

# 一种新的基于最大边缘准则的监督流形学习方法

袁 警<sup>1</sup> 杨瑞国<sup>2</sup> 原 媛<sup>2</sup> 雷迎科<sup>2</sup>

(合肥学院网络与智能信息处理重点实验室 合肥 230022)<sup>1</sup> (电子工程学院 合肥 230037)<sup>2</sup>

**摘 要** 在深入研究局部样条嵌入算法(LSE)的基础上,引入明确的线性映射关系,构建平移缩放模型和正变化特征子空间,提出了一种正交局部样条判别投影算法(O-LSDP),有效解决了原始LSE算法存在的两个主要问题:样本外点学习问题和无监督模式学习问题。该算法能够应用于模式分类问题并显著改善算法的分类识别能力。在标准人脸数据库上进行的实验比较分析验证了该算法的有效性与可行性。

**关键词** 特征提取,子空间学习,局部样条嵌入,最大边缘准则,流形学习

**中图法分类号** TP181 **文献标识码** A

## New Supervised Manifold Learning Method Based on MMC

YUAN Min<sup>1</sup> YANG Rui-guo<sup>2</sup> YUAN Yuan<sup>2</sup> LEI Ying-ke<sup>2</sup>

(Key Lab of Network and Intelligent Information Processing, Hefei University, Hefei 230022, China)<sup>1</sup>

(Electronic Engineering Institute, Hefei 230037, China)<sup>2</sup>

**Abstract** Based on the analysis of local spline embedding (LSE) method, we proposed an efficient feature extraction algorithm called orthogonal local spline discriminant projection (O-LSDP). By introducing an explicit linear mapping, constructing different translation and rescaling models for different classes as well as orthogonalizing feature subspace, O-LSDP can effectively circumvent the two major shortcomings of the original LSE algorithm, i. e., out-of-sample and unsupervised learning. O-LSDP not only inherits the advantages of LSE which uses local tangent space as a representation of the local geometry so as to preserve the local structure, but also makes full use of class information and orthogonal subspace to significantly improve discriminant power. Extensive experiments on standard face databases and plant leaf data set verify the feasibility and effectiveness of the proposed algorithm.

**Keywords** Feature extraction, Subspace learning, Local spline embedding, Maximum margin criterion, Manifold learning

## 1 引言

流形学习方法假定输入数据是嵌入在高维观测空间的低维流形上,其目的是找出高维数据中所隐藏的低维流形结构。经典的流形学习方法包括等距特征映射算法(ISOMAP)<sup>[1]</sup>、局部线性嵌入算法(LLE)<sup>[2,3]</sup>、Laplacian特征映射算法(LE)<sup>[4,5]</sup>、Hessian特征映射算法(HLLE)<sup>[6]</sup>、最大差异展开算法(MVU)<sup>[7,8]</sup>、局部切空间排列算法(LTSA)<sup>[9]</sup>、黎曼流形学习算法(RML)<sup>[10,11]</sup>和局部样条嵌入算法(LSE)<sup>[12,13]</sup>等。由于流形学习方法的非线性本质、几何直观性和计算可行性,使得它们在数据挖掘、机器学习、模式识别、图像分析和计算机视觉等研究领域受到了广泛的关注并得到了一些成功的应用。然而,当它们面临模式识别与分类任务时,不可避免地会暴露其自身的局限性:(1)无显式的映射关系,即仅仅产生一个隐含的非线性映射将训练样本集从高维输入空间映射到低维空间,而并不能产生一个明确的映射函数关系。因此,对于新的测试样本,无法获得其在低维空间的投影。为了使流形学习方法更好地应用于模式分类任务,研究人员开始尝试通

过线性化、核化和张量化扩展来提高流形学习的样本外点学习能力<sup>[14]</sup>。He等人对LE算法进行线性化处理,提出了局部保持投影算法(Locality Preserving Projection, LPP)<sup>[15]</sup>,其实质是通过对数据流形上的Laplacian-Beltrami算子的特征函数进行线性近似,从而获得高维输入空间到低维流形的显式映射关系。Yan等人在图嵌入框架理论下提出了边缘Fisher分析算法(Marginal Fisher Analysis, MFA)来解决流形学习的样本外点学习问题<sup>[14]</sup>。(2)无监督模式学习。原始的流形学习算法具有很好的维数约简能力,但因为是无监督模式学习,在实际应用中不能有效地利用样本的先验知识来提高算法的分类识别能力。如何运用已知的诸如样本的类别等先验信息,更好地处理模式识别与分类问题,是流形学习的一个重要研究方向。近年来,许多监督流形学习算法相继被提出来。Ridder和Pan等人分别使用融入了样本类别信息的欧氏距离和凸轮加权距离(Cam Weighted Distance)来提高LLE算法的判别能力,从而提出了监督局部线性嵌入算法(Supervised LLE, SLLE)<sup>[16]</sup>和加权局部线性嵌入算法(Weighted LLE, WLLE)<sup>[17]</sup>。

到稿日期:2013-06-01 返修日期:2013-08-27 本文受国家自然科学基金(61272333, 61273302, 61005010),安徽省自然科学基金(1208085MF94, 1208085MF98, 1308085MF84)资助。

袁 警(1956-),女,硕士,副教授,主要研究方向为智能信息处理,E-mail: hfuyuanmin@163.com;杨瑞国(1964-),男,讲师,主要研究方向为信号处理;原 媛(1983-),女,硕士,讲师,主要研究方向为机器学习;雷迎科(1975-),男,博士,副研究员,主要研究方向为机器学习。

受局部样条嵌入算法(LSE)<sup>[12,13]</sup>和最大边缘准则(Maximum Margin Criterion, MMC)<sup>[18]</sup>的启发,我们提出了一种新颖的监督流形学习算法——正交局部样条判别投影(Orthogonal Local Spline Discriminant Projection, O-LSDP)。O-LSDP算法从两个方面来提高原始LSE算法的分类识别能力:1)寻求一种线性映射关系,使其在保持数据流形局部几何信息的同时满足MMC准则的最大化类间平均边缘;2)通过寻找低维嵌入空间的正交基向量来进一步改进算法的判别能力。与原始的LSE算法相比,O-LSDP算法有以下几点不同:

(1)O-LSDP算法寻求一个显式的从输入观测空间到低维特征空间的线性映射关系,而LSE算法不能产生一个明确的映射关系,因此,对于新的测试样本,LSE算法无法获得其在低维嵌入空间的投影。

(2)O-LSDP算法尝试在保持流形局部几何结构的同时,有效利用样本的类别信息来提高原始LSE算法的分类识别能力。

(3)O-LSDP算法通过求解正交化的特征子空间来提取最优的分类特征。

本文第2节介绍MMC准则和LSE算法的基本原理;第3节提出O-LSDP算法;第4节通过在2个标准人脸数据库上的实验结果,验证了O-LSDP算法进行特征提取与分类的可行性和有效性;第5节对O-LSDP算法相关问题进行讨论;最后对本文进行小结。

