency by measuring the indexing time (time taken to extract and store feature vectors from all images in the database) and the response time (time taken by the retrieval system to response to user's query) of above four algorithms. We see from Table 3 that when compared with algorithms [20, 44, 47], our proposed algorithms can achieve a much quicker retrieval in term of both indexing time and response time.

We also compared the proposed methods with current state-of-the-art retrieval pipelines including traditional Local-based methods [5, 10, 18, 25, 27] and CNN-based methods [1, 4, 7, 31, 33, 34] on another three publicly available retrieval datasets, Holidays, Oxford, and UK-bench. For a fair comparison, we only report mAP on representation with relevant dimensions and exclude post-processing methods such as spatial re-ranking or query expansion. The results of retrieval accuracy (mAP) of retrieval accuracy (mAP) of Holidays, Oxford, and UK-bench are shown in Table 4, in which the bold indicate the best results in the
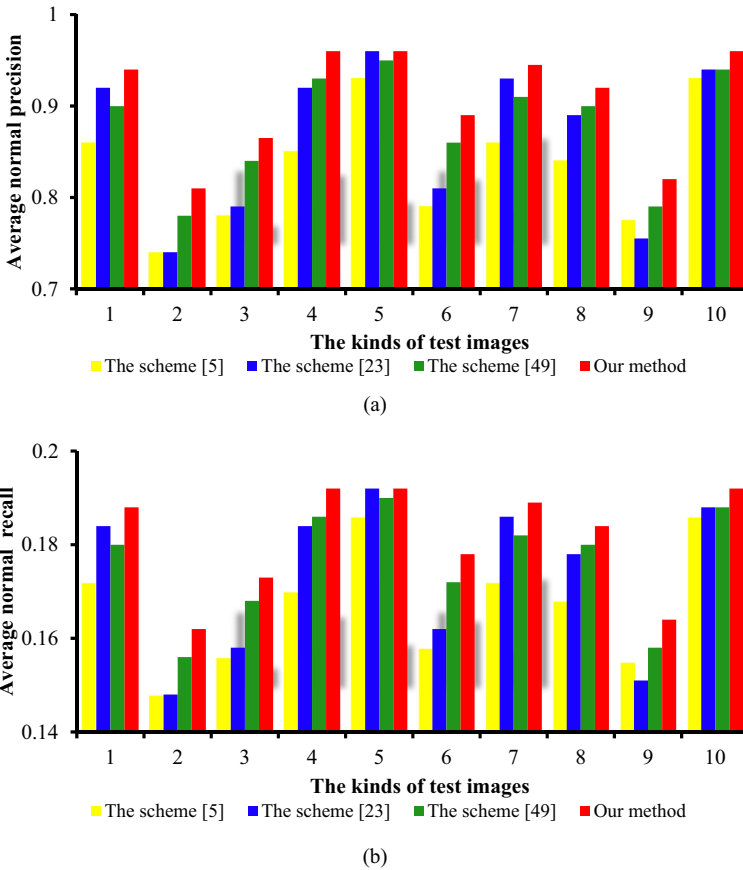
(a)



(b)

**Fig. 11** The average retrieval performance of four algorithms: **a** The average normal precision, **b** The average normal recall

comparison experiments. It is interesting find that our algorithm performs better than all local-based approaches by a large margin but perform worse than the CNN-based approaches by little margin. However, the retrieval efficiency mainly relies on the feature vector length. It is worth noting that the dimension of the feature vector used in or method is significantly lower than that of the CNN-based algorithms. That is to say, although accuracy has decreased slightly, but the retrieval efficiency has been improved under the same experimental conditions, which belongs to a more effective compromise retrieval scheme.

