

## 稀疏保持典型相关分析及在特征融合中的应用

侯书东<sup>1</sup> 孙权森<sup>1</sup>

**摘要** 稀疏保持投影 (Sparsity preserving projections, SPP) 由于保持了数据间的稀疏重构性, 因而获取的投影向量满足旋转、尺度和平移的不变性, 并能够在无标签的情况下提取样本的自然鉴别信息, 在人脸识别领域取得了较为成功的应用. 本文在典型相关分析 (Canonical correlation analysis, CCA) 的基础上引入稀疏保持项, 提出一种稀疏保持典型相关分析 (Sparsity preserving canonical correlation analysis, SPCCA). 该方法不仅实现了两组特征集鉴别信息的有效融合, 同时对提取特征间的稀疏重构性加以约束, 增强了特征的表示和鉴别能力. 在多特征手写体字符集与人脸数据集上的实验结果表明, SPCCA 比 CCA 具有更优的识别性能.

**关键词** 典型相关分析 (CCA), 稀疏保持投影 (SPP), 稀疏保持典型相关分析 (SPCCA), 特征融合

**引用格式** 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. 自动化学报, 2012, 38(4): 659–665

**DOI** 10.3724/SP.J.1004.2012.00659

## Sparsity Preserving Canonical Correlation Analysis with Application in Feature Fusion

HOU Shu-Dong<sup>1</sup> SUN Quan-Sen<sup>1</sup>

**Abstract** Sparsity preserving projections (SPP) aim to preserve the sparse reconstructive relationship among the data and have been successfully applied in face recognition. The projections are invariant to rotations, rescalings, and translations of the data, and more importantly, they contain natural discriminating information even without class labels. Enlightened by this, we propose a sparsity preserving canonical correlation analysis (SPCCA). It can not only fuse the discriminative information of two feature sets efficiently, but also constrains the sparse reconstructive relationship among each feature set in order to increase the representational power and has good discrimination capability of the feature extracted by SPCCA. Experimental results on the multiple feature databases and face databases show that the proposed SPCCA is better than CCA.

**Key words** Canonical correlation analysis (CCA), sparsity preserving projections (SPP), sparsity preserving CCA (SPCCA), feature fusion

**Citation** Hou Shu-Dong, Sun Quan-Sen. Sparsity preserving canonical correlation analysis with application in feature fusion. *Acta Automatica Sinica*, 2012, 38(4): 659–665

收稿日期 2011-04-29 录用日期 2011-07-01  
Manuscript received April 29, 2011; accepted July 1, 2011  
国家自然科学基金 (60773172), 教育部博士学科点基金 (200802880017), 江苏省自然科学基金 (BK2008411) 资助  
Supported by National Natural Science Foundation of China (60773172), Doctoral Fund of Ministry of Education of China (200802880017), and National Science Foundation of Jiangsu Province (BK2008411)  
本文责任编辑 刘一军  
Recommended by Associate Editor LIU Yi-Jun  
1. 南京理工大学计算机科学与技术学院 南京 210094  
1. School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094

特征提取作为模式识别的重要组成部分, 通常的研究对象是一组目标的一组数据, 该数据称之为单模态的原始特征数据, 如字符、人脸等. 图像识别领域中一般将像素值作为原始的特征数据, 当目标的数据表示获取之后, 再利用基于单模态的特征提取方法, 如主成分分析 (Principle component analysis, PCA)<sup>[1–2]</sup>、线性鉴别分析 (Linear discriminant analysis, LDA)<sup>[3–5]</sup> 等降维技术提取目标的统计特征, 提高特征的表示与鉴别能力, 降低噪声、干扰等不利因素造成的负面影响, 增强系统的稳定性. 实验证明在良好的限制条件下, 基于单模态的方法能够获取很好的实验结果. 然而在包括图像识别等的目标识别应用中, 基于单组特征的特征提取算法会受到各种现实退化因素的困扰, 造成在噪声环境下系统性能和稳定性的下降. 数据表示的多样性预示着特征描述方式的多样性, 每种描述方式都有其关注的侧重点, 不可能也不必包涵所有的表示信息. 如何将各种信息通过有效的途径予以融合处理, 这属于多视图学习的研究范畴<sup>[6]</sup>. 与传统的将多组特征向量组合成一个高维向量的简单方式相比, 多视角学习能够挖掘各组特征之间的相关性, 并能以较少的维数取得更好的效果.

典型相关分析 (Canonical correlation analysis, CCA)<sup>[7]</sup> 通过最大化两组变量在投影空间中的相关性而达到特征融合的目的, 已成功应用于字符与人脸等图像识别<sup>[8]</sup>. 核典型相关分析 (Kernel CCA, KCCA)<sup>[9]</sup> 通过隐性的非线性映射将两组特征分别映射到高维特征空间, 借助于核技巧, 以线性的方式提取模式的非线性特征. 此外, 局部保持典型相关分析 (Locality preserving CCA, LPCCA)<sup>[10]</sup> 以近邻样本之间的相关取代了单一的样本一一对应相关, 在数据可视化与姿态估计等实验中的结果比 CCA 与 KCCA 更加准确可靠. 子模式典型相关分析 (Sub-pattern CCA, SpCCA)<sup>[11]</sup> 将局部与全局特征矢量之间的相关特性作为有效判据信息, 既达到了融合局部和全局信息的目的, 又消除了特征之间的冗余. 传统的 CCA 在没有类标签矩阵的指导下属于非监督学习方法, 会出现识别性能不足的问题. Sun 等提出了一种鉴别型 CCA (Discriminant CCA, DCCA)<sup>[12]</sup>, 不仅考虑了同一模式的两组对应特征之间的相关, 同时也充分考虑了同类特征与非同类特征之间的相关及其对分类问题的影响. 用 DCCA 进行特征抽取, 能够实现类内相关的最大化与类间相关的最小化, 通常比 CCA 具有更好的识别性能. 但是, DCCA 抽取的最大特征维数由于受到类别数目的限制, 可能会出现投影轴数目不足的问题, 因而在一定程度上抑制了算法的应用. 广义典型相关分析 (Generalized CCA, GCCA)<sup>[13]</sup> 将最小化类内散布矩阵作为目标函数之一, 降低了特征的类内离散度, 提高了特征表示的鉴别能力. 半监督的典型相关分析 (Semi-CCA)<sup>[14]</sup> 引入成对约束的形式, 利用少量的监督信息来提高分类性能, 并在字符和人脸识别实验中得到了验证. 最近, Zhang 等提出的随机相关集成方法 (Random correlation ensemble, RCE)<sup>[15]</sup> 利用随机理论中的 bagging 和 boosting 策略, 提取类内样本不同视角间的随机相关系数, 以提高特征表示的鉴别能力.

