2차원 웨이브렛 패킷에 기반한 필기체 문자인식의 특징선택방법

(A Feature Selection for the Recognition of Handwritten Characters based on Two-Dimensional Wavelet Packet)

김 민 수 † 백 장 선 †† 이 귀 상 ††† 김 수 형 †††

(Minsoo Kim) (Jangsun Back) (Gueesang Lee) (Soohyung Kim)

요 약 본 논문에서는 문자인식의 특징선택방법으로 2차원 웨이브렛 패킷을 이용하는 새로운 방법을 제안한다. 영상자료의 특징들로부터 중심특징을 선택하기위한 차원축소 기법으로 주성분분석 기법이 주로 사용된다. 하지만, 주성분분석 기법은 고유시스템에 의존하기 때문에, 이상치나 잡음등에 민감할 뿐만 아니라, 전역적 특징만을 선택 하는 경향이 있다. 때때로, 영상자료의 중요한 특징이 가장자리 부분이나 뾰족한 부분 같은 지역적 정보일 수 있다. 이러한 경우, 주성분분석 기법은 좋은 결과를 줄 수 없다. 또한 고유시스템은 많은 계산시간을 요구한다.

본 논문에서 원 자료는 2차원 웨이브렛 패킷기저에 의해 변환되고, 최적 판별 기저가 탐색된 후, 그것으로부터 적절한 특징이 선택된다. 주성분분석 기법과 비교하여, 제안된 방법은 웨이브렛의 좋은 특성에 의해 전역적 특징뿐만 아니라 지역적 특징의 선택이 빠른 계산시간으로 이루어진다.

제안된 방법의 성능을 보이기 위해 PCA와 제안된 방법의 인식률의 실험결과가 분석되었다.

키워드 : 문자인식, 특징선택, 주성분분석, 웨이브렛

Abstract We propose a new approach to the feature selection for the classification of handwritten characters using two-dimensional(2D) wavelet packet bases. To extract key features of an image data, for the dimension reduction Principal Component Analysis(PCA) has been most frequently used. However PCA relies on the eigenvalue system, it is not only sensitive to outliers and perturbations, but has a tendency to select only global features. Since the important features for the image data are often characterized by local information such as edges and spikes, PCA does not provide good solutions to such problems. Also solving an eigenvalue system usually requires high cost in its computation.

In this paper, the original data is transformed with 2D wavelet packet bases and the best discriminant basis is searched, from which relevant features are selected. In contrast to PCA solutions, the fast selection of detailed features as well as global features is possible by virtue of the good properties of wavelets.

Experiment results on the recognition rates of PCA and our approach are compared to show the performance of the proposed method.

Key words : Character Recognition, Feature Selection, Principal Component Analysis, Wavelets

† 비 회 원 : 전남대학교 BK21 사업단 post-doc
kimms@chonnam.ac.kr
†† 비 회 원 : 전남대학교 통계학과 교수
jbaek@chonnam.ac.kr
††† 종신회원 : 전남대학교 전산학과 정보통신연구소 교수
gslee@chonnam.ac.kr
†††† 종신회원 : 전남대학교 컴퓨터정보학부 교수
shkim@chonnam.ac.kr
논문접수 : 2001년 3월 15일
심사완료 : 2002년 5월 28일

1. INTRODUCTION

The important goal of discriminant analysis, or pattern recognition, is to explore the relationship of observations originating from different classes as well as to assign unclassified entities into predefined categories. The former is concerned with separating distinct sets of objects and the latter with allocating new objects to predefined groups. Image recognition is one of the major fields in pattern recognition, which comprises the object recognition like classifying handwritten characters [1], recognition of fingerprints, and the face recognition.

A problem which frequently arises in classifying image data is that the dimensionality, i.e., the number of features to be used for the classification, is usually quite large, especially when using statistical classifiers. In such cases, the number of parameters for the statistical pattern recognition model increases rapidly and it makes the parameter estimation very difficult. The purpose of feature selection is to remove any redundant, irrelevant information from the patterns, and in the process, to reduce the dimensionality of the feature vector. Generally, for the reduction of the dimension in classification, PCA[also known as Karhunen-Loève Transform(KLT)] has been often used[2, 3, 4]. Since PCA relies on the eigenvalue system, it is not only sensitive to outliers, but has a tendency to select only global features. Often, the important features for image data, such as edges and spikes, are characterized by local information. If this is the case, PCA fails to extract those features which can be important for the classification. And also eigenvalue systems require expensive operations in solving the problem. The computational complexity of the algorithm for solving the eigenvalue system is $O(n^3)$, where n is the number of pixels in the image.

