Int J Comput Vis (2012) 98:187–201
DOI 10.1007/s11263-011-0503-6

Rotation Invariant Curvelet Features for Region Based Image Retrieval

Dengsheng Zhang · M. Monirul Islam · Guojun Lu · Ishrat Jahan Sumana

Received: 6 November 2009 / Accepted: 10 October 2011 / Published online: 27 October 2011

Abstract There have been much interest and a large amount of research on content based image retrieval (CBIR) in recent years due to the ever increasing number of digital images. Texture features play a key role in CBIR. Many texture features exist in literature, however, most of them are neither rotation invariant nor robust to scale and other variations. Texture features based on Gabor filters have been shown with significant advantages over other methods, and they are adopted by MPEG-7 as one of the texture descriptors for image retrieval. In this paper, we propose a rotation invariant curvelet features for texture representation. With systematic analysis and rigorous experiments, we show that the proposed curvelet texture features significantly outperforms the widely used Gabor texture features. A novel region padding method is also proposed to apply curvelet transform to region based image retrieval. Retrieval results from standard image databases show that curvelet features are promising for both texture and region representation.

Keywords Texture · Image retrieval · Curvelet transform · CBIR · Gabor filters

D. Zhang (✉) · M.M. Islam · G. Lu · I.J. Sumana
Gippsland School of Information Technology, Monash University, Churchill, Victoria 3842, Australia
e-mail: Dengsheng.Zhang@monash.edu

M.M. Islam
e-mail: Md.Monirul.Islam@monash.edu

G. Lu
e-mail: Guojun.Lu@monash.edu

I.J. Sumana
e-mail: Ishrat.Sumana@monash.edu

1 Introduction

In recent years, there has been a tremendous interest in image retrieval, especially in content based image retrieval (CBIR) due to the ever increasing amount of digital images (Long et al. 2003; Zhang and Lu 2004; Datta et al. 2008; Lew et al. 2006; Vasconcelos 2007; Liu et al. 2007; Manjunath et al. 2002). There are several reasons for the wide interest in CBIR . First, the number of images is simply too large to do the traditional text annotations by humans. Second, text annotations are ambiguous due to subjectivity. Many CBIR methods have been proposed in literature (Tamura et al. 1978; Niblack et al. 1993; Huang et al. 1997; Chaudhuri and Sarkar 1995; Ma and Manjunath 1995; Liu and Picard 1996; Manjunath and Ma 1996; Zhang and Lu 2000; Chen et al. 2004; Vertan and Boujemaa 2000; Manjunath et al. 2001; Ngo et al. 2001; Liu et al. 2008; Inoue 2004; Wang et al. 2001, 2006, 2007, 2008a, 2008b; Ferecatu and Boujemaa 2007; Jeon et al. 2003; Duygulu et al. 2002; Hervé and Boujemaa 2007; Lu et al. 2006, 2008; Howarth and Ruger 2004; Do and Vetterli 2002; Bhagavathy and Chhabra 2007; Suematsu et al. 2002). In all these methods, feature extraction constitutes the key component of the system. Ideally, the extracted features should agree with human perception on images. Several perceptual features have been used in literature, such as color, shape and texture. Color features are well defined, however, colors have severe limitations to identify and distinguish image objects. For example, clouds of different colors may be regarded as different objects using color features. Similarly, sky and ocean may have the same color features. Shape features are usually applied to applications where objects are well defined, such as industry object recognition, trade mark retrieval, graphical design etc. (Zhang and Lu 2004). In color image retrieval, shape features are usually modelled together

with texture features using edge information although some simple shape features can be used alone (Jeon et al. 2003; Duygulu et al. 2002). Texture features are very effective and useful, and have been used virtually in all the CBIR systems. The importance of using texture features is that texture features can differentiate objects with similar colors and shapes. Compared with colors and shapes, texture features are more difficult to model. It is known that human has certain perceptual intuitions to judge textured objects, such as, smooth/rough, regular/random, directional/chaotic, etc. In practice, these texture characteristics can be modelled using the edge singularities or corner singularities embedded in images.

The three types of features (color, shape and texture) are usually designed and tested separately and then combined together in natural image retrieval. In this paper, we focus on texture feature extraction. A variety of techniques have been developed for extracting texture features. These techniques can be broadly classified into spatial texture feature extraction methods and spectral texture feature extraction methods.

1.1 Spatial Texture Feature Extraction Methods

In spatial approach, texture features are extracted by computing the pixel statistics or finding the local pixel structures in original image. These methods include the Tamura textures (Long et al. 2003; Tamura et al. 1978), co-occurrence matrix method (Long et al. 2003; Howarth and Ruger 2004), Markov random field (MRF) method (Long et al. 2003; Liu and Picard 1996; Manjunath and Ma 1996), and fractal dimension (FD) (Chaudhuri and Sarkar 1995) method. Tamura et al. are among the earliest researchers to formally define texture features (Tamura et al. 1978). The mostly cited Tamura texture features in literature consist of six perceptual characteristics of images such as the degree of contrast, coarseness, directionality, linearity, roughness and regularity. In most of the cases, only the first three Tamura features are used as the other three features are defined based on the combinations of the first three features. Tamura features are appealing because they are high level perceptual features and suitable for texture browsing. However, it is difficult to define many types of high level features. Therefore, Tamura features are not enough to distinguish all the textures in the world.

Many statistical texture features are based on grey level co-occurrence matrices (GLCM) (Long et al. 2003) or its color counterpart of color correlogram (Huang et al. 1997). A GLCM represents how frequent every particular pair of grey levels is in the pixel pairs, separated by a certain distance d along a certain direction a. The matrices are statistics of a Markov random field with multiple pairwise pixel interactions. Various statistics can be computed from the GLCM, such as contrast, energy, entropy, homogeneity, and periodicity. However, GLCM features are very expensive to compute, and they are not efficient for image classification and retrieval. MRF texture methods model an image pixel location as a random variable, as a result, an image is considered as a random field. Each type of textures is characterised by a joint probability distribution of signals that accounts for spatial inter-dependence, or interaction among the signals. The interacting pixel pairs are usually called neighbours, and a random field texture model is characterised by geometric structure and quantitative strength of interactions among the neighbours.

Among the many MRF texture methods, the Simultaneous Auto-Regressive (SAR) model is the most widely used, as it uses fewer parameters (Long et al. 2003). In SAR, a pixel value is estimated as a linear combination of the neighboring pixel values and an additive Gaussian noise. Parameters of the SAR model can be estimated using the least square error (LSE) technique or the maximum likelihood estimation (MLE). Rotation-invariant SAR model (RISAR) has been proposed to capture textures of different orientations, where neighboring pixels are defined on circles centered at the target pixel. In order to make SAR more robust, multiresolution MRF (MRMRF) has also been proposed (Liu and Picard 1996), where an image is represented by a multi-resolution Gaussian pyramid before applying the MRF model. The basic idea of MRF is to find if there are any regular structures in an image, consequently, MRF model is not good at distinguishing the various differences between regular and random textures. The second problem with the MRF methods is how to find the adequate neighbourhood, and this nontrivial problem has no general solution. Finally, the computation of MRF features is prohibitively expensive (Liu and Picard 1996).

The FD method (Chaudhuri and Sarkar 1995) is based on the theory of fractal geometry which characterises shapes or patterns of self-similarity. It attempts to find the smallest structure which replicates the whole pattern. In practice, FD method models a grey level image as a 3D terrain surface, and a differential box counting is conducted under the surface to measure how rough the surface is. As the logarithmic number of boxes and the logarithmic box size have a linear relationship, FD can be estimated from the least-square fit of the two variables. As FD only models the roughness feature, other features like directionality and contrast are missed from FD. Therefore, in Chaudhuri and Sarkar (1995), six FDs have to be computed from a number of modified images. Furthermore, FD is not rotation invariant.

Spatial texture methods are easy to understand, many of them even have perceptual meanings. However, many of those methods involve complex search and optimisation processes which have no general solutions. Furthermore, spatial texture methods are easily affected by noise which is a natural component of digital images. These problems explain

why spatial methods usually give poor performance in most applications. It is not surprising that most recent works in literature employ spectral texture methods.

1.2 Spectral Texture Feature Extraction Methods

In spectral texture methods, images are transformed into frequency domain using certain spatial filter bank. Texture features are then extracted from the transformed spectra using statistics. Due to the large neighbourhood support of the filters, spectral methods are very robust to noise. Because the use of FFT in the implementation process, spectral methods are also much more efficient to compute than most of the spatial methods. Furthermore, spectral methods can generate sufficient number of features to distinguish variety of texture images. Common spectral methods include Fourier transform (FT) (Hervé and Boujemaa 2007), discrete cosine transform (DCT) (Ngo et al. 2001; Lu et al. 2006), Wold texture (Long et al. 2003; Liu and Picard 1996), wavelet transform (Datta et al. 2008; Wang et al. 2001; Do and Vetterli 2002; Bhagavathy and Chhabra 2007; Suematsu et al. 2002) and Gabor filters (Manjunath and Ma 1996; Zhang and Lu 2000; Liu et al. 2007).