According to the Fig. 11, Tables 2, and 3, we see that the image retrieval accuracy by the proposed method is competitive with the other methods. The effectiveness of the proposed

**Table 3** Retrieval efficiency comparisons with three multiple feature fushion-based algorithms

|  | Wang et al. [47] | Khokher and Talwar [20] | Varish et al. [44] | Proposed algorithm |
|---|---|---|---|---|
| Average indexing time (in seconds) | 0.313 | 0.361 | 0.293 | 0.289 |
| Average response time (in seconds) | 0.245 | 0.307 | 0.222 | 0.209 |

**Table 4** Accuracy comparisons with the state-of-the-art

|  | Method | Feature vector Length | Holidays | Oxford | UK-bench |
|---|---|---|---|---|---|
| Local-based approaches | LCS + RN [10] | | 0.658 | 0.517 | 0.324 |
| | Improved BOW [18] | | 0.747 | 0.546 | 0.352 |
| | VLADintra [5] | | 0.625 | 0.448 | –/– |
| | HVLAD [25] | | 0.721 | 0.638 | 0.356 |
| | Fine-residual VLAD [27] | | 0.622 | –/– | 0.343 |
| CNN-based approaches | Neural Codes [7] | 128 | 0.789 | 0.557 | 3.56 |
| | Spatial pooling [33] | 256 | 0.742 | 0.533 | –/– |
| | OxfordNet [31] | 128 | 0.816 | 0.593 | –/– |
| | NerVLAD-CNN [4] | 256 | 0.860 | 0.635 | 3.69 |
| | CRB-16 [1] | 512 | 0.854 | 0.697 | 3.56 |
| | CRB-M [1] | 512 | 0.809 | 0.582 | 3.40 |
| | Alex +GeM [34] | 256 | 0.787 | 0.601 | –/– |
| | VGG + GeM [34] | 256 | 0.831 | **0.820** | –/– |
| Proposed method | | 95 | 0.812 | 0.676 | 3.56 |

image retrieval results from: (1) Quaternion polar harmonic transforms (QPHTs) coefficients are adopted to depict the image shape, which has many desirable properties such as expression efficiency, robustness to noise, geometric invariance, fast computation, etc.; (2) image texture is represented by BKF parameters of NSST sub-bands, which are robust to illumination and image blurring, and also reduce computational complexity in the texture retrieval phase; (3) QPHTs coefficients and BKF parameters of NSST are combined effectively for image retrieval.

# 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 QPHTs coefficients and BKF parameters in NSST domain. 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.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

# References

1. Alzu'bi A, Amira A, Ramzan N (2017) Content-based image retrieval with compact deep convolutional features. Neurocomputing 249:95–105

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 Geosci Remote Sensing 52(2):3023–3034

4. Arandjelovi R, Gronat P, Torii A, Pajdla T, Sivic J (2016) NetVLAD: CNN architecture for weakly supervised place recognition. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5297–5307

5. Arandjelovic R, Zisserman A (2013) All about VLAD. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1578–1585

6. Atto AM, Berthoumieu Y, Bolon P (2013) 2-D wavelet packet spectrum for texture analysis. IEEE Trans Image Process 22(6):2495–2500

7. Babenko A, Slesarev A, Chigorin A, Lempitsky V (2014) Neural codes for image retrieval. In: European Conference on Computer vision (ECCV). Springer, pp 584–599

8. Chen WT, Liu WC, Chen MS (2010) Adaptive color feature extraction based on image color distributions. IEEE Trans Image Process 19(8):2005–2016

9. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection, vol 1. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, pp 886–893

10. Delhumeau J, Gosselin PH, Jégou H, Pérez P (2013) Revisiting the VLAD image representation. Proceedings of the 21st ACM International Conference on Multimedia, Rennes, pp 653–656

11. Fadili JM, Boubchir L (2005) Analytical form for a Bayesian wavelet estimator of images using the Bessel K form densities. IEEE Trans Image Process 14(2):231–240

12. 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

13. He Z, You X, Yuan Y (2009) Texture image retrieval based on non-tensor product wavelet filter banks. Signal Process 89(8):1501–1510

14. 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

15. Huang J, You X, Yuan Y, Yang F, Lin L (2010) Rotation invariant iris feature extraction using Gaussian Markov random fields with non-separable wavelet. Neurocomputing 73(4–6):883–894

16. Imran M, Hashim R, Khalid NEA (2014) Color histogram and first order statistics for content based image retrieval. In: Recent Advances on Soft CompHeuting and Data Mining, Springer, pp: 153–162

17. Jain V, Sahu N (2013) A survey: on content based image retrieval. Int J Eng Res Appl 3(4):1166–1169

18. Jégou H, Douze M, Schmid C (2010) Improving bag-of-features for large scale image search. Int J Comput Vis 87(3):316–336

