



一种新的特征融合方法及其在图像识别中的应用

孙全森^{a,b,*}, 曾胜根^a, 刘艳^a, 衡平安^c, 夏德森^a

^a 南京科技大学计算机系, 南京 210094, 中华人民共和国

^b 济南大学数学系, 济南 250022, 中华人民共和国

^c 香港中文大学计算机科学与工程系, 香港

2004年3月11日收到; 2004年12月17日收到修订版; 2004年12月17日接受

摘要

本文根据典型相关分析 (CCA) 的思想, 提出了一种基于特征融合的特征提取新方法。首先, 讨论了用于模式识别的CCA理论框架及其合理描述。其过程可以解释为: 提取两组具有相同模式的特征向量; 建立两组特征向量之间的相关准则函数; 提取它们的典范相关特征, 形成有效的判别向量进行识别。然后, 解决了两个总散射矩阵为奇异时的典范投影向量问题, 这样就适合于高维空间和小样本量的情况, 在这个意义上, CCA的适用范围得到了扩展。最后, 在理论上进一步分析了该方法用于识别的内在本质。在康科迪亚大学CENPARMI手写阿拉伯数字数据库和耶鲁大学人脸数据库上的实验结果表明, 识别率远远高于采用单一特征的算法或现有的融合算法。

© 2005年模式识别协会。由爱思唯尔有限公司出版。保留所有权利。

关键词: 佳能相关分析 (CCA); 特征融合; 特征提取; 手写字符识别; 人脸识别

1. 简介

信息融合技术是数据处理的新兴技术之一。在信息融合的三个层次 (像素级、特征级和决策级) 中, 由多分类器组合委托的决策级融合一直是模式识别的热点研究领域之一, 并在手写字符和人脸识别方面取得了成功的应用[1-3]。虽然特征级融合的研究始于

不像其他融合方法那样早, 它得到了令人高兴的发展[4,5]

。特征级融合的优势很明显。从同一图案中提取的不同特征向量总是反映出图案的不同特性。通过对这些不同的特征进行优化和融合, 不仅可以保持多特征的有效判别信息, 还可以在一定程度上消除多余的信息。这对分类和识别尤为重要。现有的特征融合方法有两种。一种是将两组特征向量组合成一个联合向量[4], 然后在高维的实向量空间中提取特征。另一种方法是将两组特征向量用一个复合向量结合起来[5,6], 然后在复合向量空间中提取特征。这两种特征融合方法都可以提高识别率, 特征融合

*

通讯作者。南京科技大学计算机科学系, 中华人民共和国南京市, 210094。Tel./fax: +86 5312927158。

电子邮件地址: qssun@beelink.com (Q.-S. Sun)。

基于联合向量的方法被称为串行特征融合，基于复合向量的方法被称为并行特征融合[6]。

正则相关分析（CCA）是处理两个随机向量之间相互关系的统计方法之一，它与主成分分析（PCA）和线性判别分析（LDA）在多变量统计分析中具有同样的重要性。它是

有价值的多数据处理方法之一[7,8]。近年来，CCA已被应用于多个领域，如信号处理、计算机视觉、神经网络和语音识别[9-13]。

本文提出了一种采用CCA思想的特征融合的新方法。首先，我们讨论了模式识别中使用的CCA的框架。即用同一样本提取两组特征向量，然后建立两组特征向量之间的相关准则函数，根据该准则提取它们的典型相关特征，并形成有效的判别向量用于识别。然后，解决了两个总散射矩阵是奇异的情况下的典范投影向量问题，这样就适合于高维空间和小样本量的情况，在这个意义上，CCA的适用范围得到了扩展。最后，从理论上分析了该方法用于识别的内在本质。该方法使用了以下的相关特征

形成，所以它不仅适用于信息融合。但同时消除了在该领域内的冗余信息。

纹理。这是一种新的分类和识别方法。在康科迪亚大学CEN-的实验结果

PARMI手写阿拉伯数字数据库和耶鲁大学标准人脸数据库显示，识别率远远超过了比采用单一特征或融合的结果要高。

该算法具有高效、稳健的特点。同时，仿真结果表明，该算法不仅可以实现对原始数据的压缩，而且还可以实现对其他数据的压缩。

在这个过程中，我们可以看到，我们的产品不仅具有很强的特征维度，而且还具有很好的分类能力。

性能，反映了图像的基本特征。本文的其余部分组织如下。第2节中，介绍了用于特征融合的CCA理论和方法

提出。在第3节中，对组合特征的问题

在高维度的情况下，提取的问题得到了解决。

在第4节中，我们提出了一种新的方法，即："在空间和小样本的情况下"。在第4节中，提出了Fea-

在大样本数据库

和小样本数据库上测试了图象融合方法，并与其他方法进行了比较。

在第5节中，我们在理论上讨论了这种方法的内在本质，并解释了为什么它在识别中是有效的。最后，在第5节中得出结论。

2. 特征融合的理论和方法

2.1. CCA的基本理念

在多变量统计分析中，经常需要研究两个随机向量的

将两个随机向量的相关研究转换为几对变量的相关研究，这些变量是不相关的。H.

Hotelling在1936年提出了这个想法[14]。

具体来说，考虑到两个零均值的随机向量 X 和 Y ，CCA找到一对方向 c 和 f_3 ，使投影 $a_1 = c^T X$ 和 $b_1 = f_3^T Y$ 之间的相关度达到最大。

投影 a_1 和 b_1 被称为第一个的一对典型变量。然后找到第二对典型变量 a_2 和 b_2 ，它们与典型变量 a_1 和 b_1 互不相关，同时也使它们之间的相关性最大化。就这样做，直到 X 和 Y 的所有相关特征都被提取出来。为了研究 X 和 Y 的相关性，我们只需要分析几对典型变量的相关性。

2.2. 组合特征提取的理论和算法

假设 w_1, w_2, \dots, w_c 是 c 个已知的模式类。让 $Q = \{r_i | r_i \in R^N\}$ 是一个训练样本空间。给定 $A = \{x | x \in R^p\}$ 和 $B = \{y | y \in R^q\}$ ，其中 x 和 y 是两个同一样本 r 的特征向量；由不同的分别是指。我们将讨论在转换后的训练样本特征空间 A 和 B 中的特征融合。

假设 A 和 B 被认为是两个随机矢量空间。我们的想法是根据CCA的思想来提取 x 和 y 之间的典型相关特征。

在第2.1节中描述的，我们把它们表示为 $c_1^T x$ 和 $f_3^T y$ （第一对）， $c_2^T x$ 和 $f_3^T y$ （第二对）， \dots ， $c_d^T x$ 和 $f_3^T y$ （第 d 对）。鉴于以下情况：

$$X^* = (c_1^T x, c_2^T x, \dots, c_d^T x)^T = (c_1, c_2, \dots, c_d)^T x = W_c^T x \quad (1)$$

$$Y^* = (f_3^T y, f_3^T y, \dots, f_3^T y)^T = (f_3, f_3, \dots, f_3)^T y = W_f^T y \quad (2)$$