FT method is used in Hervé and Boujemaa (2007). Two histograms are computed from the FT spectra, one with the circular partition and the other with the wedged partition. Although FT is a powerful image analysis tool, it can only capture global features which are not sufficient for texture analysis. Ngo et al. (2001) arrange the low order DCT coefficients into Mandala space which represents the partial derivatives of the original image. Texture features are computed from the variance of each of the derivative images. DCT method is also used by Lu et al. (2006), but they use mean and standard deviation as the texture features instead. Similar to FT, DCT method can only capture the global features.

Liu and Picard (1996) attempt to compute perceptual texture features similar to the Tamura features such as periodicity, directionality and randomness, but use spectral method. In this method, an image autocovariance map is first computed through inverse FT, an autocovariance energy ratio is then computed for each image to determine if the image has harmonic structure or random structure. If the image has harmonic structure, the periodicity feature is extracted by finding the harmonic peaks from the FT spectra and the directionality feature is extracted by applying the Hough transform on the FT spectra. If the image has random structure, the randomness features are extracted from the estimated MRSAR coefficients. Finally, images are ranked using both types of features and a probabilistic weighting is used to determine the joint rank. In this approach, arbitrary thresholding and weighing are used in the FT peak searching, matching and the energy ratio calculation. Complex computation is needed in both the harmonic peak searching and MRMSAR coefficient estimation. Furthermore, the retrieval is also expensive as complex peak matching and quadratic distance matching for the high dimensional MRSAR features are needed during the retrieval stage. All these constraints can make the Wold features impractical when applied to real application data.

Because of the limitations of FT in capturing local image features, space-frequency analysis methods have been introduced, such as wavelet method (Long et al. 2003; Ma and Manjunath 1995; Wang et al. 2001; Do and Vetterli 2002; Bhagavathy and Chhabra 2007; Suematsu et al. 2002) and Gabor filters method (Manjunath and Ma 1996; Zhang and Lu 2000; Chen et al. 2004). In wavelet method, images are decomposed into different frequency components using filters at different scales. Texture features are then extracted from each of the frequency components. Both the pyramid-structured wavelet transform (PWT) and tree-structured wavelet transform (TWT) can be used. Although the TWT method gives a slight better result, the difference is not significant (Ma and Manjunath 1995; Manjunath and Ma 1996). The wavelet method shows significant advantage over FT method as it captures local spectral features at multiple resolutions. However, the problem with wavelet method is that wavelets are usually sensitive to point singularities instead of edge singularities which are crucial for extracting texture features. The other problem with wavelet method is that it does not adapt to directional textures.

Gabor filters are a set of Gaussian shaped wavelets tuned to different orientations. As a result, it can capture texture features at multiple orientations and multiple scales. Another advantage of Gabor filters over wavelets is that they achieve the optimal localization in spatial and frequency domain. Daugman made an intensive investigation on Gabor filters (Daugman 1985, 1988, 1989, 1993), and showed that 2D Gabor filters best describe the 2D receptive-field profiles of simple cells found in the visual cortex of mammal animals. Manjunath and Ma systematically introduced Gabor texture feature into CBIR and made a comprehensive evaluation of Gabor feature with other major texture methods (Manjunath and Ma 1996). They showed that Gabor texture feature significantly outperforms those other major texture methods. Gabor texture feature has since been adopted by MPEG-7 as texture descriptor (Manjunath et al. 2002). Zhang et al. proposed a method to normalise Gabor features into rotation invariant texture features (Zhang and Lu 2000). So far, Gabor feature is the best texture feature in literature and has been used by most of the recent CBIR systems. However, Gabor filters also have limitations. First, the Gaussian shaped Gabor filters still emphasize point singularities instead of edge singularities. Second, due to the Gaussian shape frequency response, Gabor filters have significant overlap in the frequency plane. Consequently, Gabor filters output are not mutually orthogonal, which may result in a

significant correlation between texture features (Arivazhagan et al. 2006).

Recently, a new wavelet like method called curvelet has been introduced to overcome the limitation of wavelet (Do and Vetterli 2003; Do 2001; Starck et al. 2002; Candes et al. 2006). Curvelets are a special set of wavelets which are designed to adapt to curved edges in images. With curvelet, image edge information is captured at different orientations and scales. It appears to take the advantages of both wavelet and Gabor filters, and has been successfully used in image denoising and enhancement. There are also a few early applications in image classification (Joutel et al. 2007a, 2007b; Arivazhagan et al. 2006; Semler and Dettori 2006) and retrieval (Shekhar and Chaudhuri 2005). However, none of them has considered rotation invariance and region based image retrieval. Joutel et al. (2007a, 2007b) have created an assistance tool for the identification of ancient handwritten manuscripts using ridgelet transform. However, only curvature and orientation features are extracted which are not sufficient for image retrieval. Arivazhagan et al. (2006) and Shekhar et al. (2005) used curvelet features for color image classification and retrieval. However, color images are not homogenous and they cannot be classified without segmentation. Semler and Dettori (2006) applied curvelet in medical domain and used curvelet texture features on cropped CT images for organ classification.

During the development of curvelet transform, two versions of curvelet transform have been implemented, namely the projection-slice based version and the wrapping based version (or the first generation curvelet transform and the second generation curvelet transform) (Semler and Dettori 2006). It has been shown that the wrapping based curvelet transform not only is much more efficient but also gives better performance. All the works of Joutel et al. (2007a, 2007b), Arivazhagan et al. (2006) and Shekhar and Chaudhuri (2005) used the first generation curvelet transform in their applications, and none of the features extracted is rotation invariant. Sumana et al. (2008), show how to compute the second generation curvelet features with preliminary retrieval result from a small dataset. So far, there is no wide use of curvelet features in CBIR systems and there is no systematic/comprehensive evaluation of curvelet method using large standard datasets. In this paper, we first formally introduce efficient curvelet features for image retrieval and systematically analyse the curvelet method along with other spectral methods such as wavelet and Gabor filters. Next, a rotation invariant method is proposed to normalize the curvelet features. We then propose a novel padding method to apply curvelet features on region based color image retrieval. Rigorous performance tests are carried out to evaluate curvelet features against state of the art texture features in literature.

The rest of the paper is organized as following. In Sect. 2, we describe the concept of curvelet transform in details and compare curvelet method with wavelet method and Gabor filters. In Sect. 3, rotation invariant curvelet features are described in details. In Sect. 4, we test and compare the retrieval performance of curvelet features with both Gabor and wavelet features using standard texture databases. In Sect. 5, a novel region padding method is proposed to apply curvelet transform in region based image retrieval. Section 6 concludes the paper and indicates future research direction.

2 Description of Curvelet Transform

Basically, curvelet transform extends the ridgelet transform to multiple scale analysis. Therefore, let's start from the definition of ridgelet transform. Given an image function $f(x, y)$, the continuous ridgelet transform is given as (Starck et al. 2002):

$$\Re_f(a, b, \theta) = \iint \psi_{a,b,\theta}(x, y) f(x, y) dx dy \tag{1}$$

where $a > 0$ is the scale, $b \in R$ is the translation and $\theta \in [0, 2\pi)$ is the orientation. The ridgelet is defined as:

$$\psi_{a,b,\theta}(x, y) = a^{-\frac{1}{2}} \psi\left(\frac{x\cos\theta + y\sin\theta - b}{a}\right) \tag{2}$$

Figure 1(a) shows a typical ridgelet. It is oriented at an angle θ, and is constant along lines: $x \cdot \cos\theta + y \cdot \sin\theta = constant$ (Do and Vetterli 2003). Along the orthogonal direction it is a wavelet. It can be seen that a ridgelet is linear in edge direction and is much sharper than a conventional sinusoidal wavelet (Fig. 1(b)).

For comparison, the 2-D wavelet is given as following (Fig. 1(b)):

$$\psi_{a_1,a_2,b_1,b_2}(x, y) = a_1^{-\frac{1}{2}} a_2^{-\frac{1}{2}} \psi\left(\frac{x - b_1}{a_1}\right) \psi\left(\frac{y - b_2}{a_2}\right) \tag{3}$$

As can be seen, the ridgelet is similar to the 2-D wavelet except that the point parameters (b_1, b_2) are replaced by the line parameters (b, θ). In other words, the two transforms are related by Do and Vetterli (2003):

Wavelet: $\psi_{scale,\ point\text{-}position}$
Ridgelet: $\psi_{scale,\ line\text{-}position}$

In contrast, Gabor filters are Gaussian shaped wavelets tuned to different orientations and scales.

$$g_{a,\theta,b_1,b_2}(x, y) = a^{-\frac{1}{2}} g\left(\frac{x\cos\theta + y\sin\theta - b_1}{a}, \frac{-x\sin\theta + y\cos\theta - b_2}{a}\right) \tag{4}$$

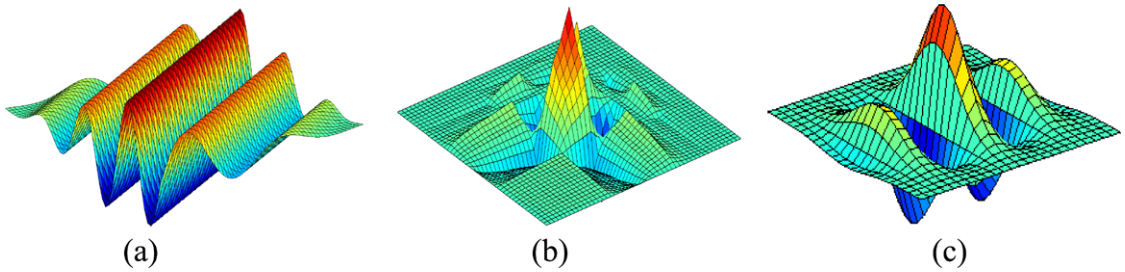


Fig. 1 (**a**) A ridgelet; (**b**) A Daubechies order 2 wavelet; (**c**) A Gabor filter

