

# Rotation-invariant and scale-invariant Gabor features for texture image retrieval

Ju Han<sup>1</sup>, Kai-Kuang Ma<sup>\*</sup>

*School of Electrical and Electronic Engineering, Nanyang Technological University, Nanyang 639798, Singapore*

Received 29 January 2004; received in revised form 8 December 2006; accepted 19 December 2006

## Abstract

Conventional Gabor representation and its extracted features often yield a fairly poor performance in retrieving the rotated and scaled versions of the texture image under query. To address this issue, existing methods exploit multiple stages of transformations for making rotation and/or scaling being invariant at the expense of high computational complexity and degraded retrieval performance. The latter is mainly due to the lost of image details after multiple transformations. In this paper, a rotation-invariant and a scale-invariant Gabor representations are proposed, where each representation only requires few summations on the conventional Gabor filter impulse responses. The optimum setting of the orientation parameter and scale parameter is experimentally determined over the Brodatz and MPEG-7 texture databases. Features are then extracted from these new representations for conducting rotation-invariant or scale-invariant texture image retrieval. Since the dimension of the new feature space is much reduced, this leads to a much smaller metadata storage space and faster on-line computation on the similarity measurement. Simulation results clearly show that our proposed invariant Gabor representations and their extracted invariant features significantly outperform the conventional Gabor representation approach for rotation-invariant and scale-invariant texture image retrieval.

© 2007 Elsevier B.V. All rights reserved.

*Keywords:* Content-based image retrieval; Texture analysis; Gabor filter; Rotation-invariant; Scale-invariant; Brodatz; MPEG-7; Metadata; Data mining

## 1. Introduction

Besides color, *texture* is another salient and indispensable feature for content-based image indexing and retrieval application through similarity matching. Periodicity, coarseness, inherent direction and pattern complexity are considered as the most perceptually distinct properties of texture. Textures are psycho-physically perceived by the *human visual system* (HVS), particularly, on the aspects of orientation and scale of texture patterns [1].

Texture analysis has been an active research in the past with numerous algorithms being developed based on different models, such as *grey-level co-occurrence* (GLC) matrices

[2], *Markov random field* (MRF) model [3], *simultaneous auto-regressive* (SAR) model [4], and *Wold decomposition* model [5], to name a few. These well-known spatial domain texture analysis models (except the GLC matrices and the multi-resolution SAR model) have a fundamental weakness that the image is analyzed at a single scale, and this aspect can be improved by employing a spatial-frequency (i.e., multi-channel) representation. In fact, it is well-known that the visual cortex of the HVS can be appropriately modeled as a set of independent channels, and each channel is tuned to a particular orientation and spatial frequency [6]. Hence, a joint spatial-frequency multi-channel representation methodology is effective in characterizing texture image features.

Wavelet theory is a unified and effective mathematical framework for multichannel image analysis (e.g., [7]). Essentially, it transforms an image to one low-pass image plus multiple high-pass images. The low-pass image is

<sup>\*</sup> Corresponding author.

*E-mail address:* [ekkma@ntu.edu.sg](mailto:ekkma@ntu.edu.sg) (K.-K. Ma).

<sup>1</sup> Present address: Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

obtained by iteratively applying a low-pass filter to the filtered baseband image obtained from the previous iteration, while the high-pass images contain the information resulted from high-pass filtering at each decomposition stage. The energy and standard deviation of the high-pass images are the most commonly used features for texture classification and segmentation [8,9].

As a relaxed wavelet transformation [10], the Gabor filter family essentially performs a multi-channel representation, which is in line with the multichannel filtering mechanism of the HVS in perceiving visual information [11–13]. This theory holds that the HVS perceives the image presented in the retina through a set of filtered images, and each image contains some unique visual information over a narrow range of orientation channel. For that, image representation through Gabor filtering has been shown to be a good fitting to the receptive field profiles (i.e., the impulse responses of the cells) in the striate cortex, while providing optimal localization of image details in a joint spatial and frequency domain [12].

The above-mentioned Gabor representation is fairly effective in texture analysis and has been beneficial to various image-based applications, such as segmentation [14–19], retrieval [20,21], and biometrics [10,23], etc. However, most of these approaches are sensitive to the changes of orientations and scales of the texture pattern. On the other hand, objects of interest under various orientations and scales are often encountered in the different applications, such as character recognition and target detection. Therefore, conventional Gabor representation and its extracted features often yields a fairly unacceptable performance in retrieving the rotated and scaled versions of the texture image under query, since the representation is *variant* to the different orientations and scales. In this paper, rotation-*invariant* and scale-*invariant* Gabor representations are proposed, from which each representation only involves a simple modification of the conventional Gabor filter family for achieving rotation invariance and scale invariance, individually.

The paper is organized as follows. In Section 2, conventional Gabor representation will be described with necessary details as the background. In Section 3 the proposed rotation-invariant and scale-invariant Gabor representation and their extracted invariant features are presented. By considering certain practical aspects, the optimum parameter selection for these Gabor-based features are studied and recommended in Section 4. The texture image retrieval performance resulted from independently exploiting the conventional and our proposed invariant Gabor-based features are compared by conducting extensive simulation experiments over the Brodatz and MPEG-7 texture databases. Section 5 concludes the paper.

