Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

PERGAMON

Pattern Recognition 37 (2004) 2097–2100

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Rapid and Brief Communication

Essence of kernel Fisher discriminant: KPCA plus LDA

Jian Yang[a,b,c,*], Zhong Jin[a,d], Jing-yu Yang[a], David Zhang[c], Alejandro F. Frangi[b]

[a] *Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, PR China*
[b] *Computer Vision Group, Aragon Institute of Engineering Research, University of Zaragoza, E-50018 Zaragoza, Spain*
[c] *Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong*
[d] *Centre of Computer Vision, University Autonoma of Barcelona, E-08193 Barcelona, Spain*

Received 7 October 2003; accepted 15 October 2003

Abstract

In this paper, the method of kernel Fisher discriminant (KFD) is analyzed and its nature is revealed, i.e., KFD is equivalent to kernel principal component analysis (KPCA) plus Fisher linear discriminant analysis (LDA). Based on this result, a more transparent KFD algorithm is proposed. That is, KPCA is first performed and then LDA is used for a second feature extraction in the KPCA-transformed space. Finally, the effectiveness of the proposed algorithm is verified using the CENPARMI handwritten numeral database.
© 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Kernel-based methods; Fisher linear discriminant analysis; Principal component analysis; Feature extraction; Handwritten numeral recognition

1. Introduction

Fisher linear discriminant analysis (LDA) is a traditional statistical technique for dimensionality reduction. It has been widely used and proven successful in a lot of real-world applications. But, due to its limitation of linearity, LDA fails to perform well for nonlinear problems. To overcome this weakness of LDA, the non-linear versions of Fisher discriminant analysis have been proposed over the last few years. Mika [1] formulated kernel Fisher discriminant (KFD) for two-class cases, while Baudat [2] developed the generalized kernel discriminant analysis (GDA) for multi-class problems. Since two-class problems are a special case of multi-class problems, in this paper, we will focus on KFD analysis for multi-class cases.

KFD turns out to be more effective than LDA in various applications; however, the existing KFD algorithms are not as simple and transparent as LDA. It is the complicated formalization of KFD algorithms that covers the intuitive characteristics of kernel discriminant analysis. In this paper, we perform an in-depth analysis on the current KFD algorithm GDA [2], and reformulate it as a two-phase process: kernel principal component analysis (KPCA) [3] plus LDA. This formalization is more transparent and simpler than the previous ones. It provides us with a new viewpoint of KFD. On the one hand, the relations between KPCA and KFD become clearer. On the other hand, KFD itself becomes more intuitive, more understandable and easier to be implemented.

2. Analysis on kernel Fisher discriminant

2.1. Fundamentals

For a given nonlinear mapping Φ, the *input data space* $\mathbb{R}^n$ can be mapped into the *feature space* $\mathcal{F}$:

$$\Phi : \mathbb{R}^n \to \mathcal{F}, \quad x \mapsto \Phi(x). \tag{1}$$

Correspondingly, a pattern in the original *input space* $\mathbb{R}^n$ is mapped into a potentially much higher dimensional feature vector in the *feature space* $\mathcal{F}$.

* Corresponding author. Tel.: +86-25-84315510; fax: +86-25-84315510.
E-mail addresses: csjyang@comp.polyu.edu.hk (J. Yang), zhongjin@cvc.uab.es (Z. Jin), yangjy@mail.njust.edu.cn (J.-y. Yang), csdzhang@comp.poly.edu.hk (D. Zhang), afrangi@unizar.es (A.F. Frangi).

