

# 局部保持多投影向量 Fisher 判别分析算法

张 召<sup>1)</sup> 业 宁<sup>1),2)</sup> 业巧林<sup>3)</sup>

<sup>1)</sup>(南京林业大学信息科学技术学院 南京 210037)

<sup>2)</sup>(山东大学计算机科学与技术学院 济南 250100)

<sup>3)</sup>(南京理工大学计算机科学与技术学院 南京 210094)

**摘 要** 特征选择是在损失较少信息的情况下处理高维图像数据的关键技术,是高维数据预处理的重要步骤.通过引入 Fisher 判别分析(Fisher Discriminant Analysis, FDA)和典型相关分析(Canonical Correlation Analysis, CCA)的思想,采用以样本的类标号形式给出的先验信息,考虑样本数据的局部性,提出了一种监督的基于 Fisher 判别信息的局部保持多投影向量分析方法(Locality Preserving Multi-projection Vector Fisher Discriminant Analysis, LPMVF).通过定义新准则,LPMVF 具有以下优点:(1)便于计算,可有效避免奇异性;(2)借助标准核映射,可快速将 LPMVF 推广到非线性的特征空间;(3)与 CCA 算法类似,LPMVF 最终得到一对投影变换,可有效嵌入样本数据,可将原始数据投影成一系列“有用的”特征形式,并使数据的投影在嵌入空间中更具可分离性;(4)与局部化的 Fisher 判别分析(Local Fisher Discriminant Analysis,简称 LFDA)相比,LPMVF 也能够有效保持数据样本间的局部近邻关系;(5)在大多数情况下,该文算法的学习能力甚至优于经典的 FDA、KFD 和 LFDA 算法.在几个标准数据集上的实验结果表明,LPMVF 及其非线性的推广算法能够提取出描述能力更强的特征信息,可有效利用类标号监督信息提高分类性能.

**关键词** 局部保持;多投影向量;特征选择;分类;判别分析

中图法分类号 TP18 DOI号: 10.3724/SP.J.1016.2010.00865

## Locality Preserving Multi-Projection Vector Fisher Discriminant Analysis

ZHANG Zhao<sup>1)</sup> YE Ning<sup>1),2)</sup> YE Qiao-Lin<sup>3)</sup>

<sup>1)</sup>(School of Information Science and Technology, Nanjing Forestry University, Nanjing 210037)

<sup>2)</sup>(School of Computer Science and Technology, Shandong University, Jinan 250100)

<sup>3)</sup>(School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094)

**Abstract** Feature selection has been an important preprocessing step in high-dimensional image data analysis without losing much intrinsic information. By introducing the ideas of Fisher Discriminant Analysis (FDA) and Canonical Correlation Analysis (CCA), the paper discusses the supervised feature selection problem where samples are accompanied with class labels and proposes a new Locality Preserving Multi-projection Vector Fisher Discriminant Analysis algorithm called LPMVF. LPMVF takes the local structure of the original data into account, so the multimodal samples data can be embedded appropriately. By defining the new guidelines, LPMVF has the following advantages: (1) LPMVF can be easily computed and can avoid the singular problems; (2) LPMVF can be easily extended to non-linear feature selection scenarios by employing the kernel trick; (3) Similar to CCA, LPMVF attempts to find two sets of basis vectors for two multivariate datasets of different classes, one for each class, which can project the original data

收稿日期:2009-07-20;最终修改稿收到日期:2010-01-17. 本课题得到江苏省自然科学基金(BK2009393)、国家自然科学基金(30671639)、江苏省高校科技创新计划项目(164070265)、南京林业大学科技创新项目(2009106)、2009年江苏省研究生创新基金(CX09S\_013Z)资助. 张 召,男,1984年生,硕士研究生,主要研究方向为机器学习、智能信息处理、模式识别. E-mail: zzhang618@gmail.com. 业 宁,男,1967年生,博士,副教授,主要研究方向为模式识别、数据挖掘、生物信息学. 业巧林,男,1982年生,博士研究生,主要研究方向为数据挖掘、机器学习、模式识别.

onto a set of more useful features in the found embedding space, which would be benefit to classification and pattern recognition; (4) The same with Local Fisher Discriminant Analysis (LFDA), LPMVF can preserve the local relationships between the data points; (5) In most cases, the learning performance of the LPMVF method is superior to those of the classical FDA, KFD and LFDA algorithms. The authors verify the feasibility and effectiveness of LPMVF by extensive visualization and classification tasks. Experimental results on the benchmark datasets show that LPMVF and its nonlinear extended algorithm can extract the good features and effectively improve the accuracy by introducing the class labels as priori knowledge.

**Keywords** locality preservation; multi-projection vector; feature selection; classification; discriminant analysis

## 1 引 言

越来越多的研究和应用领域需要处理较高维的图像数据,而要对高维的数据进行统计分析和可视化研究,必须进行降维.降维的目的是将高维的数据嵌入到一个较低维的特征空间,同时保持样本的内部结构特征.线性 Fisher 判别分析(Linear Fisher Discriminant Analysis, FDA)<sup>[1]</sup>和核 Fisher 判别分析(Kernel Fisher Discriminant Analysis, KFD)<sup>[2]</sup>在模式识别中的特征选择和数据可视化研究中得到了广泛的应用和推广<sup>[3-6]</sup>.FDA 借助方差分析的思想,选择一组最优的投影向量,使得投影到低维空间的样本数据具有较好的可分离性,同时使在投影空间中的不同类样本间的差异尽可能大,是一种单模态的分析方法.所谓单模态识别是指利用从单一信息渠道获得的观察样本进行识别的技术.FDA 算法有一个明显的局限性是只有当属于不同类的样本分布满足同方差时得到的解才是最优的.KFD 是线性 Fisher 判别分析的非线性推广算法,考虑将样本数据映射到一个更高维的核特征空间,在核空间内进行 Fisher 线性判别,隐含地实现原输入空间的非线性判别,找到具有线性决策边界的基于核的分类规则,FDA 和 KFD 都是经实验验证过有效方法.

为了更好地嵌入样本数据,保持数据的局部特征是十分重要的,这使得基于局部保持的特征选择方法<sup>[7-10]</sup>成为近年来机器学习和智能信息挖掘领域中的热点问题.局部保持投影算法(Locality Preserving Projections, LPP)<sup>[7]</sup>能有效保持数据的局部特征,使位于原始输入空间中距离较近的样本点在嵌入特征空间内仍较为紧凑,并可以有效地嵌入多模态数据.Sugiyama 考虑数据的局部性,结合了 FDA

和 LPP 算法的良好特性,提出了局部化的监督型 Fisher 判别分析(Local Fisher Discriminant Analysis, LFDA)算法<sup>[8,10]</sup>,用于特征选择.实验证明 LFDA 算法能够有效地嵌入多模态的样本数据,并使属于不同类别的样本在构造出的特征空间内更具可分离性.LFDA 与 FDA 和 KFD 类似,其目标都是找到一组最优的投影变换,使得投影在该变换上的不同类别的样本更具可分离性.典型相关分析(Canonical Correlation Analysis, CCA)<sup>[11]</sup>侧重于多模态的识别,即利用互补原理,最大化不同模态数据之间的互信息相关,减少数据之间的不确定性,从而达到增强识别能力的目的,近年来,算法 CCA 已经被成功地应用于人脸表情识别<sup>[12]</sup>、图像处理<sup>[13]</sup>和模式识别<sup>[14]</sup>等研究和应用领域.不同于 FDA 及其变型的是,对于给定的两个属于不同类别的多变量数据集  $X$  和  $Y$ ,CCA 的目标是求得一对基向量,使两个数据集之间的相关最大,然而 CCA 是一种无监督的学习方法,只关注成对样本之间的相关性,并将相关作为不同空间中样本之间的相似性度量标准,没有对样本的类信息加以利用,而在很多实际应用研究中,得到与数据样本相关的一些先验知识是至关重要的,如类标号信息和约束信息等.

本文以样本的类标号作为先验信息,考虑样本数据的局部性,提出了一种监督的基于 Fisher 信息的局部保持多投影向量判别分析算法(Locality Preserving Multi-projection Vector Fisher Discriminant Analysis, LPMVF),用于特征选择,可有效弥补上述算法中存在的不足.