在两个线性变换（3）和（4）之后。

$$Z_1 = \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} W_c^T x \\ W_f^T y \end{pmatrix} = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} \quad (3)$$

$$Z_2 = X + Y = W_x x + W_y y = \begin{pmatrix} W_x & W_y \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} \quad (4)$$

作为组合特征预测，分别用于分类，而变换矩阵为

$$W_1 = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix} \quad \text{和} \quad W_2 = \begin{pmatrix} W_x & W_y \end{pmatrix}$$

其中 $W_x = (c_1, c_2, \dots, c_d)$ ， $W_y = (f_3, f_3, \dots, f_3)$ 。

定义1.方向 c_i 和 f_{3i} 被称为 x 和 y 的第 i 对典型投射向量（CPV）， $c_i^T x$ 和 $f_{3i}^T y$

作为它们的第 i 个典型相关特征。我们也称 w_1 和 w_2 作为典范投射矩阵（CPM）， z_1 和 z_2 作为典范相关判别特征（CCDF）。

和线性变换 (3) 和 (4) 分别作为特征融合策略I (FF S I) 和特征融合策略II (FFS II)。

接下来, 我们将讨论如何获得CPV和CCDF的价值和质量。

假设 $S_{xx} \in R^{p \times p}$ 和 $S_{yy} \in R^{q \times q}$ 表示的协方差矩阵(它们的总散点矩阵是 nS_{xx} 和 nS_{yy} , 其中 n 是样本数)的 A 和 B 分别。而 $S_{xy} \in R^{p \times q}$ 表示它们的集间协方差矩阵, 那么 x 的协方差矩阵可以表示为

$$S = \begin{pmatrix} Var(x) & Cov(x, y) \\ Cov(y, x) & Var(y) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix},$$

其中 S_{xx} 和 S_{yy} 是正定的, $S_{xy}^T = S_{yx}$, 并且 $r = rank(S_{xy})$ 。

在本节中, 我们将只讨论 S_{xx} 和 S_{yy} 为非奇异的情况。对于单数的情况, 我们将在第3节讨论。

考虑到 x 和 y 的线性组合

$$c, Tx = a_1x_1 + a_2x_2 + \dots + a_px_p.$$

$$f_3Ty = b_1y_1 + b_2y_2 + \dots + b_qy_q.$$

其中 $c, \in R^p$ 和 $f_3 \in R^q$ 是两个任意的非零向量。CCA找到一对方向 c 和 f_3 , 使典范变量 c, Tx 和 f_3Ty 之间的相关性最大化。由于

$$Var(c, Tx) = c, TVar(x)c = c, TS_{xx}c,$$

$$Var(f_3Ty) = f_3TVar(y)f_3 = f_3TS_{yy}f_3,$$

$$Cov(c, Tx, f_3Ty) = c, TCov(x, y)f_3 = c, TS_{xy}f_3.$$

我们可以给出以下的标准函数。

$$J(c, f_3) = \frac{c, TS_{xy}f_3}{(c, TS_{xx}c, f_3TS_{yy}f_3)^{1/2}}. \quad (5)$$

显然, 标准函数 (5) 具有以下特点。

$$\begin{aligned} \cdot J(kc, f_3) &= J(c, f_3), \quad \forall k, l \in R. \\ \cdot J(c, f_3) &\text{的极值与长度无关。} \\ \cdot c, \text{和} f_3 &\text{, 但与它们的方向有关。} \end{aligned}$$

根据上述特点, 我们可以认为

$$c, TS_{xx}c = f_3TS_{yy}f_3 = 1. \quad (6)$$

现在问题转化为解决约束条件(6)中的CPV $c, \text{和} f_3$, 以及准则(5)的极值。假设 $(k-1)$ 对CPV $\{c_1, f_31\}, \{c_2, f_32\}, \dots, \{c_{k-1}, f_3k-1\}$ 得到, 那么第 k 个可以通过以下方式完成

解决以下优化问题。

最大 $J(c, f_3)$ 。

$$\begin{aligned} c, TS_{xx}c &= f_3TS_{yy}f_3 = 1, \\ \text{模式1} \quad c_i^T S_{xx}c &= f_{3i}^T S_{yy}f_3 = 0 \quad (i = 1, 2, \dots, k-1). \end{aligned} \quad (7)$$

我们将讨论模型I的最优解, 如下所示。使用拉格朗日乘数法来转换公式 (5)。

让

$$L(c, f_3) = c, TS_{xy}f_3 - \frac{A_1}{2}(c, TS_{xx}c - 1) - \frac{A_2}{2}(f_3^T S_{yy}f_3 - 1).$$

其中 A_1 和 A_2 是拉格朗日乘法器。

设定 $L(c, f_3)$ 关于以下的偏导数 $c, \text{和} f_3$ 为零

$$\frac{\partial L}{\partial c} = S_{xy}f_3 - A_1S_{xx}c = 0, \quad (8)$$

$$\frac{\partial L}{\partial f_3} = S_{yx}c - A_2S_{yy}f_3 = 0. \quad (9)$$

将公式 (8) 和 (9) 的两边都乘以 $c,^T$ 和 f_3^T 分别考虑约束条件 (6), 我们得到 $c, TS_{xy}f_3 =$

$$A_1c, TS_{xx}c = A_1, f_3TS_{yy}f_3 = A_2f_3TS_{yy}f_3 = A_2.$$

因为 $S_{xy}^T = S_{yx}$, 所以 $A_1 = A_2 = A$, 那么 $A_1 = A_2 = A$, 那么

$$J(c, f_3) = c, TS_{xy}f_3 = f_3TS_{yx}c = A. \quad (10)$$

这表明, 拉格朗日乘数 A_1 和 A_2 等于 c, Tx 和 f_3Ty 的相关系数。

因此, 公式 (8) 和 (9) 也可以写为

$$S_{xy}f_3 - AS_{xx}c = 0, \quad (11)$$

$$S_{yx}c - AS_{yy}f_3 = 0. \quad (12)$$

由于 S_{xx} 和 S_{yy} 都是正定的, 由公式 (11) 可知和 (12), 我们得到

$$S_{xy}S_{yy}^{-1}S_{yx}c = AS_{xx}c. \quad (13)$$

$$S_{yx}S_{xx}^{-1}S_{xy}f_3 = AS_{yy}f_3. \quad (14)$$

现在问题已经转化为解决两个

广义的特征问题。给定 $M_{xy} = S^{-1}S_{xy}S^{-1}S_{yx}$ 和 $M_{yx} = S^{-1}S_{yx}S^{-1}S_{xy}$ 则公式 (12) 和 (13) 改为

$$M_{yx}c = A^2c, \quad (15)$$

$$M_{xy}f_3 = A^2f_3. \quad (16)$$

这里是关于 "中国" 的特征值和特征向量的定理。

M_{xy} 和 M_{yx} 来自 [8]。

定理1. M_{xy} 和 M_{yx} 具有相同的非零特征值, 它满足 $\lambda^2 - A_2 = 0$, 其中 $r = \text{rank}(S_{xy})$ 。