近年来, 基于稀疏表示的特征降维和分类算法在各领域得到快速发展. 稀疏表示分类 (Sparse representation-based classification, SRC)<sup>[16]</sup> 作为一种新型的监督分类器, 能有效处理人脸识别实验中的遮挡、污损等问题, 并对特征的选取具有一定的鲁棒性. 稀疏保持投影 (Sparsity preserving projections, SPP)<sup>[17]</sup> 将稀疏表示中系数表示的稀疏性作为一种自然鉴别信息引入到特征提取问题中, 在常用的人脸数

数据集上验证了其可行性与有效性。Fan 等<sup>[18]</sup>在流形正则化的半监督分类框架下,以数据的稀疏正则化替代原来的邻接图,提出了稀疏正则化的最小二乘分类(Sparse regularized least square classification, S-RLSC)。利用 CCA 与最小二乘在特定情况下的等价性, Sun 等<sup>[19]</sup>提出了一种最小二乘典型相关分析(Least squares CCA, LS-CCA),并引入 Lasso 算法<sup>[20]</sup>,获取稀疏的投影向量,在基因表达谱数据分析等多标签分类问题中效果优于常规的 CCA。然而 LS-CCA 需要类标签矩阵的介入,无法实现两组数据的稀疏特征提取。

典型相关分析通过融合两组特征,提高了特征表示的鉴别性和鲁棒性,增强了算法的稳定性能。然而面对分类问题,如何利用样本的分布状况,提高算法的识别效果,一直以来都是模式识别等领域研究的热点问题。LPCCA 考虑了样本的近邻分布信息,有效保持了数据的局部欧氏结构,因而在数据可视化和姿态估计等应用中获得了良好的效果。但是欧氏距离在图像识别等问题中并不能很好地度量高维样本之间的相似度。受上述稀疏算法的启发,本文在 CCA 的框架内引入稀疏保持项,提出了稀疏保持典型相关分析(Sparsity preserving CCA, SPCCA),不仅实现了两组特征集鉴别信息的有效融合,同时对特征间的稀疏重构性加以约束,增强特征的表示和鉴别能力。该方法首先通过稀疏表示学习各特征集内样本之间的稀疏重构系数,依此构造稀疏重构权重矩阵,然后要求算法提取的两组特征不但满足最大相关性的要求,而且能够保证各特征集内稀疏重构关系作为一种结构信息得以保持,提高在复杂环境下算法的稳定性和鲁棒性。

本文第 1 节对 CCA 和 SPP 算法作简单介绍。第 2 节详细描述了 SPCCA 及其算法流程。第 3 节通过实验验证所提算法的性能,并对实验结果进行分析。最后对本文的工作进行总结,并展望进一步的工作。

## 1 相关工作

### 1.1 典型相关分析

给定  $N$  对已中心化的样本  $\{\mathbf{x}_i, \mathbf{y}_i\} (i = 1, \dots, N)$ , 令  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$  与  $Y = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N]$ , CCA 的目标在于寻找一对投影向量  $\mathbf{w}$  和  $\mathbf{u}$ , 满足  $\mathbf{z}_1 = \mathbf{w}^T X$  和  $\mathbf{z}_2 = \mathbf{u}^T Y$  之间具有最大的相关系数, 其判别函数为

$$\arg \max_{\mathbf{w}, \mathbf{u}} \rho = \frac{\mathbf{w}^T S_{xy} \mathbf{u}}{\sqrt{\mathbf{w}^T S_{xx} \mathbf{w}} \times \sqrt{\mathbf{u}^T S_{yy} \mathbf{u}}} \quad (1)$$

其中,  $S_{xx}$ ,  $S_{yy}$  分别表示特征集  $X$  和  $Y$  的协方差矩阵,  $S_{xy}$  表示  $X$  和  $Y$  的互协方差矩阵。采用 Lagrange 乘法子, 求解上述优化函数, 得到如下的广义特征方程:

$$\begin{pmatrix} 0 & S_{xy} \\ S_{xy}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{u} \end{pmatrix} = \lambda \begin{pmatrix} S_{xx} & 0 \\ 0 & S_{yy} \end{pmatrix} \begin{pmatrix} \mathbf{w} \\ \mathbf{u} \end{pmatrix} \quad (2)$$

其中,  $\lambda$  表示求取的两组成分  $\mathbf{z}_1$  和  $\mathbf{z}_2$  之间的相关系数, 也是作为选取投影轴的重要依据。

### 1.2 稀疏保持投影

首先给出某一样本集  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{m \times N}$ , 则对每一个样本  $\mathbf{x}_i$ , 都存在一个对应的子样本集  $M_i = [\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_N] \in \mathbf{R}^{m \times (N-1)}$ 。基于 1 范式稀疏表示的目标就是寻找一组系数  $\boldsymbol{\alpha}_i \in \mathbf{R}^{N-1}$ , 满足  $\mathbf{x}_i = M_i \boldsymbol{\alpha}_i$ , 并且要求  $\boldsymbol{\alpha}_i$  中系数的绝对值之和尽可能小, 用数学形式表示

如下:

$$\begin{aligned} \min \quad & \|\boldsymbol{\alpha}_i\|_1 \\ \text{s.t.} \quad & \mathbf{x}_i = M_i \boldsymbol{\alpha}_i \end{aligned} \quad (3)$$

在实际应用中, 式 (3) 中的限制条件一般很难成立, 因此在上述优化模型中考虑样本的重构误差, 提出一种稳定的稀疏表示, 表示如下:

$$\begin{aligned} \min \quad & \|[\boldsymbol{\alpha}_i; \mathbf{t}_i]\|_1 \\ \text{s.t.} \quad & \mathbf{x}_i = \begin{bmatrix} M_i & I \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i \\ \mathbf{t}_i \end{bmatrix} \end{aligned} \quad (4)$$

