

# 小样本情况下 Fisher 线性鉴别分析的理论及其验证

陈伏兵<sup>1),2)</sup> 张生亮<sup>1)</sup> 高秀梅<sup>1),2)</sup> 杨静宇<sup>1)</sup>

<sup>1)</sup>(南京理工大学计算机科学系,南京 210094) <sup>2)</sup>(淮阴师范学院数学系,淮安 223001)

**摘要** 线性鉴别分析是特征抽取中最为经典和广泛使用的方法之一。近几年,在小样本情况下如何抽取 Fisher 最优鉴别特征一直是许多研究者关心的问题。本文应用投影变换和同构变换的原理,从理论上解决了小样本情况下最优鉴别矢量的求解问题,即最优鉴别矢量可在一个低维空间里求得;给出了特征抽取模型,并给出求解模型的 PPCA + LDA 算法;在 ORL 人脸库 3 种分辨率灰度图像上进行实验。实验结果表明,PPCA + LDA 算法抽取的鉴别向量有较强的特征抽取能力,在普通的最小距离分类器下能达到较高的正确识别率,而且识别结果十分稳定。

**关键词** 小样本问题 主成分分析 线性鉴别分析 压缩变换 人脸识别

**中图分类号**: TP391 **文献标识码**: A **文章编号**: 1006-8961(2005)08-0984-08

## Theory of Fisher Linear Discriminant Analysis for Small Sample Size Problem and Its Verification

CHEN Fu-bing<sup>1),2)</sup>, ZHANG Sheng-liang<sup>1)</sup>, GAO Xiu-mei<sup>1),2)</sup>, YANG Jing-yu<sup>1)</sup>

<sup>1)</sup>(Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094)

<sup>2)</sup>(Department of Mathematics, Huaiyin Teachers College, Huaian 223001)

**Abstract** Linear discriminant analysis is one of the classical and popular methods used for feature extraction. In recent years many researchers have been absorbed in the problem of how to extract the optimal Fisher discriminant feature in small sample size case. By making use of the principle of projection transformation and isomorphic transformation, in this paper, we have solved the problem of how to gain the optimal discriminant vectors in the singular case. In fact these optimal discriminant vectors can be derived from a low dimension transformed subspace. Fulfilling the need of application, a new model for feature extraction has been put forward and a corresponding algorithm, called PPCA + LDA in this paper, has been established. The experiments on three kinds of resolution grayscale image for ORL face image database have been performed. The results of experiments show that the set of the discriminant vectors extracted by the proposed algorithm has powerful ability of feature extraction and the recognition results are very robust by the general minimum distance classifier.

**Keywords** small sample size problem, PCA (principal component analysis), LDA (linear discriminant analysis), compressed transformation, face recognition

## 1 引言

特征抽取是模式识别研究的关键问题,它的基本任务是找出对分类最有效的特征<sup>[1]</sup>。通常需要

一个定量的准则来衡量特征对分类的有效性。线性鉴别分析是特征抽取中最为经典和广泛使用的方法之一。

关于线性鉴别分析的研究应追溯到 Fisher 在 1936 年发表的经典论文<sup>[2]</sup>,其基本思想是选择使得

**基金项目**:国家自然科学基金项目(60472060)

**收稿日期**:2004-11-01; **改回日期**:2005-01-19

**第一作者简介**:陈伏兵(1964 ~ ),男,副教授。现为南京理工大学模式识别与智能系统专业博士研究生。主要研究方向为时间序列、模式识别、图像处理等。E-mail:cfb631123@sina.com

Fisher 准则函数达到极值的向量作为最佳投影方向,从而使得样本在该方向上投影后,达到最大的类间离散度和最小的类内离散度。在 Fisher 思想的基础上,Wilks<sup>[3]</sup>和 Duda<sup>[4,5]</sup>分别提出了鉴别矢量集的概念,即寻找一组鉴别矢量构成子空间,以原始样本在该子空间内的投影矢量作为鉴别特征用于识别。1970 年 Sammon 提出了基于 Fisher 鉴别准则的最佳鉴别平面<sup>[6]</sup>的概念,随后,Foley 和 Sammon<sup>[7]</sup>进一步提出了采用一组满足正交条件的最佳鉴别矢量集进行特征抽取的方法。1988 年 Duchene 和 Leclercq<sup>[8]</sup>给出了多类情况下最佳鉴别矢量集的计算公式。2001 年 Jin 和 Yang<sup>[9,10]</sup>从统计不相关的角度,提出了具有统计不相关性的最优鉴别矢量集的概念,与 F-S 鉴别矢量集不同的是,具有统计不相关性的最优鉴别矢量是满足共轭正交条件的,该方法被称为不相关的鉴别分析或 Jin-Yang 线性鉴别法。

以上提到的各种方法仅适用于类内散布矩阵非奇异(可逆)的情形,但实际应用中存在着大量的典型的小样本问题,比如在人脸图像识别问题中,类内散布矩阵经常是奇异的。这是因为待识别的图像矢量的维数一般较高,而在实际问题中难以找到或根本不可能找到足够多的训练样本来保证类内散布矩阵的可逆性。因此,在小样本情况下,如何抽取 Fisher 最优鉴别特征成为一个公认的难题<sup>[11-16]</sup>。近几年来关于小样本情况下线性鉴别分析方法的研究激起了人们的广泛兴趣,相继提出不少解决该类问题的方法。概括起来,这些方法可分为以下两类:

(1)从模式样本出发,即在模式识别之前,通过降低模式样本特征向量的维数达到消除奇异性的目的。基于这一思想的处理方法可分为两种:一是利用变换降维,典型的代表是 Eigenfaces 方法<sup>[17]</sup>和增强 Fisher 线性鉴别模型法<sup>[18]</sup>;二是通过降低图像的分辨率实现降维<sup>[9]</sup>。第 1 种方法保留主分量上的投影信息,抛弃了次分量上的信息;第 2 种方法无疑会丢失图像的某些细节信息。即这两种降维方法虽然可以消除奇异性,但都是以鉴别信息的损失为代价的,因此无法保证所抽取的特征是最优的。

(2)从算法本身入手,通过发展直接针对于小样本问题的算法来解决问题<sup>[12-16,19]</sup>。Hong 等人提出的扰动法<sup>[19]</sup>是一个近似算法,其基本思想是,当类内散布矩阵奇异时,通过对之进行一个小的扰动,使得扰动后的矩阵变为非奇异的,以扰动后的矩阵代替原来的类内散布矩阵进行鉴别矢量的求解,从

而将问题转化为可逆的情形加以解决。与之不同的是,Liu 给出了一个精确算法,称为正交补空间法<sup>[12]</sup>。正交补空间法的弱点在于,每求解一个最优鉴别矢量,都需要回到原始样本空间内构造子空间的正交补空间,当原始样本空间的维数很大时,这一过程是异常耗时的。Chen 等人提出的零空间法<sup>[13]</sup>是在类内散布矩阵的零空间内寻找极大化类间散布量的一组标准正交的特征向量并将其作为投影轴。该方法抛弃了类内散布矩阵零空间之外的鉴别信息;类内散布矩阵的零空间之外的空间中仍然可能包含有效的鉴别信息;Yu 等人提出了一种直接的线性鉴别分析方法<sup>[14]</sup>,此方法利用了类内散布矩阵的零空间包含的鉴别信息,但同时又舍弃了类间散布矩阵的零空间;Yang 等人提出了压缩变换方法<sup>[15]</sup>。在不损失任何有效鉴别信息的前提下,求解最优鉴别矢量只需要在低维的欧氏空间内进行。在以上这些算法中,除了 Yang 的压缩变换法,其他算法理论都存在着一个共同的弱点,那就是需要在原始维数空间中求最优鉴别矢量集,从计算上来看不适合处理高维空间中的问题。

基于广义的 Fisher 线性判别准则,Yang 等人在文献<sup>[15]</sup>,<sup>[16]</sup>中提出压缩变换的方法,其本质是奇异情况下 Fisher 最优鉴别特征的抽取过程可分为两步进行:第 1 步,利用 K-L 变换,以总体散布矩阵为产生矩阵,将高维的原始样本压缩为  $m$  维( $m$  表示总体散布矩阵的秩);第 2 步,在变换空间内,利用 Fisher 鉴别变换进行特征抽取。虽然  $m$  比原始空间维数小很多,但通常情况下, $m$  仍然会很大,而实际上,并非产生矩阵所有特征向量都有很大的保留意义。为此,秉承文献<sup>[15]</sup>,<sup>[16]</sup>变换的思想,从算法入手,对小样本情况下的最优鉴别矢量求解的问题进行了进一步的研究,给出了特征抽取的新模型,并提供了求解模型的相应算法。称为 PPCA + LDA (Projected PCA plus LDA) 方法。在 ORL 人脸库 3 种分辨率灰度图像上进行实验的结果表明,PPCA + LDA 算法抽取的鉴别向量有较强的特征抽取能力,从而完善了小样本情况下 Fisher 线性鉴别分析的理论。

## 2 广义 Fisher 线性鉴别分析

### 2.1 散布矩阵概念及其性质

设  $\omega_1, \omega_2, \dots, \omega_c$  为  $C$  个模式类,模式  $X$  为  $n$  维

实向量。类间散布矩阵  $S_b$ 、类内散布矩阵  $S_w$  和总体散布矩阵  $S_t$  分别定义为

$$S_b = \sum_{i=1}^c P(\omega_i) (m_i - m_0)(m_i - m_0)^T \quad (1)$$

$$S_w = \sum_{i=1}^c P(\omega_i) E\{(X - m_i)(X - m_i)^T / \omega_i\} \quad (2)$$

$$S_t = S_b + S_w = E\{(X - m_0)(X - m_0)^T\} \quad (3)$$

其中,  $P(\omega_i)$  为第  $i$  类训练样本的先验概率,  $m_i = E\{X/\omega_i\}$  为第  $i$  类训练样本的均值,  $m_0 = E\{X\} =$

$\sum_{i=1}^m P(\omega_i) m_i$  为全体训练样本的均值。

由式(1)、式(2)和式(3)的定义知,  $S_w, S_b, S_t$  均为非负定矩阵, 而且满足  $S_t = S_b + S_w$ 。当  $S_w$  可逆时,  $S_w$  与  $S_t$  均为正定矩阵。

引理 1<sup>[13]</sup>  $S_w, S_b$  和  $S_t$  的秩有以下性质:

- (1)  $rank(S_w) \leq \min(n, N - C)$
- (2)  $rank(S_b) \leq \min(n, C - 1)$
- (3)  $rank(S_t) \leq \min(n, N - 1)$