## 2. Conventional Gabor representation

A 2-D Gabor function  $g(x,y)$  and its Fourier transform  $G(u,v)$  can be expressed as:

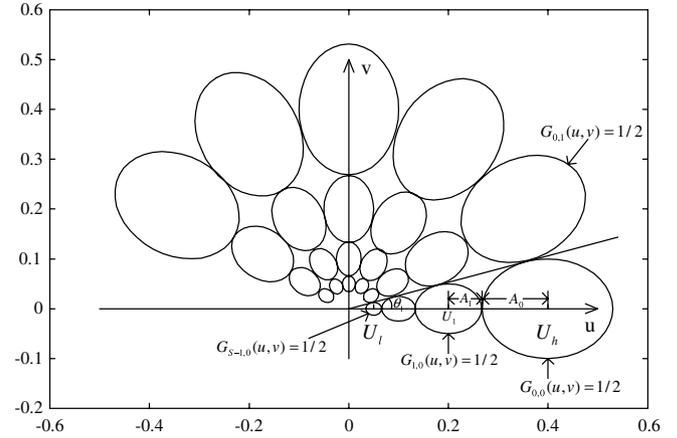


Fig. 1. The elliptical contours adjacent to each other indicate the half-peak magnitude of the filter responses in the Gabor filter family. The origin  $(u,v)=(0,0)$  is at the center of the image array. The filter parameters used here are  $U_l=0.05$ ,  $U_h=0.38$ ,  $K=6$ , and  $S=4$ .

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi jWx \right] \quad (1)$$

and

$$G(u,v) = \exp \left\{ -\frac{1}{2} \left[ \frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right\}, \quad (2)$$

respectively, where  $\sigma_u = \frac{1}{2\pi\sigma_x}$  and  $\sigma_v = \frac{1}{2\pi\sigma_y}$ . In (1) and (2),  $\sigma_x$  and  $\sigma_y$  characterize the spatial extent and frequency bandwidth of the Gabor filter, and  $(W,0)$  represents the center frequency of the filter in the frequency-domain rectilinear coordinates  $(u,v)$ .

Let  $g(x,y)$  be the mother generating function for the Gabor filter family. A set of Gabor functions  $g_{m,n}(x,y)$  can be generated by rotating and scaling  $g(x,y)$  to form an almost complete and non-orthogonal basis set, that is,

$$g_{m,n}(x,y) = a^{-2m}g(x',y') \quad (3)$$

where  $x' = a^{-m}(x \cos \theta_n + y \sin \theta_n)$ ,  $y' = a^{-m}(-x \sin \theta_n + y \cos \theta_n)$ ,  $a > 1$ ,  $\theta_n = n\pi/K$ ,  $m=0,1,\dots,S-1$ , and  $n=0,1,\dots,K-1$ . Parameter  $S$  is the total number of scales, and parameter  $K$  is the total number of orientations.

In [20], it has been pointed out that to reduce the redundancy in the filtered images, the filter parameters are chosen to ensure that the adjacent half-peak magnitude contours of the filter responses in the frequency domain are tangent to each other. In line with the same objective, however, a new result has been derived based on Fig. 1 as follows.

$$a = \left( \frac{U_h}{U_l} \right)^{\frac{1}{S-1}}, \quad \sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2 \ln 2}}, \quad \text{and } \sigma_v = \tan \left( \frac{\pi}{2K} \right) \left( \frac{U_h^2}{2 \ln 2} - \sigma_u^2 \right)^{\frac{1}{2}}, \quad (4)$$

where  $U_l$  and  $U_h (=W)$  denote the lower and upper center frequencies of interest, respectively. The detailed derivation of (4) is provided in Appendix A.

Given an image  $I(x, y)$ , its Gabor-filtered images are

$$J_{m,n}(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) g_{m,n}(x - x_1, y - y_1). \quad (5)$$

The mean and the standard deviation of the magnitude of the filtered images, which are used to construct the feature vector, are respectively defined as

$$\mu_{m,n} = \frac{1}{N} \sum_x \sum_y |J_{m,n}(x, y)|, \quad (6)$$

$$\text{and } \sigma_{m,n} = \sqrt{\frac{1}{N} \sum_x \sum_y (|J_{m,n}(x, y)| - \mu_{m,n})^2}, \quad (7)$$

where  $N$  is the total number of the image pixels. To measure the similarity of two texture images  $i$  and  $j$ , the distance can be computed over their feature vectors as follows [20]:

$$d_{m,n}(i, j) = \left| \frac{\mu_{m,n}^{(i)} - \mu_{m,n}^{(j)}}{\alpha(\mu_{m,n})} \right| + \left| \frac{\sigma_{m,n}^{(i)} - \sigma_{m,n}^{(j)}}{\alpha(\sigma_{m,n})} \right|. \quad (8)$$

where  $\alpha(\mu_{m,n})$  and  $\alpha(\sigma_{m,n})$  are the standard deviations of the respective features over the entire database and used for feature normalization.

