Gabor Feature Stabilities for Basic Image Transformations

Keiji YAMADA

Information Technology Research Laboratories, NEC Corporation
4-1-1, Miyazaki, Miyamae-ku, Kawasaki, 216, Japan
yamada@pat.cl.nec.co.jp

Abstract

This paper describes the feature stabilities for four basic image transformations, while examining an image feature extraction method using 2D Gabor filters, as an exmple. The basic transformations are intensity change, scaling, translation, and rotation. Based on the consideration about intensity change and scaling, feature normalization methods were proposed. Moreover, optimal sampling resolutions are determined according to the discussions. Scaling is related with the sampling resolution in the frequency domain and the filter wavelength should be $\sqrt{2}^j$-fold of the smallest filter wavelength, where j is an integer number. The sampling interval in an image domain is a half of a filter wavelength. The sampling interval in a rotation angle domain is $\pi/10$. The Gabor feature extraction method, whose sampling resolutions were determined based on the stability considerations, were examined on printed Japanese character recognition. When sufficient sampling resolution was not attained, the recognition rate became lower. This method can achieve a maximally 96.8% recognition rate, when every sampling resolution is satisfied.

1 Introduction

A primary visual cortex in a monkey brain detects a line segment orientation as a contour orientation[1]. Pattern recognition techniques have positively used orientation features of a target. An orientation feature extraction is easily implemented in terms of matching image parts with 3×3 pixel orientation patterns and is applied to Japanese character recognition [2]. However, such a recognition method misrecognizes patterns whose strokes are rotated and/or translated.

In order to address this problem, a feature extraction method have to generate features with proper stabilities. They should be difficult to influence by insignificant pattern deformations. Furthermore, the marginal stability must be clear for every kind of pattern transformation. Sampling resolution for features should be determined, based on the marginal stability. Too much feature stability, however, prevents patterns resembling each other from being distinguished. On the contrary, there is the possibility of missing essential features, if the feature stability is limited and the feature components are sampled too sparsely.

This paper discusses marginal feature stability and the sampling resolution. They are actually determined for features based on 2D Gabor filters, which is well known to have two dimensional receptive fields, very similar to those for a simple cell in a mammalian visual cortex[3]. Therefore, recognition results using this filter are expected to become plausible. The feature extraction methods constructed are examined in regard to printed Japanese character recognition.
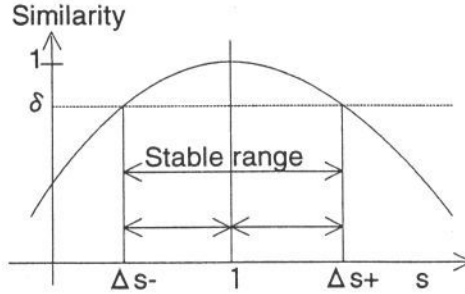
Figure 1: Example showing feature stability on scaling transformation

2   Feature stability and sampling resolution

In pattern recognition, features are extracted from an input pattern and discrimination is carried out using the extracted features. Each component feature value should be stable when its own characteristic part in an input pattern is geometrically transformed a little. When image patterns are considered, there are four kinds of local transformations. They are intensity change, scaling , translation, and rotation[4]. Every image tranformation in a local area is broken down into a combination of them. Local transformations are connected to generate all global image deformations. Therefore, the stability of a feature extraction method has to be considered in each of the local transformations.

Assume that an image pattern is represented as a function $I(\vec{x})$, which is an intensity value at a two dimensional position $\vec{x}$. An intensity change , scaling, translation , and rotation are respectively expressed as

$$
\begin{align}
I'(\vec{x}) &= \alpha I(\vec{x}) + \beta \tag{1}\\
I'(\vec{x}) &= I(s\vec{x}) \tag{2}\\
I'(\vec{x}) &= I(\vec{x} + \Delta\vec{x}) \tag{3}\\
I'(\vec{x}) &= I(\text{Rot}(\theta)\vec{x}), \tag{4}
\end{align}
$$

where $\text{Rot}(\theta)$ is a rotation matrix for an angle θ and the origin is assumed to be at a local image center. In order to detect an input image characteristic part, which corresponds to a standard image part, a feature extracted from a transformed image should be almost the same as an original feature. For example, an original feature vector and a feature vector from an enlarged local image $I(s\vec{x})$ are, respectively, $F(I(\vec{x}))$ and $F(sI(\vec{x}))$. The larger a scaling factor s is, the smaller the similarity between $F(I(\vec{x}))$ and $F(sI(\vec{x}))$ as shown in Fig.1.

