

联合平滑矩阵多变量椭圆分布的稀疏表示算法

邱虹¹ 王万良² 郑建炜²

摘要 处理部分光照和遮挡等噪声的图像重构及分类问题因其极具挑战而备受关注, 该问题的解决很大程度上取决于对误差的描述, 常见的方法以向量形式存储误差矩阵且假定其服从于独立同分布, 忽视了图像数据的内部结构信息. 针对该问题, 本文提出一种联合平滑矩阵多变量椭圆分布的稀疏表示算法 (Sparse representation with smoothed matrix multivariate elliptical distribution, SMED). 该算法强调误差矩阵中各个像素间的依赖性并假定误差矩阵作为一个随机矩阵变量服从于矩阵多变量椭圆分布; 之后, 引入辅助变量光滑目标函数, 使得模型易于获得全局最优解; 最后, 采用迭代加权最小二乘法优化求解模型. 此外, 文中对 SMED 算法的收敛性和复杂度进行了理论分析, 并讨论了模型的参数敏感性. 在 AR、ExYaleB 和 PubFig 三个公开数据集上的实验验证了所提算法具有鲁棒的鉴别力, 且其综合性能明显优于经典算法.

关键词 稀疏表示, 矩阵多变量椭圆分布, 迭代加权最小二乘法, 人脸重构和识别

引用格式 邱虹, 王万良, 郑建炜. 联合平滑矩阵多变量椭圆分布的稀疏表示算法. 自动化学报, 2019, 45(8): 1548–1563

DOI 10.16383/j.aas.2018.c170350

Sparse Representation With Smoothed Matrix Multivariate Elliptical Distribution

QIU Hong¹ WANG Wan-Liang² ZHENG Jian-Wei²

Abstract As one of the most challenging problems, partial illumination or occlusion in image reconstruction and classification has recently received considerable attention. This problem heavily depends on the characterization of representation error, while most existing approaches ignore the internal structure information of image data for the error matrix needs to be stretched into a vector with each element being assumed independently corrupted. To improve the algorithms in this regard, a new sparse representation with smoothed matrix multivariate elliptical distribution (SMED) is proposed, which emphasizes the dependency between pixels of noise and assumes that error matrix is a random matrix variate and follows a matrix multivariate elliptical distribution. Then, it introduces regularization terms to smooth the objective function, which makes the model easily to obtain a globally optimal solution. Finally, iteratively reweighted least square is adopted to solve the SMED model efficiently. Furthermore, comprehensive analyses are also performed including convergence behavior, computational complexity, together with parameter determination. Experimental results on the AR, ExYaleB and PubFig databases demonstrate that the proposed algorithm is a robust discriminative classifier with excellent performance, being superior to classical algorithms.

Key words Sparse representation, matrix multivariate elliptical distribution, iteratively reweighted least square, face reconstruction and recognition

Citation Qiu Hong, Wang Wan-Liang, Zheng Jian-Wei. Sparse representation with smoothed matrix multivariate elliptical distribution. *Acta Automatica Sinica*, 2019, 45(8): 1548–1563

收稿日期 2017-06-23 录用日期 2017-10-30
Manuscript received June 23, 2017; accepted October 30, 2017
国家自然科学基金 (61873240, 61602413), 浙江省自然科学基金 (LY19F030016, LY17F020001), 浙江省科技计划项目 (2016C31084, LGF19F020008), 浙江省教育厅一般项目 (Y201840695), 宁波市软科学项目 (2018A10024) 资助
Supported by National Natural Science Foundation of China (61873240, 61602413), Zhejiang Provincial Natural Science Foundation (LY19F030016, LY17F020001), Project of Science and Technology Plan of Zhejiang Province (2016C31084, LGF19F020008), General Project of Education Department of Zhejiang Province (Y201840695), and Soft Science Foundation of Ningbo City (2018A10024)

本文责任编辑 黎铭

Recommended by Associate Editor LI Ming

1. 浙江万里学院大数据与软件工程学院 宁波 315100 2. 浙江工业大学计算机科学与技术学院 杭州 310023

1. College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315100 2. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023

