Available online at www.sciencedirect.com
SCIENCE DIRECT®

Pattern Recognition 39 (2006) 81–88

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Ensemble learning for independent component analysis

Jian Cheng[a,b], Qingshan Liu[a], Hanqing Lu[a,*], Yen-Wei Chen[c]

[a] *National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, P.O. Box 2728, Beijing 100080, China*
[b] *Nokia Research Center, Nokia House 1, He Ping Li Dong Jie 11, Beijing 100013, China*
[c] *College of Information Science and Engineering, Ritsumeikan University, 525-8577, Japan*

Received 14 June 2004; received in revised form 14 June 2005; accepted 27 June 2005

Abstract

It is well known that the applicability of independent component analysis (ICA) to high-dimensional pattern recognition tasks such as face recognition often suffers from two problems. One is the small sample size problem. The other is the choice of basis functions (or independent components). Both problems make ICA classifier unstable and biased. In this paper, we propose an enhanced ICA algorithm by ensemble learning approach, named as random independent subspace (RIS), to deal with the two problems. Firstly, we use the random resampling technique to generate some low dimensional feature subspaces, and one classifier is constructed in each feature subspace. Then these classifiers are combined into an ensemble classifier using a final decision rule. Extensive experimentations performed on the FERET database suggest that the proposed method can improve the performance of ICA classifier.

Keywords: Independent component analysis; Ensemble learning; Random independent subspace; Face recognition; Majority voting

1. Introduction

In recent years, pattern recognition has attracted much attention because of its potential application, such as military, biometrics, human–computer interface, information security, etc. However, the data often has high dimensionality and contains much redundancy, which makes the statistical estimation very difficult and the computational complexity is large. How to reduce the redundancy and extract features are key issues in pattern recognition. To address these issues, a number of algorithms have been developed to reduce dimensionality [1–4]. Sirovich and Kirby [1] applied principal component analysis (PCA) to reduce the dimensionality of samples. Further, Turk and Pentland [2] developed it to a well-known face recognition method, known as Eigenfaces. PCA performs dimensionality reduction by projecting the original samples into a lower dimensional linear subspace spanned by the leading eigenvectors of the training data's covariance matrix. However, PCA can only encode the second-order statistical dependencies between pixels. There still exist much higher-order statistical dependencies among three or more pixels. As an extension of PCA, independent component analysis (ICA) [4–6] is capable of finding a set of linear basis vectors to make the higher-order statistical independence besides the second-order statistical independence in PCA.

Due to its generality, ICA has been applied in many fields, such as signal processing [7,8], medical image analysis [9], face representation [10,11], etc. However, ICA method often encounters two challenging problems in practical applications. One is the small sample size problem, i.e. there are only a small number of training samples available in practice. The training sample size is too small compared to the dimensionality of feature space. In case of small sample size, the ICA classifier constructed on the training sets is biased and has a large variance, which result in the unstable performance of ICA classifier. The other is the choice of basis functions (or independent components) problem. The most common approach is to throw away the eigenvectors

* Corresponding author. Tel.: +86 10 8261 4465; fax: +86 10 6255 1993.
E-mail addresses: jcheng@nlpr.ia.ac.cn (J. Cheng), qsliu@nlpr.ia.ac.cn (Q. Liu), luhq@nlpr.ia.ac.cn (H. Lu), chen@is.ritsumei.ac.jp (Y.-W. Chen).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.06.018

corresponding to little eigenvalues in pre-processing stage. However, the choice criterion of this approach satisfies the least reconstruction error in PCA, but is not optimal for ICA.

Ensemble learning techniques have become extremely popular over the last several years because they combine multiple classifiers into an ensemble classifier, which is often demonstrated to has significantly better performance than its base classifiers [12,13]. Such as boosting [14], bagging [15], and the Random Subspace method [16], etc. Boosting is a method to combine a sequence of classifiers by a final decision rule, in a weighted version of the training samples which are updated dynamically according to the errors in previous classification. The falsely classified samples get larger weights in next classifier. In bagging, a number of random independent bootstrap replicates of training samples are generated firstly, and then one classifier is constructed on each of them. The final prediction is decided by aggregating the decisions of all classifiers. The random subspace method is to combine a set of classifiers that base on resampling with replacement in the feature space of training data.