Stable features for a specific image transformation are assumed here to be obtained when a similarity is greater than the threshold value δ. In Fig.1, feature stabilities for scaling are guaranteed between Δs_+ and Δs_-. A feature extraction method should be designed so as to maintain a wide stable range for every image transformation category. For this purpose, image normalization methods have been used. To addition, feature extraction methods ought to be adjusted to generate stable features. Furthermore, a sampling resolution and a sampling interval should be optimally determined. In this example, the marginal stable range in a 2 Dimensional (2D) frequency domain is $\Delta s_+/\Delta s_-$. In order to detect stable features without missing important features, features should be sampled in a 2D
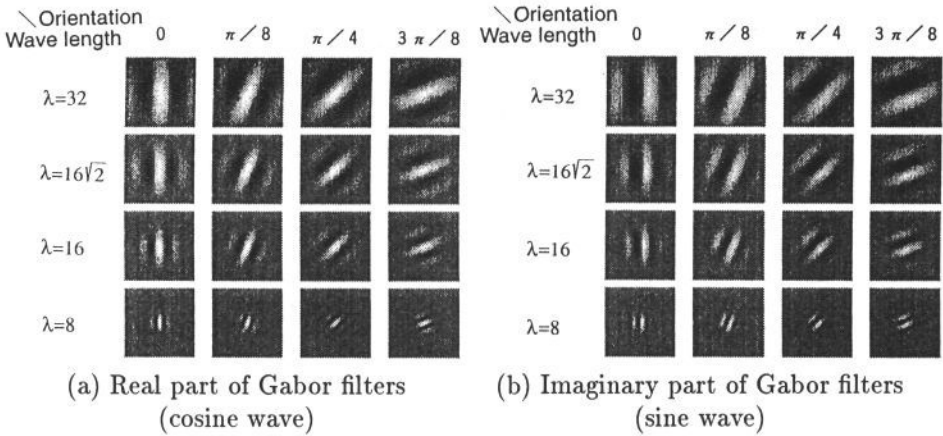
\Orientation Wave length	0	$\pi/8$	$\pi/4$	$3\pi/8$
$\lambda=32$				
$\lambda=16\sqrt{2}$				
$\lambda=16$				
$\lambda=8$				

(a) Real part of Gabor filters
(cosine wave)

(b) Imaginary part of Gabor filters
(sine wave)

Figure 2: Examples of Gabor filters with orientation θ and wavelength λ

frequency domain at the interval of the marginal stability range. In the following section, feature stabilities and sampling resolutions for feature extraction with Gabor filters will be discussed.

3 Gabor feature extraction and normalization

3.1 Multi-resolution feature by 2D Gabor filter

A 2 dimensional (2D) Gabor filter is a 2D Gaussian filter combined with a sine wave, which moves along one direction in a 2D domain[6][3]. The configuration of the 2D Gabor filter has been known to resemble a 2D receptive field for a simple cell in a cat visual cortex[3]. A recognition method, whose feature extraction is implemented using this, would be expected to realize as good a recognition as a human does. A 2D Gabor filter will be simply called a Gabor filter in the following.