本文第 2 节给出 LPMVF 算法的详细推导过程和特性分析,并借助核映射,将 LPMVF 推广到非线性的特征空间.通过引入样本数据的子模式结构<sup>[15-17]</sup>,构造子模式特征空间,提出了另一种子模式

型非线性算法;第3节通过可视化和分类实验对提出的算法进行性能测试,并对实验结果进行分析;最后对本文的工作进行总结,并展望下一步的研究目标。

## 2 局部保持多投影向量 Fisher 判别分析算法

### 2.1 线性 LPMVF

本节提出了一种线性的基于 Fisher 信息的局部保持多投影向量判别分析方法(LPMVF),用于样本特征的选择和提取工作.在  $n$  维的输入空间  $\mathbb{R}^n$  上,对给定的  $m$  个样本  $\mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2) = (\mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_{m_1}^1, \mathbf{x}_1^2, \mathbf{x}_2^2, \dots, \mathbf{x}_{m_2}^2)$ ,该集合由第 1 类样本组成的  $n \times m_1$  维的矩阵  $\mathbf{A}$  和第 2 类样本组成的  $n \times m_2$  维的矩阵  $\mathbf{B}$  构成,其中  $m_1$  和  $m_2$  分别为两类样本的数目,且满足  $m_1 = m_2$ .对给定的两个属于不同类别的多变量数据集  $\mathbf{x}^1$  和  $\mathbf{x}^2$ ,类似于典型相关分析(Canonical Correlation Analysis, CCA)<sup>[11,18]</sup> 的设计思想,LPMVF 的目标是通过求解两个简单的标准特征值问题找到一对有效的投影变换  $\mathbf{T}_{\text{LPMVF}}^1$  和  $\mathbf{T}_{\text{LPMVF}}^2$ ,使投影到变换  $\mathbf{T}_{\text{LPMVF}}^1$  或  $\mathbf{T}_{\text{LPMVF}}^2$  上的第一类数据样本离该类样本在  $\mathbf{T}_{\text{LPMVF}}^1$  或  $\mathbf{T}_{\text{LPMVF}}^2$  上的投影均值尽可能地近,而投影到  $\mathbf{T}_{\text{LPMVF}}^1$  或  $\mathbf{T}_{\text{LPMVF}}^2$  上的另一类数据样本离该投影均值尽可能地远.进而通过投影变换  $\mathbf{r}_l^i = (\mathbf{T}_{\text{LPMVF}}^l)^T \mathbf{x}_i^l (l=1,2)$  得到样本集  $\mathbf{x}^l$  的  $d \times m_l (d \leq n)$  维的低维描述  $\mathbf{r}^l$ ,实现降维,可使得低维变换后的具有不同类标号信息的数据之间更具可分离性,本文中记  $l$  为类标号.

注意到  $\mathbf{T}_x^T \mathbf{x}^l$  意味着利用变换  $\mathbf{T}_x$  能将输入模式  $\mathbf{x}^l$  投影成一系列更有用的特征集合<sup>[19]</sup>,理想地,那个特征集合应该尽可能地紧凑,即要求矩阵  $\mathbf{T}_x^T \mathbf{T}_x$  或  $\mathbf{T}_x$  具有较小的秩,并希望  $\mathbf{T}_x^T \mathbf{T}_x$  的秩等于  $\mathbf{T}_x$  的秩.目标是寻找一个关于  $\mathbf{T}_x^T \mathbf{T}_x$  的特征分解  $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ ,即矩阵  $\mathbf{T}_x$  的秩等于  $\|\mathbf{\Lambda}\|_0$ ,但直接对零范数进行最小化处理是不大实际的,所以在式(1)中的优化问题中利用欧几里德范数(L2-范数)进行近似,即  $\|\mathbf{\Lambda}\|_2 = \|\mathbf{T}_x\|_2$ .则 LPMVF 的第一个投影变换矩阵  $\mathbf{T}_{\text{LPMVF}}^1 = (\mathbf{T}_{x_1}, \mathbf{T}_{x_2}, \dots, \mathbf{T}_{x_d})$  可通过求解如下的模型得到,其中  $\mathbf{I}$  为单位矩阵.

$$\mathbf{T}_{\text{LPMVF}}^1 = \arg \min_{\mathbf{T}_x \in \mathbb{R}^{d \times m_1}} \frac{1}{2} \|\mathbf{T}_x\|_2^2 + \frac{P}{2} \sum_{i=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} \left( \mathbf{T}_x^T \mathbf{x}_i^1 - \mathbf{T}_x^T \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbf{x}_j^1 \right)^2 -$$

$$\frac{Q}{2} \sum_{i=1}^{m_2} \mathbf{M}_{i,j}^{(b)} \left( \mathbf{T}_x^T \mathbf{x}_i^2 - \mathbf{T}_x^T \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbf{x}_j^1 \right)^2$$

subject to  $\mathbf{T}_x^T \mathbf{T}_x = \mathbf{I}$  (1)

即投影到各个分量  $\mathbf{T}_{x_g} (g=1,2,\dots,d(d \leq n))$  上的第 1 类样本尽可能地离该类数据样本在向量  $\mathbf{T}_{x_g}$  上的投影均值近,而投影到  $\mathbf{T}_{x_g}$  上的另一类样本应离该投影均值尽可能地远.模型(1)中的协调参数  $P$  和  $Q$  分别被用于平衡模型(1)中后两项的贡献程度,直觉地,属于同一类别的数据样本与该类样本在  $\mathbf{T}_x$  上的投影均值之间的距离应尽可能地接近期望的尺度<sup>[20]</sup>,经验地设置  $P=1$  和  $Q>1$ ,且该模型可有效避免 Fisher 判别过程中可能产生的奇异性问题.特征矩阵  $\mathbf{M}_{i,j}^{(w)}$  和  $\mathbf{M}_{i,j}^{(b)}$  被用于实现 LPMVF 算法的局部保持特性.假设  $G$  表示一个具有  $m$  个节点的图,若样本  $\mathbf{x}_i^l (\mathbf{x}_i^2)$  和  $\mathbf{x}_j^l (\mathbf{x}_j^2)$  是局部近邻,则在二者之间设置一条边作为权重,常用的局部近邻定义方法有如下两种策略<sup>[7,21]</sup>:(1)若样本  $\mathbf{x}_i^l (\mathbf{x}_i^2)$  和  $\mathbf{x}_j^l (\mathbf{x}_j^2)$  间的距离尺度小于  $\epsilon$ ,则认为二者是局部近邻关系,其中  $\epsilon \in R$  是一个预先设定的阈值;(2)若样本  $\mathbf{x}_i^l (\mathbf{x}_i^2)$  是  $\mathbf{x}_j^l (\mathbf{x}_j^2)$  的  $K$  近邻,则样本  $\mathbf{x}_i^l (\mathbf{x}_i^2)$  和  $\mathbf{x}_j^l (\mathbf{x}_j^2)$  是局部近邻.根据局部近邻的概念,本文中由规则(2)定义不同类别内部样本间的权重系数  $\mathbf{A}_{ij}^1$  和  $\mathbf{A}_{ij}^2$ ,其中  $\|\cdot\|^2$  为 L2-范数的平方.

$$\mathbf{A}_{ij}^l = \begin{cases} \exp(-\|\mathbf{x}_i^l - \mathbf{x}_j^l\|^2 / m_l), & \text{若 } \mathbf{x}_i^l \text{ 是 } \mathbf{x}_j^l \text{ 的近邻}, l=1,2 \\ 0, & \text{否则} \end{cases}$$

(2)

规则(2)中的参数  $m_l$  为第  $l$  类样本的数量,被用于平衡权重矩阵  $\mathbf{A}_{ij}^1$  和  $\mathbf{A}_{ij}^2$ .事实上,规则(2)体现了数据样本点与其周围的样本点之间的局部近邻关系,也就是说,若  $\|\mathbf{x}_i^1 - \mathbf{x}_j^1\|$  和  $\|\mathbf{x}_i^2 - \mathbf{x}_j^2\|$  的范数越小,则权重系数  $\mathbf{A}_{ij}^1$  和  $\mathbf{A}_{ij}^2$  越大,反之,若  $\|\mathbf{x}_i^1 - \mathbf{x}_j^1\|$  和  $\|\mathbf{x}_i^2 - \mathbf{x}_j^2\|$  的范数越大,则权重  $\mathbf{A}_{ij}^1$  和  $\mathbf{A}_{ij}^2$  越小,且  $\mathbf{A}_{ij}^l (l=1,2)$  为对称矩阵.若令对角矩阵  $\mathbf{R}_{11} = \sum_i \mathbf{A}_{ij}^1 \mathbf{A}_{ij}^1, \mathbf{R}_{22} = \sum_i \mathbf{A}_{ij}^2 \mathbf{A}_{ij}^2$  和  $\mathbf{R}_{21} = \sum_i \mathbf{A}_{ij}^2 \mathbf{A}_{ij}^1 = \mathbf{R}_{12}$ ,则问题(1)中的特征矩阵  $\mathbf{M}_{i,j,l}^{(w)}$  和  $\mathbf{M}_{i,j,l}^{(b)}$  可表示为类似于文献<sup>[7-8,10]</sup>中的定义准则,如规则(3)所示,其中,  $\mathbf{M}_{i,j,l}^{(w)}$  表示对应第  $l$  类样本的类内局部特征矩阵.

$$\mathbf{M}_l^{(w)} = \mathbf{R}_{ll} - \mathbf{A}^l \mathbf{A}^l, l=1,2,$$

$$\mathbf{M}^{(b)} = \mathbf{R}_{21} - \mathbf{A}^2 \mathbf{A}^1 = (\mathbf{M}^{(b)})^T$$

(3)

**定理 1.** 算法 LPMVF 的第一个投影变换矩

阵  $\mathbf{T}_{\text{LPMVF}}^1$  可通过求解矩阵  $(\mathbf{I} + \mathbf{P}\tilde{\mathbf{A}} - \mathbf{Q}\tilde{\mathbf{B}})$  的前  $d$  个最小的特征值对应的特征向量得到。

证明. 为了便于计算, 将问题(1)中的优化模型化简为下列形式:

$$\begin{aligned} \mathbf{T}_{\text{LPMVF}}^1 = \arg \min_{\mathbf{T}_x \in \mathbb{R}^{d \times m_1}} & \frac{1}{2} \|\mathbf{T}_x\|^2 + \frac{P}{2} \mathbf{T}_x^T \tilde{\mathbf{A}} \mathbf{T}_x - \frac{Q}{2} \mathbf{T}_x^T \tilde{\mathbf{B}} \mathbf{T}_x \\ \text{subject to } & \mathbf{T}_x^T \mathbf{T}_x = \mathbf{I} \end{aligned} \quad (4)$$