In this paper, we apply the random subspace method to overcome the small sample size problem and the choice problem of basis functions for ICA classifier. By resampling the features with replacement, a set of low dimensional feature subspaces is generated. Then one ICA classifier is constructed in each feature subspace. For a test sample, each ICA classifier gives a prediction. The final predictions are decided by aggregating all predictions using a final decision rule. We name the proposed method random independent subspace (RIS). There are two resampling ways in our method. The first resampling is done in the original feature space. The second resampling is done in the whitened feature space. To verify, respectively, the effect of the two resampling, RIS is divided into three schemes. Scheme I only resamples in the original feature space of the training set. The resampling in the original feature space is equivalent to reduce the dimensionality of feature space and increase the virtual samples. Scheme II only resamples in the whitened feature space whose bases are eigenvectors of covariance matrix of the training set. In scheme II, we can avoid the manual choice of basis functions (or independent components), which is still an unsatisfactorily resolved issue. Integrating the merits of schemes I and II, scheme III adopts a two-level cascade resampling structure which resamples the original feature space of the training set and the whitened feature space at the same time.

2. Independent component analysis

Independent component analysis (ICA) [4–6] technique is capable of finding a set of linear basis vectors to make the higher-order statistical independence besides the second-order statistical independence in PCA. Initially, ICA is intimately related to the blind source separation (BSS) problem in early works [7,8]. Its goal is to separate the observed signals into a linear combination of latent independent signals. Later, this technique is related to the principle of redundancy reduction suggested by Barlow [17] as a coding strategy. Because no analytic solution can be obtained, a number of learning algorithms have been developed to learn approximate solutions for ICA. Bell and Sejnowski proposed a simple learning algorithm, InfoMax [5,18], which maximize the mutual information between the inputs and outputs of a neural network. InfoMax is an unsupervised learning rule that contains two stages, whitening stage and rotation stage.

Assume a set of training set $X = [x_1, x_2, \dots, x_n]$, where each column vector x_i represents a N-dimensional sample and the number of training samples is n. The general model of ICA can be described as follows:

$$X = A * S, \tag{1}$$

where $S = [s_1, s_2, \dots, s_n]$ is the coefficient. A is a square mixing matrix and its column vectors are basis functions. The independent component analysis is to find a separating matrix W_I, so that $U_I = W_I * X$ approximates the independent component S, possibly permuted and rescaled.

At the first stage, the mean is subtracted from the training set, and the first- and the second-order statistical correlations are removed. This stage is known as whitening or sphering. Principal Component Analysis gives an orthogonal solution to this task.

$$X * X^{\mathrm{T}} * E = E * \Lambda, \tag{2}$$

where E is the eigenvector matrix and Λ is the eigenvalue matrix. When whitening operator $W_P = \Lambda^{-1/2} * E^{\mathrm{T}}$, the transformed data $\tilde{X} = W_P * X$ are decorrelated, i.e. $\tilde{X} * \tilde{X}^{\mathrm{T}} = I$, where I is an identity matrix.

At the second stage, the whitened data are rotated to make components as independent as possible. InfoMax uses the following iterative learning procedure to realize the rotation:

$$U = W * \tilde{X},$$

$$\Delta W = [I + H * U^{\mathrm{T}}] * W,$$

$$\hat{W} = W + \rho * \Delta W \to W, \tag{3}$$

where H is a matrix whose entry $H_{ij} = 1 - 2/(1 + \mathrm{e}^{-U_{ij}})$, and ρ is learning rate.

The full transform matrix W_I is the product of the whitening matrix and the rotation matrix learned by (3), $W_I = W * W_P$.