其中,  $N$  是训练样本总数。

### 2.2 准则函数及其方法

Fisher 准则函数:

$$J_f(\xi) = \frac{\xi^T S_b \xi}{\xi^T S_w \xi} \quad \xi \neq 0 \quad (4)$$

广义的 Fisher 准则函数:

$$J(\xi) = \frac{\xi^T S_b \xi}{\xi^T S_t \xi} \quad \xi \neq 0 \quad (5)$$

当类内散布矩阵  $S_w$  非奇异时, 式(4)、式(5)两准则完全等价; 当  $S_w$  奇异时, 准则式(5)是准则式(4)的合理拓广<sup>[12]</sup>。这里采用准则式(5)。

首先, 给出 F-S 线性鉴别和 J-Y 线性鉴别的统一描述。

F-S 最优鉴别矢量集是满足以下正交条件且使得准则函数式(4)、式(5)达到极值的一组矢量  $\xi_1, \xi_2, \dots, \xi_d$ :

$$\xi_j^T \xi_i = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad i, j = 1, \dots, d \quad (6)$$

具体地讲, 该最优鉴别矢量集的第 1 个鉴别矢量  $\xi_1$  取为 Fisher 最优投影方向; 当前  $i$  个鉴别矢量  $\xi_1, \xi_2, \dots, \xi_i$  取定后, 第  $i+1$  个鉴别矢量可由求解下面模型 1 最优化问题得到。

模型 1  $\max(J(\xi)), (\xi_j, \xi) = 0, j = 1, \dots, i, \xi \in R^n$

其中,  $(\xi_j, \xi) = \xi_j^T \xi$ 。

J-Y 最优鉴别矢量集  $\xi_1, \xi_2, \dots, \xi_d$  满足以下共轭正交条件

$$\xi_j^T S_t \xi_i = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad i, j = 1, \dots, d \quad (7)$$

J-Y 最优鉴别矢量集的第 1 个鉴别矢量  $\xi_1$  取为 Fisher 最优投影方向; 当前  $i$  个鉴别矢量  $\xi_1, \dots, \xi_i$  取定后, 第  $i+1$  个鉴别矢量仍可利用模型 1 确定, 不过此时模型中的内积定义为  $(\xi_j, \xi) = \xi_j^T S_t \xi$ 。

当总体散布矩阵  $S_t$  可逆时, 文献[9], [10]给出了 J-Y 最优鉴别矢量集的具体算法。

## 3 奇异情况下线性鉴别分析

### 3.1 投影变换原理

定义 1(投影变换) 设  $V_1, V_2$  是欧氏空间  $R^n$  的两个子空间,  $R^n = V_1 \oplus V_2, \xi \in R^n, \xi = X + \eta, X \in V_1, \eta \in V_2$ , 规定对应法则  $L: \xi \rightarrow X$ , 此对应法则  $L$  称为欧氏空间  $R^n$  到它的子空间  $V_1$  的投影变换。

容易证明投影变换具有以下性质。

定理 1 投影变换  $L$  是线性变换。

因为  $S_t$  是实对称矩阵, 所以  $S_t$  必有  $n$  个标准正交的特征向量, 分别表示为  $\beta_1, \beta_2, \dots, \beta_m, \dots, \beta_n$ , 其中,  $m = rank(S_t), \beta_1, \dots, \beta_m$  为  $S_t$  的非零特征值所对应的特征向量,  $\beta_{m+1}, \dots, \beta_n$  为  $S_t$  的零特征值所对应的特征向量。令  $V_{S_t}$  表示  $S_t$  的非零特征值所对应的特征子空间, 即  $V_{S_t} = span\{\beta_1, \dots, \beta_m\}$ , 则  $V_{S_t}^\perp = span\{\beta_{m+1}, \dots, \beta_n\}$ , 且  $R^n = V_{S_t} \oplus V_{S_t}^\perp$ 。

对于任意  $\xi \in R^n$ , 令  $\xi = X + \eta$ , 其中,  $X \in V_{S_t}, \eta \in V_{S_t}^\perp$ , 规定  $R^n$  到它的子空间  $V_{S_t}$  的投影映射  $L: R^n \rightarrow V_{S_t}, \xi \rightarrow X$ , 则在此投影映射下有以下结论。

定理 2<sup>[15]</sup> (投影变换原理)  $J_f(\xi) = J_f(X), J(\xi) = J(X)$ 。

根据投影变换原理, 模型 1 等价于以下模型 2。

模型 2  $\max J(X), (X_j, X) = 0, j = 1, \dots, i, X \in V_{S_t}$ 。

### 3.2 同构变换原理

引理 2<sup>[20]</sup> 有限维线性空间同构的充分必要条件是它们有相同的维数。

由于  $\dim V_{S_t} = m$ , 根据引理 2,  $V_{S_t}$  同构于  $m$  维欧氏空间  $R^m$ , 相应的同构变换可定义为

$$X = PY \quad (8)$$

其中,  $P = (\beta_1, \beta_2, \dots, \beta_m), Y \in R^m$

该变换是从  $R^m$  到  $V_s$  上的一一变换。

在同构变换  $X = PY$  下, 准则函数  $J(X)$  变为

$$J(X) = \frac{Y^T (P^T S_b P) Y}{Y^T (P^T S_1 P) Y}$$

定义函数

$$\tilde{J}(Y) = \frac{Y^T \tilde{S}_b Y}{Y^T \tilde{S}_1 Y} \quad (9)$$

其中,  $\tilde{S}_b = P^T S_b P, \tilde{S}_1 = P^T S_1 P$ 。

易证明  $\tilde{S}_b, \tilde{S}_1$  均为  $m$  阶非负定阵, 故  $\tilde{J}(Y)$  可视为类似于  $J(X)$  的一个准则函数。此外, 由于  $\tilde{S}_1$  是可逆矩阵, 故  $\tilde{S}_1$  是正定矩阵。

