A Study of Feature Extraction and Selection Using Independent Component Analysis

Seiichi Ozawa

Div. of System Function Science
Graduate School of Science and Technology
Kobe University
Kobe 657-8501, JAPAN
ozawa@eedept.kobe-u.ac.jp

Manabu Kotani

Dept. of Computer and Systems Engineering
Faculty of Engineering
Kobe University
Kobe 657-8501, JAPAN
kotani@cs.kobe-u.ac.jp

Abstract

This paper demonstrates some consideration results on feature extraction and selection for hand-written Japanese Hiragana characters using independent component analysis (ICA). In some ICA algorithms where whitening of input signals is introduced as preprocessing, one can consider that the process of feature extraction using ICA consists of two types of transformations: one is the transformation from an input image to its principal components (PCs), and the other is the transformation from PCs to independent components (ICs). From this fact, two types of feature selection can be applied to outputs of these transformations (i.e. PCs and ICs). Furthermore, as criteria of useful features, cumulative proportion can be adopted in the former type of feature selection, and kurtosis can be adopted in the latter. Thus, we present five different feature selection methods in this paper. To discuss the effectiveness of these methods, recognition experiments using hand-written Japanese Hiragana characters are carried out. As a result, we show that a hybrid method, in which feature selection is carried out for ICs as well as for PCs, has attractive characteristics if small dimensions of feature vectors are preferred in classification.

1 Introduction

Recently, independent component analysis (I-CA) has been widely noticed as a decorrelation technique based on higher-order moment of input signals[1]. Using such characteristics, ICA has been so far applied to problems of blind signal separation such as sound/image separation and EEG signal separation. On the other hand, feature extraction of images and sounds has been also focused as one of prominent applications of ICA[2, 3, 4, 5]. Bartlett & Sejnowski extracted feature vectors from images of human faces with ICA, and showed that these feature vectors had greater viewpoint invariance for human faces as compared with PCA (principal component analysis) ones[6][1]. Since PCA decorrelates only the second order statistics of input signals, this result indicates that higher-order features are useful for capturing invariant features of face patterns as well as conventional second-order features. Such invariant characteristics of ICA feature vectors might be attractive for other pattern recognition problems.

In our previous work, ICA feature vectors are utilized for recognizing hand-written characters in order to study the usefulness of ICA as a feature extraction method[7]. An ICA feature vector is given as a coefficient vector of the bases that is obtained by applying an ICA algorithm to a training set of character images. From the experimental results, we showed that the recognition accuracy greatly depended on input dimensions; in general, small dimensions of inputs tend to attain good classification performance. In this work, input dimensions were changed with the size of subimages, which are separated by imposing small size of windows to a character image. Although such reduction of input dimensions is actually effective, one can say that only local features of a character image are considered in this approach. To achieve higher performance, we should use global features as well as local ones; that is, we should reduce dimensions by selecting useful features extracted from the whole of a character image.

In this paper, we will study feature selection methods in order to obtain high-performance ICA feature vectors through some recognition experiments for hand-written Japanese Hiragana characters. In Section 2, we will briefly refer to an ICA algorithm and its objective function. In Section 3, a method of feature extraction using ICA will be described, then we will propose some criteria for feature selection. In Section 4, recognition experiments will be carried out, and we will discuss what criteria are suitable for feature selection. In Section

[1] For notational convenience, we denote feature vectors obtained by using ICA and PCA as *ICA feature vectors* and *PCA feature vectors*, respectively.

5, we will give conclusions.

2 Independent Component Analysis

Several ICA algorithms have been proposed so far, which are different in objective functions (or contrast functions) for statistical independence and how to derive ICA algorithms[1, 8, 9, 10]. In general, estimated independent components obtained by using these algorithms are different each other. However, it is difficult to discuss which algorithms are most appropriate for feature extraction of characters in the present circumstances. Hence, in the followings, we shall adopt Fast ICA algorithm proposed by Hyvärinen and Oja[11] from its convergence speed.

