

模糊标号典型相关分析及其在人脸识别中应用

苏志勋¹, 刘艳艳², 刘秀平^{*1}

(1. 大连理工大学 应用数学系, 辽宁 大连 116024;

2. 防灾科技学院 基础部, 河北 三河 065201)

摘要: 把样本分布信息融于特征提取过程将有助于提高特征的分类能力. 利用模糊隶属度概念, 提出一种基于模糊标号典型相关分析的特征提取新方法. 构造模糊标号刻画样本的分布情况, 并将其与典型相关分析结合, 能提取综合灰度信息和分布信息的有效判别特征. 此外, 针对样本不足导致的小特征值包含较多干扰信息的问题, 基于矩阵理论及双空间分析思想, 进一步提出双空间模糊标号典型相关分析算法, 缓解了过小特征值对算法性能的影响. 在 ORL 和组合人脸数据库上的实验结果表明新特征具有良好的分类能力, 证实了所提算法的有效性及应用价值.

关键词: 典型相关分析; 模糊隶属度; 小样本问题; 特征提取; 人脸识别

中图分类号: TP391.4 **文献标志码:** A

0 引言

典型相关分析(CCA)是研究两组随机变量之间相关关系的一种统计分析方法^[1,2], 近年来在模式识别等领域受到越来越多的关注^[3~6]. CCA 能同时对两组变量进行处理, 若取其中一组变量为样本的类别标号, CCA 可用做有监督特征提取, 特别地, 用只含 0、1 的二值向量表示样本所属类别, CCA 等价于经典特征提取方法 Fisher 线性判别分析(FLDA)^[5~7]. 但是, 由于图像样本受外界条件诸如光照、拍摄角度、噪声等的影响非常大, 各类图像间并不存在严格的分界线, 而是有重叠区域和离群样本的存在, 二值标号向量只是把样本指定为绝对属于或不属于某一类, 并没有考虑各类图像间的模糊联系, 显然丢失了有用的样本分布信息.

基于上述讨论, 本文引入模糊隶属度的思想, 提出一种新的模糊标号典型相关分析算法, 根据图像样本的分布特点定义隶属度函数, 用于构造描述样本分布信息的模糊标号, 并将其与 CCA 结合提取综合灰度信息与分布信息的特征, 该特征充分利用了样本的分布信息, 用于分类更具判

别能力. 本文称用二值向量表示类别的方法为标号 CCA, 记为 LCCA, 相应地, 将新方法记为 FLCCA.

在模式识别领域, 高维小样本问题常会导致样本的协方差矩阵奇异. 孙权森等^[4]利用主成分分析(PCA)对数据降维以保证协方差矩阵非奇异, 并证明 CCA 不会损失任何有效判别信息, 但此时需利用全部非零特征值对应的特征向量对数据进行 PCA 处理. 然而, 样本不足会使小特征值包含较多干扰信息^[8], 直接利用会使结果存在较大误差, 易导致过拟合. 双空间分析方法将特征空间划分为两部分进行处理以提高算法性能, 在 FLDA 中已有了成功应用^[9~11]. 本文基于矩阵理论^[12]及双空间分析思想, 提出双空间 FLCCA 算法, 在避免小特征值对应特征向量对数据降维的同时, 最大限度地保留样本所包含的有用信息, 以缓解过小特征值对 FLCCA 算法性能的影响.

人脸识别是图像识别领域的重大研究课题, 有相当重要的理论和应用价值^[13], 本文选用人脸图像进行识别实验, 从实验的角度进一步验证本文对所提方法理论分析的正确性.

收稿日期: 2006-12-03; 修回日期: 2008-11-27.

基金项目: 新世纪优秀人才支持计划资助项目(NCET-05-0275); 国家自然科学基金资助项目(60673006, 60873181).

作者简介: 苏志勋(1965-), 男, 教授, 博士生导师; 刘秀平*(1964-), 女, 博士, 副教授, E-mail: xpliu@comgi.com.