同构映射  $X = PY$  具有以下性质:

**定理 3<sup>[15]</sup>** (同构变换原理) 设  $X = PY$  是  $Y$ -空间到  $X$ -空间的同构变换, 且  $X^* = PY^*$ , 则  $X^*$  是准则函数  $J(X)$  极值点的充分必要条件是  $Y^*$  是准则函数  $\tilde{J}(Y)$  的极值点。

**定理 4<sup>[15]</sup>** 设矩阵  $P = (\beta_1, \beta_2, \dots, \beta_m), X_i = PY_i, X_j = PY_j$ , 则

- (1)  $X_i$  与  $X_j$  正交当且仅当  $Y_i$  与  $Y_j$  正交;
- (2)  $X_i$  与  $X_j$  关于  $S_1$  共轭正交当且仅当  $Y_i$  与  $Y_j$  关于  $\tilde{S}_1$  共轭正交。

在同构变换  $X = PY$  下, 模型 2 变换为以下模型 3。

**模型 3**  $\max \tilde{J}(Y), (Y_j, Y) = 0, j = 1, \dots, i, Y \in R^m$

根据定理 3、定理 4 不难得出以下结论:

**推论 1** 设  $Y_1, \dots, Y_d (d \leq m)$  为模型 3 (令  $i = 1, \dots, d - 1$ ) 的最优解, 则  $X_1 = PY_1, \dots, X_d = PY_d$  为模型 2 的最优解。

根据矩阵理论中广义特征值性质<sup>[21]</sup>, 模型 3 是可解的。

以上解决了模型可解的问题, 求解模型 3 的目的是想通过式 (8) 获得模型 2 优化问题的解, 在模式识别中就是要获得最优鉴别矢量。但在模式识别中, 由于特征的相关性, 太多的鉴别特征反而会使正确识别率下降, 所以有必要进一步降低特征向量的维数, 减少计算量, 提高识别效率。压缩变换的思想满足了这个需求。

### 3.3 压缩变换

**定义 2** 设  $Q \in R^{m \times s}$ , 且  $Q^T Q = E$  (单位矩阵), 则线性变换  $Z = Q^T Y$  称为压缩变换。

利用压缩变换  $Y = QZ$  (即  $Z = Q^T Y$ ), 对模型 3 进行变换, 使其转化为模型 4。

**模型 4**  $\max \hat{J}(Z), (Z_j, Z) = 0, j = 1, \dots, i, Z \in V$

这里  $\hat{J}(Z) = \frac{Z^T \hat{S}_b Z}{Z^T \hat{S}_1 Z}, \hat{S}_b = Q^T \tilde{S}_b Q, \hat{S}_1 = Q^T \tilde{S}_1 Q, V$  是  $Q$  的列空间。由于  $\hat{S}_1$  是可逆矩阵, 故  $\hat{S}_1$  是正定矩阵。

与模型 3 类似, 模型 4 也是可解的。根据定理 3 和定理 4, 模型 4 的解具有以下性质:

**定理 5** 设  $Z_1, Z_2, \dots, Z_d$  为模型 4 (令  $i = 1, \dots, d - 1$ ) 的最优解, 则  $Y_1 = QZ_1, \dots, Y_d = QZ_d$  为模型 3 的最优解。

由上面的定理 2、定理 3、定理 4、推论 1 和定理 5, 建立了奇异情况下求解基于准则式 (5) 的最优鉴别矢量集理论, 即首先进行投影变换, 再进行同构变换, 最后施行压缩变换并求解压缩变换后的模型 4。压缩变换实质是利用 PCA 方法对投影变换后的模式进一步降维, 以便加快处理速度或降低问题复杂度, 解模型 4 实质是利用 LDA 方法进行鉴别特征抽取, 以便获得鉴别矢量, 因此, 称这种求得最优鉴别集的方法为 PPCA + LDA (Projected PCA plus LDA) 方法。需要指出的是, 在进行压缩变换  $Z = Q^T Y$  时, 系数矩阵  $Q$  的选取有一定的灵活性, 视具体情况和需要而定。基本原则是  $Q$  应选择与样本信息有关的矩阵。理论上, 当  $Q$  的列空间的维数等于  $m$  时, 模型 4 与模型 3 是等价的, 但这种变换没有任何意义, 因此, 一般地, 取  $Q$  的列空间的维数小于  $m$ 。

### 3.4 PPCA + LDA 算法

根据同构变换与压缩变换的理论, PPCA + LDA 算法可描述如下:

- (1) 计算训练样本的总体散布矩阵  $S_t = \frac{1}{N} \sum_{i=1}^c \sum_j^{l_i} (x_{ij} - m_0)(x_{ij} - m_0)^T$ ;
- (2) 计算  $S_t$  的非零特征值对应的特征向量  $\beta_1, \dots, \beta_m$ , 这里  $m = rank(S_t)$ ; 得到  $R^n$  的子空间  $V_s = span\{\beta_1, \beta_2, \dots, \beta_m\}$ , 利用投影变换原理, 将模型 1 的求解问题转化为解模型 2;
- (3) 进行同构变换  $X = PY$ , 模型 2 的求解问题转化为解模型 3;
- (4) 进行压缩变换  $Z = Q^T Y$ , 模型 3 的求解问题转化为解模型 4;
- (5) 求得模型 4 的最优解:  $Z_1, Z_2, \dots, Z_d \in V, (d \leq s)$ , 从而获得模型 2 的最优解  $PQ(Z_1, Z_2, \dots, Z_d)$ 。