0031-3203/$30.00 © 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2003.10.015

The idea of KFD is to solve the problem of LDA in the *feature space* $\mathscr{F}$, thereby yielding a set of nonlinear discriminant vectors in *input space*. This can be achieved by maximizing the following Fisher criterion:

$$J^{\Phi}(\boldsymbol{\varphi}) = \frac{\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{S}_b^{\Phi}\boldsymbol{\varphi}}{\boldsymbol{\varphi}^{\mathrm{T}}\mathbf{S}_t^{\Phi}\boldsymbol{\varphi}}, \quad \boldsymbol{\varphi} \neq \mathbf{0}, \tag{1}$$

where $\mathbf{S}_b^{\Phi}$ and $\mathbf{S}_t^{\Phi}$ are the between-class and total scatter matrices defined in *feature space* $\mathscr{F}$:

$$\mathbf{S}_b^{\Phi} = \frac{1}{M}\sum_{i=1}^{c} l_i(\mathbf{m}_i^{\Phi} - \mathbf{m}_0^{\Phi})(\mathbf{m}_i^{\Phi} - \mathbf{m}_0^{\Phi})^{\mathrm{T}}, \tag{2}$$

$$\mathbf{S}_t^{\Phi} = \frac{1}{M}\sum_{i=1}^{M} (\Phi(\mathbf{x}_i) - \mathbf{m}_0^{\Phi})(\Phi(\mathbf{x}_i) - \mathbf{m}_0^{\Phi})^{\mathrm{T}}. \tag{3}$$

Here, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M$ is a set of M training samples in *input space*; l_i is the number of training samples of class i and satisfies $\sum_{i=1}^{c} l_i = M$; $\mathbf{m}_i^{\Phi}$ is the mean vector of the mapped training samples of class i; $\mathbf{m}_0^{\Phi}$ is the mean vector across all mapped training samples.

2.2. Kernel Fisher optimal discriminant vectors

The optimal discriminant vectors with respect to the Fisher criterion are actually the eigenvectors of the generalized equation $\mathbf{S}_b^{\Phi}\boldsymbol{\varphi} = \lambda\mathbf{S}_t^{\Phi}\boldsymbol{\varphi}$. Since any of its eigenvector can be expressed by a linear combination of the observations in *feature space*, we have

$$\boldsymbol{\varphi} = \sum_{j=1}^{M} a_j\Phi(\mathbf{x}_j) = \mathbf{Q}\boldsymbol{\alpha}, \tag{4}$$

where $\mathbf{Q} = [\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_M)]$ and $\boldsymbol{\alpha} = (a_1, \ldots, a_M)^{\mathrm{T}}$.

Substituting Eq. (4) into Eq. (1), the Fisher criterion is converted to [2]:

$$J^{\mathrm{K}}(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{KWK})\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{KK})\boldsymbol{\alpha}}, \tag{5}$$

where the matrix $\mathbf{K}$ is defined as

$$\mathbf{K} = \tilde{\mathbf{K}} - \mathbf{1}_M\tilde{\mathbf{K}} - \tilde{\mathbf{K}}\mathbf{1}_M + \mathbf{1}_M\tilde{\mathbf{K}}\mathbf{1}_M. \tag{6}$$

Here, $\mathbf{1}_M = (1/M)_{M\times M}$; $\tilde{\mathbf{K}} = \mathbf{Q}^{\mathrm{T}}\mathbf{Q}$ is an $M \times M$ matrix and, its elements are determined by

$$\tilde{\mathbf{K}}_{ij} = \Phi(\mathbf{x}_i)^{\mathrm{T}}\Phi(\mathbf{x}_j) = (\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)) = \mathrm{k}(\mathbf{x}_i, \mathbf{y}_j), \tag{7}$$

where $\mathrm{k}(\mathbf{x}, \mathbf{y})$ is the kernel function corresponding to a given nonlinear mapping Φ. And, $\mathbf{W} = \mathrm{diag}(\mathbf{W}_1, \ldots, \mathbf{W}_c)$, where $\mathbf{W}_j$ is a $l_j \times l_j$ matrix with terms all equal to $1/l_j$. Thereby, $\mathbf{W}$ is an $M \times M$ block diagonal matrix.