1 典型相关分析的基本理论

典型相关分析(CCA)是一种用于分析两组随机变量之间线性相关关系的统计方法,最早由 Hotelling 提出^[1],其基本思想是寻找两组随机变量各自的线性组合,使线性组合后的两个随机变量的相关性达到最大。

给定随机变量 $x \in \mathbf{R}^p$ 与 $y \in \mathbf{R}^q$, CCA 的目标是寻找 d 对投影方向 $\alpha_i \in \mathbf{R}^p$ 和 $\beta_i \in \mathbf{R}^q$, 使投影 $x_i^* = \alpha_i^T x$ 与 $y_i^* = \beta_i^T y$ 之间的相关系数最大, 且 x_i^* 之间及 y_i^* 之间均不相关。一般地, 该问题等价于求解如下特征方程的前 d 个大特征值对应的特征向量^[2]:

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \alpha = \lambda^2 \alpha,$$

$$\Sigma_{yy}^{-1} \Sigma_{yx} \Sigma_{xx}^{-1} \Sigma_{xy} \beta = \lambda^2 \beta$$

其中 Σ_{xx} 、 Σ_{yy} 和 Σ_{xy} 分别表示 x 和 y 的协方差矩阵和互协方差矩阵, 且 $\Sigma_{yx} = \Sigma_{xy}^T$ 。在实际问题中, 协方差矩阵一般是未知的, 只能根据样本估计进行计算:

$$C_{xx} = \frac{1}{N-1} \mathbf{X} \mathbf{X}^T, \quad C_{yy} = \frac{1}{N-1} \mathbf{Y} \mathbf{Y}^T,$$

$$C_{xy} = \frac{1}{N-1} \mathbf{X} \mathbf{Y}^T$$

其中 $\mathbf{X} = (x_1 - \mu_x \quad \cdots \quad x_N - \mu_x)$, $\mathbf{Y} = (y_1 - \mu_y \quad \cdots \quad y_N - \mu_y)$, $\mu_x = \frac{1}{N} \sum_{i=1}^N x_i$, $\mu_y = \frac{1}{N} \sum_{i=1}^N y_i$, x_i 、 y_i 分别表示 x 和 y 的 N 个样本。

2 模糊标号典型相关分析

把样本的分布信息有效地融合到特征提取过程将有助于得到更多的分类信息, 但是, 如何合理刻画样本的分布是一个关键问题。既然外界干扰模糊了各类图像之间的分界, 使图像样本与各图像类之间都存在关系, 那么利用隶属度来刻画这种模糊联系会是一个很好的选择。本章首先定义了一种新的隶属度函数来构造模糊标号, 用于描述样本的分布信息, 随后讨论了将其融合于图像特征提取过程的 CCA 算法。

2.1 模糊标号的定义

设有 c 类图像, 每类图像有 $N_i (i = 1, \dots, c)$ 幅训练样本, 将 $m \times n$ 维图像按每列首尾相连的方式转化为 $p = mn$ 维列向量, 得到训练样本集

$$\Omega = \{x_{ij} \in \mathbf{R}^p, i = 1, \dots, c; j = 1, \dots, N_i\}$$

其中 x_{ij} 表示第 i 类图像的第 j 个样本, 空间 Ω 刻画了图像的灰度信息。由前文, 为提高特征的分类能力, 需合理定义集合 ψ , 用于刻画 Ω 中样本的分布信息。

设 $m_k (k = 1, \dots, c)$ 为第 k 类图像 $\Omega_k = \{x_{kj} \in \Omega, j = 1, \dots, N_k\}$ 的样本均值, $\forall x_{ij} \in \Omega, d_{ij}^k = \|x_{ij} - m_k\|_2$ 为 x_{ij} 到 Ω_k 中心的欧式距离, 则定义样本 x_{ij} 归属于 Ω_k 的隶属度为

$$\omega_{ij}^k = \frac{d_{ij}^k - \max_{v=1, \dots, c} (d_{ij}^v)}{\min_{v=1, \dots, c} (d_{ij}^v) - \max_{v=1, \dots, c} (d_{ij}^v)} \quad (1)$$