上述算法步骤如图 1 所示。

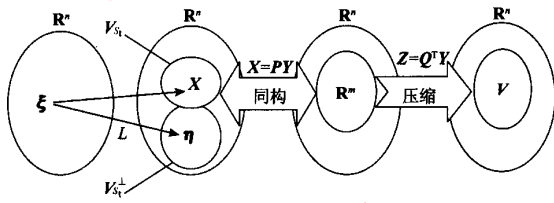


图 1 PPCA + LDA 法求解广义 Fisher 鉴别准则最优鉴别矢量一般求解过程

Fig. 1 Illustration of the process of finding the optimal discriminant vectors based on the generalized Fisher discriminant criterion for PPCA + LDA algorithm

PPCA + LDA 的算法可通过 F-S 最优鉴别法, J-Y 最优鉴别法和 Yang 的组合最优鉴别法解模型 4 实现。

总的说来,与 Liu<sup>[12]</sup>, Chen<sup>[13]</sup>, Yu<sup>[14]</sup>, Hong<sup>[19]</sup>, Yang<sup>[15]</sup> 等人的方法相比,PPCA + LDA 法求解最优鉴别矢量集的思想具有明显的优势,即最优鉴别矢量集的计算只需要在  $s$  维空间  $V$  内进行。根据引理 1 性质 3,由于  $s \leq m \leq N - 1$ ,而训练样本数  $N$  远远小于原始样本特征的维数  $n$ ,故本文方法极大地降低了计算量,提高了求解速度。以 ORL 人脸图像库为例,图像的分辨率为  $92 \times 112$ ,图像总数为 400 幅,一般地,训练样本数取为 200。在该情况下, Liu<sup>[12]</sup>, Chen<sup>[13]</sup>, Yu<sup>[14]</sup>, Hong<sup>[19]</sup> 的方法需要在  $92 \times 112 = 10\ 304$  维的空间内进行, Yang<sup>[15]</sup> 的方法需要在 199 维的空间内进行,这必然耗费大量的计算时间,而本文算法在 30 维的空间内进行获得的结果已相当满意。

### 4 试验结果与分析

利用 ORL 人脸图像数据库进行人脸识别试验。ORL 人脸图像数据库共由 40 人每人 10 幅图像组成。实验分别对分辨率为  $92 \times 112$ ,  $46 \times 56$  和  $23 \times 28$  的灰度图像进行。以每人的前 5 幅图像作为训练样本,后 5 幅作为测试样本,这样训练样本和测试样本总数均为 200。因此本实验属于高维小样本问题。

首先计算出原始样本的类间散布矩阵、类内散布矩阵和总体散布矩阵的秩,分别为 39, 160 和 199。3 种分辨率情况下类间散布矩阵、类内散布矩阵和总体散布矩阵都是奇异的。采用 PCA 方法,分别求出总体散布矩阵的非零特征值所对应的 199 个标准正交的特征向量,用这些特征向量构成投影变换矩阵  $P$ ,进行投影变换  $Y = P^T X$ ,分别将 10 304 维、

2 576 维和 644 维的人脸图像矢量压缩为 199 维。在变换后的 199 维空间中,总体散布矩阵  $\tilde{S}_t$  是正定的,但类内散布矩阵  $\tilde{S}_w$  仍然是奇异的。采用 F-S 鉴别法、J-Y 鉴别法和组合鉴别法分别解模型 3,抽取鉴别特征,然后,求出  $\tilde{S}_t$ ,  $\tilde{S}_w$  的特征值。将特征值由大到小排序,对应的特征向量也相应排序,取如下两种压缩变换,一是  $Q$  取  $\tilde{S}_t$  的前 30 个特征向量为列(这时鉴别特征抽取是在 30 维的空间内进行),二是  $Q$  取  $\tilde{S}_w$  的前 39 个特征向量为列(这时鉴别特征抽取是在 39 维的空间内进行),再相应的用 PPCA + LDA 方法分别解模型 4。对得到的最优鉴别矢量集,用最小距离分类器进行分类,实验结果如表 1 和表 2 所示。

表 1 是  $Q$  取  $\tilde{S}_t$  的前 30 个特征向量为列构成压缩变换时用 F-S 鉴别法、J-Y 鉴别法和组合鉴别法分别解模型 3 和对应的用 PPCA + LDA 方法解模型 4(对 3 种分辨率 ORL 图像)得到的最优鉴别矢量集识别结果,表中的数据,如 69(20),括号外的数表示解模型 3 的误识结果,括号内的数表示解模型 4 的误识结果;最后一行的数据表示各维误识结果均值。从表 1,总体上可以看出 PPCA + LDA 方法的结果优于其他方法的结果。具体地,PPCA + LDA 方法的结果明显好于 F-S 鉴别法和 J-Y 鉴别法的结果,平均误识率低;PPCA + LDA 方法的结果与组合鉴别法结果几乎相当。这是由于,一方面该情况下模型 4 利用了  $\tilde{S}_w$  的鉴别信息,模型 3 没有利用  $\tilde{S}_w$  的鉴别信息,另一方面  $\tilde{S}_t$  的前 30 个非零特征值(它们的和占有所有特征值之和的 97.8%)对应的特征向量包含了  $\tilde{S}_t$  的绝大部分鉴别信息,选取这些特征向量构成压缩矩阵丢失的鉴别信息比降低抽取特征的相关性释放出的鉴别信息要小得多。

