

基于低秩表示的判别特征提取算法

苏雅茹¹, 许智杰¹, 吴小惠²

(1. 福州大学数学与计算机科学学院, 福建 福州 350108; 2. 厦门大学航空航天学院, 福建 厦门 361005)

摘要: 为了更好地获取高维数据的特征, 提出一种特征提取算法——低秩判别映射. 首先基于低秩表示构造代表样本关联性的关联矩阵, 然后利用关联矩阵应用判别准则. 低秩表示以样本作为基函数, 利用所有样本构建关联矩阵, 其构造特点决定了关联矩阵能够很好地体现样本集的全局结构和样本之间的判别关系. 人脸数据集的实验表明, 低秩判别映射优于其他广泛应用的特征提取方法.

关键词: 低秩表示; 判别准则; 特征提取

中图分类号: TP391.1

文献标识码: A

Low-rank representation based discriminant feature extraction

SU Yaru¹, XU Zhijie¹, WU Xiaohui²

(1. College of Mathematics and Computer Science, Fuzhou University, Fuzhou, Fujian 350108, China;

2. School of Aerospace Engineering, Xiamen University, Xiamen, Fujian 361005, China)

Abstract: A novel feature extraction approach, namely low-rank representation discriminant projections, is proposed to better capture the feature information of high-dimensional data. In low-rank representation, the affinity matrix whose entries measure similarities among data points is constructed by first. Then the discriminate criterion is employed based on the input affinity matrix. Through the affinity matrix derived by low-rank representation, it can well capture the global structure of data and the discriminative property. Experimental results based on face data sets show that LRDP is superior to other widely-used feature extraction methods.

Keywords: low-rank representation; discriminant criteria; feature extraction

0 引言

随着数据类型越来越丰富, 数据结构也越来越复杂. 比如, 数据库文本、Web 数据、各种各样的图像数据和视频数据等, 这给数据的分析、处理和应用带来了困难与挑战. 这些数据往往具有很高的维数和复杂的结构, 数据形式通常是稀疏的, 带有大量的冗余信息, 数据的这些特点不仅增加了分析处理负担, 使得时间和空间复杂度迅速上升从而导致算法性能下降, 甚至容易掩盖数据的真实结构从而导致错误的分析处理结果. 高维复杂数据的诸多应用需求使得对高维数据的分析处理在科学研究领域占据着越来越重要的地位.

充分利用高维复杂数据的结构对学习问题进行建模是利用数据信息的关键所在. 关联矩阵是描述数据结构的一种形式, 用于度量样本和样本之间的相关性. 相关性高的样本通常会被聚集到一块, 关联矩阵的理想表现是假设样本按类别排好, 其对应的关联矩阵是块对角矩阵. 近年来, 利用低秩表示构造的关联矩阵被应用于聚类^[1-6]、维数约简^[7-9]、噪声人脸识别^[10]、闭塞人脸识别^[11-12]等问题, 并取得很好的效果.

本研究提出一种低秩表示判别映射(low-rank representation discriminant projections, LRDP), 基于低秩表示(low-rank representation, LRR)构造关联矩阵并用于判别特征提取, 并利用人脸数据集的实验证明了LRDP的特征提取有效性.

收稿日期: 2018-06-04

通讯作者: 苏雅茹(1983-), 讲师, 主要从事智能算法方面研究, yarusu@fzu.edu.cn

基金项目: 国家自然科学基金资助项目(61673323); 福建省自然科学基金资助项目(2016J01295, 2017J01068)

<http://xbzrb.fzu.edu.cn>

1 低秩判别映射

首先, 给定一组来自若干相邻子空间的样本, 低秩表示把样本向量看成其它所有样本向量的线性组合, 并寻找最低秩表示, 低秩表示的构造特点决定了 LRDP 的良好数据全局结构表达能力和判别结构表达能力^[1]. 然后, 基于表达数据关系结构的低秩表示引入充分利用数据判别信息的判别准则, 这些特点使得 LRDP 很好地利用了数据样本信息.