$\omega_{ij}^k \in [0, 1]$, 其本质上是利用距离刻画样本的归属程度。但是外界环境的影响可能使某些样本偏离所属图像类, 使得其归属于其他类的程度更大, 若直接采用此定义描述样本分布, 错误的类别信息将影响对样本的最终分类。因此考虑通过引入惩罚项将这些样本拉回所属区域, 定义如下改进隶属度:

$$\tilde{\omega}_{ij}^k = \begin{cases} 1; & k = i \\ \frac{\omega_{ij}^k}{\omega_{ij}^i + \tau_{ij}}; & k \neq i \end{cases} \quad (2)$$

其中惩罚项

$$\tau_{ij} = \begin{cases} \frac{\max_{v \neq i} (\omega_{ij}^v) - \omega_{ij}^i}{\theta}; & \frac{\max_{v \neq i} (\omega_{ij}^v)}{\omega_{ij}^i} > \theta \\ 0; & \text{其他} \end{cases}$$

阈值 $\theta \in (0, 1)$ 与 τ_{ij} 成反比, 用于决定其大小。通过引入惩罚项, 改进后的隶属度满足 $\tilde{\omega}_{ij}^i = 1, \tilde{\omega}_{ij}^k \leq \theta (k \neq i)$, 即 x_{ij} 归属于 Ω_i 的程度最大, 归属于其他类的程度相对较小, 且 $\tilde{\omega}_{ij}^k / \tilde{\omega}_{ij}^i = \omega_{ij}^k / \omega_{ij}^i (k, l \neq i)$ 保证了样本归属于其他类的程度比例不变, 最大限度保留了其与各图像类之间的联系。

综上所述, $\tilde{\omega}_{ij}^k (k = 1, \dots, c)$ 分别描述了样本 x_{ij} 归属于各类图像的程度, 因此, 对 $\forall x_{ij} \in \Omega$, 定义其模糊标号 $y_{ij} = (\tilde{\omega}_{ij}^1 \quad \tilde{\omega}_{ij}^2 \quad \cdots \quad \tilde{\omega}_{ij}^c)^T$, 在某种意义上 y_{ij} 刻画了 x_{ij} 在 Ω 中的粗略位置, 从而

$$\psi = \{y_{ij} \in \mathbf{R}^c, i = 1, \dots, c; j = 1, \dots, N_i\}$$

近似描述了 Ω 中训练样本的分布情况。

此外, 改进隶属度中的惩罚项需要适度选取, 太小不足以抗拒噪声等的干扰, 过大又会使样本的原有分布信息丢失, 特别地, 若 θ 趋近于 0, 此时惩罚项 τ_{ij} 趋近于 ∞ , y_{ij} 退化为二值标号向量, 与 CCA 结合将等价于 FLDA(LCCA)^[5~7], 进而失

去引入模糊的意义,第4章的实验表明一般选择 $\theta \in (0.1, 0.4)$.