对 ORL 人脸库的 3 种分辨率灰度图像进行实验,还发现,随着图像分辨率的降低,识别的正确率不仅没有降低反而会提高。如组合鉴别法得到的正确识别率,对模型 4 的结果(即 PPCA + LDA 方法的结果)来说,  $92 \times 112$  的最好结果是 94.5% (维数取 16),  $46 \times 56$  的最好结果是 95% (维数取 16),  $23 \times 28$  的最好结果是 97% (维数取 15)。

关于表 1 的实验所选取的压缩变换没有使组合鉴别法的优点得到充分发挥,因为这时  $\tilde{S}_w$  是非奇异的,本质上这时组合鉴别法等价于统计不相关方法,但是,由于组合鉴别法利用了  $\tilde{S}_w$  非零特征值对应的特征向量提供的鉴别信息,从而导致其识别结果优

表 1 3 种分辨率 ORL 人脸识别错误数对照表 ( $Q$  取  $\tilde{S}_b$  的前 30 个特征向量)

Tab.1 Comparison of the error recognition results of three kind resolutions for ORL face image database ( $Q$  consists of the first 30 eigenvectors of  $\tilde{S}_b$ )

维数	F-S 鉴别法 (PPCA + LDA 方法)			J-Y 鉴别法 (PPCA + LDA 方法)			组合鉴别 (PPCA + LDA 方法)		
	92 × 112	46 × 56	23 × 28	92 × 112	46 × 56	23 × 28	92 × 112	46 × 56	23 × 28
11	69(20)	67(22)	64(19)	86(21)	63(21)	63(13)	19(19)	18(17)	14(15)
12	60(15)	62(16)	60(14)	80(18)	60(15)	54(11)	15(15)	15(14)	13(13)
13	60(16)	59(14)	57(14)	76(17)	59(15)	46(9)	16(16)	14(17)	16(10)
14	57(18)	46(14)	53(14)	76(16)	61(13)	44(8)	16(13)	14(13)	15(9)
15	52(16)	43(15)	52(15)	69(17)	59(14)	43(10)	16(13)	16(13)	15(6)
16	42(15)	31(14)	45(17)	62(13)	52(12)	42(10)	16(11)	16(10)	14(6)
17	31(16)	31(16)	41(17)	59(17)	54(14)	43(11)	14(12)	13(12)	14(9)
18	31(15)	31(17)	40(16)	55(17)	51(14)	39(12)	16(12)	12(14)	14(9)
19	26(16)	28(15)	35(15)	53(16)	54(15)	37(11)	13(14)	11(13)	13(8)
20	22(13)	22(15)	30(15)	45(17)	47(11)	36(11)	12(14)	10(13)	12(8)
21	17(14)	22(14)	23(13)	38(17)	49(11)	33(12)	13(13)	11(11)	13(9)
22	21(14)	21(14)	23(13)	36(20)	44(14)	29(11)	11(13)	11(12)	12(8)
23	20(13)	19(13)	21(12)	35(16)	43(14)	31(11)	12(14)	11(12)	10(9)
24	18(15)	17(16)	22(14)	33(15)	38(12)	29(9)	11(11)	11(11)	11(8)
25	15(13)	16(14)	20(14)	30(20)	38(15)	28(9)	12(16)	11(14)	11(9)
26	13(12)	17(13)	17(15)	33(19)	32(15)	28(9)	13(15)	9(13)	10(9)
27	17(15)	16(16)	15(17)	27(17)	31(13)	29(8)	10(14)	10(14)	10(9)
28	14(20)	20(18)	15(19)	27(17)	30(15)	28(8)	8(16)	9(15)	10(9)
29	14(18)	19(18)	11(19)	29(18)	28(15)	28(9)	8(18)	9(16)	10(10)
30	15(23)	16(22)	11(22)	27(23)	29(18)	27(15)	8(21)	8(19)	10(14)
均值	30.7(15.9)	30.2(15.8)	32.8(15.7)	48.8(17.6)	46.1(14.3)	36.9(10.4)	13(14.5)	12(13.7)	12.45(9.4)

注:识别结果中,如 69(20),括号外(内)的数表示 F-S 鉴别法 (PPCA + LDA 方法) 的误识结果;最后一行的数表示各维误识结果均值;待识别样本数为 200。

表 2 3 种分辨率 ORL 人脸识别错误数对照表 ( $Q$  取  $\tilde{S}_b$  的前 39 个特征向量)

Tab.2 Comparison of the error recognition results of three kind resolutions for ORL face image database ( $Q$  consists of the first 39 eigenvectors of  $\tilde{S}_b$ )

