Local and global Gabor features for raised character recognition

Jianmei Li (李建美)[1], Changhou Lu (路长厚)[1], Xueyong Li (李学勇)[1], and Guoping Li (李国平)[2]

[1] *Department of Mechanical Engineering, Shandong University, Ji'nan 250061*

[2] *Department of Mechanical Engineering, Ji'nan University, Ji'nan 250022*

Received August 20, 2007

Conventional Gabor representation and its extracted features often yield a fairly poor performance in extracting the invariance features of objects. To address this issue, a global Gabor representation method for raised characters pressed on label is proposed in this paper, where the representation only requires few summations on the conventional Gabor filter responses. Features are then extracted from these new representations to construct the invariant features. Experimental results clearly show that the obtained global Gabor features provide good performance in rotation, translation, and scale invariance. Also, they are insensitive to illumination conditions and noise changes. It is proved that Gabor filters can be reliably used in low-level feature extraction in image processing and the global Gabor features can be used to construct robust invariant recognition system.

OCIS codes: 100.0100, 100.3010, 100.5010, 100.5760.

Since raised or indented characters are more changeless for spot and time-change than printed characters, they are widely used on some industrial products. However, images of such characters are of poor contrast because the characters have the same color and material as the background. So recognition of raised or indented characters is difficult than that of conventional optical characters.

To date, only a few literatures appear on studying raised or indented characters and two different approaches have been developed. The first one is to binarize the gray-scale image by choosing an appropriate threshold and then extract feature on the binary images[1]. However, for low-quality images, the binarization process will inevitably result in information loss and generate a lot of broken strokes and noise into the binary images. The other approach works directly on gray-scale images. This approach includes the following categories: 1) direct topographical feature extraction[2] or edge detection from gray-scale images[3]; 2) global feature extraction by circular projection and discrete cosine transform[4]. These methods obtain some improvement in some cases, claiming success on various data. However, in practical applications, some problems still exist. For example, topographical features and edge detectors have poor performance on images with noise or dirty background. Circular projection is sensitive to the accuracy of center point location. The novel feature extraction method based on Gabor filters proposed tries to solve these problems.

Gabor filters[5] have optimal joint localization, or resolution in both the spatial and the spatial-frequency domains. Also, the Gabor-filter-based features seem to be similar to features extracted by humans and, thus, may be effective to be classified. Because of these properties, Gabor filters have been widely applied to computer vision, texture analysis, face recognition, and optical character recognition[6−10].

The two-dimensional (2D) Gabor functions are Gaussian functions modulated by complex valued sinusoids of given frequencies and orientations in the space domain, and shifted Gaussians in the spatial-frequency domain. These Gabor functions are parameterized by five values which control the orientation, the radial frequency bandwidth and the center frequency.

The 2D Gabor function is defined as

$$G^{\phi}(x,y) = G(x,y,\sigma_x,\sigma_y,f_x,f_y,\theta)$$

$$= \mathrm{e}^{-((\frac{X(x,y,\theta)}{\sigma_x})^2+(\frac{Y(x,y,\theta)}{\sigma_y})^2)} \cdot \mathrm{e}^{i(f_x\cdot x+f_y\cdot y)}, \quad (1)$$

where $\phi = (\sigma_x, \sigma_y, f_x, f_y, \theta)$ is the vector form of the five Gabor parameters and $X(x,y,\theta)$, $Y(x,y,\theta)$ are the rotated (x,y) coordinates,

$$X(x,y,\theta) = x\cos(\theta) + y\sin(\theta),$$

$$Y(x,y,\theta) = -x\sin(\theta) + y\sin(\theta).$$

This 2D Gabor filter consists of a Gaussian envelope multiplied by a complex sinusoid. The Gaussian envelope has a spatial extent and a bandwidth determined by (σ_x, σ_y) with the major axes rotated by an angle θ about the z-axis. The complex sinusoid has 2D frequency (f_x, f_y) and an orientation $\gamma = \tan^{-1}(f_y/f_x)$.