To overcome such drawbacks, there have been many approaches[5,6] which use one-dimensional (1D) wavelets to 1D applications and have proven to be very effective. Such 1D applications include signal processing on electrocardiogram or on-line handwritten character signals. Even though these methods are designed for 1D applications, they can be extended to 2D applications if 2D data set is mapped to 1D data like a sequence of data elements. In this case, however, spatial information contained in the data set of 2D applications cannot be fully utilized. For example, correlations with pixels located in vertical, horizontal and diagonal directions in an image data cannot be properly represented if the pixels are collapsed into 1D data formats.

In this paper, we develop a procedure called 2D Wavelet Packet Discriminant Bases(2DWP-DB) using 2D wavelet packet for the recognition of image data, specially of hand-written characters. To select relevant features for the recognition of images, the original data is transformed with 2D wavelet packet bases and the best discriminant basis is searched. Based on the best discriminant basis found, relevant features are selected and irrelevant features are discarded. The main characteristic of wavelets is that they provide localized frequency information about image data. Such information is particularly beneficial for classification where detailed features make substantial differences. Also, the whole procedure can be carried out much faster because of the low complexity of computing the wavelet packet basis. The computational complexity of computing the best 2D wavelet packet basis is known as is $O(n\log n)$ for an image which is composed of n pixels [7]. In section 2, basic concepts of 2D wavelet packet functions are given. In section 3, the selection algorithm of 2DWP-DB is explained. Section 4 gives the result of experiments and finally conclusions follow.

2. TWO-DIMENTIONAL WAVELET PACKET FUNCTIONS

In the late 1980's, wavelet with its accompanying multiresolution analysis emerged in the field of applied mathematics, which has made a tremendous impact on fields such as image analysis and signal processing. The wavelet technology introduced in [8], consists of an orthonormal basis for the function space $L^2(R)$, in which all basis elements are formed from a single function by two simple operations of dyadic dilations and integer translations :

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$$

for given the function(the mother wavelet) φ, all wavelets are constructed from the mother wavelet.

Wavelets are functions that satisfy certain requirements. The very name wavelet comes from the requirement that they should integrate to zero, "waving" above and below the x-axis. The diminutive connotation of wavelet suggest the function has to be well localized. There are many kinds of wavelets. One can be chosen among wavelets, compactly supported wavelets, wavelets with simple mathematical expressions, and wavelets with simple associated filters. The most simple one is the Haar wavelet.

$$\varphi(x) = \begin{cases} 1, & 0 \le x < \frac{1}{2} \\ -1, & \frac{1}{2} \le x < 1 \\ 0, & o.w. \end{cases}$$

Like sines and cosines in Fourier analysis, wavelets are used as basis functions in the representation of other functions. Once the wavelet $\varphi(x)$ is fixed, all "wavelets" are constructed from mother wavelet by means of two operations: dilations and translations.

Wavelet packets, a generalization of wavelets bases, are alternative bases that are formed by taking linear combinations of usual wavelet functions. These bases inherit properties such as orthonormality and smoothness from their corresponding wavelet functions. In this paper, the important reason using wavelet packets is to search the best basis can be chosen from the library of basis functions. With respect to wavelet packets, the references include [9,10].

Wavelet packet function is a function with three indices : $w^m_{j,k}(t)$.

$$w^m_{j,k}(x) = 2^{j/2} w^m(2^j x - k)$$

Just as with usual wavelets, integers j and k index dilation and translation operations, respectively. And the extra index $m = 0, 1, 2, \cdots$ is called the modulation parameter.