其中, 矩阵  $\tilde{\mathbf{A}}$  和  $\tilde{\mathbf{B}}$  满足:

$$\tilde{\mathbf{A}} = \sum_{i=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} \left( \mathbf{x}_i^1 - \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbf{x}_j^1 \right) \left( \mathbf{x}_i^1 - \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbf{x}_j^1 \right)^T \quad (5)$$

$$\tilde{\mathbf{B}} = \sum_{i=1}^{m_2} \mathbf{M}_{i,j}^{(b)} \left( \mathbf{x}_i^2 - \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbf{x}_j^1 \right) \left( \mathbf{x}_i^2 - \frac{1}{m_1} \sum_{j=1}^{m_1} \mathbf{x}_j^1 \right)^T \quad (6)$$

构建模型(4)的 Lagrange 函数, 可以得到

$$\begin{aligned} L(\mathbf{T}_x, \lambda^1) = & \frac{1}{2} \mathbf{T}_x^T \mathbf{T}_x + \frac{P}{2} \mathbf{T}_x^T \tilde{\mathbf{A}} \mathbf{T}_x - \\ & \frac{Q}{2} \mathbf{T}_x^T \tilde{\mathbf{B}} \mathbf{T}_x - \frac{\lambda^1}{2} (\mathbf{T}_x^T \mathbf{T}_x - \mathbf{I}) \end{aligned}$$

通过计算  $\frac{\partial L(\mathbf{T}_x, \lambda^1)}{\partial \mathbf{T}_x} = 0$ , 可以得到如下的标准

特征值问题:

$$(\mathbf{I} + \mathbf{P}\tilde{\mathbf{A}} - \mathbf{Q}\tilde{\mathbf{B}}) \mathbf{T}_x = \lambda^1 \mathbf{T}_x \quad (7)$$

式(7)中的标准特征值问题可通过特征值分解(例如奇异值分解等方法)快速有效地完成求解, 即矩阵  $(\mathbf{I} + \mathbf{P}\tilde{\mathbf{A}} - \mathbf{Q}\tilde{\mathbf{B}})$  的前  $d$  个最小的特征值  $\lambda_1^1 \geq \lambda_2^1 \geq \dots \geq \lambda_d^1$  所对应的特征向量即为模型(4)中的 LPMVF 优化准则下的第一个最优变换矩阵  $\mathbf{T}_{\text{LPMVF}}^1$  的各投影分量. 此时, 定理 1 得证. 证毕.

同理可得投影变换  $\mathbf{T}_{\text{LPMVF}}^2$  的各分量  $\mathbf{T}_{y_g}$ ,  $g=1, 2, \dots, d$ , 并使得投影到  $\mathbf{T}_{\text{LPMVF}}^2$  上的第一类样本尽可能地离该类样本在变换  $\mathbf{T}_{\text{LPMVF}}^2$  上的投影均值近, 而投影到  $\mathbf{T}_{\text{LPMVF}}^2$  上的另一类样本离该投影均值尽可能地远. 第二个投影变换  $\mathbf{T}_{\text{LPMVF}}^2 = (\mathbf{T}_{y_1}, \mathbf{T}_{y_2}, \dots, \mathbf{T}_{y_d})$  可通过求解如模型(8)中的最小化问题得到:

$$\begin{aligned} \mathbf{T}_{\text{LPMVF}}^2 = \arg \min_{\mathbf{T}_y \in \mathbb{R}^{d \times m_2}} & \frac{1}{2} \|\mathbf{T}_y\|^2 + \frac{P}{2} \mathbf{T}_y^T \tilde{\mathbf{C}} \mathbf{T}_y - \frac{Q}{2} \mathbf{T}_y^T \tilde{\mathbf{D}} \mathbf{T}_y \\ \text{subject to } & \mathbf{T}_y^T \mathbf{T}_y = \mathbf{I} \end{aligned} \quad (8)$$

类似地, 矩阵  $\tilde{\mathbf{C}}$  和  $\tilde{\mathbf{D}}$  满足:

$$\begin{cases} \tilde{\mathbf{C}} = \sum_{i=1}^{m_2} \mathbf{M}_{i,j,2}^{(w)} \left( \mathbf{x}_i^2 - \frac{1}{m_2} \sum_{j=1}^{m_2} \mathbf{x}_j^2 \right) \left( \mathbf{x}_i^2 - \frac{1}{m_2} \sum_{j=1}^{m_2} \mathbf{x}_j^2 \right)^T \\ \tilde{\mathbf{D}} = \sum_{i=1}^{m_1} (\mathbf{M}_{i,j}^{(b)})^T \left( \mathbf{x}_i^1 - \frac{1}{m_2} \sum_{j=1}^{m_2} \mathbf{x}_j^2 \right) \left( \mathbf{x}_i^1 - \frac{1}{m_2} \sum_{j=1}^{m_2} \mathbf{x}_j^2 \right)^T \end{cases} \quad (9)$$

通过求解问题(8), 可以得到矩阵  $(\mathbf{I} + \mathbf{P}\tilde{\mathbf{C}} -$

$\mathbf{Q}\tilde{\mathbf{D}})$  的前  $d$  个较小的特征值  $\lambda_1^2 \geq \lambda_2^2 \geq \dots \geq \lambda_d^2$  所对应的特征向量, 而这些特征向量则构成了第二个投影变换  $\mathbf{T}_{\text{LPMVF}}^2$  的各个分量. 变换  $\mathbf{T}_{\text{LPMVF}}^2$  的计算过程和求解  $\mathbf{T}_{\text{LPMVF}}^1$  的过程相似, 这里不再给出详细的过程. 实验中的最优特征数量  $d$  的值不是固定的, 而是与每个数据集有着直接的关系, 本文将通过在不同的数据集上的仿真实验确定最优的投影维数.

## 2.2 基于核空间的非线性 KLPMVF

上述内容都只是涉及到线性的 LPMVF 算法, 本节中借助核映射<sup>[22]</sup>将线性 LPMVF 算法推广为非线性的 KLPMVF 算法. 我们知道, 核映射的效果与其核函数的选取有着直接的关系, 按照文献[23]所述, 核是输入空间的一个函数, 又是特征空间中通过核函数非线性映射的内积表示. 特别地, 核函数可定义为如下形式:  $K(\mathbf{x}, \mathbf{y}) = \langle \phi(\mathbf{x}), \phi(\mathbf{y}) \rangle = (\phi(\mathbf{x}))^T \phi(\mathbf{y})$ . 对每个核函数, 都存在核映射  $\phi$  可将样本数据  $\mathbf{x}_i$  投影成高维的  $\phi(\mathbf{x}_i)$ ,  $i=1, 2, \dots, m$ . 在本文中, 我们定义了如下特征映射:  $\phi: \mathbb{R}^n \rightarrow \mathbb{H}^p$  ( $p > n$ ), 在核特征空间  $\mathbb{H}$  中, 原始的输入模式可用  $\phi(\mathbf{x}^l) = (\phi(\mathbf{x}_1^l), \phi(\mathbf{x}_2^l), \dots, \phi(\mathbf{x}_{m_l}^l))$ ,  $l=1, 2$  进行描述. 在实验中, 总是选取核参数  $\sigma=10$  的 RBF 核函数用于数据的投影, RBF 核可表示如下:

$$K(\mathbf{x}, \mathbf{x}^T) = \exp(-\|\mathbf{x} - \mathbf{x}^T\|^2 / 2\sigma^2) \quad (10)$$

Mika 等在文献[2]中已经证明, 任何一个属于核特征空间  $\mathbb{H}$  中的向量, 均可按照被映射的样本数据进行展开, 也就是说, 位于高维核空间  $\mathbb{H}$  中的投影变换  $\mathbf{T}_x$  和  $\mathbf{T}_y$  可被重写如下形式:

$$\mathbf{T}_x^\phi = \sum_{i=1}^{m_1} \alpha_{x_i} \phi(\mathbf{x}_i^1) = \phi(\mathbf{x}^1) \alpha_x, \quad (11)$$

$$\mathbf{T}_y^\phi = \sum_{i=1}^{m_2} \alpha_{y_i} \phi(\mathbf{x}_i^2) = \phi(\mathbf{x}^2) \alpha_y$$

其中,  $\phi(\mathbf{x}^1) = (\phi(\mathbf{x}_1^1), \phi(\mathbf{x}_2^1), \dots, \phi(\mathbf{x}_{m_1}^1))$ ,

$$\phi(\mathbf{x}^2) = (\phi(\mathbf{x}_1^2), \phi(\mathbf{x}_2^2), \dots, \phi(\mathbf{x}_{m_2}^2)).$$

规则(11)中的标号  $m_l$ ,  $l=1, 2$  分别为两类样本的数目, 且需要最终求取的投影变换  $\alpha_x$  和  $\alpha_y$  为

$$\alpha_x = (\alpha_{x_1}, \alpha_{x_2}, \dots, \alpha_{x_d}), \quad \alpha_y = (\alpha_{y_1}, \alpha_{y_2}, \dots, \alpha_{y_d})$$

