基于稀疏表示和非参数判别分析的降维算法*

杜 春,孙即祥,周石琳,王亮亮,赵晶晶
(国防科技大学 电子科学与工程学院,湖南 长沙 410073)

摘 要:针对人脸识别问题提出一种新的监督降维算法。算法首先基于稀疏表示理论,利用同类样本间的稀疏重构来构建图。此方案不仅可以克服传统图构造方法中参数选择的困难,而且能够更好地刻画类内信息。然后,算法采用非参数类间离差来刻画类间信息,非参数类间离差在处理复杂分布数据时相比于参数类间离差更具判别力。最后,算法通过保持类内稀疏重构关系的同时最大化非参数类间离差来求得最优的投影矩阵。在ORL和Extended Yale B公共人脸数据库的实验表明,该算法能够获得较好的识别结果。

关键词:降维;稀疏表示;非参数判别分析
中图分类号:TP391.41 **文献标志码**:A **文章编号**:1001-2486(2013)02-0143-05

Dimensionality reduction based on sparse representation and nonparametric discriminant analysis

DU Chun, SUN Jixiang, ZHOU Shilin, WANG Liangliang, ZHAO Jingjing
(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Aiming at the face recognition problem, a new supervised dimensionality reduction algorithm is presented. On the basis of sparse representation theory, the proposed algorithm uses the within-class sparse construction to construct graph. This scheme can avoid the difficulty of parameter selection in traditional graph construction methods, and characterize the within-class information well. Furthermore, the multi-class nonparametric discriminant scatter is applied to characterize the between-class information, which will be more discriminative than parametric discriminant scatter in dealing with complex-distributed data. By maximizing the nonparametric between-class scatter and preserving the within-class sparse reconstructive relationship, the proposed algorithm can seek for the optimal projection matrix. Experimental results on ORL and Extended Yale B dataset show that the proposed method can achieve good recognition effect.

Key words: dimensionality reduction; sparse representation; nonparametric discriminant analysis

降维技术在模式识别、机器学习、数据挖掘等领域具有广泛的应用。近年来,随着微分几何等数学理论的发展,以流形学习为代表的非线性降维技术成为研究的热点。本质上,流形学习致力于将数据从高维观测空间映射到一个低维空间,发掘隐藏在高维观测数据中的低维线性或非线性流形结构。典型的流形学习算法包括局部线性嵌入(Locally Linear Embedding, LLE)[1]、等度规映射(Isometric Mapping, ISOMAP)[2]、邻域保持嵌入(Neighborhood Preserving Embedding, NPE)[3]、局部保持映射(Locality Preserving Projections, LPP)[4],等。

经典流形学习算法可以统一到图嵌入框架[5]当中,一般先通过近邻关系来构建邻接矩阵或拉普拉斯矩阵,然后应用谱图理论来求解低维嵌入。这些算法简洁明确,能够发现隐藏在高维数据中的内在结构。然而,如何选择最优的模型参数(例如邻域尺寸,热核权值参数等)仍是一个难点问题。最近,Qiao等基于稀疏分解理论提出了稀疏保持投影(Sparsity Preserving Projections, SPP)[6]算法。与保持局部邻域几何关系的LPP和NPE算法不同,SPP试图保持高维和低维空间的稀疏重构关系。SPP通过求解L_1范数优化问题来构建图,无需选择任何模型参数,因而具有较好的适应性。然而,作为一种非监督的降维算法,SPP算法没有利用数据的类别信息,这在一定程度上限制了其在模式识别中的应用。

为了进一步提高算法性能,本文在SPP算法基础上提出了一种基于稀疏表示和非参数判别分析(Nonparametric Discriminant Analysis, NDA)[7-8]的监督降维算法。该算法首先利用同类样本的稀疏表示来构造图,用于描述样本间的稀疏重构关