### 3. Invariant Gabor representations

To overcome the rotation-variance and scale-variance drawbacks encountered in conventional Gabor representation, two new *invariant* representations are proposed in this paper for significantly improving texture image retrieval performance. Besides performance, computation is another merit point, since each new representation can be easily obtained through some summations over conventional Gabor representation. Furthermore, the dimension of the resulted invariant feature space (or vector) is much reduced, thus yielding faster computation on similarity measurement.

#### 3.1. Basic idea

Let us consider a set of identical texture images with the same texture content, except under different orientations. For each texture image, although the resulted signal energy distribution at each scale level (i.e., within each circular or fan-shaped “band” as shown in Fig. 2(a)) would be different from band to band, however, the total energy of the Gabor filters yielded in each band tends to be quite constant, regardless the orientation angle of the texture pattern and the number of scales involved.

Likewise, the Gabor filter responses under different scales, but along the same orientation direction, could be summed up for achieving scale-invariance as shown in Fig. 2(b). However, unlike rotation-invariance, scale-

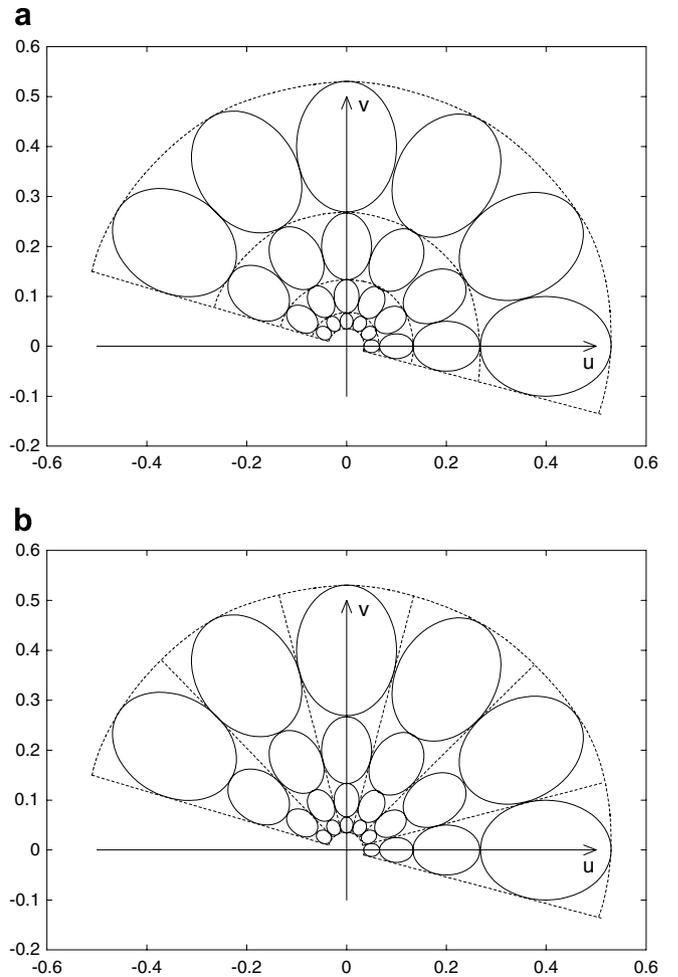


Fig. 2. Regions bounded by the dotted lines are established for (a) rotation-invariant Gabor representation, and (b) scale-invariant Gabor representation.

invariance is inherently much more complicated. Note that rotation does not change the texture pattern. However, any drastically down-scaling could result in aliasing and greatly alter the original texture content. Therefore, generally speaking, scale-invariance could be reasonably achieved only when the scaling factor is not too large and before the aliasing being incurred. The scale-invariance claimed in this paper is imposed by this assumption.

#### 3.2. Rotation-invariant Gabor representation and feature vector

By summing all the  $K$  filters in (3) with different orientations at each scale level, our proposed rotation-invariant Gabor filter family  $\{g_m^{(R)}(x, y)\}$  is obtained. That is,

$$g_m^{(R)}(x, y) = \sum_{n=0}^{K-1} g_{m,n}(x, y), \quad m = 0, 1, \dots, S-1. \quad (9)$$

Each  $g_m^{(R)}(x, y)$  is a filter that extracts features from a specific scale band covering all the orientations within the entire half-plane frequency spectrum as shown in Fig. 2(a),

thus yielding a *rotation-invariant* Gabor representation. Hence, the transformation of image  $I(x, y)$  is

$$J_m^{(R)}(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) g_m^{(R)}(x - x_1, y - y_1),$$

$$m = 0, 1, \dots, S - 1, \quad (10)$$

that represents the properties of  $I(x, y)$  within a specific scale band covering the entire orientation span of the half-plane frequency spectrum. Therefore, the mean value  $\mu_m^{(R)}$  and the standard deviation  $\sigma_m^{(R)}$  of the magnitude of the transform coefficients are defined as

$$\mu_m^{(R)} = \frac{1}{N} \sum_x \sum_y |J_m^{(R)}(x, y)|, \quad (11)$$

$$\sigma_m^{(R)} = \sqrt{\frac{1}{N} \sum_x \sum_y (|J_m^{(R)}(x, y)| - \mu_m^{(R)})^2}, \quad (12)$$

respectively, and they are rotation-invariant as well. The rotation-invariant Gabor feature vector  $\mathbf{f}^{(R)}$  is now constructed for texture image retrieval as

$$\mathbf{f}^{(R)} = [\mu_0^{(R)}, \sigma_0^{(R)}, \mu_1^{(R)}, \sigma_1^{(R)}, \dots, \mu_{S-1}^{(R)}, \sigma_{S-1}^{(R)}]. \quad (13)$$