则矩阵  $\tilde{\mathbf{A}}$  和  $\tilde{\mathbf{B}}$  在核空间  $\mathbb{H}$  中可表示为如下形式:

$$\begin{aligned} (\tilde{\mathbf{A}})^\phi &= \sum_{i=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} \left( \phi(\mathbf{x}_i^1) - \frac{1}{m_1} \sum_{j=1}^{m_1} \phi(\mathbf{x}_j^1) \right) \cdot \\ & \left( \phi(\mathbf{x}_i^1) - \frac{1}{m_1} \sum_{j=1}^{m_1} \phi(\mathbf{x}_j^1) \right)^T \\ (\tilde{\mathbf{B}})^\phi &= \sum_{i=1}^{m_2} \mathbf{M}_{i,j}^{(b)} \left( \phi(\mathbf{x}_i^2) - \frac{1}{m_1} \sum_{j=1}^{m_1} \phi(\mathbf{x}_j^1) \right) \cdot \end{aligned} \quad (12)$$

$$\left( \phi(\mathbf{x}_i^2) - \frac{1}{m_1} \sum_{j=1}^{m_1} \phi(\mathbf{x}_j^1) \right)^\top \quad (13)$$

此时令  $\bar{\theta}_l^\phi = \frac{1}{m_l} \sum_{j=1}^{m_l} \phi(\mathbf{x}_j^l)$ ,  $l=1, 2$ ,  $\mathbf{R}_{ii} = \sum_j \mathbf{M}_{i,j}^{(w)}$ , 其中,  $\mathbf{M}_{i,j}^{(w)}$  为矩阵  $\mathbf{M}_i^{(w)}$  的第  $(i, j)$  个输入元素, 将规则(12)和(13)代入  $(\mathbf{T}_x^\top \hat{\mathbf{A}} \mathbf{T}_x)^\phi$  和  $(\mathbf{T}_x^\top \hat{\mathbf{B}} \mathbf{T}_x)^\phi$  后, 可得

$$\begin{aligned} & (\mathbf{T}_x^\top \hat{\mathbf{A}} \mathbf{T}_x)^\phi = \\ & \alpha_x^\top (\varphi(\mathbf{x}^1))^\top \sum_{i=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} (\varphi(\mathbf{x}_i^1) - \bar{\theta}_1^\phi) \cdot \\ & (\varphi(\mathbf{x}_i^1) - \bar{\theta}_1^\phi)^\top \varphi(\mathbf{x}^1) \alpha_x = \\ & \alpha_x^\top (\varphi(\mathbf{x}^1))^\top \left( \sum_{i=1}^{m_1} \left( \sum_{j=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} \right) \varphi(\mathbf{x}_i^1) (\varphi(\mathbf{x}_i^1))^\top + \right. \\ & \left. \sum_{j=1}^{m_1} \left( \sum_{i=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} \right) \bar{\theta}_1^\phi (\bar{\theta}_1^\phi)^\top - \sum_{i=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} \varphi(\mathbf{x}_i^1) (\bar{\theta}_1^\phi)^\top - \right. \\ & \left. \sum_{i=1}^{m_1} \mathbf{M}_{i,j,1}^{(w)} \bar{\theta}_1^\phi (\varphi(\mathbf{x}_i^1))^\top \right) \varphi(\mathbf{x}^1) \alpha_x = \\ & \alpha_x^\top (\varphi(\mathbf{x}^1))^\top \left( \left( \frac{1}{m_1^2} + 1 \right) \sum_i \varphi(\mathbf{x}_i^1) \mathbf{R}_{ii} (\varphi(\mathbf{x}_i^1))^\top - \right. \\ & \left. \frac{2}{m_1} \varphi(\mathbf{x}^1) \mathbf{M}_1^{(w)} (\varphi(\mathbf{x}^1))^\top \right) \varphi(\mathbf{x}^1) \alpha_x = \\ & \alpha_x^\top (\varphi(\mathbf{x}^1))^\top \varphi(\mathbf{x}^1) \left( \left( \frac{1}{m_1 \times m_1} + 1 \right) \mathbf{R} - \frac{2}{m_1} \mathbf{M}_1^{(w)} \right) \cdot \\ & (\varphi(\mathbf{x}^1))^\top \varphi(\mathbf{x}^1) \alpha_x \\ & = \alpha_x^\top \mathbf{K}_{11} \tilde{\Psi} \mathbf{K}_{11}^\top \alpha_x \end{aligned} \quad (14)$$

其中,  $\mathbf{K}_{li} = (\phi(\mathbf{x}^l))^\top \phi(\mathbf{x}^l)$ ,  $l=1, 2$ ,  $\mathbf{K}_{12} = (\phi(\mathbf{x}^1))^\top \cdot \phi(\mathbf{x}^2)$ . 再令  $\mathbf{Q}_{jj} = \sum_i \mathbf{M}_{i,j}^{(b)}$ ,  $\mathbf{U}_{ii} = \sum_j \mathbf{M}_{i,j}^{(b)}$ , 则  $(\mathbf{T}_x^\top \hat{\mathbf{B}} \mathbf{T}_x)^\phi$  可转化为

$$\begin{aligned} & (\mathbf{T}_x^\top \hat{\mathbf{B}} \mathbf{T}_x)^\phi = \alpha_x^\top (\phi(\mathbf{x}^1))^\top \sum_{i=1}^{m_2} \mathbf{M}_{i,j}^{(b)} (\phi(\mathbf{x}_i^2) - \bar{\theta}_1^\phi) \cdot \\ & (\phi(\mathbf{x}_i^2) - \bar{\theta}_1^\phi)^\top \phi(\mathbf{x}^1) \alpha_x \\ & = \alpha_x^\top \left( \mathbf{K}_{12} \mathbf{Q} \mathbf{K}_{12}^\top + \frac{1}{m_1 \times m_1} \mathbf{K}_{11} \mathbf{U} \mathbf{K}_{11} - \right. \\ & \left. \frac{2}{m_1} \mathbf{K}_{12} \mathbf{M}^{(b)} \mathbf{K}_{11} \right) \alpha_x \\ & = \alpha_x^\top \tilde{\Phi} \alpha_x \end{aligned} \quad (15)$$

**定理 2.** 算法 KLPMVF 的投影变换  $\alpha_x$  可通过求解矩阵  $(\mathbf{K}_{11} + \mathbf{P} \mathbf{K}_{11} \tilde{\Psi} \mathbf{K}_{11}^\top - \mathbf{Q} \tilde{\Phi})$  的前  $d$  个最小的特征值对应的特征向量得到.

证明. 将规则(14)和(15)代入模型(4)后, 模型(4)可重写为

$$\begin{aligned} & \min_{\alpha_x} \frac{1}{2} \alpha_x^\top \mathbf{K}_{11} \alpha_x + \frac{P}{2} \alpha_x^\top \mathbf{K}_{11} \tilde{\Psi} \mathbf{K}_{11}^\top \alpha_x - \frac{Q}{2} \alpha_x^\top \tilde{\Phi} \alpha_x \\ & \text{subject to } \alpha_x^\top \mathbf{K}_{11} \alpha_x = I \end{aligned} \quad (16)$$

这里, 类似于求解线性 LPMVF 的两个投影变

换的策略, 此时引入模型(16)的 Lagrange 函数, 并对  $\alpha_x$  求偏导数并置 0 后, 可得

$$(\mathbf{K}_{11} + \mathbf{P} \mathbf{K}_{11} \tilde{\Psi} \mathbf{K}_{11}^\top - \mathbf{Q} \tilde{\Phi}) \alpha_x = \lambda_1^\phi (\mathbf{I} + \mathbf{K}_{11}) \alpha_x \quad (17)$$

其中,  $\beta$  为正的泛化因子, 通过计算问题(17)可得投影变换  $\alpha_x$ , 且计算过程可避免奇异性. 定理 2 得证.

证毕.

同理, 为了获得投影变换  $\alpha_y$ , 令  $\mathbf{G}_{ii} = \sum_j \mathbf{M}_{i,j,2}^{(w)}$ ,  $\mathbf{E}_{jj} = \sum_i (\mathbf{M}_{i,j}^{(b)})^\top$ ,  $\mathbf{F}_{ii} = \sum_j (\mathbf{M}_{i,j}^{(b)})^\top$ , 则  $\alpha_y$  可通过与问题(17)相类似的特征值问题(18)计算得到

$$(\mathbf{K}_{22} + \mathbf{P} \mathbf{K}_{22} \tilde{\mathbf{M}} \mathbf{K}_{22}^\top - \mathbf{Q} \tilde{\mathbf{N}}) \alpha_y = \lambda_2^\phi (\mathbf{I} + \mathbf{K}_{22}) \alpha_y \quad (18)$$

其中,

$$\tilde{\mathbf{N}} = \left( \mathbf{K}_{12}^\top \mathbf{E} \mathbf{K}_{12} + \frac{1}{m_2 \times m_2} \mathbf{K}_{22} \mathbf{F} \mathbf{K}_{22} - \frac{2}{m_2} \mathbf{K}_{12}^\top \mathbf{M}^{(b)} \mathbf{K}_{22} \right),$$

$$\tilde{\mathbf{M}} = \left( \frac{1}{m_2 \times m_2} + 1 \right) \mathbf{G} - \frac{1}{m_2} \mathbf{M}_2^{(w)}.$$

KLPMVF 的有效性也将通过实验进行验证. 接下来, 将在实现 KLPMVF 算法的基础上, 考虑样本数据的子模式结构, 将 KLPMVF 推广为另外一种形式的非线性方法.