* **收稿日期**:2012-04-22
基金项目:国家自然科学基金资助项目(40901216)
作者简介:杜春(1983—),男,云南玉溪人,博士研究生,E-mail:yxduchun@gmail.com;
孙即祥(通信作者),男,教授,硕士,博士生导师,E-mail:sjx75461@hotmail.com

系,然后在保持类内稀疏重构关系的过程中利用非参数离差矩阵来提取类间鉴别信息,最后通过求解广义特征值问题来求得投影矩阵。在ORL和Extended Yale B人脸数据库上的实验表明了本文算法的有效性。

1 基于稀疏表示的图构造

给定一个采样于 d 维流形 M 的 m 维样本集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}\in\mathbb{R}^{m\times N}$, $d<m$, N 为样本个数,则可以构造一个无向图 $G=(V,E)$,其中图的顶点集 V 对应样本,边集 E 对应样本间的权值。在传统的降维算法中,图构造一般采用 k-NN或者 ε-ball来搜索邻域(确定存在边的顶点),然后利用热核函数 $W_{ij}=\exp(-\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2/\sigma^2)$ 来计算图上顶点 $\boldsymbol{x}_i$ 与 $\boldsymbol{x}_j$ 的边的权值。这种方法直观简单,但是需要引入近邻个数 k,距离半径 ε 和热核参数 σ 等模型参数。通常,降维算法对这些参数非常敏感,参数的自动选取非常困难。

基于稀疏表示的图构造方法采取另外一个思路,其首先把整个训练样本集看作一个字典,求得每一个样本在字典上的稀疏表示。然后,把具有非零重构系数的样本组成邻域,把重构系数作为权值。具体而言,基于稀疏表示的图构造可以通过求解式(1)的 L_1 范数优化问题来实现[6]

$$\begin{aligned}&\min\|\boldsymbol{w}_i\|_1\\ &\text{s.t.}\quad \boldsymbol{x}_i=\boldsymbol{X}\boldsymbol{w}_i\\ &\qquad\sum_j w_i^j=1\end{aligned}\tag{1}$$

式中,$\|\cdot\|_1$ 为 L_1 范数算子,$\boldsymbol{w}_i=(w_i^1,w_i^2,\cdots,w_i^{i-1},0,w_i^{i+1},\cdots,w_i^N)^{\mathrm{T}}$ 是关于样本 $\boldsymbol{x}_i$ 的稀疏重构系数向量。图的权值矩阵设置为 $W=(w_1,w_2,\cdots,w_N)$。

2 SPP算法

基于稀疏表示构建的图在一定程度上能够反映数据的内在结构[6]。为了在降维后最大限度地保留这种结构,SPP算法通过保持降维前后样本间的稀疏重构关系来进行降维,其目标函数定义如下

$$\begin{aligned}&\min_{A}\sum_{i=1}^{N}\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_i-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}_i\|^2\\ &\text{s.t.}\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\end{aligned}\tag{2}$$

式中,$\boldsymbol{A}$ 为投影矩阵,$\boldsymbol{w}_i$ 为对应于 $\boldsymbol{x}_i$ 的稀疏重构系数向量,$\boldsymbol{I}$ 为单位矩阵,加约束项的目的在于避免退化解。通过简单的数学推导,上述问题等价于求解

$$\begin{aligned}&\min\ \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{W}-\boldsymbol{W}^{\mathrm{T}}+\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})\\ &\text{s.t.}\quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\end{aligned}\tag{3}$$

式中,tr(·)为矩阵求迹的算子。为更简洁,记 $\boldsymbol{S}_\beta=\boldsymbol{W}+\boldsymbol{W}^{\mathrm{T}}-\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}$,可将式(3)转化为求解最大化问题

$$\begin{aligned}&\max\ \mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}_\beta\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})\\ &\text{s.t.}\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\end{aligned}\tag{4}$$