维数	F-S 鉴别法 (PPCA + LDA 方法)			J-Y 鉴别法 (PPCA + LDA 方法)			组合鉴别 (PPCA + LDA 方法)		
	92 × 112	46 × 56	23 × 28	92 × 112	46 × 56	23 × 28	92 × 112	46 × 56	23 × 28
20	22(17)	22(17)	30(18)	45(15)	47(13)	36(11)	12(13)	10(15)	12(11)
21	17(16)	22(17)	23(15)	38(13)	49(14)	33(11)	13(14)	11(15)	13(10)
22	21(18)	21(17)	23(17)	36(13)	44(14)	29(10)	11(13)	11(13)	12(9)
23	20(19)	19(16)	21(17)	35(12)	43(14)	31(10)	12(14)	11(14)	10(9)
24	18(17)	17(18)	22(13)	33(13)	38(15)	29(10)	11(14)	11(14)	10(10)
25	15(22)	16(20)	20(17)	30(13)	38(11)	28(10)	12(14)	11(12)	11(11)
26	13(18)	17(18)	17(16)	33(19)	32(13)	28(9)	13(15)	9(14)	10(11)
27	17(19)	16(18)	15(16)	27(15)	31(12)	29(9)	10(13)	10(11)	10(11)
28	14(19)	20(18)	15(17)	27(14)	30(11)	28(9)	8(13)	9(12)	10(10)
29	14(18)	19(18)	11(18)	29(13)	28(12)	28(12)	8(15)	9(11)	10(10)
30	15(21)	16(19)	11(19)	27(13)	29(13)	27(10)	8(12)	8(12)	10(10)
31	12(23)	16(22)	10(20)	26(14)	24(13)	24(10)	8(13)	7(12)	10(10)
32	11(23)	16(21)	10(21)	20(16)	22(12)	25(10)	7(14)	7(11)	10(10)
33	13(20)	15(20)	11(21)	21(15)	22(13)	25(10)	7(14)	6(11)	10(9)
34	13(21)	12(20)	9(21)	19(16)	21(13)	27(10)	7(16)	6(12)	10(9)
35	12(22)	11(20)	8(21)	19(18)	22(13)	23(10)	8(17)	6(14)	8(10)
36	11(23)	12(21)	9(21)	20(18)	20(13)	21(10)	8(17)	7(14)	8(11)
37	9(22)	13(23)	8(21)	20(17)	17(13)	20(10)	7(17)	7(14)	8(11)
38	9(25)	10(24)	11(23)	21(17)	19(14)	20(10)	7(17)	7(14)	8(10)
39	8(25)	7(24)	9(24)	20(17)	19(14)	21(10)	8(17)	7(14)	9(10)
均值	14.2(20.4)	15.9(19.6)	14.7(19)	27.3(15.1)	29.8(13)	26.6(10.1)	9.3(14.6)	8.5(13)	10(10.2)

注:识别结果中,如 22(17),括号外(内)的数表示 F-S 鉴别法 (PPCA + LDA 方法) 的误识结果;最后一行的数表示各维误识结果均值;待识别样本数为 200。

于统计不相关方法识别结果。这与文献[14]的结果一致,说明  $\tilde{S}_w$  的零空间之外的空间中仍然包含有效的鉴别信息。

表 2 是  $Q$  取  $\tilde{S}_1$  的前 39 个特征向量为列构成压缩变换时用 F-S 鉴别法、J-Y 鉴别法和组合鉴别法分别解模型 3 和用 PPCA + LDA 方法解模型 4(对 3 种分辨率 ORL 图像)得到的最优鉴别矢量集识别结果。从表 2 可以看出,PPCA + LDA 方法的结果没有 F-S 鉴别法和组合鉴别法的结果好,这是由于一方面该情况下  $\tilde{S}_1$  没有为模型 4 提高更多的鉴别信息;另一方面  $\tilde{S}_1$  的前 39 个非零特征值,它们的和仅占有所有特征值之和的 82.6%,选取这些特征值对应的特征向量构成压缩矩阵丢失了  $\tilde{S}_1$  的部分信息。另外,还对  $Q$  取  $\tilde{S}_1$  的前 180 个特征向量(它们对应的特征值之和占有所有特征值之和的 99.46%)为列

构成压缩变换的情形进行了试验,这时,PPCA + LDA 方法的结果相当好。由此可见,压缩变换并非是随意选取的。

实验是使用 Matlab 语言编程在 Inter(R) 1.80 GHz、内存 384MB 计算机上进行的。表 3 是 3 种鉴别法解模型 3 与模型 4 获得最高正确识别率时最优鉴别矢量维数、误识数目、正确识别率以及特征抽取和识别花费的时间对照表。表中的数据是对分辨率为  $23 \times 28$  的图像进行试验得到的(模型 4 的结果参见表 1)。对 F-S 鉴别法来说,虽然正确识别率有所下降(仍能达到 94%),但由于维数的降低,减少了计算量,节约了计算时间;对 J-Y 鉴别法和组合鉴别法来说,不仅最优鉴别矢量的维数降低了,抽取鉴别特征的时间减少了,而且正确识别率也得到了提高。

表 3 模型 3 与模型 4 最优鉴别维数、误识结果和正确识别率对照表

Tab.3 Comparison of the optimal dimension, error recognition numbers, correct recognition ratio and CPU time for feature extraction and recognition for Model 3 and Model 4

鉴别法	模型 3				模型 4			
	维数	误识数目	正确识别率(%)	时间(s)	维数	误识数目	正确识别率(%)	时间(s)
F-S 鉴别法	48	7	96.5	31.66	23	12	94	4.08
J-Y 鉴别法	37	20	90	4.32	14	8	96	3.74
组合鉴别法	43	7	96.5	4.3	15	6	97	3.75