The parameters of the Gabor filters were chosen according to the characteristics of the raised characters. Since the width of the character strokes has relation to the frequency, only one properly selected frequency f is needed to extract the stroke information of the characters. Wang *et al.*[10] found that the Gabor response got its maximum value when the frequency value was two times of the width of the Chinese characters. Something different to be noted is that because of the uneven gray distribution of the images, when the width w is defined as half value of the distance between the inner and outer outlines of the character strokes, the Gabor filters are most sensitive. So, the frequency f_x, f_y is defined as

$$f_x = f_y = f_0 = \frac{1}{2w}, \quad (2)$$

where w is half value of the distance between the inner and outer outlines of the character stroke.

The selection of discrete rotation angles θ_i has already been demonstrated by Park *et al.*[11], where it was shown that orientations must be spaced uniformly, that is,

$$\theta_i = \frac{i\pi}{N}, \qquad i = \{1, 2, \cdots, N-1\}, \tag{3}$$

where N is the number of orientations to be used. According to the statistical information of stroke directions of the characters, four orientations are enough to extract the information[10]. Since the stroke information of the raised characters is the same as the Chinese characters, θ_i are also set as 0°, 45°, 90°, 135°.

The response of the form in Eq. (1) at any location (x, y) for an image $I(x, y)$ can be calculated with the convolution as

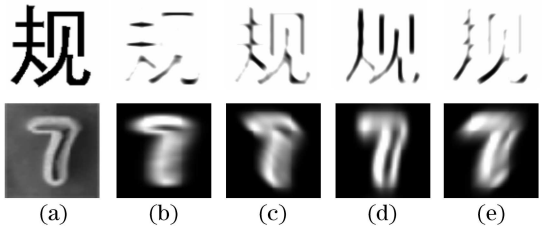
$$I(x, y; f, \theta) = \left| I(x, y) \otimes G^{\phi}(x, y; f, \theta) \right|, \tag{4}$$

where $|\bullet|$ represents a magnitude operator, $\otimes$ represents a convolution operator. The Gabor filter responses in Eq. (4) represent the local features of the image at different scale levels and orientations.

Figure 1 shows the local Gabor responses of a Chinese character "规" and a raised character "7" with the above parameters. The Gabor response in each orientation represents the stroke information of the same direction. The sub-images in Figs. 1(b)—(e) give the stroke information in horizontal, left-diagonal, vertical, and right-diagonal directions respectively. Also, it can be found that the Gabor filters with one single frequency and four orientations are enough to extract the stroke information of the characters.

To overcome the rotation-variance and scale-variance drawbacks encountered in local Gabor features, Han *et al.*[12] proposed an invariant Gabor representation method. Considering a set of images with the same content, except under different orientations, for each image, although the resulted signal energy distribution at each scale level would be different from band to band, the total energy of the Gabor filters yielded in each band tends to be quite constant, regardless the orientation angle of the image 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.

However, unlike rotation invariance, scale invariance is inherently much more complicated. Note that rotation does not change the image. Any drastically down-scaling could result in aliasing and greatly alter the original image 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 is imposed by this assumption. In this paper, the Gabor filters designed for raised characters have only one single scale and four orientations, which satisfy the above assumption.



(a) (b) (c) (d) (e)

Fig. 1. Local Gabor responses of Chinese character "规" and raised character "7" with four orientations (0°, 45°, 90°, 135°) at a single scale level. (a) Original images; (b)—(e) local Gabor responses. It is noted that the frequency f_0 chosen for these two kinds of characters is different.

By calculating all the filters in Eq. (4) with four orientations (θ_i, $i = 1, 2, 3, 4$) at a single scale (f_0), the local Gabor responses are obtained,

$$I(x, y; f_0, \theta_i) = \left| I(x, y) \otimes G(x, y; f_0, \theta_i) \right|, \quad i = 1, 2, 3, 4. \tag{5}$$

A global Gabor feature space can be formed by gathering the local Gabor responses with different frequencies and orientations[13−16]. According to Eq. (5) and the parameters chosen as the above, the global feature matrix is constructed as

$$G_{f_0} = (I(x, y; f_0, \theta_0) \quad I(x, y; f_0, \theta_1) \quad I(x, y; f_0, \theta_2) \quad I(x, y; f_0, \theta_3)). \tag{6}$$