### 2.3 子模式判别型 SpKLPMVF

子模式判别方法在人脸识别<sup>[15-16]</sup>和矩阵分解<sup>[17]</sup>等研究领域得到了广泛的应用和推广. 在本节中, 我们先将原始样本数据矩阵进行重组, 采用样本数据的子模式作为新的输入形式, 构造子模式特征空间, 试图进一步增强非线性 KLPMVF 算法的学习性能, 提出了一种子模式判别型非线性 KLPMVF 算法, 本文中简记为 SpKLPMVF. 下面将给出 SpKLPMVF 算法的详细设计和实现过程:

对给定的  $m$  个对象, 若其属性值由维数大小为  $n \times m$  的矩阵  $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_m)$  来描述, 矩阵  $\mathbf{Z}$  的每一列表示一个样本对象. 若维数  $n$  可被整数  $t$  等分, 则原始的数据矩阵  $\mathbf{Z}$  可转化为大小为  $(n/t) \times mt$  的新矩阵  $\mathbf{Y}$ , 可表示如下

$$\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_t, \mathbf{y}_{t+1}, \dots, \mathbf{y}_{2t}, \dots, \mathbf{y}_{(m-1)t+1}, \dots, \mathbf{y}_m) \quad (19)$$

此时, 矩阵  $\mathbf{Z}$  中的每一个对象  $z_i$  能够被更新如规则(20)中的形式:

$$\mathbf{z}_i = (\mathbf{y}_{(i-1)t+1}^\top, \dots, \mathbf{y}_{it}^\top)^\top, \quad i=1, 2, \dots, m \quad (20)$$

根据式(19)中的运算规则, 可通过核映射计算得到维数大小为  $mt \times mt$  的核矩阵, 而该矩阵可描述为  $\mathbf{K} = \langle \phi(\mathbf{Y}), \phi(\mathbf{Y}) \rangle = (\phi(\mathbf{Y}))^\top \phi(\mathbf{Y})$ . 此外, 可利用式(17)和(18)中的规则对核矩阵  $\mathbf{K}$  进行特征分解, 得到维数大小为  $d \times mt$  的低维描述  $\mathbf{R} =$

$\{r_i\}_{i=1}^m$ , 其中,  $d$  是在投影空间中选择的特征数量, 则可将矩阵  $\mathbf{R}$  转换如规则(21)中给出的维数大小为  $dt \times m$  的新矩阵  $\mathbf{V}$ :

$$\mathbf{V} = \{(\mathbf{r}_{(i-1)p+1}^T, \dots, \mathbf{r}_{ip}^T)^T\}_{i=1}^m \quad (21)$$

类似地, 对任何一个新的数据样本  $\mathbf{u}$ , 首先将样本  $\mathbf{u}$  转换如规则(22)中给出的维数大小为  $(n/t) \times t$  的新矩阵  $\mathbf{Y}_{\text{new}}$ :

$$\mathbf{Y}_{\text{new}} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_p) \quad (22)$$

这里,  $\mathbf{u}$  满足  $\mathbf{u} = (\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_p^T)^T$ . 同样地, 根据式(17)和(18)中的规则对式(22)中给出的维数大小为  $mt \times t$  的核矩阵进行分解, 得到  $d \times t$  的低维描述  $\mathbf{R}_{\text{new}} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_p)$ . 由此, 可将矩阵  $\mathbf{R}_{\text{new}}$  转换为如规则(23)中给出的大小为  $dt \times 1$  的新向量  $\mathbf{V}_{\text{new}}$ , 向量  $\mathbf{V}_{\text{new}}$  可描述如下:

$$\mathbf{V}_{\text{new}} = (\mathbf{r}_1^T, \mathbf{r}_2^T, \dots, \mathbf{r}_p^T) \quad (23)$$

最后, 式(21)和(23)可用于进行实现降维操作和分类. 在实现中, 首先变换得到训练样本和测试样本对应的子模式作为新的输入形式, 进而利用本文算法进行特征的有效选择. 下面, 将通过大量的可视化和分类实验验证本文算法的学习性能和有效性.

### 3 仿真实验与分析

在本节中, 几个标准的 UCI 数据集<sup>[24]</sup> 和由 VTT Building Technology 实验室提供的木材样本图像数据库<sup>[25-26]</sup> 被作为实验的主要依据.

在实验阶段, 我们进行了一系列的实验来检验线性 LPMVF 及非线性 KLPMVF 和 SpKLPMVF 的学习性能, 首先通过两个典型的 UCI 数据集直观地考察利用本文算法提取出的特征的分布情况, 接着在 4 个 UCI 数据库上测试本文算法的分类能力, 最后在木材图像数据集上检验用算法 LPMVF 以及 KLPMVF 和 SpKLPMVF 进行特征选择后对识别能力的影响. 表 1 给出了用于实验的 UCI 数据集和木材数据集的描述. 参数  $Q$  被用于调节算法的性

能, 对于不同的特征选择方法和数据集, 参数  $Q$  的值是不一致的, 为了取得较好的学习性能, 参数  $Q$  将通过在各个数据集上的随机数测试得到, 即每组产生 1000 个随机数, 在 2~200 之间进行调节, 在每个数据集上进行 10 组实验, 选取最佳的参数用于实验. 在实验中, 泛化因子  $\beta$  总是取 0.0001. 对于所有数据集, 为了保证两类样本数据的数量相等, 实验前我们从数量较少的那类样本随机抽取相应数量的数据样本进行补充. 此外, 为了便于分类器的设计, 完成特征选择后,  $K$  近邻分类器(KNN)被用于分类. 实验环境: 装有 Matlab 7.1 的 Windows XP 系统, 机器配置为 Intel (R) Pentium(R) D CPU 2.80 GHz, 2.79 GHz, 512MB.

表 1 用于测试的 UCI 数据集的统计描述

数据集	样本维数	样本数量	类别数
Ionosphere	34	351	2
Iris	4	150	3
Diabetes	8	878	2
Vote	16	435	2
Soybean	8	4177	3
Tic-tac-toe1	27	958	2
wood image database	38	800	2

#### 3.1 数据可视化实验

在本实验中, 我们将本文方法与典型的 FDA、KFD 和 LFDA 在 Iris 和 Soybean 数据集上进行可视化比较实验, 直观地验证算法提取特征的描述能力和有效性. 用于测试的 Iris 和 Soybean 测试集都是 3 类的数据集, 而本文算法是针对两类问题提出的. 实验中分别利用(“×”)类和(“○”)类、(“×”)类和(“\*” )类、(“\*” )类和(“○”)类创建两类问题, 实现数据可视化后将实验结果绘制于同一幅示意图中. 图 1(a)和(b)分别给出了 Iris 和 Soybean 中的原始数据的分布情况. 图 2 和图 3 分别给出了采用算法 FDA、KFD、LFDA、LPMVF、KLPMVF 和 SpKLPMVF 在 Iris 和 Soybean 数据集上提取的特征在二维空间中的分布情况.