### 3.3. Scale-invariant Gabor representation and feature vector

Similarly, by summing all the  $S$  filters in (3) with different scales under each orientation, our proposed scale-invariant Gabor filter family  $\{g_n^{(S)}(x, y)\}$  can be obtained as

$$g_n^{(S)}(x, y) = \sum_{m=0}^{S-1} g_{m,n}(x, y), \quad n = 0, 1, \dots, K - 1. \quad (14)$$

Each  $g_n^{(S)}(x, y)$  is a filter that extracts features from a specific orientation band covering all the scales with the frequency spectrum as shown in Fig. 2(b), thus yielding a *scale-invariant* Gabor representation. Hence, the transformation of image  $I(x, y)$  is

$$J_n^{(S)}(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) g_n^{(S)}(x - x_1, y - y_1), \quad n = 0, 1, \dots, K - 1, \quad (15)$$

that represents the properties of  $I(x, y)$  within a specific orientation band that covers  $S$  scales. Therefore, the mean value  $\mu_n^{(S)}$  and the standard deviation  $\sigma_n^{(S)}$  of the magnitude of the transform coefficients are

$$\mu_n^{(S)} = \frac{1}{N} \sum_x \sum_y |J_n^{(S)}(x, y)|, \quad (16)$$

$$\sigma_n^{(S)} = \sqrt{\frac{1}{N} \sum_x \sum_y (|J_n^{(S)}(x, y)| - \mu_n^{(S)})^2}. \quad (17)$$

Note that these quantities are also scale-invariant. The scale-invariant Gabor feature vector  $\mathbf{f}^{(S)}$  is thus constructed for texture image retrieval as

$$\mathbf{f}^{(S)} = [\mu_0^{(S)}, \sigma_0^{(S)}, \mu_1^{(S)}, \sigma_1^{(S)}, \dots, \mu_{K-1}^{(S)}, \sigma_{K-1}^{(S)}]. \quad (18)$$

## 4. Performance evaluation of texture image retrieval

Two popular texture databases are used in our experiments: the Brodatz and MPEG-7 texture databases, consisting of 112 and 52 different types of texture images, respectively. For the Brodatz texture database, each original image with a size of  $512 \times 512$  is evenly divided into sixteen  $128 \times 128$  non-overlapping sub-images, thus creating a database of  $112 \times 16 = 1792$  Brodatz texture images. If any one of them is imposed as the query image, the 16 texture images divided from the same original image are viewed as the images from the same class and targeted to be retrieved as the ground truth.

Each image in the database is processed to generate a feature vector as its metadata for performing similarity matching. The distance  $d(i, j)$  between two feature vectors according to (8) is computed (say, image  $i$  is the query image, and image  $j$  is one of the database images under matching). Those images with the highest scores (i.e., shortest distances) are retrieved and displayed according to their ranked scores. In the ideal case, the top 16 images being displayed should be those images from the same class (i.e., the ground truth), including the query image (supposed to be ranked as the top match).

The retrieval performance is assessed in terms of the commonly used *precision* and *recall* defined as follows [24]:

$$\text{Precision}(N) = \frac{C_N}{N} \quad \text{and} \quad \text{Recall}(N) = \frac{C_N}{M}, \quad (19)$$

where  $N$  is the total number of retrieved images with the highest ranking scores,  $C_N$  is the total number of the ground-truth images from the same class that appear in the retrieved  $N$  images, and  $M$  is the actual number of the ground-truth images from the same class of the query image. In our experiment,  $M = 16$  and  $0 \leq C_N \leq 16$ ; thus,  $0 \leq \text{Recall}(N) \leq 1$ . Note that  $C_N$  is a function of  $N$ ; intuitively, the larger the  $N$  value is imposed, the more the ground-truth images will be retrieved (i.e., larger  $C_N$ ). After exhaustively experimenting each of the database images as the query image, the curve of the average precision or the average recall can be used to assess the retrieval performance. Experimental results on the MPEG-7 texture database are obtained in the same way.

### 4.1. Decomposition setting for conventional Gabor representation

Besides the retrieval performance, there are two other important aspects in determining the number of filters for generating conventional Gabor features: (1) the metadata storage space and (2) on-line retrieval time required in computing the feature distance. Extensive experiments using conventional Gabor features under all possible combinations of the scale numbers ( $S = 3, 4,$  and  $5$ ) and the orientation numbers ( $K = 4, 5, 6, 7, 8,$  and  $9$ ) are conducted on the Brodatz and MPEG-7 texture databases,

respectively. Comparing the obtained results, the following facts are observed:

- Conventional Gabor features with  $S = 4$  or  $S = 5$  yields better retrieval performance than that with  $S = 3$ ;
- Conventional Gabor features with  $S = 4$  and different  $K$  values produce almost the same retrieval performance as that of using  $S = 5$  under the same  $K$  values, respectively.