3. Random independent subspace

Redundancy reduction and feature selection are key issues in pattern recognition. To reduce the dimensionality of feature space and computational complexity, many techniques have been developed, such as PCA (eigenfaces), LDA, etc. As an extension of PCA, ICA has been demonstrated to be an effective method [9–11,18]. However, there are often only a small number of training samples are available for ICA in practice, i.e. the number of training samples is far less than the dimensionality, which make the covariance matrix singular in ICA. It induces ICA classifier to be unstable. On the other hand, the choice of basis functions is still an open problem. The most common approach is to throw away the eigenvectors corresponding to the little eigenvalues in whitening stage. However, the choice criterion of this approach satisfies the least reconstruction error in PCA, but is not optimal for ICA.

In order to improve the performance and overcome the shortcoming of ICA classifier, we proposed an enhanced ICA method adopting randomly resampling strategy, named as random independent subspace (RIS). Let $X = [x_1, x_2, \ldots, x_n]$ be the training set matrix, where n is the number of training samples. Each column $x_i = (x_{i1}, x_{i2}, \ldots, x_{iN})^{\mathrm{T}} \in R^N$ is an N-dimensional feature representation for a training sample. Usually, N is very large. We randomly resample K r-dimensional feature subspaces from the N-dimensional feature space, where $r < N$. These sampling is repeated with replacement. Therefore the new training set are $X^k = [x_1^k, x_2^k, \ldots, x_n^k]$, $k = 1, 2, \ldots, K$, where $x_i^k = \left(x_{i1}^k, x_{i2}^k, \ldots, x_{ir}^k\right)$. Then, one and only one ICA classifier C^k is constructed using the r-dimensional sampled training set X^k. For a new test y, it is projected into each subspace and given a prediction $C^k(y)$ by each ICA classifier C^k. The final prediction is decided by aggregating all predictions using a final decision rule, $C(y) = aggre\left\{C^1(y), C^2(y), \ldots, C^K(y)\right\}$.

We can also resample in the whitened feature space as well as the original feature space. To verify the effect of the two resampling respectively, we develop three different schemes for RIS. The first one is to only resample in the original feature space, which is named RIS1. The second one is to only resample in the whitened feature space in whitening stage, which is named RIS2. The third one is a two-level cascade resampling structure that resamples both the original feature space and the whitened feature space, which is named RIS3. The whole flow graph is depicted in Fig. 1. The pseudocodes of algorithms are shown in Figs. 2–4.