When a 2D sine wave is $\vec{\phi} = \frac{2\pi}{\lambda}(\cos\theta, \sin\theta)$, where θ is an angle of a direction in which the wave moves and λ is a wavelength, a Gabor filter is defined as

$$\Psi(\vec{\phi}, \vec{x}) = \frac{1}{\eta}\exp(i\vec{\phi}\cdot\vec{x})\exp(-\frac{2\pi^2(\vec{x}, \vec{x})}{\sigma^2\lambda^2}), \tag{5}$$

where $\vec{x}$ is a point in a filter and the filter center is the origin. The variance parameter for a Gaussian window is $\sigma\lambda/2\pi$. In this paper, the value of a parameter σ is determined as a fixed value π. Figure 2 shows examples of Gabor filters. A response $G(\vec{\phi}, \vec{x})$ is obtained by convolving a filter $\Psi(\vec{\phi}, \vec{x})$ to an image.

$$G(\vec{\phi}, \vec{x_0}) = \int I(\vec{x})\Psi(\vec{\phi}, \vec{x_0} - \vec{x})d\vec{x} \tag{6}$$

This is a feature related with an orientation θ and a wavelength(resolution) λ. For canceling phase shift, a feature value at a point is determined as $F(\vec{\phi}, \vec{x_0}) = |G(\vec{\phi}, \vec{x_0})|$. This increases a stability for translation. A feature vector at a point is a set of $F(\vec{\phi}, \vec{x})$ with several orientations and several wavelengths.

3.2 Normalization for intensity change

Intensity change is a kind of image transformation, which is modeled as Eq.(1). When an intensity at the point $\vec{x}_0$ is changed, both the real part and the imaginary part of a Gabor filter response is obtained below.

$$\Re(G'(\vec{\phi}, \vec{x}_0)) = \alpha \Re(G(\vec{\phi}, \vec{x}_0)) + \beta \frac{\sigma^2 \lambda^2}{2\eta\pi^2} \exp(-2\sigma^2) \tag{7}$$

$$\Im(G'(\vec{\phi}, \vec{x}_0)) = \alpha \Im(G(\vec{\phi}, \vec{x}_0)) \tag{8}$$

When an image intensity is biased with β, the imaginary part of a Gabor filter response remains, but its real part increases. The bias component which includes an intrinsic image bias can be obtained at the Fourier domain origin. The real part of the response can be corrected by subtracting the bias component from the real response parts. After the correction, the feature value is

$$\tilde{F}(\vec{\phi}, \vec{x}_0) = \alpha \hat{F}(\vec{\phi}, \vec{x}_0). \tag{9}$$

The coefficient α is ignored when a recognition process uses the simple similarity, which is an angle between an input feature vector and a standard feature vector.

3.3 Normalization for scaling

When an image is enlarged $1/s$ times, the scaling is modeled in Eq. (2). At that time, the same feature value as the original has to be extracted by using a Gabor filter with $1/s$ times as large a wavelength as an original. When the wave with $1/s$ times the wavelength is expressed as

$$\vec{\phi}' = \frac{2\pi s}{\lambda}(\cos\theta, \sin\theta) = s\vec{\phi}, \tag{10}$$

the response from the enlarged image is expressed as

$$G(\vec{\phi}', \vec{x_0}) = G(s\vec{\phi}, s\vec{x}) = \frac{1}{s^2}G(\vec{\phi}, \vec{x}_0). \tag{11}$$

In order to satisfy the condition that $G(\vec{\phi}', \vec{x_0})$ should be the same as $G(\vec{\phi}, \vec{x}_0)$, the coefficient η in Eq.(5) should be replaced with $1/s^2$.

4 Gabor feature stabilities for transformation

4.1 Stability for scaling transformation

A feature stability for scaling is investigated, when a Gabor filter with a specific wavelength λ is used. A filter response for a sine wave image with the wavelength λ' is calculated. The object patterns are shown in Fig.3. Type A is a sine wave, bounded within one wavelength. Type B is a wave confined within double wavelengths. The response results are shown in Fig.4.