## 2 相关研究工作

### 2.1 最大边缘准则(MMC)

MMC算法<sup>[18]</sup>基于最大化类间平均边缘来寻找最优的线性子空间。 $S_w$ 和 $S_b$ 分别表示样本的类内散度矩阵和类间散度矩阵:

$$S_w = \sum_{i=1}^c \sum_{j=1}^{n_i} (x_j^i - m_i)(x_j^i - m_i)^T \quad (1)$$

$$S_b = \sum_{i=1}^c n_i (m_i - m)(m_i - m)^T \quad (2)$$

式中, $c$ 是样本类别数, $m$ 是总的样本均值向量, $m_i$ 是第 $i$ 类样本的均值向量, $n_i$ 表示第 $i$ 类样本数, $x_j^i$ 表示第 $i$ 类的第 $j$ 个样本。MMC算法在投影矩阵 $W$ 下的目标函数可以表示为:

$$J(W) = \text{tr}\{W^T(S_b - S_w)W\} \quad (3)$$

限制 $W$ 的列向量为单位向量,那么MMC算法所对应的最优 $W$ 可通过求解下面的特征方程获得:

$$(S_b - S_w)w = \lambda w \quad (4)$$

与LDA算法相比,MMC算法不需要对类内散度矩阵求逆,因此不仅在一定程度上提高了算法的计算效率,而且有效避免了小样本问题。

### 2.2 局部样条嵌入(LSE)

局部样条嵌入算法(LSE)是一种局部特性保持方法,对于嵌入在高维输入空间的低维流形,LSE算法首先通过主分量分析计算每个样本点局部邻域在切空间上的投影以获得该邻域所有样本的局部坐标,从而保持流形的局部结构信息;然后从所有这些有交叠的局部坐标系出发,采用Sobolev空间的一组样条函数把每个样本点的局部坐标映射成全局唯一的低维坐标。LSE算法包含以下3个步骤:

(1)选取邻域。对于每个样本点 $x_i$ ,采用k-NN准则确定其邻域 $X_i = [x_{i_1}, x_{i_2}, \dots, x_{i_k}] \in \mathbb{R}^{D \times k}$ 。

(2)局部切空间投影

LSE算法在每个样本点 $x_i$ 处由其 $k$ 近邻 $X_i$ 来构造局部切空间,即对中心化的局部邻域 $X_i$ 执行奇异值分解:

$$X_i H_k = U_i \begin{bmatrix} \sum_i \\ 0_{(D-k) \times k} \end{bmatrix} V_i^T, i=1, \dots, n \quad (5)$$

式中, $H_k = I - e_k e_k^T / k$ 是中心化算子, $I$ 是 $k \times k$ 的单位矩阵, $e_k$ 是元素全为1的 $k$ 维列向量 $e_k = [1, 1, \dots, 1]^T \in \mathbb{R}^k$ 。 $\sum_i = \text{diag}(\sigma_1, \dots, \sigma_k)$ 是对角化矩阵,包含中心化局部邻域矩阵 $X_i H_k$ 的 $k$ 个按降序排列的奇异值。 $U_i \in \mathbb{R}^{D \times D}$ 和 $V_i \in \mathbb{R}^{k \times k}$ 分别为 $X_i$ 的左右奇异向量矩阵。因此邻域 $X_i$ 的局部切空间投影 $\Theta_i = [\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_k^{(i)}]$ , $i=1, \dots, n$ 可表示为:

$$\Theta_i = (\tilde{U}_i)^T X_i H_k = \tilde{\sum}_i (\tilde{V}_i)^T = [\theta_1^{(i)}, \theta_2^{(i)}, \dots, \theta_k^{(i)}] \quad (6)$$

式中, $\theta_j^{(i)}$ 是样本点 $x_i$ 第 $j$ 个近邻的局部切空间坐标, $\tilde{U}_i \in \mathbb{R}^{D \times d}$ 和 $\tilde{V}_i \in \mathbb{R}^{k \times d}$ 由 $U_i$ 和 $V_i$ 的前 $d$ 个列向量组成, $\tilde{\sum}_i = \text{diag}(\sigma_1, \dots, \sigma_d)$ 包含矩阵 $X_i H_k$ 的 $d$ 个最大奇异值。注意,式(5)和式(6)中隐含着 $D \geq k \geq d$ 。

(3)排列全局坐标

对于第 $i$ 个局部切空间投影 $\Theta_i$ ,假设 $Y_i = [y_{i_1}, y_{i_2}, \dots, y_{i_k}] \in \mathbb{R}^{d \times k}$ 包括 $\Theta_i$ 中 $k$ 个样本点的全局低维坐标,而且 $[y_{i_1}^{(r)}, y_{i_2}^{(r)}, \dots, y_{i_k}^{(r)}]$ 表示 $Y_i$ 的第 $r$ 行坐标分量,我们构建 $d$ 个样条函数 $g_i^{(r)}: \mathbb{R}^d \mapsto \mathbb{R}$ , $r=1, 2, \dots, d$ ,使得:

$$y_{i_j}^{(r)} = g_i^{(r)}(\theta_j^{(i)}), j=1, 2, \dots, k \quad (7)$$

式中,我们知道 $y_{i_j}^{(r)}$ 和 $g_i^{(r)}$ 均是未知的,因此求解 $g_i^{(r)}$ 的问题是一个不适定问题。为使式(7)有解,所求的样条函数不仅要满足式(7)的要求,而且重构误差能够明确地由 $Y_i$ 表示。Xiang等人发现在Sobolev空间的样条函数能够满足我们的需求。现在我们的目标是构造 $d$ 个样条函数 $g_i^{(1)}, \dots, g_i^{(d)}$ ,使下面正则化重构误差达到极小:

$$E(Y_i) = \sum_{j=1}^k \sum_{r=1}^d (y_{i_j}^{(r)} - g_i^{(r)}(\theta_j^{(i)}))^2 + \lambda \sum_{r=1}^d J_r^d(g_i^{(r)}) \quad (8)$$

式中, $\lambda > 0$ 是正则化参数, $J_r^d(g_i^{(r)})$ 控制样条函数 $g_i^{(r)}$ 在 $d$ 维空间的光滑程度,具体定义见参考文献[12, 13, 19, 20, 21]。Duchon等人<sup>[19]</sup>证明在一定的约束条件下,Sobolev空间中具有如下形式的样条函数 $g(t)$ 能够使式(8)的正则化重构误差达到极小:

$$g(t) = \sum_{i=1}^l \beta_i p_i(t) + \sum_{j=1}^k \alpha_j \phi_j(t), r=1, 2, \dots, d \quad (9)$$

式中, $l = (d+s-1)! / (d!(s-1)!)$ , $\{p_i(t)\}_{i=1}^l$ 是 $\mathbb{R}^d$ 空间的多项式集合, $\phi_j(t)$ 是格林函数<sup>[24]</sup>。

$$\phi_j(t) = \begin{cases} (\|t - t_j\|)^{2s-d} \cdot \log(\|t - t_j\|), & \text{如果 } d \text{ 是偶数} \\ (\|t - t_j\|)^{2s-d}, & \text{如果 } d \text{ 是奇数} \end{cases} \quad (10)$$