其中,  $I \in \mathbf{R}^{m \times m}$  表示与特征维数同维的单位矩阵。求解上述两种优化模型可以转化为线性规划和二阶锥规划等凸规划问题, 代表性的算法有:  $l_1$ -magic<sup>1</sup>, SPGL1<sup>2</sup> 和 SLEP<sup>3</sup> 等。本文以时间复杂度为选择依据, 采用 SPGL1 算法。当获取样本  $\mathbf{x}_i$  在对应子集  $M_i$  上的稀疏表示系数  $\boldsymbol{\alpha}_i$  后, 据此构造当前的稀疏重构权重矩阵  $S = [\mathbf{s}_1, \dots, \mathbf{s}_N]$ , 其中,  $\mathbf{s}_i = [\alpha_{i,1}, \dots, \alpha_{i,i-1}, 0, \alpha_{i,i}, \dots, \alpha_{i,N-1}]^T$ ,  $\alpha_{i,j}$  表示  $\boldsymbol{\alpha}_i$  的第  $j$  个系数。稀疏表示由于能够在无监督信息指导的情况下, 赋予被表示样本具有全局结构信息的样本(如同类样本)以更多的权重<sup>[16]</sup>。因此, 这种样本间的稀疏重构关系可以作为度量特征表示与鉴别能力的一种重要衡量指标。设降维后的特征为  $\mathbf{z} = \mathbf{w}^T X$ , 则此时的稀疏重构误差  $\|\mathbf{z} - \mathbf{z}S\|^2 = \|\mathbf{w}^T X - \mathbf{w}^T X S\|^2 = \mathbf{w}^T X (I - S)(I - S)^T X^T \mathbf{w}$  表征了投影轴  $\mathbf{w}$  的稀疏保持能力。结合以上分析, 可以得出稀疏保持投影的目标函数为

$$\max_{\mathbf{w}} \frac{\mathbf{w}^T X (S + S^T - SS^T) X^T \mathbf{w}}{\mathbf{w}^T X X^T \mathbf{w}} \quad (5)$$

优化上述模型, 令  $S_\beta = S + S^T - SS^T$ , 可以得到 SPP 的最优投影矩阵  $W = [\mathbf{w}_1, \dots, \mathbf{w}_d]$  是广义特征方程  $X S_\beta X^T \mathbf{w} = \lambda X X^T \mathbf{w}$  的前  $d$  个最大广义特征值对应的广义特征向量。

## 2 稀疏保持典型相关分析

稀疏保持典型相关分析的基本思想是首先利用稀疏表示算法获取样本间的全局稀疏重构权重, 并将此视作数据的一种自然鉴别信息, 然后利用优化策略将其融入 CCA 的理论框架中, 实现特征的鉴别融合。

给定同一模式的两组特征集表示  $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{m \times N}$  和  $Y = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbf{R}^{n \times N}$ , 采用 SPGL1 算法, 分别构造样本  $\mathbf{x}_i$  和  $\mathbf{y}_i$  在相应子样本集上的稀疏表示  $\mathbf{r}_i$  和  $\mathbf{s}_i$ , 据此可以得到两组特征集上的稀疏重构权重矩阵  $R = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N] \in \mathbf{R}^{m \times N}$  和  $S = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \in \mathbf{R}^{n \times N}$ 。SPCCA 旨在寻找两组投影向量  $\mathbf{w}$  和  $\mathbf{u}$ , 满足提取后的两组特征之间不仅具有最大的相关系数, 同时要求投影后两特征集内的稀疏重构误差尽可能小。综合以上分析, SPCCA 的鉴别函数可以表示为如下的多目标优化问题:

$$\begin{cases} \max_{\mathbf{w}, \mathbf{u}} & \mathbf{w}^T S_{xy} \mathbf{u} \\ \min_{\mathbf{w} \neq 0} & \mathbf{w}^T X (I - R)(I - R)^T X^T \mathbf{w} \\ \min_{\mathbf{u} \neq 0} & \mathbf{u}^T Y (I - S)(I - S)^T Y^T \mathbf{u} \end{cases} \quad (6)$$

<sup>1</sup> www.acm.caltech.edu/l1magic

<sup>2</sup> www.cs.ubc.ca/labs/scl/spgl1/index.html

<sup>3</sup> www.public.asu.edu/~jye02/Software/SLEP/index.htm

$S_{xy}$  的定义同上, 表示  $X$  和  $Y$  的互协方差矩阵, 描述了两组变量之间的相关性. 由于多目标优化问题 (6) 无法直接求解, 本文运用评价函数法<sup>[21]</sup> 中的乘除法将模型 (6) 转化为以下的单目标优化模型:

$$\arg \max_{\mathbf{w}, \mathbf{u} \neq 0} \rho = \frac{\mathbf{w}^T S_{xy} \mathbf{u}}{\sqrt{\mathbf{w}^T \tilde{S}_{xx} \mathbf{w}} \times \sqrt{\mathbf{u}^T \tilde{S}_{yy} \mathbf{u}}} \quad (7)$$

此模型中  $\tilde{S}_{xx} = X(I - R)(I - R)^T X^T$  和  $\tilde{S}_{yy} = Y(I - S)(I - S)^T Y^T$  分别为特征集  $X$  和  $Y$  的稀疏保持散度矩阵, 需要同时满足矩阵的正定性, 否则可能出现分母为 0 而导致模型退化的情况. 从定义已知  $\tilde{S}_{xx}$  和  $\tilde{S}_{yy}$  都是实对称半正定矩阵. 因此, 当  $\tilde{S}_{xx}$  和  $\tilde{S}_{yy}$  出现奇异时, 借鉴扰动分析理论, 在稀疏保持散度矩阵上加一个很小的扰动<sup>[22]</sup>, 以保证  $\tilde{S}_{xx}$  和  $\tilde{S}_{yy}$  的正定性和模型 (7) 的良好性. 另外,  $\rho$  的极大化与投影向量  $\mathbf{w}$  和  $\mathbf{u}$  的模值无关, 仅决定于其方向的选择. 因此采用 CCA 模型优化中经常使用的 Lagrange 乘子法, 可以将问题转化为如下的两个广义本征方程特征分解问题:

$$\begin{cases} S_{xy} \tilde{S}_{yy}^{-1} S_{xy}^T \mathbf{w} = \lambda^2 \tilde{S}_{xx} \mathbf{w} \\ S_{xy}^T \tilde{S}_{xx}^{-1} S_{xy} \mathbf{u} = \lambda^2 \tilde{S}_{yy} \mathbf{u} \end{cases} \quad (8)$$

并且  $w$  与  $u$  之间满足

$$\begin{cases} S_{xy} \mathbf{u} = \lambda \tilde{S}_{xx} \mathbf{w} \\ S_{xy}^T \mathbf{w} = \lambda \tilde{S}_{yy} \mathbf{u} \end{cases} \quad (9)$$

模型 (7) 的优化并不需要同时求解式 (8) 中的两个广义特征分解问题, 可以选取求解其中具有维数  $\min(m, n)$  的广义特征方程的分解问题获取其中一组投影向量, 然后利用式 (9) 求解另一组投影向量, 从而降低算法的时间复杂度. 当获取前  $d$  组特征向量  $\mathbf{w}_i, \mathbf{u}_i (i = 1, 2, \dots, d)$  后, 构造基向量矩阵  $\{W, U\}$ , 其中  $W = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d] \in \mathbf{R}^{m \times d}$ ,  $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d] \in \mathbf{R}^{n \times d}$ , 并且有如下定理成立.