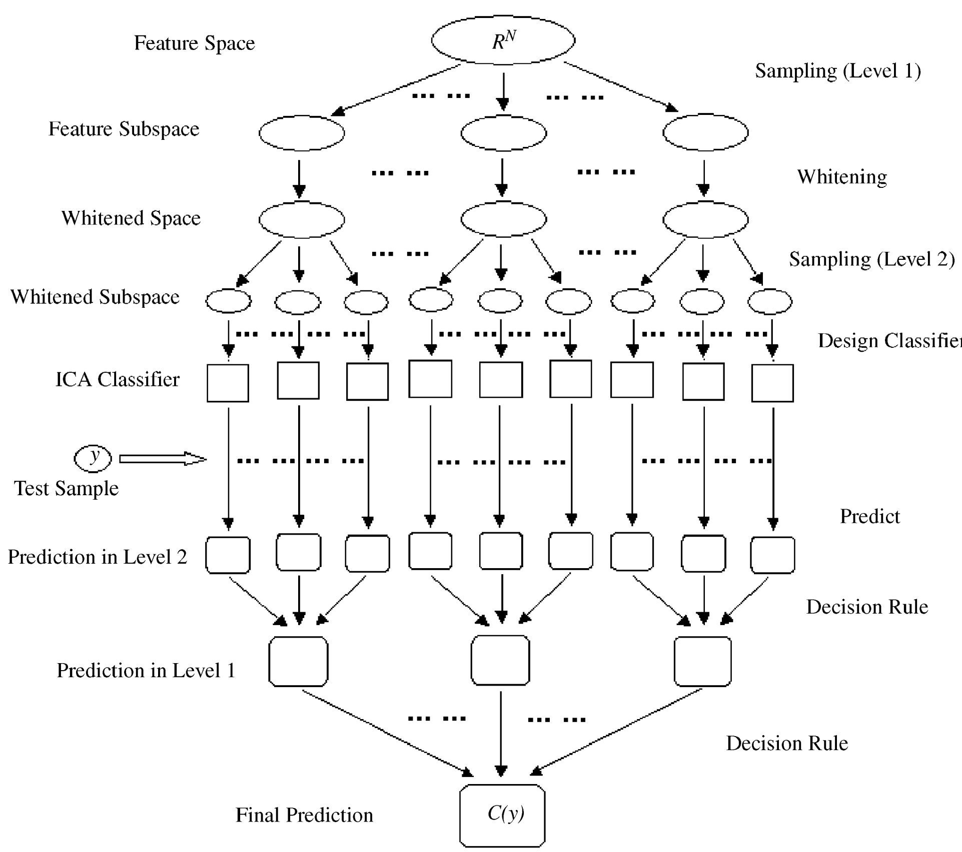

Fig. 1. Flow graph of RIS3.

RIS1: Step 1: Repeat for $k = 1, 2, \cdots, K$

(a) Select randomly r features from the original N-dimensional feature space. The new training set $X^k = [x_1^k, x_2^k, \cdots, x_n^k]$, where $x_i^k = (x_{i1}^k, x_{i2}^k, \cdots, x_{ir}^k)$.

(b) Perform whitening on the new training set X^k, then construct an ICA classifier C^k on the whitened data $\tilde{X}^k$.

Step 2: Combine classifiers C^k, $k = 1, 2, \cdots, K$, into an ensemble classifier C using the final decision rule.

Fig. 2. Random independent subspace scheme I.

RIS2: Step 1: Perform whitening on the original training set X, and denote the whitened training set $\tilde{X}$.

Step 2: Repeat for $k = 1, 2, \cdots, K$

(a) Select randomly r features from the whitened feature space $\tilde{X}$. The new training set $\tilde{X}^k = [\tilde{x}_1^k, \tilde{x}_2^k, \cdots, \tilde{x}_n^k]$, where $\tilde{x}_i^k = (\tilde{x}_{i1}^k, \tilde{x}_{i2}^k, \cdots, \tilde{x}_{ir}^k)$.

(b) Construct an ICA classifier C^k on $\tilde{X}^k$.

Step 3: Combine classifiers C^k, $k = 1, 2, \cdots, K$, into an ensemble classifier C using the final decision rule.

Fig. 3. Random independent subspace scheme II.

RIS3: Step 1: Repeat for $k = 1, 2, \cdots, K$

(a) Select randomly r features from the original N-dimensional feature space. The new training set $X^k = [x_1^k, x_2^k, \cdots, x_n^k]$, where $x_i^k = (x_{i1}^k, x_{i2}^k, \cdots, x_{ir}^k)$.

(b) Perform whitening on the new training set X^k, and denote whitened training set $\tilde{X}^k$.

(c) Repeat for $h = 1, 2, \cdots, H$

(1) Select randomly s features from the whitened feature subspace $\tilde{X}^k$. The new training set $\tilde{X}^{kh} = [\tilde{x}_1^{kh}, \tilde{x}_2^{kh}, \cdots, \tilde{x}_n^{kh}]$, where $\tilde{x}_i^{kh} = (\tilde{x}_{i1}^{kh}, \tilde{x}_{i2}^{kh}, \cdots, \tilde{x}_{is}^{kh})$.

(2) Construct an ICA classifier C^{kh} on $\tilde{X}^{kh}$.

(d) Combine classifiers C^{kh}, $h = 1, 2, \cdots, H$, into an ensemble classifier C^k using the final decision rule.

Step 2: Combine classifiers C^k, $k = 1, 2, \cdots, K$, into an ensemble classifier C using the final decision rule.

Fig. 4. Random independent subspace scheme III.

In this paper, we use majority voting as the final decision rule, which simply assigns the test sample with the class label that appears most frequently in $C^k(y)$, $k=1,2,\ldots,K$.