Therefore, for conventional Gabor representations, it can be concluded that  $S = K = 4$  is the most appropriate decomposition setting for the texture image retrieval application.

#### 4.2. Experimental results based on rotation-invariant Gabor features

In order to evaluate the retrieval performance using our proposed rotation-invariant Gabor features on the rotated textures, the ground-truth image sets (containing sixteen  $128 \times 128$  images in each set) are prepared as follows. Each original  $512 \times 512$  Brodatz image is first rotated with 16 equally spaced angles, from 0 to  $15\pi/16$  with an incremental step size of  $\pi/16$ . The sixteen  $128 \times 128$  images are then obtained by partitioning each rotated image from the image's center using a  $128 \times 128$  window. These 16 rotated images are then added back to the Brodatz database to replace the original 16 sub-divided images.

The retrieval performance using our proposed rotation-invariant Gabor feature over the Brodatz and MPEG-7 texture databases are evaluated, individually. In our experiments, the number of scales  $S$  is set to be  $S = 3, 4, 5$ , and 6. The number of orientations  $K$  is set to be  $K = 4, 5, 6, 7, 8$ , and 9. Comparing these results,  $S = 4$  and  $S = 6$  yields the best retrieval performance for the Brodatz and MPEG-7 databases, respectively. Furthermore, the performances resulted by using  $S = 4, 5$ , and 6 are very close to each other, and, as expected, almost irrelevant to the value of  $K$  imposed. Therefore,  $S = K = 4$  is considered as the most proper decomposition parameter setting for our proposed rotation-invariant Gabor representation. The retrieval results of using conventional Gabor features and our proposed rotation-invariant Gabor features are shown in Fig. 3, respectively. Note that all the 16 rotated images (i.e., the ground truth) are retrieved as the highest 16 matches (i.e., ideal performance) using our rotation-invariant Gabor features, while only 5 out of 30 images being displayed are correctly retrieved by using the conventional Gabor representation.

#### 4.3. Experimental results based on scale-invariant Gabor features

In order to evaluate the retrieval performance of our proposed scale-invariant Gabor features on the scaled texture images, the ground-truth image sets (containing six-

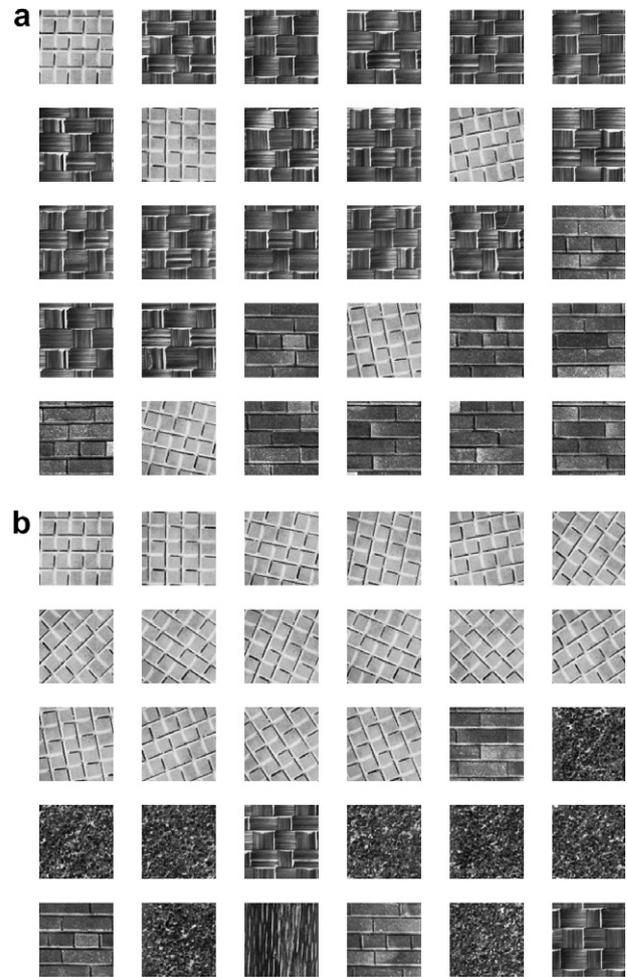


Fig. 3. The texture image retrieval results (with  $S = K = 4$ ) yielded by exploiting: (a) conventional Gabor representation, in which only 5 out of 16 ground-truth images are correctly retrieved and ranked within top 26 images and (b) our proposed rotation-invariant Gabor representation, which achieves an ideal retrieval. (The top-left image is the query image.)

teen  $128 \times 128$  images in each set) are prepared as follows. For each original  $512 \times 512$  Brodatz image, different sizes of square windows were applied to crop the image. Each cropped image is then re-sized to  $128 \times 128$  image by first applying anti-aliasing low-pass filtering followed by a down-sampling process. The resulted 16 sub-images are then added back to the Brodatz database to replace the original 16 sub-divided images.