**定理 1.** 投影矩阵  $W$  和  $U$  满足

$$\begin{cases} \mathbf{w}_i^T \tilde{S}_{xx} \mathbf{w}_j = \delta_{i,j} \\ \mathbf{u}_i^T \tilde{S}_{yy} \mathbf{u}_j = \delta_{i,j} \\ \mathbf{w}_i^T S_{xy} \mathbf{u}_j = \lambda_i \delta_{i,j} \end{cases} \quad (10)$$

其中,  $\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} (i, j = 1, 2, \dots, d)$ .

**证明.** 因为  $\tilde{S}_{xx}$  和  $\tilde{S}_{yy}$  都是实对称的正定矩阵, 采用 Cholesky 分解, 则式 (8) 中的广义特征方程可以改写为式 (11). 由  $(\tilde{S}_{yy}^{-\frac{1}{2}} S_{xy}^T \tilde{S}_{xx}^{-\frac{1}{2}}) (\tilde{S}_{xx}^{-\frac{1}{2}} S_{xy} \tilde{S}_{yy}^{-\frac{1}{2}})^T (\tilde{S}_{xx}^{\frac{1}{2}} \mathbf{w}, \tilde{S}_{yy}^{\frac{1}{2}} \mathbf{u})$  和  $\lambda$  分别是矩阵  $\tilde{S}_{yy}^{-\frac{1}{2}} S_{xy}^T \tilde{S}_{xx}^{-\frac{1}{2}}$  的左右奇异向量和奇异值. 由奇异向量之间的归一正交性及式 (9), 得

$$\begin{cases} (\tilde{S}_{xx}^{-\frac{1}{2}} S_{xy} \tilde{S}_{yy}^{-\frac{1}{2}}) (\tilde{S}_{yy}^{-\frac{1}{2}} S_{xy}^T \tilde{S}_{xx}^{-\frac{1}{2}}) (\tilde{S}_{xx}^{\frac{1}{2}} \mathbf{w}) = \lambda^2 (\tilde{S}_{xx}^{\frac{1}{2}} \mathbf{w}) \\ (\tilde{S}_{yy}^{-\frac{1}{2}} S_{xy}^T \tilde{S}_{xx}^{-\frac{1}{2}}) (\tilde{S}_{xx}^{-\frac{1}{2}} S_{xy} \tilde{S}_{yy}^{-\frac{1}{2}}) (\tilde{S}_{yy}^{\frac{1}{2}} \mathbf{u}) = \lambda^2 (\tilde{S}_{yy}^{\frac{1}{2}} \mathbf{u}) \end{cases} \quad (11)$$

$$\begin{cases} (\tilde{S}_{xx}^{\frac{1}{2}} \mathbf{w}_i)^T (\tilde{S}_{xx}^{\frac{1}{2}} \mathbf{w}_j) = \mathbf{w}_i^T \tilde{S}_{xx} \mathbf{w}_j = \delta_{i,j} \\ (\tilde{S}_{yy}^{\frac{1}{2}} \mathbf{u}_i)^T (\tilde{S}_{yy}^{\frac{1}{2}} \mathbf{u}_j) = \mathbf{u}_i^T \tilde{S}_{yy} \mathbf{u}_j = \delta_{i,j} \\ \mathbf{w}_i^T S_{xy} \mathbf{u}_j = \mathbf{w}_i^T \times (\lambda_j \tilde{S}_{xx} \mathbf{w}_j) = \lambda_j \mathbf{w}_i^T \tilde{S}_{xx} \mathbf{w}_j = \lambda_j \delta_{i,j} \end{cases} \quad (12)$$

□

获取投影矩阵  $W$  和  $U$  后, 文献 [22] 介绍了三种融合策略 (并行组合、串行组合和基于特征矩阵的方式), 并验证了后两种融合方式分类性能相当, 稍优于第一种融合方式. 本文在此处采用第三种策略, 即按式 (13) 对样本的两组原特征  $\{\mathbf{x}, \mathbf{y}\}$  进行特征提取并做融合处理.

$$M = [\mathbf{z}_1 \ \mathbf{z}_2] = [W^T \mathbf{x} \ U^T \mathbf{y}] \quad (13)$$

假设  $\zeta_1, \zeta_2, \dots, \zeta_N$  是所有的训练样本, 与之对应的相关特征矩阵分别是  $M_1, M_2, \dots, M_N$ . 对任意的测试样本  $\zeta$ , 其相关特征矩阵是  $M = [\mathbf{z}_1 \ \mathbf{z}_2]$ . 如果  $d(M, M_i) = \min_j d(M, M_j)$ , 并且  $M_i \in \omega_k (\zeta_i \in \omega_k)$ , 则  $M \in \omega_k (\zeta \in \omega_k)$ , 其中  $d(\cdot, \cdot)$  表示欧氏距离函数,  $\omega_k$  表示第  $k$  类特征样本集合.

综上所述, SPCCA 的算法流程可总结如下:

**步骤 1.** 输入两组已标准化的训练特征集  $X$  和  $Y$ ;

**步骤 2.** 根据式 (4) 构造特征集  $X$  和  $Y$  的稀疏重构权重矩阵  $R$  和  $S$ ;

**步骤 3.** 依据模型 (7) 构造特征集  $X$  和  $Y$  的稀疏保持散度矩阵  $\tilde{S}_{xx}$  和  $\tilde{S}_{yy}$ ;

**步骤 4.** 按照式 (8) 和式 (9) 计算投影矩阵  $W$  和  $U$ ;

**步骤 5.** 利用式 (13) 计算并融合样本特征  $M_i (i = 1, \dots, N)$ , 采用最近邻分类器对测试样本进行分类.

### 3 实验分析

为了验证本文提出的 SPCCA 算法在特征融合中的有效性, 我们在两个多特征手写体数据集和常用的人脸数据库上进行了充分的实验, 比较了 SPCCA 和 SPP, CCA 等相关算法的分类识别性能.

#### 3.1 手写体识别实验

本实验首先采用多特征手写体数据集来检验 SPCCA 用于特征提取和融合的性能. MFEAT 数据集<sup>4</sup> 属于 UCI 机器学习知识库中的一个组成部分, 包含 0~9 共 10 个数字的 6 组特征数据, 每类 200 个样本, 共 2000 个样本. 这 6 组特征包括轮廓相关特征<sup>5</sup> (FAC, 216)、傅里叶系数 (FOU, 76)、K-L 展开系数 (KAR, 64)、形态学特征 (MOR, 6)、像素均值 (PIX, 240) 和 Zernike 矩 (ZER, 47). 表 1 列出了 SPP 算法在不同特征上的识别结果, 说明了不同特征之间用于识别的差异性.