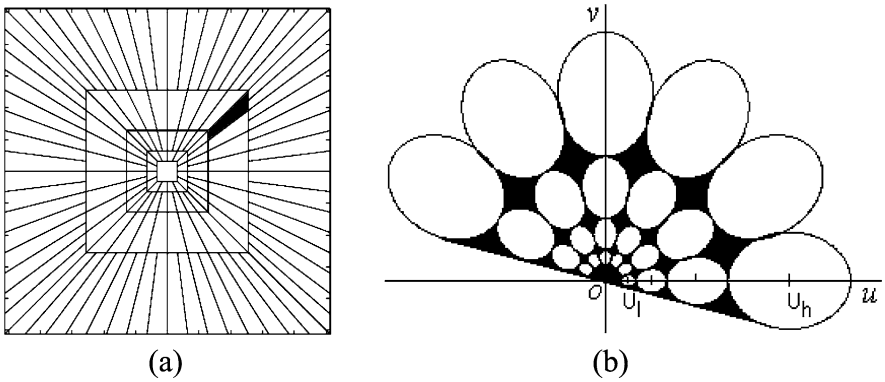


Fig. 2 (**a**) The tiling of frequency plane by curvelets (Candes et al. 2006); (**b**) The tiling of half frequency plane by Gabor filters, the ovals are the covered spectrum and the black areas are the lost frequency information

where the mother wavelet $g(x, y)$ is a Gaussian envelope modulated by a sinusoid wave (Fig. 1(c)):

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right]\cdot\exp(j2\pi Wx) \tag{5}$$

Ridgelet can also be tuned to different orientations and different scales to create the curvelets, in the similar way to Gabor filters. But different from Gabor filters which only cover part of the spectrum in the frequency domain (Manjunath and Ma 1996), curvelets have a complete cover of the spectrum in frequency domain. This is because the frequency response of a curvelet is a wedge which can seamlessly cover the frequency plane, while the frequency response of Gabor filter is an oval which cannot cover the frequency plane without leaving holes. That means, there is no loss of information in curvelet transform in terms of capturing the frequency information from the images. Figure 2(a) shows the curvelet tiling and cover of the frequency plane with 5 scales (Candes et al. 2006). The shaded wedge shows the frequency response of a curvelet at orientation 4 and scale 4. It can be seen, the spectrum cover by curvelets is complete. In contrast, there are many holes in the frequency plane of Gabor filters (Fig. 2(b)). The loss of frequency information by Gabor filters is significant across the spectrum.

3 Rotation Invariant Curvelet Feature Extraction

3.1 Curvelet Feature Extraction

The discrete curvelet transform is taken on a 2-D Cartesian grid $f[m, n]$, $0 \le m < M$, $0 \le n < N$,

$$C^D(a, b, \theta) = \sum_{\substack{0\le m<M\\ 0\le n<N}} f[m, n]\psi^D_{a,b,\theta}[m, n] \tag{6}$$

$$CT^D(a, b, \theta) = IFFT(FFT(f[m, n]) \times FFT(\psi^D_{a,b,\theta}[m, n])) \tag{7}$$

The discrete curvelet transform is implemented using the fast discrete curvelet transform. Basically, it is computed in the spectral domain to employ the advantage of FFT as shown in (7). Given an image, both the image and the curvelet are transformed into Fourier domain, then the convolution of the curvelet with the image in spatial domain becomes the product in Fourier domain. Finally, the curvelet coefficients are obtained by applying inverse Fourier transform on the spectral product. But due to the frequency response of a curvelet is a non-rectangular wedge, the wedge needs to be wrapped into a rectangle to perform the inverse Fourier transform. The wrapping is done by periodic tiling of the wedge spectrum in frequency plane, and then collecting the rectangular coefficient area in the centre

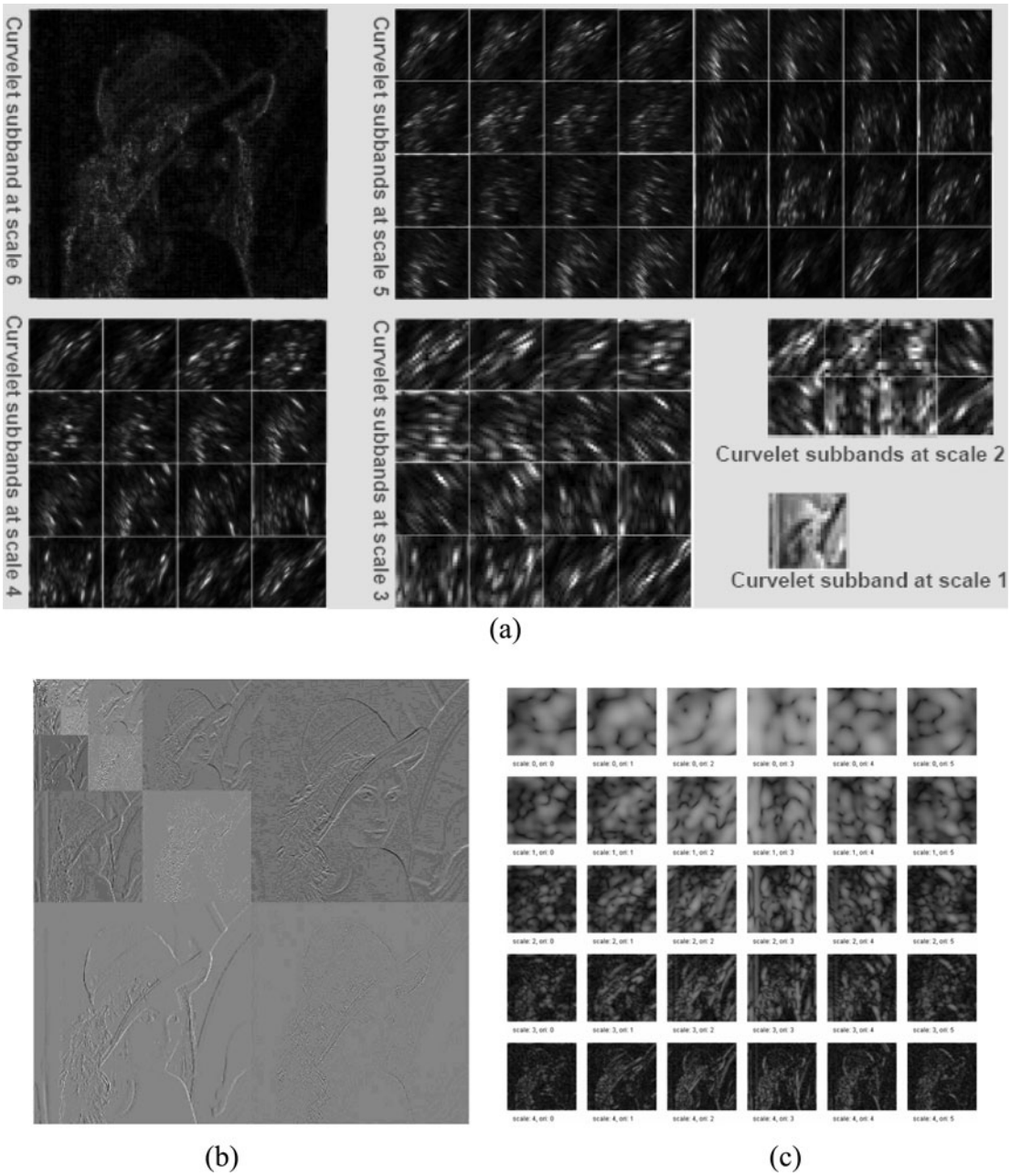


Fig. 3 (**a**) Curvelet spectra; (**b**) Wavelet spectra; (**c**) Gabor filters spectra

(Candes et al. 2006; Zhang and Lu 2003). Through this periodic tiling, the rectangular region collects the wedge's corresponding portions from the surrounding periodic wedges.

The subband spectra of curvelet transform on the Lena image is shown in Fig. 3(a). For contrast, the subband spectra of wavelet transform and Gabor filters are also shown in Fig. 3(b) and (c). It can be seen from the subband spectra of the three space-frequency methods that curvelets capture just the right texture (edge) information as needed. Both wavelet and Gabor filters capture significant amount of redundant information, this reduces their discriminative capability in image retrieval. Wavelet is the least accurate in terms of picking up the high frequency texture information.

Once the curvelet transform is applied and the coefficients are obtained at each scale and orientation, the mean μ and standard deviation σ are computed from each of the subbands as following.