19. Jian M, Lam KM (2014) Face-image retrieval based on singular values and potential-field representation. Signal Process 100:9–15

20. Khokher A, Talwar R (2017) A fast and effective image retrieval scheme using color-, texture-, and shape-based histograms. Multimed Tools Appl 76(20):21787–21809

21. Lasmar NE, Berthoumieu Y (2014) Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms. IEEE Trans Image Process 23(5):2246–2261

22. Li X (2003) Image retrieval based on perceptive weighted color blocks. Pattern Recogn Lett 24(12):1935–1941

23. Li S, Lee MC, Pun CM (2009) Complex Zernike moments features for shape-based image retrieval. IEEE Trans Syst, Man Cybern, Part A: Systems Humans 39(1):227–237

24. Lim WQ (2010) The discrete shearlet transform: A new directional transform and compactly supported shearlet frames. IEEE Trans Image Process 19(5):1166–1180

25. Liu Z, Li H, Zhou W, Rui T, Tian Q (2015) Uniforming residual vector distribution for distinctive image representation. IEEE Trans Circuits Syst Video Technol 99:1

26. Liu M, Vemuti BC, Amari SI, Nielsen F (2012) Shape retrieval using hierarchical total Bregman soft clustering. IEEE Trans Pattern Anal Mach Intell 34(12):2407–2419

27. Liu Z, Wang S, Tian Q (2016) Fine-residual VLAD for image retrieval. Neurocomputing 173:1183–1191

28. Liu GH, Yang JY (2013) Content-based image retrieval using color difference histogram. Pattern Recogn 46(1):188–198

29. Liu Y, Zhang DS, Lu GJ, Ma WY (2007) A survey of content-based image retrieval with high-level semantics. Pattern Recogn 40(1):262–282

30. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110

31. Ng J, Yang F, Davis L S (2015) Exploiting local features from deep networks for image retrieval. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 53–61

32. Park U, Park J, Jian AK (2014) Robust keypoint detection using higher-order scale space derivatives: application to image retrieval. IEEE Signal Process Lett 21(8):962–965

33. Paulin M, Douze M, Harchaoui Z, Mairal J, Perronnin F, Schmid C (2015) Local convolutional features with unsupervised training for image retrieval. Proceedings of the IEEE International Conference on Computer Vision, pp 91–99

34. Radenović F, Tolias G, Chum O (2017) Fine-tuning CNN Image Retrieval with No Human Annotation. arXiv preprint arXiv:1711.02512

35. Rakvongthai Y, Oraintara S (2013) Statistical texture retrieval in noise using complex wavelet. Signal Process Image Commun 28(10):1494–1505

36. 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):727–729

37. Shahdoosti HR, Khayat O (2016) Image denoising using sparse representation classification and non-subsampled shearlet transform. Signal, Image Video Process 10(6):1081–1087

38. Shu X, Wu XJ (2011) A novel contour descriptor for 2D shape matching and its applications to image retrieval. Image Vis Comput 29(4):286–294

39. 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

40. Smeulders AWM, Worring M, Santini S, Gupta A (2000) Content-based image retrieval at the end of the early years. IEEE Trans Pattern Anal Mach Intell 22(12):1349–1380

41. 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

42. The USC-SIPI Image Database. http://sipi.usc.edu/services/database/Database.html

43. Van De Sande K, Gevers T, Snoek C (2010) Evaluating color descriptors for object and scene recognition. IEEE Trans Pattern Anal Mach Intell 32(9):1582–1596

44. Varish N, Pradhan J, Pal AK (2017) Image retrieval based on non-uniform bins of color histogram and dual tree complex wavelet transform. Multimedia Tools & Applications 76(14):1–37

45. Vogel J, Schiele B (2006) Performance evaluation and optimization for content-based image retrieval. Pattern Recogn 39(5):897–909

46. Wan J, Wang D, SCH H, Wu PC (2014) Deep learning for content-based image retrieval: a comprehensive study. Proceedings of the 22nd ACM international Conference on Multimedia, Orlando, pp 157–166

47. Wang XY, Liang LL, Li YW, Yang HY (2016) Image retrieval based on exponent moments descriptor and localized angular phase histogram. Multimed Tools Appl 76(6):7633–7659