The retrieval performance of our proposed scale-invariant Gabor features on the Brodatz and MPEG-7 texture databases are evaluated separately. For that, several scale numbers  $S = 3, 4, 5$ , and 6, together with the orientation numbers  $K = 4, 5, 6, 7, 8$ , and 9 are experimented, respectively. Comparing their yielded results, it can be observed that  $S = K = 4$  is also the most appropriate filter setting to our proposed scale-invariant Gabor representation. The retrieval results using conventional Gabor features and our scale-invariant Gabor features are demonstrated in Fig. 4, respectively. Note that all the 16 scaled images (i.e., the ground truth) are retrieved among the highest 26 matches using our scale-invariant Gabor features. On the

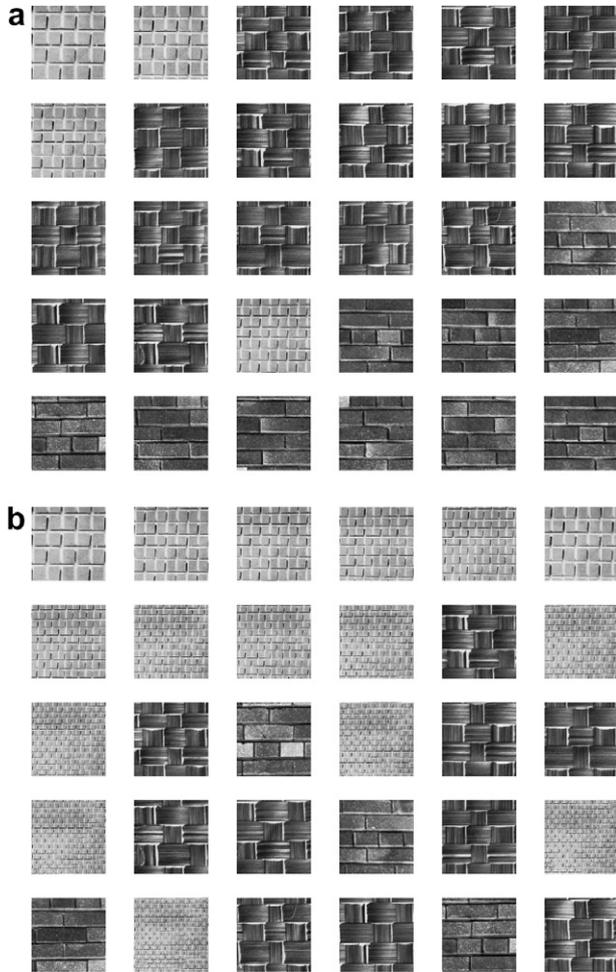


Fig. 4. The texture image retrieval results (with  $S = K = 4$ ) yielded by exploiting: (a) conventional Gabor feature, from which only four ground-truth images are correctly retrieved and ranked in the 1st, 2nd, 7th, and 21st, respectively; and (b) our proposed scale-invariant Gabor feature, resulting in total 16 ground-truth images being retrieved and ranked among the top 26 images. (The top-left image is the query image.)

other hand, only four ground-truth images are successfully retrieved among the top-ranked 30 images by using the conventional Gabor representation.

#### 4.4. Observations and comments

Inherently, scale-invariant texture image retrieval is much more difficult than rotation-invariant texture image retrieval, because the dynamic range of scale changes can be fairly large, while image rotation angle is limited to  $2\pi$ , and mostly to  $\pi$ , as a majority of texture image patterns are often symmetrical and oriented in one direction. For example, in our proposed invariant Gabor representation approach, one can see that all the 16 rotated texture images are retrieved and ranked on the top 16 positions in Fig. 3(b). On the other hand, the 16 scaled ground-truth texture images are scattered within the top 26 retrieved images in Fig. 4(b). As mentioned earlier, this is due to the fact that drastically down-scaling the texture image

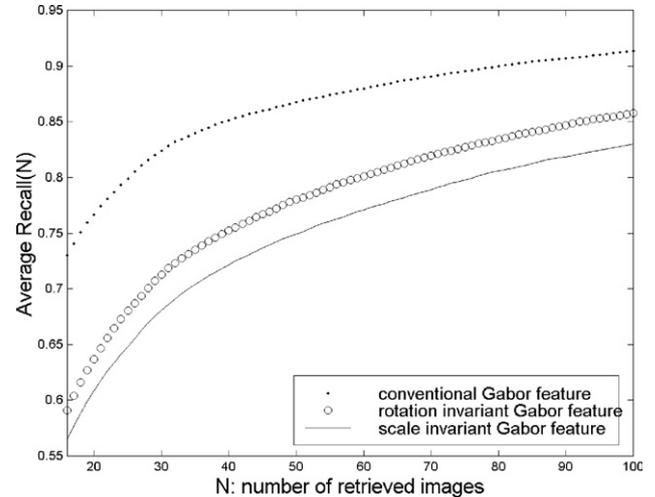


Fig. 5. The retrieval performance using conventional Gabor feature and our proposed rotation-invariant and scale-invariant Gabor features (with  $S = K = 4$ ) on the Brodatz texture database, when there is neither rotation nor scaling issue involved.

could destroy its original texture patterns owing to under-sampling aliasing effect.