4. Experiments

Face recognition is a typical pattern recognition problem. Without loss of generality, we use face recognition to assess the feasibility of RIS. Experiments are performed on FERET face database [19]. We use 1002 front view images as training set. The gallery FA and probe FB contain 1195 objects. There is one and only one image of individual in FA and FB. These images are acquired under variable illumination, facial expression, and time (duplicate images). All the images have been reduced to 48×54 by eye location. The coordinates of two eyes are set as (12,14) and (36,14). Histogram equalization is performed as pre-processing on all images. A few original and preprocessed samples are shown in Fig. 5.

For the subsequent analysis, each image is represented using a 2592-dimensional vector given by the luminance value at each pixel location. We organize the training set matrix X so that the images are in columns and the pixels are in rows, i.e., X has 2592 rows and 1002 columns. Gallery FA and probe FB are also organized as the training set.

Same as ICA, RIS is an unsupervised method. For comparability, we compare RIS with two popular unsupervised methods besides ICA, i.e., PCA (eigenfaces) and Kernel PCA on face recognition. Cosine distance has been verified to be superior to other distance metric in [20], so all experiments are implemented with cosine distance and nearest neighbor classifier:

$$d(y_i, y_j) = 1 - \frac{y_i^{\mathrm{T}} \cdot y_j}{\|y_i\| \cdot \|y_j\|}. \tag{4}$$

If the nearest neighbor from the gallery is of the same object as the probe, then the trial is a success. Otherwise it is a failure.

4.1. Random Independent Subspace scheme I

As described in Section 3, we adopt resampling strategy only in the original feature space whose dimensionality is $48 \times 54 = 2592$. The number of training samples is 1002. We select randomly 10 subspaces with replacement, each of which contains 1296 (half of 2592) features from the original feature space. In subspace, the new training set X^k, $k=1,2,\ldots,10$, has 1296 rows and 1002 columns. One ICA classifier is constructed on each subspace. For a new test sample, it is projected into each subspace and given a prediction by each ICA classifier. Then 10 classifiers are combined into an ensemble classifier using majority voting. The RIS1 increases virtual samples and reduces the dimensionality of the original feature space through resampling.

Fig. 5. Upper row is the original images and bottom row is pre-processed images.

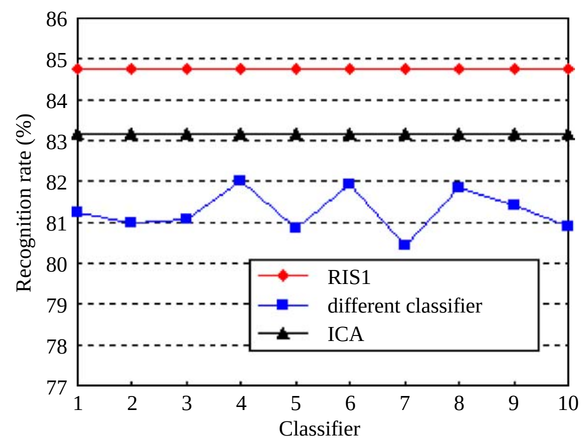

Fig. 6. The accuracy of 10 different classifiers, ICA and RIS1.

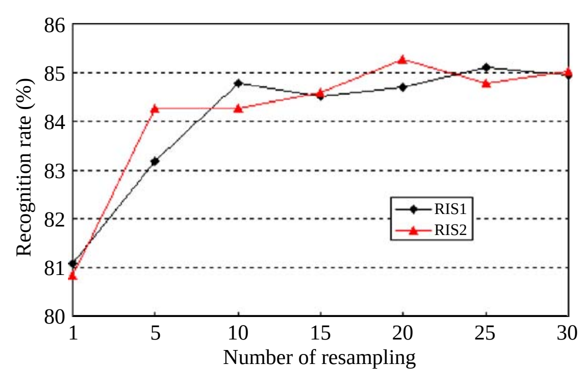

Fig. 7. Accuracy varying with the number of resampling.