Suppose that we observe a m-dimensional zero-mean input signal at time t, $\boldsymbol{v}(t) = \{v_1, \cdots, v_m\}'$, where $'$ means the transposition of matrices and vectors. Then the n-dimensional whitening signal, $\boldsymbol{x}(t)$, is given by the following equation:

$$\boldsymbol{x}(t) = \boldsymbol{M}\boldsymbol{v}(t) = \boldsymbol{D}^{-1/2}\boldsymbol{E}'\boldsymbol{v}(t), \qquad (1)$$

where $\boldsymbol{M}$ means a $n \times m$ ($n \le m$) whitening matrix that is given by a matrix of eigenvalues, $\boldsymbol{D}$, and a matrix of eigenvectors, $\boldsymbol{E}$. Here, assume that $\boldsymbol{v}(t)$ is composed of n statistically independent signals, $\boldsymbol{s}(t) = \{s_1(t), \cdots, s_n(t)\}'$. Then, the following linear transformation from $\boldsymbol{x}(t)$ to $\boldsymbol{s}(t)$ exists:

$$\boldsymbol{s}(t) = \boldsymbol{W}\boldsymbol{x}(t). \qquad (2)$$

$\boldsymbol{W}=\{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n\}'$ is often called a separating matrix, and it can be acquired through the training of a two-layer feedforward neural network. This neural network has n outputs denoted as $\tilde{\boldsymbol{s}}(t)$ $=\{\tilde{s}_1(t), \cdots, \tilde{s}_n(t)\}'$ and the ith row vector, $\boldsymbol{w}_i'(i = 1, \cdots, n)$, of $\boldsymbol{W}$ corresponds to a weight vector from inputs to the ith output, $\tilde{s}_i$.

The term 'independent' is used here according to the following definition in statistics:

$$p[s_1(t), \cdots, s_n(t)] = \prod_{i=1}^{n} p_i[s_i(t)], \qquad (3)$$

where $p[\cdot]$ is a probability density function. Since the above probability density function is not preliminary unknown, suitable objective functions should be devised such that neural outputs, $\tilde{s}_i$, are satisfied with Eq.(3) as much as possible, i.e. $\tilde{\boldsymbol{s}}(t) \simeq \boldsymbol{s}(t)$. Karhunen and Oja have proposed the following objective function[10], $J(\cdot)$, to be maximized in terms of output signals, $\tilde{\boldsymbol{s}}$:

$$J(\tilde{\boldsymbol{s}}) = \sum_{i=1}^{n} |E\{\tilde{s}_i^4\} - 3[E\{\tilde{s}_i^2\}]^2|, \qquad (4)$$

where $E\{\cdot\}$ means expectation. As well known, Eq.(4) corresponds to the fourth-order cumulants of $\tilde{s}_i(t)$, called *kurtosis*. Learning algorithms for a separation matrix, $\boldsymbol{W}$, are derived from the gradient of Eq.(4). In the followings, we adopt Fast ICA algorithm proposed by Hyvärinen & Oja in which fixed points of the gradient are obtained on-line (see [11] for details).

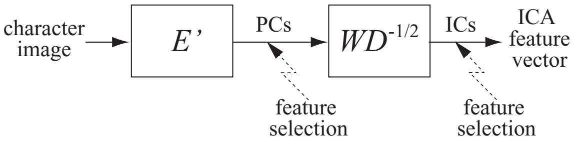


Figure 1: A block diagram of feature extraction of characters using ICA.

3 Extraction and Selection of Features

3.1 Feature Extraction

As described in Section 2, ICA algorithms allow us to decompose input signals into their independent components such that they are satisfied with Eq.(3) as much as possible. Such characteristics of ICA can be applied to feature extraction of handwritten characters.

Based on Eqs. (1) and (2), the relation between inputs and outputs of ICA is given by

$$\tilde{\boldsymbol{s}}(t) = \tilde{\boldsymbol{W}}\boldsymbol{D}^{-1/2}\boldsymbol{E}'\boldsymbol{v}(t) = \boldsymbol{B}\boldsymbol{v}(t), \qquad (5)$$