1.1 低秩表示

低秩表示的数学描述如下^[1]:

给定一个样本集 $X = [X_1, X_2, \dots, X_n] \in \mathbf{R}^{d \times n}$, n 表示样本个数, d 表示样本维数, 每个样本 $X_i \in \mathbf{R}^d$ 都可以表示为“字典” $A = [a_1, a_2, \dots, a_m]$ 中的基的线性组合, 则模型描述为:

$$X = AZ \tag{1}$$

其中: $Z = [Z_1, Z_2, \dots, Z_m]$ 是一个由系数向量组成的系数矩阵, 系数向量 Z_i 对应着样本 X_i .

因为 A 的复杂性, 样本集 X 的系数矩阵 Z 通常有多个可行解, 上述模型可以通过以下的优化问题来求解:

$$\min_Z \text{rank}(Z); \quad \text{s.t. } X = AZ \tag{2}$$

其中 “最低秩表示” Z^* 表示最优解.

由秩 $\text{rank}(Z)$ 的非凸性, 上述模型的优化问题通常很难解决. 于是, 采用矩阵计算方法^[13-14], 优化问题(2)可以替换为以下凸优化问题:

$$\min_Z \|Z\|_*; \quad \text{s.t. } X = AZ \tag{3}$$

其中: $\|Z\|_*$ 表示矩阵 Z 的奇异值之和, 称为矩阵 Z 的核范数^[15].

将问题扩展, 给定样本集 $X = [X_{11}, \dots, X_{1n_1}, \dots, X_{K1}, \dots, X_{Kn_K}] \in \mathbf{R}^{d \times n}$, 其含有 K 个类, 第 k 类含有 n_k 个样本, 总共有 $n = \sum_{k=1}^K n_k$ 个样本, 每个样本是一个 d 维向量. 为了对样本集进行分割, 需要对样本集的子空间结构建模, 即构造代表样本关联性的关联矩阵. 因此, 低秩表示的 A 不采用标准基, 而采用自身一样本集 X , 于是, 凸优化问题(3)转化为:

$$\min_Z \|Z\|_*; \quad \text{s.t. } X = XZ \tag{4}$$

经过理论证明^[1], 当满足两个条件: 样本数据充足, 即 $n_k > \text{rank}(X_k) = d_k$, 且子空间独立时, 优化问题(4)可以得到一个块对角阵解:

$$Z^* = \begin{bmatrix} Z_1^* & 0 & 0 & 0 \\ 0 & Z_2^* & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & Z_k^* \end{bmatrix}_{k \times k} \tag{5}$$

其中: Z_k^* 是一个 $n_k \times n_k$ 矩阵, 它满足 $\text{rank}(Z_k^*) = d_k, \forall i$.

低秩表示用样本自身来表示数据向量, 其原因是解空间为 $I - \text{null}(X)$, 当 $\text{rank}(X) < n$ 时, 总是存在非平凡解. 所以, 即使没有足够的样本数据, 优化问题(4)仍然存在可行解. 如果是稀疏表示, 情况就不一样, 当稀疏表示用样本自身表示数据向量时, 会得到一个平凡解.

1.2 目标函数

给定样本集 $X = [X_1, X_2, \dots, X_n] \in \mathbf{R}^{d \times n}$, 假定均值为零, 包含 K 个类别的样本, 并得到最低秩表示 Z^* :

$$Z^* = \begin{bmatrix} Z_{1,1} & \cdots & Z_{n,1} \\ \vdots & \ddots & \vdots \\ Z_{1,n} & \cdots & Z_{n,n} \end{bmatrix} \in \mathbf{R}^{n \times n} \tag{6}$$

定义样本 X_i 的最低秩表示向量: $Z_i = [Z_{i,1}, Z_{i,2}, \dots, Z_{i,n}]^T$, 并设 k 个类别中的第 l 个样本的最低秩表示向量为 Z_{kl} .