为了使式(8)的解具有唯一性,给出如下边界约束:

$$\sum_{j=1}^k \alpha_j p_i(\theta_j) = 0, i=1, 2, \dots, l \quad (11)$$

现在,根据式(10)和式(11),Sobolev空间中的 $d$ 个样条函数 $g_i^{(1)}, \dots, g_i^{(d)}$ 的系数向量 $\alpha^r = [\alpha_1^r, \alpha_2^r, \dots, \alpha_k^r]^T \in \mathbb{R}^k$ 和 $\beta^r = [\beta_1^r, \beta_2^r, \dots, \beta_l^r]^T \in \mathbb{R}^l$ , $r=1, 2, \dots, d$ 能够通过下面的线性方程组来求解:

$$A_i \cdot \begin{pmatrix} \alpha^1, \dots, \alpha^d \\ \beta^1, \dots, \beta^d \end{pmatrix} = \begin{pmatrix} Y_i^T \\ 0 \end{pmatrix} \quad (12)$$

其中

$$A_i = \begin{pmatrix} K & P \\ P^T & 0 \end{pmatrix} \in \mathbb{R}^{(k+d) \times (k+d)} \quad (13)$$

$K$ 是元素为 $K_{\alpha} = \phi(\|\theta_j^{(i)} - \theta_l^{(i)}\|)$ 的 $k \times k$ 对称阵, $P$ 是元素为 $P_{\alpha} = p_i(\theta_j^{(i)})$ 的 $k \times l$ 矩阵。

为了使重构误差  $E(Y_i)$  能够明确地由  $Y_i$  表示,对式(8)作如下近似:

$$E(Y_i) \approx \sum_{r=1}^d \sum_{j=1}^k (y_i^{(r)} - g_i^{(r)}(\alpha_j^{(r)}))^2 + \lambda \sum_{r=1}^d (\alpha^r)^T K \alpha^r \quad (14)$$

如果  $\lambda$  足够小,可以忽略式(14)右边的第一项,因此有:

$$E(Y_i) \propto \sum_{r=1}^d (\alpha^r)^T K \alpha^r = \text{tr}(Y_i B_i Y_i^T) \quad (15)$$

式中,  $B_i$  是由  $A_i^{-1}$  的左上角  $k \times k$  元素组成的子矩阵。计算所有样本的重构误差,即:

$$E(Y) = \sum_{i=1}^n \text{tr}(Y_i B_i Y_i^T) \quad (16)$$

设  $S_i \in \mathbb{R}^{n \times k}$  是满足  $Y S_i = Y_i$  的 0-1 选择矩阵,则式(16)可以转化为:

$$E(Y) = \text{tr}(Y S B S^T Y^T) = \text{tr}(Y M Y^T) \quad (17)$$

式中,  $S = [S_1, \dots, S_n]$ ,  $B = \text{diag}(B_1, \dots, B_n)$ ,  $M = S B S^T$ 。

为了防止出现退化解,对  $Y$  加上标准化约束  $Y Y^T = I$ , 那么全局最优的低维嵌入  $Y$  由  $M$  的第 2 至第  $d+1$  个最小特征值所对应的特征向量组成。

### 3 正交局部样条判别投影(O-LSDP)

针对原始的 LSE 算法存在样本外点学习问题和分类性能差的缺点,本节提出了一种有监督的 LSE 算法——正交局部样条判别投影算法(O-LSDP)。该方法在 LSE 算法的基础上,通过引入类别信息和正交化投影子空间来提高 LSE 算法的分类性能。

#### 3.1 O-LSDP 算法动机

对于数据的可视化而言,维数约简的目的是把原始数据集投影到一个 2 维或 3 维空间,并且尽可能忠实地保持数据的本征结构。而对于数据的分类来说,维数约简的主要目的是把数据映射到一个特征空间,使得同类样本的差异尽可能小,而不同类样本的差异尽可能大。LSE 算法作为一种基于局部特性保持的流形学习方法,能够有效挖掘嵌入在高维空间的低维流形结构,因此它是一种有效的数据可视化方法,但它不是一种很好的面向分类的特征提取方法。我们通过一个例子来说明这一点。图 1 是应用 LSE 算法对一个植物叶片图像数据集进行 2 维嵌入的结果。该数据集由瑞典植物叶片数据库(Swedish Leaf Database)<sup>[22]</sup> 的第 1 类和第 3 类共 150 幅叶片图像(每类 75 幅)构成,对每幅图像进行预处理并归一化为  $32 \times 32$  像素大小的灰度图。在图 1 中,左右两边分别列举了每类 9 张代表性的叶片图像。从可视化的角度来看,这两类数据的 2 维嵌入结果具有良好的流形分布;从分类的角度来看,我们不难发现 2 维嵌入结果被粗略地划分为两大部分,其中星形表示一类,圆形表示另一类,但是这两类当中有部分数据相互混叠在一起。这表明 LSE 算法在保持流形局部几何结构时强调数据自然的聚类,但所提取的特征并不是叶片分类最优的投影方向。

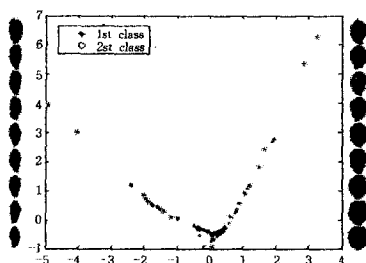


图 1 基于 LSE 的植物叶片数据集的 2 维嵌入结果

为了改进原始 LSE 算法的分类识别能力,我们提出了 O-LSDP 监督流形学习算法。O-LSDP 算法的目标是利用样本的类别信息和正交化特征子空间来改进 LSE 算法的判别能力。一方面在保持流形局部几何结构的同时,为不同的类构建不同的平移和缩放因子,从而使同类样本拉得更近而不同类样本分得更开。我们继续使用图 1 的例子来说明这一点。图 2 显示了经过平移变换后 LSE 的 2 维嵌入结果。从图中可以看出,不同类的类心之间由于采用了平移变换明显被拉开,两类之间不再有数据重叠,显然其判别能力得到提高。就分类而言,光拉开不同类的类心是不够的,同时还必须使同类的数据分布尽可能集中,因此还需要对同类数据进行适当的缩放变换。图 3 是经过平移和缩放变换后 LSE 的 2 维嵌入结果。从图 3 中发现经过平移和缩放变换后,同类数据之间拉得更近而不同类数据之间分得更远,从而明显提高了数据的可分性。

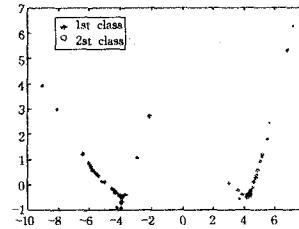


图 2 经过平移变换后 LSE 的 2 维嵌入结果

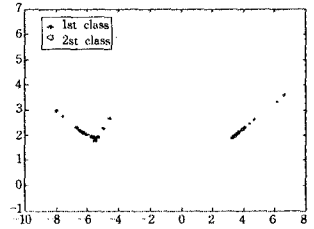


图 3 经过平移和缩放变换后 LSE 的 2 维嵌入结果

通过上面的分析,我们发现构建合适的平移和缩放模型能够显著改进算法的分类识别能力。然而在实际应用中如何对每类样本选择不同的平移和缩放因子是一个公开的问题。为了解决这个问题,采用最大边缘准则(MMC)来为每类样本寻找最优的平移和缩放模型,提高数据的可分性能。图 4 为基于 MMC 准则寻找最优的平移和缩放模型后 LSE 的 2 维嵌入结果。从图中可以看出,基于 MMC 准则的 LSE 算法能够完全正确地分离两类植物叶片图像。