When a filter with 16 pixel wavelength operates on Type A, the wave image with 13.5 pixel wavelength produces the maximum response value. When δ is 0.9 and feature value change by 10% is permitted, the stable range is between 10 pixel

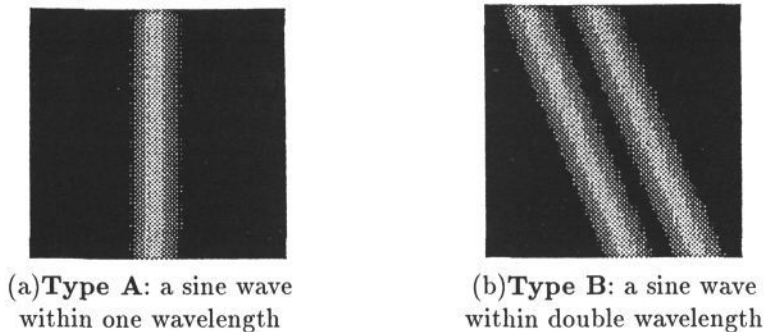| (a)**Type A**: a sine wave within one wavelength | (b)**Type B**: a sine wave within double wavelength |

Figure 3: Sine wave images examined on the feature stability

wavelength and 20 pixel wavelength. The filter response for the sine wave image within one wavelength is comparatively insensitive to the wavelength change.

From Fig.4(b), a filter with 16 pixel wavelength produces the maximum response value, when an image wavelength is the same. This response is more sensitive than the response in the Type A case. Feature value changes by 10% are confined between 13.5 pixel wavelength and 18.5 pixel wavelength. This means that the stable range is about 1.4 ($\sqrt{2}$). This gives the sampling intervals in frequency domain as $\sqrt{2}^j$-fold of the basic wavelength λ.

4.2   Stability for translation

When a translation is modeled as Eq.(3), the response is as follows.

$$
G(\vec{\phi}, \vec{x_0}) = \frac{1}{\eta} \int I(\vec{x'}) \exp i(\vec{\phi} \cdot (\vec{x_0} - \vec{x'})) \exp(-\frac{2\pi^2}{\sigma^2\lambda^2}|\vec{x_0} - \vec{x'}|^2) \exp(i\vec{\phi} \cdot \Delta\vec{x})
$$
$$
\exp(-\frac{4\pi^2}{\sigma^2\lambda^2}\Delta\vec{x} \cdot (\vec{x_0} - \vec{x'})) \exp(-\frac{2\pi^2}{\sigma^2\lambda^2}|\Delta\vec{x}|^2)d\vec{x'} \tag{12}
$$

This response value depends on an input pattern configuration. A part $\exp(i\vec{\phi}\cdot\Delta\vec{x})$ in this equation indicates that, when a translation direction is vertical to a wave ridge, the response value rapidly decreases for the translation. When the pattern is translated along the wave ridge, the response value changes very little.

Filter responses are examined for both Type A and Type B in Fig.3. The results are shown in Fig.5. A response value by a filter with 16 pixel wavelength varies by 10%, when an original image moves in 4 pixels. Totally, a half of a wavelength is a stable range for translation of Type A patterns. Results on Type B patterns indicate that the stable range for translation of double wavelength patterns is twice as large as that in the former case. The worse case is in the Type A case. Therefore, sampling intervals in an image pattern domain should be a half of a filter wavelength.

4.3   Stability for rotation

Filter responses are examined when input sine wave patterns are rotated in angle ψ. Both Type A and Type B patterns were also used in this experiment. In both cases, an input pattern wavelength was 16 pixels. Filter wavelengths were 8, 16,
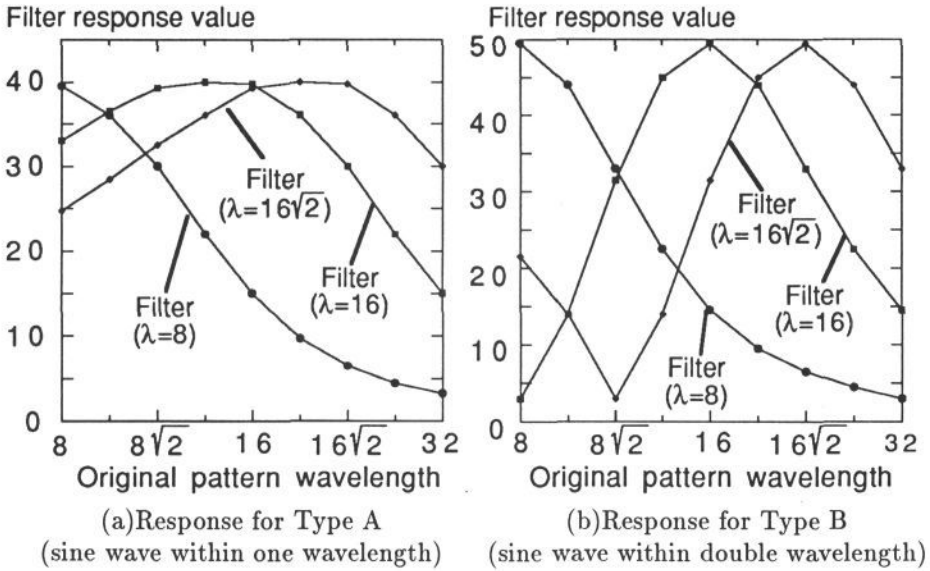
Filter response value