给定线性映射 $Y = V^T X$, 对于单个样本 $X_i \in \mathbf{R}^d$, 可以从 d 维空间映射到 d_r 维空间, 得到 $Y_i \in \mathbf{R}^{d_r}$; 对

于样本集 X , 也可以从 d 维空间映射到 d_r 维空间, 得到 Y , 其中 $d_r \ll d$. 在 d_r ($d_r \ll d$) 维的映射空间中, 以最低秩表示向量 Z_{kl} 为重构系数, 采用第 k 个类别的所有样本重构第 k 个类别中的第 l 个样本 Y_{kl} .

定义映射空间样本集的内类残差:

$$\tilde{R}_W = \frac{1}{n} \sum_{k=1}^K \sum_{l=1}^{n_k} (Y_{kl} - Y\delta_k(Z_{kl})) (Y_{kl} - Y\delta_k(Z_{kl}))^T \tag{7}$$

其中: $\delta_k(Z_{kl})$ 实现只保留最低秩表示向量 Z_{kl} 中第 k 个类别的所有样本系数的功能. 同时, 定义映射空间样本集的内类残差:

$$\tilde{R}_B = \frac{1}{n(K-1)} \sum_{k=1}^K \sum_{l=1}^{n_k} \sum_{k' \neq k} (Y_{kl} - Y\delta_k(Z_{kl})) (Y_{kl} - Y\delta_{k'}(Z_{kl}))^T \tag{8}$$

其中: $\delta_{k'}(Z_{kl})$ 实现只保留最低秩表示向量 Z_{kl} 中第 k' 个类别的所有样本系数的功能.

判别准则的目标是尽量减小类内残差并尽量增大类间残差, 通过最大化以下目标函数来实现:

$$J = \text{tr}(\beta\tilde{R}_B - \tilde{R}_W) \tag{9}$$

其中: β 系数用于平衡类内残差和类间残差.

因为 $Y = V^T X$ 是线性映射, 则有 $\tilde{R}_W = V^T R_W V$ 和 $\tilde{R}_B = V^T R_B V$, 以及:

$$R_W = \frac{1}{n} \sum_{k=1}^K \sum_{l=1}^{n_k} (X_{kl} - X\delta_k(Z_{kl})) (X_{kl} - X\delta_k(Z_{kl}))^T \tag{10}$$

$$R_B = \frac{1}{n(K-1)} \sum_{k=1}^K \sum_{l=1}^{n_k} \sum_{k' \neq k} (X_{kl} - X\delta_k(Z_{kl})) (X_{kl} - X\delta_{k'}(Z_{kl}))^T \tag{11}$$

$$J(V) = \text{tr}(V^T(\beta R_B - R_W)V) \tag{12}$$

这里要求 $V = [V_1, V_2, \dots, V_d]$ 且 $V_i^T V_i = 1$ ($i = 1, 2, \dots, d$). 因此, LRDP 的目标函数为:

$$\max \sum_{i=1}^d V_i^T (\beta R_B - R_W) V_i; \quad \text{s.t. } V_i^T V_i = 1 \quad (i = 1, 2, \dots, d) \tag{13}$$

采用拉格朗日乘子法求解上述目标函数:

$$L(V_i, \lambda_i) = \sum_{i=1}^d V_i^T (\beta R_B - R_W) V_i - \lambda_i (V_i^T V_i - 1) \tag{14}$$

$$\frac{\partial L}{\partial V_i} = ((\beta R_B - R_W) - \lambda_i I) V_i = 0 \tag{15}$$

其中, λ 是拉格朗日乘数. 那么

$$(\beta R_B - R_W) V_i = \lambda_i V_i \tag{16}$$

其中: λ_i ($i = 1, 2, \dots, d$) 是 $(\beta R_B + R_W)$ 和 XX^T 的广义特征值, V_i ($i = 1, 2, \dots, d$) 是相应的广义特征向量. 于是, 目标函数的最优解, 即最优变换矩阵 V 由以上广义特征分解得到的最大的 d 个特征值对应的特征向量组成, 且 $J(V) = \sum_{i=1}^d V_i^T (\beta R_B - R_W) V_i = \sum_{i=1}^d \lambda_i V_i^T V_i = \sum_{i=1}^d \lambda_i$.

