A Tutorial of Wavelet for Pattern Recognition

r99942126 Guan-Chen Pan 潘冠臣
E-mail: r99942126@ntu.edu.tw
Graduate Institute of Communication Engineering
National Taiwan University, Taipei, Taiwan, ROC

Abstract

Wavelet theory is a powerful tool with rich mathematical contents and great applications. It can be employed in lots of fields and applications, such as signal processing, image analysis, communication systems, time frequency analysis, image compression, etc. It is a novel way to do with some fields, and it actually does well in some applications. Pattern recognition is an important technique for us. It also can do lots of applications such as recognize systems or security systems. Today I will introduce the basic concept of how we use wavelet for pattern recognition. I will focus on the pattern recognition on symbols and alphabets, and then state the steps of the process.

Chapter 1 Introduction

Wavelet is a relative recent development mathematic in 1980s, and it can be applied in lots of field, like JPEG2000. JPEG2000 is a new technique for image compression. In the standard JPEG, we use discrete Fourier transform (DCT), and in the JPEG2000, we use discrete wavelet transform to replace DCT. It not only increase the compression ratio but also has better performance and in image. So we try to use wavelet on every field. Now i will introduce wavelet for pattern recognition. Pattern recognition can do with not only wavelet transform but also other transforms and mathematics. As for the applications of wavelet theory to pattern recognition, we can consider them to be two ways：

1. System-component-oriented
2. Application-task-oriented

And Fig. 1.1 displays the two ways. It is clear that the two sides are related to each other. Each group of the right side relates to the component of left side. For instance, singularity detector and invariant representation are related to feature extraction.

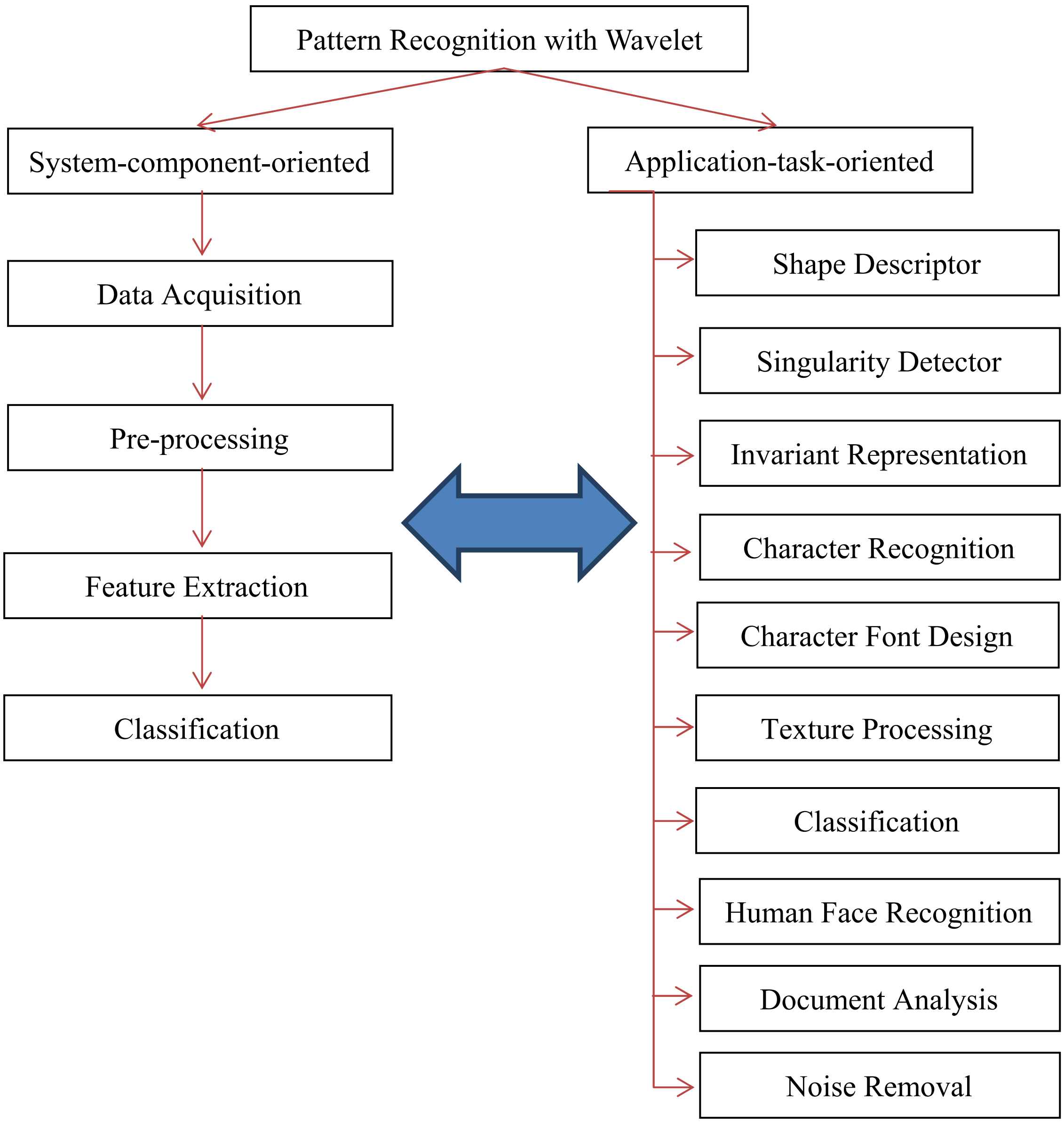


Fig. 1.1 Pattern recognition with wavelet

Pattern recognition can subdivide into face recognition, image recognition, handwritten and printed character recognition, etc. Pattern recognition is an advanced knowledge and there are lots of new techniques arising nowadays, and there is not an absolute answer for pattern recognition. Today i will focus on the pattern recognition in handwritten and printed character recognition, and introduce the details in the below. In this tutorial, there is a basic concept for wavelet theory in Chapter 2, and then Chapter 3 and Chapter 4 are the cores about pattern recognition. At last there is a conclusion in Chapter 5.