为了得到约束条件(7)下的解决方案, 超假设: 1.

$$\begin{aligned} G_1 &= S^{-1/2} S_{xx} S_{xy} S_{yy}^{-1} S_{yx} S_{xx}^{-1} S^{-1/2}, \\ G_2 &= S^{-1/2} S_{yx} S_{xy}^{-1} S_{xx} S_{xy}^{-1} S^{-1/2}. \end{aligned} \quad (17)$$

那么我们可以得到, M_{xy} 和 G_1 具有相同的非零特征值, 根据矩阵的相关定理, 特征值, 所以做 M_{yx} 和 G_2 。这样, 非零特征值都为 A_2 。设 $H = S^{-1/2} S_{xy} S^{-1/2}$, 则 $g_1 = hh^T$, $g_2 = hh^T$ 。使用奇异值分解法

(对矩阵 H 的 SVD 定理, 我们得到 $H = U \Lambda V^T$, 其中 $\Lambda = \begin{pmatrix} A_1 & & \\ & A_2 & \\ & & 0 \end{pmatrix}$ 的全部非零特征值。

G_1 和 G_2 的 u_i 和 $v_i (i = 1, 2, \dots, r)$ 是正交的 G_1 和 G_2 的特征向量对应于非零特征值 A_2 , 分别。

从上面, 我们可以推断出一个重要的定理2, 即如下。

定理2. 给定 $c_i = S^{-1/2} u_i, f_{3i} = S^{-1/2} v_i, i = 1, 2, \dots, r$ 。那么

- (1) c_i 和 f_{3i} 是与 A_2 相关的 M_{xy} 和 M_{yx} 的特征向量。
 - (2) $\begin{pmatrix} c_i^T S_{xx} c_j & c_i^T S_{xy} f_{3j} \\ c_i^T S_{yx} f_{3j} & c_i^T S_{yy} f_{3j} \end{pmatrix} = \begin{pmatrix} A_2 & 0 \\ 0 & A_2 \end{pmatrix} \delta_{ij}$ 。
- 其中 $\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}, i, j = 1, 2, \dots, r$ 。

事实上, 定理2给出了在以下情况下的解决方案准则(5)的极值, 也是为了模型1的优化解。

根据公式 (10) 和定理2, 我们可以得到如下的推论。

推论1. 给定 G_1 或 M_{xy} 的所有特征值为 A_2 , 则 G_1 或 M_{xy} 的非零特征值为 A_2 。那么, 该

分离函数 (5) 有 $J(c_i, f_{3i}) = A_i (i = 1, \dots, p)$ 。

推论2. 根据上述特征融合策略II (FFS II), 提取的组合特征是不相关的, 而且组合的投射向量是 M 正态的。其中

$$M = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}.$$

证明. 设 $X^* = W^T X = (x_1^*, x_2^*, \dots, x_d^*)^T$, $Y^* = W^T Y = (y_1^*, y_2^*, \dots, y_d^*)^T$ 。

从定理2, 我们知道, $\forall i \neq j$
 $\text{cov}(x_i^*, x_j^*) = c_i^T S_{xx} c_j = 0$ 。

$$\text{cov}(y_i^*, y_j^*) = f_{3i}^T S_{yy} f_{3j} = 0.$$

在公式 (4) 中。

$$\text{cov}(x_i^* + y_i^*, x_j^* + y_j^*) = \text{cov}(x_i^*, x_j^*) + 2\text{cov}(x_i^*, y_j^*) + \text{cov}(y_i^*, y_j^*) = 0.$$

所以, 特征向量 z_2 的组成部分是不相关的。正弦波

$$\begin{pmatrix} c_i^T & f_{3i}^T \end{pmatrix} \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix} \begin{pmatrix} c_j \\ f_{3j} \end{pmatrix} = 2(1 + A_i) \delta_{ij}$$

所以 $\begin{pmatrix} c_i \\ f_{3i} \end{pmatrix}$ 和 $\begin{pmatrix} c_j \\ f_{3j} \end{pmatrix}$ 是 M 正态, 其中 $i, j = 1, 2, \dots, r$ 。

定理3. 在标准 (5) 下, 满足约束条件 (6) 和 (7) 的有效 CPV 的数量为 r 对, 在最, 其中 $r = \text{rank}(S_{xy})$, 并得到 $d(r)$ 对 CPV

是由两个特征方程 (15 和 16) 的前四个最大特征值所对应的特征向量组成的。满足公式 (18) 的。

证明. 从定理1和推论1, 我们知道

$$J(c_i, f_{3i}) = A_i, i = 1, 2, \dots, \min(p, q). \text{ 其中 } A_2 = \begin{pmatrix} A_1 & & \\ & A_2 & \\ & & 0 \end{pmatrix}$$

所以, $J(c_i, f_{3i}) = 0 (i = r + 1, \dots, \min(p, q))$ 。在这种情况下, 有效的 CPV 不能被提取出来。这意味着有效 CPV 的数量最多只有 r 对。那么从定理来看, 2, 我们知道, $d(r)$ 对 CPV 可以由两个特征方程 $M_{xy} c = A_2 c$, 和 $M_{yx} f = A_2 f$ 的前 d 个最大特征值对应的特征向量组成。分别, 并满足条件 (18)。

2.3. 该算法的步骤

第1步: 提取两个不同的特征向量的集合

用相同的模式样本形成由原始模式样本空间 Q 转换而来的训练样本空间 A 和 B 。

第2步: 计算 A 和 B 中样本的协方差矩阵 S_{xx} 和 S_{yy} , 以及它们的套间协方差矩阵 S_{xy} 。

第三步: 根据公式 (17) 计算 G_1 和 G_2 , 然后找到它们的非零特征值 A_2 。

并对应于正态特征向量 u_i 和 $v_i (i = 1, 2, \dots, r)$ 。

第4步: 得到 CPV c_i 和 $f_{3i} (i = 1, 2, \dots, r)$, 根据定理2, 选择第一个 d 对向量来做 CPT。

第5步: 使用FFS I或FFS II提取CCDF, 这是, 用于分类。

从公式 (11) 和 (12) 可以看出, 我们只需要解决每一对CPV中的一个, 其他的可以通过公式 (11) 或 (12) 解决。一般来说, 我们可以选择低阶矩阵G1或G2来寻找其特征值和特征向量, 这样就可以降低计算的复杂性。

