

一种基于指数降维的监督型稀疏保持 典型相关分析算法

蒋文, 齐林

(郑州大学 信息工程学院, 河南 郑州 450001)

摘要:提出一种基于指数降维的监督型稀疏保持典型相关分析算法. 通过将样本的类别信息与样本特征相融合, 克服以往引入监督信息导致重建误差增大的缺陷, 同时实现类内相关的最大化与类间相关的最小化; 针对传统算法处理稀疏信号的高维小样本问题的瓶颈, 改进算法对总体散布矩阵做指数化的处理, 既保留有效信息, 又将总体散布矩阵非奇异化, 克服 PCA 预处理散布矩阵导致有效信息流失的缺陷. 依据 ORL, Yale, AR 和 FERET 人脸数据库而进行的仿真实验表明, 该算法比其他的典型相关分析方法具有更好的识别效果.

关键词:典型相关分析(CCA); 稀疏保持(SPP); 指数降维; 特征提取; 人脸识别

中图分类号: O235 **文献标志码:** A **DOI:** 10.3969/j.issn.2095-476X.2015.5/6.019

An improved supervised sparsity preserving CCA algorithm based on exponential dimensionality reduction

JIANG Wen, QI Lin

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: An improved supervised sparsity preserving canonical correlation analysis algorithm based on exponential dimensionality reduction was proposed. The problem that the fitting error increased while adding supervised information to the SPPCA was solved by the fusion of the class label information and sample feature. The local manifold structure of the data was realized at the same time. Aimed at the problem of traditional algorithm in dealing with small sample of high-dimensiona sparse signal, index scattering matrix was used to retain effective information while building the non-singular scattering matrix. It overcame the default of effective information losses while using PCA to extract principal features of the scattering matrix. The experimental results on ORL, Yale, AR and FERET face databases showed that the proposed algorithm was better than related canonical correlation analysis methods in recognition effect.

Key words: canonical correlation analysis(CCA); sparsity preserving projection(SPP); exponential dimensionality reduction; feature extraction; face recognition

收稿日期: 2015-04-16

基金项目: 国家自然科学基金项目(61210005, 61331021)

作者简介: 蒋文(1991—), 男, 四川省安岳县人, 郑州大学硕士研究生, 主要研究方向为计算机图像处理技术.

通信作者: 齐林(1961—), 男, 河南省郑州市人, 郑州大学教授, 博士研究生导师, 主要研究方向为通信系统及其信号处理、信号检测与估计、多媒体信号处理、情感及生物特征识别.

0 引言

随着现代社会人类掌握的技术手段越来越先进,人们可以从多种信息源获得海量信息,如何去隐藏其中的大量的冗余信息,已经成为当前数据分析和处理的关键.近十年来,典型相关分析(CCA)被广泛地应用于模式识别、计算机视觉及生物医学等领域中,同时在诸多领域中取得突破,例如人脸识别、行为分类和疾病诊断等^[1-4].经过科研人员的不懈努力,CCA不断地进行着改进,取得了许多成果:能够更好地解决人脸识别中的非线性问题的核典型相关分析(KCCA)^[5];能够实现类内相关的最大化与类间相关的最小化且比CCA具有更好的识别性能的判别型典型相关分析(DCCA)^[6];使投影后样本在最小化类内离散度的同时,两组特征之间具有最大的相关性,从而有利于投影后同类样本聚类信息的保持的广义典型相关分析(GCCA)^[7];能有效避免小样本问题的发生,较好地描述非线性的人脸识别问题的子模式典型相关分析(SpCCA)^[8];能利用少量的监督信息来提高分类性能的半监督的典型相关分析(Semi-CCA)^[9];能够捕获数据的局部流形结构特性,在数据可视化和姿态估计中获得比CCA更好的实验结果的局部保持的典型相关分析(LPCCA)^[10].另外,侯书东等^[11-12]分别提出SPCCA和S2PCCA两种稀疏保持典型相关分析方法,能实现两组不同特征的融合的基础上,对样本间的稀疏重构性进行约束,增强鉴别能力.然而,上述这些稀疏表示算法中,没有区分对待不同类别的稀疏表示能力,从而无法利用现实中存在的大量类别信息,只是在传统SPP算法中引入监督信息,人为地把其他类样本的稀疏系数设置为零^[13-15],从表面上消除了异类样本的稀疏表示误差.但实际上恰恰相反,一部分稀疏表示的信息损失,最终导致稀疏重建误差的增大.此外,用CCA对两组稀疏信号进行融合时,有时会出现识别率下降的问题,通过大量实验发现并验证,此问题源于对总体散布矩阵(或自协方差矩阵)的预处理方面.孙权森等^[4,7]的处理方法是用PCA先对总体散布矩阵做特征提取,构造出非奇异的散布矩阵,由于稀疏信号的总体散布矩阵特征值较少(即只有少数较大特征值),提取后导致一部分有效信息流失,识别率下降.