Chapter 2 Wavelet Theory

The wavelet transform is a useful tool, and the first wavelet transform is Haar wavelet about 1910. But the concept of wavelet is not popular until 1981, Jean Morlet proposed. After that, the development of wavelet became flourishing. In the 1980s, there are lots of studies about wavelet, such as Daubechies's systematical method to constructs the compact support orthogonal wavelet, and Stephane Mallat and Meyer's multi-resolution concept. However, the complexity of wavelet is still too mass for computation until 1989 Mallat proposed the fast wavelet transform.

- 1910, Haar families.
- 1981, Morlet, wavelet concept.
- 1988, Mallat and Meyer, multiresolution.
- 1988, Daubechies, compact support orthogonal wavelet.
- 1989, Mallat, fast wavelet transform.

2.1. Continuous Wavelet Transform

The definition of continuous wavelet transform is (2.1).

$$X_w(a,b) = \frac{1}{\sqrt{b}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-a}{b}\right)dt \tag{2.1}$$

Where x(t) is the input, and ψ(t) is the mother wavelet. The constant a is location which is any real number, and b is scaling, which is any positive real number. There are lots of types of mother wavelet can be used, such as Haar basis, and Mexican hat function, etc. Mexican hat function is in (2.2), and its figure is in Fig. 2.1.

$$\psi(t) = \frac{2^{5/4}}{\sqrt{3}}\left(1-2\pi t^2\right)e^{-\pi t^2} \tag{2.2}$$

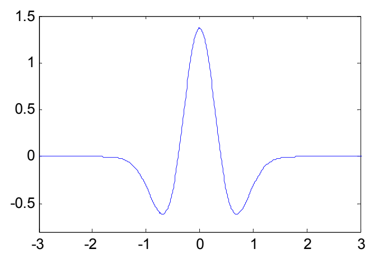


Fig. 2.1 The figure of Mexican hat function

Because the figure of the function is just like a Mexican hat, so it is called

Mexican hat function. The scaling function is in (2.3).

$$\phi(t) = \int_{-\infty}^{\infty} \Phi(f) e^{j2\pi f t} df \tag{2.3}$$

$$\text{where } \left|\Phi(f)\right|^2 = \int_{f}^{\infty} \frac{|\Psi(f_1)|^2}{|f_1|} df_1 \quad \text{for } f > 0, \quad \Phi(-f) = \Phi^*(f).$$

In general, the scaling function is regarding as low pass filter, and the mother wavelet function is regarding as high pass filter. There are two major problems for continuous wavelet transform. One is hard to implement, and another is hard to find $\phi(t)$. To solve the two problems, we use discrete wavelet transform.

2.2. Discrete Wavelet Transform

The concept of discrete wavelet transform is shown in Fig. 2.2.

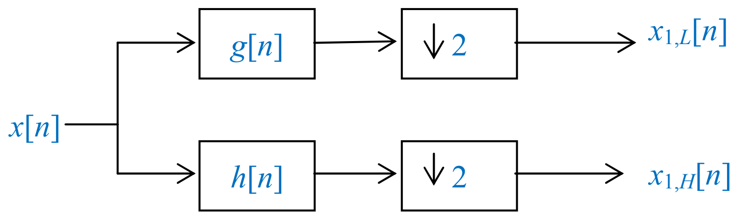


Fig. 2.2 The concept of discrete wavelet transform

Where $x[n]$ is the input, $h[n]$ is the high pass filter, $g[n]$ is the low pass filter, and $\downarrow 2$ is the down-sampling by the factor of 2, $x_{1,L}[n]$ is the output of the low pass filter, and $x_{1,H}[n]$ is the output of the high pass filter. Where the $g[n]$ is just like the mother wavelet function in continuous wavelet transform, and $h[n]$ is the just like the scaling function in continuous wavelet transform. The coefficients of Daubechies filters are usually used for $h[n]$ and $g[n]$. $x_{1,L}[n]$ is the rough part of the input $x[n]$, and $x_{1,H}[n]$ is the detail part of input. In image compression, we usually keep $x_{1,L}[n]$ and discard $x_{1,H}[n]$ to achieve the compression.

2.3. 2D Wavelet Transform

2D wavelet transform is shown in Fig. 2.3. 2D wavelet transform is the combination of two 1D wavelet transform. First we do the 1D wavelet transform along n, and then do the 1D wavelet transform along m.

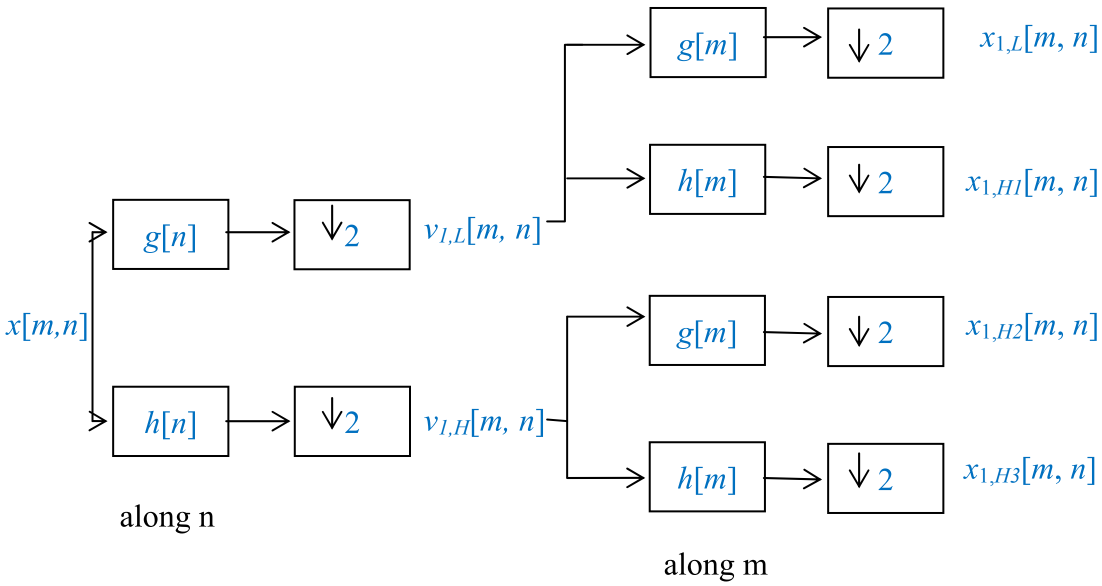


Fig. 2.3 The concept of 2D discrete wavelet transform

When we use the 2D discrete wavelet transform in an image, we will obtain 4 part of output, which the size of each part is one fourth of the original size. Fig. 2.4 is the output of Lena processed by 2D discrete wavelet transform.