We can define 2D wavelet packet function by the tensor product of two 1D wavelet packet functions(horizontal function and vertical function).

$$w^{m_h m_v}_{(j,k_h,k_v)}(x_1,x_2) \equiv w^{m_h}_{j_h,k_h}(x_1) \otimes w^{m_v}_{j_v,k_v}(x_2), \quad j = j_h = j_v$$

Any smooth function can be expressed with wavelet packet functions. Denoting I by a suitably chosen set of indices, decomposition of 2D function $f(x_1, x_2)$ into its wavelet packet components is given by

$$f(x_1,x_2) = \sum_{(m_h m_v, j)\in I} \sum_{k \in Z} a^{m_h,m_v}_{j,k_h,k_v} w^{m_h m_v}_{(j,k_h,k_v)}(x_1,x_2),$$

where the coefficients are computed via

$$a^{m_h,m_v}_{j,k_h,k_v} = \iint w^{m_h m_v}_{(j,k_h,k_v)} f(x_1,x_2)\,dx_1 dx_2.$$

The example of 2D wavelet decomposition of image '4' is given in Figure 1.

Suppose we can express the wavelet packet transform in a vector-matrix form as

$$\boldsymbol{\alpha}_j = \mathbf{W}_j^T \mathbf{x}$$

where $\boldsymbol{\alpha}_j \in \Re^{n\times n}$ contains the wavelet packet coefficients, and $\mathbf{W}_j \in \Re^{n\times n}$ is an orthogonal wavelet packet transform matrix. Then $\boldsymbol{a}^{2'} = \boldsymbol{a}_j \cdot \boldsymbol{a}_j$ (where product " · " means Hadamard product) means the energy at each position on level j, it has been used in the wavelet scalogram[11, 12]

2-D wavelet packet coefficients using Haar filter

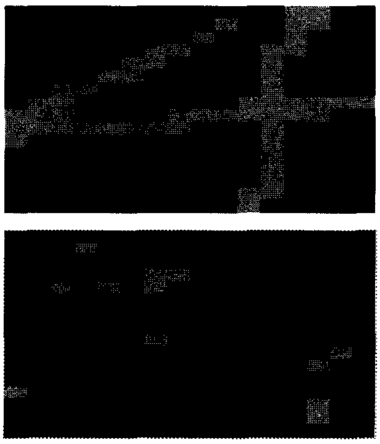

Fig. 1 An example of 2-D wavelet decomposition for a image '4'
The original data (left), 2-D wavelet packet coefficients(level 1) of original data (right)

for the image '4' of 8×8pixels are shown in Figure 2. Going from left to right, top to bottom, the first picture means the original data. Also, it is considered as wavelet coefficients $a_0^{0,0}$ under $w_0^{0,0}$ basis ($w_0^{0,0} \equiv w_0^0 \otimes w_0^0$, that is, the tensor product of a horizontal 1-D wavelet packet function w_0^0 and a vertical 1-D wavelet packet function w_0^0, and $w_j^m = (w_{j,0}^m, w_{j,1}^m, \cdots, w_{j,8\times 2^j-1}^m)$). The second one is $a_{-1}^{0,0}$, $a_{-1}^{0,1}$, $a_{-1}^{1,0}$, $a_{-1}^{1,1}$, the expression by combinations of 4 - partial basis functions $w_{-1}^{0,0}$, $w_{-1}^{0,1}$, $w_{-1}^{1,0}$, $w_{-1}^{1,1}$. And third one and forth one are $a_{-2}^{0,0}, a_{-2}^{0,1}, \cdots, a_{-2}^{3,3}$ and $a_{-3}^{0,0}, a_{-3}^{0,1}, \cdots, a_{-3}^{7,7}$, respectively. Those are calculated under basis functions $w_{-2}^{0,0}, w_{-2}^{0,1}, \cdots, w_{-2}^{3,3}$ and $w_{-3}^{0,0}$, $w_{-3}^{0,1}, \cdots, w_{-3}^{7,7}$, respectively.

From a set of basis functions, we can select any special basis if we want. Figure 3 is an example wavelet packet coefficients with arbitrarily supposed basis. If this is the case, new basis consists of following partial basis functions of each level, $w_{-1}^{1,1}$ (level 1), $w_{-2}^{2,0}$, $w_{-2}^{3,0}$, $w_{-2}^{2,1}$, $w_{-2}^{0,1}$, $w_{-2}^{1,1}$, $w_{-2}^{0,2}$, $w_{-2}^{0,3}$, $w_{-2}^{1,3}$ (level 2), $w_{-3}^{0,0}$, $w_{-3}^{1,0}$, $w_{-3}^{0,1}$, $w_{-3}^{1,1}$, $w_{-3}^{2,0}$, $w_{-3}^{3,0}$, $w_{-3}^{2,1}$, $w_{-3}^{3,1}$, $w_{-3}^{6,2}$, $w_{-3}^{7,2}$, $w_{-3}^{6,3}$, $w_{-3}^{7,3}$, $w_{-3}^{2,4}$, $w_{-3}^{3,4}$, $w_{-3}^{2,5}$, $w_{-3}^{3,5}$ (level 3).