3 稀疏NDA

SPP算法是无监督方法,无法利用样本的先验类别信息。为了进一步提高降维性能并应用于模式识别,本文从两个方面对SPP算法进行扩展:在类内信息的刻画方面,本文算法仅用同类样本构造字典来进行稀疏重构,可以有效地利用先验信息;在类间信息的刻画方面,本文算法采用NDA的思路,可以更好地提取对分类有用的信息。在3.3节,给出了完整的算法推导与求解。

3.1 类内稀疏重构保持

在基于稀疏表示构建的图中,权值 w_i^j 能够反映样本 $\boldsymbol{x}_j$ 对 $\boldsymbol{x}_i(i\neq j)$ 进行稀疏重构的贡献,w_i^j 越大,则 $\boldsymbol{x}_j$ 与 $\boldsymbol{x}_i$ 相似程度就越高,它们来自同一类的可能性就越大。现有研究表明,能对 $\boldsymbol{x}_i$ 进行最稀疏表示的样本通常与 $\boldsymbol{x}_i$ 来自同一类别[6,9]。换句话说,在稀疏重构系数向量 $\boldsymbol{w}_i$ 中,非零系数通常对应于与 $\boldsymbol{x}_i$ 同类的样本,而与 $\boldsymbol{x}_i$ 异类的样本对 $\boldsymbol{x}_i$ 的稀疏重构贡献较小。受此启发,本文在对样本进行稀疏表示时仅考虑将同类样本作为字典,按式(5)求解稀疏表示

$$\begin{aligned}&\min\|\boldsymbol{w}_{k,i}\|_1\\ &\text{s.t.}\ \boldsymbol{x}_{k,i}=\boldsymbol{X}_{k,i}\boldsymbol{w}_{k,i}\\ &\qquad\sum_j w_{k,i}^j=1\end{aligned}\tag{5}$$

式中,$\boldsymbol{x}_{k,i}$ 表示第 k 类第 i 个样本,$i=1,2,\cdots,N_k$,$k=1,2,\cdots,c$,N_k 为第 k 类样本的个数,c 为样本类数;字典 $\boldsymbol{X}_{k,i}$ 中包含第 k 类的所有样本;所求得的稀疏重构系数向量 $\boldsymbol{w}_{k,i}=(w_{k,i}^1,w_{k,i}^2,\cdots,w_{k,i}^{i-1},0,w_{k,i}^{i+1},\cdots,w_{k,i}^{N_k})^{\mathrm{T}}$。令 $\boldsymbol{w}'_{k,i}=(\boldsymbol{0}_{(N_1+N_2+\cdots+N_{k-1})\times1},\boldsymbol{w}_{k,i}^{\mathrm{T}},\boldsymbol{0}_{(N_{k+1}+N_{k+2}+\cdots+N_c)\times1})^{\mathrm{T}}$,则得到图的权值矩阵为 $\boldsymbol{W}_M=(\boldsymbol{w}'_{1,1},\cdots,\boldsymbol{w}'_{1,N_1},\boldsymbol{w}'_{2,1},\cdots,\boldsymbol{w}'_{2,N_2},\cdots,\boldsymbol{w}'_{c,1},\cdots,\boldsymbol{w}'_{c,N_c})$。

与SPP算法类似,我们希望在降维前后保持同类样本间的稀疏重构关系,由此构造如下的目标函数

$$\begin{aligned}&\min_{A}\sum_{k=1}^{c}\sum_{i=1}^{N_k}\|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}_{k,i}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{w}'_{k,i}\|^2\\ &\text{s.t.}\quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{I}\end{aligned}\tag{6}$$

按式(2)、(3)进行推导，式(6)转化为求解最大化问题

$$\max \operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}) \quad \text{s.t.} \quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I} \tag{7}$$

式中，$\boldsymbol{S}_M = \boldsymbol{W}_M + \boldsymbol{W}_M^{\mathrm{T}} - \boldsymbol{W}_M^{\mathrm{T}}\boldsymbol{W}_M$。