本文在多特征融合的基础上结合局部化思想,

综合考虑类别信息与局部邻域信息,针对以往引入监督信息导致重建误差增大和PCA预处理散布矩阵导致有效信息流失的缺陷,提出一种基于指数降维的监督型稀疏保持典型相关分析算法,以期克服上述缺陷,获得更好的识别效果.

1 基于指数降维的监督型稀疏保持典型相关分析算法

1.1 改进型监督稀疏保持投影算法理论

本文在多特征融合的基础上结合局部化思想,将样本的类别信息与样本特征相融合,以克服文献[13-14]提出的算法因人为消除了异类样本的稀疏表示误差而导致误差增大的缺陷,并实现类内相关的最大化与类间相关的最小化.

设 $\mathbf{X}_1 = [x_1, x_2, \dots, x_N] \in R^{p \times N}$ 和 $\mathbf{X}_2 = [y_1, y_2, \dots, y_N] \in R^{q \times N}$ 为两组特征样本,样本数为 N , \mathbf{X}_1 的特征维数为 p , \mathbf{X}_2 的特征维数为 q .

构建类信息矩阵 $\mathbf{X}_3 = [l_1, l_2, \dots, l_N] \in R^{C \times N}$,样本 x_i 和 y_i 的类标号向量为 l_i ,模式的类数为 C .对于 C 类($C \geq 2$)分类问题,目前已有多种类标号编码方式,本文采用One-of-C编码,即 i 类样本的类标号均为 $(\underbrace{0, 0, \dots, 0}_i, 1, 0, \dots, 0)^T \in R^C$.

因此鉴别函数可写为

$$\max_{w_x, w_y \neq 0} \frac{\sum_{i=1}^3 \sum_{j=1}^3 \alpha_i^T \mathbf{S}_{ij} \alpha_j}{\sqrt{\alpha_1^T \mathbf{S}_{11} \alpha_1} \times \sqrt{\alpha_2^T \mathbf{S}_{22} \alpha_2} \times \sqrt{\alpha_3^T \mathbf{S}_{33} \alpha_3}}$$

其中, \mathbf{S}_{ij} 为 \mathbf{X}_i 与 \mathbf{X}_j 的互协方差矩阵; $\mathbf{S}_{11} = \mathbf{X}_1(I - R)(I - R)^T \mathbf{X}_1^T$ 和 $\mathbf{S}_{22} = \mathbf{X}_2(I - S)(I - S)^T \mathbf{X}_2^T$ 分别为 \mathbf{X}_1 和 \mathbf{X}_2 的稀疏保持散度矩阵,且均满足矩阵的正定性; $\mathbf{S}_{33} = \mathbf{X}_3 \mathbf{X}_3^T$ 为 \mathbf{X}_3 的自协方差矩阵.

又因为 $\alpha_i^T \mathbf{S}_{ij} \alpha_j = (\alpha_i^T \mathbf{S}_{ij} \alpha_j)^T = \alpha_j^T \mathbf{S}_{ji} \alpha_i$,所以

$$\begin{aligned} \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i^T \mathbf{S}_{ij} \alpha_j &= \alpha_1^T \mathbf{S}_{11} \alpha_1 + \alpha_1^T \mathbf{S}_{12} \alpha_2 + \\ &\alpha_1^T \mathbf{S}_{13} \alpha_3 + \alpha_2^T \mathbf{S}_{21} \alpha_1 + \alpha_2^T \mathbf{S}_{22} \alpha_2 + \\ &\alpha_2^T \mathbf{S}_{23} \alpha_3 + \alpha_3^T \mathbf{S}_{31} \alpha_1 + \alpha_3^T \mathbf{S}_{32} \alpha_2 + \alpha_3^T \mathbf{S}_{33} \alpha_3 = \\ &\alpha_1^T \mathbf{S}_{11} \alpha_1 + \alpha_2^T \mathbf{S}_{22} \alpha_2 + \alpha_3^T \mathbf{S}_{33} \alpha_3 + \\ &2\alpha_1^T \mathbf{S}_{12} \alpha_2 + 2\alpha_1^T \mathbf{S}_{13} \alpha_3 + 2\alpha_2^T \mathbf{S}_{23} \alpha_3 \end{aligned} \quad (1)$$