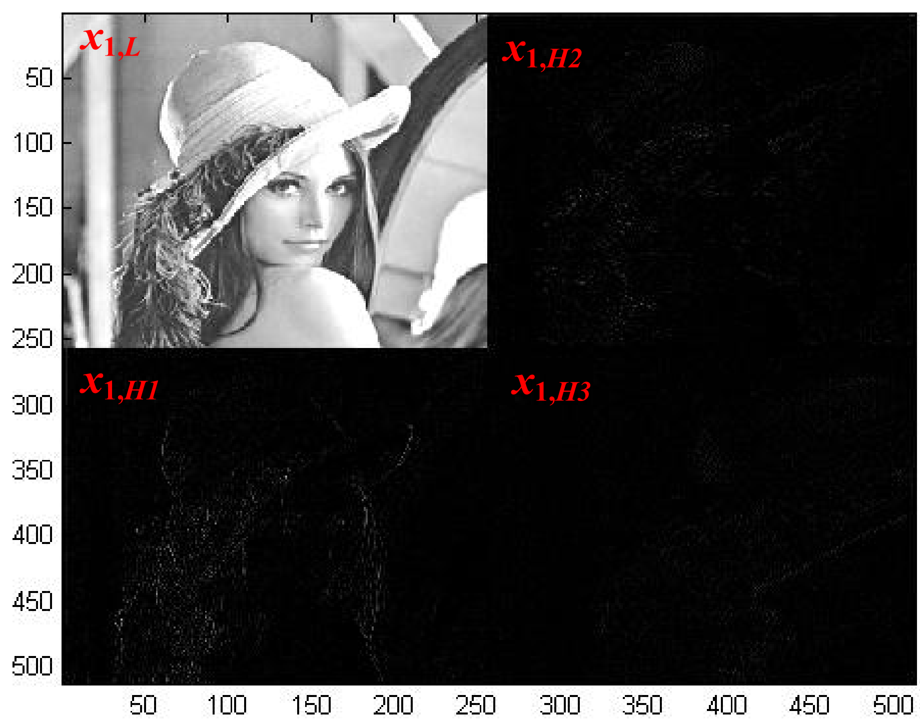


Fig. 2.4 Lena processed by 2D discrete wavelet transform

We can see that the $x_{1,L}$ is just like the original image, and the $x_{1,H1,}$ $x_{1,H2,}$ $x_{1,L}$ are respective corresponding to the horizontal edges, vertical edges, and corners. We can use the characteristic to do image compression.

Chapter 3 Data Acquisition and Data Pre-Processing

In pattern recognition, the entire task can be divided into four parts, just like the left part of Fig. 1.1, data acquisition, data pre-processing, feature extraction and classification. First is the data acquisition.

3.1. Data Acquisition

The first step of pattern recognition is data acquisition. In the real world, the analog data from the physical world are acquired through a transducer, and further digitized to discrete format suitable for computer processing. For example, we photograph the scenery by digital camera, and we get the photo. In digital camera, it capture what the landscape we want to photograph, and change the analog picture into digital data, then save in our memory card. The procedure from we press down the photograph button till save the data in memory card is a kind of data acquisition.

3.2. Data Pre-Processing

A major function of the data pre-processing part is to modify the measured data obtained from the data acquisition part so that those data can be more suitable for the further processing in feature extraction and classification. There are many modifications in the data pre-processing part. For example, some of them are listed below：

- Gray-level histogram modification
- Smoothing and noise elimination
- Edge sharpening
- Boundary detection and contour tracing
- Thinning
- Segmentation
- Morphological processing
- Texture object from textural background
- Approximation of curves and surfaces

In the above modifications, many items, such as noise, edges, boundaries, surfaces, textures, curves, etc., are of singularities in different patterns.

Chapter 4 Feature Extraction

Feature extraction is the heart of a pattern recognition system. In pattern recognition, features are utilized to identify one class of pattern from another. The pattern space is usually of high dimensionality. The objective of the feature extraction is to characterize the object, and further, to reduce the dimensionality of the measurement space to a space suitable for the application of pattern classification techniques. Feature extraction can be view as a mapping, which maps a pattern space into a feature space, and the dimensionality of the feature space has to be smaller than pattern space.

4.1. Methods of feature selection

There are many methods to feature selection, and they can be categorized into the following four parts：

1. **Entropy minimization**

   Entropy is a statistical measure of uncertainty. It can be used as a suitable criterion in the design of optimum feature selection. It is based on the assumption that the pattern classes under consideration are normally distribution.

2. **Orthogonal expansion**

   When the assumption of the normally distribution is not valid, the method of orthogonal expansion is an alternative approach for the feature extraction. Its advantage is that it does not need to know the various probability density functions.

3. **Functional approximation**

   If the features of a class of objects can be characterized by a function that is determined on the basis of observed data, then it is called functional approximation.

4. **Divergence**

   It can be used to determine the feature ranking and to evaluate the effectiveness of class discrimination

4.2. Ring-Projection-Wavelet-Fractal Method

In this chapter, we use ring-projection-wavelet-fractal signatures (RPWFS), and this approach can be categorized into the second method of feature selection listed above. Because of the two dimensional wavelet analysis is not as effective as one dimensional wavelet analysis. So we do the ring-projection-wavelet-fractal. The main idea of the ring-projection-wavelet-fractal method is to reduce the problem of two dimensional patterns into one dimensional pattern, then we can use the 1D wavelet transform to deal with the pattern recognition problem. The diagram of ring-projection-wavelet-fractal method is in Fig. 4.1.