The global Gabor feature at a location (x, y) in Eq. (6) can be summed over the whole image as

$$G_{\mathrm{s}} = \sum_x \sum_y (I(x, y; f_0, \theta_0) \quad I(x, y; f_0, \theta_1) \quad I(x, y; f_0, \theta_2) \quad I(x, y; f_0, \theta_3)). \tag{7}$$

Because the Gabor responses are summed over the whole image, the global Gabor feature in Eq. (7) is translation invariant. Also, because the scale information is retained in the magnitudes of the features, but the ratio between the magnitudes remains over different scales, a scale invariant feature can be constructed by normalizing the global feature matrix in Eq. (6) as

$$G' = \frac{G_{f_0}}{G_{\mathrm{s}}}. \tag{8}$$

In addition to scale invariance, normalization makes the feature also illumination invariant. By summing the Gabor feature in Eq. (8) with different orientations (θ_i, $i = 1, 2, 3, 4$) at a single scale level f_0, a rotation-invariant feature can be obtained[12]. So, the rotation-invariant feature can be calculated as

$$G = \sum_{\theta_i} G', \quad i = 1, 2, 3, 4. \tag{9}$$

Thus, a global Gabor feature with translation, scale and translation invariance is formed. Also, it is insensitive to illumination changes. In order to verify these properties, experiment was carried out with the raised character images of different conditions. The size of the character images is normalized to 50×50. The final global Gabor feature is obtained through calculating Eqs. (5)—(9). Figure 2 shows the results of the proposed global Gabor features.

Figure 2(a) gives the global Gabor features of the character "3" with different rotation angles. It is found that

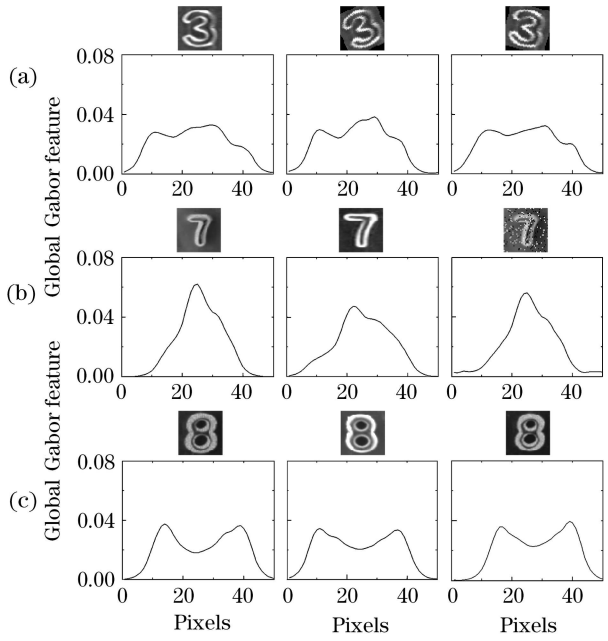


Fig. 2. Global Gabor feature representations of different raised characters. (a) Features of the character "3" with different rotation angles; (b) features of "7" with different sizes and "salt and pepper" noise; (c) features of "8" with different illumination and position conditions.

the features provide good rotation invariance. Figure 2(b) shows the proposed Gabor representations of "7" with different sizes and "salt and pepper" noise. The small change of the Gabor feature tells its good performance in scale invariance and insensitivity to noises. Figure 2(c) demonstrates the features of "8" with different illumination and position conditions, which verify the performance in translation invariance and robustness to illumination conditions.

The recognition experiment was carried out with the images of the raised characters pressed on label. In order to find an optimal classifier for the global Gabor features, the characteristics of the global Gabor feature space is analyzed further.

Figure 3 shows the Gabor features of the raised character "7" and its rotated versions in Eq. (6) respectively. The left column shows the original images and other four

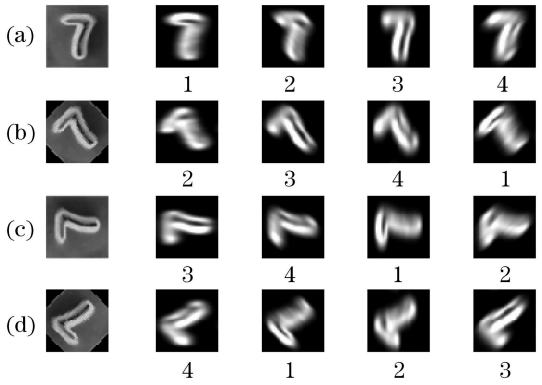


Fig. 3. Gabor features of character "7" and its rotated versions with four orientations (0°, 45°, 90°, 135°). (a) Original image and its Gabor responses; (b)—(d) anti-clockwise rotated versions by 45°, 90°, and 135° respectively.