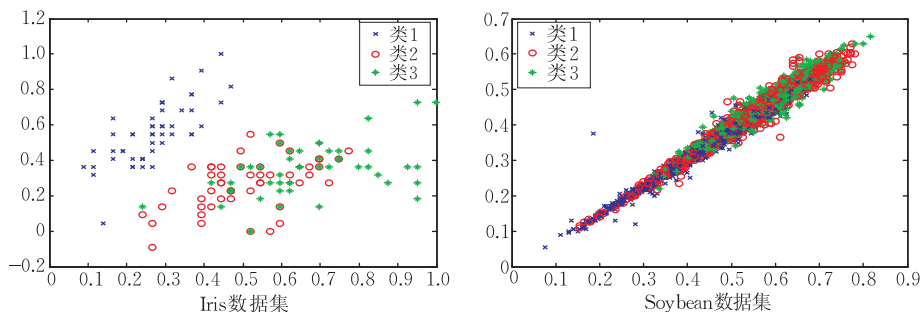


图 1 Iris 和 Soybean 数据集中的原始数据在二维空间中的分布, 符号 ×, ○, \* 分别表示 3 个不同类样本对应的特征

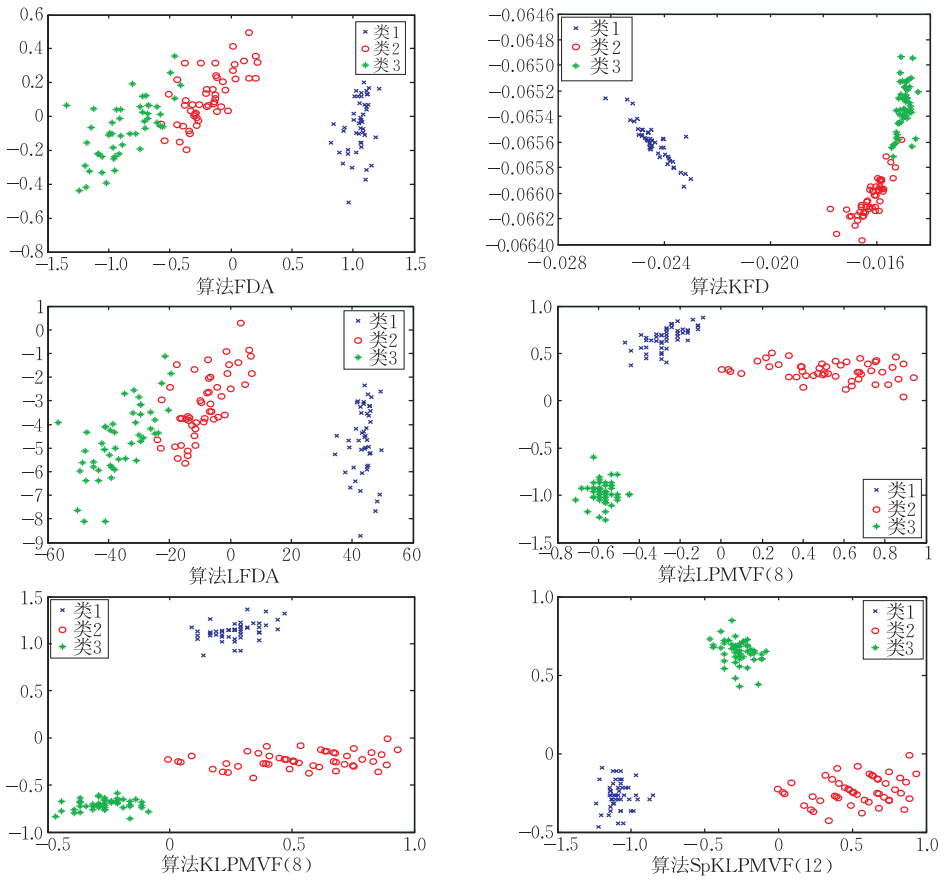


图2 算法FDA、KFD、LFDA、LPMVF(8)、KLPMVF(8)和 SpKLPMVF(12)在 Iris 数据集上提取的特征在二维空间中的分布

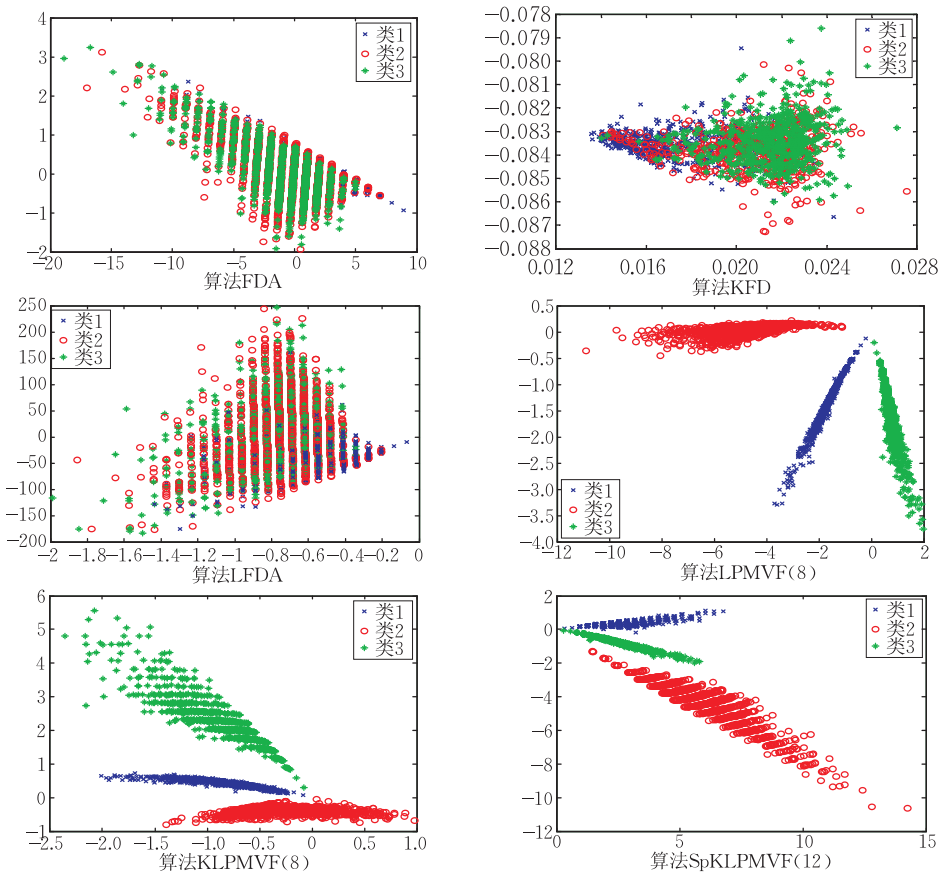


图3 算法FDA、KFD、LFDA、LPMVF(8)、KLPMVF(8)和 SpKLPMVF(12)在 Soybean 数据集上提取的特征在二维空间中的分布

可以看出:(1)从图 1(a)可以发现,Iris 数据集中的(“×”)类和(“\*”,“○”)类内的样本对应的数据点在原始的二维空间中接近线性可分,此时 FDA,KFD 和 LFDA 均取得了比较好的实验效果,即可将(“×”)类和(“\*”,“○”)类分离开,然而,符号(“○”)和(“\*”)代表的两类样本数据之间存在部分的重叠现象,这将会在分类中产生错分。(2)从图 1(b)可以看出,数据集 Soybean 中的 3 类样本((“×”),(“○”)和(“×”))在原始的二维空间中完全混叠在一起,几乎失去了可分离性.数据经过算法 FDA,KFD 和 LFDA 变换后,不同类的样本仍然不能被有效分开,不同类别的样本之间仍存在严重的混叠现象,将直接影响分类的精度。(3)算法 LPMVF,KLPMVF 和 SpKLPMVF 在 Iris[2×2] 和 Soybean[4×2]两个数据集上均取得了较好的实

验效果,不同类的样本数据可被有效分离开来,且算法能有效保持数据的局部特征,说明本文算法可提取出较好的描述能力较强的特征信息.其中,中括号内的数据是 SpKLPMVF 算法中重组矩阵的大小,小括号内的数值为参数  $Q$  的选取,在后面章节将不再做作出特别说明.可视化实验结果预示着用本文算法提取的特征更有利于作分类和模式识别.

### 3.2 基于 UCI 数据集的分类实验

3.1 节通过在两个典型数据集上的数据可视化实验直观地评价了本文算法的有效性,本实验中选取 Diabetes,Ionosphere,Vote 和 Tic-tac-toe1 4 个数据集验证 LPMVF,KLPMVF 和 SpKLPMVF 的分类能力.在本文的所有实验中,对于每个测试集,选取每类样本数量的一半作为训练样本,另一半作为测试样本,本实验中采用 1-NN 分类器进行分类.

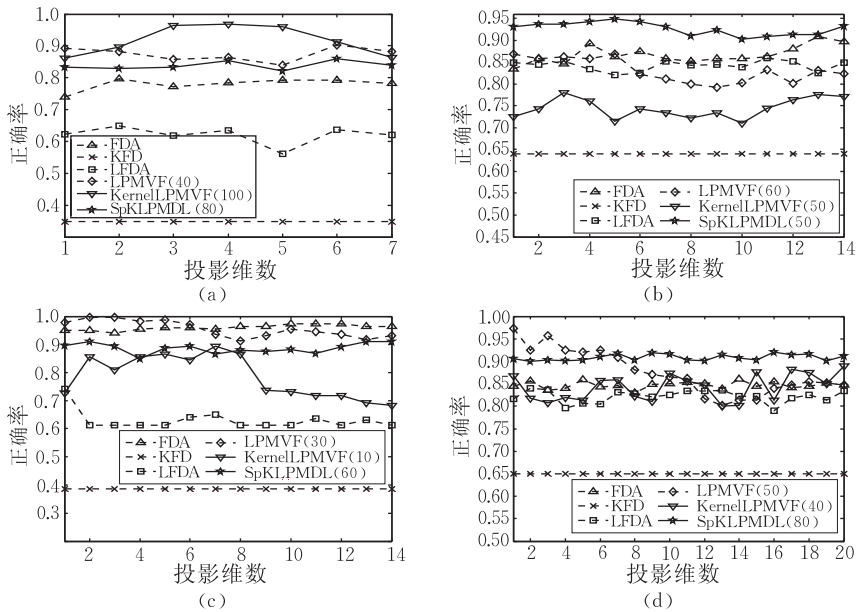


图 4 针对不同数量的选择的特征,各种算法在 Diabetes,Ionosphere,Vote 和 Tic-tac-toe1 4 个数据集上的实验比较

图 4 给出了 FDA,KFD,LFDA,LPMVF,KLPMVF 和 SpKLPMVF 算法随着选取的特征数量的不同,在 Diabetes,Ionosphere,Vote 和 Tic-tac-toe1 4 个 UCI 数据集的实验结果.在实验中,算法 SpKLPMVF 在 4 个分类数据集上的重组矩阵分别选取为 $[4 \times 2]$ , $[17 \times 2]$ , $[4 \times 4]$ 和 $[9 \times 3]$ .可以看出:(1)KFD 算法在 4 个数据集上的分类精度均处于底线,分析原因主要是 KFD 算法的目标是找到特征空间中的具有线性决策边界的基于核的分类规则,使得不同类样本间存在较严重的混叠.此时,FDA 算法取得了较高的分类精度,且在 Vote 数据集上取得了与 LPMVF 极为接近的分类结果,

LPMVF 在 Ionosphere 上的表现相对较差,仅优于 KFD;(2)针对抽取出的特征数量的不同,算法 LPMVF,KLPMVF 和 SpKLPMVF 体现出广泛的适应性,性能较为优越;(3)从图 4(b)和 4(d)中的结果可以看出,随着抽取的特征数量的不同,算法 SpKLPMVF 在 Ionosphere 和 Tic-tac-toe1 两个数据集上总是取得最高的分类精度,此时,LFDA 也取得了较好的实验结果,性能较好.