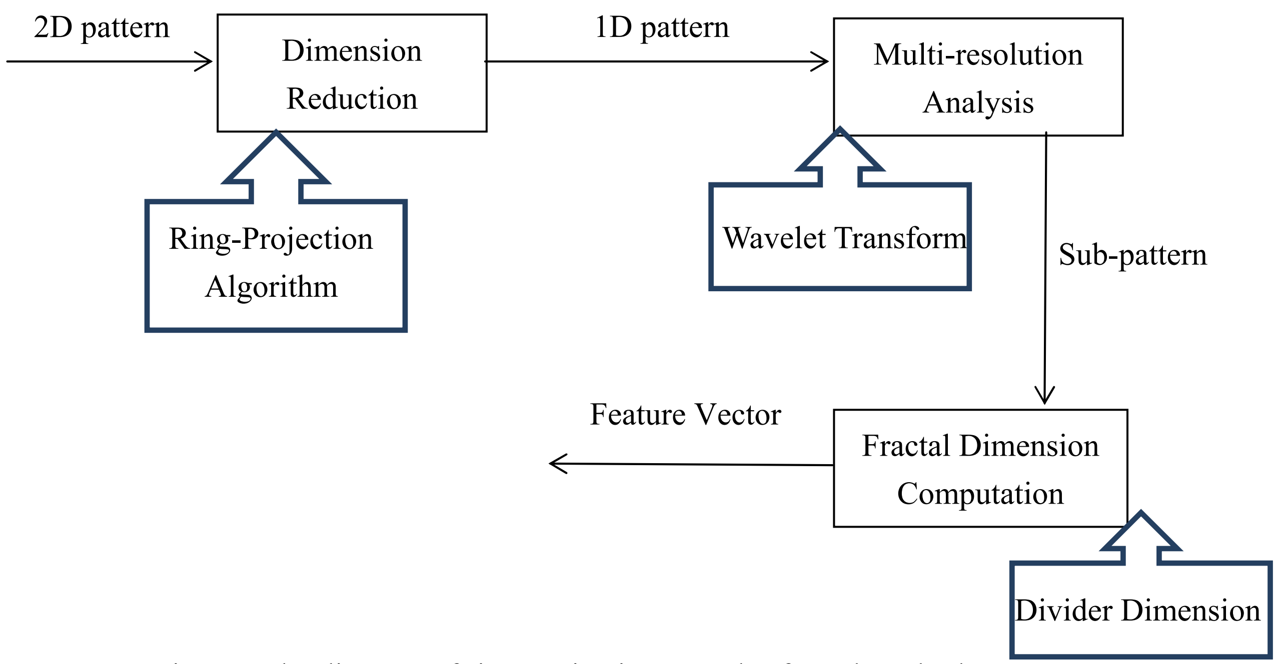


Fig. 4.1 The diagram of ring-projection-wavelet-fractal method

Do the ring-projection-wavelet-fractal method, and we can get the feature vector in the output. Then we can utilize the feature vector and the database to do the classification.

4.3. Dimensionality Reduction by Ring-Projected

At first we suppose that a two dimensional pattern such as an alphanumeric symbol has been represented in a binary image, and its gray scale image $p(x, y)$

can be represented as (4.1) and Fig. 4.2

$$p(x, y) = \begin{cases} 1 & if\,(x,y)\in D \\ 0 & otherwise \end{cases} \tag{4.1}$$

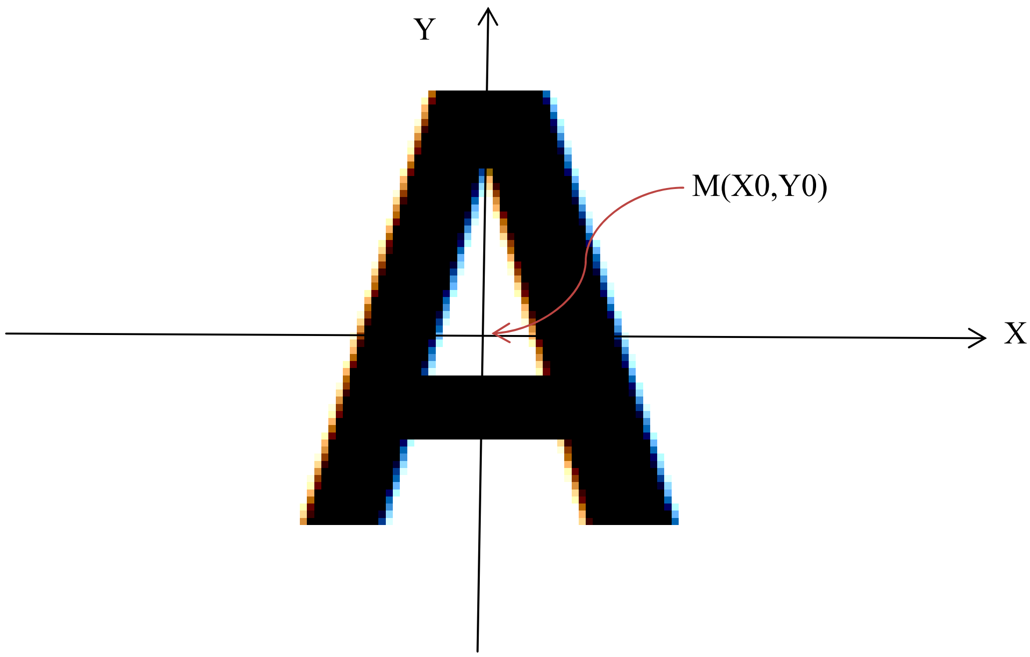


Fig. 4.2 The alphabet A on the x-y plane

where domain D corresponds to the black region of latter. The above function $p(x, y)$ can be viewed as a two dimensional density function of mass distribution on the plane. It is noted that the corresponding density function in (4.1) is uniform distribution. That is to say, the mass is homogeneously distributed over the region D. We let

$$\mathrm{M} = \max_{N\in D} \left| N(x,y) - m(x_0,y_0) \right| \tag{4.2}$$

Where $\left| N(x,y) - m(x_0,y_0) \right|$ represent the Euclidean distance between two points, N and m, on the plane. Then we transform the original x-y space to $\gamma - \theta$ space.

$$\begin{cases} x = \gamma\cos\theta \\ y = \gamma\sin\theta \end{cases} \tag{4.3}$$