Filter response value

(a)Response for Type A
(sine wave within one wavelength)

(b)Response for Type B
(sine wave within double wavelength)

Figure 4: Original pattern wavelength λ' and filter wavelength λ

Table 1: Gabor feature stabilities for basic image transformations

Image trans-formation	Domain type	Stable range for one wave-length pattern	Stable range for double wave-length pattern	Sampling interval
scaling	log frequency	$\pm(\log 2)/2$	$\pm(\log 2)/4$	$(\log 2)/2$
translation	image pattern	$\pm\lambda/4$	$\pm\lambda/2$	$\lambda/2$
rotation	rotation angle	$\pm\pi/20$	$\pm\pi/20$	$\pi/10$

and $16\sqrt{2}$. The responses are shown in Fig.6. The response decreases little when an input pattern is rotated by 10 degree. In the case of $\pi/16$ rotation, a filter response value decreases by 15%. The results on the Type B pattern is almost the same as the former. In many recognition methods, which use orientation features, orientations are segregated into 8 classes. If an orientation domain is sampled in such sampling resolution, a response change by 15% should be permitted.

4.4 Consideration on sampling resolution

This section reports results obtained by investigating Gabor feature stabilities for basic image transformations for sine wave patterns with one wavelength and double wavelength. The results are listed in Table 1. With regard to scaling, a stable range for double wavelength patterns (Type B) is a half as wide as one for one wavelength patterns(Type A). Therefore, a sampling interval in a log frequency domain is determined to be $\frac{1}{2}\log 2$, according to the narrower stable range. For translation, a sampling interval in an image pattern domain is determined to be half of a wavelength, based on a result for a Type A pattern. In the same way, a
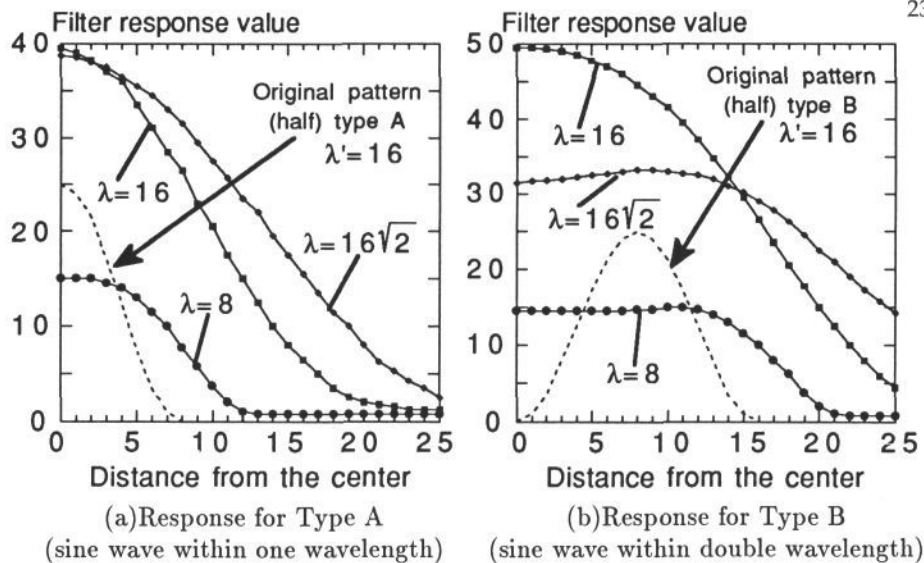
(a)Response for Type A
(sine wave within one wavelength)

(b)Response for Type B
(sine wave within double wavelength)

Figure 5: Filter response value change by pattern translation

sampling interval in orientation angle domain is $\pi/10$.