信号稀疏表示是过去近 20 年来信号处理方向一个非常引人注目的研究领域, 目的是在给定的超完备字典中用尽可能少的原子来表示信号, 从而获得信号更为简洁的表示方式, 方便进一步对信号进行压缩、编码等加工处理. 随着压缩感知理论^[1]提出在所求系数足够稀疏时, 最小化 l_0 范数的 NP 难题^[2]可以转化为最小化 l_1 范数来进行求解, 稀疏表示已广泛应用于机器学习、计算机视觉、模式识别等领域.

在模式识别领域中, 具有代表性的是 Wright 等^[3]提出的一种基于稀疏表示的分类器 (Sparse representation-based classifier, SRC), 该分类器将稀疏表示成功应用于人脸识别, 有效解决了人脸图

Hangzhou 310023

像中遮挡、光照等问题. 其核心思想是通过选取完整的输入样本作为训练样本构造出超完备字典, 并由此得到测试样本的稀疏表示, 然后逐类构造出测试样本的近似线性表示, 最后寻找最小重构误差的类别标签作为测试样本的类别归属得以完成分类任务. Wright 指出 SRC 对数据缺损不敏感, 当重构编码系数足够稀疏时, 特征空间的选取变得不再重要, 这些特点使得 SRC 成为一种优秀的分类算法, 且在图像分类研究中获得了成功应用. 但是, 该算法仍然存在两个关键问题有待关注: 1) 选择 l_1 范数作为编码重构系数的稀疏约束是否是比较合适的策略; 2) 选择 Frobenius 范数 (F 范数) 作为线性重构表示的约束是否对图像表示足够有效. 基于此, 众多科研工作者致力于这两个问题的研究.

Yang 等^[4] 重新探究了 l_1 范数在模式识别任务中的作用, 发现 l_1 范数比 l_0 范数能提供更多有意义的分类信息, 即 l_0 范数只能实现稀疏, 而 l_1 范数不仅能实现稀疏, 还能保持数据间的协作关系. 对此, Zhang 等^[5] 通过理论分析和实验验证解释了 SRC 分类器获得优秀分类性能的根本原因并非它的稀疏性, 而是数据间的协作关系, 并提出了协作表示分类器 (Collaborative representation-based classifier, CRC). 其中, CRC 用 l_2 范数替代 SRC 中的 l_1 范数, 在不降低识别率的前提下大幅度提升了算法的运行效率. 此外, 为了加快稀疏编码过程, 相继提出了许多基于加权 l_1 或 l_2 范数的算法^[6-8].

上述算法均采用 F 范数作为线性重构表示的约束. 诸多学者认为, 找到合适的方式来描述重构误差的特性有利于重构及分类性能的提升. Lu 等^[9] 考虑到 F 范数对大规模异常点的处理较弱, 采用 l_1 范数对误差重构进行评估, 实验表明 l_1 范数比 F 范数对误差的评估更可靠. Naseem 等^[10] 则利用 Huber 估计处理一些随机的光照和噪声像素将所提线性回归分类器 (Linear regression classification, LRC) 扩展为强健的线性回归分类器 (Robust LRC, RLRC)^[11], 提升了 LRC 在多光照及噪声下的鲁棒性. Yang 等^[12] 采用 M 估计匹配一般的噪声, 提出了强健的稀疏编码算法 (Robust sparse coding, RSC). 随后, Yang 等^[13] 从样本特征的概率分布形式出发, 对不同的特征添加贡献度因子, 提出正则化鲁棒编码算法 (Regularized robust coding, RRC), 其具有更高的抗噪能力, 却丢失了样本的局部分布特性. 而 He 等^[14] 利用相关熵对处理噪声和噪点的稳定性优势, 提出一种基于相关熵的稀疏表示算法 (Correntropy-based sparse representation, CESR). 随后, He 等^[15] 又构建了半二次框架, 该框架结合两种现有的稀疏健壮回归模型: 以 SRC 为代表的应对误差修复的相加模型和以 CESR 和 RSC

为代表的应对误差检测的相乘模型. 通过分而治之的策略, He 等^[16] 将健壮人脸识别的过程分解为异常值检测和人脸识别两个阶段, 使得稀疏表示适用于大规模数据库.