2.2 基于模糊标号 CCA 的特征提取

将 Ω 和 ψ 中的元素看做随机变量,利用 CCA 可以提取综合两个空间关联信息的有效特征,此时,只需求解特征方程 $\mathbf{C}_{xx}^{-1}\mathbf{C}_{xy}\mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\boldsymbol{\alpha} = \lambda^2\boldsymbol{\alpha}$. 但是,在模式识别领域,小样本问题常导致样本协方差矩阵奇异.文献[4]证明,利用 PCA 对高维数据降维,CCA 不会损失任何有效判别信息.具体地,设 \mathbf{C}_{xx} 的非零特征值及相应标准正交特征向量分别为 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_s > 0$ 和 $\mathbf{u}_i \in \mathbf{R}^p (i = 1, \dots, s, s = \text{rank}(\mathbf{C}_{xx}))$. 令 $\mathbf{P} = (\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_s)$, 对原始数据降维,可得 $\hat{\mathbf{C}}_{xx} = \mathbf{P}^T\mathbf{C}_{xx}\mathbf{P} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_s)$, $\hat{\mathbf{C}}_{xy} = \mathbf{P}^T\mathbf{C}_{xy}$, 易知 $\hat{\mathbf{C}}_{xx}$ 非奇异; 求 $\hat{\mathbf{C}}_{xx}^{-1}\hat{\mathbf{C}}_{xy}\mathbf{C}_{yy}^{-1}\hat{\mathbf{C}}_{yx}$ 前 d 个大特征值对应的特征向量 $\mathbf{g}_i \in \mathbf{R}^s, \boldsymbol{\alpha}_i = \mathbf{P}\mathbf{g}_i$ 即为使 Ω 与 ψ 相关性最大的前 d 个投影方向. 综上,对任一图像样本 $\mathbf{x} \in \mathbf{R}^p, \mathbf{z} = (\boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2 \ \dots \ \boldsymbol{\alpha}_d)^T \mathbf{x}$ 即所提取的 d 维特征向量,它综合利用了图像的灰度信息和分布信息,用于分类更具判别能力,其中 $d \leq \text{rank}(\hat{\mathbf{C}}_{xy})$.

3 双空间模糊标号典型相关分析

小样本问题还会导致 \mathbf{C}_{xx} 过小特征值包含较多干扰信息,直接利用会使结果存在较大误差,而产生过拟合现象.本章将针对该问题展开讨论并提出一种解决方案.

如 2.2 节所述,对于小样本问题,CCA 无需讨论类内散布及类间散布矩阵的主元空间与零空间所包含判别信息的区别^[9~11],只需利用 PCA 将数据降维至 \mathbf{C}_{xx} 的值域空间,即可获得全部判别信息,可见,CCA 用于特征提取比 FLDA 具有优势.但是,样本不足会导致 \mathbf{C}_{xx} 较小特征值更多的是对噪声的估计,并且数值趋于相等^[8],若利用其对应特征向量对数据降维,会使所求 CCA 投影方向主要由样本中的噪声决定,无法反映两个空间真正的潜在关系.为避免过拟合,最直接的办法就是将小特征值对应特征向量舍弃^[4],但其包含的判别信息也会随之丢失;或者对 CCA 进行正则化处理,给 $\hat{\mathbf{C}}_{xx}$ 加脊惩罚项 $\mu\mathbf{I} (\mu > 0, \mathbf{I}$ 是与 $\hat{\mathbf{C}}_{xx}$ 同维的单位矩阵),在引入小偏差的同时,使结果对样本选择及噪声干扰不敏感^[14],但 μ 的最优值

并不容易确定,不适当的值甚至会使结果变差^[3].

综合上述讨论,为缓解过小特征值对 FLCCA 算法性能的影响,不仅要避免正则化,还需要综合利用全部特征向量以最大限度地保留判别信息.由文献[12]可知,若矩阵的特征值互相接近,其对应的特征向量对于小的干扰非常敏感,但对由一组相近特征值对应的特征向量所张成的子空间而言却是稳定的,在这个子空间内求得的特征向量对于扰动较不敏感.基于此,本节利用双空间分析思想^[9,10],将 \mathbf{C}_{xx} 的特征空间分为两部分处理,提出一种双空间模糊标号 CCA 方法,记为 Dual-FLCCA,算法步骤如下:

(1) 分别构造集合 Ω 和 ψ , 得到样本矩阵 \mathbf{X} 和 \mathbf{Y} ;

(2) 计算 Ω 和 ψ 中样本的协方差矩阵 $\mathbf{C}_{xx}, \mathbf{C}_{yy}$ 及互协方差矩阵 \mathbf{C}_{xy} ;

(3) 计算 \mathbf{C}_{xx} 的非零特征值 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_s$ 及其标准正交特征向量组 $\mathbf{u}_1, \dots, \mathbf{u}_s, s = \text{rank}(\mathbf{C}_{xx})$;