Hence,

$$p(x, y) = p(\gamma\cos\theta, \gamma\sin\theta) \tag{4.4}$$

Where, $\gamma \in [0,\infty)$, $\theta \in (0, 2\pi]$

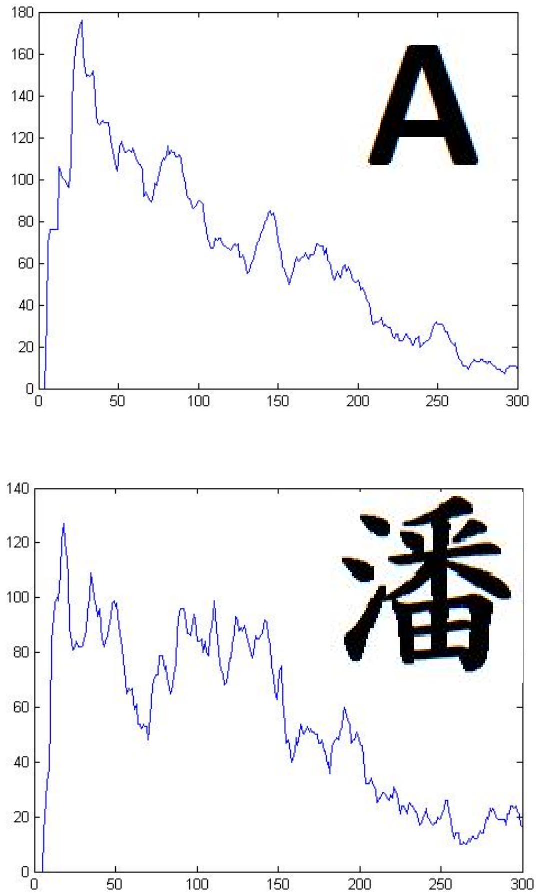
For any fixed $\gamma \in \left[0, M\right]$, we can compute the following integral

$$f(\gamma) = \int_0^{2\pi} p(\gamma\cos\theta, \gamma\sin\theta)d\theta \tag{4.5}$$

The resulting $f(\gamma)$ is one dimensional, and $\gamma \in [0, M]$. So we successfully transform the two dimensional into one dimensional. Due to the facts that the centroid of the mass distribution is invariant to rotation, so the ring-projection is rotation-invariant. Besides, we can get the discrete form for the (4.5) in the (4.6).

$$f(\gamma) = \sum_{k=0}^{M} p(\gamma\cos\theta_k, \gamma\sin\theta_k) \tag{4.6}$$

There are two example for the ring-projection to transform the two dimensional image into one dimensional signal in Fig. 4.3



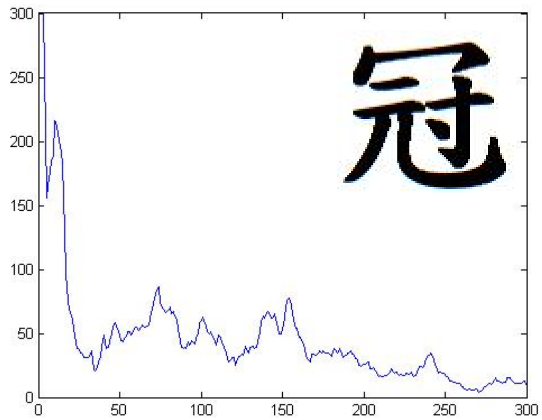


Fig. 4.3 The ring-projection of the image of ‘A’, ‘潘’, and’冠’

4.4. Wavelet Orthonormal Decomposition

Because the property of wavelet transform, the part V_{j-1} can be decomposed as follow

$$V_{j-1} = V_j \oplus W_j \tag{4.7}$$

where $\cdots V_{j+1} \subset V_j \subset V_{j-1} \cdots$.

With this property we can do many times of wavelet transform and decompose the input signal many times, too. We use the Fig. 4.3 to be an example and decompose its signal three times by 1D wavelet transform. The result is Fig. 4.4 and Fig. 4.5

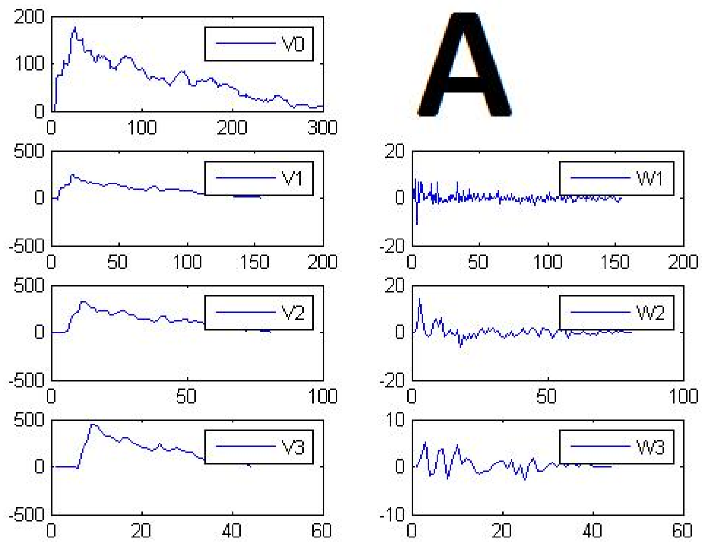


Fig 4.4 the decomposition of the image of ‘A’