3.2 类间鉴别信息提取

SPP 算法是无监督方法，无法利用先验类别信息。在本文中，采用 NDA 来提取类间鉴别信息。与 LDA 不同，NDA 算法没有对数据高斯分布的假设，在计算离差矩阵时也无需估计二阶统计量(比如均值)，因而能够更好地描述高维复杂数据的结构。

从原理上讲，NDA 就是求每个样本和其类间最近邻的差值组成的类间离差矩阵 $\boldsymbol{S}_E$，然后用得到的 $\boldsymbol{S}_E$ 来提取类间鉴别信息。对于第 k 类第 i 个样本 $\boldsymbol{x}_{k,i}, i = 1,2,\cdots,N_k, k = 1,2,\cdots,c$，文献[8]定义 $\boldsymbol{x}_{k,i}$ 的类内和类间最近邻分别为

$$\boldsymbol{x}_{k,i}^{I} = \{\boldsymbol{x}' \in C_k / \|\boldsymbol{x}' - \boldsymbol{x}_{k,i}\| \leqslant \|\boldsymbol{z} - \boldsymbol{x}_{k,i}\|, \forall \boldsymbol{z} \in C_k\} \tag{8}$$

$$\boldsymbol{x}_{k,i}^{E} = \{\boldsymbol{x}' \in \overline{C}_k / \|\boldsymbol{x}' - \boldsymbol{x}_{k,i}\| \leqslant \|\boldsymbol{z} - \boldsymbol{x}_{k,i}\|, \forall \boldsymbol{z} \in \overline{C}_k\} \tag{9}$$

令 $\boldsymbol{\Delta}_{k,i}^{E} = \boldsymbol{x}_{k,i} - \boldsymbol{x}_{k,i}^{E}, \boldsymbol{\Delta}_{k,i}^{I} = \boldsymbol{x}_{k,i} - \boldsymbol{x}_{k,i}^{I}$，则非参数类间离差阵为

$$\boldsymbol{S}_E = \frac{1}{N}\sum_{k=1}^{c}\sum_{i=1}^{N_k} w_{k,i}(\boldsymbol{\Delta}_{k,i}^{E})(\boldsymbol{\Delta}_{k,i}^{E})^{\mathrm{T}} \tag{10}$$

式中，$\boldsymbol{x}_{k,i}$ 对应的样本权重定义为 $w_{k,i} = \min\{\|\boldsymbol{\Delta}_{k,i}^{E}\|, \|\boldsymbol{\Delta}_{k,i}^{I}\|\} / (\|\boldsymbol{\Delta}_{k,i}^{E}\| + \|\Delta_{k,i}^{I}\|)$。

通过观察式(8)～(10)可以看出，上述的 NDA 方法通过“非此即彼”的思想来求得类间最近邻差值，其更适合于对两类问题的分析。为了更好地描述多类问题中的类间差距，本文取样本 $\boldsymbol{x}_{k,i}$ 与每一类的类间最近邻差值来构造非参数离差矩阵，即令

$$\boldsymbol{S}_E = \frac{1}{N(c-1)}\sum_{k=1}^{c}\sum_{j\neq k}^{c}\sum_{i=1}^{N_k} w_{k,i,j}(\boldsymbol{\Delta}_{k,i,j}^{E})(\boldsymbol{\Delta}_{k,i,j}^{E})^{\mathrm{T}} \tag{11}$$

式中，$\boldsymbol{x}_{k,i,j}^{E} = \{\boldsymbol{x}' \in C_j / \|\boldsymbol{x}' - \boldsymbol{x}_{k,i}\| \leqslant \|\boldsymbol{z} - \boldsymbol{x}_{k,i}\|, \forall \boldsymbol{z} \in C_j, j \neq k\}$，$\boldsymbol{\Delta}_{k,i,j}^{E} = \boldsymbol{x}_{k,i} - \boldsymbol{x}_{k,i,j}^{E}$，其对应的权重定义为 $\boldsymbol{w}_{k,i,j} = \min\{\|\boldsymbol{\Delta}_{k,i,j}^{E}\|, \|\boldsymbol{\Delta}_{k,i}^{I}\|\} / (\|\boldsymbol{\Delta}_{k,i,j}^{E}\| + \|\boldsymbol{\Delta}_{k,i}^{I}\|)$。