When generating texture features (as metadata) in an image database, we may need to accommodate various texture querying objectives. Some content-based retrieval applications may prefer rotation/scale sensitive results, while others might not. Hence, a practical approach is to generate the metadata for both rotation/scale sensitive and rotation/scale invariant cases. Note that the additional computation cost to having both metadata is, generally speaking, not a concern. This is due to the fact that the computing of the proposed rotation/scale invariant texture features can continuously proceed on the results of the conventional Gabor representation via summations.

One should bear in mind that neither rotation-invariant nor scale-invariant is always desirable in the image query application. For example, in browsing the texture images in terms of their classes, any image can be considered as a representative of its associated class. Thus, any other images from the same class with either rotated or scaled version are, in fact, undesirable to be retrieved and displayed. In this scenario, the options of rotation-invariant and scale-invariant modes should be switched off for reducing: (1) the amount of computation time spent on similarity matching, (2) the number of irrelevant images being displayed, and (3) the user's time spent on evaluating the retrieved results.

It is important to point out that when neither rotation nor scaling is part of the application's requirements, conventional Gabor feature is supposed to achieve superior performance to that of the other two cases, individually. This is because that the cardinality of the conventional Gabor feature vector is  $K$ -times and  $S$ -times that of rotation-invariant and scale-invariant Gabor feature vectors, respectively. Therefore, it is expected that the more the useful features are exploited, the better the retrieval performance will be yielded. See Fig. 5 for  $S = K = 4$ , for example.

## 5. Conclusion

In this paper, two *invariant* Gabor representations are introduced: (1) rotation-invariant, and (2) scale-invariant. The feature vectors constructed from these representations are individually exploited for texture image retrieval application. Unlike existing rotation-invariant or scale-invariant methods that require a cascade of transformations, each of our proposed representations only requires few summations over the filter responses of conventional Gabor filter family, individually. That is, for rotation-invariant, all the filter impulse responses with different orientations but at the same level of scale are summed together. For scale-invariant, all the filter impulse responses with different scales but along the same orientation are summed together. Furthermore, compared with scale invariance, rotation invariance is much easier to handle, since there is no aliasing issue involved.

Extensive texture image retrieval experiments were conducted over the Brodatz and MPEG-7 texture databases using the conventional and our proposed rotation-invariant and scale-invariant Gabor features. It has been concluded that  $S=K=4$  is the most appropriate filter parameter setting for texture image retrieval, with considerations of the following practical aspects: (1) retrieval performance, (2) metadata storage space and (3) on-line computation of feature distance for retrieval. The texture image retrieval performance results clearly show that our proposed rotation-invariant and scale-invariant Gabor representations are quite insensitive to rotation or scale changes of texture patterns, and the retrieval performance is much superior to that of conventional Gabor representation, when these invariant options are desirable.

## Appendix A. Parameters design for the Gabor filter family

The mother generating function of the Gabor filter family is

$$g_{0,0}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j U_h x \right],$$

and its frequency response locates at the center frequency  $(U_h, 0)$ :

$$G_{0,0}(u, v) = \exp \left\{ -\frac{1}{2} \left[ \frac{(u - U_h)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right\}.$$

A set of Gabor functions  $g_{m,n}(x, y)$  can be generated by rotating and scaling  $g_{0,0}(x, y)$  to ensure that the half-peak magnitude contours of their frequency responses are tangent to each other as shown in Fig. 1 [20]:

$$g_{m,n}(x, y) = a^{-2m} g_{0,0}(x', y') \quad (\text{A.1})$$

where  $x' = a^{-m}(x \cos \theta_n + y \sin \theta_n)$ ,  $y' = a^{-m}(-x \sin \theta_n + y \cos \theta_n)$ ,  $a > 1$ ,  $\theta_n = n\pi/K$ ,  $m = 0, 1, \dots, S-1$ , and  $n = 0, 1, \dots, K-1$ . Parameter  $S$  is the total number of scales, and parameter  $K$  is the total number of orientations.  $U_l$  and

$U_h$  denote the lower and upper center frequencies of interest, respectively. Let  $G_{m,n}(u, v)$  be the frequency response of  $g_{m,n}(x, y)$ . According to the properties of the Fourier transform, we have

$$G_{m,0}(u, v) = G_{0,0}(a^m u, a^m v), \quad m = 0, 1, \dots, S-1. \quad (\text{A.2})$$

The corresponding half-peak magnitude contour  $G_{m,0}(u, v) = \frac{1}{2}$  can be represented as

$$\frac{(u - U_m)^2}{A_m^2} + \frac{v^2}{B_m^2} = 1, \quad (\text{A.3})$$

where  $\frac{U_m}{A_m} = U_h/a^m$ ,  $A_m = \sqrt{2 \ln 2} \sigma_u/a^m$ , and  $B_m = \sqrt{2 \ln 2} \sigma_v/a^m$ . Thus, we have

$$U_S = U_h/a^{S-1} = U_l \Rightarrow a = \left( \frac{U_h}{U_l} \right)^{\frac{1}{S-1}}. \quad (\text{A.4})$$

According to the half-peak constraint, elliptical contours  $G_{0,0}(u, v) = \frac{1}{2}$  and  $G_{1,0}(u, v) = \frac{1}{2}$  should be tangent to each other; this leads to

$$U_0 - U_1 = A_0 + A_1 \Rightarrow \sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2 \ln 2}}. \quad (\text{A.5})$$