上述准则函数的求解可表述为优化模型

$$\begin{aligned} \max_{\alpha_1, \alpha_2, \alpha_3} \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i^T \mathbf{S}_{ij} \alpha_j \\ \text{s. t.} \quad \alpha_1^T \mathbf{S}_{11} \alpha_1 = \alpha_2^T \mathbf{S}_{22} \alpha_2 = \alpha_3^T \mathbf{S}_{33} \alpha_3 = 1 \end{aligned}$$

利用Lagrange乘子法最终可以将上式转化为

$$L(\alpha_1, \alpha_2, \alpha_3) = \sum_{i=1}^3 \sum_{j=1}^3 \alpha_i^T S_{ij} \alpha_j - \lambda(\alpha_1^T \tilde{S}_{11} \alpha_1 - 1) - \lambda(\alpha_2^T \tilde{S}_{22} \alpha_2 - 1) - \lambda(\alpha_3^T S_{33} \alpha_3 - 1)$$

令 $\frac{\partial L}{\partial \alpha} = 0, i = 1, 2, 3$, 得到广义特征方程

$$\begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \lambda \begin{pmatrix} \tilde{S}_{11} & 0 & 0 \\ 0 & \tilde{S}_{22} & 0 \\ 0 & 0 & S_{33} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}$$

求解该前 d 个最大广义特征值对应的广义特征向量 $\{\alpha_{1i}, \alpha_{2i}, \alpha_{3i}\}, i = 1, 2, \dots, d$, 从而可构造基向量组 $\{W_1, W_2, W_3\}$, 其中

$$\begin{aligned} W_1 &= \{\alpha_{11}, \alpha_{12}, \dots, \alpha_{1d}\} \\ W_2 &= \{\alpha_{21}, \alpha_{22}, \dots, \alpha_{2d}\} \\ W_3 &= \{\alpha_{31}, \alpha_{32}, \dots, \alpha_{3d}\} \end{aligned} \quad (2)$$

将线性变换 (3) 作为投影后的融合特征用于分类

$$Z = [W_1 \quad W_2 \quad W_3]^T \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \quad (3)$$

其中投影矩阵为 $W = [W_1 \quad W_2 \quad W_3]$.

综上所述, 算法流程总结如下.

步骤 1 输入两组特征集 X_1 和 X_2 ;

步骤 2 构建类信息矩阵 X_3 ;

步骤 3 根据式 (1) 构建 $\sum_{i=1}^3 \sum_{j=1}^3 \alpha_i^T S_{ij} \alpha_j$, 其中 S_{ij}

为 X_i 与 X_j 的互协方差矩阵;

步骤 4 构建 X_1 和 X_2 的稀疏保持散度矩阵 \tilde{S}_{11} 和 \tilde{S}_{22} , 并构建 X_3 的自协方差矩阵 S_{33} ;

步骤 5 根据式 (2) 计算投影矩阵 W ;

步骤 6 根据式 (3) 计算融合特征 Z ;

步骤 7 对测试样本进行分类.

1.2 自协方差矩阵的高维小样本情况分析

CCA 虽然已经应用于诸多领域, 但前提是自协方差矩阵可逆. 文献[22]虽然解决了求解问题, 但其所用的广义逆并非理论意义上的精确解, 从而限制了其应用. 文献[4, 7]的处理方法是用 PCA 先对总体散布矩阵做特征提取, 构造出非奇异的散布矩阵. 然而此方法在在处理两组稀疏信号时, 由于稀疏信号的总体散布矩阵特征值较少(即只有少数较大特征值), 如果用 PCA 的方法进行提取就会导致有效信息流失, 从而导致识别率下降.

本文受文献[15]的启发, 用指数化总体散布矩

阵的方法, 将高维的原始样本的特征空间变换为指数空间, 既可保留有效信息, 又将总体散布矩阵非奇异化.