1.3 低秩判别映射(LRDP)

1.3 低秩判别映射(LRDP)

低秩判别映射(LRDP)的算法步骤如表 1 所示.

为了对比低秩表示和稀疏表示对数据结构的表达能力, 提出一种基于稀疏表示的判别特征提取算法-稀疏表示判别映射(sparse representation discriminant projections, SRDP). SRDP 和 LRDP 的不同在于, SRDP 的样本关联矩阵是基于稀疏表示构造的, 而 LRDP 基于低秩表示. 稀疏表示可以通过以下的凸优化问题求解^[16-17]:

$$\min_{\alpha_i} \|\alpha_i\|_1; \quad \text{s.t. } X_i = X\alpha_i \tag{17}$$

表 1 低秩判别映射

Tab.1 low-rank discriminant projections

输入: 样本集 $X \in \mathbf{R}^{d \times n}$; 输出: 变换矩阵 V
第一步: 由式(4) 计算样本集 X 的最低秩表示 Z^* .
第二步: 由式(10) ~ (11) 得到目标函数(13).
第三步: 由式(16) 得到变换矩阵 V .

其中: $X = [X_1, X_2, \dots, X_n] \in \mathbf{R}^{d \times n}$ 为一个样本集; n 为样本个数; d 为样本维数; 每个样本 $X_i \in \mathbf{R}^d$ 都可以表示为样本集中所有样本的线性组合; $\|\alpha_i\|_1$ 表示 l_1 范数, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ 就是所求的稀疏表示 S .

2 实验结果与分析

本实验采用 Yale 和 UMIST 人脸数据集, 数据集信息如表 2 所示. 实验数据集被分为训练集和测试集两部分, 假设实验为 P -train, 则每类随机取 P 个样本作为训练样本, 其余作为测试样本. 实验要选择合适的 P 值, 因为 P 值太大信息不充分, 太小容易过拟合.

对每组实验数据, 分别采用主成分分析(PCA)、线性判别分析(LDA)、谱回归判别分析(SRDA)、边沿 Fisher 分析(MFA)、SRDP、LRDP 六种算法进行特征提取, 统一采用 SRC 分类方法的分类准确率来评价各种特征提取算法的性能. 为了准确地评估六种算法的性能, 对每个 P 值进行 20 次随机实验, 并取 20 次随机实验的分类准确率的平均值作为分类准确率.

假设样本类别数为 K , 样本数为 n , 则 LDA 和 SRDA 最多可以取 $K-1$ 个特征, 而其它算法的特征数可以达到 $n-1$. 实验结果给出了分类准确率(分类性能)随特征数变化的曲线图(详见图 1~2), 并在表 3~4 中给出最佳分类准确率(最佳分类性能)、相应的标准差和特征数.