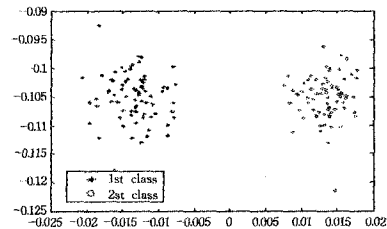


图 4 基于 MMC 准则寻找平移和缩放模型后 LSE 的 2 维嵌入结果

另一方面,我们发现基于 MMC 准则的 LSE 算法求解的低维特征子空间并不是正交的, Ye 等人证明了正交化投影子空间可以改善算法的分类能力<sup>[23]</sup>。因此我们拟通过正交化低维特征子空间来消除数据的噪声影响以进一步提高算法的分类性能(见图 5)。

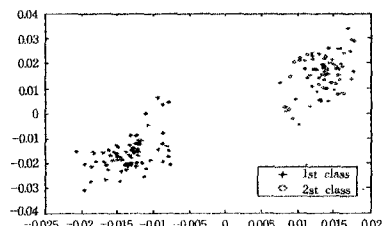


图 5 正交化特征子空间后的 2 维嵌入结果

### 3.2 O-LSDP 算法推导

#### 3.2.1 线性化 LSE 算法

为了解决 LSE 算法的样本外点学习问题,构造显式的从输入高维观测样本  $X$  到输出低维嵌入  $Y$  的线性映射关系去近似 Sobolev 空间上的样条函数,即  $Y=V^T X$ 。因此,原始的 LSE 算法目标函数式(17)转变为下面的形式:

$$J_1(Y) = \min \operatorname{tr}(YMY^T) = \min \operatorname{tr}(V^T XMX^T V) \quad (18)$$

一旦线性变换矩阵  $V$  确定,新的样本点在低维特征空间的投影问题就迎刃而解。我们称这个算法为线性化 LSE 算法(Linearized LSE, LLSE)。考虑一个新的测试样本  $x_i$ ,其在低维特征空间的投影可表示为:

$$y_i = V^T x_i$$

#### 3.2.2 获得最优的线性判别投影

基于上面的分析,我们知道线性化 LSE 算法(LLSE)的目标是尽可能地保持近邻所确定的局部几何结构。为了使 LSE 算法获得更好的分类性能,我们向 LLSE 算法引入 MMC 准则来构建最优的平移和缩放模型,使得它在保持 LSE 局部几何结构的同时,能够拥有 MMC 准则强大的判别能力,也就是说,通过式(18)所确定的线性变换矩阵  $V$  同时满足式(3)。那么现在的问题可以描述成下面的多目标优化问题:

$$\begin{cases} \min \operatorname{tr}\{V^T XMX^T V\} \\ \max \operatorname{tr}\{V^T (S_b - S_w)V\} \end{cases} \quad (19)$$

而且,在 LSE 算法中有两个约束,即单位化约束和中心化约束:

$$V^T XX^T V = I \quad (20)$$

$$Y_e = 0 \quad (21)$$

中心化约束式(21)实际上是消除低维输出结果  $\{y_i\}_{i=1}^n$  的平移自由度,而在我们提出的算法中,需要利用由 MMC 准则隐含所确定的平移模型去提高算法的分类识别能力,因此应当忽略中心化约束,算法的优化问题可以转化为下面带约束的目标函数:

$$\begin{cases} \min \operatorname{tr}\{V^T XMX^T V\} \\ \max \operatorname{tr}\{V^T (S_b - S_w)V\} \end{cases} \quad (22)$$

$$\text{s. t. } V^T XX^T V = I$$

这个约束的多目标优化问题旨在寻找一个既能使 LSE 的重构误差最小又能同时最大化类间平均边缘的线性变换。对式(22)进行线性化操作:

$$\min \operatorname{tr}\{V^T (XMX^T - \beta(S_b - S_w))V\} \quad (23)$$

$$\text{s. t. } V^T XX^T V = I$$

式中,  $\beta \in [0, 1]$  是尺度因子。使用 Lagrangian 乘子法来求解上面的优化问题,即:

$$\frac{\partial}{\partial V} \operatorname{tr}\{V^T (XMX^T - \beta(S_b - S_w))V - \lambda(V^T XX^T V - I)\} = 0 \quad (24)$$

式(24)可转化为求解下面的广义特征值问题:

$$(XMX^T - \beta(S_b - S_w))v = \lambda XX^T v \quad (25)$$

式中,  $\lambda$  是广义特征方程式(25)的特征值,  $v$  是其特征向量。

假设  $v_1, v_2, \dots, v_d$  是广义特征对  $(XMX^T - \beta(S_b - S_w), XX^T)$  的前  $d$  个最小特征值所对应的特征向量,那么使目标

函数式(23)达到极小化的线性变换矩阵  $V$  可表示为:

$$V = [v_1, v_2, \dots, v_d] \quad (26)$$

我们称这种新的线性子空间方法为局部样条判别投影算法(Local Spline Discriminant Projection, LSDP)。注意在实际应用中,经常会碰到小样本问题(SSS),因此导致  $XX^T$  是奇异的,为了避免出现该问题,我们需要对数据进行 PCA 预处理,将数据投影到 PCA 子空间并选取合适的投影维度,使得  $XX^T$  非奇异。

#### 3.2.3 求解正交化特征向量

由式(25)所求解的广义特征向量是非正交的,我们采用 Gram-Schmidt 正交化来产生一组正交的低维特征子空间以消除噪声影响,从而进一步提高算法的分类识别能力。我们称该算法为正交局部样条判别投影(Orthogonal Local Spline Discriminant Projection, O-LSDP)。

令  $g_1 = v_1$ , 假定前  $k-1$  个正交基向量为  $g_1, g_2, \dots, g_{k-1}$ , 则第  $k$  个正交基向量计算方法如下:

$$g_k = v_k - \sum_{i=1}^{k-1} \frac{g_i^T v_k}{g_i^T g_i} g_i \quad (27)$$

因此  $G = [g_1, g_2, \dots, g_d]$  即为所求的  $d$  维特征子空间的正交基向量。

### 3.3 O-LSDP 算法步骤

根据上面的分析,我们提出的 O-LSDP 算法主要步骤如下:

第 1 步 对每个样本点  $x_i$ , 通过 KNN 标准确定其  $k$  个近邻点  $X_i = [x_{i_1}, x_{i_2}, \dots, x_{i_k}]$ ;

第 2 步 计算中心化邻域  $X_i H_k$  的  $d$  个左奇异向量矩阵  $U_i$ , 并按式(6)计算邻域  $X_i$  的局部切空间坐标  $\Theta_i$ ;

第 3 步 根据式(13)计算矩阵  $A_i$ ;

第 4 步 按照如下局部求和的方式计算样条排列矩阵  $M$ :

$$M(I_i, I_i) \leftarrow M(I_i, I_i) + B_i, i = 1, 2, \dots, n$$

其中,  $M$  初始化为零矩阵,  $I_i = \{i_1, \dots, i_k\}$  表示样本  $x_i$  的  $k$  个近邻对应的下标索引集合,  $B_i$  是  $A_i^{-1}$  左上角  $k \times k$  的子矩阵;