Now, let us consider the QR decomposition of matrix $\mathbf{K}$. Suppose $\gamma_1, \gamma_2, \ldots, \gamma_m$ are $\mathbf{K}$'s orthonormal eigenvectors corresponding to m (m is the rank of $\mathbf{K}$) nonzero eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$. Then, $\mathbf{K}$ can be expressed by

$$\mathbf{K} = \mathbf{P}\Lambda\mathbf{P}^{\mathrm{T}}, \text{ where } \mathbf{P} = (\gamma_1, \gamma_2, \ldots, \gamma_m) \text{ and}$$

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$$

Obviously, $\mathbf{P}^{\mathrm{T}}\mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

Substituting $\mathbf{K} = \mathbf{P}\Lambda\mathbf{P}^{\mathrm{T}}$ into Eq. (5), we have

$$J^{\mathrm{K}}(\boldsymbol{\alpha}) = \frac{(\Lambda^{1/2}\mathbf{P}^{\mathrm{T}}\boldsymbol{\alpha})^{\mathrm{T}}(\Lambda^{1/2}\mathbf{P}^{\mathrm{T}}\mathbf{WP}\Lambda^{1/2})(\Lambda^{1/2}\mathbf{P}^{\mathrm{T}}\boldsymbol{\alpha})}{(\Lambda^{1/2}\mathbf{P}^{\mathrm{T}}\boldsymbol{\alpha})^{\mathrm{T}}\Lambda(\Lambda^{1/2}\mathbf{P}^{\mathrm{T}}\boldsymbol{\alpha})}. \tag{8}$$

Let

$$\boldsymbol{\beta} = \Lambda^{1/2}\mathbf{P}^{\mathrm{T}}\boldsymbol{\alpha}. \tag{9}$$

Then, Eq. (8) becomes

$$J(\boldsymbol{\beta}) = \frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_b\boldsymbol{\beta}}{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_t\boldsymbol{\beta}}, \tag{10}$$

where

$$\mathbf{S}_b = \Lambda^{1/2}\mathbf{P}^{\mathrm{T}}\mathbf{WP}\Lambda^{1/2} \text{ and } \mathbf{S}_t = \Lambda. \tag{11}$$

It is easy to know that $\mathbf{S}_t$ is positive definite and $\mathbf{S}_b$ is semi-positive definite. So, Eq. (10) is a standard generalized Rayleigh quotient. By maximizing this Rayleigh quotient, we can obtain a set of optimal solutions $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_d$, which are actually the eigenvectors of $\mathbf{S}_t^{-1}\mathbf{S}_b$ corresponding to d ($d \leqslant c-1$) largest eigenvalues.

From Eq. (9), we know that for a given $\boldsymbol{\beta}$, there exists at least one $\boldsymbol{\alpha}$ satisfying $\boldsymbol{\alpha} = \mathbf{P}\Lambda^{-1/2}\boldsymbol{\beta}$. Thus, after determining $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_d$, we can obtain a set of optimal solutions $\boldsymbol{\alpha}_j = \mathbf{P}\Lambda^{-1/2}\boldsymbol{\beta}_j$ ($j = 1, \ldots, d$) with respect to the criterion in Eq. (5). Thereby, the optimal discriminant vectors with respect to the Fisher criterion in Eq. (1) in *feature space* are

$$\boldsymbol{\varphi}_j = \mathbf{Q}\boldsymbol{\alpha}_j = \mathbf{QP}\Lambda^{-1/2}\boldsymbol{\beta}_j, \quad j = 1, \ldots, d. \tag{12}$$

2.3. Essence of KFD transformation: KPCA + LDA

Given a sample $\mathbf{x}$ and its mapped image $\Phi(\mathbf{x})$, we can obtain the discriminant feature vector $\mathbf{z}$ by the following KFD transformation:

$$\mathbf{z} = \Psi^{\mathrm{T}}\Phi(\mathbf{x}), \tag{13}$$

where $\Psi = (\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \ldots, \boldsymbol{\varphi}_d) = (\mathbf{QP}\Lambda^{-1/2}\boldsymbol{\beta}_1, \mathbf{QP}\Lambda^{-1/2}\boldsymbol{\beta}_2, \ldots, \mathbf{QP}\Lambda^{-1/2}\boldsymbol{\beta}_d) = (\mathbf{QP}\Lambda^{-1/2})(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_d)$.