实验中, 选取任意两个数据集作为  $X$  集和  $Y$  集, 共有 15 种组合方式. 对每种组合, 选取每类前 100 对样本作为训练集, 后 100 对样本作为测试集. 实验结果如表 2 所示. 另外, 本小节也列出了传统串行融合方式的实验结果 (将两组特征进行标准化后组合成一个高维向量), 并记作 Baseline. 从表 2 可以看出, Baseline 通过简单的组合处理, 能够获得不错的实验结果, 但是主要存在以下两点问题:

1) 特征表示的维数过多, 加剧小样本问题的出现, 并导致时间和空间复杂度高;

<sup>4</sup> <http://archive.ics.uci.edu/ml/datasets/Multiple+Features>

<sup>5</sup> 括弧中依次为其特征的英文缩写与特征维数

2) 不同的特征表示通常具有不同的数据表示及物理含义, 简单的组合方式可能会忽略和丢失某些重要变量, 如组合 *c7* 中 Baseline 方法的实验结果出现了严重的退化.

CCA 能以有效的方式融合两组特征, 减少特征表示维数的同时改善基于单特征集的识别性能, 然而在面对两组特征集之间维数具有明显差异时, 容易出现由于抽取的特征维数不足而导致的性能下降问题, 如组合 *c3*. SPCCA 在特征融合时考虑了样本的稀疏重构关系, 降低了特征维数的敏感性, 在一定程度上弱化了 CCA 的上述缺点. 因此除组合 *c2* 略低于 CCA 及组合 *c3* 略低于 Baseline 方法以外, SPCCA 在其余 13 种组合中均能取得最好的识别结果, 明显优于 CCA. 另外, 与表 1 中 SPP 的实验结果相比, SPCCA 明显地消除了由于特征选择的差异性而导致的识别结果的不稳定, 在大多数组合中的识别率超过 95%. 这表明在面对多特征识别时, SPCCA 既能融合具有互补性的特征, 提高特征鉴别能力, 也能降低具有较低鉴别能力的特征对整个系统的影响, 增强多特征系统对于复杂环境的鲁棒性.

为了增强实验结果的说理力, 表 2 同时列出了各组合中两组不同特征分别利用 SPP 算法进行分类器组合时的成对差异性度量. 文献 [23] 详细介绍了一些差异性度量方法以及之间的相互关系, 本文采用常用的 Kappa 统计 (评价内部一致性) 度量两组特征之间的差异性. 从实验结果可以看出  $\kappa$  的值越小说明特征之间的差异性越大, 从而利用 SPCCA 算法进行特征融合后的实验结果越好, 如在组合 *c10* 和 *c13* 中 SPCCA 的实验结果要明显优于其他两种方法; 而在组合 *c2*, *c4* 和 *c11* 中  $\kappa$  值明显偏大, 表明两组特征之间存在较大的冗

余, 因此 SPCCA 的实验结果与 CCA 相比并没有明显的改善. 这一现象为 SPCCA 的应用场合选择提供了良好的参考依据.

在下一个实验中, 采用 CENPARMI 手写体数据集, 该数据集包含 0~9 共 10 类数字样本, 每类包含训练样本 400 个, 测试样本 200 个. 文献 [24] 通过对数字图像进行预处理, 提取出如下四组图像特征: Gabor 特征 (GAB, 256)、Legendre 特征 (LEG, 121)、Pseudo-Zernike 特征 (PSE, 36) 和 Zernike 特征 (ZNK, 30). SPP 算法在各特征的实验结果见表 1, 同样可以看出不同特征之间性能的差异性. 将上述特征两两组合分别进行融合与分类的实验结果见表 3.

在此数据集上, 用 CCA 进行特征融合的分类结果普遍劣于 Baseline 方法, 这主要由于 Baseline 没有特征降维的过程, 从而保留了所有的可用信息. 然而由于 Baseline 仅采用简单的特征组合方式, 所以在基于两组特征的融合识别时结果要弱于表 1 中 SPP 基于单组特征的最好实验结果. SPCCA 集中了稀疏重构和特征融合的双重优点, 其识别性能优于 SPP, Baseline 和 CCA, 并且只需要很少的维数便可达到最大识别率, 有利于减少分类时间的消耗和空间存储需求. 类似于上一个实验, 同样地采用  $\kappa$  参数度量两组特征在利用 SPP 算法进行分类时的差异性. 从表 3 可以看出, 组合 *c2* 和 *c3* 的  $\kappa$  明显低于其他组合, SPCCA 的融合结果明显优于未融合时的结果以及其他两种融合方法的结果; 然而在组合 *c6* 中  $\kappa$  明显高于其他组合, 利用 SPCCA 进行特征融合的效果较其他方法并无明显的改善.

表 1 SPP 算法在两组多特征手写体数据集上的识别率 (相应的特征维数)

Table 1 Recognition rates of SPP in two multiple feature databases (corresponding dimensionality)

	MFEAT						CENPARMI			
	FAC	FOU	KAR	MOR	PIX	ZER	GAB	LEG	PSE	ZNK
SPP	0.949 (45)	0.841 (25)	0.971 (44)	0.601 (5)	0.978 (38)	0.815 (18)	0.884 (40)	0.936 (57)	0.757 (34)	0.750 (30)

表 2 多特征手写体数据集 MFEAT 上的识别率 (相应的特征维数)

Table 2 Recognition rates in multiple feature database MFEAT (corresponding dimensionality)

<i>c</i>	Combination of datasets		$\kappa$	Recognition rates		
	<i>X</i> - <i>Y</i>			Baseline	CCA	SPCCA
1	FAC-FOU		0.030	0.948 (292)	0.880 (14)	0.980 (16)
2	FAC-KAR		0.473	0.948 (280)	0.972 (23)	0.970 (26)
3	FAC-MOR		-0.142	0.934 (222)	0.763 (6)	0.928 (6)
4	FAC-PIX		0.411	0.949 (456)	0.958 (20)	0.970 (25)
5	FAC-ZER		0.086	0.961 (263)	0.864 (37)	0.965 (43)
6	FOU-KAR		0.014	0.967 (142)	0.928 (18)	0.981 (23)
7	FOU-MOR		0.105	0.425 (82)	0.752 (6)	0.821 (6)
8	FOU-PIX		0.022	0.972 (316)	0.851 (21)	0.977 (26)
9	FOU-ZER		0.396	0.804 (123)	0.833 (20)	0.872 (19)
10	KAR-MOR		-0.171	0.697 (70)	0.834 (6)	0.951 (6)
11	KAR-PIX		0.658	0.973 (304)	0.971 (22)	0.976 (49)
12	KAR-ZER		0.058	0.818 (114)	0.884 (29)	0.964 (35)
13	MOR-PIX		-0.170	0.778 (246)	0.739 (6)	0.959 (6)
14	MOR-ZER		0.202	0.783 (53)	0.743 (6)	0.787 (5)
15	PIX-ZER		0.0678	0.829 (287)	0.831 (26)	0.962 (21)

### 3.2 人脸识别实验

分别在 ORL, Yale 和 AR 三个人脸数据库上进行人脸识别实验, 并与 CCA, SPP 进行对比. 在实验中, 将图像进行 Daubechies 正交小波分解获取低频分量作为样本的第二特征集, 然后对两组特征进行 PCA 降维, 以增强算法的稳定性和消除小样本问题的出现. 同时, 为了增强实验比较性, 实验中还实现了将两组特征串联后再采用 SPP 的方法, 记为 SPP\*.