第 5 步 将数据集  $X$  投影到 PCA 子空间  $X_{PCA} = V_{PCA}^T X$ ;

第 6 步 计算矩阵  $X_{PCA} M X_{PCA}$ ;

第 7 步 计算样本集  $X_{PCA}$  的类间散度矩阵  $S_b$ 、类内散度矩阵  $S_w$  以及它们的差值  $S_b - S_w$ ;

第 8 步 根据式(27)计算特征子空间的  $d$  个正交基向量  $G = [g_1, g_2, \dots, g_d]$ , 并获得样本集  $X$  在  $d$  维正交特征子空间的投影  $Y = G^T X_{PCA}$ 。

## 4 实验结果

为了评估正交局部样条判别投影算法(O-LSDP)的有效性和分类能力,本文在 2 个不同的人脸数据库: Yale database<sup>1)</sup> 和 Olivetti Research Laboratory (ORL) database<sup>2)</sup> 上进行了大量实验。为了更好地分析实验结果,在人脸数据库上将提出的 O-LSDP 算法与传统子空间降维和特征提取方法:最大边缘准则(MMC)<sup>[18]</sup>、线性判别分析(LDA)<sup>[24]</sup> 以及基于流形学习的算法:监督局部保持投影(Supervised Locality Preserving Projection, SLPP)<sup>[25]</sup>、监督线性局部切空间排列

<sup>1)</sup> <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

<sup>2)</sup> <http://www.cam-orl.co.uk/facedatabase.html>

算法 (Supervised LLTSA, SLLTSA)<sup>[26]</sup>、边缘 Fisher 分析 (MFA)<sup>[14]</sup>、线性化局部样条嵌入算法 (LLSE)、局部样条判别投影算法 (LSDP) 进行了仿真实验和比较。除了将所有人脸图像裁剪成  $32 \times 32$  的灰度图以外, 没有进行其它的预处理。

5 种基于流形学习的算法即 SLPP, SLLTSA, LLSE, LSDP 和 O-LSDP 在构建近邻图时, 考虑到同类样本应当位于同一个局部几何结构中, 因此最近邻参数  $k$  选择为  $l-1$  ( $l$  表示每类的训练样本数)。在 LSDP 和 O-LSDP 算法中, 尺度因子  $\beta=1$ ; 在 MFA 算法中, 类内最近邻数  $k_1$  和类间最近邻样本对数  $k_2$  分别取  $[1, l-1]$  和  $[l, 8c]$  之间的最佳值。为了更好地比较这些特征提取方法的性能, 所有实验的距离度量均采用欧氏距离, 分类器均采用最简单的最近邻分类器。

#### 4.1 Yale 人脸库上的实验结果

Yale 人脸库包含有 15 个人的 165 幅灰度图像 (每个人有 11 幅图像)。对于每一个人, 其图像都是在不同的光照和表情条件下采集的, 其中光照变化的图像有 3 幅, 即按照人脸从左-中-右变化; 人脸的面部表情变化, 包含有正常、高兴、悲哀、困倦、惊讶和眨眼的表情; 戴眼镜和没有戴眼镜的图像。图 6 是 Yale 人脸数据库中 1 个人的样本图像。



图 6 Yale 人脸库的样本图像

首先测试特征空间维数的变化对识别率的影响。对于 LDA 算法, 由于其至多有  $c-1$  个非零的广义特征值, 因此特征空间维数上界为  $c-1$ 。对于每个人, 随机选择 3、5 和 7 幅图像作为训练样本集, 其余的作为测试集, 不同特征提取方法的识别率随特征子空间维数的变化曲线如图 7 所示。总体来说, 所有这些方法的性能都是随着特征子空间维数的变化而变化。从图中可以看出, 开始阶段识别率是随着特征维数的

增加而增加。但这一趋势并不是一直保持, 当特征维数达到某一阈值后, 识别率将会呈下降趋势或出现波动现象。

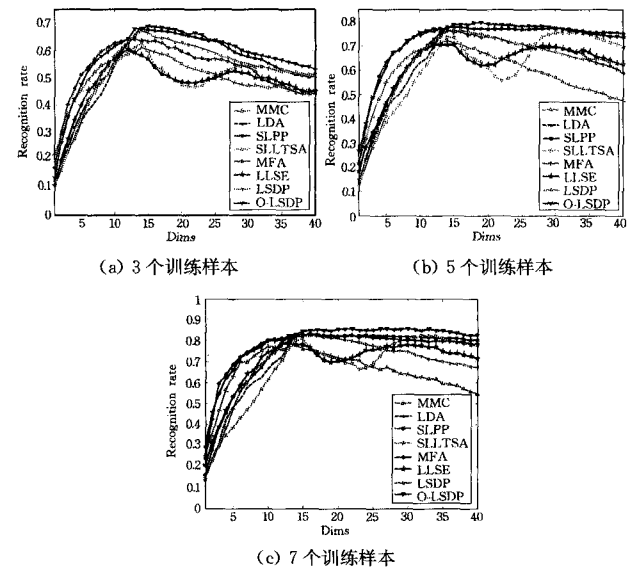


图 7 MMC, LDA, SLPP, SLLTSA, MFA, LLSE, LSDP 和 O-LSDP 在 Yale 人脸库上的识别率随特征空间维数变化曲线

其次测试 8 种特征提取方法的识别效果。对于每个人, 随机选择 7 幅图像作为训练样本, 剩余的 4 幅图像作为测试样本。图 8 显示了 20 次重复实验的最大平均识别率, 可以看出, MMC, LDA, SLPP, LLTSA, MFA, LLSE, LSDP 和 O-LSDP 算法的最大平均识别率分别为  $77.83(\pm 5.02)\%$ 、 $82.25(\pm 4.72)\%$ 、 $82.92(\pm 4.71)\%$ 、 $82.25(\pm 5.52)\%$ 、 $83.42(\pm 4.03)\%$ 、 $78.42(\pm 4.24)\%$ 、 $82.33(\pm 4.73)\%$  和  $85.50(\pm 3.29)\%$ 。可以看出在这 8 种方法中, O-LSDP 获得了最高的平均识别率。

表 1 BASELINE, MMC, LDA, SLPP, SLLTSA, MFA, LLSE, LSDP 和 O-LSDP 在 Yale 人脸库上取不同训练样本数时的最大平均识别率、标准差及相应的特征空间维数

Method	3 Train	4 Train	5 Train	6 Train	7 Train
Baseline	50.79±4.53	54.19±4.94	56.00±5.44	59.20±4.61	60.58±4.87
MMC	61.08±4.84(14)	67.43±5.20(13)	72.11±3.42(14)	75.40±4.29(14)	77.83±5.02(14)
LDA	68.25±4.21(14)	74.86±5.51(14)	77.22±3.50(14)	81.73±5.05(14)	82.25±4.72(14)
SLPP	67.92±4.25(14)	75.14±5.46(16)	77.22±3.50(14)	81.60±4.94(14)	82.92±4.71(14)
SLLTSA	62.75±4.30(14)	69.71±5.24(15)	75.72±4.62(33)	80.27±4.69(32)	82.25±5.52(32)
MFA	64.04±5.63 (12, 2, 117)	71.48±5.82 (12, 3, 115)	77.06±4.19 (13, 4, 77)	80.80±4.82 (17, 5, 99)	83.42±4.03 (16, 4, 109)
LLSE	60.25±4.54(12)	65.14±5.15(13)	70.89±5.77(12)	76.00±3.82(13)	78.42±4.24(12)
LSDP	67.17±5.05(14)	73.67±5.82(14)	76.50±3.52(15)	80.93±4.70(15)	82.33±4.73(16)
O-LSDP	68.75±4.76(15)	76.48±4.43(16)	79.72±3.69(19)	84.27±3.95(22)	85.50±3.29(26)