The above transformation can be divided into two items:

$$\mathbf{y} = (\mathbf{QP}\Lambda^{-1/2})^{\mathrm{T}}\Phi(\mathbf{x}) \tag{14}$$

and

$$\mathbf{z} = \mathbf{G}^{\mathrm{T}}\mathbf{y} \quad \text{where} \quad \mathbf{G} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_d). \tag{15}$$

Let us consider the transformation in Eq. (14) first. Since $\mathbf{Q} = [\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_M)]$, $\mathbf{P} = (\gamma_1, \gamma_2, \ldots, \gamma_m)$ and

$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, Eq. (14) can be rewritten as

$$\mathbf{y} = \left(\frac{\gamma_1}{\sqrt{\lambda_1}}, \ldots, \frac{\gamma_m}{\sqrt{\lambda_m}}\right)^{\mathrm{T}} (\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_M))^{\mathrm{T}} \Phi(\mathbf{x}),$$

$$= \left(\frac{\gamma_1}{\sqrt{\lambda_1}}, \ldots, \frac{\gamma_m}{\sqrt{\lambda_m}}\right)^{\mathrm{T}} [\mathrm{k}(\mathbf{x}_1, \mathbf{x}), \ldots, \mathrm{k}(\mathbf{x}_M, \mathbf{x})]. \quad (16)$$

Since $\gamma_1, \gamma_2, \ldots, \gamma_m$ are $\mathbf{K}$'s orthonormal eigenvectors and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the associated nonzero eigenvalues, the transformation in Eq. (16) is exactly the KPCA transformation [3], which transforms *feature space* $\mathcal{F}$ into Euclidean space $\mathbb{R}^m$.

Now, let us view the issues in the KPCA-transformed space $\mathbb{R}^m$. Looking back at Eq. (10) and considering the matrices $\mathbf{S}_b$ and $\mathbf{S}_t$ within the function $J(\boldsymbol{\beta})$, it is easy to verify that they are actually the between-class and total scatter matrices in the KPCA-transformed space $\mathbb{R}^m$. Since the expression of $\mathbf{S}_b$ in Eq. (11) is not so intuitive, we can construct it directly in $\mathbb{R}^m$, based on KPCA-transformed features:

$$\mathbf{S}_b = \frac{1}{M} \sum_{i=1}^{c} l_i (\mathbf{m}_i - \mathbf{m}_0)(\mathbf{m}_i - \mathbf{m}_0)^{\mathrm{T}}, \quad (17)$$

where l_i is the number of training samples in class i; $\mathbf{m}_i$ is the mean vector of the training samples in class i; $\mathbf{m}_0$ the mean vector across all training samples.

Since $\mathbf{S}_b$ and $\mathbf{S}_t$ are the between-class and total scatter matrices in the KPCA-transformed space $\mathbb{R}^m$, the function $J(\boldsymbol{\beta})$ is actually the (linear) Fisher criterion in such space and, its stationary points $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_d$ are the associated Fisher optimal (linear) discriminant vectors. Correspondingly, the transformation in Eq. (15) is essentially the Fisher linear discriminant transformation in the KPCA-transformed space.

Up to now, the essence of KFD has been revealed. That is, KPCA is first employed to reduce the dimension of *feature space* to m, and then LDA is used for further feature extraction in the KPCA-transformed space $\mathbb{R}^m$.

2.4. Two-phase KFD algorithm

As a summary of the preceeding discussion, the two-phase KFD algorithm is given below:

Step 1: *Perform KPCA in input space* $\mathbb{R}^n$. Construct the centralized inner product matrix $\mathbf{K}$ using Eq. (6) and calculate its orthonormal eigenvectors $\gamma_1, \gamma_2, \ldots, \gamma_m$ corresponding to m largest nonzero eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$. Carry out KPCA transformation using Eq. (16).