实际上, 以上算法均属于多元分析的范畴. 众所周知, 多元分析是研究多个自变量与因变量相互关系的一组统计理论和方法. 其应用的限制条件是, 各个因素每一水平的样本必须是独立的随机样本, 其重复观测的数据服从正态分布, 且各总体方差相等. 总体来说, 上述这些算法均假设误差向量满足正态性或独立性, 但这并不适用于现实场景, 特别是在部分随机误差向量的分布呈现出重尾现象的情况下. 对于这种情况, Kibria 等^[17] 提出基于多元 t 误差的线性模型. Basu 等^[18] 使用多元幂指数分布作为语音识别领域的重尾分布. Liu^[19] 则假定观测的数据是相互依赖的, 且将多元幂指数回归模型扩展为矩阵变量幂指数回归模型. 这些研究意味着假定观测数据之间相互依赖且服从重尾分布对于描述一些实际观测具有现实意义. 事实上, 人脸图像中例如光照、遮挡或表情等噪声数据之间高度相关, 并不完全服从于独立同分布. 因此, 找到一种合适的针对矩阵变量的分布来描述误差矩阵的特征至关重要.

本文旨在解决部分遮挡和受光照影响的图像的重构及分类问题. 传统方法往往将图像以向量的形式进行存储分类, 忽视了图像数据的内部结构信息, 且它们依据最大似然估计的观点假设噪声数据服从独立同分布对误差进行处理, 不适用于现实场景. 本文所提模型强调误差矩阵中各个像素间的依赖性并假定该误差矩阵作为一个随机矩阵变量服从于矩阵多变量椭圆分布. 同时, 假定模型中的编码重构系数服从拉普拉斯分布或高斯分布. 由于通过假定所得的模型具有非光滑性, 不利于问题的优化求解, 本文引入辅助变量光滑模型, 随后采用迭代加权最小二乘法优化求解模型. 至此, 本文得到一种新的稀疏表示算法, 称为联合平滑矩阵多变量椭圆分布的稀疏表示算法 (Sparse representation with smoothed matrix multivariate elliptical distribution, SMED).

本文后续结构安排如下. 第 1 节介绍多变量分析的相关工作, 包括多元分布和矩阵多变量椭圆分布. 第 2 节具体描述 SMED 算法的目标函数, 并给出模型优化求解的方案. 第 3 节对 SMED 算法的收敛性以及复杂度进行分析. 第 4 节选择代表性的人脸数据库验证 SMED 的实际重构和分类性能. 最后, 第 5 节总结全文工作并给出后续展望.

1 相关工作

本节比较几种常用的多元分布并对矩阵多变量

椭圆分布进行简要介绍.

1.1 多元分布比较

不同于以往的算法, 本文旨在找到一种合适的能够描述人脸图像中例如光照、遮挡或表情等噪声数据特征的分布. 由于现实图像中的噪声数据均表现出尖峰重尾的特征, 这就需要找到一种更具适应性的分布来满足这一需求, 以减少异常值的影响, 提高鲁棒性. 图 1 给出了几种常用多元分布的比较图. 由图 1(d) 可知, 二维拉普拉斯分布及二维幂指数分布较二维高斯分布更具尖峰重尾性, 且图 1(c) 中二维幂指数分布的重尾范围大于图 1(b) 中的二维拉普拉斯分布. 此外, 通过三者在 XOY 面上的投影图可知, 二维高斯分布的投影面呈圆形形状, 二维拉普拉斯分布的投影面呈菱形形状, 二维幂指数分布的投影面呈椭圆形形状, 而椭圆形形状能够很好地兼

具圆形和菱形的特点, 这使得二维幂指数分布更具灵活性, 且椭圆形形状较圆形和菱形使得分布具有更肥的尾部区域, 故而对于异常值更具鲁棒性. 综上所述, 相对于二维高斯分布和二维拉普拉斯分布, 二维幂指数分布能更好地描述噪声数据的特征.