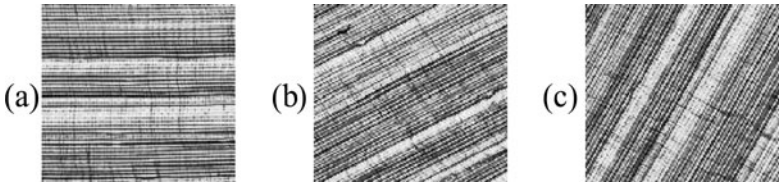


Fig. 4 Three similar textures with rotation of 0, 30 and 60 degrees respectively

$$\mu_{s\theta} = \frac{E(s,\theta)}{m \times n}$$
$$\sigma_{s\theta} = \frac{\sqrt{\sum_x \sum_y (|C_{s\theta}(x,y)| - \mu_{s\theta})^2}}{m \times n} \tag{8}$$

where s is the scale, θ is the orientation, m and n are the dimensions of the corresponding subband, and $E(s,\theta) = \sum_x \sum_y |C_{s\theta}(x,y)|$ is the total spectral energy of the subband.

Therefore, for each curvelet, two texture features are obtained. If l curvelets are used for the transform, $2l$ texture features are obtained. Thus, a $2l$ dimensional texture feature vector is used to represent each image in the database for image retrieval. To mitigate the dynamic range of the spectral energy, both the mean and standard deviation features are normalised using the maximum values of the corresponding features in the database.

Based on the curvelet subband division in Fig. 2(a), with 4 levels curvelet decomposition, 50 $(= 1 + 16 + 32 + 1)$ subbands of curvelet coefficients are computed. However, curvelet at angle θ produces the same coefficients as curvelet at angle $\theta + \pi$. Therefore, half of the subbands at scale 2 and 3 are discarded due to this symmetry. As a result, 26 $(= 1 + 8 + 16 + 1)$ subbands are preserved, and a 52 dimensional feature vector is generated for each image in the database. Similarly, for 5 levels curvelet decomposition, 42 $(= 1 + 8 + 16 + 16 + 1)$ subbands of curvelet coefficients are computed. Therefore, an 84 dimension feature vector is generated for each image in the database.

3.2 Rotation Normalisation

The curvelet features acquired this way are not rotation invariant. When texture images are rotated, their feature vectors differ significantly. Consider the three texture images in Fig. 4, they are the same texture at different orientation. However, Table 1 shows that the feature vectors of these images are quite different in their natural orders.

By analysing the feature patterns of the three feature vectors in Table 1, we observe that the feature values at each scale are circularly shifted when the textures are rotated. It is also observed, for rotated texture patterns, the spectral energy at each scale is concentrated at the dominant directions, indicated as bold numbers in the table. Based on these observations, we apply circular shift at each scale to shift the highest spectral energy to the first position and the second highest spectral energy to the second position, so that the feature patterns at each scale are preserved for rotated images. Using scale 2 section in Table 1 as an example, the scale 2 energy values of the three transformed images are respectively:

(1.312077, 3.104991, **4.112442**, 1.969191, 1.710743, 2.022882, 2.135686, 2.628652)
(**4.214702**, 3.32811, 1.325157, 2.838665, 2.307907, 2.093898, 2.019883, 2.142651)
(2.268232, 1.970233, 2.077396, 2.372323, 2.715669, 1.387268, 3.537925, **4.400911**)

Table 1 Curvelet subband energy distribution of the three images in Fig. 4

Scale (s)	Orientation (θ)	Fig. 4(a)	Fig. 4(b)	Fig. 4(c)
s_1	1	6.612492	6.734992	6.773879
s_2	1	1.312077	**4.214702**	2.268232
	2	***3.104991***	***3.32811***	1.970233
	3	**4.112442**	1.325157	2.077396
	4	1.969191	2.838665	2.372323
	5	1.710743	2.307907	2.715669
	6	2.022882	2.093898	1.387268
	7	2.135686	2.019883	***3.537925***
	8	2.628652	2.142651	**4.400911**
s_3	1	1.777388	2.274384	1.443845
	2	1.061707	**5.339085**	1.467926
	3	1.329857	***3.289273***	1.385565
	4	***2.642727***	1.615015	1.797558
	5	**5.142773**	1.303992	1.333557
	6	1.876816	1.900946	1.492066
	7	1.344805	2.21836	1.221638
	8	1.367682	1.59207	1.234507
	9	1.27489	1.271694	1.483058
	10	1.539031	1.168736	2.340187
	11	1.35263	1.370046	1.874732
	12	1.369954	1.351483	1.149372
	13	1.198996	1.684908	1.43349
	14	1.209949	1.420957	***2.780489***
	15	1.282413	1.670626	**5.306259**
	16	2.281955	1.704048	1.931466
s_4	1	2.726421	2.83314	2.834186

After shifting the highest value to the first position, the second highest value to the second position, and circularly shifting other values accordingly, the normalised scale 2 energy of the three images are respectively:

(**4.112442**, 3.104991, 1.312077, 2.628652, 2.135686, 2.022882, 1.710743, 1.969191)
(**4.214702**, 3.32811, 1.325157, 2.838665, 2.307907, 2.093898, 2.019883, 2.142651)
(**4.400911**, 3.537925, 1.387268, 2.715669, 2.372323, 2.077396, 1.970233, 2.268232)

Scale 3 energy values are normalized in the same way. The standard deviation features are also shifted in the same way in the feature vectors. By this rotation normalisation process, we are able to acquire and index rotation invariant curvelet features offline.

4 Retrieval Performance Tests on Texture Image Retrieval

In this section, we test the performance of the curvelet features in terms of retrieval accuracy, robustness and computation efficiency. The retrieval performance of curvelet features is compared with both wavelet and Gabor texture features.

4.1 Databases

Two datasets are used for the experiments. The first is the standard Brodatz texture dataset which consists of 112 different categories of natural and man made texture images (Ma and Manjunath 1995; Liu and Picard 1996; Manjunath and Ma 1996; Zhang and Lu 2000; Chen et al. 2004). Each category consists of 16 similar images cropped from one of the 112 original images, and images in each category have the same size of 128 × 128 pixels. In total, the dataset has 1,792 images from 112 categories. The second standard dataset is the VisTex dataset which consists of 1,264 color texture images (Li et al. 2000). The VisTex dataset has been manually organized into 71 categories by MIT. Each image in the VisTex dataset also has the size of 128 × 128 pixels, and all the images in this dataset are converted to gray level images. In order to test the retrieval performance and robustness of curvelet features, 4 image databases are created from the two datasets. The ground truth of both the datasets are therefore known.

- DB-I. The original Bradatz dataset is used to test retrieval performance as is widely done in literature.
- DB-II. To test the robustness of curvelet features on wider variety of texture images, the Brodatz and VisTex texture datasets are merged into a larger database. This generates a database of 3,056 texture images.
- DB-III. In order to test the tolerance to scale distortions, 200 scaled images are created in addition to the Brodatz dataset. In the first, 40 images are randomly selected from the Brodatz dataset, each of which belongs to a different category of textures. We then scale each of the 40 images by 10% to 50%, which creates 200 scaled images and they are mixed with the original Brodatz texture dataset. This generates a database of 1,992 images.
- DB-IV. To test the rotation invariance of curvelet features, each of the 112 original images in the Brodatz dataset is rotated 6 times with 30 degree in each increment. For each rotated mother image, 8 sub-images are created. As a result, $6 \times 8 \times 112 = 5{,}376$ rotated images are generated. The 5,376 rotated images are mixed with the original Brodatz dataset, which generates a database of 7,168 rotated images.

4.2 Performance Measurement

The standard performance measurement precision-recall pair is used for the evaluation of retrieval performance. The precision P and recall R are defined as following:

$$P = \frac{\textit{No. of relevant images retrieved}}{\textit{Total no. of images retrieved}}$$
$$R = \frac{\textit{No. of relevant images retrieved}}{\textit{Total no. of relevant images in DB}} \tag{9}$$

4.3 Distance Measure

Given the query feature vector $\mathbf{Q} = \{Q_0, Q_1, \ldots, Q_{2n-1}\}$, it is compared with each target feature vector $\mathbf{T} = \{T_0, T_1, \ldots, T_{2n-1}\}$ in the indexed list using the L_2 distance:

$$d(\mathbf{Q}, \mathbf{T}) = \left(\sum_{i=0}^{2n-1} (Q_i - T_i)^2 \right)^{\frac{1}{2}} \tag{10}$$

The reason of choosing L_2 distance instead of other complex distance measure is because L_2 distance is very robust and is the most widely used in literature (Long et al. 2003; Zhang and Lu 2003). The other reason is that L_2 distance is also very efficient which is important for online image retrieval. Finally, the images in the database are ranked according to their distance d to the query image, and the ranked list of images is returned to the user.

4.4 Standard Retrieval Performance Test

To benchmark the performance of curvelet texture features, we compare the retrieval performance of curvelet features with that of the two widely used spectral texture features in literature, i.e., the Gabor filter feature which is adopted by MPEG-7 and the dual-tree rotated complex wavelet filter (DT-RCWF) feature described in Kokare et al. (2006).