2.4. 特征融合的预处理

通常情况下, 当两组相同的特征被提取出来时, 可以根据上述算式直接提取CCDF。根据定理3, 我们可以得出结论, 最多可以有 r 对有效的CPV。因此, 提取的CCDF的维数不会超过 r 。当两组特征的维数 p 与 q 相差很大时, 提取的CCDF是否满足分类的需要? 维数较高的一组特征是否会损失大量的信息? 在直接感知中, 这种可能性是存在的。因此, 我们可以采用以下方法进行预处理, 以减少特征维数差异大所带来的不利影响。对于

两组特征向量 $x \in R^p$ 和 $y \in R^q$, 融合

首先按照串行模式进行, 形成一个 $(p+q)$ 维的联合向量。然后, 数据被重新分配到

得到两组新的特征向量 $x^* \in R^{\frac{p+q}{2}}$ 和 $y^* \in R^{\frac{p+q}{2}}$ (或 $x^* \in R^{\frac{p+q+1}{2}}$ 和 $y^* \in R^{\frac{p+q-1}{2}}$)。最后, 使用上述公式提取 x^* 和 y^* 之间的CCDF。

这种预处理的效果将在以后的实验中得到证明。

3. 高维情况和小样本量问题

3.1. 算法和理论

在模式识别领域, 有许多典型的小样本问题。例如, 在人脸识别中, 我们总是需要处理高维情况和小样本量的问题。在这些问题中, 常见的是样本的总散射矩阵是单数。因为要识别的图像向量的维度很高, 要找到足够的训练样本以保证总散射矩阵的可逆性是非常困难甚至不可能的。在这种情况下, 本文讨论的主要问题是

本节将介绍如何在特征空间中获得CPV。如第2.2节中提到的A和B。

作为一种重要的多数据处理方法, CCA在许多领域都有应用。然而, 这些应用仅仅局限于两个随机向量所属的特征空间的协方差矩阵为非奇异的情况[8-15]。一种处理奇异协方差矩阵和控制复杂性的方法

在 S_{xx} 和 S_{yy} 中加入一个倍数的识别矩阵 AI , $A>0$; 这个操作只是将特征值移动了 A , 因此, 如果 A 被选择得足够大, 将使两个矩阵都是正定的[16]。这是

另一种方法, 即用general reversibility来解决CPV的解的问题[7]。如上所述, 该解在理论上不是一个精确的解。因此它的应用是有限的。我们的想法是在不丢失任何有效信息的前提下, 将原始样本的高维特征空间转换为低维的欧几里得空间。而协方差矩阵在低维欧几里得空间中是非星形的, 所以我们只需要提取CPV在空间。

假设 S_{xx} 和 S_{yy} 中至少有一个是单数, 例如, S_{xx} 是单数, 而 S_{yy} 是非单数 (当 S_{xx} 和 S_{yy} 都是单数时, 方法类似), 我们可以从协方差矩阵的定义知道 S_{xx} 是半正定的, 而 S_{yy} 是正定的。然后我们给出了如何得到CPV解的方法。

假设 y_1, y_2, \dots, y_p 是 S_{xx} 的 p 个正态特征向量, 那么 $R^p = \text{span}\{y_1, y_2, \dots, y_p\}$ 。

定义2. 给定一个子空间 $\langle l = \text{span}\{y_1, y_2, \dots, y_m\}$, 其正补是 $\langle l^\perp = \text{span}\{y_{m+1}, \dots, y_p\}$, 其中 $m = \text{rank}(S_{xx})$ 。 y_1, y_2, \dots, y_m 为特征向量, 对应于导致 S_{xx} 的非零特征值。

从定义2我们可以知道, $R^p = \langle l + \langle l^\perp$, 对于 $\forall c, \in R^p$, c 可以表示为 $c = c_\perp + c_\parallel$, 其中 $c_\perp \in \langle l^\perp$, $c_\parallel \in \langle l$, 然后定义映射 $f: c \rightarrow c_\parallel$ 。很明显, f 是一个从 R^p 到 $\langle l$ 的线性变换。

根据线性代数的关系定理, 我们可以很容易地得出如下结论

悖论。如果 S_{xx} 是奇异的, 那么 $c, TS_{xx}c = 0$, 当且仅当 $S_{xx}c = 0$ 。

根据Lemma, 我们可以很容易地发现, $\langle l^\perp$ 是无效空间的 S_{xx} 。

定理4. 有了上述线性变换 $f: c \rightarrow c_\parallel$, 则 $J(c, \beta) = J(c_\parallel, \beta)$ 。

证明。借定理和 $\langle l^\perp$ 的定义表明, c_\perp

$$c, TS_{xx}c = c_\perp TS_{xx}c_\perp = 0. \quad (19)$$

所以

$$c, TS_{xx}c = c, TS_{xx}c_\parallel + 2c, TS_{xx}c_\perp + c_\perp TS_{xx}c_\perp = c, TS_{xx}c_\parallel. \quad (20)$$

假设

$$X = (x_1 - \mu_x, x_2 - \mu_x, \dots, x_n - \mu_x)$$

$$Y = (y_1 - \mu_y, y_2 - \mu_y, \dots, y_n - \mu_y)$$

$$Wx = (c, 1, \dots, c, d) = (P\tilde{a}, \dots, Pc, \tilde{d}) \\ = P(\tilde{a}, \dots, c, \tilde{d}) = PW^{\sim}x.$$

所以

$$W_1 = \begin{pmatrix} W_x^{\sim} & 0 \\ 0 & E \end{pmatrix} = P \begin{pmatrix} W_x^{\sim} & 0 \\ 0 & E \end{pmatrix} P^T \\ = \begin{pmatrix} P & 0 \\ 0 & E \end{pmatrix} \begin{pmatrix} W_x^{\sim} & 0 \\ 0 & E \end{pmatrix} P^T, \\ W_2 = \begin{pmatrix} W_x^{\sim} & 0 \\ 0 & E \end{pmatrix} = P \begin{pmatrix} W_x^{\sim} & 0 \\ 0 & E \end{pmatrix} P^T = \begin{pmatrix} P & 0 \\ 0 & E \end{pmatrix} \begin{pmatrix} W_x^{\sim} & 0 \\ 0 & E \end{pmatrix} P^T,$$