ORL 人脸数据库<sup>[25]</sup> 包括 40 个人的 10 幅人脸图像, 每人 10 幅图像, 分别拍摄于不同时间和光照条件下, 具有不同的表情和面部细节. 随机选择每人的 4 幅或 5 幅图像进行训练, 其余的 6 幅或 5 幅用作测试. 实验中 PCA 降维到 42 维. Yale 人脸数据库<sup>[26]</sup> 包含 15 个人的 165 幅灰度图像, 每人 11 幅, 包括光照方向、眼睛、表情变化. 选取每人的 4 幅或 5 幅图像作为训练, 其余 7 幅或 6 幅用作测试. 实验中 PCA 降维到 48 维. AR 人脸数据库<sup>[27]</sup> 由 126 个人 (男 70 人, 女 56 人) 的 4000 余幅彩色图像构成, 每人 26 张图像, 分为 2 组, 每组 13 张, 拍摄时间间隔为 2 周, 分别反映了人脸的表情、光照和遮挡的变化. 在本文中, 我们选取每人的 14 张无遮挡灰度图像进行人脸识别实验. 选取每人的 5 幅或 7 幅图像作为训练, 其余 9 幅或 7 幅用作测试. 由于类别数较多, 实验中 PCA 降维到 256 维.

随机实验分别独立进行 10 次. 实验结果见表 4 和图 1 所示. 文献 [17] 中已验证 SPP 的实验结果要好于传统的 PCA, LPP<sup>[28]</sup> 和 NPE<sup>[29]</sup> 算法, 因此本文仅列出了 SPP 的实验结果. 分析上述实验结果, 可以得出如下的结论:

1) SPP 算法借鉴了稀疏表示中的稀疏重构信息, 因而在常用的人脸识别数据集中优于经典的特征提取方法<sup>[17]</sup>. 本实验中, SPP 在 ORL 数据集中能够获得比 CCA 更好的识别结果. 在 Yale 数据集中, 当训练样本为 4 时识别率与 CCA 相差无几, 但是当训练样本为 5 时, 识别效果要明显地低于 CCA 和 SPCCA; 特别地, 在 AR 数据集中这种现象尤为明显, 这体现了多特征融合在复杂环境下相对于单特征系统的

优越性.

2) SPP\* 采用简单的组合方式实现了特征的融合处理, 然而这一方式忽略了不同组特征之间的相互关系, 导致在计算稀疏重构系数时增加了估计的不准确性, 同时算法的时间和空间复杂度也急剧上升. 从表 4 的实验结果可以得到当两组特征的维数都较少时, 这种组合方式可以略微提高算法的识别性能. 但是在 AR 数据集上当两组特征的维数都偏高时, SPP\* 的实验性能反而略低于基于第一组特征的 SPP 方法. 由于串行组合方式没有考虑两组特征之间的相关性, 因而在三组人脸数据库上的识别性能都低于本文算法, 并且相应的特征维数要明显高于其他方法. 鉴于以上考虑, 同时为了增强图示的可视性, 在图 1 中并没有列出 SPP\* 的识别率随维数变化的示意图.

3) CCA 在大多数情况下都能够以较少的维数达到稳定的识别率, 并且满足随着维数增加而保持在一个稳定的水平. 但是在如 AR 等类别个数较多的人脸数据库中, 识别结果会随着维数的增加在达到一个峰值之后, 随着维数的增加反而引起识别效果的大幅度下降. 这主要是因为 CCA 在抽取低维主要特征后 (较大特征值所对应的投影向量), 由于没有监督信息的指引, 高维特征包含的含噪和不稳定信息影响了低维特征的鉴别能力, 从而导致识别性能下降.

4) SPCCA 在 CCA 的框架内, 同时考虑了两组特征内的稀疏重构权重矩阵, 增强了特征融合的鉴别性, 在人脸数据集上的实验结果一致地优于 SPP, SPP\* 和 CCA. 从 AR 人脸数据集上识别率随维数变化的情况, 可以看出本文方法并没有出现 CCA 在选取高维特征时识别率剧烈下降的问题, 说明了 SPCCA 能够有效抑制高维特征中存在的噪声信息, 提高了算法提取鉴别信息的稳定性.

## 4 总结与展望

在典型相关分析的框架内, 本文结合了稀疏表示的思想, 提出了一种新的特征提取和融合技术 — 稀疏保持典型相关

表 3 多特征手写体数据集 CENPARMI 上的识别率 (相应的特征维数)

Table 3 Recognition rates in multiple feature database CENPARMI (corresponding dimensionality)

Combination of datasets			Recognition rates		
$c$	$X - Y$	$\kappa$	Baseline	CCA	SPCCA
1	GAB-LEG	0.137	0.905 (377)	0.899 (23)	0.941 (19)
2	GAB-PSE	0.096	0.887 (292)	0.833 (17)	0.900 (22)
3	GAB-ZNK	0.093	0.893 (286)	0.830 (16)	0.912 (26)
4	LEG-PSE	0.125	0.917 (157)	0.868 (20)	0.934 (21)
5	LEG-ZNK	0.168	0.920 (151)	0.862 (20)	0.931 (27)
6	PSE-ZNK	0.523	0.777 (66)	0.769 (30)	0.783 (30)

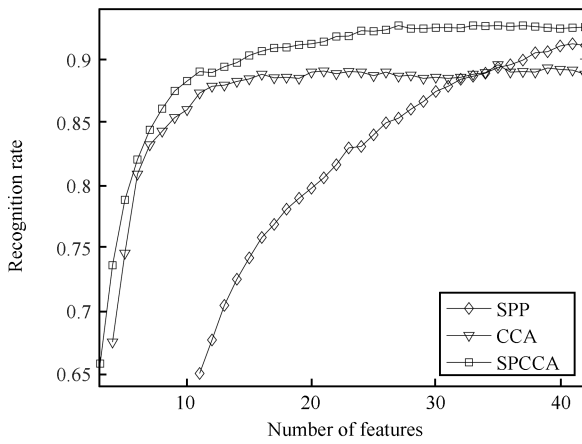
表 4 在各人脸数据集上的最优识别率比较 (相应的特征维数)