Where $V_0 = W_1 \oplus V_1$

$= W_1 \oplus W_2 \oplus V_2$

$$= W_1 \oplus W_2 \oplus W_3 \oplus V_3$$

$$= W_1 \oplus W_2 \oplus W_3 \oplus W_4 \oplus V_4$$

$$= \cdots\cdots$$

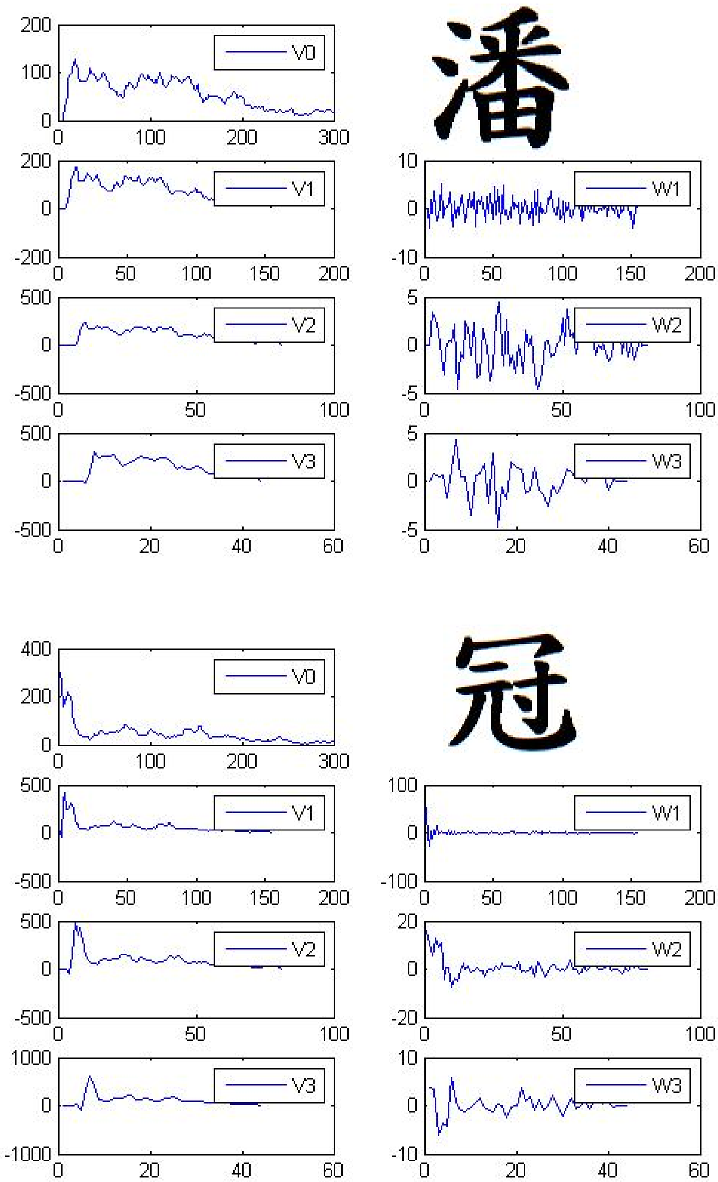


Fig 4.5 the decomposition of the image of the ‘潘’ and ‘冠’

4.5. The Divider Dimension of one dimensional patterns

In the above sections, we have known how to carry out ring-projections to reduce an original two dimensional pattern into one dimensional pattern. Furthermore, we use the 1D wavelet transform to deal with the one dimensional signal to get the sub-patterns, and sub-patterns are non-self-intersecting curves. In this section, we will compute the divider dimension of those curves, and thereafter, use the computed divider dimension to construct a feature vector for the original two dimensional patterns for pattern recognition.

Suppose that C is a non-self-intersecting curve, and $\delta > 0$. Let $M_\delta(C)$ be the maximum number of ordered sequence of points $x_0, x_1, \ldots, x_M$ on curve C, and $|x_k - x_{k-1}| = \delta$ for k=1,2,….,M. Then the divider dimension of curve C can be defined as follow

$$\dim_D C = \lim_{\delta \to 0} \frac{\log M_\delta(C)}{-\log \delta} \tag{4.8}$$

where $|x_k - x_{k-1}|$ represents the magnitude of difference between two vectors.

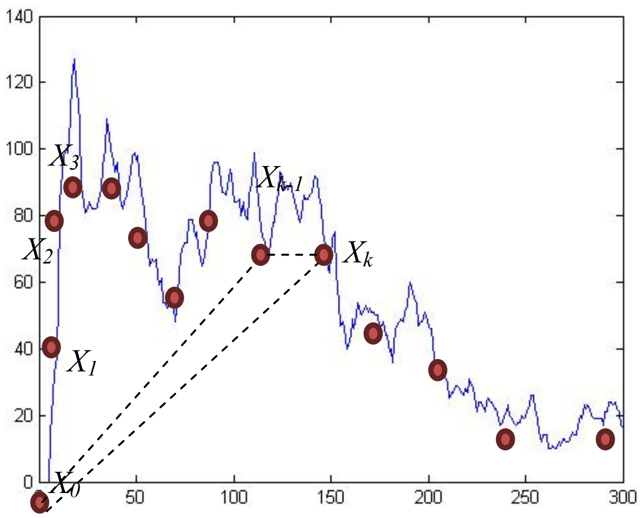


Fig. 4.6 The difference between two vectors on curve C

4.6. Summary

To sum up the above statements, there are three steps for feature extraction as follow.

Step1 Ring-projection of two dimensional patterns

We denote each of the two dimensional patterns by $p(x, y)$. Thus, the ring-projection of $p(x, y)$ can be expressed as (4.9)

$$f(x_k) = \sum_{i=0}^{M} p(x_k \cos\theta_i, x_k \sin\theta_i) \tag{4.9}$$

Step2 Wavelet transform of the one dimensional patterns

Each output of Step 1 can be decomposed by wavelet transform. For example, if we decompose a two dimensional input V_0 three times, it will become as (4.10)

$$V_0 = W_1 \oplus W_2 \oplus W_3 \oplus V_3 \tag{4.10}$$

And we totally have 1 pattern and 6 sub patterns, $V_0, V_1, V_2, V_3, W_1, W_2, W_3$.

Step3 Computation of divider dimensions for wavelet transformed sub-patterns

With the output of Step 2, we compute each patterns and sub patterns to obtain the divider dimensions. The divider dimension is just like (4.8). The above example performs three consecutive wavelet transform and has 7 patterns and sub patterns. Each pattern is a curve, so for each of seven curves, we further compute its divider dimension, and therefore, relate each symbol with a feature vector.

Chapter 5 Classification

After feature extraction, we obtain feature vector. How can we do pattern recognition with feature vector? The answer is classification! With the above procedure, we get the feature vector. To do the pattern recognition, we must have a data bank to compare. To do the above procedure with our data bank, we can also obtain lots of feature vector. Now the problem is how to compare the feature vector of input and the feature vector of data bank. To solve this problem, we need a classifier to classify whether it is the correct or not. In this chapter, we will discuss about the classifier. On approximately, there are three kinds of concept in classifier design as follow

- Minimum average-loss classifier design
- Minimum error-probability classifier design
- Orthogonal wavelet series in classifier design

5.1 Bayes decision theory

In statistical pattern recognition, since the feature vector for a pattern is a random variable, so there is always a probability of committing errors no matter which decision scheme we choose to apply classification. The Bayes decision theory can be represent as (5.1)

$$\frac{p(x|2)}{p(x|1)} \underset{x\in B_2}{\overset{x\in B_1}{\substack{<\\ \geq}}} \frac{P_1(C_{21}-C_{11})}{P_2(C_{12}-C_{22})} \tag{5.1}$$

Where B_1 is the region of 1, and B_2 is the region of 2.C_{ij} is referred to as the lose cause by misclassifying pattern j into pattern i. Bayes decision theory is the basis of the below ways of classifier design.