5 Printed Japanese character recognition

5.1 Features for Japanese characters

The optimal sampling intervals for Gabor features were discussed for a log frequency domain, an image pattern domain, and a rotation angle domain in the previous sections. The features obtained, based on these considerations, are examined regarding printed Japanese character recognition. Japanese characters are comprised of alpha-numerics, symbols, Hiragana characters (cursively written consonant and vowel symbols), Katakana characters (printed consonant and vowel symbols), and Kanji characters (Chinese ideographs). They total about 3000.

Many Kanji characters are composed of 10 and more horizontal and vertical lines with a few slant lines. If the narrowest line width and the narrowest line space width are assumed to be two, the smallest wavelength λ_0 is 4 pixels. Hence, Gabor filter wavelengths should be determined as $\sqrt{2}^j \lambda_0$, where j is an integer number. When a Gabor fileter has a wavelength λ_j, the $\lambda_j/2$ movement of lines can be absorbed in a Gabor feature extraction. If lines were to move in more than a half of the wavelength, points where feature vectors are extracted should be shifted $\lambda_j/2$ by $\lambda_j/2$. Almost all Kanji characters mainly consist of lines in four orientations. Therefore, 8 orientations of Gabor filters are used for feature extraction, though 10 orientations are necessary for 10% divergence.

5.2 Recognition experiments

The following experiments were carried out, in order to examine how great the toleration for the transformations were for a Gabor feature extraction, whose pa-

232

(a)Response for Type A
(sine wave within one wavelength)

(b)Response for Type B
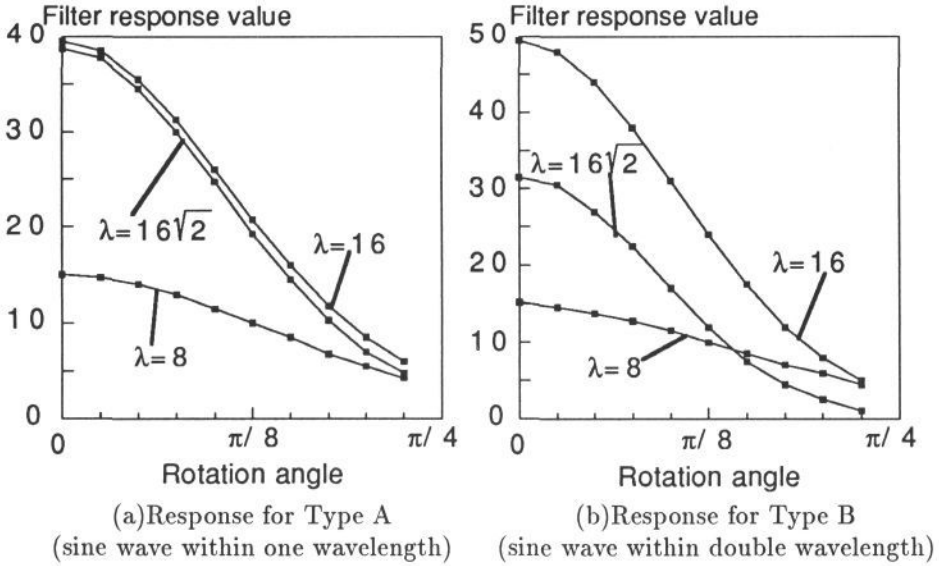(sine wave within double wavelength)

Figure 6: Filter response value change by pattern translation

rameters were determined according to the previous discussions. Printed patterns in one font set are used for making standard patterns. Their examples are listed in Fig.7. A character image size is 64 × 64 pixels.