引理 1 定义特征样本集 $X = [x_1, x_2, \dots, x_N] \in R^{p \times N}$, 其指数表示 e^X 定义为

$$e^X = I + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \dots + \frac{X^N}{N!} + \dots$$

因此, 指数化的鉴别函数可以定义为

$$\max_{w_x, w_y \neq 0} \frac{\sum_{i=1}^3 \sum_{j=1}^3 \alpha_i^T S_{ij} \alpha_j}{\sqrt{\alpha_1^T e^{S_{11}} \alpha_1} \times \sqrt{\alpha_2^T e^{S_{22}} \alpha_2} \times \sqrt{\alpha_3^T e^{S_{33}} \alpha_3}}$$

综上所述, 本文算法综合考虑类别信息与局部邻域信息, 将样本的类别信息与样本特征相融合, 并使得提取出的特征能够实现同类样本之间相关最大化, 同时对总体散布矩阵做指数化的处理, 既可保留有效信息, 又将总体散布矩阵非奇异化, 从而克服以往引入监督信息导致重建误差增大和 PCA 预处理散布矩阵导致有效信息流失的缺陷.

2 实验结果分析

本文在 FERET, ORL, Yale 和 AR 这 4 个人脸数据库中识别人脸识别仿真实验, 并将本文算法与 DCCA, LPCCA, SLPPCA, SPCCA, S2PCCA 进行对比, 同时也将鉴别分析算法 LDA 和稀疏保持投影算法 SPP 的实验结果引入, 进行对比分析.

在 ORL 人脸数据库上的仿真实验中, 对 40 个人的 400 幅灰度图像进行分组, 在每个人的 10 幅图片中选取 5 幅图像作为训练, 其余 5 张作测试.

在 Yale 人脸数据库上的仿真实验中, 对 15 个人的 165 幅灰度图像进行分组, 在每个人的 11 幅图片中选取 5 幅图像作为训练, 其余 6 张作测试.

在 AR 人脸数据库上的仿真实验中, 对 126 个人的 4 000 多幅彩色度图像进行分组, 在每个人的 26 幅图片中选取 13 幅图像作为训练, 其余 13 张作测试.

为了增强算法的可对比性, 在 FERET 人脸数据库的一个子库上进行仿真实验, 对 200 个人的 1 400 幅灰度图像进行分组, 在每个人的 7 幅图片中选取 4 幅图像作为训练, 其余 3 张作测试.

实验结果见表 1 (括号内为取得最大识别率对应的维数).

表 1 各算法在 4 个人脸数据库上的

识别率比较 %

算法	FERET	ORL	Yale	AR
LDA	85.9(17)	94.5(20)	92.9(14)	96.2(24)
SPP	84.1(17)	94.7(20)	93.1(15)	96.2(26)
LPCCA	44.0(100)	95.5(41)	77.0(38)	86.4(34)
DCCA	80.9(17)	94.5(17)	92.4(14)	94.1(16)
SLPCCA	45.8(77)	94.8(40)	87.6(45)	82.8(39)
SPCCA	86.2(19)	95.3(17)	92.9(14)	92.9(14)
S2PCCA	87.8(20)	95.0(22)	94.1(14)	96.1(20)
本文算法	89.4(27)	96.8(23)	94.2(14)	97.3(34)

由表 1 可见,在 4 个人脸库中,本文算法的识别率均优于其他相关方法.实验结果表明,SPP 识别率与 LDA 相当,本文算法明显强于这两种算法.从 SPCCA 与 S2PCCA 的识别率对比中可知,监督信息的加入确实提高了识别性能,并且 S2PCCA 比较稳定,因而不会导致 SPP 中某次实验出现较低识别率的情况,但是这种改进的效果并不明显,证明此种引入监督信息的方式丢失了一些信息,影响了识别效果.实验结果显示,本文算法与 S2PCCA 识别率相当,究其原因,这是因为 Yale 库中人脸图像数据较少,所以本文算法提取到的鉴别信息并不比 S2PCCA 丰富,导致两种方法的识别率相当.

图 1—图 4 分别给出各算法在 ORL, Yale, AR, FERET 人脸库中识别率随特征维数变化的实验结果.鉴于在 4 个人脸库上仿真实验中,SPP 与 LDA 两种算法的识别性能相当,所以在上述 4 图中仅给出 SPP 的识别率随维数变化的曲线.同理,由于实验结果显示 LPCCA 与其他算法结果差距明显,且 LPP 的最初目的并非模式分类,因此未在图中显示其结果.鉴于 SLPCCA 分类效果不佳的实验结果,其实验结果也在图中略去.