(4) 定义满足 $\sigma_i/\sigma_{i+1} < 1 + \epsilon$ 的特征值为相近特征值, $i = 1, \dots, s-1, \epsilon$ 是个很小的数. \mathbf{C}_{xx} 的非零特征值被分为两部分,相应地,特征空间被划分为主空间 $\mathbf{F} = \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_M\}$ 和补空间 $\bar{\mathbf{F}}$;

(5) 令 $\tilde{\mathbf{P}} = (\mathbf{u}_1 \ \dots \ \mathbf{u}_M)$, 计算 $\tilde{\mathbf{C}}_{xx} = \tilde{\mathbf{P}}^T\mathbf{C}_{xx}\tilde{\mathbf{P}}$, $\tilde{\mathbf{C}}_{xy} = \tilde{\mathbf{P}}^T\mathbf{C}_{xy}$ 及样本数据在 $\bar{\mathbf{F}}$ 空间的投影 $\mathbf{Z} = (\mathbf{I} - \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T)\mathbf{X}$;

(6) 计算 \mathbf{ZZ}^T 的非零特征值及相应的标准正交特征向量组 $\mathbf{v}_1, \dots, \mathbf{v}_r, r = \text{rank}(\mathbf{ZZ}^T)$, 令 $\bar{\mathbf{P}} = (\mathbf{v}_1 \ \dots \ \mathbf{v}_r)$, 计算 $\bar{\mathbf{C}}_{xx} = \bar{\mathbf{P}}^T\mathbf{ZZ}^T\bar{\mathbf{P}}$, $\bar{\mathbf{C}}_{xy} = \bar{\mathbf{P}}^T\mathbf{ZY}^T$;

(7) 求 $\tilde{\mathbf{C}}_{xx}^{-1}\tilde{\mathbf{C}}_{xy}\mathbf{C}_{yy}^{-1}\tilde{\mathbf{C}}_{yx}$ 前 \bar{d} 个大特征值对应的特征向量 $\tilde{\mathbf{g}}_i$, 计算 \mathbf{F} 空间的投影方向 $\tilde{\boldsymbol{\alpha}}_i = \tilde{\mathbf{P}}\tilde{\mathbf{g}}_i, \bar{d} \leq \text{rank}(\tilde{\mathbf{C}}_{xy})$;

(8) 求 $\bar{\mathbf{C}}_{xx}^{-1}\bar{\mathbf{C}}_{xy}\mathbf{C}_{yy}^{-1}\bar{\mathbf{C}}_{yx}$ 前 \bar{d} 个大特征值对应的特征向量 $\bar{\mathbf{g}}_i$, 计算 $\bar{\mathbf{F}}$ 空间的投影方向 $\bar{\boldsymbol{\alpha}}_i = (\mathbf{I} - \tilde{\mathbf{P}}\tilde{\mathbf{P}}^T)\bar{\mathbf{P}}\bar{\mathbf{g}}_i, \bar{d} \leq \text{rank}(\bar{\mathbf{C}}_{xy})$;

(9) 对任一图像样本 $\mathbf{x} \in \mathbf{R}^p, \mathbf{z} = \begin{pmatrix} (\tilde{\boldsymbol{\alpha}}_1 \ \dots \ \tilde{\boldsymbol{\alpha}}_{\bar{d}})^T \mathbf{x} \\ \chi(\bar{\boldsymbol{\alpha}}_1 \ \dots \ \bar{\boldsymbol{\alpha}}_{\bar{d}})^T \mathbf{x} \end{pmatrix}$ 即所提取的 $\bar{d} + \bar{d}$ 维组合特

征向量, 其中 \bar{F} 空间特征的权系数 $\chi = (s - M)/s$.