Step 2: *Perform LDA in KPCA-transformed space* $\mathbb{R}^m$. Construct the between-class scatter matrix $\mathbf{S}_b$ using Eq. (17) and the total scatter matrix $\mathbf{S}_t = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$. Calculate $\mathbf{S}_t^{-1}\mathbf{S}_b$'s eigenvectors $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_d$ corresponding to d ($d \leqslant c-1$) largest eigenvalues. Carry out LDA transformation using Eq. (15).

Here, a remark on the above algorithm should be given. For numerical robustness, in Step 1, m is usually selected as a number strictly less than the real rank of $\mathbf{K}$ (generally, the rank of $\mathbf{K}$ is $M-1$, where M is the total number of training samples) in practical applications.

3. Experiments

The experiment was performed on Concordia University CENPARMI handwritten numeral database. The database contains 6000 samples of 10 numeral classes (each class has 600 samples). Here, our experiment was performed based on two sets of original features: 256-dimensional Gabor transformation features and 121-dimensional Legendre moment features.

In the experiment, we chose the first 200 samples of each class for training, the remaining 400 samples for testing. Thus, the total number of training samples is 2000 while the total number of testing samples is 4000. The standard LDA, GDA [2] and the proposed algorithm KPCA plus LDA are, respectively, utilized to extract $c-1=9$ features based on each set of original features. For GDA and KPCA plus LDA, two popular kernels are involved. One is the second-order polynomial kernel $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y} + 1)^2$, and the other is Gaussian kernel, $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x}-\mathbf{y}\|^2/\delta)$, where the parameter δ is chosen as $0.3 \times n$ (n is the dimension of *input space*). And, m is selected as 800 by experience for both kernel-based methods mentioned above. Note that here, we implement GDA using the authors' original codes available at the website [4]. Finally, a minimum distance classifier is employed for classification. The classification results are listed in Table 1. Table 1 shows the results of KPCA plus LDA are almost the same as those of GDA. The small difference between them possibly results from the precision in calculation. These results are in accordance with what we expected, since KPCA plus LDA and GDA are essentially equivalent. Also, the method of KFD is shown to be more powerful than LDA in our experiments.

4. Discussion

In this paper, we proposed a new strategy, KPCA plus LDA, to implement KFD analysis. This strategy is exactly consistent with the existing two-phase linear Fisher discriminant analysis algorithm, PCA plus LDA, which has been a popular framework of LDA for small sample size problems [5]. Actually, this consistency is easily interpreted. If a linear kernel, i.e., $k(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$, is adopted instead of nonlinear kernels, the proposed algorithm would degenerate to be a PCA plus LDA algorithm like Fisherfaces [5].

Acknowledgements

This work was supported by grants TIC2002-04495-C02 from the Spanish Ministry of Science and Technology (MCyT) as well as the National Science Foundation of China under Grant No. 60072034. Zhong Jin and Alejandro F. Frangi are also supported by a Ramón y Cajal Research Fellowship from Spanish Ministry of Science and Technology.

Table 1
Comparison of the recognition rates (%) of LDA, GDA and the proposed KPCA plus LDA algorithm

Feature set	LDA	GDA (Gaussian)	KPCA + LDA (Gaussian)	GDA (Polynomial)	KPCA + LDA (Polynomial)
Gabor	81.3	91.5	91.5	90.2	90.2
Legendre	88.9	94.8	94.7	94.5	94.5

References

[1] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, K.-R. Müller, Fisher discriminant analysis with kernels, IEEE International Workshop on Neural Networks for Signal Processing IX, Madison, USA, August, 1999, pp. 41–48.

[2] G. Baudat, F. Anouar, Generalized discriminant analysis using a kernel approach, Neural Comput. 12 (10) (2000) 2385–2404.