5.2 Minimum average-loss classifier design

In order to make our discussion concisely, we assume that p_j, $p(x|j)$, and C_{ij} are known. We will further assume that the damage caused by misclassifications is greater than that by correct classification. So $C_{12} > C_{22}$ and $C_{21} > C_{11}$. Under that assumption, we try to find the optimal decision function. Because $C_{12} > C_{22}$ and $C_{21} > C_{11}$, so we can get the following relationships from Bayes decision theory.

$$\text{if } \frac{p(x|2)}{p(x|1)} < \frac{P_1(C_{21}-C_{11})}{P_2(C_{12}-C_{22})}, \quad \text{then } x \in B_1$$
$$\text{if } \frac{p(x|2)}{p(x|1)} \geq \frac{P_1(C_{21}-C_{11})}{P_2(C_{12}-C_{22})}, \quad \text{then } x \in B_2 \tag{5.2}$$

where $\frac{p(x|2)}{p(x|1)}$ and $\frac{P_1(C_{21}-C_{11})}{P_2(C_{12}-C_{22})}$ are referred to as likelihood ratio and decision threshold. It is very similar to the Bayes decision theory, and it will enable us to design a minimum average lose classifier.

5.3 Minimum error-probability classifier design

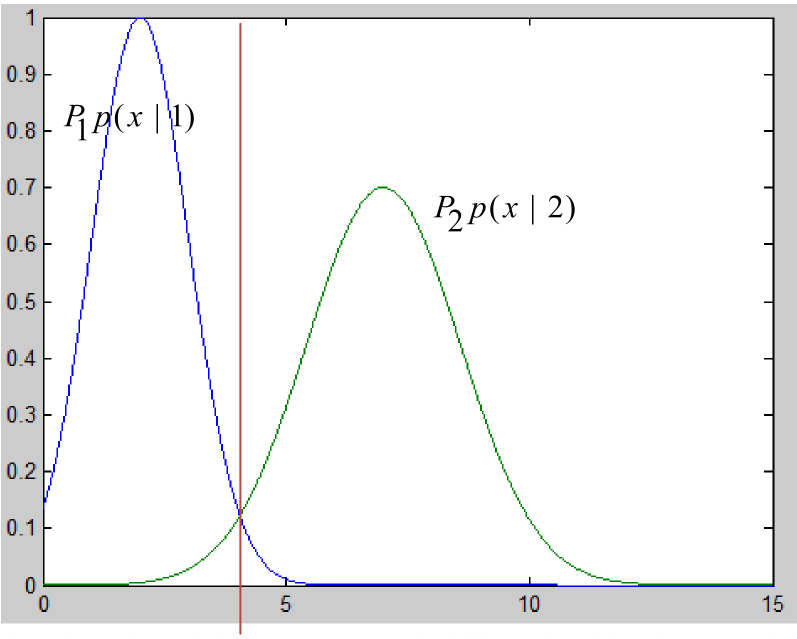
In minimum error-probability classifier design, we will assume that $C_{12} = C_{21}$=1, and $C_{11} = C_{22}$=0. It is just like what we learn in the communication theory. So with the assumption above, we can simplify the Bayes theory to (5.3)

$$\text{if } \frac{p(x|2)}{p(x|1)} < \frac{P_1}{P_2}, \quad \text{then } x \in B_1$$
$$\text{if } \frac{p(x|2)}{p(x|1)} \geq \frac{P_1}{P_2}, \quad \text{then } x \in B_2 \tag{5.3}$$

the probability of total error will be equal to the sum of the probabilities of the two errors in (5.4)

$$R = \int_{B_2} P_1 p(x \mid 1)dx + \int_{B_1} P_2 p(x \mid 2)dx \tag{5.4}$$

We can also use a figure to find the threshold just like Fig. 5.1. In Fig. 5.1, there are two Gaussian distribution function, and the threshold is the red line.



We may notice that the minimum error probability classifier is a special case of the minimum average lose classifier when $C_{12} = C_{21}$=1, and $C_{11} = C_{22}$=0. However, in real life application, it may not be the case that both p_j and $p(x \mid j)$ are known. As far as the problems of unknown $p(x \mid j)$ are concerned, we shall discuss how to estimate $p(x \mid j)$ by applying a method of orthogonal series approximation from the theory of statistical non- parametric estimation.

Orthogonal wavelet series in classifier design is a method that we do not need to know the p_j and $p(x \mid j)$. in the case where $p(x \mid j)$ is unknown, we have to estimate $p(x \mid j)$ based on a set of sample feature vectors. Since the estimation of $p(x \mid j)$ is the same as the estimation of general probability density function $p(x)$. Before discuss about orthogonal wavelet series in classifier design, there are 2 ways to design a classifier without the assumption

- Histogram method
- Orthogonal series probability density estimators

5.4 Histogram method

In density function estimation, the easiest way and the most commonly-used method is histogram method. In this method, we use a series of point, ….< a_{-1}< a_0 < a_1 ….,to subdivide a real domain into a set of disjoint intervals (a_i , a_{i-1}). At each interval, a probability value cab be estimated based on the following calculation：

$$\frac{\#(\{j, a_i < X_j < a_{i+1}\})}{N(a_{i+1} - a_i)} \tag{5.5}$$

where #(A) returns the numbers of elements in set A.