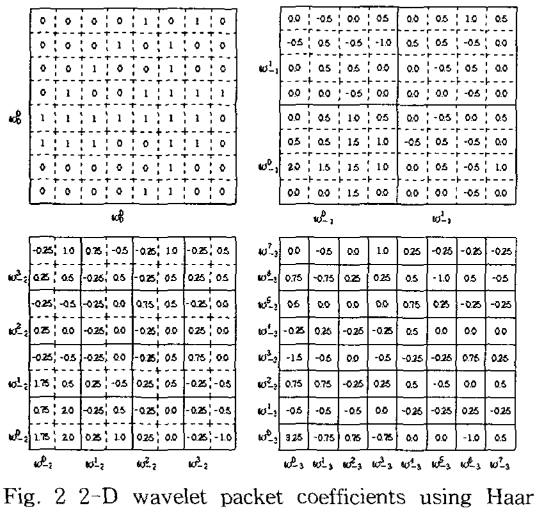

Fig. 2 2-D wavelet packet coefficients using Haar filter for the image '4' of pixels. Going from left to right, top to bottom, original data(level 0), level -1, level -2, level -3, $w_j^m = (w_{j,0}^m, w_{j,1}^m, \cdots, w_{j,8\times 2^j-1}^m)$.

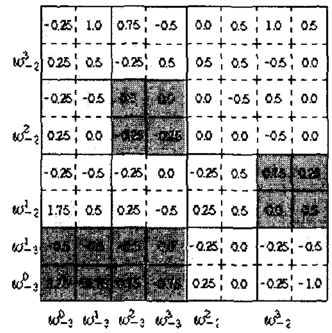

Fig. 3 An example of coefficients of new basis in 2D wavelet packet basis

3. PROPOSED ALGORITHM : 2DWP-DB

The approach to the feature selection proposed in this paper is guided by the so-called best-basis paradigm in [9]. That is :

- Select the best basis(or coordinate system) for the problem from a set of bases.
- Sort the features by importance for the problem and discard unimportant features.
- Use the surviving features to solve the problem.

The criteria for selecting the best basis and the definition of important features can be different by the requirement of the given problem. If we want to apply the above approach to the signal compression, a basis for the global components of the coordinate vectors is selected so that we can discard the other detail components without much signal degradation. For the regression, a basis through which we can see the essential relationships between input signals and output responses of interest should be selected. For the classification purpose, a basis in which classes are maximally-separated is the choice. Since we focus on the classification of images in this paper, we identify maximally-separated classes.

To identify the wavelet basis which transforms the original images into maximally-separated group of images, we need to consider a discriminant measure among groups of images. Although there are many choices for the discriminant measure [13], we use $W=\|\mathbf{p}-\mathbf{q}\|^2=\sum_{i=1}^{n}(p_i-q_i)^2$ as discriminant measure for simplicity, where $\mathbf{p}=\{p_i\}_{i=1}^{n}$, $\mathbf{q}=\{q_i\}_{i=1}^{n}$ be two nonnegative sequences with $\sum p_i=\sum q_i=1$. The discriminant measure W is known to be adequate when images are normalized and sizes of sample images are same. Also, the algorithm for classification become fast when the discriminant measure is additive.

Definition 1. The discriminant measure $D(\mathbf{p},\mathbf{q})$ is said to be additive if

$D(\mathbf{p},\mathbf{q})=\sum_{i=1}^{n}D(p_i,q_i)$, for $\mathbf{p}=p_1,p_2,\cdots,p_n$, $\mathbf{q}=q_1,q_2,\cdots,q_n$,

It is obvious the measure W is additive.