Furthermore, the elliptical contour  $G_{0,0}(u, v) = \frac{1}{2}$  should be tangent to the elliptical contour  $G_{0,1}(u, v) = \frac{1}{2}$  as well as the line

$$v = u \tan \theta_1, \quad \theta_1 = \frac{\pi}{2K}. \quad (\text{A.6})$$

Substituting (A.6) into  $G_{0,0}(u, v) = \frac{1}{2}$  for replacing  $v$ , we obtain

$$\frac{(u - U_h)^2}{\sigma_u^2} + \frac{(u \tan \theta_1)^2}{\sigma_v^2} = 2 \ln 2. \quad (\text{A.7})$$

Since the line (A.6) is tangent to  $G_{0,0}(u, v) = \frac{1}{2}$ , (A.7) has only one unique solution for  $u$ . That is,

$$\left( -\frac{2U_h}{\sigma_u^2} \right)^2 = 4 \left( \frac{1}{\sigma_u^2} + \frac{\tan^2 \theta_1}{\sigma_v^2} \right) \left( \frac{U_h^2}{\sigma_u^2} - 2 \ln 2 \right) \quad (\text{A.8})$$

Therefore, we have

$$\sigma_v = \tan \left( \frac{\pi}{2K} \right) \left( \frac{U_h^2}{2 \ln 2} - \sigma_u^2 \right)^{\frac{1}{2}}. \quad (\text{A.9})$$

## References

- [1] A. Del Bimbo, Visual Information Retrieval, Morgan Kaufmann, San Francisco, California, 1999.
- [2] J. Nystuen, F. Garcia, Sea ice classification using SAR backscatter statistics, IEEE Transactions on Geoscience and Remote Sensing 30 (3) (1992) 502–509.
- [3] G. Cross, A. Jain, Markov random field texture models, IEEE Transactions on Pattern Analysis and Machine Intelligence 5 (1) (1983) 25–39.
- [4] J. Mao, A. Jain, Texture classification and segmentation using multiresolution simultaneous autoregressive models, Pattern Recognition 25 (2) (1992) 173–188.
- [5] F. Liu, R. Picard, Periodicity, directionality, and randomness: Wold features for image modeling and retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (7) (1996) 722–733.

- [6] J. Beck, A. Sutter, R. Ivry, Spatial frequency channels and perceptual grouping in texture segmentation, *Computer Vision, Graphics, Image Processing* 37 (1987) 299–325.
- [7] G.V. de Wouwer, P. Scheunders, D.V. Dyck, Statistical texture characterization from discrete wavelet representations, *IEEE Transactions on Image Processing* 8 (4) (1999) 592–598.
- [8] T. Chang, C.-C. Kuo, Texture analysis and classification with tree-structured wavelet transform, *IEEE Transactions on Image Processing* 2 (4) (1993) 429–441.
- [9] J. Smith, S. Chang, Image indexing and retrieval based on human perceptual color clustering, *Proceeding of IEEE International Conference on Image Processing* 3 (1994) 407–411.
- [10] J. Daugman, High confidence visual recognition of persons by a test of statistical independence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15 (11) (1993) 1148–1161.
- [11] J. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles, *Vision Research* 20 (1980) 847–856.
- [12] J. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, *Journal of the Optical Society of America, A* 2 (7) (1985) 1160–1169.
- [13] M. Turner, Texture discrimination by Gabor functions, *Biological Cybernetics* 55 (1986) 71–82.
- [14] A. Jain, F. Farrokhnia, Unsupervised texture segmentation using Gabor filters, *Pattern Recognition* 24 (12) (1991) 1167–1186.
- [15] M. Clark, A. Bovik, W. Geisler, Texture segmentation using Gabor modulation/demodulation, *Pattern Recognition Letters* 6 (1987) 261–267.
- [16] I. Fogel, D. Sagi, Gabor filters as texture discriminator, *Biological Cybernetics* 61 (1989) 103–113.
- [17] T.N. Tan, Texture edge detection by modelling visual cortical channels, *Pattern Recognition* 28 (9) (1995) 1283–1298.
- [18] T. Weldon, W. Higgins, D. Dunn, Gabor filter design for multiple texture segmentation, *Optical Engineering* 35 (10) (1996) 2852–2863.
- [19] P. Kruizinga, N. Petkov, Non-linear operator for oriented texture, *IEEE Transactions on Image Processing* 8 (10) (1999) 1395–1407.
- [20] B. Manjunath, W. Ma, Texture features for browsing and retrieval of image data, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18 (8) (1996) 837–842.
- [21] S. Grigorescu, N. Petkov, P. Kruizinga, Comparison of texture features based on Gabor filters, *Optical Engineering* 11 (10) (2002) 1160–1167.
- [23] T. Lourens, N. Petkov, P. Kruizinga, Large scale natural vision simulations, *Future Generation Computer Systems* (10) (1994) 351–358.
- [24] L. Gravano, H. Garcia-Molina, A. Tomasic, Precision and recall of gloss estimators for database discovery, in: *Proceedings of the Third International Conference on Parallel and Distributed Information Systems*, 1994, pp. 103–106.