A feature vector is calculated at a grid position in a character image. When 8 orientation and 6 resolution Gabor filters operate at a grid position, an obtained feature vector has forty-eight feature components. Examples of the filter responses are shown in Fig.8. Feature vectors at all grid points are put together into a vector which is used for similarity calculation.

Grid points in a 64×64 character image are sampled in seven pixel intervals. Totally, 9× 9 grid points are selected. At each grid point, a feature vector is generated. When Gabor filters are used with a wavelength of $\sqrt{2}^j \lambda_0$,($j \in \{0,1\}$, $\lambda_0 = 4$), the sampling interval is larger than the essential sampling interval. Hence, saliency feature vectors may possibly be missed. A relaxation matching method is useful in order to detect the essential feature vector. In this case, the smallest wavelength is 4 pixels. Therefore, feature vectors are measured at points two pixels away from the original grid point. The recognition results are shown in Table 2.

When a fixed point matching is used, the recognition rates in a five-wavelength-case and a four-wavelength-case are larger than that when using the smallest wave-length Gabor filter. This indicates that features with the four pixel wavelength are intolerant of line translation in character patterns. This can be recovered by feature point shift in the relaxation matching. When four orientation Gabor filters are used, features are resampled with a half of the normal resampling interval 2. This recognition rate is almost the same as the recognition rate when feature points are resampled, based on the normal sampling resolution. When the orientation number varies, the recognition rate in the case of eight orientation Gabor filters is better than that in the case of four orientations. The reason is that the latter case doesn't satisfy the necessary sample density in rotation angle domain.
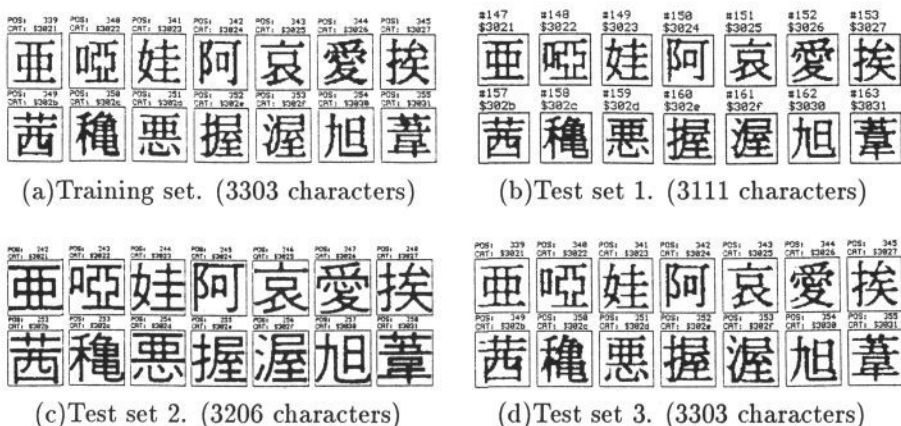
(a)Training set. (3303 characters)

(b)Test set 1. (3111 characters)

(c)Test set 2. (3206 characters)

(d)Test set 3. (3303 characters)

Figure 7: Examples of Japanese characters used for experiments
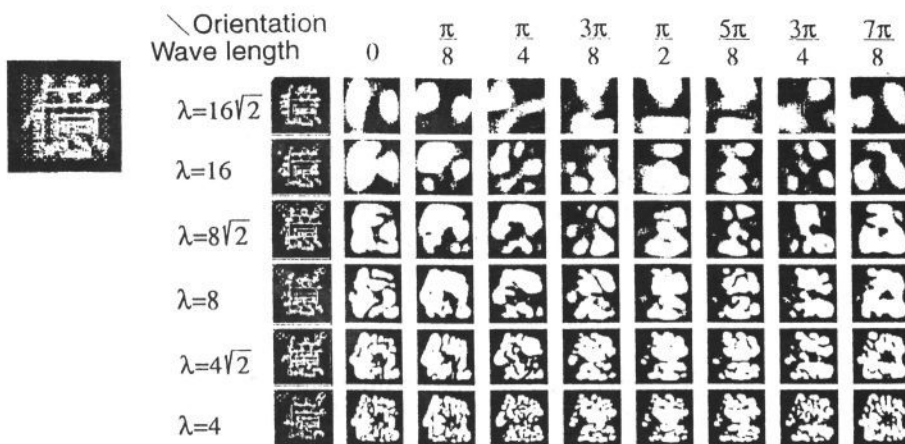