Suppose that there are L image groups, $\mathbf{p}^1,\mathbf{p}^2,\cdots,\mathbf{p}^L$, $\mathbf{p}^i$ where denotes ith group of sample images. And to measure differences among L image groups, we may consider another measure Z which is sum of pairwise combinations of measure D.

$$Z(\mathbf{p}^1,\mathbf{p}^2,\cdots,\mathbf{p}^L)\equiv\sum_{i=1}^{L-1}\sum_{j=i+1}^{L}D(\mathbf{p}^i,\mathbf{p}^j)$$

For inputs to the discriminant measure Z, the energy function of each image is used. Suppose the wavelet coefficients for a image for the resolution level j is computed as in Figure 2. Then the value of the energy function of at an arbitrary position (a,b) is computed by c^2/T, where, T is an arbitrary position (a,b), is the total energy of the original image or $T=\sum_i\sum_j p_{ij}^2$ where p_{ij} is the value of the original pixel in position i,j.

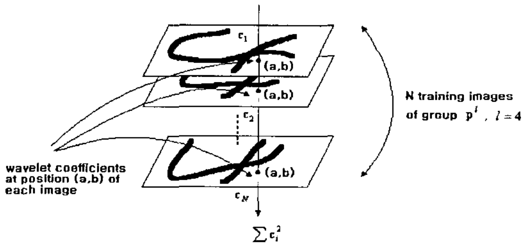

Fig. 4 A diagram for calculation of energy function value at position (a,b) at each image

Now, like Figure 4, suppose that we have N training images in the group $\mathbf{p}^l=\{p_1^l,p_2^l,\cdots p_N^l\}$, then the value of the energy function at an arbitrary position of the group $\mathbf{p}^l$ is computed by

$$\frac{\sum_{k=1}^{n}c_k^2}{T^*},\text{ where } T^*=\sum_{k=1}^{N}T_k.$$

The energy function can be defined formally as follows.

Definition 2. Suppose that sample images are divided into a set of groups $\mathbf{p}^1,\mathbf{p}^2,\cdots,\mathbf{p}^L$. Each image $\mathbf{x}_i$ is size $n\times n$. Let $\{\mathbf{x}_1^l,\mathbf{x}_2^l,\cdots,\mathbf{x}_N^l\}$ be a set of training images belonging to class $\mathbf{p}^l$. Then the energy function of $\mathbf{p}^l$ for j th resolution level with translation parameters, k_h, k_v, modulation parameters, m_h, m_v, denoted by $\Gamma_l(j,k_h,k_v,m_h,m_v)$ is a set of real values specified as

$$\Gamma_l(j,k_h,k_v,m_h,m_v)\equiv\frac{\sum_{i=1}^{N}((\mathbf{w}_{j,k_h,k_v}^{m_h,m_v})^T\mathbf{x}_i^l)^2}{\sum_{i=1}^{N}<\mathbf{x}_i^l>^2}.$$

where $j=0,-1,-2,\cdots,-lob_2\frac{n}{2}$, $k_h,k_v=0,\cdots,n\times2^j-1$, size of $\mathbf{x}_i^l=n\times n$, $\sum_{i=1}^{N}\langle\mathbf{x}_i^l\rangle^2=\sum_{i=1}^{N}\sum_a^n\sum_b^n\mathbf{x}_i^{l^2}(a,b)$.

$\Gamma_l(j,k_h,k_v,m_h,m_v)$ is computed by accumulating the squares of wavelet coefficients of images at each position and divided by the total energy of the images belong to class l. This normalization is important if there are significant differences in number of samples among classes.

If an image of $n_1\times n_2$ pixels is given, it is converted to a square image of $n\times n$ pixels with uniform positions, sizes, angles, which is called 'shape normalization'. In this paper, nonlinear shape normalization using inscribed circle [14] is used. The examples of normalization for numeral '8' are shown in Figure 5.

Our algorithm selects 2DWP-DB which maximizes the discriminant measure on the energy functions of classes. Let $B_j^{m_h,m_v}$ denote a set of basis matrices at subspace $\mathbf{w}_j^{m_h,m_v}$. Let $A_j^{m_h,m_v}$ denote a set of discriminant basis matrix. Also, let $\Delta_j^{m_h,m_v}$ be a work matrix containing the discriminant measure of subspace $\mathbf{w}_j^{m_h,m_v}$.