Dual-FLCCA 避免了 C_{xx} 小特征值对应的特征向量对数据的直接降维, 从而缓解了过小特征值对 FLCCA 性能的影响, 但是, 尽管在 \bar{F} 空间求解的 v_1, \dots, v_r 相对于 u_{M+1}, \dots, u_s 而言是稳定的, 但其毕竟只是对 \bar{F} 空间中特征向量的近似逼近, 不可避免地包含干扰信息, 因此在 \bar{F} 空间提取的特征不能与 F 空间的特征等同. 注意到, 如果 \bar{F} 中包含 C_{xx} 的非零特征值越多, 该空间所包含的判别信息也越多, 所以在步骤(9)中引入以 $s - M$ 定义的权系数 χ .

综上所述, Dual-FLCCA 算法通过将两个空间提取的特征向量加权组合用于后续分类, 在缓解小特征值影响的同时, 最大限度地利用了样本所包含的分类判别信息, 较好地解决了高维小样本问题, 因此, 在处理干扰及相近特征值较多的情况时, Dual-FLCCA 是一种相对有效的方法.

4 实验及分析

人脸识别是一类典型的高维多类图像识别问题. 为验证本文所提算法的有效性及应用价值, 设计如下两组实验. 由于本文关注的是特征提取方法, 只采用简单的最近邻分类器和 Euclidian 距离进行分类.

4.1 组合人脸数据库

为验证算法在高维小样本以及样本采集条件变化较大情况下的有效性和稳定性, 本文构造了一个组合人脸数据库, 包含 ORL、Yale 以及本教研室自拍人脸数据库中的图像. ORL 标准人脸数据库由 40 人, 每人 10 幅、分辨率为 92×112 的图像组成, 有中等程度的姿态与尺度变化. Yale 标准人脸数据库由 15 人, 每人 11 幅、分辨率为 243×320 的图像组成, 有较大的表情变化、光照变化以及部分缺损. 图形图像教研室人脸数据库由本教研室 30 名学生, 每人 15 幅、分辨率为 128×128 的图像组成, 人脸的尺度、姿态和表情的变化程度较大, 部分包含光照和饰物的变化. 从各数据库中取每人 10 幅图像, 简单截取脸部区域归一化为 92×112 大小, 此外不做任何光照、旋转等预处理, 组合人脸数据库共 85 人, 包含光照、姿态、表情、饰物等多种变化, 样本情况较复杂, 如图 1 所示.

4.2 关于 FLCCA 的实验

FLCCA 通过将样本分布信息融入特征提取过程来提高算法性能, 本节设计如下实验, 用于验证 FLCCA 特征的分类能力和说明阈值 θ 的作用.

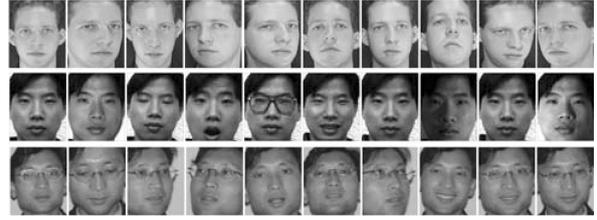
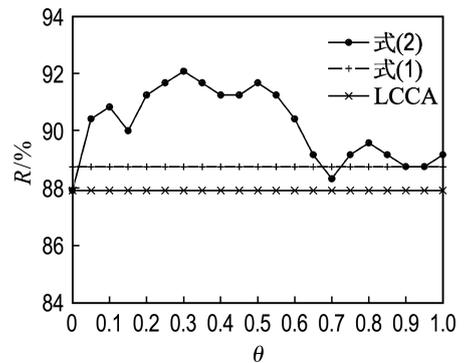