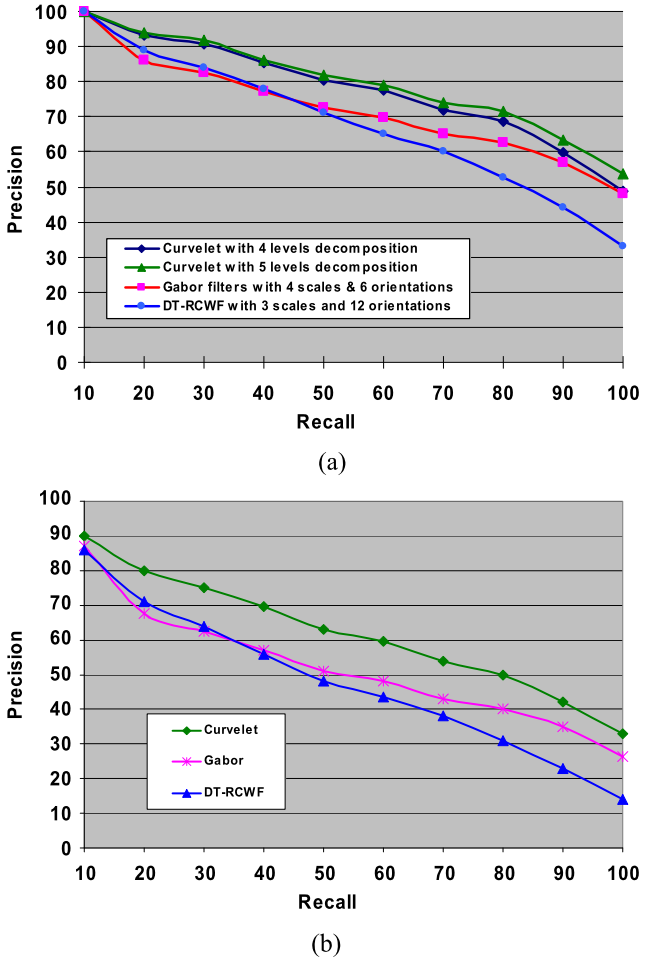


Fig. 5 (**a**) Average retrieval result of 1,792 queries from DB-I. (**b**) Average retrieval result of 3,056 queries from DB-II

For Gabor feature extraction, we use the same method and parameters (4 scales and 6 orientations) as they are used in Manjunath and Ma (1996) (code available from the authors' web site). The Gabor filters with this configuration generate 48 texture features for each input image, which is comparable with the 52 texture features of the 4 levels curvelet transform. For the DT-RCWT feature, images are decomposed into 24 subbands which consist of 3 scales and 8 oriented subbands at each scale. This gives 48 DT-RCWT features using both mean and standard deviation representations for each subband. For online retrieval efficiency, we use the L_2 distance instead of the expensive quadratic distance used in Manjunath and Ma (1996). The comparison is done on both DB-I and DB-II. Because the ground truth of both databases is known, all images in each database are used as queries. For each query, the precision of the retrieval at each level of the recall is obtained. These precision values are then averaged to produce the average precision-recall curve for both the databases. The retrieval performance of curvelet, wavelet and Gabor features are shown in Fig. 5.

As can be seen from both the graphs, the retrieval performance of curvelet features is significantly higher than that of both the Gabor features and the wavelet features. The curvelet transform with 5 levels decomposition has even better retrieval performance than that with 4 levels decomposition, however, at the cost of higher feature dimension which is 84. For efficient online retrieval, lower feature dimension is desirable. On the larger database DB-II, though all the three types of features drops performance, curvelet features still give significant higher performance than both the Gabor and the wavelet features. This demonstrates that the curvelet features are more robust than both the Gabor and the wavelet features. In both tests, wavelet features have a much sharper drop of performance than both the Gabor and the curvelet features.

The reasons why curvelet feature has significantly higher retrieval performance than Gabor features are due to curvelet has several advantages over Gabor filter. First, due to the half-height truncation of the filters to remove the overlap in spectral domain, many holes are created in the spectrum, that is, the spectrum is not completely covered by Gabor filters (Liu et al. 2007). As a result, significant amount of frequency information in an image is lost due to the incomplete spectrum cover. Secondly, curvelet transform uses ridgelets as the filters which are very sensitive to edge information; while Gabor uses Gaussian shape wavelets as the filters which are not so sensitive to edge information especially in the low scaled filters. Thirdly, curvelet transform is a true wavelet transform, while Gabor filter is not because it does not scale image during the filtering (only filters are scaled). Due to the adaptation to image size, curvelet can tolerate more scale distortions in images. This can be observed from Fig. 3 and the results from the following scale tolerance test. Compared with wavelets, curvelets capture high frequency features more effectively and accurately.

4.5 Scale Tolerance Test

For texture features, tolerance to mild scale distortions is an important attribute as natural images are usually captured with certain scale difference. Although scale invariance is itself a research topic in texture feature design, a test of the features' scale robustness is valuable information to image retrieval. The test is done on DB-III. In the first phase of the experiment, we perform texture retrieval using the selected 40 categories of original images as queries, so the total number queries in this phase is $40 \times 16 = 640$ (each category contains 16 original image). The retrieval performance from this phase of test is used as the reference for the scaled image retrieval test in the next phase. In the second phase of retrieval test, we combine the 200 scaled images and the original 640 images ($640 + 200 = 840$) as queries. The main objective of this phase is to find out curvelets' tolerance to

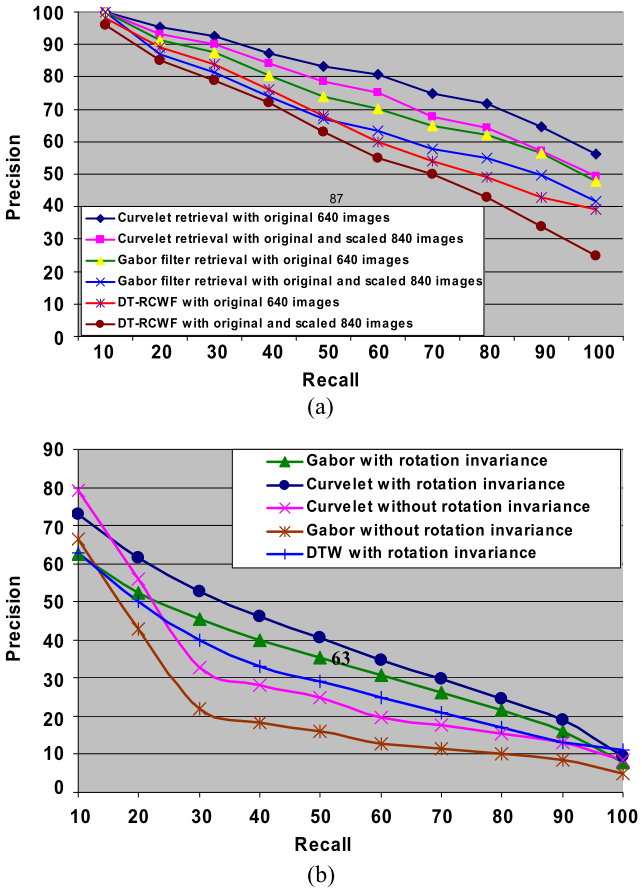


Fig. 6 (**a**) Average retrieval result from DB-III; (**b**) Average retrieval result from DB-IV

scaled distortions. All the three types of features are tested for scale tolerance. The result is shown in Fig. 6(a). It can be observed that while curvelet retrieval performance on the scaled database drops gradually which is normal, Gabor retrieval performance drops about consistently across all the recall levels, wavelet features has the sharpest drop of performance. It demonstrates that both the Gabor features and wavelet features are more sensitive to scale distortions than the curvelet features.

4.6 Rotation Invariance Test

Objects and images could be in different orientations due to camera tilt or post processing. Most texture features in literature are not rotation invariant. As a result, MPEG-7 recommends the use of best match when texture images are retrieved (Manjunath et al. 2002). However, the best match is too expensive for online image retrieval. In this paper, we normalize curvelet features into rotation invariance during indexing, leading to simple online matching. We compare the rotation invariant curvelet features with the rotation invariant Gabor features which are also rotation normalized using the method described in Zhang and Lu (2000) and the rotation invariant dual-tree wavelet (DTW) features described in Kokare et al. (2006). The average retrieval result from the rotation invariance test is shown in Fig. 6(b). It can be seen that, overall, curvelet features outperform both Gabor and wavelet features with a significant margin. The graph also shows that without rotation normalization, the retrieval performance of both curvelet and Gabor features is unacceptably low.

Table 2 Feature extraction time

Method	Total time taken for 1792 images (seconds)	Average time taken for each image (seconds)
Curvelet (4 level)	589.59	0.3290
Curvelet (5 level)	613.68	0.3425
Gabor filters (48 features)	674.49	0.3764
DT-RCWF	577.47	0.3222

4.7 Computation Efficiency Comparison

The time for curvelet feature extraction has also been recorded and compared with Gabor filters (Table 2). The reported time is obtained on the Windows platform of a PC with 3.00 GHz Pentium processor and 2 GB memory. It can be observed that curvelet feature extraction is more efficient than Gabor filter, while comparable with the dual-tree wavelet feature extraction

5 Region Based Image Retrieval Using Curvelet Features

Most of recent CBIR systems are region based, where images are segmented into regions of homogenous texture (Datta et al. 2008; Liu et al. 2007, 2008; Wang et al. 2001, 2006, 2007, 2008b; Jeon et al. 2003; Duygulu et al. 2002; Suematsu et al. 2002). The segmented regions have irregular shapes. However, spectral transforms need to be applied on squared shapes for efficiency and accuracy. Furthermore, many of the segmented regions are not large enough for effective spectral transform. In this section, a novel padding method is proposed to extract sizable square regions from irregular regions. The retrieval performance of the curvelet features is compared with that of Gabor features.