The best accuracies of 10 different classifiers are illustrated in Fig. 6 . The results show that the accuracy of individual classifier is no more than 82%, but the ensemble classifier (RIS1) combining 10 classifiers can increase near 3%. Fig. 7 shows that when the number of resampling increases from 1 through 10, the recognition rate increases quickly from 81.07% to 84.77%, then there are no obvious improvements when the number of resampling increased. It suggests that ensemble of 10 classifiers is appropriate and efficient. Fig. 9 reports the performance of RIS1 in detail. RIS1 achieves 84.77% recognition accuracy while that of ICA is 82.76%. These experimental results show that resam-

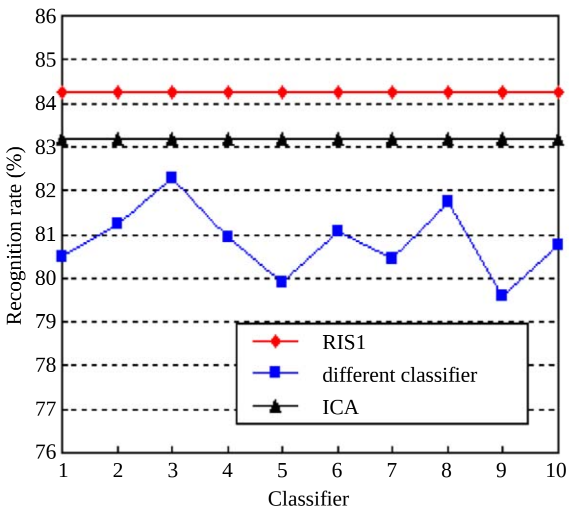

Fig. 8. The accuracy of 10 different classifiers, ICA and RIS2.

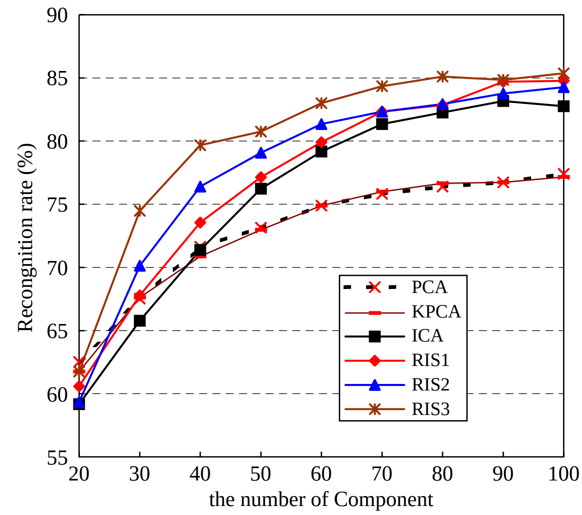

Fig. 9. Comparison RIS with ICA, PCA, and Kernel PCA.

pling in the original feature space can improve the performance of ICA classifier.

4.2. Random Independent Subspace scheme II

In scheme 2, PCA technique is used firstly to whiten the original data, and then we resample the whitened feature space whose bases are the eigenvectors. Because zero and little eigenvalues usually correspond to noise, we only retain the eigenvectors corresponding to the 150 leading eigenvalues as bases. In order to improve the accuracy of each classifier, we fix these bases corresponding to the 20 largest eigenvalues, and the others are sampled randomly from the residual 130 bases. Then ICA classifier is constructed on the subspace. Compared with conventional choice strategy in standard ICA, the scheme 2 avoids loss of some important bases corresponding little eigenvalues.

Fig. 8 shows that the performances of 10 individual classifiers are unstable. The worst recognition rate is 79.58% while the best recognition rate is 82.25%. The ensemble classifier (RIS2) can get 84.27% combining the 10 classifiers using majority vote. Fig. 7 shows that the number of resampling has significant effect on the performance. When the number of resampling increases from 1 through 5, the recognition rate increases from 80.82% to 84.27%. However, there are not obvious improvement when continue to increase the number of resampling. In Fig. 9, comparison between RIS2 and ICA suggests that resampling in the whitened feature space can also improve the performance of classifier.

4.3. Random Independent Subspace scheme III

In scheme I, resampling the original feature space alleviates the negative effect brought by the small sample size problem, but the choice problem of basis functions is still unresolved. On the contrary, scheme II only resolves the choice problem of basis functions while the small sample size problem is unresolved. To addresses the two problems at the same time, scheme III adopts a two-level cascade resampling structure. The first resampling is in the original feature space (Level 1). The second resampling is in the whitened feature space (Level 2). All classifiers in Level 2 give voting to their parent classifiers in Level 1, and then all classifiers in Level 1 vote to the final prediction.