where $\tilde{\boldsymbol{W}}$ is a separation matrix trained by an ICA algorithm and $\boldsymbol{B} = \tilde{\boldsymbol{W}}\boldsymbol{D}^{-1/2}\boldsymbol{E}'$ is a $n \times m$ matrix. When an ICA input, $\boldsymbol{v}(t)$, corresponds to the tth presentation of character images, the ICA output, $\tilde{\boldsymbol{s}}(t)$, can be considered as its feature vector. Here, the ith row vector, $\boldsymbol{b}_i'$ $(i = 1, \cdots, n)$, of $\boldsymbol{B}$ corresponds to a base vector spanning n-dimensional feature space (such base vectors are called ICA-bases for convenience). Since $\boldsymbol{E}'\boldsymbol{v}(t)$ corresponds to principal components (PCs) of $\boldsymbol{v}(t)$, one can say that an ICA feature vector is given with transformation $\tilde{\boldsymbol{W}}\boldsymbol{D}^{-1/2}$ of PCs. Therefore, we can consider that the process of feature extraction consists of two types of transformations (see Fig. 1). One is the transformation from an input image to its PCs, and the other is the transformation from PCs to independent components (ICs). Different feature selection (dimension reduction) can be applied to outputs of these transformations: that is, we can reduce dimensions of a feature vector after PCs are obtained and/or after ICs are obtained. Finally, an ICA feature vector for the tth character image, $\tilde{\boldsymbol{s}}(t)$, is obtained through these two types of feature selection.

3.2 Feature Selection

3.2.1 Feature Selection Based on Cumulative Proportion

Cumulative proportion has been often used in feature selection as a criterion of determining useful features. For convenience, eigenvalues of a covariance matrix of training samples are denoted in order of their magnitude: $\lambda_1 \geq \cdots \geq \lambda_m$. Then, the cumulative proportion, c_ν, is defined as follows:

$$c_\nu = \frac{\sum_{i=1}^{\nu} \lambda_i}{\sum_{i=1}^{m} \lambda_i}, \tag{6}$$

where ν is the number of large eigenvalues considered here. Let us introduce an upper bound of cumulative proportion, c_0, that gives a threshold value of determining what feature vectors should be adopted, then the largest value of ν can be determined such that $c_\nu \leq c_0$ holds. We select ν eigenvectors with the first ν eigenvalues as PCA-bases; that is, we consider a ν-dimensional subspace spanned by eigenvectors with $\lambda_1 \cdots \lambda_\nu$. After this feature selection is carried out, a ν-dimensional vector of PCs, $\boldsymbol{E}'\boldsymbol{v}(t)$, is obtained, then a ν-dimensional ICA feature vector, $\tilde{\boldsymbol{s}}(t)$, is calculated from Eq. (5). For convenience, we denote such a feature selection method as $\mathrm{FS_{cp}}$ in the following section.

3.2.2 Feature Selection Based on Kurtosis

When feature selection is carried out for ICs, a proper choice for the criteria is to adopt kurtosis that is a measure for signal independence and is often used in learning of ICA-bases. Since we merely obtain ICA-bases such that their coefficients (i.e. ICs) are satisfied with the independent condition of Eq. (3) as much as possible, kurtosis of ICs should have variety in their magnitude. Considering that an absolute value of kurtosis indicates the degree of independence, ICA features can be ranked in order of their degrees of independence. Here, let us assume that strongly independent features are useful for classification and we want ν-dimensional ICA feature vectors, $\tilde{\boldsymbol{s}}(t)$; then we should select ICA features, $\tilde{s}_i(t)$, with ν large absolute values of kurtosis. Since kurtosis becomes zero when the distribution of ICs is Gaussian, one can say that ICs with non-Gaussian distribution are preferred to be selected as ICA features.

To discuss the usefulness of kurtosis as a criterion for feature selection, we evaluate kurtosis for all ICs when 46 Japanese Hiragana hand-written characters of 50 people in ETL-4 database are adopted as training samples for ICA-bases. The character images with 76×72 binary values are preprocessed through centralization and size normalization before they are converted to 16×16 16-bit gray-scale images (i.e. dimensions of inputs, $\boldsymbol{v}(t)$, are 256). Fast ICA algorithm is applied to these preprocessed character images, then ICA-bases are trained.