注: 对于 MFA 算法, 括号中第 1 个数表示特征空间维数, 第 2 和第 3 个数分别表示参数类内最近邻数  $k_1$  和类间最近邻样本对数  $k_2$ 。

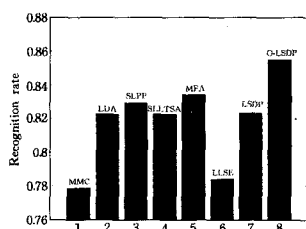


图 8 MMC, LDA, SLPP, SLLTSA, MFA, LLSE, LSDP 和 O-LSDP 在 Yale 人脸库上进行 20 次实验的最大平均识别率

最后检验训练样本数目对识别性能的影响。对于每种方法, 分别随机选取 3、4、5、6 和 7 幅图像作为训练样本集, 其余

的作为测试样本集。每个实验独立重复 20 次, 计算平均识别率。注意其中 Baseline 方法表示不做任何维数约简, 直接在原始的 1024 维图像空间中测试识别率。表 1 为在不同训练样本条件下最大的平均识别率和标准差, 以及获得最大识别率时的特征空间维数。从表 1 可以看出, O-LSDP 算法在各种情况下都获得了最佳的分类识别性能。

#### 4.2 ORL 人脸库上的实验结果

英国剑桥大学的 ORL 人脸库包括 400 幅人脸图像, 来自 40 个不同的人, 每人 10 幅图像, 分别在不同时间、不同光照、不同姿态、不同人脸表情 (睁眼和闭眼、微笑和不微笑) 和不同的人脸细节变化 (有无眼镜) 条件下采集得到。ORL 人脸库

中原始图像为  $92 \times 112 = 10384$  维灰度图像, 灰度值标准化在 0 至 1 之间。图 9 是一个人的 10 幅图像。



图 9 ORL 人脸库的样本图像

实验设计与上一节相同。每个人的训练样本分别选择 3、5 和 7 幅图像, 其余的图像作为测试样本。记录的识别率为 20 次实验的平均识别率。图 10 显示了在 ORL 数据集上 MMC、LDA、SLPP、SLLTSA、MFA、LLSE、LSDP 和 O-LSDP 在不同训练样本数目条件下识别率随特征子空间维数变化的曲线。从图中可以看出, 随着特征子空间维数的增加, 这些方法的识别率也增加。当特征子空间的维数高于某个阈值时, 识别率基本保持平稳。

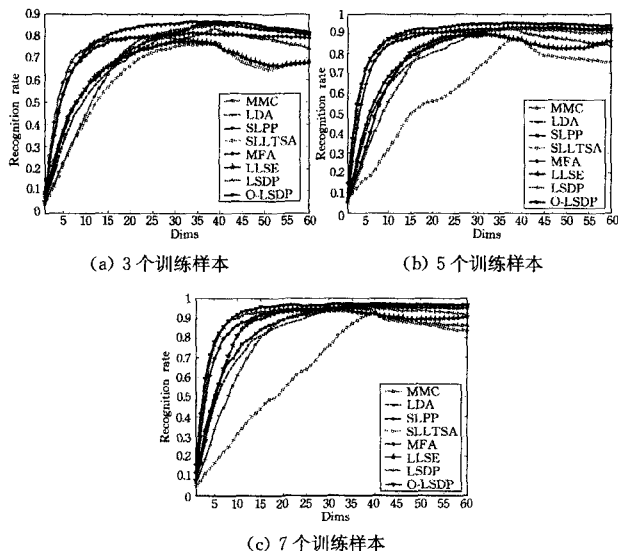


图 10 MMC、LDA、SLPP、SLLTSA、MFA、LLSE、LSDP 和 O-LSDP 在 ORL 人脸库上的识别率随特征空间维数变化曲线

表 2 BASELINE、MMC、LDA、SLPP、SLLTSA、MFA、LLSE、LSDP 和 O-LSDP 在 ORL 人脸库上取不同训练样本数时的最大平均识别率、标准差及相应的特征空间维数

Method	3 Train	4 Train	5 Train	6 Train	7 Train
Baseline	77.14±2.23	82.48±2.22	86.90±2.12	89.34±2.55	90.92±2.26
MMC	83.45±1.90(39)	88.83±2.15(38)	92.47±2.13(37)	93.88±2.34(34)	94.92±1.77(36)
LDA	85.86±1.68(38)	90.33±1.66(39)	93.23±1.91(39)	94.62±1.96(39)	95.71±1.86(39)
SLPP	85.84±1.62(39)	90.25±1.63(39)	93.28±1.92(39)	94.66±1.91(39)	95.79±1.90(39)
SLLTSA	81.00±2.13(81)	86.02±2.06(83)	90.30±1.64(95)	92.34±2.6(99)	94.21±1.92(111)
MFA	81.38±2.79 (34, 2, 305)	89.85±2.26 (42, 3, 319)	93.47±2.35 (43, 4, 215)	95.44±1.69 (44, 5, 246)	96.46±1.97 (51, 6, 242)
LLSE	77.54±3.25(32)	84.92±2.93(32)	89.43±2.22(33)	92.44±2.23(27)	94.04±2.50(30)
LSDP	85.29±1.55(39)	89.88±1.70(39)	92.80±1.79(38)	94.50±1.94(39)	95.42±2.00(39)
O-LSDP	86.73±1.55(40)	93.15±1.72(43)	95.40±1.37(37)	96.22±1.51(35)	97.08±2.12(34)

注: 对于 MFA 算法, 括号中第 1 个数表示特征空间维数, 第 2 和第 3 个数分别表示参数类内最近邻数  $k_1$  和类间最近邻样本对数  $k_2$ 。

## 5 讨论

在 Yale 和 ORL 人脸数据库上的实验结果表明:

(1) O-LSDP 算法在所有的实验条件下实现了最好的分类性能。这主要源于以下几个方面的原因: 1) 基于流形学习的线性特征提取方法能够有效探测非线性流形结构。传统的线性特征提取方法如 PCA、MMC 和 LDA 仅能发现平坦的欧氏结构, 并不能捕捉到非线性的流形结构。而对于基于流形学习的方法如 SLPP、SLLTSA 和 O-LSDP 算法, 尽管它们也

图 11 显示了每个人随机选择 7 幅图像作为训练样本时重复实验 20 次的最大平均识别率。从图 11 可以看出, MMC、LDA、SLPP、SLLTSA、MFA、LLSE、LSDP 和 O-LSDP 在特征子空间维数分别为 36、39、39、111、51、30、39 和 34 时达到最大平均识别率 94.92%、95.71%、95.79%、94.21%、96.46%、94.04%、95.42% 和 97.08%。标准差分别为 1.77%、1.86%、1.90%、1.92%、1.97%、2.50%、2.00% 和 2.12%。我们不难发现, O-LSDP 算法的识别率明显超过其它 7 种方法, 其中 SLLTSA 性能最差, 而 LDA 和 SLPP 算法几乎获得同样的识别率。