Table 4 Comparison of best recognition rates in each face dataset (corresponding dimensionality)

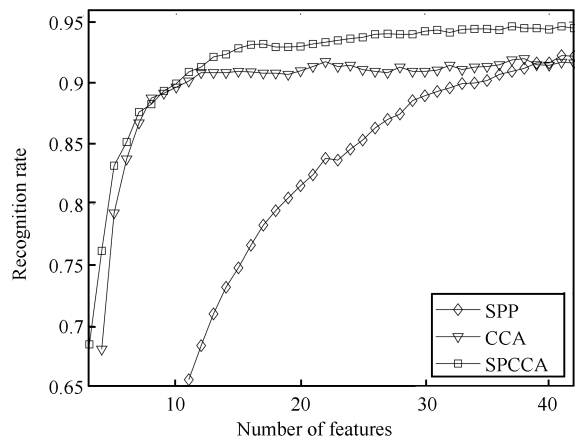
Dataset		SPP	SPP*	CCA	SPCCA
ORL	4 trainings	0.9121 (41)	0.9179 (72)	0.8954 (35)	0.9267 (37)
	5 trainings	0.9225 (41)	0.9385 (82)	0.9200 (38)	0.9465 (37)
Yale	4 trainings	0.7648 (48)	0.7733 (58)	0.7686 (34)	0.8286 (46)
	5 trainings	0.7800 (48)	0.7933 (73)	0.8622 (34)	0.8911 (47)
AR	5 trainings	0.8364 (256)	0.8129 (403)	0.9370 (86)	0.9450 (242)
	7 trainings	0.8871 (256)	0.8780 (398)	0.9683 (71)	0.9762 (167)

分析 (SPCCA). SPCCA 由于考虑了样本间的稀疏重构性, 因而能够在利用 CCA 进行特征提取和融合时无监督的方式引入一种自然鉴别信息, 并在字符和人脸数据集中验证了

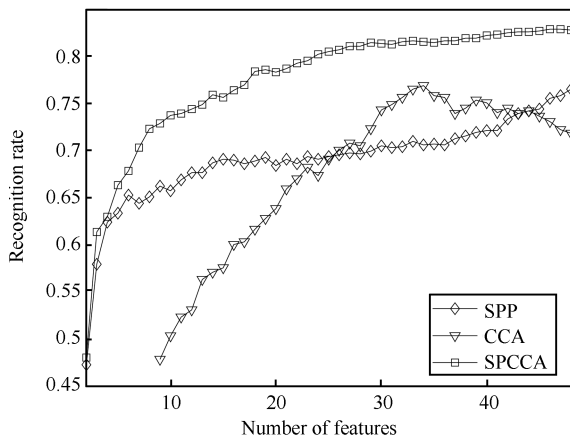
算法的可行性. 然而面对现实生活中大量的非线性问题, 如何利用稀疏表示的有关技术研究和解决非线性的和稀疏的 CCA 将是以后研究的重点内容.



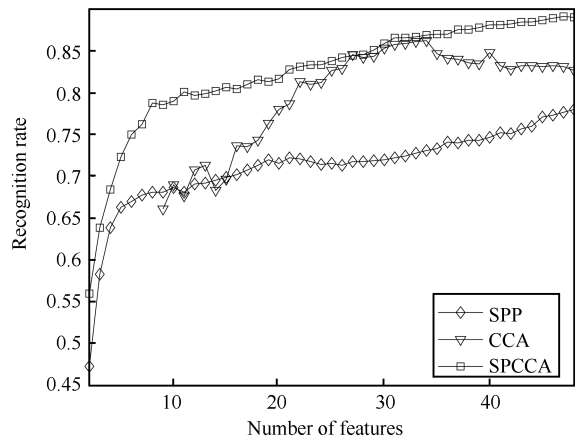
(a) 每个人 4 个训练样本 (ORL-4)  
(a) 4 trainings per person (ORL-4)



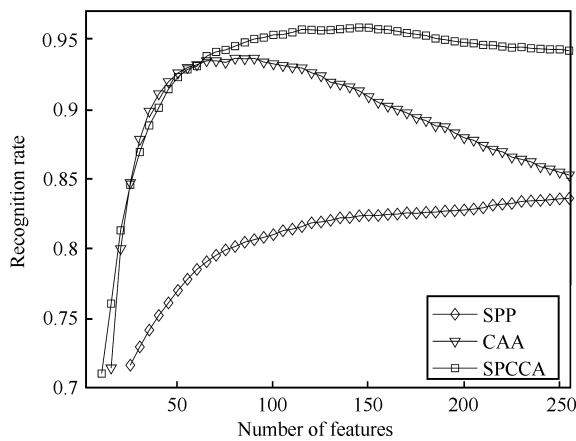
(b) 每个人 5 个训练样本 (ORL-5)  
(b) 5 trainings per person (ORL-5)



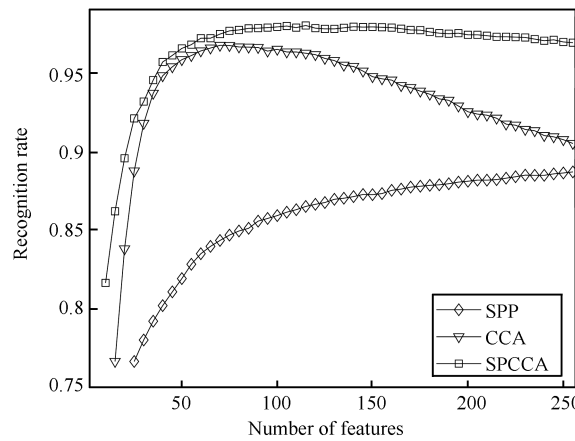
(c) 每个人 4 个训练样本 (Yale-4)  
(c) 4 trainings per person (Yale-4)



(d) 每个人 5 个训练样本 (Yale-5)  
(d) 5 trainings per person (Yale-5)



(e) 每个人 5 个训练样本 (AR-5)  
(e) 5 trainings per person (AR-5)



(f) 每个人 7 个训练样本 (AR-7)  
(f) 7 trainings per person (AR-7)