48. Wang XY, Liu YN, Li S (2016) Robust image watermarking approach using polar harmonic transforms based geometric correction. Neurocomputing 174:627–642

49. Wang XY, Liu YN, Xu H, Wang AL (2016) Blind optimum detector for robust image watermarking in nonsubsampled shearlet Domain. Inf Sci 372:634–654

50. Wang CP, Wang XY, Li YW, Xia ZQ, Zhang C (2018) Quaternion polar harmonic Fourier moments for color images. Inf Sci 450:141–156

51. Wang CP, Wang XY, Zhang C, Xia ZQ (2016) Geometrically invariant image watermarking based on fast Radial Harmonic Fourier Moments. Signal Process-Image Commun 45:10–23

52. Wang CP, Wang XY, Zhang C, Xia ZQ (2017) Geometric correction based color image watermarking using fuzzy least squares support vector machine and Bessel K form distribution. Signal Process 134:197–208

53. Wang XY, Wu JF, Yang HY (2010) Robust image retrieval based on color histogram of local feature regions. Multimed Tools Appl 49(2):323–345

54. Wang XY, Yu YJ, Yang HY (2011) An effective image retrieval scheme using color, texture and shape features. Comput Stand Interfaces 33(1):59–68

55. Xie L, Shen J, Zhu L (2016) Online cross-modal hashing for web image retrieval. In: Proc. AAAI Conf. Artif. Intell, pp. 294–300

56. Yap PT, Jiang X, Kot AC (2010) Two-dimensional polar harmonic transforms for invariant image representation. IEEE Trans Pattern Anal Mach Intell 32(7):1259–1270

57. Yap PT, Paramesran R (2006) Content-based image retrieval using Legendre chromaticity distribution moments. IEE Proc-Vis, Image. Signal Process 153(1):17–24

58. Yu J, Liu DQ, Tao DC, Seah HS (2012) On combining multiple features for cartoon character retrieval and clip synthesis. IEEE Trans Syst, Man Cybern, Part B: Cybern 42(5):1413–1427

59. Zhao ZJ, Tian Q, Sun HD, Guo JX (2016) Content based image retrieval scheme using color, texture and shape features. Int J Signal Process Image Process. Pattern Recogn 9(1):203–212

60. Zhu L, Shen J, Xie L, Cheng Z (2017) Unsupervised topic hypergraph hashing for efficient mobile image retrieval. IEEE Transactions on Cybernetics 47(11):3941–3954

61. Zhu Z, You X, Chen CLP, Tao D, Ou W, Jiang X, Zou J (2015) An adaptive hybrid pattern for noise-robust texture analysis. Pattern Recogn 48:2592–2608

**Yu-Nan Liu** received his B.E. and M.S.E degrees from the school of Computer and Information Technology, Liaoning Normal University, China, in 2014 and 2017 respectively. Currently, he is pursuing his Ph.D. degree in Computer Science and Technology from Nanjing University of Science and Technology, Nanjing, China. His research interests include visual information retrieval, pattern recognition, digital watermarking, and forgery detection.



**Shan-Shan Zhang** received the master's degree in signal and information processing from Tongji University in 2011, and the PhD degree in computer Science from University of Bonn in 2015. From Jan. 2015 until Dec. 2016, she was a postdoc researcher at the department of Computer Vision and Multimodal Computing, Max Planck Institute for Informatics, Germany. Since Dec. 2016, she has been a professor at Nanjing University of Science and Technology, China. Her main research interests include computer vision and pattern recognition and their applications in intelligent transportation systems.

**Yu Sang** received the Master's degree and the PhD degree in computer science and technology from Liaoning Normal University, China, and Dalian University of Technology, China, in 2008 and 2012 respectively. From Apr. 2015 to Dec. 2017, he was a postdoc researcher at Postdoctoral Program, Dalian University of Technology, China, and Postdoctoral Workstation, Liaohe Petroleum Exploitation Bureau, China. Currently, he has been an associate professor at Liaoning Technical University. His main research interests include Artificial Intelligent, Computer Vision, and Seismic Signals Processing.



**Si-Miao Wang** received her B.E. and M.S.E degrees from the school of Computer and Information Technology, Liaoning Normal University, China, in 2014 and 2017 respectively. Her research interests include visual information retrieval, pattern recognition, and digital watermarking.