We resample 10 times in the original feature space and 5 times in the orthogonal whitened feature space, respectively. As shown in Table 1, by aggregating 5 classifiers in Level 2 (in whitened feature space), the classifiers in Level 1 are obvious better than their child classifiers in Level 2. For example, the first classifier in Level 1 get 80.42%, which is combined from 5 child classifiers (in Level 2) whose accuracies are 75.56%, 76.73%, 76.65%, 77.82% and 78.57%. Further, the 10 classifiers in Level 1 are aggregated into an ensemble classifier that obtains a best accuracy rate 85.36%. These experimental results suggest that the two-level cascade resampling structure can get a stronger classifier. As in RIS1 and RIS2, we also consider the effect of the number of resampling. Experiments show that there is no significant improvement with more resampling.

A comprehensive comparison is reported in Fig. 9. The horizontal axis indicates the number of features used by classifiers, and the vertical axis indicates recognition rate. ICA, PCA and Kernel PCA, are compared with three schemes of RIS with 20–100 features. Here, Kernel PCA adopts Gaussian kernel function and kernel parameters are set by cross-validation. The results show that the three schemes of RIS have significant improvement compared with ICA. At the same time, the performance of RIS is also superior to PCA and KPCA. These experimental results further suggest that the RIS is an effective method.

Table 1
Level 1 is accuracy of 10 classifiers each of which is aggregated from the corresponding 5 child classifiers in Level 2. Level 2 is accuracy of 5 classifiers constructed in sampled whitened feature subspace. The final is the best accuracy of RIS3 aggregated from the 10 classifiers in Level 1.

Accuracy rate (%)		1	2	3	4	5	6	7	8	9	10
Level 2	1	75.56	78.91	76.48	77.48	77.82	77.99	77.99	78.99	77.48	77.99
	2	76.73	76.82	77.57	77.90	77.90	77.99	76.65	78.49	77.82	77.99
	3	76.65	78.91	77.32	77.90	76.48	77.07	76.56	77.40	76.31	78.07
	4	77.82	77.82	76.90	75.64	77.32	78.91	77.82	78.41	78.32	78.49
	5	78.57	76.56	75.56	78.82	79.24	78.07	77.40	78.66	77.57	77.99
Level 1		80.42	81.42	79.83	80.42	81.00	80.84	80.25	81.51	80.59	80.17
Final		85.36									

5. Discussions and conclusions

In pattern classification using ICA classifier, we often encounter two challenging problems: small sample size problem and the choice problem of basis functions, which results in the unstable performance of ICA classifier. In order to improve the performance of ICA classifier, we proposed an enhanced ICA algorithm, Random Independent Subspace (RIS), which adopted random sampling strategy to produce a set of subspaces, and one ICA classifier is constructed on each subspace. Then all ICA classifiers are aggregated into an ensemble classifier using majority voting. In this paper, we give three schemes for RIS. Scheme I resamples in the original feature space, which reduces the dimensionality of feature space and increases the virtual samples for ICA classifiers so that the negative effect brought by small sample size is alleviated. The scheme II resamples in whitened feature space, which can effectively deal with the choice problem of basis functions. In order to possess the merits of schemes I and II, the scheme III adopts a two-level cascade resampling structure, which not only resamples in the original feature space, but also resamples in whitened feature space. Although resampling twice in scheme III increase the computational complexity, most of computation exists in training stage which is offline implemented, the test stage of scheme III just increases little time consumption compared with schemes I and II. The experimental results of face recognition suggest that the scheme III can significantly improve the performance of ICA classifier.

Acknowledgements

We would like to thank the anonymous reviewers for their useful comments. The work is supported by grants from Natural Sciences Foundation of China under Grant No. 60405005 and 60121302, and the joint fund of NSFC-RGC under Grant No. 60318003. The first author is also supported in part by postdoctoral fellowship from Nokia Research Center, Beijing.