事实上, 二维幂指数分布作为 Kotz-type 分布^[20] 的特例是椭圆分布家族中的一员. 近年来, 椭圆分布被诸多学者证明是一种作为多元高斯分布、多元拉普拉斯分布的有效替代, 它在统计学、金融学、信号处理以及模式识别等领域中发挥着重要作用, 尤其在鲁棒性研究以及数据包含异常值或呈现出重尾现象的情况下表现更加突出^[21]. 依据文献^[22-24] 可知, 在一些随机现象真实场景中, 部分随机向量的分布数据往往包含异常值或呈现出长尾、重尾现象, 而椭圆分布对于这些现象具有较好的适应性. 例如: Liu 等^[25] 针对一般的非正态数据, 提出

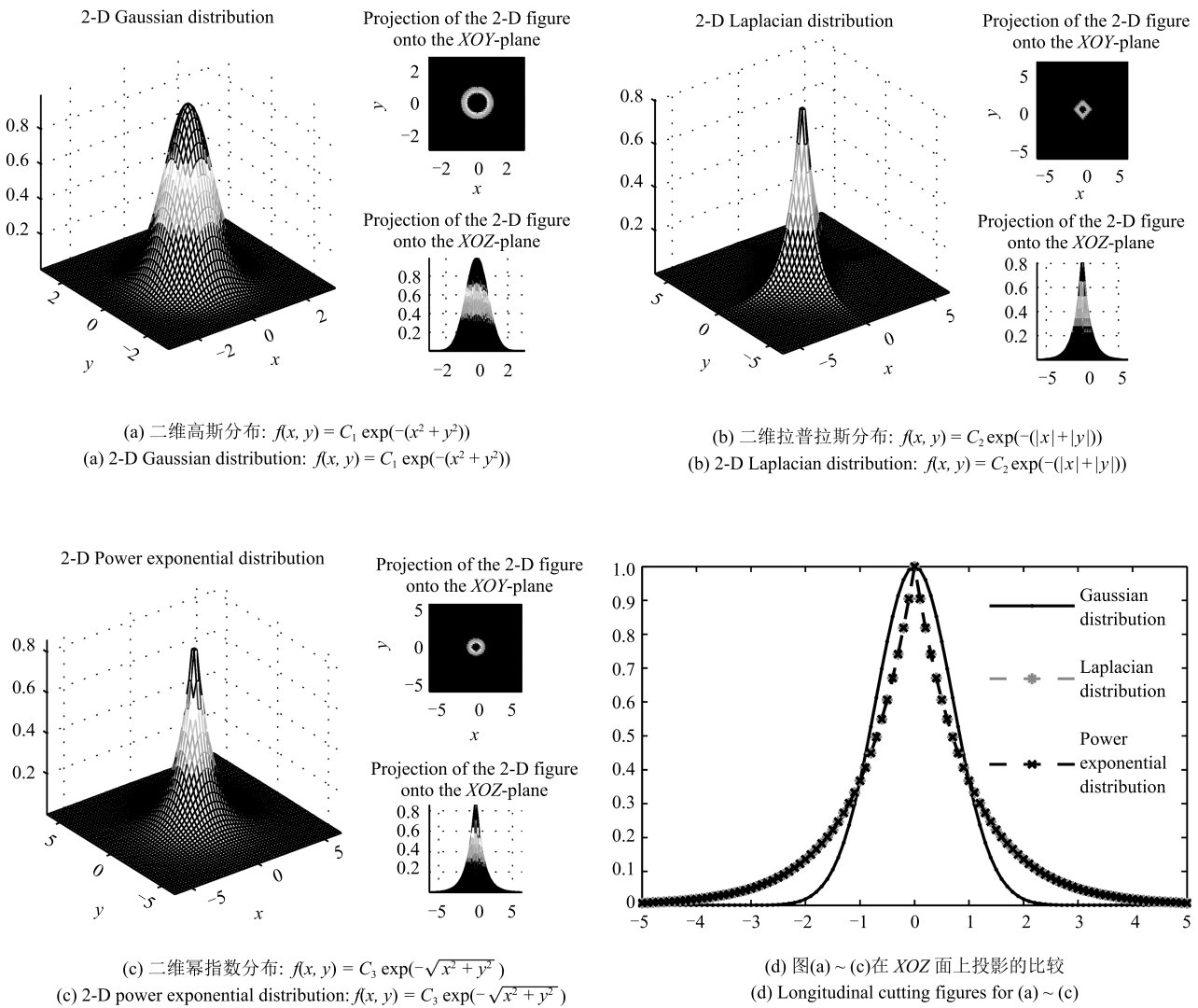


图 1 多元分布比较图 (常数取 $C_i = 1, i = 1, 2, 3$)

Fig. 1 The comparison chart of the multivariate distribution (where the constants are $C_i = 1, i = 1, 2, 3$)

基于幂指数分布的多元回归模型, 扩展了对于随机误差的一般分布假设. Sun 等^[26] 利用重尾椭圆分布对数据的结构协方差矩阵提供鲁棒估计. 此外, Fang 等^[27] 提出除了多元正态分布, 椭圆分布家族中的其他多元分布均能很好地描述随机矩阵中各个元素间的依赖关系. 因此, 假设误差矩阵服从椭圆分布来描述图像中的噪声数据能够更接近数据的固有特性, 即噪声像素间的依赖关系.

1.2 矩阵多变量椭圆分布

椭圆分布^[28] 作为概率分布家族中的一员, 允许同时考虑大量非高斯分布, 并且对典型未知数据的协方差矩阵提供最佳的估计, 能够有效构建误差结构. 本小节简要介绍多变量椭圆分布和矩阵多变量椭圆分布的定义.

定义 1. 给定 $\mathbf{x} \in \mathbf{R}^n$ 是一个 n 维随机向量, 如果对于某个 $\boldsymbol{\mu} \in \mathbf{R}^n$ 和某个对称正定矩阵 $\Sigma \in \mathbf{R}^{n \times n}$, \mathbf{x} 的特征函数具有如下形式:

$$\phi_{\mathbf{x}}(\mathbf{p}) = \mathbb{E}[\exp(\mathbf{i}\mathbf{p}^T \mathbf{x})] = \exp(\mathbf{i}\mathbf{p}^T \boldsymbol{\mu}) \phi(\mathbf{p}^T \Sigma \mathbf{p}) \quad (1)$$

其中, $\mathbb{E}[\cdot]$ 代表期望值, $\phi(\cdot)$ 是一个标量函数, 称为特征生成元, $\phi(\cdot): \mathbf{R} \rightarrow \mathbf{R}$, 且 $\mathbf{p} \in \mathbf{R}^n$. 则称 \mathbf{x} 服从具有分布参数 $\boldsymbol{\mu}$, Σ 和 ϕ 的 n 维多变量椭圆分布, 记为 $\mathbf{x} \sim E_n(\boldsymbol{\mu}, \Sigma, \phi)$.

定义 2. 给定 $X \in \mathbf{R}^{m \times n}$ 是一个 $m \times n$ 维随机矩阵, 如果对于某个矩阵 $M \in \mathbf{R}^{m \times n}$ 和某个对称正定矩阵 $\Sigma \in \mathbf{R}^{m \times m}$, X 的特征函数具有如下形式:

$$\Phi_X(P) = \exp(\text{tr}(\mathbf{i}P^T M)) \Phi(P^T \Sigma P) \quad (2)$$

其中, $\Phi(\cdot)$ 是特征生成元, $\Phi(\cdot): \mathbf{R}^{n \times n} \rightarrow \mathbf{R}$. 则称 X 服从具有分布参数 M , Σ 和 Φ 的 $m \times n$ 维矩阵多变量椭圆分布, 记为 $X \sim E_{m,n}(M, \Sigma, \Phi)$.