We can construct a small region [x-h_n , x+h_n), and (5.5) can be (5.6)

$$\hat{p}_N(x) = \frac{\#(\{j, 1 \le j \le N, x - h_N < X_j < x + h_N\})}{2Nh_N} \tag{5.6}$$

Now let us define a uniform probability density function, $\hat{K}(x)$, over[-1,1),as follow

$$\hat{K}(x) = \begin{cases} 1/2 & -1 \le x < 1 \\ 0 & otherwise \end{cases} \tag{5.7}$$

and then the (5.6) can be rewrite as (5.8)

$$\hat{p}_N(x) = \frac{1}{Nh_N} \sum_{i=1}^{N} \hat{K}\left(\frac{x - X_i}{h_N}\right) \tag{5.8}$$

5.5 Orthogonal series probability density estimators

As an alternative to the histogram method, we can also use the orthogonal series of a function to estimate the density function. Let $\{\phi_j(x)\}$ be an orthonormal basis, and $\hat{p}(x)$ be an estimator of density function p(x), then we can get (5.9), (5.10)

$$p(x) = \sum_{j=-\infty}^{\infty} c_j \phi_j(x) \tag{5.9}$$

$$\hat{p}(x) = \sum_{j=1}^{m} c_j \phi_j(x) \tag{5.10}$$

We can get c_j by calculating the minimum mean square error γ_e.

$$\gamma_e = \int_{-\infty}^{\infty} \left| p(x) - \hat{p}(x) \right|^2 \partial x$$

$$\hat{p}_m(x) = \sum_{n=-\infty}^{\infty} \hat{a}_{mn} 2^{m/2} \phi(2^m x - n) \tag{5.11}$$

By $\frac{\partial \gamma_e}{\partial c_k} = 0$, k=1,2,…,m, we can get

$$\int_{-\infty}^{\infty} 2(p(x) - \sum_{j=1}^{m} c_j \phi_j(x))\phi_j(x) \partial x = 0 \tag{5.12}$$

Hence

$$\sum_{j=1}^{m} c_j \int_{-\infty}^{\infty} \phi_j(x)\phi_k(x) dx = \int_{-\infty}^{\infty} \phi_k(x) p(x) dx \tag{5.13}$$

We can get (5.14), and it is just like what we learn the projection in linear algebra.

$$c_k = \int_{-\infty}^{\infty} \phi_k(x) p(x) dx \tag{5.14}$$

By calculate c_k using the mean value of N samples, we can obtain (5.15) and

$$\hat{c}_k = \frac{1}{N} \sum_{j=1}^{m} \phi_j(X_i) \tag{5.15}$$

Therefore

$$\hat{p}(x) = \sum_{j=1}^{m} \hat{c}_j \phi_j(x) \tag{5.16}$$

5.6 Orthogonal wavelet series in classifier design

First, we define a bivariate function *L(x ,y)* and $L_m(x,y)$ in (5.17) and (5.18).

$$L(x,y) = \sum_{-\infty}^{\infty} \phi(x-n)\phi(y-n) \tag{5.17}$$

$$L_m(x,y) = 2^m L(2^m x, 2^m y) \tag{5.18}$$

where $\phi(x)$ is a scaling function. From the geometrical point of view, $L_m(x,y)$ is an impulse function scaled from *L(x ,y)*. Let *p(x)* be the probability function and $p_m(x)$ be the orthogonal projection of *p(x)* in space V_m, we can get (5.19)

$$p_m(x) = \sum_{n=-\infty}^{\infty} a_{mn} 2^{m/2} \phi(2^m x-n) \tag{5.20}$$

The minimum mean square error estimator of $p_m(x)$ will be written as (5.21)

$$\hat{p}_m(x) = \sum_{n=-\infty}^{\infty} \hat{a}_{mn} 2^{m/2} \phi(2^m x-n) \tag{5.21}$$

Where

$$\hat{a}_{mn} = \frac{1}{N} \sum_{i=-\infty}^{N} 2^{m/2} \phi(2^m X_i-n) \tag{5.22}$$

To rewrite (5.21) by (5.18), we can get the result as (5.23)

$$\hat{p}_m(x) = \sum_{n=-\infty}^{\infty} [\frac{1}{N}\sum_{i=1}^{N} 2^{m/2} \phi(2^m X_i-n)] 2^{m/2} \phi(2^m x-n)$$

$$= \frac{1}{N}\sum_{i=1}^{N} \sum_{n=-\infty}^{\infty} 2^m \phi(2^m x-n)\phi(2^m X_i-n)$$

$$= \frac{1}{N}\sum_{i=1}^{N} L_m(x, X_i) \tag{5.23}$$

It is just like a combination of histogram method and orthogonal series probability density estimators in the orthogonal wavelet series in classifier design.

Chapter 6 Conclusion

Wavelet theory is a novel mathematic and is also a powerful tool with rich mathematical contents and great applications. Its applications are abundant, and pattern is just one of the applications. For example, image analysis, signal processing, communication systems, time frequency analysis are also applications of wavelet theory.

In this tutorial, we know the basic concept of pattern recognition, and we can roughly categorize the pattern recognition into four steps: data Acquisition, pre-processing, feature extraction, and classification. In the four steps, feature extraction is the most important part for pattern recognition. There are lots of ways in feature extraction, and it can also categorize into four kinds, such as entropy minimization, orthogonal expansion, functional approximation, and divergence. In this tutorial, we use ring-projection-wavelet-fractal method, and it is part of orthogonal expansion. The ring-projection-wavelet-fractal method is good at recognize the alphabets, gestures, and some little icons. However, it is not good at recognize the human face or scenery.

We use the ring-projection-wavelet-fractal method to reduce the problem of two dimensional patterns into one dimensional pattern, then we can use the 1D wavelet transform to deal with the pattern recognition problem. Finally, we get feature vectors and go to classify whether it is belonging to.

There are lots of ways to design the classifier. In the tutorial we introduce minimum average-loss classifier design, minimum error-probability classifier design, and orthogonal wavelet series in classifier design. With the feature vectors and the appropriate classifier, we can recognize the input correctly with high probability!

References

[1] Y. Y. Tang, L. H. Yang, J. Liu, H. Ma, *Wavelet Theory and Its*

Applications to Pattern Recognition, World Scientific

[2] S. Theodoridis, K. Koutroumbas, *Pattern Recognition*, fourth edition, Academic press

[3] J. J. Ding, "Time Frequency Analysis and Wavelet Transform", available in http://disp.ee.ntu.edu.tw/tutorial.php

[4] W.Y. Wei, "Image Compression", available in http://disp.ee.ntu.edu.tw/tutorial.php

[5] C. L. Liu, "A Tutorial of the Wavelet Transform", available in http://disp.ee.ntu.edu.tw/tutorial.php

[6] S. Pittner, S. V. Kamarthi, "Feature Extraction From Wavelet Coefficients for Pattern Recognition Tasks", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 1, January, 1999

[7] B. S. Manjunah, W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", EEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 8, august, 1996