3.3 稀疏非参数判别分析

综合3.1节和3.2节的分析，本文提出在保持类内稀疏重构关系的同时最大化非参数类间离差信息为目标，构造如下的最优化问题

$$\arg\max_{A} \operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A}) + \mu \cdot \operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{S}_E\boldsymbol{A}) \quad \text{s.t.} \quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{I} \tag{12}$$

式中，规则化系数 μ 用于平衡稀疏保持项和非参数类间离差项。

采用拉格朗日乘子法对式(12)进行求解，即令

$$\frac{\partial}{\partial \boldsymbol{A}}\operatorname{tr}(\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}} + \mu\boldsymbol{S}_E)\boldsymbol{A} - \lambda(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A} - \boldsymbol{I})) = \boldsymbol{0} \tag{13}$$

求得

$$(\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}} + \mu\boldsymbol{S}_E)\boldsymbol{A} = \lambda\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A} \tag{14}$$

容易看出，式(14)等价于一个关于 $\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}} + \mu\boldsymbol{S}_E$ 和 $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ 的广义特征值分解问题，所求投影矩阵 $\boldsymbol{A}$ 就对应 d 个最大特征值对应的特征向量，其中 d 为降维的维数。

3.4 时间复杂度分析

本文算法的时间复杂度主要包含稀疏表示求解，非参数类间离差矩阵计算和广义特征值分解三个方面，算法运行时间与训练样本个数 N，维数 m 等有关。采用标准线性规划来求解式(1)的复杂度为 $4km^2/3 + kmN + O(kN)$，其中 k 为标准线性规划迭代的次数[11]。据此，SPP 算法中稀疏表示求解的计算时间为 $N(4km^2/3 + kmN + O(kN))$，本文根据式(5)求解稀疏表示的计算时间为 $\sum_{i=1}^{c} N_i(4km^2/3 + kmN_i + O(kN_i))$。当 $N \gg N_i$ 时，$i = 1,2,\cdots,c$，本文算法可以较多地节省计算时间。此外，算法在计算非参数类间离差矩阵时的计算复杂度为 $O(mN^2)$，在求解广义特征值分解时的计算复杂度为 $O(m^3)$。

4 实验结果与分析

为了验证算法的有效性，在 ORL 和 Extended Yale B 两个公共人脸数据库上进行实验，并与 LPP、NPE、LDA、SPP 等降维算法进行比较。ORL 数据库包含 40 个人，每个人 10 幅图像，包含不同的光照、姿态、表情和细节变化。Extended Yale B 数据库包含 38 个人，每个人大约 64 幅图像，在变化的光照(实验室控制)下采集。为了保证所有算法在公平的环境下进行，本文把所有样本的尺寸归一化到 32×32(1024 维)，把样本的灰度值归一化到[0,1]，并采用最近邻分类器进行分类。为避免 LDA 算法中类内离差矩阵奇异导致无法计算，本文提前对所有样本用 PCA 进行预处理，保留 100% 的主成分。对于 LPP 和 NPE 算法，设置近邻