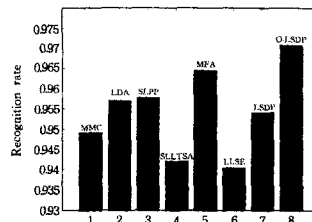


图 11 MMC、LDA、SLPP、SLLTSA、MFA、LLSE、LSDP 和 O-LSDP 在 ORL 人脸库上进行 20 次实验的最大平均识别率

最后, 检验了训练样本数目对识别率的影响。从每个人的 10 幅图像中分别随机选择 3、4、5、6 和 7 幅图像作为训练样本集, 其余的图像作为测试样本集。我们独立重复 20 次实验并计算平均识别率。表 2 为选择不同训练样本数目条件下的最大平均识别率和标准差, 以及获得最大识别率时的特征子空间维数。正如表 2 所列, 在 8 种方法中, 我们提出的 O-LSDP 算法在选择 5 种不同的训练样本数目时都实现了最好的分类性能, 而 LLSE 算法产生了最低的识别率, MMC 的识别率超过了 SLLTSA, 并且随着训练样本数目的增加, MMC 实现了与 LDA 相当的识别率; 除此, 随着训练样本数目的增加, MFA 与 O-LSDP 的识别率接近。

是线性特征提取方法, 但是我们认为从非线性流形学习的观点看待它们更合适。因为原始的流形学习方法仅仅将训练样本集从高维输入空间嵌入到低维空间, 并不能产生一个明确的映射函数关系, 流形学习线性化的主要目的就是通过线性近似去获得一个明确的映射关系。如 SLPP 算法通过对数据流形上的 Laplacian-Beltrami 算子的特征函数进行线性近似, 从而获得高维输入空间到低维流形的显式映射关系<sup>[4,5]</sup>。也就是说, 本质上这些基于流形学习的线性方法并不等价于传统的线性特征提取方法。众所周知, PCA 的主要目标是保持

数据的全局几何结构, LDA 旨在保持数据的判别信息, 而这些基于流形学习的线性方法通过构建近邻图明确地考虑了数据的非线性流形结构, 因此, 它们能够发现服从流形分布的内在非线性结构。实验中所采用的 Yale 和 ORL 人脸库的图像随姿态、光照和表情等因素变化而变化, 而这些因素已经被研究者们证明是流形的本征特征<sup>[1-3,27]</sup>, 所以应用基于流形学习的特征提取方法能够有效地探测到这些本征特征。从这个意义来说, 基于流形学习的特征提取方法明显优于其它传统的线性特征提取方法。2) 有效利用样本类别信息指导构建平移和缩放模型。O-LSDP 算法在保持数据流形局部几何结构的同时, 有效利用样本的类别信息为不同类数据构建最优的平移和缩放因子, 从而提高算法的分类识别能力。当然 O-LS-DP 算法并不像 LDA 和 MMC 算法那样一味地去追求“最小化类内离散度和最大化类间离散度”, 而是在保持流形局部几何结构与判别信息之间寻找最佳的结合点。在许多实际的分类问题中, 保持流形的局部几何结构比保持全局欧氏结构更为重要<sup>[15]</sup>。因此我们认为在面向服从流形分布的高维数据分类任务时, 最优的特征提取方法应当既要充分利用样本的先验信息指导分类又要同时保持数据的局部几何结构, 只有这样才能提取最优的分类特征。3) 正交化特征子空间有助于消除冗余噪声, 进一步提高算法的分类能力。

(2) O-LDSP 算法在 Yale 和 ORL 人脸数据库上的分类识别性能明显优于 SLLTSA 算法。一方面尽管 O-LDSP 和 SLLTSA 算法都通过逼近每个样本点邻域的局部切空间来保持低维流形的局部几何特性, 但是它们对局部坐标的全局排列方法有着本质的不同。SLLTSA 算法通过使用仿射变换来排列局部坐标而 O-LDSP 算法通过构建 Sobolev 空间的光滑样条函数来排列局部坐标。与仿射变换相比, 光滑样条函数能够实现更小的重构误差, 从而更好地保持流形的局部几何。另一方面, 正交化特征子空间有助于改善 O-LDSP 算法的分类能力。

(3) 在 Yale 和 ORL 人脸库上, 随着训练样本数目的增加, MFA 算法的识别率接近 O-LSDP 算法的识别率。这主要是因为 MFA 算法可以通过设置合适的  $k_1$  和  $k_2$  参数有效地捕获数据的局部几何信息和判别信息, 然而前提是 MFA 算法必须要遍历  $k_1$  和  $k_2$  可能的取值来进行模型选择。因此, 随着训练样本数目的增加, MFA 算法的计算复杂度会显著提高。

**结束语** 原始的流形学习算法在面临模式分类任务时存在泛化能力差和无监督学习的缺点, 导致它们无法直接应用于模式识别问题。本文首先详细介绍了最大边缘准则 (MMC) 和局部样条嵌入算法 (LSE), 然后针对 LSE 算法在数据分类中的不足, 提出了一种新的监督流形学习算法——正交局部样条判别投影 (O-LSDP)。O-LSDP 算法通过线性化手段解决样本外点学习问题, 并通过构建平移缩放模型以及正交化特征子空间来提高算法的分类识别能力。最后, 在 Yale 和 ORL 人脸数据库上进行实验比较分析, 验证了该算法的可行性与有效性。

## 参 考 文 献

[1] Tenenbaum J B, de Silva V, Langford J C. A global geometric

framework for nonlinear dimensionality reduction [J]. *Science*, 2000, 290(5500): 2319-2323

[2] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. *Science*, 2000, 290(5500): 2323-2326

[3] Saul L K, Roweis S T. Think globally, fit locally: unsupervised learning of low dimensional manifolds [J]. *The Journal of Machine Learning Research*, 2003, 4: 119-155

[4] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering [C]// *Proceedings of Advances in Neural Information Processing Systems*. 2002, 14: 585-591

[5] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. *Neural Computation*, 2003, 15(6): 1373-1396

[6] Donoho D L, Grimes C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data [C]// *Proceedings of the National Academy of Sciences of the United States of America*. 2003: 5591-5596

[7] Weinberger K Q, Saul L K. Unsupervised learning of image manifolds by semidefinite programming [C]// *Proceedings of CVPR-04*. 2004: 988-995

[8] Weinberger K Q, Saul L K. An introduction to nonlinear dimensionality reduction by maximum variance unfolding [C]// *Proceedings of AAAI-06*. 2006: 1683-1686

[9] Zhang Z, Zha H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment [J]. *SIAM J. Scientific Computing*, 2005, 26(1): 313-338

[10] Lin T, Zha H. Riemannian manifold learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 30(5): 796-809

[11] Lin T, Zha H, Lee S. Riemannian manifold learning for nonlinear dimensionality reduction [C]// *Proceedings of ECCV2006*. 2006: 44-55