依据定义 2 可知, 矩阵多变量椭圆分布是一种对称概率分布, 也可通过概率密度函数对其进行定义

$$f(X) = Cg((X - M)^T \Sigma^{-1}(X - M)) \quad (3)$$

其中, C 是正数, 特征生成元 $g(\cdot): \mathbf{R}^{n \times n} \rightarrow \mathbf{R}$.

随机向量、随机矩阵的特征函数总是存在的, 并且与概率密度函数存在一一对应的关系. 因而椭圆分布的特征生成元一旦确定, 其分布的概率密度函数形式就确定了, 例如: 当特征生成元 $\phi(\cdot)$ 的形式为 $\phi(\mathbf{x}) = \exp(-\mathbf{x}/2)$ 时, \mathbf{x} 服从多元正态分布. 椭圆分布包括的多元分布有: 多元正态分布、多元柯西分布、多元拉普拉斯分布、多元 logistic 分布、多元 student 分布以及部分多元稳定分布等. 本节参考文献 [29] 中的特征生成元, 选择 $g(\cdot) = \exp(\text{tr}(-(\cdot)^{k/2}))$, $k > 0$, 该 $g(\cdot)$ 不仅保持了拉普

拉斯分布对异常值的鲁棒性, 而且具有长尾特性. 将 $g(\cdot) = \exp(\text{tr}(-(\cdot)^{k/2}))$ 代入式 (3) 中, 可得:

$$f(X) = C \exp(\text{tr}(-(X - M)^T \Sigma^{-1}(X - M))^{k/2}) \quad (4)$$

本文将以此概率密度函数为基础构建目标函数, 并辅以平滑目标函数、加快迭代最优过程等优势.

2 SMED 算法描述

本节提出一种联合平滑矩阵多变量椭圆分布的稀疏表示算法, 简称 SMED. 首先对该算法目标函数的构建过程进行描述, 然后给出模型求解的优化方案.

2.1 目标函数构建

探索人脸图像数据中误差矩阵的内部结构是重构及分类识别任务的一项成功策略, 受椭圆分布启发, 采用矩阵多变量椭圆分布描述误差矩阵, 具有更直观的物理意义和求解思路.

给定 n 个 $p \times q$ 维图像矩阵 $X_1^1, X_2^1, \dots, X_{n_1}^1, X_1^2, X_2^2, \dots, X_{n_2}^2, \dots, X_1^C, X_2^C, \dots, X_{n_C}^C$, 将其作为训练样本. 其中, X_i^c 代表第 c 类的第 i 个样本矩阵, 样本总类别数为 C , n_i 是第 i 类的样本个数, 且 $n = n_1 + n_2 + \dots + n_C$ 是总样本数, 测试样本表示为图像矩阵 $Y \in \mathbf{R}^{p \times q}$.

将任意测试样本 Y 通过训练样本进行线性表示, 可得

$$Y = X_1^1 \alpha_1 + X_2^1 \alpha_2 + \dots + X_{n_1}^1 \alpha_{n_1} + \dots + X_1^C \alpha_{n-n_C} + \dots + X_{n_C}^C \alpha_n + E \quad (5)$$

其中, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_n] \in \mathbf{R}^n$ 是训练样本所对应的重构编码系数, 为 $n \times 1$ 维列向量, E 为误差矩阵.

定义线性映射 $\mathbf{R}^d \rightarrow \mathbf{R}^{p \times q}$:

$$X(\boldsymbol{\alpha}) = X_1^1 \alpha_1 + X_2^1 \alpha_2 + \dots + X_{n_1}^1 \alpha_{n_1} + \dots + X_1^C \alpha_{n-n_C} + \dots + X_{n_C}^C \alpha_n \quad (6)$$

通过式 (6), 可将式 (5) 简化为

$$Y = X(\boldsymbol{\alpha}) + E \quad (7)$$

对应经典的线性向量回归模型^[30], 式 (5) 和式 (7) 给出了线性矩阵回归模型的一般形式.