个数 $k = p-1$，p 为训练样本中每一类选取的样本个数，热核参数 σ 为训练样本范数的均值。

4.1 参数 μ 的设置

由式(14)可以看出，规则化系数 μ 用于平衡 $\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}}$ 和 $\boldsymbol{S}_E$，其取决于具体的训练样本数据。类似于文献[10]的做法，本文把 μ 看做是平衡 $\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}}$ 和 $\boldsymbol{S}_E$ 能量变化的因子。假设 $\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}}$ 和 $\boldsymbol{S}_E$ 的能量变化之比为 β，则可以用 β 与一个非负实数 D 的乘积来近似 μ。为简便，这里用 $\boldsymbol{X}\boldsymbol{S}_M\boldsymbol{X}^{\mathrm{T}}$ 和 $\boldsymbol{S}_E$ 的最大特征值来近似其能量变化，此时就可以得到参数 $\mu \simeq D\cdot\beta = D\cdot\lambda_{\max}(\boldsymbol{X}\boldsymbol{S}_\beta\boldsymbol{X}^{\mathrm{T}})/\lambda_{\max}(\boldsymbol{S}_E)$，其中，$\lambda_{\max}(\cdot)$ 表示求矩阵最大特征值算子。关于 D 的确定，实验中首先根据经验给定它的一个取值范围 $D \in \{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}\}$，然后在训练样本集上进行五折交叉验证，测试每一个候选参数，最终选取平均识别率最高对应的值作为 D。经过交叉验证，本文 ORL 实验中的 D 取值为 10^{-3}，Extended Yale B 实验中的 D 取值为 10^{-4}。

表 1 ORL 数据的最高平均识别率和对应的维数

Tab. 1 Maximal average recognition rates and the corresponding dimensionality on ORL

ORL	LPP	NPE	LDA	SPP	Proposed
2Train	0.575(62)	0.680(79)	0.715(38)	0.753(79)	0.821(39)
3Train	0.650(82)	0.756(119)	0.833(36)	0.804(119)	0.862(40)
4Train	0.701(89)	0.831(159)	0.889(38)	0.843(159)	0.918(41)
5Train	0.769(105)	0.867(199)	0.929(39)	0.868(186)	0.945(86)

表 2 Extended Yale B 数据的最高平均识别率和对应的维数

Tab. 2 Maximal average recognition rates and the corresponding dimensionality on Extended Yale B

Extended Yale B	LPP	NPE	LDA	SPP	Proposed
5Train	0.582(189)	0.659(189)	0.730(37)	0.769(188)	0.815(166)
10Train	0.698(379)	0.739(379)	0.848(37)	0.884(379)	0.902(217)
15Train	0.712(569)	0.847(569)	0.898(37)	0.906(561)	0.946(291)

表 3 非均匀随机抽样下的实验结果

Tab. 3 Experimental result of non-uniform random sampling

数据集	LPP/s	NPE	LDA	SPP	Proposed
ORL	0.675(96)	0.818(199)	0.873(35)	0.855(194)	0.925(117)
Extended Yale B	0.582(199)	0.691(199)	0.720(35)	0.762(167)	0.795(148)

4.2 识别率比较

在 ORL 实验中，每次从各类人脸随机抽取 i($i=2,3,4,5$)个样本为训练集，剩下的作为测试集，重复进行 20 次测试，取 20 次实验的平均值为最后识别率。四组实验分别表示为 2Train，3Train，4Train，5Train。表 1 给出了 5 种算法在 ORL 人脸数据库的最高平均识别率和对应的维数。

在 Extended Yale B 实验中，每次从各类人脸随机抽取 i($i=5,10,15$)个样本为训练集，剩下的为测试集，重复进行 20 次测试，取 20 次实验的平均值为最后识别率。三组实验分别表示为 5Train，10Train，15Train。表 2 给出了 5 种算法的最高平均识别率和对应的维数。

在上述实验中，训练样本集中各类样本个数均相同。为了验证算法在非均匀随机抽样下的性能，本文从 ORL 和 Extended Yale B 中分别随机抽取 200 个样本作为训练集(其中，每类样本个数不再要求相同，但考虑到求解同类稀疏表示时字典元素不能少于 2 个，这里要求随机抽取后每类的样本个数需大于 1)，剩下的作为测试集，重复进行 20 次实验，取 20 次实验的平均值为最后的识别率。表 3 分别给出非均匀随机抽样条件下的 ORL 和 Extended Yale B 数据集的最高平均识别率及其对应的维数。