There are at least two approaches to the kurtosis evaluation. As we might easily expect, one approach is that kurtosis for each IC is evaluated using all of the character samples independent of categories. Another approach is that kurtosis for each IC is evaluated using only character samples in a single class. In this approach, ICs are ranked in their independence for each category; hence, feature selection for every category is carried out separately. In both approaches, ICA features with larger absolute values of kurtosis are preferred to be selected. The former and latter methods of feature selection based on such kurtosis evaluation are respectively denoted as $\mathrm{FS_{kurt}^{all}}$ and $\mathrm{FS_{kurt}^{class}}$ for convenience. Figure 2 shows the distribution of kurtosis when the hand-written characters of ETL-4 are adopted to evaluate kurtosis. In Fig. 2, we can see that kurtosis of some ICs are large and the rest of them are rather small; hence features with large absolute values of kurtosis are easily selected. Figures 3(a)-(d) respectively exemplify kurtosis distributions for four classes of Hiragana characters: (a) 'A', (b) 'I', (c) 'U', and (d) 'E'. These distributions are evaluated for 50 samples per class that are the same samples as we used in the previous evaluation. As you can seen in Figs. 3(a)-(d), the differences in absolute values of kurtosis become distinctive as compared with the previous result in Fig.2. Therefore, one can say that feature selection can be carried out easier by using $\mathrm{FS_{kurt}^{class}}$. However, in this approach, we should notice that we define different classifiers for every class depending on which features should be selected.

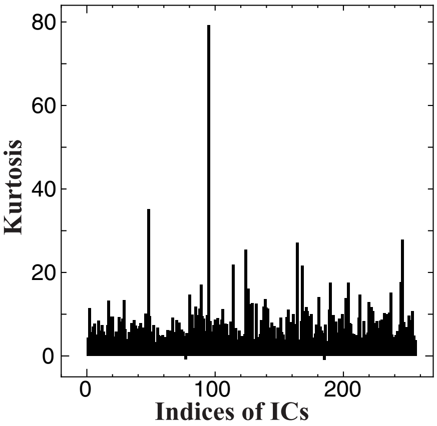


Figure 2: Kurtosis distribution for 256 ICs that are calculated from 46 Japanese Hiragana hand-written characters of 50 people.

4 Recognition Experiments

To evaluate the effectiveness of feature selection methods presented in the previous section, recognition experiments are carried out for 100 datasets of 46 hand-written Japanese Hiragana characters stored in the ETL-4 database. The original character images with 76×72 pixels are converted into 16×16 16-bit gray-scale images through the same processing as we adopted in 3.2.2. The 100 sets of Hiragana characters are divided into halves: 50 sets are adopted as training patterns and the

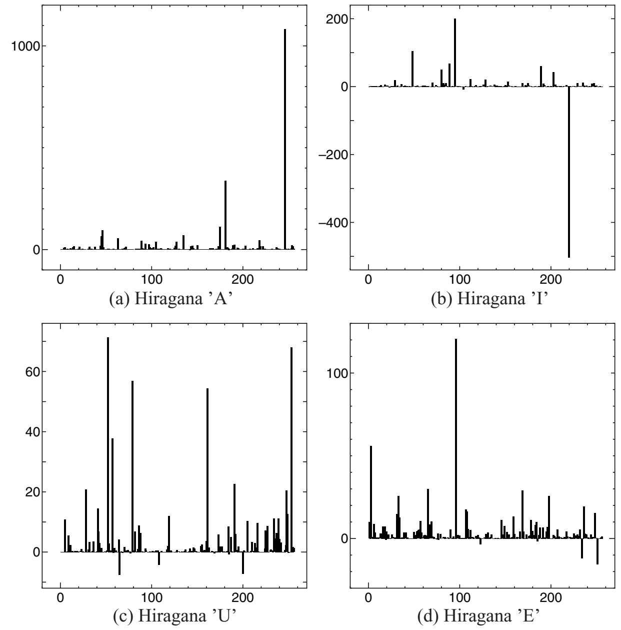


Figure 3: Kurtosis distributions for 256 ICs when hand-written characters in four different categories ((a) 'A', (b) 'I', (c) 'U', and (d) 'E') are adopted to evaluate kurtosis. For each class, 50 samples collected from 50 different people (ETL-4 database) are used for evaluation. Horizontal and vertical axes correspond to indices of ICs and their values of kurtosis, respectively.