图 1 三种算法在各人脸数据集中识别率随维数变化示意图

Fig. 1 The recognition rates of the three algorithms on the face databases

## References

- 1 Jolliffe I T. *Principle Component Analysis*. New York: Springer, 1986
- 2 Barshan E, Ghodsi A, Azimifar Z, Jahromi M Z. Supervised principal component analysis: visualization, classification and regression on subspaces and submanifolds. *Pattern Recognition*, 2011, **44**(7): 1357–1371
- 3 Kyperountas M, Tefas A, Pitas I. Weighted piecewise LDA for solving the small sample size problem in face verification. *IEEE Transactions on Neural Networks*, 2007, **18**(2): 506–519
- 4 Lin Yu-E, Gu Guo-Chang, Liu Hai-Bo, Shen Jing, Zhao Jing. An orthogonal feature extraction method based on the within-class preserving for small sample size problem. *Acta Automatica Sinica*, 2010, **36**(5): 644–649  
(林玉娥, 顾国昌, 刘海波, 沈晶, 赵靖. 适用于小样本问题的具有类内保持的正交特征提取算法. *自动化学报*, 2010, **36**(5): 644–649)
- 5 Cheng Zheng-Dong, Zhang Yu-Jin, Fan Xiang, Zhu Bin. Study on discriminant matrices of commonly-used Fisher discriminant functions. *Acta Automatica Sinica*, 2010, **36**(10): 1361–1370  
(程正东, 章毓晋, 樊祥, 朱斌. 常用 Fisher 判别函数的判别矩阵研究. *自动化学报*, 2010, **36**(10): 1361–1370)
- 6 Wang Z, Chen S C, Sun T K. MultiK-MHKS: a novel multiple kernel learning algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(2): 348–353
- 7 Hotelling H. Relations between two sets of variates. *Biometrika*, 1936, **28**(3–4): 321–377
- 8 Sun Quan-Sen, Zeng Sheng-Gen, Heng Pheng-Ann, Xia De-Shen. The theory of canonical correlation analysis and its application to feature fusion. *Chinese Journal of Computers*, 2005, **28**(9): 1524–1533  
(孙权森, 曾生根, 王平安, 夏德深. 典型相关分析的理论及其在特征融合中的应用. *计算机学报*, 2005, **28**(9): 1524–1533)
- 9 Melzer T, Reiter M, Bischof H. Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 2003, **36**(9): 1961–1971
- 10 Sun T K, Chen S C. Locality preserving CCA with applications to data visualization and pose estimation. *Image and Vision Computing*, 2007, **25**(5): 531–543
- 11 Hong Quan, Chen Song-Can, Ni Xue-Lei. Sub-pattern canonical correlation analysis with application in face recognition. *Acta Automatica Sinica*, 2008, **34**(1): 21–30  
(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. *自动化学报*, 2008, **34**(1): 21–30)
- 12 Sun T K, Chen S C, Yang J Y, Shi P F. A novel method of combined feature extraction for recognition. In: *Proceedings of the 8th IEEE International Conference on Data Mining*. Pisa, Italy: IEEE, 2008. 1043–1048
- 13 Sun Q S, Liu Z D, Heng P A, Xia D S. A theorem on the generalized canonical projective vectors. *Pattern Recognition*, 2005, **38**(3): 449–452
- 14 Peng Yan, Zhang Dao-Qiang. Semi-supervised canonical correlation analysis algorithm. *Journal of Software*, 2008, **19**(11): 2822–2832  
(彭岩, 张道强. 半监督典型相关分析算法. *软件学报*, 2008, **19**(11): 2822–2832)
- 15 Zhang J C, Zhang D Q. A novel ensemble construction method for multi-view data using random cross-view correlation between within-class examples. *Pattern Recognition*, 2011, **44**(6): 1162–1171
- 16 Wright J, Yang A Y, Ganesh A, Sastry S S, Ma Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(2): 210–227
- 17 Qiao L S, Chen S C, Tan X Y. Sparsity preserving projections with applications to face recognition. *Pattern Recognition*, 2010, **43**(1): 331–341
- 18 Fan M Y, Gu N N, Qiao H, Zhang B. Sparse regularization for semi-supervised classification. *Pattern Recognition*, 2011, **44**(8): 1777–1784
- 19 Sun L, Ji S W, Ye J. Canonical correlation analysis for multilabel classification: a least squares formulation, extensions and analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(1): 194–200
- 20 Tibshirani R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, 1996, **58**(1): 267–288
- 21 Wei Quan-Ling, Wang Ri-Shuang, Xu Bing, Wang Jia-Yun, Bai Wen-Lin. *Mathematical Programming and Optimization Design*. Beijing: National Defense Industry Press, 1984  
(魏权龄, 王日爽, 徐冰, 汪家芸, 白文林. *数学规划和优化设计*. 北京: 国防工业出版社, 1984)
- 22 Sun Quan-Sen. *Research on Feature Extraction and Image Recognition Based on Correlation Projection Analysis* [Ph.D. dissertation], Nanjing University of Science and Technology, China, 2006  
(孙权森. *基于相关投影分析的特征抽取与图像识别研究* [博士学位论文], 南京理工大学, 中国, 2006)
- 23 Kuncheva L I, Whitaker C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 2003, **51**(2): 181–207
- 24 Hu Zhong-Shan, Lou Zhen, Yang Jing-Yu, Liu Ke, Sun Jing-Yi. Handwritten digit recognition basis on multi-classifier combination. *Chinese Journal of Computers*, 1999, **22**(4): 369–374  
(胡钟山, 娄震, 杨靖宇, 刘克, 孙靖夷. 基于多分类器组合的手写体数字识别. *计算机学报*, 1999, **22**(4): 369–374)
- 25 Samaria F S, Harter A C. Parameterisation of a stochastic model for human face identification. In: *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*. Sarasota, USA: IEEE, 1994. 138–142
- 26 Belhumeur P N, Hefanpha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711–720
- 27 Martinez A M, Benavente R. The AR Face Database, CVC Technical Report #24, Purdue University, USA, 1998
- 28 He X F, Niyogi P. Locality preserving projections. In: *Proceedings of the Advances in Neural Information Processing System*. Vancouver, Canada: MIT Press, 2003. 153–160
- 29 He X F, Cai D, Yan S C, Zhang H. Neighborhood preserving embedding. In: *Proceedings of the 10th IEEE International Conference on Computer Vision*. Beijing, China: IEEE, 2005. 1208–1213

侯书东 南京理工大学博士研究生. 主要研究方向为模式识别和图像处理. E-mail: hou\_shudong@hotmail.com  
(HOU Shu-Dong Ph.D. candidate at Nanjing University of Science and Technology. His research interest covers pattern recognition and image processing.)

孙权森 南京理工大学教授. 主要研究方向为模式识别, 图像处理与遥感信息处理. 本文通信作者. E-mail: qssun@126.com  
(SUN Quan-Sen Professor at Nanjing University of Science and Technology. His research interest covers pattern recognition, image processing, and remote sensing information processing. Corresponding author of this paper.)