上述部分已对各种算法的分类能力和有效性加以了验证和评价,下面将对算法的时间复杂性加以分析和评价.表 2 和表 3 分别给出了对应于图 4 中的在上述 4 个数据集上的平均分类精度和总体运行

时间比较,可以看出:(1)随着投影维数的不断变化,算法 LPMVF、KLPMVF 和 SpKLPMVF 几乎总是取得了优于另外几种算法的分类精度,性能较好;(2)从各种算法的总体运行时间看,在大多数情况下,非线性的 KLPMVF 和 SpKLPMVF 运行需要花费的时间比线性 LPMVF 多,且运行时间接近 KFD 算法.KFD 在 Vote 数据集上花费了较低的运行时间,且运行时间少于 KLPMVF 和 SpKLPMVF

算法.线性 FDA 的时间复杂性保持最低,优于另外几种算法.LFDA 的运行时间几乎总是介于线性 FDA 和非线性 KFD 之间,除了 Vote 数据集,而 LPMVF 的时间复杂度与 LFDA 相接近.综上所述,当样本维数比样本数量相对较少时,本文算法可通过特征值问题有效完成求解,且整体性能较好,且时间复杂度较低.

表 2 不同算法在 4 个数据集上的平均分类精度结果比较

数据集	精度/%					
	FDA	KFD	LFDA	LPMVF	KLPMVF	SpKLPMVF
Diabetes	78.76	35.45	62.12	88.65	91.27	84.62
Ionosphere	86.63	64.58	79.38	83.81	75.30	92.56
Vote	96.55	39.38	62.27	97.63	79.12	89.57
Tic-tac-toe1	85.42	65.00	82.70	85.65	85.26	89.84

表 3 不同算法在 4 个数据集上的总体运行时间比较

数据集	运行时间/s					
	FDA	KFD	LFDA	LPMVF	KLPMVF	SpKLPMVF
Diabetes	0.5017	1.2280	0.9957	1.1840	1.6286	1.5363
Ionosphere	4.5586	5.3547	4.8240	5.1986	5.7879	5.6216
Vote	0.9076	2.1905	3.9956	4.2263	4.5224	4.3204
Tic-tac-toe1	4.9751	5.4829	4.9270	5.1217	5.7109	5.6296

### 3.3 木材节子缺陷识别实验

为了进一步检验 LPMVF 及非线性 (Sp) KLPMVF 算法的性能,在 VTT Building Technology 提供的木材图像数据集<sup>[25-26]</sup>上进行木材缺陷识别实验,并与经典的 FDA<sup>[1]</sup>、KFD<sup>[2]</sup>和 LFDA<sup>[8,10]</sup>算法作了对比.在实验中,先将原始木材图像进行 3 层对偶树复小波变换 (Dual-Tree Complex Wavelet Transform, DTCWT)<sup>[27]</sup>,选取基于不同方向、不同频率的能量统计特征作为实验分类的依据,完成特征抽取后,用  $K$  近邻法分类.该木材数据库包括 738 幅木材图片和 438 个已明确标定的节子缺陷类型,主要包括腐朽、油脂、木节和虫眼等,实验中通过手工方式将木材的原始图片剪裁成各种像素大小的子图像,接着采用 DTCWT 变换提取 38 维的特征作为实验的主要依据,其中包含有缺陷训练样本 400 个 (即正类样本,标定为 1) 和 400 个无缺陷训练样本 (即负类样本,标定为 -1). DTCWT 变换提取出的木材纹理特征可表示如下:

$$\sum_{L=1}^3 \sum_{h=1}^6 (FV\_e(L, h), FV\_std(L, h)) + \sum_{s=1}^6 (FV\_e(s), FV\_std(s)) \quad (24)$$

其中,  $L$  为分解层数,  $h$  为高频子带数量,  $s$  为低频子带数量,  $FV\_e$  和  $FV\_std$  分别为第  $L$  层逼近子带高斯分布的均值和标准差,该特征能有效地刻画木材图像的局部和全局特征.若矩阵  $\mathbf{A}_i = (\boldsymbol{\psi}_{i1}, \boldsymbol{\psi}_{i2}, \dots, \boldsymbol{\psi}_{in})$ ,  $\mathbf{A}_j = (\boldsymbol{\psi}_{j1}, \boldsymbol{\psi}_{j2}, \dots, \boldsymbol{\psi}_{jn})$  为两个特征集合,则  $\mathbf{A}_i$  和  $\mathbf{A}_j$  之间的距离尺度可定义为

$$dist(\mathbf{A}_i, \mathbf{A}_j) = \sum_{t=1}^n \|\boldsymbol{\psi}_{it} - \boldsymbol{\psi}_{jt}\|^2 \quad (25)$$

若矩阵  $\mathbf{A}_1$  和  $\mathbf{A}_2$  分别被用于存储有缺陷样本和无缺陷样本对应的特征,分别被赋予类标签  $l=1$  和  $-1$ ,此时,对给定的任意一个新的木材样本图像  $A$ ,若  $dist(A, \mathbf{A}_1) = \min dist(A, \mathbf{A}_j)$ ,且  $\mathbf{A}_1$  属于有缺陷样本 (即类别标签  $l=1$ ),则将样本  $A$  判别为有缺陷区域,反之认为该样本为正常区域.图 5 给出各种算法从木材特征集中提取的特征的分布,可以看出有缺陷和无缺陷样本特征在 LPMVF、KLPMVF 和 SpKLPMVF 的嵌入空间中更加易于分类,由此可训练得到性能优良的分类器模型,提高识别率.此时, KFD 算法将两类数据投影成为一个孤立点,这里不给出示意图.

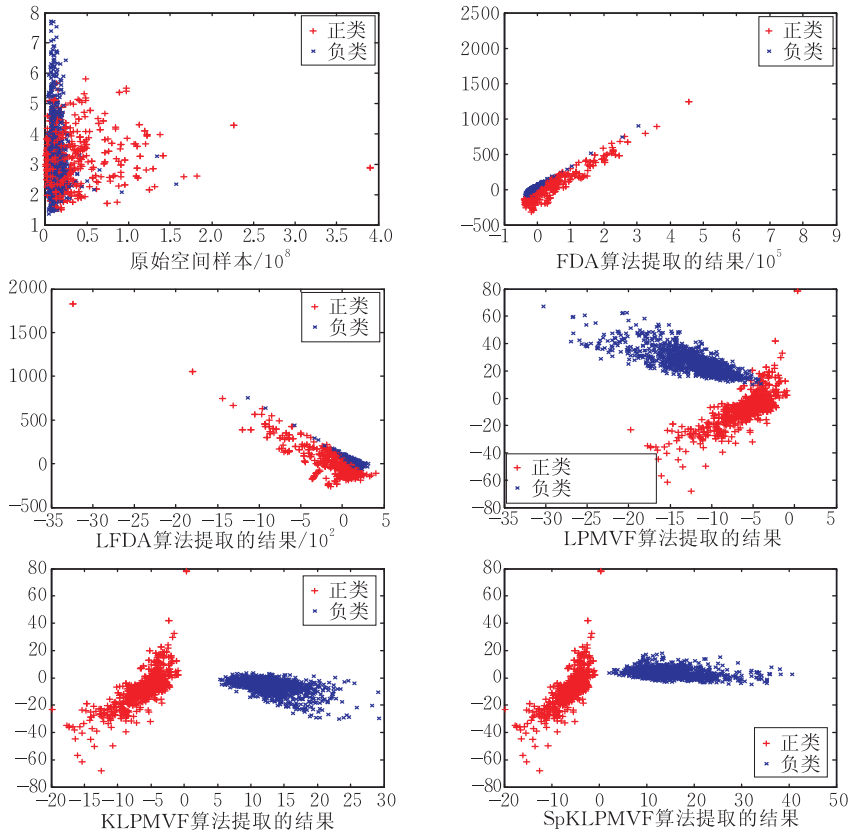


图 5 原始数据和各种算法从该木材数据集上提取出的特征在二维空间中的分布情况

图 6 给出了 FDA、KFD、LFDA、LPMVF、KLPMVF 和 SpKLPMVF 算法在不同的投影维数和近邻数量条件下的缺陷识别率. 在实验中, 算法 SpKLPMVF 的重组矩阵取  $[19 \times 2]$ . 可以看出: (1) 随着投影维数的不同, LPMVF、KLPMVF 和 SpKLPMVF 取得了较高的缺陷识别率, 且算法保持稳定特性, 此外, FDA 在该木材数据库上表现较好, 接近本文算法的识别能力, 实验结果如图 6(a) 所示, 图中的  $N$  是原始样本数量,  $D$  是原始样本维

数; (2) 图 6(b) 给出了各种算法在不同数量的近邻下的木材缺陷识别率, 其中的  $d$  为算法选择的特征维数. 在相同测试条件下, 本文算法几乎总是取得了最高的缺陷识别率, 明显优于其他几种算法, 且随着近邻数量的不断增加, 体现出良好的适应性; (3) 算法 LFDA 在该数据集上的表现较差, 性能较为不稳定, 且在大多数情况下, 缺陷识别率达到最低, 性能明显劣于 FDA 和本文算法.