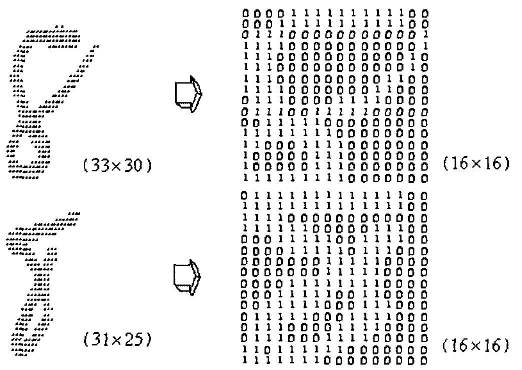

Fig. 5 The example of normalization using NSN method for image '8'

In step 1, we calculate the energy functions Γ_l, $l=1,\cdots,L$ of groups $\mathbf{p}^1, \mathbf{p}^2, \cdots \mathbf{p}^L$ as defined in Definition 2. In step 2, initially $B_0^{0,0}$ is assigned to $A_0^{0,0}$, $\Delta_0^{0,0}$ is calculated. For $j=0$, there is only one basis for $\mathbf{w}_0^{m_h,m_v}$(which is $B_0^{m_h,m_v}$) because $j=0$ is the maximum depth of decomposition. In step 3, for $j=-1$ let $A_{-1}^{m_h,m_v}$ be a basis, then either $A_{-1}^{m_h,m_v}=B_{-1}^{m_h,m_v}$ or $A_{-1}^{m_h,m_v} = A_{-2}^{m_h\times 2,m_v\times 2} \oplus A_{-2}^{m_h\times 2,m_v\times 2+1} \oplus A_{-2}^{m_h\times 2+1,m_v\times 2} \oplus A_{-2}^{m_h\times 2+1,m_v\times 2+1}$.

If $\Delta_{-1}^{m_h,m_v} \geq \Delta_{-2}^{m_h\times 2,m_v\times 2} + \Delta_{-2}^{m_h\times 2,m_v\times 2+1} + \Delta_{-2}^{m_h\times 2+1,m_v\times 2} + \Delta_{-2}^{m_h\times 2+1,m_v\times 2+1}$ then $A_{-1}^{m_h,m_v}=B_{-1}^{m_h,m_v}$ else $A_{-1}^{m_h,m_v}= A_{-2}^{m_h\times 2,m_v\times 2} \oplus A_{-2}^{m_h\times 2,m_v\times 2+1} \oplus A_{-2}^{m_h\times 2+1,m_v\times 2} \oplus A_{-2}^{m_h\times 2+1,m_v\times 2+1}$. Then by inductive rule, the best 2DWP-DB $A_j^{m_h,m_v}$ is determined, in the decreasing order of $j(j=-1,-2,\cdots,-\log_2\frac{n}{2})$.

After this step, we have a 2DWP-DB. From the repetition of STEP 3, once 2DWP-DB is selected, we can use all expansion coefficients of images in the basis as features.

However, the two subsequent steps are still necessary because we want to reduce the dimensionality of features. In STEP 4, we have to calculate a measure of discriminant power of an individual basis function. STEP 5 reduces the dimensionality of the problem from to without losing the discriminant information in terms of energy functions among classes.

Algorithm

Given a training data set consisting of L classes of normalized images $[\{\mathbf{x}_i^l\}_{i=1}^{N_l}]_{l=1}^{L}$, $\mathbf{x}_i^l$ is an image of $n\times n$

STEP 1: Construct the energy functions Γ_l, $l=1,\cdots,L$

STEP 2: SET $A_J^{m_h,m_v}= B_J^{m_h,m_v}$,

$\Delta_J^{m_h,m_v} = D(\{\Gamma_l(J,\cdot,\cdot,m_h,m_v)\}_{l=1}^{L})$,

initially, $J=0, m^h, m^v=0$

STEP 3: Determine the best subspace $A_j^{m_h,m_v}$.

for $j=-1,-2,\cdots,-\log_2\frac{n}{2}$, $m_h, m_v=0,\cdots,2^{-j}-1$, by the following rule.

SET $\Delta_j^{m_h,m_v}=D([\Gamma_l(j,\cdot,\cdot,m_h,m_v)]_{l=1}^{L})$

IF $\Delta_j^{m_h,m_v}\geq\Delta_{j-1}^{m_h\times 2,m_v\times 2}+\Delta_{j-1}^{m_h\times 2,m_v\times 2+1}+\Delta_{j-1}^{m_h\times 2+1,m_v\times 2}+\Delta_{j-1}^{m_h\times 2+1,m_v\times 2+1}$

THEN $A_j^{m_h,m_v}=B_j^{m_h,m_v}$

(where $\Delta_j^{m_h\times 2,m_v\times 2}$ is smooth level of $\Delta_j^{m_h,m_v}$,