5.1 Transforming Irregular Regions to Square Regions

Basically, the shape transform is to use mirror padding to extend an extracted rectangle to a square region of a given size so that the texture property of the original region is best preserved in the transformed region. An irregular shaped

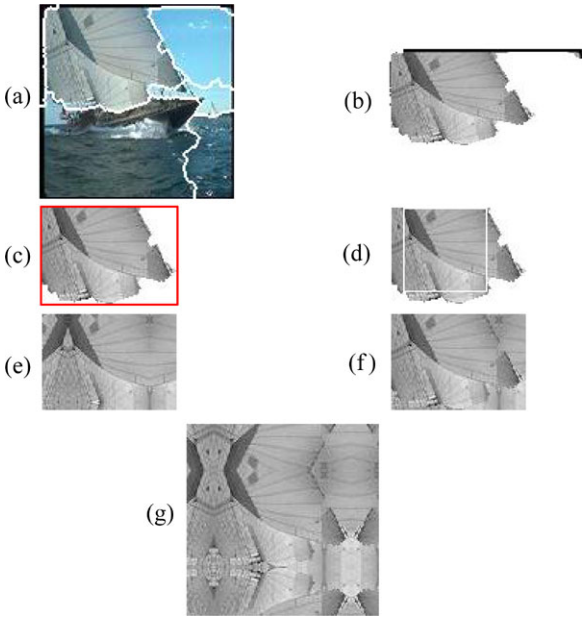


Fig. 7 Transform irregular regions to regular regions. (**a**) A segmented image; (**b**) the top left segmented region in the segmented image in (**a**); (**c**) region pre-processing; (**d**) the extracted largest internal square; (**e**) mirror padding of the internal square in (**d**); (**f**) superimpose the region in (**c**) on the padded region in (**e**); (**g**) mirror padded square from (**f**)

region can be transformed into a regular shaped region by finding either the bounding box or the largest internal rectangle. Though a bounding box is easy to find, it always includes non-region pixels which need to be filled in before spectral transform. The accuracy of the features extracted from a bounding box depends on the values in the non-region pixels. The most commonly used technique is 'zero-padding' which fills these positions with zeros. However, zero padded regions are so different from the original region that the overall texture information of the region is significantly changed. In contrast, an internal square consists of only valid region-pixels. Thus the features extracted from an internal square are more accurate than the features extracted from a bounding box. However, an internal square may not be useful unless it is large enough to cover a significant portion of a region. Therefore, we propose a method to use both the bounding box and the largest internal square to extract a square shape with reasonable size. The algorithm of the shape transform is outlined in the following. Basically, the algorithm extracts the largest internal square from the region and uses the pattern in the internal square to fill the blank area of the bounding box, and the bounding box is then mirror padded into a square region.

Algorithm 1

1. Remove boundary pixels using pre-processing.
2. Find the bounding box of the pre-processed region.
3. Find the largest internal square of the desired region.
4. Extend the internal square to the size of the bounding box using mirror padding.
5. Super impose the bounding box over the mirror padded rectangle.
6. Extend the superimposed rectangle to a square of given size using mirror padding.

We describe the main idea of the algorithm using an example in Fig. 7. The region in the upper left corner of Fig. 7(a) is to be transformed into a square region. Figure 7(b) shows the region in grey scale. As segmented regions often include boundary pixels, these pixels need to be removed for the regions. Therefore, a pre-processing is applied to remove these non-image pixels. Figure 7(c) shows the region after applying pre-processing on the region of Fig. 7(b). The red boundary is the bounding box. The largest internal square is then found and is shown in Fig. 7(d). The internal square is first enlarged to the size of the bounding box through mirror padding which preserves the natural transition (Fig. 7(e)). The region of Fig. 7(c) is then superimposed on the mirror padded region in Fig. 7(e), so the original region is completely preserved in the output region (Fig. 7(f)). The filled space now carries the same texture patterns of the original region. The rectangular region of Fig. 7(f) is further extended to a 128×128 region using mirror padding. Figure 7(g) shows the final transformed square region which has close texture pattern to the original region. Figure 8 shows a few more examples of the transformed square regions from segmented regions.

Homogenous region padding is an important issue in image retrieval due to the limitation of current segmentation techniques. In many situations, image operators need to be applied on squared region of sufficient size. It is shown in Islam et al. (2009) that the proposed padding outperforms the conventional zero padding significantly in terms of accuracy of retrieving regions.

5.2 Retrieval Performance Tests

Once a segmented region is transformed into a squared region, curvelet transform and Gabor filters are applied to extract the texture features. To compare the performance of both curvelet feature and Gabor filter feature in representing regions, we perform region retrieval using both types of the features. At first, 10,000 images are collected from 100 categories. Among them, 5,000 images of 50 categories are from the Corel5K dataset (Liu et al. 2008; Wang et al. 2008a; Li et al. 2000). The other 5,000 images are collected from the Google image search using 50 keyword queries, 100 similar images from each query are selected. The 10,000 im-

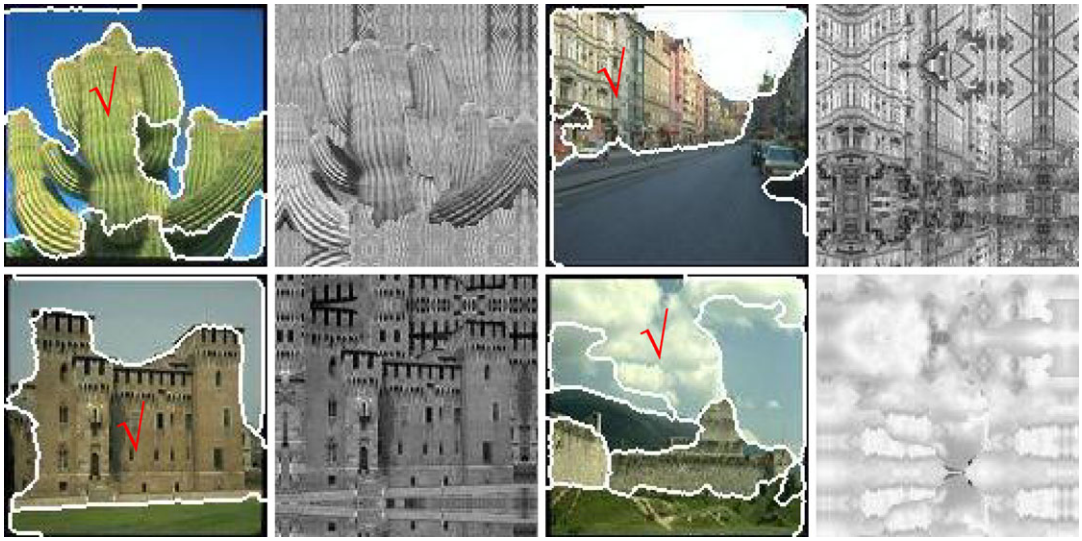


Fig. 8 Examples of segmented regions and their transformed square regions, regions to be transformed are marked with a '√'

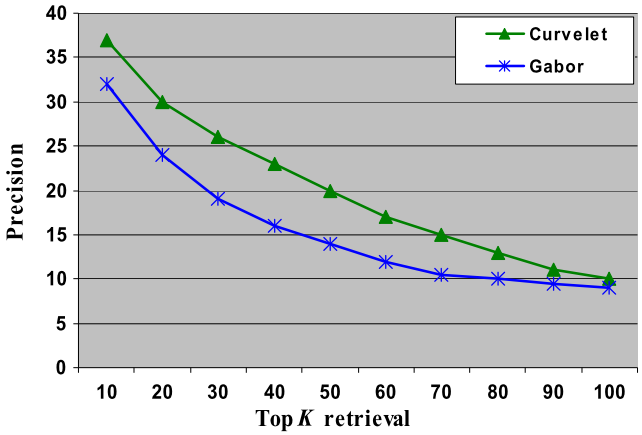


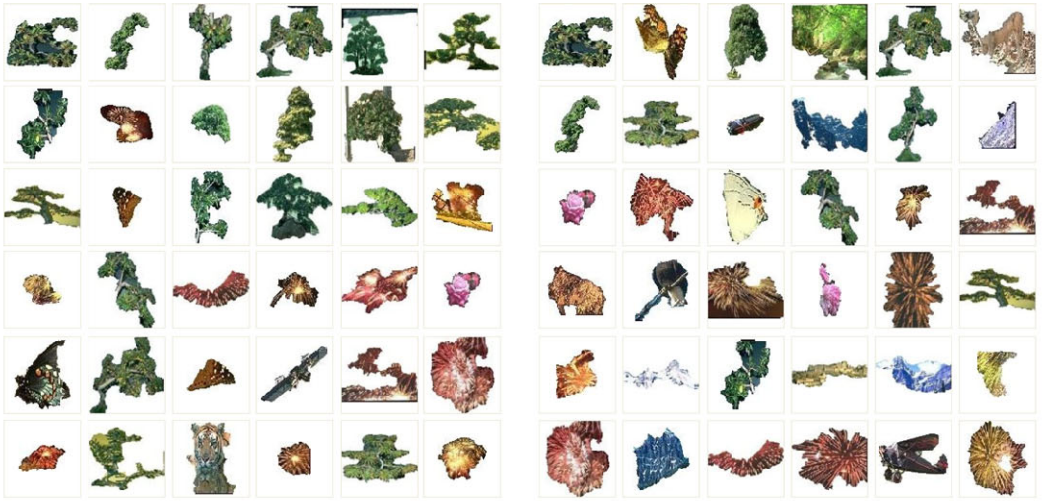
Fig. 9 Average region retrieval performance of curvelet and Gabor filters

ages in the database are segmented into 55,870 regions using one of the state-of-the-art segmentation algorithms, namely the JSEG method (Deng and Manjunath 2001). 5,000 regions from 22 categories are selected as queries. These regions are chosen because they clearly represent some concepts in the database. The regions are from the following concepts: ape, balloon, bear, bird, butterfly, car, copter, deer, elephant, fighter plane, fireworks, flower, fox, horse, plane, tiger, beach, sunset, boat, office and tree. All the collected regions are transformed to a squared region and the curvelet feature of 52 dimensions is extracted using the technique described in Sect. 3.