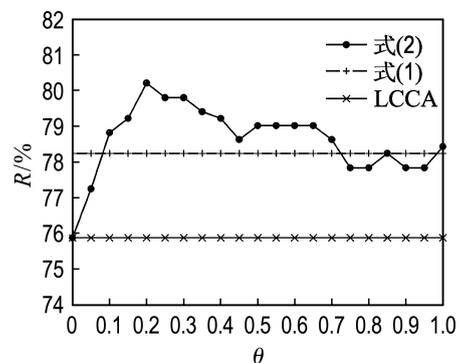
图 1 组合人脸数据库图像示例

Fig. 1 Sample images in the combined database

在 ORL 和组合数据库分别进行实验. 选定每人前 4 幅图像作为训练样本, 剩余 6 幅用于测试; 分别利用式(1)和(2)提取 FLCCA 特征, 以步长 0.05 改变 θ , 并与 LCCA^[5]方法比较; 为使结果更好地反映识别率 R 随 θ 的变化, 仅保留 PCA 大特征值对应的特征向量对数据降维, 取 $\epsilon = 5 \times 10^{-3}$. 结果如图 2 所示.



(a) ORL 人脸库



(b) 组合人脸库

图2 关于 FLCCA 的实验

Fig. 2 Experiments about FLCCA

由结果可知,FLCCA 总体来说能取得理想的识别效果,表明模糊标号可以更合理地描述样本的分布,有利于识别任务的完成.但是,由式(1)构造的模糊标号对样本分布的估计存在偏差,识别率的增长幅度并不大.式(2)引入由 θ 决定的惩罚项来改进模糊标号对样本分布的描述能力,两数据库均在 $\theta \in (0, 1, 0.4)$ 时获得最高识别率.由结果还可看出,当 θ 接近 0 时,FLCCA 退化为 LCCA;当 θ 增大到某程度后,过小的惩罚会因不能抗拒干扰而使识别率变差.因此,选择适当的 θ ,有助于获得更具判别能力的特征.值得指出的是,ORL 和组合人脸库分别在 $\theta=0.3$ 和 $\theta=0.2$ 时获得最高识别率,可见,对于越复杂的训练样本数据,所给的惩罚项也应该越大.

4.3 关于 Dual-FLCCA 的实验

本实验用于验证 Dual-FLCCA 对高维小样本问题的解决能力.取组合人脸数据库中每人的前 $k(2 \leq k \leq 9)$ 幅图像作为训练样本,剩余后 $10-k$ 幅测试.令 R-FLCCA 表示用正则化处理的 FLCCA,令 P-FLCCA 表示直接舍弃小特征值的处理,即 $\chi=0$.实验中,取 $\theta=0.25$,R-FLCCA 分别加 μ 等于 10^{-6} 、 10^{-7} 和 10^{-8} 的正则项,结果见图 3.其中 P-FLCCA 的特征维数与其保留特征值的个数相等,R-FLCCA 与 Dual-FLCCA 的特征维数均为 85.

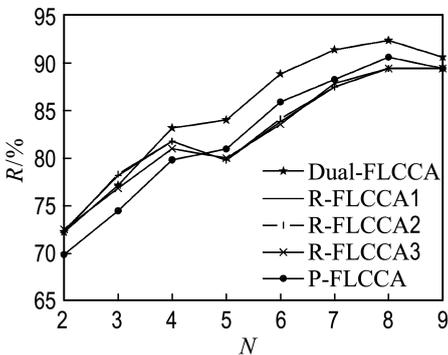


图3 关于 Dual-FLCCA 的实验

Fig. 3 Experiments about Dual-FLCCA

由图 3 可知, μ 的选取对 R-FLCCA 的性能有一定影响,识别率会随着正则项的改变发生波

动;而且随着训练样本的增加, C_{xx} 相近特征值对算法性能的影响越来越强烈, $k \geq 5$ 时,P-FLCCA 的识别率已高于 R-FLCCA;但在 k 较小时,由于舍弃太多有用信息,P-FLCCA 的效果较差. Dual-FLCCA 将在两个空间提取的特征加权组合用于分类,在解决相近特征值影响的同时,最大限度地保留有用信息,因此,在 k 较大时能取得良好的识别效果, k 较小时也能取得与 R-FLCCA 相当的识别率,较好地缓解了过小特征值对算法性能的影响.