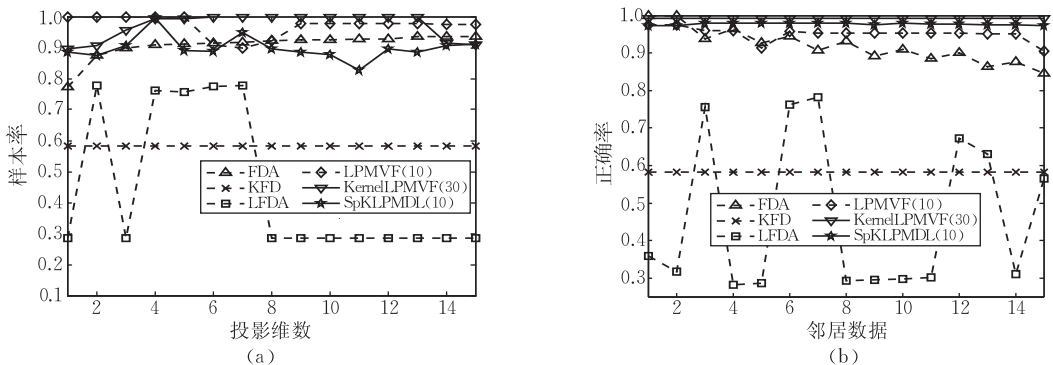


图 6 在不同的投影维数和近邻数量条件下, 各种算法在该木材图像数据库上的识别精度

## 4 结论与展望

在 Fisher 判别分析 FDA 和典型相关分析 CCA 的基础上, 采用以样本的类标号形式给出的先验知识, 考虑数据的局部性, 提出了一种新的多模态识别方法——监督 LPMVF. LPMVF 可通过求解两个标准特征值问题完成, 最终求取得到一对有效的投影变换矩阵或向量. LPMVF 可抽取出描述能力较强的特征信息, 使得原始数据的投影在构造的嵌入空间中便于分类, 并可有效避免奇异性, 同时 LPMVF 可有效保持样本数据间的局部近邻关系. 借助核映射, LPMVF 可快速被推广为非线性的 KLPMVF 算法. 本文中试图进一步增强非线性 KLPMVF 算法的学习能力, 构造原始数据的子模式结构空间, 对原始数据矩阵进行重新组合, 将 KLPMVF 推广为另一种形式的非线性分析方法——子模式 KLPMVF. 用 LPMVF 及其非线性的推广算法进行有效特征的选择和提取后, 能够有效利用类标号监督信息提高算法的分类性能和在模式识别应用中的识别率. 在大多数情况下, 本文算法体现出的整体性能接近于甚至优于传统经典的 FDA、KFD 和 LFDA 算法. 目前, 半监督学习和多模态识别算法是机器学习领域中的研究热点, 在未来的研究中, 可将半监督以及多模态的思想方法与本文算法结合起来, 实现多模态半监督特征选择算法. 此外, 考虑如何有效选取非线性 KLPMVF 算法中的核函数和最优核参数也是一个值得研究的方向.

**致 谢** 在此, 我们向对本文工作予以帮助和建议的老师和同学表示感谢, 并向对本文工作中的不足之处提出宝贵评审意见的审稿专家表示衷心的感谢!

### 参 考 文 献

[1] Duda R O, Hart P E, Stork D. Pattern Classification. New York: Wiley-Interscience, 2000

[2] Mika S, Ratsch G, Weston J et al. Fisher discriminant analysis with kernels//Hu Y H, Larsen J, Wilson E, Douglas S. Neural Networks for Signal Processing, IX, 1999: 41-48

[3] Yang J, Frangi A F, Yang J Y, Zhang D, Jin Z. KPCA plus LDA: A complete kernel Fisher discriminant framework for

feature extraction and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(2): 230-244

[4] Xiong T, Ye J P, Li Q, Cherkassky V, Janardan R. Efficient kernel discriminant analysis via QR decomposition//Advances in Neural Information Processing Systems 17, 2005

[5] Yang M H. Kernel eigenfaces vs. kernel Fisherfaces: Face recognition using kernel methods//Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition. Washington, DC, IEEE Computer Society, 2002: 215-220

[6] Lu J W, Plataniotis K N, Venetsanopoulos A N. Face recognition using kernel direct discriminant analysis algorithms. IEEE Transactions on Neural Networks, 2003, 14(1): 117-126

[7] Min W, Lu K, He X. Locality preserving projection. Pattern Recognition, 2004, 37(4): 781-788

[8] Masashi Sugiyama. Local Fisher discriminant analysis for supervised dimensionality reduction//Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh, Pennsylvania, 2006: 905-912

[9] Kim Tae-Kyun, Kittler J. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 318-327

[10] Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. Journal of Machine Learning Research, 2007, 8(5): 1027-1061

[11] Borga M, Knutsson H. Canonical correlation analysis in early vision Processing//Proceedings of the 9th European Symposium on Artificial Neural Networks. Bruges, Belgium, 2001: 309-314

[12] Zheng W M, Zhou X Y, Zou C R, Zhao L. Facial expression recognition using kernel canonical correlation analysis (KCCA). IEEE Transactions on Neural Networks, 2006, 17(1): 233-238

[13] Loog M, van Ginneken B, Duin R P W. Dimensionality reduction by canonical contextual correlation projections//Proceedings of the European Conference on Computer Vision. Prague Czech Republic: Springer, 2004: 562-573

[14] Sun Quan-Sen, Zeng Sheng-Gen, Liu Yan, Heng Pheng-Ann, Xia De-Shen. A new method of feature fusion and its application in image recognition. Pattern Recognition, 2005, 38(12): 2437-2448

[15] Hsieha Ping-Cheng, Tung Pi-Cheng. A novel hybrid approach based on sub-pattern technique and whitened PCA for face recognition. Pattern Recognition, 2009, 42(5): 978-984

[16] Tan Keren, Chen Songcan. Adaptively weighted sub-pattern PCA for face recognition. Neurocomputing, 2005, 64(3): 505-511

- [17] Zhang D, Zhou Z-H, Chen S. Non-negative matrix factorization on kernels//Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence (PRICAI'06). Guilin, China, LNAI, 2006, 4099: 404-412
- [18] Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 2004, 16(12): 2639-2664
- [19] Tsang Ivor W, Kwok James T. Distance metric learning with kernels//Proceedings of the International Conference on Artificial Neural Networks. Istanbul, Turkey: Springer, 2003: 126-129
- [20] Zhang Dao-Qiang, Zhou Zhi-Hua, Chen Song-Can. Semi-supervised dimensionality reduction//Proceedings of the 7th SIAM International Conference on Data Mining (SDM'07). Minneapolis, MN, 2007: 629-634
- [21] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering//Proceedings of the Conference Advances in Neural Information Processing System 15. Cambridge, MA: MIT Press, 2001: 585-591
- [22] Schölkopf B, Smola A J. *Learning with Kernels*. Cambridge, MA: MIT Press, 2002
- [23] Zhang D Q, Chen S C. Clustering incomplete data using kernel-based fuzzy c-means algorithm. *Neural Processing Letters*, 2003, 18(3): 155-162
- [24] Blake C, Keogh E, Merz C J. UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, 1998
- [25] Niskanen Matti, Silven Olli, Kauppinen Hannu. Experiments with SOM based inspection of wood//Proceedings of the International Conference on Quality Control by Artificial Vision. Le Creusot, Burgundy, 2001: 311-316
- [26] Silven Olli, Niskanen Matti, Kauppinen Hannu. Wood inspection with non-supervised clustering. *Machine Vision and Applications*, 2003, 13(5): 275-285
- [27] Kingsbury N G. Complex wavelets for shift invariant analysis and filtering of signal. *Journal of Applied and Computational Harmonic Analysis*, 2001, 10(3): 234-253



**ZHANG Zhao**, born in 1984, master candidate. His research interests include machine learning, intelligent information processing and pattern recognition.

**YE Ning**, born in 1967, Ph. D., associate professor. His current research interests include pattern recognition, machine learning, data mining and bioinformatics.

**YE Qiao-Lin**, born in 1982, Ph. D. candidate. His current research interests include data mining, machine learning and pattern recognition.

## Background

This work was supported by the National Science Foundation of China under grant No. 30671639, the Natural Science Foundation of Jiangsu Province of China under grant No. BK2009393, the Innovation Foundation 2008 of Jiangsu Province of China under grant No. 164070265, the Innovation Foundation 2009 of Nanjing Forestry University under grant No. 2009106 and the Scientific Research Foundation 2009 of Jiangsu Province of China under grant No. CX09S\_013Z.

More and more research and application areas need to deal with the larger high-dimensional data, which leads to a hot of studying the methods of feature selection and feature selection has been an important preprocessing step, which leads us to consider methods of feature selection that allow one to represent the data points in a low-dimensional embedding space. In order to embed the samples data well, it is important to preserve the spatial local structure of the original data points. Many locality based approaches have been continuously proposed for feature selection in recent several years, such as LPP, LFDA, etc. The successful applications

of the locality preservation based methods for feature selection and pattern recognition inspire us to pay more attention to the locality based techniques.

In this paper, the authors consider the supervised feature selection problem where samples are accompanied with class labels and propose an effective feature selection approach called locality preserving multi-projection vector Fisher discriminant analysis (LPMVF). LPMVF takes the local structure of the original samples data into account, thus the data points can be embedded appropriately. By defining the new guidelines, the proposed methods have several obvious advantages over some established classical feature selection methods. Experimental results show that the proposed methods can extract the good features and effectively improve the accuracies by introducing the class labels as priori knowledge. Furthermore, viewing from the good property for data visualization, it is very suitable for pattern recognition researches and applications.