columns give their filter responses with 0°, 45°, 90° and 135° individually. It can be found from the sub-images in Fig. 3 that when the image is rotated to a certain degree, the local Gabor responses also change to the same angle without changing the shape of Gabor features of the original image. This means that the response of the Gabor filter for a rotated image is equal to the response of the correspondingly rotated filter for the original image without rotation.

It also can be found in Fig. 3 that the sub-images with the same indexing numbers 1, 2, 3, 4 have the same shape but different orientations. So, this makes it possible to find invariant feature of different orientations by using simple matrix shift operations as

$$G^{(\theta+k)} = \left(G\left(1:m, k:n\right) G\left(1:m, 1:k-1\right)\right), \qquad (10)$$

where $G\left(i:j, u:v\right)$ represents the sub-matrix of G containing rows $i, \cdots, j$ and columns $u, \cdots, v$, m is the number of scales and n is the number of the orientations. $k = 1, 2, 3, \cdots, n$ denotes the rotation index. Now, using Eq. (10), a rotation-invariant squared Euclidean distance measure can be defined as

$$d(G_1, G_2) = \min_k \left\{ \sum_{\theta_i} \left[G_1'\left(f_0, \theta_i\right) - G_2'^{(\theta+k)}(f_0, \theta_i) \right]^2 \right\}, \qquad (11)$$

where G_1 and G_2 are feature vectors from two inspected images. $G'\left(f_0, \theta_i\right)$ is the normalized global Gabor feature matrix and $G'^{(\theta+k)}\left(f_0, \theta_i\right)$ is the result of column-wise shift operation in Eq. (10).

The character database is established by collecting the numerals $0-9$ with different orientations and illumination conditions, 30 samples for each numerals. Also, ten samples of each numeral are added with "salt and pepper" noise to generate some new samples. All the images of the samples are normalized to the size of 50×50. In each class of numerals, one undisturbed image was selected to generate a class-specific feature vector. Then the features of a character were extracted and classified to the most similar class by measuring the distance to all class-specific feature vectors according to Eq. (11).

Also, in order to verify the effectiveness of the proposed method, recognition experiments using the existing methods[1−4] were also carried out. Table 1 gives the description of different features and the recognition results.

It can be found from Table 1 that the global Gabor feature performs excellently compared with other methods. The classification accuracy of 98.3% achieved with the proposed global Gabor feature tells that the proposed features have good separate capability. The samples with different conditions successfully classified verify that the proposed feature have good performance in rotation, translation and scale invariance and also provide robustness to illumination and noise changes.

Having the Gaussian shape both in the frequency and spatial domains, the Gabor filters can be considered as spatially concentrated band-pass filters that are localized in the two domains. Thus, disturbances in a distinctly different location do not affect the filter responses. In

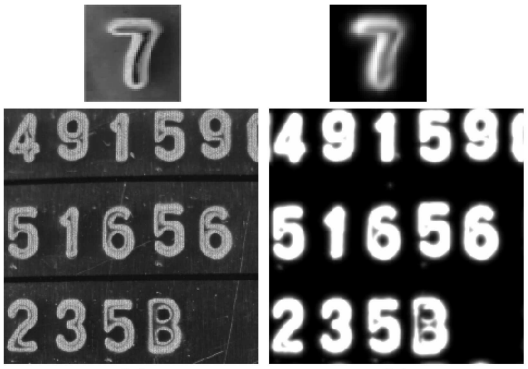
Table 1. Recognition Experimental Results

Feature	Description	Number of Total Samples	Number of Characters Successfully Classified	Recognition Rate (%)
BF	Features Based on Binary Images[1]	400	364	91.0
TF	Topographical Features[2]	400	383	95.8
EF	Edge Features[3]	400	369	92.3
DCT	DCT Transform-Based Method[4]	400	386	96.5
LGAB	Local Gabor Features	400	327	81.8
GGAB	Global Gabor Features	400	393	98.3

DCT: discrete cosine transform.