观察表 1，表 2 和表 3，可以发现：(1)在同等条件下，SPP 和 LDA 算法的识别率高于 LPP 和 NPE 两种非监督降维算法。其原因主要是：SPP 算法引入了稀疏表示，能够较好地反映数据的内在结构；LDA 方法属于监督降维算法，能够有效利用数据的先验类别信息。(2)对于 ORL 和 Extended Yale B 数据，在均匀随机抽样和非均匀

随机抽样条件下,本文算法均优于其他四种降维算法。以 ORL 数据集中非均匀随机抽样下的结果为例,本文算法最高平均识别率达到 92.5%,分别比 LPP、NPE、LDA 和 SPP 提高了 25%、10.7%、5.2%和 7.0%。由于在降维过程中同时兼顾类内局部结构保持与类间鉴别信息的提取,本文算法能够获得较高的识别正确率。

4.3 计算代价比较

以 ORL 数据集上的实验为例,本文比较了五种算法的计算时间,如表 4 所示。其中,N 表示训练样本个数,m 表示经 PCA 预处理后数据的维数。由表 4 可以看出,由于算法中求解稀疏表示的步骤比较耗时,SPP 和本文算法的计算代价要大于 LPP,NPE 和 LDA 算法。然而,相比于 SPP,本文算法仅用同类样本构造字典来进行稀疏表示,能够在保证识别率的前提下大大降低计算复杂度。

表 4　运行时间比较

Tab. 4　Comparison of computation time

数据集	N	m	LPP/s	NPE/s	LDA/s	SPP/s	Proposed/s
2Train	80	79	0.20	0.14	0.13	3.21	0.51
3Train	120	119	0.24	0.17	0.15	8.44	1.11
4Train	160	159	0.27	0.26	0.18	17.07	1.77
5Train	200	199	0.33	0.28	0.23	33.95	2.73

5　结论

本文针对人脸识别问题,基于稀疏表示和非参数判别分析提出一种新的监督降维算法。该算法的特点在于:首先,利用同类样本的稀疏重构保持来刻画类内信息,有效避免了 LPP 和 NPE 算法的模型参数选择问题;其次,利用非参数离差矩阵来提取类间鉴别信息,可以更精细地刻画多类问题中的类间差别,获得更高的识别率。由于需要进行稀疏表示求解并采用交叉验证设置参数,本文算法计算复杂度偏大,如何加快算法运算速度是下一步需要研究的问题。

参考文献(References)

[1] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290: 2323 – 2326.

[2] Tenenbaum J B, Silva V d, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290: 2319 – 2323.

[3] He X F, Cai D, Yan S C, et al. Neighborhood preserving embedding[C]//Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing: IEEE, 2005:1208 – 1213.

[4] He X F, Yan S C, et al. Face recognition using Laplacianfaces [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27:328 – 340.

[5] Yan S C, Xu D, et al. Graph embedding and extensions: a general framework for dimensionality reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 31(2): 40 – 51.

[6] Qiao L S, Chen S C, Tan X Y. Sparsity preserving projections with applications to face recognition[J]. Pattern Recognition, 2010, 43: 331 – 341.

[7] Fukunaga K, Mantock J. Nonparametric discriminant analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1983, 5: 671 – 678.

[8] Bressan M, Vitria J. Nonparametric discriminant analysis and nearest neighbor classification [J]. Pattern Recognition Letters, 2003, 24(15): 2743 – 2749.

[9] Wright J, Yang A Y, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31: 210 – 227.

[10] Liu Q, Tang X, et al. Face recognition using kernel scatter-difference-based discriminant analysis[J]. IEEE Transactions on Neural Networks, 2006, 17(4): 1081 – 1085.

[11] Donoho D L, Tsaig Y. Fast solution of L1-norm minimization problems when the solution may be sparse [J]. IEEE Transactions on Information Theory, 2008, 54(11): 4789 – 4812.