Figure 8: Examples of Gabor filter responses on Kanji characters

6 Concluding remarks

This paper described an image feature extraction method using 2D Gabor filters. At first, the stabilities for the features were discussed for four basic image transformations, which were intensity change, scaling, translation, and rotation. Based on the consideration about intensity change and scaling, the feature normalization methods were proposed. The discussion on the feature stabilities for scaling, translation, and rotation clarified that (1)the stable range for scaling is a $\sqrt{2}$ times as large as a wavelength of a Gabor filter, (2) the stable range for translation is a half of a wavelength of a Gabor filter, and (3) the stable range for rotation is $\pi/10$. These stable ranges directly lead sampling resolutions for a frequency domain, a 2D image pattern domain, and a rotation angle domain.

The Gabor feature extraction method, whose sampling resolutions were determined based on the stability considerations, were examined on printed Japanese character recognition. The recognition rates when the sampleing resolusions determined above were used, were greater than those when the sampling were sparsely

Table 2: Recognition result by fixed point matching and relaxation matching
(F: Fixed point matching, R: Relaxation matching)

Number of wavelengths and (wavelength set)	Number of orienta- tions	Dis. between original grid and additional points	F / R	Recog- nition rate(%)
6 (4, 4√2, 8, 8√2, 16, 16√2)	8	0	F	95.4
6 (4, 4√2, 8, 8√2, 16, 16√2)	8	(2,0,-2)	R	96.6
5 (4√2, 8, 8√2, 16, 16√2)	8	0	F	96.1
5 (4√2, 8, 8√2, 16, 16√2)	8	(2,0,-2)	R	96.8
4 (4√2, 8, 8√2, 16)	8	0	F	96.0
4 (4√2, 8, 8√2, 16)	8	(2,0,-2)	R	96.7
6 (4, 4√2, 8, 8√2, 16, 16√2)	4	(2,0,-2)	R	96.2
6 (4, 4√2, 8, 8√2, 16, 16√2)	4	(2,1,0,-1,-2)	R	96.1

done. The former results are almost equal to the recognition rates when the sampling resolutions were unnecessarily large.

This experiments could not clarify that the frequency sampling resolutions were optimal. The author should examine it. Furthermore, he will investigate stabilities of other features and determine optimal sampling resolutions for them , and will clarify the relations among many kinds of visual feature extractions.

References

[1] D.H.Hubel, T.N.Wiesel., "Functional architecture of macaque monkey visual cortex", Proc. R. Soc. Lon. B, vol. 198, pp.1-59, 1977

[2] J.Tsukumo and H.Tanaka "Classification of Handprinted Chinese Characters Using Non-linear Normalization And Correlation Methods", 9th ICPR, pp.168-171, 1988

[3] J.G.Daugman, "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", IEEE trans. on ASSP, vol.36, No.7, pp.1169-1179, 1988

[4] T.Wakahara, "on-line cursive script recognition using local affine transformation", Proc. 9th Int. Conf. Pattern Recognition, pp.1133-1137, 1988

[5] M.lades, J.C.Vorbruggen, J.Buhmann, J.Lange, C v.d. Malsburg, R.P.Wurtz, and W.Konen, "Distortion Invariant Object Recognition in the Dynamic link Architecture, IEEE Trans. COMPUTER, vol.42, No.3, pp.300-311, 1993

[6] D.Gabor, "Theory of Communication", J. Institute of Elec. Eng., vol.93, pp.429-456, 1947

[7] J.Tsukumo, "Handprinted Kanji Character Recognition based on Flexible Template Matching", 11th ICPR, pp.483-486, 1992