5 结 语

本文通过定义隶属度函数来构造模糊标号,用其描述样本分布,并与 CCA 结合提取综合灰度信息和分布信息的有效特征,从而提出模糊标号 CCA 特征提取方法.此外,针对小样本情形导致的对协方差矩阵较小特征值估计不足的问题,提出基于双空间分析的改进算法,以缓解过小特征值对算法性能的影响.人脸识别实验结果证实了新算法的有效性及应用价值.

复杂事物间的关系往往是非线性的,如何将本文算法推广至 CCA 的非线性形式^[3,5],是本文下一步的工作;另外,如何更合理地刻画样本的分布,以及如何更好地解决高维小样本问题与过拟合现象^[11],也是需要进一步考虑的问题.

参考文献:

- [1] HOTELLING H. Relations between two sets of variates [J]. *Biometrika*, 1936, **28**:321-377
- [2] 孙文爽,陈兰祥.多元统计分析[M].北京:高等教育出版社,1994
- [3] MELZER T. Generalized canonical correlation analysis for object recognition [D]. Vienna: Vienna University of Technology, 2002
- [4] 孙权森,曾生根,王平安,等.典型相关分析的理论及其在特征融合中的应用 [J]. *计算机学报*, 2005, **28**(9):1524-1533
- [5] HE Y H, ZHAO L, ZOU C R. Face recognition based on PCA/KPCA plus CCA [C] // *ICNC 2005, LNCS 3611*. Berlin: Springer-Verlag, 2005
- [6] SUN T K, CHEN S C. Class label versus sample label-based CCA [J]. *Applied Mathematics and*

- Computation, 2007, **185**:272-283
- [7] BARKER M, RAYENS W. Partial least squares for discrimination [J]. **Journal of Chemometrics**, 2003, **17**(3):166-173
- [8] MOGHADDAM B, PENTLAND A. Probabilistic visual learning for object representation [J]. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 1997, **19**(7):696-709
- [9] YANG J, YANG J Y. Why can LDA be performed in PCA transformed space [J]. **Pattern Recognition**, 2003, **36**:563-566
- [10] WANG X G, TANG X O. Dual-space linear discriminant analysis for face recognition [C] // **Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)**. Washington: IEEE Computer Society, 2004
- [11] ZUO W M, WANG K Q, ZHANG D, *et al.* Regularization of LDA for face recognition: a post-processing approach [C] // **AMFG 2005, LNCS 3723**. Berlin: Springer-Verlag, 2005
- [12] STEWART G W. **Introduction to Matrix Computations** [M]. New York: Academic Press, 1973
- [13] ZHAO W, CHELLAPPA R, ROSENFELD A, *et al.* Face recognition: a literature survey [J]. **ACM Computing Surveys**, 2003, **35**(4):399-459
- [14] DE BIE T, CRISTIANINI N, ROSIPAL R. Eigenproblems in pattern recognition [M] // **Handbook of Computational Geometry for Pattern Recognition, Computer Vision, Neurocomputing and Robotics**. Heidelberg: Springer-Verlag, 2004

Fuzzy label canonical correlation analysis and its application to face recognition

SU Zhi-xun¹, LIU Yan-yan², LIU Xiu-ping^{*1}

(1. Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China;

2. Department of Basic Courses, Institute of Disaster Prevention Science and Technology, Sanhe 065201, China)

Abstract: Incorporating the sample distribution information into the process of feature extraction is beneficial to promoting the classification performance of features. A fuzzy label canonical correlation analysis (CCA) algorithm is proposed for image feature extraction. Fuzzy class labels in the form of membership degrees are designed elaborately to represent the sample distribution. Then the fuzzy labels are embedded in CCA to extract more discriminative features which combine the information about gray level and distribution together. Furthermore, according to the matrix theory and dual-space idea, an improved method named dual-space fuzzy label CCA is proposed to counteract the effect of small eigenvalues which are poorly estimated due to finite samples. The experimental results on ORL and combined face databases show that the features have a powerful ability of recognition, and that the proposed methods are efficient and practical.

Key words: canonical correlation analysis; fuzzy membership degree; small sample size problem; feature extraction; face recognition