本文假定误差矩阵 $E = Y - X(\boldsymbol{\alpha}) \sim E_{m,n}(M, \Sigma, \Phi)$, 即服从矩阵多变量椭圆分布, 且训练样本所对应的重构编码系数 $\alpha_1, \alpha_2, \dots, \alpha_n$ 独立同分布于高斯分布 ($l = 2$) 或拉普拉斯分布 ($l = 1$). 则依据最大后验概率估计可得

$$\boldsymbol{\alpha} = \arg \max_{\boldsymbol{\alpha}} \ln P(\boldsymbol{\alpha}|Y) + \ln P(\boldsymbol{\alpha}) \quad (8)$$

其中, $P(\boldsymbol{\alpha}|Y) = C_1 \exp(\text{tr}(-(Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})))^{k/2})$, $P(\boldsymbol{\alpha}) = C_2 \prod_{i=1}^n \exp(-\boldsymbol{\alpha}_i^l/\beta)$. 具体目标函数描述为

$$\min_{\boldsymbol{\alpha}} \xi(\boldsymbol{\alpha}) = \text{tr}((Y - X(\boldsymbol{\alpha}))^T \times (Y - X(\boldsymbol{\alpha})))^{\frac{k}{2}} + \frac{1}{\beta} \sum_{i=1}^n \|\boldsymbol{\alpha}\|_i^l \quad (9)$$

令 $\lambda = 2/\beta$, 且已知矩阵 X 的核范数表示为 $\|X\|_* = \text{tr}(X^T X)^{1/2}$, 向量 \boldsymbol{x} 的 l_p 范数表示为 $\|\boldsymbol{x}\|_p = (\sum_{i=1}^n |\boldsymbol{x}_i|^p)^{1/p}$, 则可将式 (9) 调整为

$$\min_{\boldsymbol{\alpha}} \xi(\boldsymbol{\alpha}) = \|Y - X(\boldsymbol{\alpha})\|_*^k + \frac{1}{2} \lambda \|\boldsymbol{\alpha}\|_i^l \quad (10)$$

其中, $k > 0$, $l = 1$ 或 2 . 当 $k = 1$, $l = 1$ 或 2 时, 上式的右边第 1 项为核范数, 第 2 项为 l_1 范数或 l_2 范数的平方, 这两项均为非光滑凸函数. 此外, 当 $0 < k < 1$ 时, 右边第 1 项为非光滑凹函数; 当 $k > 1$ 时, 两项均为非光滑凸函数.

然而, 由于式 (10) 具有非光滑性, 不利于目标函数的优化求解, 本文通过引入辅助变量光滑上式来解决这一问题, 则可将一般的平滑矩阵多变量椭圆分布模型归纳为如下优化问题:

$$\min_{\boldsymbol{\alpha}} \xi(\boldsymbol{\alpha}, \mu) = \left\| \begin{bmatrix} Y - X(\boldsymbol{\alpha}) \\ \mu I \end{bmatrix} \right\|_*^k + \frac{1}{2} \lambda \left\| \begin{bmatrix} \boldsymbol{\alpha} \\ \mu \mathbf{1} \end{bmatrix} \right\|_i^l \quad (11)$$

上式能处理的参数范围为 $\mu > 0$, $k > 0$, $k \neq 2$, $l = 1$ 或 2 , $I \in \mathbf{R}^{q \times q}$ 是单位矩阵且 $\mathbf{1} \in \mathbf{R}^n$ 是元素值均为 1 的 $n \times 1$ 维列向量.

文献 [31] 和文献 [32] 中的算法通过不同方法对非光滑目标函数进行了优化求解. 相比较而言, 式 (11) 采用 μI 和 $\mu \mathbf{1}$ 两项辅助变量来光滑目标函数的操作拥有以下优势: 1) 当 $\mu > 0$ 时, $\xi(\boldsymbol{\alpha}, \mu)$ 是光滑的, 它较非光滑的目标函数更容易进行优化求解; 2) 当 $k \geq 1$ 时, 依据式 (11) 两项的凸性质可知 $\xi(\boldsymbol{\alpha}, \mu)$ 是一个凸函数, 则最小化目标函数即为凸优化问题, 可得全局最优解.