[3] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.

[4] G. Baudat, F. Anouar, "Generalized Discriminant Analysis, codes for Matlab 5", http://www.kernel-machines.org/index.html.

[5] P.N. Belhumeur, J.P. Hespanha, D.J. Kriengman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Machine Intell. 19 (7) (1997) 711–720.

About the Author—JIAN YANG was born in Jiangsu, China, on 3rd June 1973. He obtained his B.S. degree in Mathematics at the Xuzhou Normal University in China in 1995. He then continued on to complete a Masters of Science degree in Applied Mathematics at the Changsha Railway University in 1998 and his Ph.D. at the Nanjing University of Science and Technology in the Department of Computer Science on the subject of Pattern Recognition and Intelligence Systems in 2002. From March to September in 2002, he worked as a research assistant at department of computing, Hong Kong Polytechnic University. Now, he is a postdoctoral research fellow at the University of Zaragoza and is affiliated with the Division of Bioengineering of the Aragon Institute of Engineering Research (I3A). He is the author of more than 20 scientific papers in pattern recognition and computer vision. His current research interests include pattern recognition, computer vision and machine learning. Dr. Yang has recently been awarded the Ramón y Cajal Research Fellowship, a national program of the Spanish Ministry of Science and Technology.

About the Author—ZHONG JIN was born in Jiangsu, China, on 4 December 1961. He received the B.S. degree in mathematics, M.S. degree in applied mathematics and Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology (NUST), China in 1982, 1984 and 1999, respectively. He is a professor in the Department of Computer Science, NUST. He visited the Department of Computer Science and Engineering, The Chinese University of Hong Kong from January 2000 to June 2000, and from November 2000 to August 2001 and visited Laboratoire HEUDIASYC, UMR CNRS 6599, Universite de Technologie de Compiegne, France from October 2001 to July 2002. Dr. Jin is now visiting Centre de Visio per Computador, Universitat Autonoma de Barcelona, Spain, as a Ramón y Cajal program Research Fellow. His current interests are in the areas of pattern recognition, computer vision, face recognition, facial expression analysis, and content-based image retrieval.

About the Author—JING-YU YANG received the B.S. Degree in Computer Science from Nanjing University of Science and Technology (NUST), Nanjing, China. From 1982 to 1984 he was a visiting scientist at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. From 1993 to 1994 he was a visiting professor at the Department of Computer Science, Missuria University and in 1998, he worked as a visiting professor at Concordia University in Canada. He is currently a professor and Chairman in the department of Computer Science at NUST. He is the author of over 100 scientific papers in computer vision, pattern recognition, and artificial intelligence. He has won more than 20 provincial awards and national awards. His current research interests are in the areas of pattern recognition, robot vision, image processing, data fusion, and artificial intelligence.

About the Author—DAVID ZHANG graduated in computer science from the Peking University in 1974 and received his MSc and PhD degrees in computer science & engineering from the Harbin Institute of Technology (HIT) in 1983 and 1985, respectively. In 1988, he joined the University of Windsor, Ontario, Canada, as a visiting research fellow in electrical engineering. He received his second PhD in electrical and computer engineering at the University of Waterloo, Ontario, Canada, in 1994. After that, he was an associate professor at the City University of Hong Kong and a professor at the Hong Kong Polytechnic University. Currently, he is a founder and director of the Biometrics Technology Centre supported by the UGC of the Government of the Hong Kong SAR, and holds the same position for the Biometrics Technology Centre supported by the National Natural Scientific Foundation (NSFC) of China. He is the Founder and Editor-in-Chief of the International Journal of Image and Graphics, and an Associate Editor in some international journals such as IEEE Trans. on SMC-C, Pattern Recognition, and International Journal of Pattern Recognition and Artificial Intelligence. His research interests include automated biometrics-based identification, neural systems and applications, and image processing and pattern recognition. So far, he has published over 180 articles, including seven books, around his research areas.