其中 E 是一个身份矩阵。

上述转化可以分解为两个转化。

$$\begin{pmatrix} x^{\sim} \\ y^{\sim} \end{pmatrix} = \begin{pmatrix} P^T & 0 \\ 0 & E \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (24)$$

$$Z_1 = \begin{pmatrix} W_x^{\sim} & 0 \\ 0 & W_y^T \end{pmatrix} \begin{pmatrix} x^{\sim} \\ y^{\sim} \end{pmatrix} \quad (25)$$

$$Z_2 = (W_x^T W_y)^T \begin{pmatrix} x^{\sim} \\ y^{\sim} \end{pmatrix} \quad (26)$$

在变换(24)中, $x^{\sim} = P$

Tx , $y^{\sim} = y$, 特征空间 B 没有发生变化, 但特征空间 A 成为一个新的 m 维特征子空间。由于

P 的列向量是对应于 S_{xx} 的非零特征值的特征向量, 所以这个变换被称为K-L变换, 在K-

L变换空间, 训练样本的协方差为

所以很明显, 由模型3确定的最优解 $\tilde{a}, \dots, c, \tilde{d}; f_{31}, \dots, f_{3d}$ (在 $K-L$ 变换空间 R^m 和特征空间 B 中基于准则 $J^{\sim}(c, \tilde{d}, f_3)$ 的CPV)。

从上面的分析中, 我们可以得出CCDF在奇异性条件下的提取过程(考虑到常见的情况, 即 S_{xx} 和 S_{yy} 都是奇异的)。

- 高维原始样本的维度可以通过K-L变换降低为 $rank(S_{xx})$ 和 $rank(S_{yy})$ 。
- 使用公式(25)或(26)来提取反式特征空间中的CCDF。

以上从理论上讨论了当 S_{xx} 和 S_{yy} 都是单数时的CPV求解问题。在实践中, 特别是针对模式分类的问题, 有时为了提高特征提取的速度, 在PCA(或K-L变换)的步骤中, 我们会选择与 S_{xx} (或 S_{yy})相对较大的非零特征值相对应的特征向量来组成变换矩阵 W_{PCA} 。然而, 一些对应于太小的非零特征值的特征向量将被放弃。所以得到的CCDF不会影响分类效果。

因为在模式分类中, 太小的非零特征值可能包括更多的干扰信息, 这将导致过拟合问题[17]。

4. 实验和分析

4.1. 关于CENPARMI手写数字的实验数据库

本实验的目的是测试以下内容的正确性
本报告中提出的大样本的算法。

每。我们采用了康科迪亚大学的CENPARMI手写阿拉伯数字数据库, 该数据库在世界范围内很流行。在这个数据库中, 有10个类别, 即10个数字(从0到9), 每个类别有600个样本。训练样本和测试样本分别为4000和2000。Hu等人[18]做了一些预处理工作, 提取了以下四种特征。 X^G : 256维的Gabor变换特征[19]。

X^L : 21维的Legendre时刻特征[20]。

X^P : 36维的Seudo-Zernike力矩特征[21]。

X^Z : 30维Zernike时刻特征[22]。

4.1.1. 关于两个特征融合的实验

将上述四个特征中的任意两个特征在原特征空间进行组合, 采用2.3节所述的算法, 得到CPM

w_1 和 w_2 , 然后通过FFS I和FFS II提取CCDF。采用最小距离分类器进行分类, 分类错误率如表1所示。

注意在获得CPV时, 我们应该先获得低维的CPV, 然后再根据公式(11)或(12)获得高维的。例如, 当我们

想结合 X^G 和 X^P 两个功能, 以获得

CPV有36维和256维, 我们只需要从36维矩阵 GXP 中获得特征值和相应的特征向量。

将此算法与现有的两个特征水平进行比较融合方法[4,6], 这里给出了实验结果。首先, 每个样本应该被归一化。然后, 根据文献[4]中描述的串行特征融合方法, 我们将两组特征向量组合成一个联合向量。[4], 并根据文献[6]中描述的并行特征融合方法, 将两组特征向量通过一个复合向量组合起来。[6]最后, 我们将合并后的特征放在合并空间中, 即对其训练样本的总散点矩阵进行K-L变换。分类是由最小距离分类器完成的。分类结果显示在表1中。

在以上四组特征中, 不同组的特征向量的维数差异可能很大。为了分析第2.4节中描述的预处理方法的有效性, 我们对本文提出的算式进行了对比实验。例如, 对于特征组合的组

表1

不同特征融合方法中基于两个特征的分类错误率

结合特点	串行融合	尺寸	平行融合	尺寸	FFS I	尺寸	FFS II	尺寸
XG, XL	0.1920	96	0.1925	48	0.1170	242	0.1290	85
XG, XP	0.2280	128	0.2285	80	0.2050	72	0.2181	31
XG, XZ	0.2295	116	0.2290	74	0.2230	60	0.2255	20
XL, XP	0.2400	102	0.2410	61	0.1810	72	0.2160	34
XL, XZ	0.2505	79	0.2505	59	0.2110	60	0.2312	30
XP, XZ	0.4760	51	0.4785	31	0.3215	56	0.3295	30

表2

通过第2.4节的预处理，基于两个特征的分类错误率

结合特点	FFS I	缩略语 狮子座	FFS II	缩略语 狮子座
$(x^g + x^l)/2, (x^g + x^l)/2$	0.1175	376	0.1415	172
$(x^g + x^p)/2, (x^g + x^p)/2$	0.1430	292	0.1765	138
$(x^g + x^z)/2, (x^g + x^z)/2$	0.1500	286	0.1800	143
$(x^l + x^p)/2, (x^l + x^p)/2$	0.1400	156	0.1955	78
$(x^l + x^z)/2, (x^l + x^z)/2$	0.1495	150	0.2100	73

表3

基于Gabor特征和Legendre特征的单一特征的分类错误率

单一功能	参考文献[23]。			原始特征
	ULDA	FSLD	徐勇	
XG	0.199	0.274	0.198	0.269
XL	0.141	0.270	0.150	0.479

XG 和 XP ，我们首先将相应的两组特征向量按照串行模式组合成286维的特征向量，
tor。然后，从每个特征向量中抽取一半，形成两组143维的特征向量。在这两组特征向量上，本文所描述的运算法则被用来提取CCDF。对于这种组合模式，我们简要记录如下。

$(x^g + x^p)/2 - (x^g + x^p)/2$ 。在实验中，我们采用最小距离分类器和相应的实验结果见表2。

这种特征融合的优势，我们在表3中列出了近年来在同一数据库中，基于单一特征进行识别的实验结果。如徐勇等人提出的ULDA和FSLDA算法[23]。此外，我们还给出了基于原始特征的识别结果。实验都是用最小距离分类器完成的。