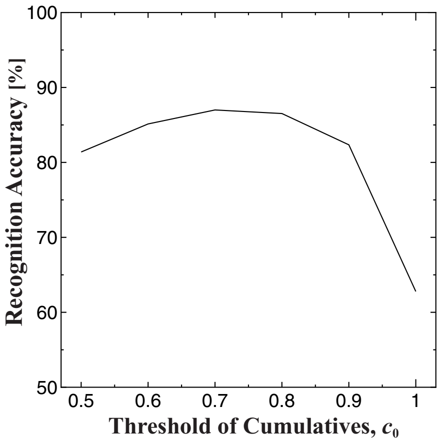


Figure 4: Recognition accuracy for ICA features whose dimensions are reduced with feature selection based on cumulative proportion: $\mathrm{FS}_{\mathrm{pc}}$. Threshold values of cumulative proportion, c_0, are varied from 0.5 to 1.0. Classification is carried out using k-NN classifier (k=5).

Table 1: The dimensions, ν, of ICA feature vectors obtained with $\mathrm{FS}_{\mathrm{pc}}$ when threshold values of cumulative proportion, c_0 are varied from 0.5 to 1.0.

c_0	0.5	0.6	0.7	0.8	0.9	1.0
ν	12	17	26	39	69	256

other 50 sets are adopted as test patterns. Namely, the total numbers of training and test patterns are respectively 2300. Training patterns are used for generating reference vectors as well as learning ICA-bases. In classification, we adopt the following similarity as a measure of pattern matching:

$$S(c,i) = \frac{\tilde{\boldsymbol{s}}'\boldsymbol{p}_i^c}{|\tilde{\boldsymbol{s}}||\boldsymbol{p}_i^c|}, \qquad (7)$$

where $\tilde{\boldsymbol{s}}$ is an ICA feature vector extracted from a test pattern and $\boldsymbol{p}_i^c$ is the ith reference vector of class c. After calculating similarities of $S(c,i)$ for all c and i, classification is carried out based on the k-nearest neighbor (k-NN) rule. The value of k is set to 5 in the following experiments.

Figure 4 demonstrates the recognition accuracy when only $\mathrm{FS}_{\mathrm{cp}}$ is applied in feature selection, where vertical and horizontal axes correspond to recognition accuracy and threshold values of cumulative proportion, c_0, respectively. c_0 is varied from 0.5 to 1.0 to examine the optimal number of ICA features. The dimensions, ν, of ICA feature vectors that are obtained with $\mathrm{FS}_{\mathrm{cp}}$ are shown in Table 1. As you can see in Fig. 4, feature selection based on cumulative proportion is useful for selecting ICA features. In this experiment, the optimal value of c_0 was 0.7.

As described in 3.2.2, we can adopt the other two types of feature selection that are different in what samples are utilized for kurtosis evaluation: $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ and $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$. In $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$, kurtosis is evaluated for all training samples independent of categories; on the other hand, in $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$ the kurtosis evaluation is separately carried out for each class. In both approaches, ICA features with larger absolute values of kurtosis are preferred to be selected. Figure 5 shows the recognition performances of $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ and $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$ where horizontal axis corresponds to the number of selected ICA features. From Fig. 5, although there are not so much difference in performances for both selection methods, one can say that $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$ outperforms $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ when the number of selected features is small. However, since the performance monotonously deteriorates with a decrease in features' dimensions in both methods, we conclude that these selection methods are not so much effective in recognition.

We can also consider a hybrid feature selection based on cumulative proportion and kurtosis evaluation: that is, feature selection based on kurtosis evaluation is carried out after features are selected based on cumulative proportion. For convenience, a hybrid method combining $\mathrm{FS}_{\mathrm{cp}}$ and $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ is simply denoted as $\mathrm{FS}_{\mathrm{cp}}+\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$; similarly a hybrid method combining $\mathrm{FS}_{\mathrm{cp}}$ and $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$ is denoted as $\mathrm{FS}_{\mathrm{cp}}+\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$. Figures 6(a)(b) show the recognition performances of $\mathrm{FS}_{\mathrm{cp}}+\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ and