Table 2. Recognition Rates (in %) of Character Images with Gaussian Noises

Gaussian Noise $(0, \sigma^2)$	0	0.05	0.1	0.5
GGAB	98.3	97.3	96.8	95.5
TF[2]	95.8	93.0	82.3	75.8
DCT[4]	96.5	94.8	91.5	88.3



(a) (b)

Fig. 4. Reconstruction of image using Gabor filter coefficients. (a) Original images of a single character "7" and a whole image; (b) reconstructed images.

order to verify this property of Gabor filters, samples are also added with Gaussian noise $(0, \sigma^2)$ with different variance. The results shown in Table 2 tell that the global Gabor feature has the best performance for images with noises.

Since the Gabor filters can be able to detect the local stroke information of a character, the whole information of the character can also be reconstructed by gathering the responses of the Gabor filters. An iterative threshold technique[17] that minimizes the reconstruction error is introduced to recover the stroke shape from the Gabor responses. The representation power of the Gabor features is demonstrated in Fig. 4 ,which shows the reconstruction images of a single character "7"and a whole image. It can be seen that the Gabor filters are able to capture the key information of the character. Also, the images reconstructed emphasize more details about the character itself and weakened the background simultaneously.

In this paper, the 2D Gabor filters are applied to the raised character recognition. Based on the local Gabor feature of the raised characters, the global Gabor feature is established. Experiment results show that the global Gabor feature has good separate capability and good performance in rotation, translation and scale invariance. Also, the proposed feature has great robustness to illumination and noise changes. Moreover, the Gabor feature has great reconstruction power. Another good point found in this work is that the proposed global Gabor feature extraction method only requires few summations over the filter responses of the conventional Gabor filters, so it costs less computation time compared with other methods. The studies in this paper verify that Gabor filters can be reliably used in low-level feature extraction in image processing and the filter responses can be used to construct robust invariant recognition system.

J. Li's e-mail address is lijianmei@sdu.edu.cn.

References

1. B. Shen, W. Wu, Y. Zhang, G. Shen, and L. Yang, Chin. J. Mech. Eng. **18,** 467 (2005).
2. W. Quan, N. Zheng, and J. Xue, Mini-Micro Systems (in Chinese) **25,** 271 (2004).
3. Y. K. Ham, M. S. Kang, H. K. Chung, R.-H. Park, and G. T. Park, Opt. Eng. **34,** 102 (1995).
4. J. Cao and C. Lu, J. Optoelectron·Laser (in Chinese) **15,** 477 (2004).
5. J. G. Daugman, J. Opt. Soc. Am. A **2,** 1160 (1985).
6. N. Chen and J. Zhong, Chin. Opt. Lett. **1,** 648 (2003).
7. A. Bodnarova, M. Bennamoun, and S. Latham, Pattern Recogn. **35,** 2973 (2002).
8. L.-L. Huang, A. Shimizu, and H. Kobatake, Pattern Recogn. Lett. **26,** 1641 (2005).
9. Y. Hamamoto, S. Uchimura, M. Watanabe, T. Yasuda, Y. Mitani, and S. Tomita, Pattern Recogn. **31,** 395 (1998).
10. X. Wang, X. Ding, and C. Liu, Pattern Recogn. **38,** 369 (2005).
11. H. J. Park and H. S. Yang, Pattern Recogn. Lett. **22,** 869 (2001).
12. J. Han and K.-K. Ma, Image Vis. Comput. **25,** 1474 (2007).
13. V. Kyrki, J.-K. Kamarainen, and H. Kälviäinen, Pattern Recogn. Lett. **25,** 311 (2003).
14. S. Feng, Y. Bao, S. Nie, and L. Wang, Chin. J. Lasers (in Chinese) **34,** 952 (2007).
15. B. Liu and J. Peng, Acta Opt. Sin. (in Chinese) **27,** 1419 (2007).
16. L. Li, B. Guo, and K. Shao, Chin. Opt. Lett. **5,** 332 (2007).
17. Y.-M. Su and J.-F. Wang, Pattern Recogn. **36,** 635 (2003).