从表1-3可以看出，我们的方法（FFS I和FFS II）的识别错误率低于串行和并行融合的方法。当结合

表4

在FFS I和FFS II中基于多特征的分类错误率

结合特点	FFS I 尺寸	FFS II 尺寸
$XG, (XL + XP)$	0.0820 300	0.0925 151
$XG, (XL + XZ)$	0.0810 292	0.0935 132
$(x^g + x^p), (x^l + x^z)$	0.0830 302	0.0920 118
$(x^g + x^z), (x^l + x^p)$	0.0790 304	0.0895 140

在Gabor特征和Legendre特征之间，FFS I和FFS II比其他基于单一特征的方法表现更好。这表明本文提出的算法是一种有效的特征融合方法。此外，从表1-2中我们可以看出，结合不同的特征会影响识别率。因此，融合前的特征筛选是最重要的。

从表1-2可以看出，在融合 XG 之前或 XL 与 XP 或 XZ ，首先进行预处理（2.4节），然后提取CCDF，在不同程度上改善了分类结果。以 XG 与 XZ 为例，在两种特征融合策略（FFS I和FFS II）下，预处理后的分类错误率分别降低了7.3%和4.55%，显示了2.4节所述预处理模式的有效性。此外，我们还可以看到，在对 XG 和 XL 进行同样的预处理后，分类结果没有发生明显变化。我们认为，两组特征向量的维数相差不大，可以直接提取CCDF；只有在维数相差较大时，才适合采用第2.4节的预处理方式。

4.1.2. 多特征融合的实验

这个实验给出了几组特征同时参与融合的结果。例如，对于 XG 、 XL 和 XZ 三组特征的融合，有一种组合模式是：先将 XL 与 XZ 按序列模式组合成一组特征向量，再与另一组 XG 组合；根据本文（第2.3节）的算式提取CCDF，用于分类。这种组合模式简称为 $XG, (XL + XZ)$ 。在

表4，图1和图2提供了实验的结果，其中几组特征在不同的情况下参与融合。

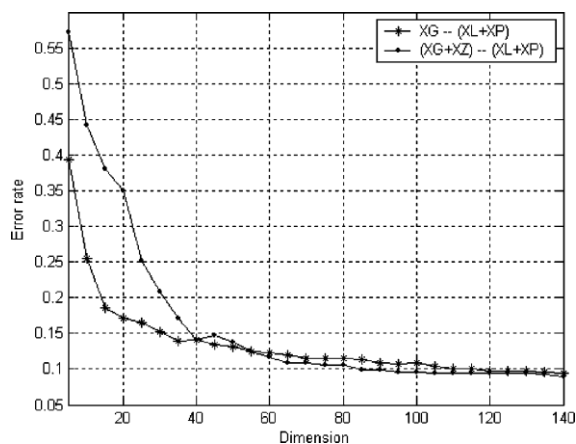


图1. $XG - (XL + XP)$ 和 XG 的分类错误率
 $(XG + XZ) - (XL + XP)$ 在 FFS II 中。

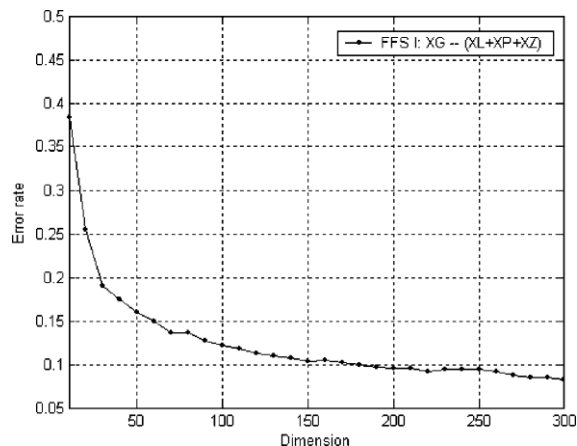


图3. FFS I 中 $XG - (XL + XP + XZ)$ 的分类错误率。

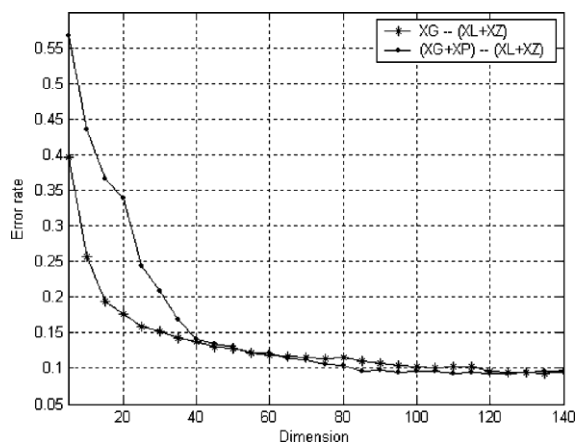


图2. $XG - (XL + XZ)$ 和 XZ 的分类错误率。
 $(XG + XP) - (XL + XZ)$ 在 FFS II 中。

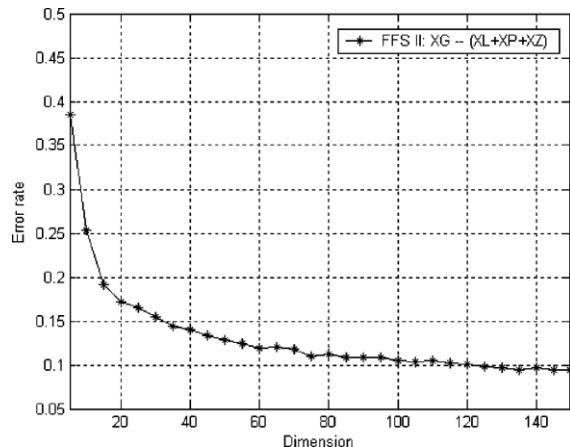


图4. FFS II 中 $XG - (XL + XP + XZ)$ 的分类错误率。

模式。所有这些实验都是在最小距离分类器下进行的。

在上述所有实验中，当使用本文的算术方法获得 CDV 时，第 2.2 节中提到的 S_{xx} 和 S_{yy} 都是正定的条件得到了满足，因此，CCDF 可以直接使用第 2.3 节中描述的算术程序来提取。下面我们给出四组特征的其他组合形式，例如

$XG - (XL + XP + XZ)$ ，从而产生了奇异的 S_{yy} ，重新引用第三节提供的算术原理。假设由 XG 和 $(XL + XP + XZ)$ 形成的训练样本空间的特征为 $A = \{x \mid x \in R^{256}\}$ 和 $B = \{y \mid y \in R^{187}\}$ ，由于 $r = \text{rank}(S_{yy}) = 179$ ，我们知道 S_{yy} 是单一的。根据第 3.2 节的讨论，我们首先