表 2 Yale、UMIST 数据集

Tab.2 Yale、UMIST datasets

公共数据集	样本数	特征数	类别数
Yale	165	1 024	15
UMIST	564	1 024	20

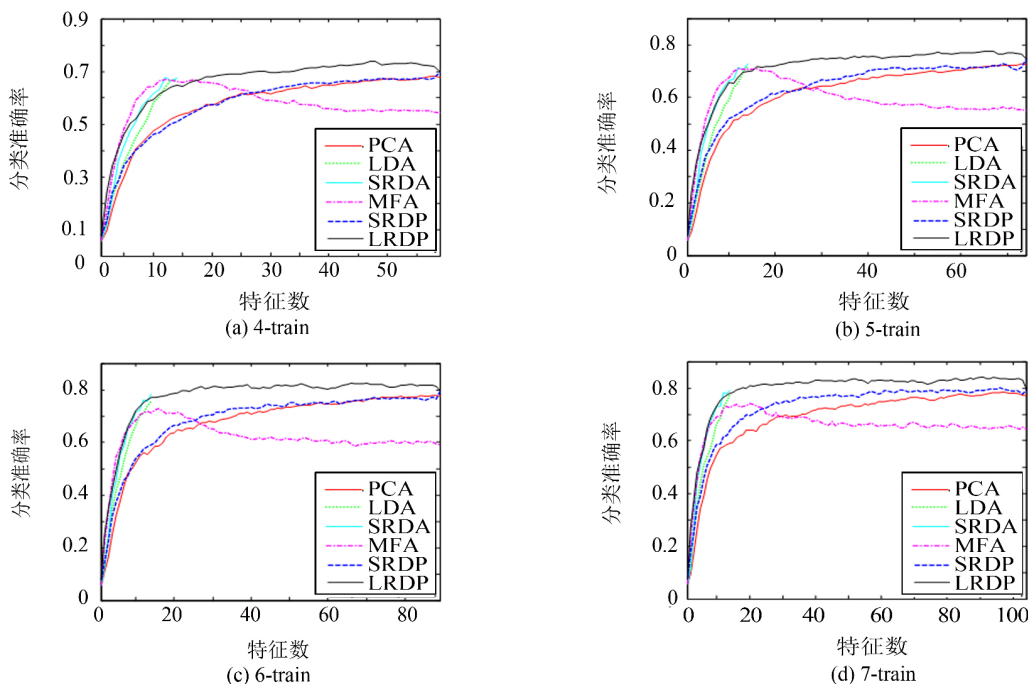


图 1 Yale 数据上分类性能随特征数变化的曲线

Fig.1 Recognition accuracy vs. feature dimension on Yale data set

表 3 Yale 数据上的最佳分类性能及相应的标准差和特征数

Tab.3 Optimal recognition rate as well as corresponding standard deviation and feature dimension on Yale data set

算法	4-train	5-train	6-train	7-train
PCA	69.19±4.44(58)	73.42±3.99(73)	77.80±3.40(87)	79.50±4.96(97)
LDA	67.67±6.11(14)	72.83±3.13(14)	75.93±5.24(14)	78.02±3.80(14)
SRDA	68.76±4.91(12)	73.37±2.97(14)	78.40±5.07(14)	80.01±4.04(14)
MFA	67.86±4.37(12)	71.94±5.56(16)	73.87±4.25(16)	75.17±5.16(20)
SRDP	68.86±4.34(59)	73.49±3.95(74)	78.13±4.06(89)	79.25±3.88(96)
LRDP	73.81±4.31(47)	77.50±3.66(65)	82.47±3.46(66)	84.03±4.81(94)

从图 1~2 可以看出, LRDP 的性能几乎总是优于其它的算法. 对各个 P 值, 各个算法在特征数比较少时, 分类性能随着特征数的增加而快速上升, 当特征数达到一定值后, 性能上升变得缓慢甚至有少许下降, 说明随着特征数的增加, 判别信息的增加变得不明显, 其原因为受到冗余特征的负面影响. 随着 P 值的增加, 即训练样本数的增加, 各个算法的性能都有一定的提高, 说明训练样本的特征是判别信息的主要来源. 从表 3~4 可以看出, LRDP 最佳性能总是优于其它算法(包括 MFA).