References

[1] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, J. Opt. Soc. Amer. A 4 (3) (1987) 519–524.
[2] M.A. Turk, A.P. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.
[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[4] P. Comon, Independent component analysis—a new concept?, Signal Process. 36 (1994) 287–314.
[5] A.J. Bell, T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.
[6] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for Independent Component Analysis, Neural Comput. 9 (7) (1997) 1483–1492.
[7] J. Herault, J. Jutten, Space or time adaptive signal processing by neural network models, in: J.S. Denker (Ed.), Neural Networks for Computing: AIP Conference Proceedings, vol. 151, American Institute for Physics, New York, 1986.
[8] C. Jutten, J. Herault, Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture, Signal Process. 24 (1991) 1–10.
[9] M.J. Mckeown, T.J. Sejnowski, Independent component analysis of FMRI data: examining the assumptions, Hum. Brain Map. 6 (1998) 368–372.
[10] M.S. Bartlett, H.M. Lades, T.J. Sejnowski, Independent component representations for face recognition, Proc. SPIE 3299 (1998) 528–539.
[11] M.S. Bartlett, Face Image Analysis by Unsupervised Learning, Kluwer Academic Publisher, Princeton, NJ, 2001.
[12] J. Kittler, M. Hatef, R.P.W. Duin, J. Matas, On combining classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 20 (3) (1998) 226–239.
[13] X. Wang, X. Tang, Random sampling LDA for face recognition, IEEE Conference on Computer Vision and Pattern Recognition (CVPR '04), Washington, DC, USA, 2004, pp. 259–265.
[14] Y. Freund, R.E. Schapire, Experiments with a new boosting algorithm, in: Proceedings of 13th International Conference on Machine Learning (ICML '96), Bari, Italy, 1996, pp. 48–156.
[15] L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140.
[16] T.K. Ho, The Random Subspace Method for constructing decision forests, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 832–844.
[17] H.B. Barlow, Possible principles underlying the transformation of sensory messages, in: W.A. Rosenblith (Ed.), Sensory Communication, MIT Press, Cambridge, MA, 1961.
[18] A.J. Bell, T.J. Sejnowski, The 'Independent Components' of natural scenes are edge filters, Vision Research 37 (23) (1997) 3327–3338.

[19] P.J. Philips, H. Wechsler, J. Huang, P. Rauss, The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing 16 (1998) 295–306.

[20] B.A. Draper, K. Baek, M.S. Bartlett, J.R. Beveridge, Recognizing faces with PCA and ICA, Comput. Vision Image Understand. 91 (2003) 115–137.

About the Author—JIAN CHENG received his B.S. degree and M.S. degree in Mathematics from Wuhan University in 1998 and 2001, respectively. Then he obtained his Ph.D. degree in Pattern Recognition and Intelligence Systems from Institute of Automation, Chinese Academy of Sciences in 2004. Currently, he is a postdoctoral at Nokia Research Center. His research interests include face recognition, image retrieval, machine learning, etc.

About the Author—QINGSHAN LIU was born in 1975. He received his B.E. degree from Chinese Textile University (now DongHua University) in 1997, M.S. degree from South East University in 2000. He received his Ph.D. degree in 2003 from the National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Science, where now he is an assistant professor. His current research interests are image and video analysis, object tracking and recognition, and machine learning.

About the Author—HANQING LU professor. He received his B.E. degree in 1982 and his M.E. degree in 1985 from Harbin Institute of Technology, and Ph.D. degree in 1992 from Huazhong University of Sciences and Technology. Currently, he is a deputy director of National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interests include image and video analysis, medical image processing, object recognition, etc. He has published more than 100 papers in these fields.

About the Author—YEN-WEI CHEN received a B.E. degree in 1985 from Kobe University, a M.E. degree in 1987 and a D.E. degree in 1990, both from Osaka University, Japan. He is currently a professor in college of Information Science and Engineering, Ritsumeikan University. His research interests include intelligent signal and image processing, radiological imaging and soft computing. He has published more than 100 research papers in these fields.