The performance of the curvelet features is compared with that of the Gabor features. As the ground truth of the database regions are known, each of the query regions is used as a query. The average retrieval performance for both the methods is shown in Fig. 9. It should be noted that color regions are difficult to retrieve using texture feature alone. In actual CBIR systems, texture feature is combined with color to achieve good retrieval performance.

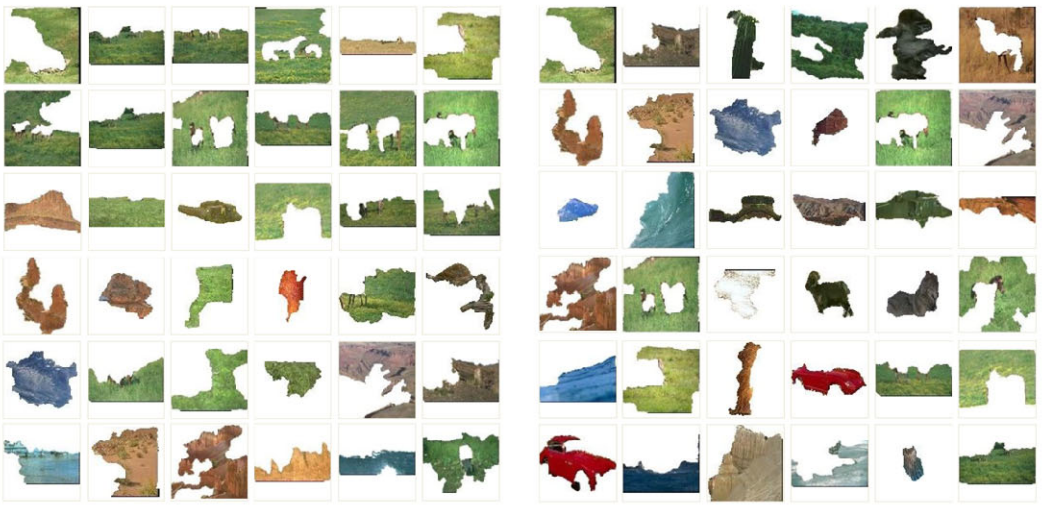
As shown in Fig. 9, the performance of curvelet feature is significantly higher than that of Gabor feature in retrieving natural image-regions. There are several reasons behind this high performance of curvelet feature. Some reasons are due to the characteristics of natural images, while some others are due to the characteristics of curvelet itself. The textures of natural images are random, they cover the entire spectrum—from smooth to rough, fine to coarse and non-directional to directional. Thus the filter bank should cover the entire frequency spectra to represent natural textures, and must be robust enough to represent the different variations of natural textures. Curvelet has several advantages over Gabor in this regard as discussed in Sect. 4. Another advantage of curvelet over Gabor filters is that curvelet transform has higher number of sub-bands at finer scales than coarser scales, while Gabor transform has equal number of sub-bands at all scales. As a result, curvelet captures high frequency information more accurately than Gabor filters, and high frequency information is crucial for texture representation.

Figure 10 shows a few examples of region retrieval using both curvelet feature (left) and Gabor feature (right). In all cases, the difference between the two is significant. For example, in the first case, when the top 18 images are considered, curvelet retrieves 15 tree regions, while Gabor only retrieves 8 trees. When the top 36 images are considered, curvelet's precision is 19 out of 36, while Gabor's precision is 10 out of 36.



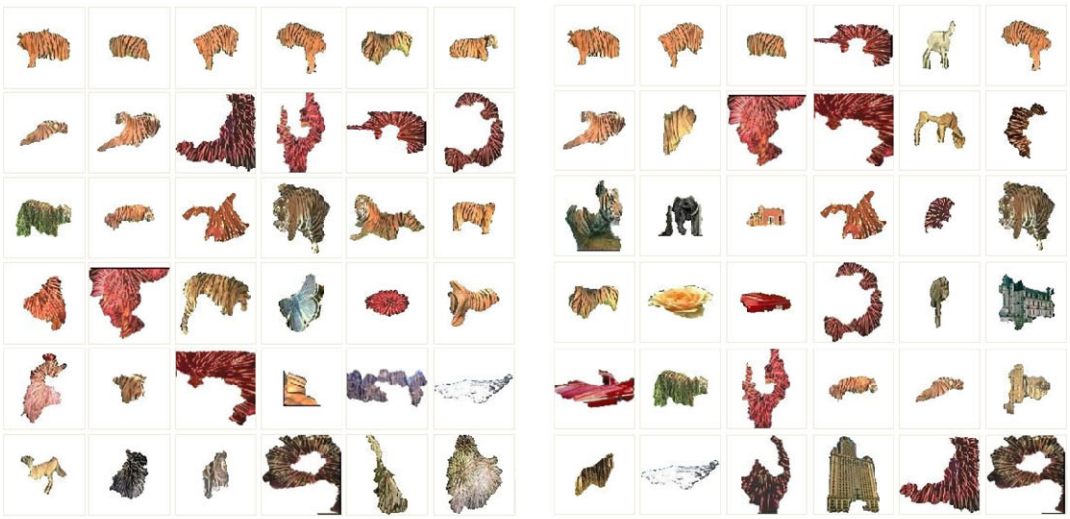
(a) 15 trees out of 18; 19 trees out of 36.

(b) 8 trees out of 18; 10 trees out of 36.



(c) 17 grass out of 18; 24 grass out of 36.

(d) 6 grass out of 18; 12 grass out of 36.



(e) 13 tigers out of 18; 18 tigers of 36.

(f) 9 tigers out of 18; 14 tigers out of 36.

Fig. 10 (**a**) Tree retrieval using curvelet; (**b**) tree retrieval using Gabor; (**c**) grass retrieval using curvelet; (**d**) grass retrieval using Gabor; (**e**) tiger retrieval using curvelet; (**f**) tiger retrieval using Gabor

6 Conclusions

Texture feature is an essential component of any image retrieval system due to its perceptual recognition capability of images. In this paper, rotation invariant curvelet texture features are proposed for region based image retrieval. Several contributions have been made. Firstly, we conduct a comprehensive review of state of the art texture methods in literature. Secondly, curvelet transform and curvelet texture features based on low order statistics are formally introduced for image retrieval and systematically analysed. Thirdly, a rotation invariant texture feature based on curvelet transform is proposed. Rigorous performance tests have been carried out on standard databases to evaluate the proposed method against state of the art texture methods in literature. Results have been systematically analysed and discussed. We demonstrate that curvelet texture method has several advantages over the existing spectral texture methods. The fourth contribution of the paper is the development of a novel region padding method and the application of curvelet transform on region based image retrieval. Results show the rotation invariant curvelet texture features significantly outperforms the Gabor texture features and are very promising for image retrieval. Since Gabor features have been adopted by ISO MPEG-7 standard, this result will have a significant impact on future CBIR research. In future, semantic learning using curvelet features, shape features and color features will be investigated.

References

http://www.curvelet.org, accessed on 23rd December (2008).

Arivazhagan, S., Ganesan, L., & Kumar, T. G. S. (2006). Texture classification using Curvelet statistical and co-occurrence features. In *Proc. of the 18th international conference on pattern recognition* (ICPR06), Washington, DC, August 20–24 (Vol. 2, pp. 938–941).

Bhagavathy, S., & Chhabra, K. (2007). *A wavelet-based image retrieval system* (Technical Report—ECE278A). Vision Research Laboratory, University of California, Santa Barbara.

Candes, E., Demanet, L., Donoho, D., & Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling and Simulation*, *5*(3), 861–899.

Chaudhuri, B. B., & Sarkar, N. (1995). Texture segmentation using fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(1), 72–77.

Chen, L., Lu, G., & Zhang, D. S. (2004). Effects of different Gabor filter parameters on image retrieval by texture. In *Proc. of IEEE 10th international conference on multi-media modelling*, Australia, 2004 (pp. 273–278).