注:模型 4 的结果为 PPCA + LDA 方法的结果,待识别样本数为 200。

### 5 结 论

将投影变换、同构变换和压缩变换相结合,从理论上解决了小样本情况下广义的 Fisher 鉴别准则最优鉴别向量的求解问题,完善了小样本情况下线性鉴别分析理论。给出了有效实用算法。利用这个算法,基于推广的 Fisher 鉴别准则最优鉴别向量的计算只需要在低维空间里进行,减少了计算量,提高了求解的效率。由于给出的算法不是一个精确的算法,故提高算法的精度显得尤为重要,这可以通过选择合适的压缩变换来实现。本文选取  $\tilde{S}_w$  的前 30 个特征值对应的特征向量构成压缩变换系数矩阵,效果十分显著,能否通过其他途径找到更好的压缩变换,有待进一步研究。

### 参考文献 (References)

1 Bian Z Q, Zhang X G. Pattern recognition (2nd edition) [M]. Beijing: Tsinghua University Press, 1999: 176 ~ 177. [边肇祺, 张

学工. 模式识别(第二版)[M]. 北京: 清华大学出版社, 1999: 176 ~ 177. ]  
 2 Fisher R A. The use of multiple measurements in taxonomic problems [J]. Annals of Eugenics, 1936, 7: 179 ~ 188.  
 3 Wilks S S. Mathematical Statistics [M]. New York: Wiley, 1962: 577 ~ 578.  
 4 Duda Richard O, Hart Peter E, Stork David G. Pattern Classification [M]. Li H D, Yao T X, Beijing: China Machine Press, 2003: 94 ~ 98. [Duda Richard O, Hart Peter E, Stork David G. 模式分类 [M]. 李宏东, 姚天翔译. 北京: 机械工业出版社, 2003: 94 ~ 98. ]  
 5 Duda R, Hart P. Pattern Classification and Scene Analysis [M]. New York: Wiley, 1973: 113 ~ 120.  
 6 Sammon J W. An optimal discriminant plane [J]. IEEE Transactions on Computer, 1970, 19: 826 ~ 829.  
 7 Foley D H, Sammon J W Jr. An optimal set of discriminant vectors [J]. IEEE Transactions on Computer, 1975, 24(3): 281 ~ 289.  
 8 Duchene J, Leclercq S. An optimal transformation for discriminant and principal component analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988, 10(6): 978 ~ 983.  
 9 Jin Zhong, Yang J Y, Hu Z S, et al. Face recognition based on uncorrelated discriminant transformation [J]. Pattern Recognition, 2001, 34(7): 1405 ~ 1416.

- 10 Jin Z, Yang J Y, Tang Z M, *et al.* A theorem on uncorrelated optimal discriminant vectors [ J ]. *Pattern Recognition*, 2001, **34**(10):2041 ~ 2047.
- 11 Belhumeur Peter N, Hespanha Joao P, Kriegman David J, *et al.* Eigenfaces vs. Fisherfaces: recognition using class specific linear projection [ J ]. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 1997, **19**(7): 711 ~ 720.
- 12 Liu Ke, Cheng Y Q, Yang J Y, *et al.* An efficient algorithm for Foley-Sammon optimal set of discriminant vectors by algebraic method [ J ]. *International Journal of Pattern Recognition and Artificial Intelligence*, 1992, **6**(5): 817 ~ 829.
- 13 Chen Li-fen, Liao H-Y Mark, Ko M-T, *et al.* A new LDA-based face recognition system which can solve the small sample size problem [ J ]. *Pattern Recognition*, 2000, **33**(10): 1713 ~ 1726.
- 14 Yu Hua, Yang Jie. A direct LDA algorithm for high-dimensional data—with application to face recognition [ J ]. *Pattern Recognition*, 2001, **34**(10): 2067 ~ 2070.
- 15 Yang J, Yang J Y, Ye H, *et al.* Theory of fisher linear discriminant analysis and its application [ J ]. *Acta automatic Sinica*, 2003, **29**(4): 482 ~ 493. [ 杨键, 杨静宇, 叶晖等. Fisher 线性鉴别分析的理论研究及其应用 [ J ]. *自动化学报*, 2003, **29**(4): 482 ~ 493. ]
- 16 Yang Jian, Yang Jing-Yu. Why can LDA be performed in PCA transformed space? [ J ]. *Pattern Recognition*, 2003, **36**(2): 563 ~ 566.
- 17 Turk M, Pentland A. Eigenfaces for recognition [ J ]. *Journal of Cognitive Neuroscience*, 1991, **3**(1): 71 ~ 86.
- 18 Cheng Jun Liu, Harry Wechsler. A shape and texture based enhanced Fisher classifier for face recognition [ J ]. *IEEE Transactions on Image Processing*, 2001, **10**(4): 598 ~ 608.
- 19 Hong Z Q, Yang J Y, Optimal discriminant plane for a small number of samples and design method of classifier on the plane [ J ]. *Pattern Recognition*, 1991, **24**(4): 317 ~ 324.
- 20 Algebra Teaching and Research group of Geometry and Algebra Staff Room of Mathematics Department of Peking University. *Advanced Algebra(2nd)* [ M ]. Beijing: Higher Education Press, 1988: 272. [ 北京大学数学系几何与代数教研室代数小组编. 高等代数(第二版) [ M ]. 北京: 高等教育出版社, 1988: 272. ]
- 21 Xu Z, Zhang K Y, Lu Q, *et al.* *Matrix Theory* [ M ]. Beijing: Science Press, 2002: 140 ~ 143. [ 徐仲, 张凯院, 陆全等编著. 矩阵论简明教程 [ M ]. 北京: 科学出版社, 2002: 140 ~ 143. ]