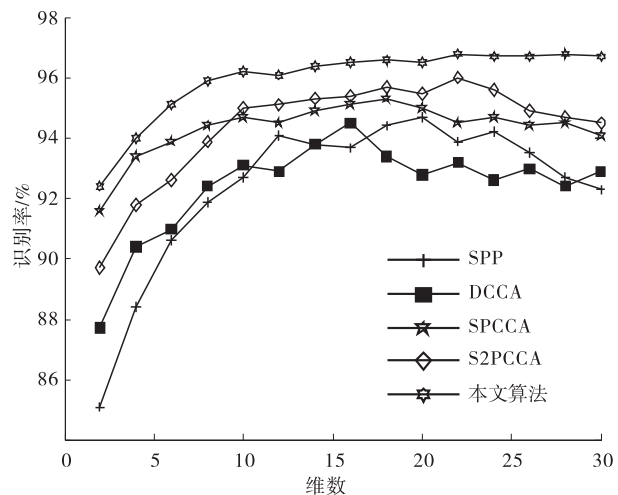


图 1 各算法在 ORL 人脸数据库上的识别率

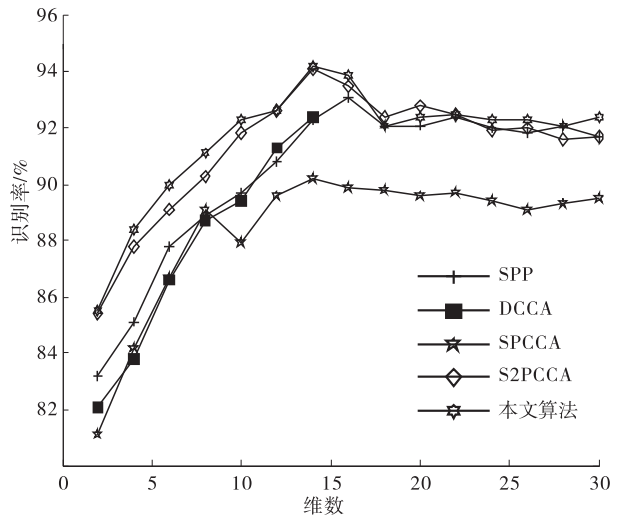


图 2 各算法在 Yale 人脸数据库上的识别率

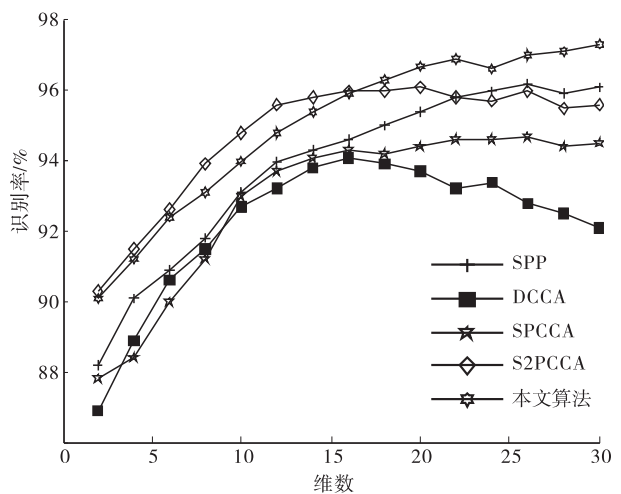


图 3 各算法在 AR 人脸数据库上的识别率

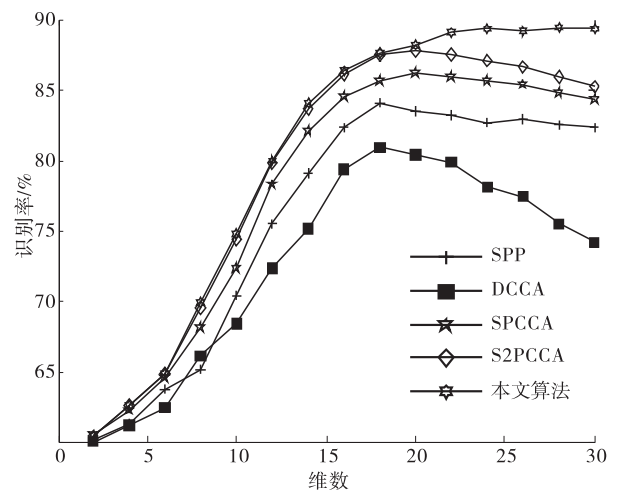


图 4 各算法在 FERET 人脸数据库上的识别率

从图 1 和图 3 中可以看出,随着特征维数的增加,多数算法的识别率会显著增加,但当其增加到

特定维数时便趋于稳定.然而在图2和图4中,这种趋势略微改变,最初阶段,识别率随着维数的增加而增加,但到达峰值时,随着特征维数的增加高维特征的识别率反而会有所减少.本文算法既保留了有效信息,又将总体散布矩阵非奇异化,克服了以往引入监督信息导致重建误差增大和PCA预处理散布矩阵导致有效信息流失的缺陷,不仅比其他的典型相关分析方法具有更好的识别效果,其识别率也会随着维数的增加而保持在稳定的状态.