$\Delta_{j-1}^{m_h\times 2,m_v\times 2+1}$ is horizontal detail level of $\Delta_j^{m_h,m_v}$,

$\Delta_{j-1}^{m_h\times 2+1,m_v\times 2}$ is horizontal detail level of $\Delta_j^{m_h,m_v}$,

$\Delta_{j-1}^{m_h\times 2+1,m_v\times 2+1}$ is diagonal detail level of $\Delta_j^{m_h,m_v}$)

ELSE

$A_j^{m_h,m_v}= A_{j-1}^{m_h\times 2,m_v\times 2}\oplus A_{j-1}^{m_h\times 2,m_v\times 2+1}\oplus A_{j-1}^{m_h\times 2+1,m_v\times 2}\oplus A_{j-1}^{m_h\times 2+1,m_v\times 2+1}$

$\Delta_j^{m_h,m_v} = \Delta_{j-1}^{m_h\times 2,m_v\times 2} + \Delta_{j-1}^{m_h\times 2,m_v\times 2+1} + \Delta_{j-1}^{m_h\times 2+1,m_v\times 2}+\Delta_{j-1}^{m_h\times 2+1,m_v\times 2+1}$

STEP 4: Order the wavelet packet coefficients by their power of discrimination.

STEP 5: Select $g(g\leq n)$ wavelet packet coefficients for classifiers.

4. EXPERIMENTS AND CONCLUDING REMARKS

Although the method we develop can be applied to many different types of images, we focus our experimental attention on character recognition. We carried out the experiment to compare the recognition rates of PCA and 2DWP-DB(Two-

Dimensional Wavelet Packet based Discriminant Bases). The 6000 hand-written numerals from CENPARMI at Concordia University were used for the experiment. The 4000 digits were used for training and the remaining 2000 digits were used for test, which consist of 10 groups(from 0 to 9 of 200 digits).

In experiments, Linear Discriminant Analysis (LDA) of statistical classifiers and Neural Network model(NN) were used for the calculation of recognition rates.

The illustrations of Figure 6 and Figure 7 are summarized as follows: On the whole, the recognition rate of the 2DWP-DB method is superior to that of PCA. Especially, for the case with the highest recognition rate, the proposed method gives better recognition. In addition, as the number of features increases over 70(in LDA case), the recognition rate decreases. That is , as the number of features increases excessively, the recognition rate of statistical classifiers decreases on the contrary. In other words, for statistical classifiers, dimension problems become very important. Without feature selection, when we use every feature, recognition rates are 89.1%(in LDA case) and 94.8%(in NN case), respectively. Therefore, we believe suggested method is useful in handwritten character recognition, especially when using statistical classifiers.

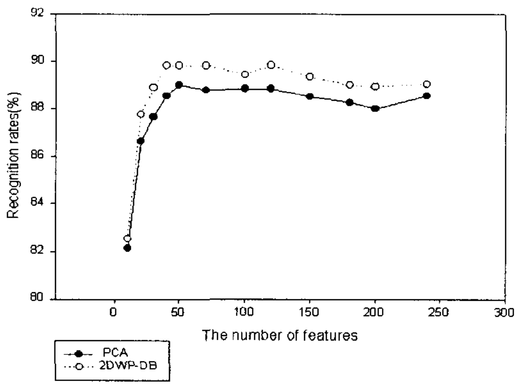

Fig. 6 Comparison of the recognition rates between PCA and 2DWP-DB using LDA

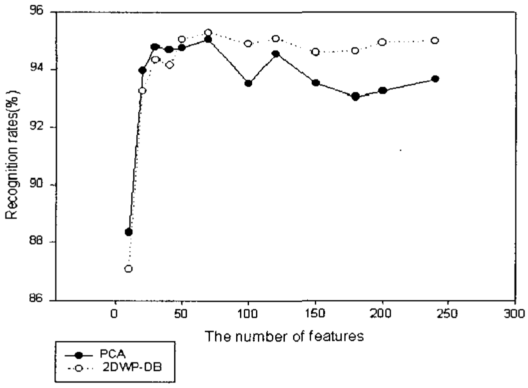

Fig. 7 Comparison of the recognition rates between PCA and 2DWP-DB using NN

REFERENCES

[1] Kim, M. S. and Baek, J. S., "Feature Extraction and Statistical Pattern Recognition for Image Data using Wavelet Decomposition," *The Korean Communications in Statistics*, Vol. 6, No. 3, pp. 831-841, 1999.