使用 K-L 变换将原始图像的 187 维减少到 179 维，形成新的训练样本空间。

$B \sim \{y \mid y \in R^{179}\}$ ，那么第 2.3 节的算术被用来提取训练样本空间 A 和 B 上的 CCDF。

仍然使用最小距离分类器。识别结果请参考图 3 和图 4。

上述实验结果表明，多组特征参与融合后，由于信息量的增加，提取的 CCDF 包含了更多有效的犯罪信息，从而使识别结果得到极大的改善。在本文提出的两种融合策略下，即 FFS I 和 FFS II，采用最小距离分类器，最佳识别率分别可以达到 92% 和 91%。此外，分类错误率下降得非常快。此外，从实验结果可以看出，FFS I 的识别效果略优于 FFS II。然而，由 FFS II 提取的 CCDF 在维数上有优势。



图5.耶鲁大学数据库中一个人的11张脸部图像的典型例子。



图6.一个原始图像，低频图像和双小波变换的高频图像。

4.2. 在耶鲁大学人脸图像数据库上进行的实验

本实验采用了耶鲁大学的人脸数据库。有15个人，分别有11张面部图像，所以总共有165张图像。每张图像的大小为 120×91 ，每个像素有256个灰度等级。这些图像

从不同的角度拍摄，有不同的变化。

表情和光照，以及部分图像不是整体。图5显示了一个人的图像的典型例子。

在这个实验中，我们使用每个人的前五张图像进行训练，其余六张用于测试。因此，训练样本和测试样本的总数分别为75和90。

首先，我们使用Daubechies正交小波对原始图像进行双小波变换（图6为双小波变换后的原始图像、低频图像和高频图像），提取低频图像的特征向量，组成第一特征空间。

的样本 $A = \{x \mid x \in R^{690}\}$ ；然后提取奇异的我们将原始图像的特征值向量合并为样本的第二特征空间 $B = \{y \mid y \in R^{91}\}$ ；最后我们将这两组特征合并。

我们将上述两组特征结合起来的原因是，低频子图像包含更多的形状信息，而从原始图像中提取的奇异值包含更多的纹理信息。所以这两组特征的相互补充将有利于提取有效的CCDF。

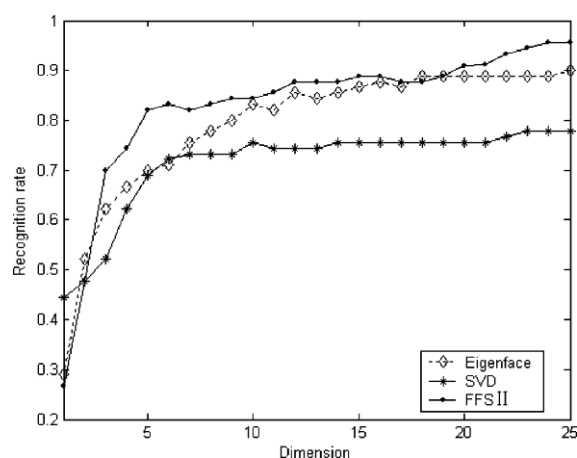


图7.eigenface、SVD方法和FFS II在最近邻分类器中的识别率。

在这个实验中，首先计算了 S_{xx} 和 S_{yy} 的等级，它们都等于74。然后利用SVD理论计算出30个正交特征向量，这些特征向量对应于 S_{xx} 和 S_{yy} 的非零特征值。因此变换矩阵 P 和 Q 是由这些获得的特征向量组成的。使用K-L变换 $\tilde{x} = P^T x$ 和 $\tilde{y} = Q^T y$ ，低频图像向量 $30 \times 23 = 690$ 的尺寸将减少为 $m_1 = 30$ 维，而原始图像的91维奇异值特征向量将被减少到 $m_2 = 30$ 维。在转换后的两个特征子空间中，两组特征将被归一化，在本文中该算法被用来获得CPM。使用FFS II来提取CCDF并使用最近邻分类器，识别结果如图7所示。

为了解释组合特征的有效性，我们给出了基于单一特征的Eigenface (PCA) [24]方法和SVD方法[25]。图7中也分别给出了采用最近邻分类器的识别结果。

表5
Eigenface、SVD方法、FFS I和FFS II在不同分类器中的识别率

分类器	Eigenface	SVD方法	FFS I	FFS II
最短距离	0.8556	0.5333	0.9444	0.9333
最近的邻居	0.9111	0.7889	0.9667	0.9556

此外，对于上述所有方法，我们也显示了在不同分类器中的识别结果。识别结果如表5所示。

从图7和表5可以看出，采用组合特征提取后的分类结果比采用单一特征提取的结果有了很大的提高。FFS I的识别率在96%以上，超过PCA方法的5%，SVD方法的26%。因此，我们，可以得出结论：组合CCDF具有更强大的识别能力，是一种高效的组合特征提取方法。此外，实验结果还表明，与其他方法相比，本文的算法对光照和表情不那么敏感。

5. 对算法有效性的分析

在模式识别的理论中，特征提取的共同原则是所选特征的统计相关性越小越好。提取的特征最好是不相关的。钟进等人[26,27]提出了一种不相关的最优判别向量的方法，并被应用于人脸识别和字符识别领域，得到了良好的效果。该理论的精髓在于判别向量的分量是不相关的，而且投射向量关于总散射矩阵是正态的。

从定理2和推论2来看，CPV分别关于 S_{xx} 、 S_{yy} 和 M 是正交的，也就是说，CCDF的分量是不相关的。因此，这种投影变换是。

通常，所花费的时间在很大程度上取决于投影向量的计算过程，使用代数方法来提取判别特征。当矩阵的等级很高时，特征值的计算就会受到影响。判断特征向量是很耗时的。在同样的模式下，假设特征向量 $x \in R^p$ 和 $y \in R^q$ ，参考文献[4]的投射向量是在 $(p+q)$ 维的实数向量空间中完成的。参考文献[4]是在 $(p+q)$ 维的实向量空间中完成的，而参考文献[5,6]是在最大 (p, q) 维的复向量空间中完成的。但我们的一个方法只在最小 (p, q) 维的实向量空间中进行。当 p 和 q 很大时，在计算速度上的优势是很明显的。例如，在第4.1节中，当我们要结合两个特征 XG 和 XZ 时，有三种结合方法选择了以下维度：286，256和30。

此外，基于Fisher准则的线性判别分析（FLDA）是特征提取和识别的最佳有效方法之一。FLDA也是CCA的一种特殊情况，可以通过理论来解决，一些基于Fisher准则的算法可以转化为本文介绍的方法（我们应该在其他论文中讨论）。所以将CCA应用于模式识别是比较普遍的，也是比较有发展潜力的。

6. 总结

本文首次将CCA的思想应用于特征融合和图像识别。提出了一种新的特征融合方法，它利用两组特征的相关特征作为有效的判别信息，因此它不仅适合于信息融合，而且可以消除特征内部的冗余信息。这为两组特征融合应用于判别提供了一种新的方法。