[12] Xiang S, Nie F, Zhang C. Spline embedding for nonlinear dimensionality reduction. *Machine Learning [C]// Proceedings of ECML2006*. 2006: 825-832

[13] Xiang S, Nie F, Zhang C. Nonlinear dimensionality reduction with local spline embedding [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 21(9): 1285-1298

[14] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: A general framework for dimensionality reduction [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 5(1): 40-51

[15] He X, Yan S, Hu Y, et al. Face recognition using laplacianfaces [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(3): 328-340

[16] De Ridder D, Kouropteva O, Okun O, et al. Supervised locally linear embedding [C]// *Proceedings of the 2003 Joint International Conference on Artificial Neural Networks and Neural Information Processing*. 2003: 333-341

[17] Pan Y, Ge S S, Mamun A. Weighted locally linear embedding for dimension reduction [J]. *Pattern Recognition*, 2009, 42(5): 798-811

保证这一前提,因此算法存在一些问题有待进一步研究,即一些类别的多样本性,一些类别的无样本性。这些问题导致空间结构信息的不平衡,影响特征提取效果。

### 参考文献

- [1] Peng H, Li B, Luo X, et al. A learning-based audio watermarking scheme using kernel Fisher discriminant analysis [J]. *Digital Signal Processing*, 2013, 23(1): 382-389
- [2] Meng Y, Zou J, Gan X, et al. Adaptive WNN aerodynamic modeling based on subset KPCA feature extraction [J]. *Journal of Central South University*, 2013, 20: 931-941
- [3] 胡益, 王丽, 马贺贺, 等. 基于核 PLS 方法的非线性过程在线监控[J]. *化工学报*, 2011, 62(9): 2555-2561
- [4] Gan J Y, Liu J F. Fusion and recognition of face and iris feature based on wavelet feature and KFPCA [C]//International Conference on Wavelet Analysis and Pattern Recognition, 2009. IC-WAPR 2009. IEEE, 2009: 47-50
- [5] Zhang X, Yan W, Zhao X, et al. Nonlinear biological batch process monitoring and fault identification based on kernel fisher discriminant analysis [J]. *Process biochemistry*, 2007, 42(8): 1200-1210
- [6] 郑建伟, 王万良, 姚晓敏, 等. 张量局部 Fisher 判别分析的人脸识别[J]. *自动化学报*, 2012, 38(9): 1485-1495
- [7] Jin L W, Ding K, Huang Z B. Incremental learning of FDA model for Chinese writer adaptation [J]. *Neurocomputing*, 2010, 73(10): 1614-1623
- [8] Xiong T, Ye J, Li Q, et al. Efficient kernel discriminant analysis via QR decomposition [J]. *Advances in neural information processing systems*, 2005, 17: 1529-1536
- [9] Golub G H, Van Loan C F. *Matrix computations* [M]. Johns Hopkins University Press, 1996
- [10] Zhang X, Yan W W, Zhao X, et al. Performance monitoring and fault identification based on DKPCA and PSVM [J]. *Journal of Shanghai Jiaotong University*, 2008, 42(2): 181-185
- [11] Chiang L H, Russell E, Braatz R D. *Fault detection and diagnosis in industrial systems* [M]. Springer Verlag, 2001
- [12] Bo C M, Zhang S, Zhang G M, et al. Fault identification of Tennessee Eastman process based on FS-KPCA [J]. *Journal of Chemical Industry and Engineering*, 2008, 59(7): 1783-1789
- [13] Ricker N L, Lee J H. Nonlinear model predictive control of the Tennessee Eastman challenge process [J]. *Computers & chemical engineering*, 1995, 19(9): 961-981
- [14] 张熠卓, 徐光华, 梁霖, 等. 利用增量式非线性流形学习的状态监测方法[J]. *西安交通大学学报*, 2011, 45(001): 64-68
- [15] 王迎, 王新明, 赵小强. 基于小波去噪与 KPCA 的 TE 过程故障检测研究[J]. *化工机械*, 2011, 38(1): 49-53
- [16] [http://blog.sina.com.cn/s/blog\\_02b6f23d0100eq88.html](http://blog.sina.com.cn/s/blog_02b6f23d0100eq88.html)

(上接第 279 页)

- [18] Li H, Jiang T, Zhang K. Efficient and robust feature extraction by maximum margin criterion [J]. *IEEE Transactions on Neural Networks*, 2006, 17(1-3): 157-165
- [19] Duchon J. Splines minimizing rotation-invariant semi-norms in Sobolev spaces [C]//Constructive theory of functions of several variables. 1977: 85-100
- [20] Meinguet J. Multivariate interpolation at arbitrary points made simple [J]. *Journal of Applied Mathematics and Physics*, 1979, 30: 292-304
- [21] Wahba G. *Spline models for observational data* [M]. SIAM Press, 1990
- [22] Söderkvist O. Computer vision classification of leaves from Swedish trees [D]. Master's Thesis, Linköping University, 2001
- [23] Ye J. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems [J]. *Journal of Machine Learning Research*, 2006, 6: 483-502
- [24] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces; Recognition using class specific linear projection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711-720
- [25] Cai D, He X, Han J. Using graph model for face analysis [R]. Univ. Illinois Urbana-Champaign, Urbana, IL, Department of Computer Science, Technical Report, 2005
- [26] Zhang T, Yang J, Zhao D, et al. Linear local tangent space alignment and application to face recognition [J]. *Neurocomputing*, 2007, 70(7-9): 1547-1553
- [27] De Silva V, Tenenbaum J B. Global versus local methods in nonlinear dimensionality reduction [C]//Proceedings of Advances in Neural Information Processing Systems. 2003, 15: 721-728

(上接第 286 页)

- [13] Cai Guo-liang, Hu Pei, Li Yu-xiu. Modified function lag projective synchronization of a financial hyperchaotic system [J]. *Nonlinear Dynamics*, 2012, 69(3): 1457-1464
- [14] 刘恒, 余海军, 向伟. 带未知扰动的多涡卷混沌系统修正函数时滞投影同步 [J]. *物理学报*, 2012, 61(18): 180503
- [15] Gao Yan-bo, Sun Bing-hua, Lu Guo-ping. Modified function projective lag synchronization of chaotic systems with disturbance estimations [J]. *Applied Mathematical Modelling*, 2013, 37(7): 4993-5000
- [16] Chai Xiu-li, Gan Zhi-hua, Shi Chun-xiao. Adaptive modified function projective lag synchronization of uncertain hyperchaotic dynamical systems with the same or different dimension and structure [J]. *Mathematical Problems in Engineering*, 2013, 2013: 282064
- [17] Fu Gui-yuan. Robust adaptive modified function projective synchronization of different hyperchaotic systems subject to external disturbance [J]. *Communications in Nonlinear Science and Numerical Simulation*, 2012, 17(6): 2602-2608
- [18] Yu Fei, Wang Chun-hua, Hu Yan, et al. Antisynchronization of a novel hyperchaotic system with parameter mismatch and external disturbances [J]. *Pramana-Journal of physics*, 2012, 79(1): 81-93
- [19] 邓玮, 方洁, 吴振军, 等. 含有不确定项的混沌系统自适应修正函数投影同步 [J]. *物理学报*, 2012, 61(14): 140503