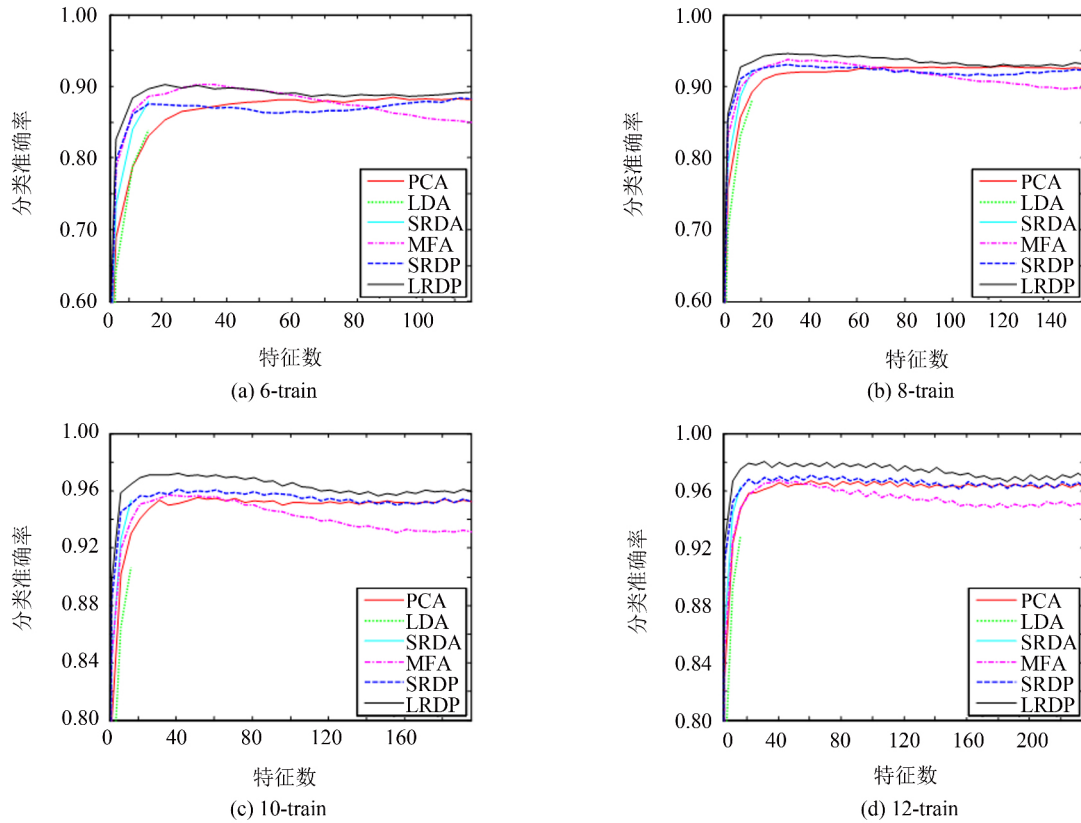


图 2 UMIST 数据上分类性能随特征数变化的曲线

Fig.2 Recognition accuracy vs. feature dimension on UMIST data set

表 4 UMIST 数据上的最佳分类性能及相应的标准差和特征数

Tab.4 Optimal recognition rate as well as corresponding standard deviation and feature dimension on UMIST data set

算法	6-train	8-train	10-train	12-train
PCA	84.44±2.34(91)	92.84±2.15(119)	95.60±1.21(67)	96.81±1.53(103)
LDA	85.40±1.98(19)	89.49±2.65(19)	92.16±1.46(19)	94.09±1.86(19)
SRDA	88.09±2.08(19)	92.84±2.16(19)	96.00±1.24(19)	96.73±1.61(19)
MFA	90.32±2.35(34)	93.83±2.18(33)	95.77±1.49(59)	96.85±1.27(39)
SRDP	88.48±2.23(115)	93.10±2.46(33)	96.12±1.37(49)	97.15±1.36(67)
LRDP	90.34±2.15(23)	94.68±2.34(29)	97.23±1.20(39)	98.05±1.42(31)

3 小结

本研究利用低秩表示和判别准则, 提出低秩表示判别映射 LRDP. 经过理论分析和对比实验, 发现存在以下三个因素: 1) 在子空间独立的假设下, 通过低秩表示获得关联图, 其类内关联系数大, 类间关联系数小, 表现出一定的判别结构表达能力; 2) LRDP 和 SRDP 都不是采用标准基作为基函数, 而是利用样本自身作为基函数, 稀疏表示单独为每个样本寻找稀疏表示, 而低秩表示对所有样本寻找低秩表示, 因此, 低秩表示对数据全局结构的表达能力比稀疏表示更强; 3) 采用缩小类内残差和增大类间残差的判别