讨论了在图像识别中使用CCA的理论和方法。当两个总散射矩阵为奇异时，CPV的问题可以得到解决，这样就可以适应高维空间和小样本量的情况，因此CCA的适用范围在理论上得到了扩展。

此外，本文提出的方法与现有的两种特征融合方法进行了理论上的比较，并提出了该方法在识别中的内在本质。从大样本和小样本的实验结果来看，提取CCDF可以实现原始特征维度的降低，并能很好地反映图像本质特征的分类性能。本文的方法是一种在特征层面上进行信息融合的好方法。我们应该改进这种方法。

最后，应该注意的是，第三节中得到的结论，即CCA的完善和发展，完全适用于使用CCA的其他领域。

鸣谢

这项工作得到了香港特别行政区研究资助局的专项研究资助（项目编号：CUHK4185/00E）的支持。

参考文献

[1] Y.S. Huang, C.Y. Suen, Method of combining multiple experts for the recognition of unconstrained handwritten numerals, IEEE Trans.模式分析. Mach.Intell.7 (1) (1995) 90-94.

[2] A.S. Constantinidis, M.C. Fairhurst, A.F.R. Rahman, A new multi-expert decision combination algorithm and its application to detection of circumscribed masses in digital mammograms, Pattern Recognition 34 (8) (2001) 1528-1537.

[3] X.-Y.Jin, D. Zhang, J.-Y.Yang, 基于群体决策组合方法的人脸识别, 模式识别36 (7) (2003) 1675-1678.

- [4] C.J. Liu, H. Wechsler, A shape-and texture-based enhanced Fisher classifier for face recognition, *IEEE Trans.Image Process.* 10 (4) (2001) 598–608.
- [5] J.Yang, J.-Y.Yang, 基于广义K-L变换的 组合特征提取, 模式识别 35 (1) (2002) 295–297.
- [6] J.Yang, J.Y. Yang, D. Zhang, J.F. Lu, 特征融合：并行策略与串行策略, 模式识别36 (6) (2003) 1369–1381.
- [7] X.T. Zhang, K.T. Fang, *Multivariate Statistical Introduction*, Sciences Press, Beijing, 1999.
- [8] W.S. Sun, L.X. Chen, 《多元统计分析》，高等教育出版社，北京，1994.
- [9] M.Borga, 学习多维信号处理，林雪平科技研究，论文，第531卷，林雪平大学电子工程系，瑞典林雪平，1998年。
- [10] S.-J. Yu, Direct blind channel equalization via the programmable canonical correlation analysis, *Signal Process.* 81 (2001) 1715–1724.
- [11] C.Fyfe, P.L. Lai, Canonical correlation analysis neural networks, *International Conference on Pattern Recognition*, vol. 2, Barcelona, 2000, pp.977–980.
- [12] H.C. Choi, R.W. King, Speaker adaptation through spectral transformation for HMM based speech recognition, *IEEE Int.Symp.Speech Image Process.* 神经网络 2 (1994) 686–689.
- [13] D.Weenink, Canonical Correlation Analysis, *Institute of Phonetic Sciences, University of Amsterdam, Proceedings*, vol. 25, 2003, pp.81–99.
- [14] H.Hotelling, 两组变量之间的关系, *Biometrika* 8 (1936) 321–377.
- [15] T.Melzer, M. Reiter, H. Bischof, 基于核函数相关分析的外观模型, 模式识别36 (9) (2003) 1961–1971.
- [16] J.Friedman, Regularized discriminant analysis, *J. Am.Statist.Assoc.* 84 (405) (1989) 65–175.
- [17] C.J. Liu, H. Wechsler, Robust coding schemes for indexing and retrieval from large face databases, *IEEE Trans.Image Process.* 9 (1) (2000) 132–137.
- [18] Z.S. Hu, Z. Lou, J.Y. Yang, K. Liu, C.Y. Suen, Handwritten digit recognition basis on multi-classifier combination, *Chinese J. Comput.* 22 (4) (1999) 369–374.
- [19] H.Yoshihiko, et al., Recognition of handwriting numerals using Gabor features, *Proceedings of the Thirteenth ICPR*, pp.250–253.
- [20] S.X. Liao, M. Pawlak, On image analysis by moments, *IEEE Trans.Pattern Anal.Mach.Intell.* 18 (3) (1996) 254–266.
- [21] R.R. Bailey, S. Mandyam, Orthogonal moment feature for use with parametric and nonparametric classifiers, *IEEE Trans.Pattern Anal.Mach.Intell.* 18 (4) (1996) 389–398.
- [22] K.Alireza, H. Yawhua, Invariant image recognition by Zernike moments, *IEEE Trans.Pattern Anal.Mach.Intell.* 12 (1990) 489–497.
- [23] Y.Xu, J.-Y.Yang, Z. Jin, Fisher判别分析的一种新方法, 模式识别37 (2) (2004) 381–384.
- [24] M.Turk, A. Pentland, Eigenfaces for recognition, *J. Cognitive Neurosci.* 3 (1991) 71–86.
- [25] 洪志强, 用于识别的图像代数特征提取, 模式识别24 (3) (1991) 211–219.
- [26] Z.Jin, J.-Y.Yang, Z.-M.Tang, Z.-S.胡, 关于不相关最佳判别向量的定理, 模式识别34 (7) (2001) 2041–2047.
- [27] Z.Jin, J.Y. Yang, Z.S. Hu, 基于不相关判别变换的人脸识别, 模式识别34 (7) (2001) 1405–1416.

作者简介-

孙泉森是暨南大学数学系的副教授。同时，他正在攻读南京科技大学模式识别与智能系统专业的博士学位。他目前的兴趣包括模式识别、图像处理、计算机视觉和数据融合。

作者简介：曾胜根，2004年获南京科技大学计算机专业博士学位。他目前的兴趣包括模式识别、图像处理和遥感图像。

作者简介：刘彦，2004年在中国南京科技大学获得计算机专业硕士学位。他目前的兴趣包括模式识别、图像处理。

作者简介：彭安恒，1992年在美国印第安纳大学获得计算机科学博士学位。他现在是香港中文大学计算机科学与工程系的教授。他是香港中文大学虚拟现实、可视化和成像研究中心的主任。他的研究兴趣包括虚拟现实在医学中的应用，科学可视化，三维医学成像，用户界面，渲染和建模，互动图形和动画。

作者简介：夏德胜，1987年在法国鲁昂大学获得模式识别与智能系统博士学位。他现在是法国ESIGELEC公司的名誉教授，也是国科大计算机系的教授和博士生导师。他是国家科技大学图像处理、分析和识别实验室的主任。他的研究兴趣在图像处理、遥感、医学图像分析和模式识别领域。