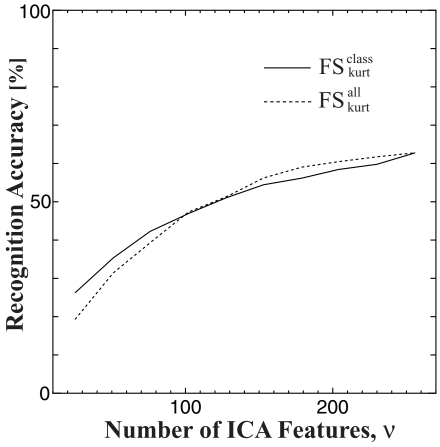
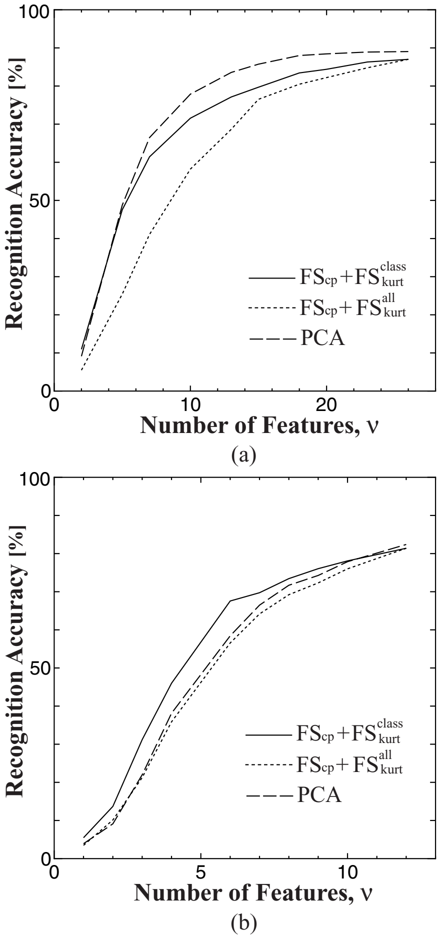


Figure 5: Recognition accuracy for ICA feature vector whose dimensions, are reduced through two types of feature selection: $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ and $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$. Classification is carried out using k-NN classifier (k=5). Horizontal axis means the number of selected ICA features, ν.

$\mathrm{FS}_{\mathrm{cp}}$+$\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$ when threshold values, c_0, are set to 0.7 and 0.5, respectively. For comparative purposes, the recognition performances for PCA features are also depicted in Figs. 6(a)(b). The selection for PCA features are carried out based on cumulative proportion. When $c_0 = 0.7$ (Fig. 6(a)), the performances for ICA feature vectors selected by both hybrid methods are inferior to the performance for PCA feature vectors if the number of selected features is not too small. However, as you can see in Fig. 6(b), ICA features selected by $\mathrm{FS}_{\mathrm{cp}}$+$\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$ realize higher performance as compared with PCA features. This result means that useful information for classification could be condensed in small number of ICA features if adequate threshold value, c_0, is set in $\mathrm{FS}_{\mathrm{cp}}$. Therefore, one can say that ICA features should be adopted in case that small dimensions of feature vectors are preferred; for example, in cases that high-speed processing and/or small size of memories are required.

5 Summary

In this paper, we study some methods of feature extraction and selection for hand-written Japanese Hiragana characters using independent component analysis (ICA). We show that the process of feature extraction consists of two types of transformations: one is the transformation from an input image to its principal components (PCs), and the other is the transformation from PCs to independent components (ICs). Hence, two types of feature selection (dimension reduction) can be applied to outputs of these transformations: that is, we can reduce dimensions of a feature vector after PCs are obtained and/or after ICs are obtained. As criteria for selecting features, cumulative proportion can be adopted in the former type of feature selection (denoted as $\mathrm{FS}_{\mathrm{pc}}$), and kurtosis can be adopted in the latter. For the latter, we present two approaches to the kurtosis evaluation: one is that kurtosis for each IC is evaluated using all of the character samples independent of categories, and another approach is that kurtosis is evaluated using only character samples in a single class. Feature selection methods based on these kurtosis evaluations are denoted as $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ and $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$, respectively. Furthermore, two hybrid approaches to feature selection are also considered: the first approach is that $\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ is carried out after $\mathrm{FS}_{\mathrm{pc}}$ (denoted as $\mathrm{FS}_{\mathrm{pc}}$+$\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$), the second approach is that $\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$ is done after $\mathrm{FS}_{\mathrm{pc}}$ (denoted as $\mathrm{FS}_{\mathrm{pc}}$+$\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$).