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: ideas, influences, and trends of the new age. *ACM Computing Surveys*, *40*(2), 5:1–60.

Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientationoptimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, *2*(7), 1160–1169.

Daugman, J. G. (1988). Complete discrete 2-D Gabor transform by neural networks for image analysis and compression. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *36*(7), 1169–1179.

Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Transactions on Biomedical Engineering*, *36*(1), 107–114.

Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*(11), 1148–1161.

Deng, Y., & Manjunath, B. S. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE PAMI*, *23*(8), 800–810.

Do, M. N. (2001). Directional multiresolution image representations. Ph.D. thesis, EPFL.

Do, M. N., & Vetterli, M. (2002). Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance. *IEEE Transactions on Image Processing*, *11*(2), 146–158.

Do, M. N., & Vetterli, M. (2003). The finite ridgelet transform for image representation. *IEEE Transactions on Image Processing*, *12*(1), 16–28.

Duygulu, P., Barnard, K., de Freitas, N., & Forsyth, D. (2002). Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In *Proc. of the 7th European conf. on computer vision* (pp. 97–112).

Ferecatu, M., & Boujemaa, N. (2007). Interactive remote-sensing image retrieval using active relevance feedback. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(4), 818–826.

Hervé, N., & Boujemaa, N. (2007). Image annotation: which approach for realistic databases? In *Proc. of the 6th ACM international conf. on image and video retrieval*, Amsterdam, Netherlands (pp. 70–177).

Howarth, P., & Ruger, S. (2004). Evaluation of texture features for content-based image retrieval. *Lecture Notes*, *3115*, 326–334.

Huang, J., Kumar, S. R., Mitra, M., Zhu, W.-J., & Zabih, R. (1997). Image indexing using color correlograms. In *Proc. of IEEE international conf. on computer vision and pattern recognition*, San Juan, Puerto Rico, 17–19 June 1997 (pp. 762–768).

Inoue, M. (2004). On the need for annotation-based image retrieval. In *Proc. of SIGIR workshop on information retrieval in context* (IRiX04), Sheffield, UK, 29th July (pp. 44–46).

Islam, M., Zhang, D., & Lu, G. (2008). Automatic categorization of image regions using dominant color based vector quantization. In *Proc. of digital image computing: techniques and applications* (DICTA08), Canberra, Australia, 1–3 December (pp. 191–198).

Islam, M. M., Zhang, D., & Lu, G. (2009). Region based color image retrieval using curvelet transform. In *Proc. of the 9th Asian conference on computer vision* (ACCV2009), Xian, China, Sept. 23–27.

Jeon, J., Lavrenko, V., & Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proc. of the 26th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 119–126).

Joutel, G., Eglin, V., Bres, S., & Emptoz, H. (2007a). Curvelets based feature extraction of handwritten shapes for ancient manuscripts classification. In *SPIE: Vol. 6500. Proc. of SPIE-IS&T electronic imaging* (65000D).

Joutel, G., Eglin, V., Bres, S., & Emptoz, H. (2007b). Curvelets based queries for CBIR application in handwriting collections. In *Ninth international conference on document analysis and recognition* (ICDAR 2007) (pp. 649–653).

Kokare, M., Biswas, P. K., & Chatterji, B. N. (2006). Rotation-invariant texture image retrieval using rotated complex wavelet filters. *IEEE Transactions on Systems, Man and Cybernetics. Part B*, *36*(6), 1273–1282.

Lew, M. S., Sebe, N., Djeraba, C., & Jain, R. (2006). Content-based multimedia information retrieval: state of the art and challenges.

ACM Transactions on Multimedia Computing Communications and Applications, *2*(1), 1–19.

Li, S. Z., Chan, K. L., & Wang, C. (2000). Performance evaluation of the nearest feature line method in image classification and retrieval. *IEEE PAMI*, *22*(11), 1335–1339.

Liu, F., & Picard, R. W. (1996). Periodicity, directionality, and randomness: wold features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(7), 722–733.

Liu, Y., Zhang, D. S., & Lu, G. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, *40*(1), 262–282.

Liu, Y., Zhang, D., & Lu, G. (2008). Region-based image retrieval with high-level semantics using decision tree learning. *Pattern Recognition*, *41*(8), 2554–2570.

Long, F., Zhang, H. J., & Feng, D. D. (2003). Fundamentals of content-based image retrieval. In D. Feng (Ed.), *Multimedia information retrieval and management*. Berlin: Springer.

Lu, Z., Li, S., & Burkhardt, H. (2006). A content-based image retrieval scheme in JPEG compressed domain. *International Journal of Innovative Computing, Information and Control*, *2*(4), 831–839.

Lu, Y., Zhang, L., Tian, Q., & Ma, W.-Y. (2008). What are the high-level concepts with small semantic gaps? In *Proc. of international conf. on computer vision and pattern recognition* (CVPR08), 23–28 June 2008 (pp. 1–8).

Ma, W. Y., & Manjunath, B. S. (1995). A comparison of wavelet transform features for texture image annotation. In *Proc. of the IEEE international conference on image processing* (ICIP), Washington, DC, Oct. 23–26 (Vol. 2, pp. 256–259).

Manjunath, B. S., & Ma, W. Y. (1996). Texture features for browsing and retrieval of large image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(8), 837–842.

Manjunath, B. S., Ohm, J., Vasudevan, V. V., & Yamada, A. (2001). Color and texture descriptors. *IEEE Transactions CSVT*, *11*(6), 703–715.

Manjunath, B. S., Salembier, P., & Sikora, T. (2002). *Introduction to MPEG-7*. New York: Wiley.

Ng, C. R., Lu, G., & Zhang, D. (2005). A new approach to texture retrieval. In *Proc. of IEEE international workshop on multimedia signal processing* (MMSP05), Shanghai, China, Oct. 30 to Nov. 2.

Ngo, C. W., Pong, T. C., & Chin, R. T. (2001). Exploiting image indexing techniques in DCT domain. *Pattern Recognition*, *34*(9), 1841–1851.

Niblack, W., et al. (1993). The QBIC project: querying image by content using color, texture and shape. *Proceedings of SPIE Storage and Retrieval for Image and Video Databases*, *1908*, 173–187.

Semler, L., & Dettori, L. (2006). Curvelet-based texture classification of tissues in computed tomography. In *Proc. of the IEEE international conference on image processing*, 8–11 Oct. (pp. 2165–2168).

Shekhar, R., & Chaudhuri, S. (2005). In *Lecture notes in computer science: Vol. 3776. Use of contourlets for image retrieval* (pp. 563–569).

Starck, J., Candès, E. J., & Donoho, D. L. (2002). The curvelet transform for image denoising. *IEEE Transactions on Image Processing*, *11*(6), 670–684.

Suematsu, N., Ishida, Y., Hayashi, A., & Kanbara, T. (2002). Region-based image retrieval using wavelet transform. In *Proc. 15th international conf. on vision interface*, May 2002.

Sumana, I., Islam, M., Zhang, D., & Lu, G. (2008). Content based image retrieval using curvelet transform. In *Proc. of IEEE international workshop on multimedia signal processing* (MMSP08), Cairns, Australia, October 8–10 (pp. 11–16).

Tamura, H., Mori, S., & Yamawaki, T. (1978). Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, *8*(6), 460–473.

Vasconcelos, N. (2007). From pixels to semantic spaces: advances in content-based image retrieval. *IEEE Computer*, *40*(7), 20–26.

Vertan, C., & Boujemaa, N. (2000). Upgrading color distributions for image retrieval: Can we do better? In *Proc. int. conf. VISUAL*, Nov. 2000 (pp. 178–188).

Wang, J. Z., Li, J., & Wiederhold, G. (2001). SIMPLIcity: Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(9), 947–963.

Wang, C., Jing, F., Zhang, L., & Zhang, H. (2006). Image annotation refinement using random walk with restarts. In *Proc. of the 14th ACM international conference on multimedia*, Santa Barbara, CA, USA, Oct. 23–27 (pp. 647–650).

Wang, C., Jing, F., Zhang, L., & Zhang, H.-J. (2007). Content-based image annotation refinement. In *Proc. of international conf. on computer vision and pattern recognition* (CVPR07) (pp. 1–8).

Wang, C., Zhang, L., & Zhang, H. (2008a). Learning to reduce the semantic gap in web image retrieval and annotation. In *Proc. of the 31st annual international ACM SIGIR conf. on research and development in information retrieval* (SIGIR08), Singapore, 20–24 July 2008 (pp. 355–362).

Wang, C., Zhang, L., & Zhang, H. (2008b). Scalable Markov model-based image annotation. In *Proc. of international conference on content-based image and video retrieval* (CIVR08), Canada, 07–09 July 2008 (pp. 113–118).

Zhang, D., & Lu, G. (2000). Content-based image retrieval using Gabor texture features. In *Proc. of first IEEE pacific-rim conference on multimedia* (PCM00), Sydney, Australia, December 13–15 (pp. 1139–1142).

Zhang, D., & Lu, G. (2003). Evaluation of similarity measurement for image retrieval. In *Proc. of IEEE international conference on neural networks & signal processing* (ICNNSP03), Nanjing, China, Dec. 14–17 (pp. 928–931).

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, *37*(1), 1–19.