准则,更好地利用了样本的类别信息.有鉴于此,本研究认为 LRDP 具有良好的稳定性和鲁棒性.总之,LRDP 是一种有效的特征提取算法,是一种全局性算法,同时还有较强的判别性,能够比较准确地描述数据集的全局结构和数据样本的判别关系.下一步工作将考虑引入一些约束,用非凸替换来求解目标函数,或者从矩阵扩展到张量形式的分析.

参考文献:

- [1] LIU G, LIN Z, YU Y. Robust subspace segmentation by low-rank representation [C]// International Conference on Machine Learning. Haifa: Omnipress, 2010: 663-670.
- [2] WANG J, SHI D, CHENG D, *et al.* LRSR: low-rank-sparse representation for subspace clustering [J]. *Neurocomputing*, 2016, 214: 1026-1037.
- [3] XIE X, GUO X, LIU G, *et al.* Implicit block diagonal low-rank representation [J]. *IEEE Transactions on Image Processing*, 2018, 27(1): 477-489.
- [4] YIN M, GAO J, LIN Z, *et al.* Laplacian regularized low-rank representation and its applications [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(3): 504-517.
- [5] SONG Y, WU Y. Subspace clustering based on latent low rank representation with Frobenius norm minimization [J]. *Neurocomputing*, 2018, 275: 2479-2489.
- [6] 陈晓云, 林莉媛, 叶先宝. 基于光滑近邻表示的基因表达数据子空间聚类 [J]. *控制与决策*, 2017, 32(7): 1235-1240.
- [7] CHONG P, ZHAO K, QIANG C. Integrating feature and graph learning with low-rank representation [J]. *Neurocomputing*, 2017, 249: 106-116.
- [8] SHU Z, FAN H, HUANG P, *et al.* Multiple laplacian graph regularised low-rank representation with application to image representation [J]. *IET Image Processing*, 2017, 11(6): 370-378.
- [9] 薛潇宇, 马小虎. 基于双邻接图正交近邻保持投影的人脸识别算法 [J]. *计算机科学*, 2017, 44(8): 31-35.
- [10] NEJATI M, SAMAVI S, DERKSEN H, *et al.* Denoising by low-rank and sparse representations [J]. *Journal of Visual Communication and Image Representation*, 2016, 36: 28-39.
- [11] GAO G, YANG J, SHEN X, *et al.* Learning robust and discriminative low-rank representations for face recognition with occlusion [J]. *Pattern Recognition*, 2017, 66: 129-143.
- [12] ILIADIS M, WANG H, MOLINA R, *et al.* Robust and low-rank representation for fast face identification with occlusions [J]. *IEEE Transactions on Image Processing*, 2017, 26(5): 2203-2218.
- [13] EMMANUEL C, BENJAMIN R. Exact matrix completion via convex optimization [J]. *Foundations of Computational Mathematics*, 2009, 9(6): 717-772.
- [14] KESHAVAN R H, MONTANARI A, OH S. Matrix completion from noisy entries [J]. *Journal of Machine Learning Research*, 2010, 11(3): 2057-2078.
- [15] MARYAM SARJOU F. Matrix rank minimization with applications [D]. Palo Alto: Stanford University, 2002.
- [16] ELHAMIFAR E, VIDAL R. Sparse subspace clustering [C]//IEEE Conference on Computer Vision and Pattern Recognition. Miami: IEEE, 2009, 35(11): 2790-2797.
- [17] SU Y, LI C, WANG R, *et al.* Sparse representation-based approach for unsupervised feature selection [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2014, 28(3): 433-451.

(责任编辑: 林晓)