3 结语

本文提出了一种基于指数降维的改进型稀疏保持典型相关分析算法,将样本的类别信息与样本特征相融合,实现了类内相关的最大化与类间相关的最小化;此外,对总体散布矩阵做指数化的处理,既保留了有效信息,又将总体散布矩阵非奇异化,同时为后续分类提供充足的有效信息.依据ORL, Yale, AR和FERET4个人脸数据库进行仿真实验,表明本文算法既保留了有效信息,又将总体散布矩阵非奇异化,克服了以往引入监督信息导致重建误差增大和PCA预处理散布矩阵导致有效信息流失的缺陷,比其他的典型相关分析方法具有更好的识别效果.

参考文献:

- [1] Sun Q S, Zeng S G, Liu Y, et al. A new method of feature fusion and its application in image recognition[J]. *Pattern Recognition*, 2005, 38(12): 2437.
- [2] Kim T K, Cipolla R. Canonical correlation analysis of video volume tensors for action categorization and detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(8): 1415.
- [3] Correa N M, Eichele T, Adali T, et al. Mult-set canonical correlation analysis for the fusion of concurrent single trial EPR and functional MRI[J]. *NeuroImage*, 2010, 50(4): 1438.
- [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.
- [5] Zheng W M, Zhou X Y, Zou C R, et al. Facial expression recognition using kernel canonical correlation analysis[J]. *IEEE Transactions on Neural Networks*, 2006, 17(1): 233.
- [6] Sun T K, Chen S C, Yang J Y, et al. A novel method of combined feature extraction for recognition[C]// *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Piscataway: IEEE*, 2008: 1043.
- [7] Sun Q S, Liu Z D, P. A. Heng P A, et al. A theorem on the generalized canonical projective vectors[J]. *Pattern Recognition*, 2005, 38(3): 449.
- [8] 洪泉, 陈松灿, 倪雪蕾. 子模式典型相关分析及其在人脸识别中的应用[J]. *自动化学报*, 2008, 34(1): 21.
- [9] 彭岩, 张道强. 半监督典型相关分析算法[J]. *软件学报*, 2008, 19(11): 2822.
- [10] Sun T K, Chen S C. Locality preserving CCA with applications to data visualization and pose estimation[J]. *Image and Vision Computing*, 2007, 25(5): 531.
- [11] 侯书东, 孙权森. 稀疏保持典型相关分析及在特征融合中的应用[J]. *自动化学报*, 2012, 38(4): 659.
- [12] 侯书东. 基于相关投影分析的特征提取研究及在图像识别中的应用[D]. 南京: 南京理工大学, 2012.
- [13] 相文楠, 赵建立. 监督型稀疏保持投影[J]. *计算机工程与应用*, 2011, 47(29): 186.
- [14] 王静. 基于稀疏表达的人脸识别算法研究[D]. 大连: 大连理工大学, 2011.
- [15] Wang S J, Yan S C, Yang J, et al. A General Exponential Framework for Dimensionality Reduction[J]. *IEEE Transaction on Image Processing*, 2014, 23(2): 920.
- [16] Barker M, Rayens W. Partial least squares for discrimination[J]. *Journal of Chemometrics*, 2003, 17(3): 166.
- [17] Franklin J. The Elements of Statistical Learning: data mining, inference and prediction[J]. *The Mathematical Intelligencer*, 2005, 27(2): 83.
- [18] Horikawa Y. Use of autocorrelation kernels in kernel canonical correlation analysis for texture classification[J]. *Lecture Notes in Computer Science*, 2004, 3316: 1235.
- [19] Johansson B. On classification: simultaneously reducing dimensionality and finding automatic representation using canonical correlation[R]. [s.l]: [s.n], 2001.
- [20] Lee Y, Lin Y, Wahba G. Multicategory support vector machines, theory and application to the classification of microarray data and satellite radiance data[J]. *Journal of the American Statistical Association*, 2004, 99: 67.
- [21] An S J, Liu W Q, Venkatesh S. Face recognition using kernel ridge regression[C]// *IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE*, 2007: 1-7.
- [22] 张尧庭, 方开泰. 多元统计分析引论[M]. 北京: 科学出版社, 1999.