(a)

(b)

Figure 6: Recognition results when k-NN classifier (k=5) is applied to ICA features whose dimensions are reduced by $\mathrm{FS}_{\mathrm{cp}}$+$\mathrm{FS}^{\mathrm{all}}_{\mathrm{kurt}}$ and $\mathrm{FS}_{\mathrm{cp}}$+$\mathrm{FS}^{\mathrm{class}}_{\mathrm{kurt}}$. Threshold values of cumulative proportion, c_0, are set to (a)0.7 and (b)0.5. For comparative purposes, the recognition performances for PCA features (denoted as 'PCA') are also displayed. Horizontal axis means the number of selected ICA (or PCA) features, ν.

Recognition experiments using hand-written Japanese Hiragana characters are carried out in order to evaluate the above five approaches. As a result, although high-performance feature vectors are generated with $\mathrm{FS}_{\mathrm{pc}}$, both $\mathrm{FS}_{\mathrm{kurt}}^{\mathrm{all}}$ and $\mathrm{FS}_{\mathrm{kurt}}^{\mathrm{class}}$ are not so much effective in the selection of ICA features. Two hybrid methods are worth adopting as feature selection methods; especially, $\mathrm{FS}_{\mathrm{pc}}+\mathrm{FS}_{\mathrm{kurt}}^{\mathrm{class}}$ is considerably effective if small dimensions of feature vectors are preferred.

Acknowledgments

The authors would like to express sincere thanks to Professor Shigeo Abe of Graduate School of Science and Technology, Kobe University for encouragement and instructive comments. This work was partially supported by the Grant-in-Aid for Encouragement of Young Scientists of the Ministry of Education, Science, Sports and Culture of Japan under Grant 10780233 (1998-1999).

References

[1] A. Hyvärinen: “Survey on independent component analysis”, *Neural Computing Surveys*, **2**, 94-128, 1999.

[2] A. J. Bell and T. J. Sejnowski: “Edges are the ’independent components’ of natural scenes”, *Advances in Neural Information Processing Systems*, **9**, MIT Press, Cambridge, MA, 831-837, 1997.

[3] J. Karhunen, A. Hyvärinen, R. Vigario, J. Hurri, and E. Oja: “Applications of neural blind separation to signal and image processing”, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 131-134, 1997.

[4] M. Kotani, Y. Shirata, S. Maekawa, S. Ozawa, and K. Akazawa, “Application of independent component analysis to feature extraction of speech”, *Proc. of Int. Joint Conf. on Neural Networks* (IJCNN99-Washington DC), CD-ROM #704, 1999.

[5] M. Kotani, S. Maekawa, S. Ozawa, and K. Akazawa: ”Signal processing of speech using independent component analysis based on information maximization algorithm”, *Transaction of The Society of Instrument and Control Engineers* (in Japanese), **36**, 5, 456-458, 2000.

[6] M. S. Bartlett and T. J. Sejnowski: “Viewpoint invariant face recognition using independent component analysis and attractor networks”, *Advances in Neural Information Processing Systems*, **9**, MIT Press, Cambridge, MA, 817-823, 1997.

[7] S. Ozawa, T. Tsujimoto, M. Kotani, and N. Baba: “Application of independent component analysis to hand-written Japanese character recognition”, *Proc. of International Joint Conf. on Neural Networks* (IJCNN99-Washington DC), CD-ROM #462, 1999.

[8] A. J. Bell and T. J. Sejnowski: “An information maximization approach to blind separation and blind deconvolution”, *Neural Computation*, **7**, 1129-1159, 1995.

[9] S. Amari, A. Chichocki, and H. Yang: “A new learning algorithm for blind signal separation”, *Advances in Neural Information Processing Systems*, **8**, MIT Press, Cambridge, MA, 757-763, 1996.

[10] J. Karhunen and E. Oja: “A class of neural networks for independent component analysis”, *IEEE Trans. on Neural Networks*, **8**, 3, 486-503, 1997.

[11] A. Hyvärinen and E. Oja: “A fast fixed-point algorithm for independent component analysis”, *Neural Computation*, **9**, 1483-1492, 1997.