[2] Hotelling, H., "Analysis of a complex of statistical variables into principal components," *J. Educ. Psych*, 24, 1933

[3] Karhunen, K., *"Über Linearen Methoden in der Wahrscheinlichkeitsrechnng," Ann. Acad. Sci. Fennicae*, Ser. A 37, no 1., 1947.

[4] Watanabe, S., "Karhunen-Loève expansion and factor analysis: theoretical remarks and applications," *Trans. 4th Prague Conf. Inform. Theory, Statist. Decision Functions, Random Processes (Prague)*, Publishing House of the Czechoslovak Academy of Sciences, pp.645-660, 1967.

[5] Saito, N. and Coifman, R. R., "Local discriminant bases," *Mathematical Imaging : Wavelet Applications in Signal and Image Processing*, 1994.

[6] Learned, R. E. and Wilsky, A. S., "A Wavelet Packet Approach to Transient Signal Classification," *Applied and Computational Harmonic Analysis*, vol 2, pp.265-278., 1995.

[7] Wickerhauser, M. V., *Adaptive Wavelet Analysis: From Theory to Software*. AK peters: Boston, 1994.

[8] Daubechies, I., "Orthonormal bases of compactly supported wavelets," *Communications in Pure and Applied Mathematics*, 41, pp. 909-996, 1988.

[9] Coifman, R. R. and Saito, N., "Constructions of

local orthonormal bases for classification and regression," *in Comptes Rendus Acad. Sci*, Paris, Serie I 319, *no.* 2, *pp.* 191-196, 1994.
[10] Coifman, R. R. and Wickerhauser, M. V., "Entropy -based algorithms best basis selection," *IEEE Trans. Inform.* Theory 38, no.2, pp. 713-719. 1992
[11] Collineau, S., "Some remarks about the scalograms of wavelet transform coefficients," In *Wavelets and Their Applications*, Byrnes, J. S., Byrnes, J. L., Hargreaves, K. A., and Berry, K. Kluwer Academic Publications: Dordrecht, The Netherlands, 1994.
[12] Arino, M. A. and Vidakovic, B., "On wavelet scalograms and their applications in economic time series". *Discussion Paper 95-21, ISDS, Duke University*, Durham, North Carolina, 1995.
[13] Basseville, M., "Distance measures for signal processing and pattern recognition," *Signal Processing*, 18, no.4, pp. 349-369, 1989.
[14] Yamada, H., Yamamoto, K. and Saito, T., "A Linear Normalization for Hand -printed Kanji Character Recognition - Line Density Equalization," *Pattern Recognition*, Vol. 23, No. 9, pp. 1023-1029, 1990.

김 민 수
1994년 전남대학교 전산학과(학사). 1996년 전남대학교 전산통계학과(이학석사). 2000년 전남대학교 전산통계학과(이학박사). 2000년-2002년 전남대학교 BK21사업단 post-doc. 관심분야는 통계적 패턴인식, 다변량 통계분석, wavelets

백 장 선
1981년 연세대학교 응용통계학과(학사). 1984년 연세대학교 응용통계학과(석사). 1991년 미국 Texas A&M 대학교 통계학과(박사). 1991년 ~ 1993년 Southern Methodist University 통계학과 Post-doc.. 1993년 ~ 현재 전남대학교 통계학과 부교수. 관심분야는 비모수 함수추정, 다변량 통계분석

이 귀 상
1980년 서울대학교 전기공학과(학사). 1982년 서울대학교 전자계산기공학과(석사). 1991년 Pennsylvania 주립대 전산학과(박사). 1984년 ~ 현재 전남대학교 전산학과, 정보통신연구소 교수. 관심분야는 멀티미디어통신, 영상압축, VLSI/CAD

김 수 형
1986년 서울대학교 컴퓨터공학과(학사). 1988년 한국과학기술원 전산학과 졸업(공학석사). 1993년 한국과학기술원 전산학과 졸업(공학박사). 1993년 ~ 1996년 삼성전자 멀티미디어연구소 선임연구원. 1997년 ~ 현재 전남대학교 컴퓨터정보학부 부교수. 관심분야는 패턴인식, 문서영상처리, 필적 및 서명감정, WBI.