此外, 根据式 (12) 可知 $\xi(\boldsymbol{\alpha}, \mu) \geq \xi(\boldsymbol{\alpha})$, 且当且仅当 $\mu = 0$ 时, $\xi(\boldsymbol{\alpha}, \mu) = \xi(\boldsymbol{\alpha})$. 这意味着 $\xi(\boldsymbol{\alpha}, \mu)$ 是 $\xi(\boldsymbol{\alpha})$ 通过 μ 的优化函数, 如果 $\xi(\boldsymbol{\alpha}, \mu)$ 是递减函数则 $\xi(\boldsymbol{\alpha})$ 亦为递减函数.

$$\left\| \begin{bmatrix} Y - X(\boldsymbol{\alpha}) \\ \mu I \end{bmatrix} \right\|_*^k = \sum_{i=1}^n (\sigma_i((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I))^{\frac{k}{2}} =$$

$$\sum_{i=1}^n (\sigma_i((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha}))) + \mu^2)^{\frac{k}{2}} \geq \sum_{i=1}^n (\sigma_i((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha}))))^{\frac{k}{2}} = \|Y - X(\boldsymbol{\alpha})\|_*^k \quad (12)$$

其中, $\sigma_i(\cdot)$ 代表矩阵的第 i 个正奇异值.

2.2 模型优化求解

诸多方法可用来优化求解模型 (11), 例如: 半正定规划法 (Semi-definite programming, SDP)^[33] 是在满足约束“对称矩阵的仿射组合半正定”的条件下使线性函数极大(极小)化的问题. 加速近端梯度法 (Accelerated proximal gradient, APG)^[34] 是一个求解 l_1 优化问题较快的方法, 其收敛速度为二阶收敛, 相比较其他方法具有明显的速度优势. 交替方向乘子法 (Alternating direction method of multipliers, ADMM)^[35] 又称增广拉格朗日法, 是求解约束优化问题的一类重要方法, 具有不需要惩罚参数趋向于无穷大等优点^[36]. 然而, SDP 因其算法复杂度高不适用于大规模数据 (对于 $n \times n$ 维的矩阵, SDP 的算法复杂度是 $O(n^6)$). APG 需要目标函数具有 Lipschitz 连续梯度, 这使得该方法的使用受到限制. 相比较 SDP 和 APG, ADMM 以其较好的性能和严格的理论保证, 成为众多学者研究的热点. 但是, ADMM 需要在非光滑目标函数中加入多个辅助变量乘子, 这大大降低了算法的收敛速度, 且其在优化求解目标函数的过程中需要应用奇异值分解 (Singular value decomposition, SVD)^[37], 增加了算法的复杂度. 本节采用迭代加权最小二乘法 (Iteratively reweighted least squares, IRLS)^[38] 对模型进行优化求解, 避免了 ADMM 需要加入多个辅助变量乘子和进行奇异值分解的缺陷.

依据相关公式, 式 (11) 可优化调整为

$$\min_{\boldsymbol{\alpha}} \text{tr}((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I)^{\frac{k}{2}} + \frac{1}{2} \lambda \left(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2 \right)^{\frac{l}{2}} \quad (13)$$

其中, $\boldsymbol{\alpha}_i$ 代表向量 $\boldsymbol{\alpha}$ 的第 i 个元素.

令

$$\gamma(\boldsymbol{\alpha}) = \text{tr}((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I)^{\frac{k}{2}}$$

$$\delta(\boldsymbol{\alpha}) = \left(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2 \right)^{\frac{l}{2}}$$

则

$$\xi(\boldsymbol{\alpha}, \mu) = \gamma(\boldsymbol{\alpha}) + \frac{\lambda}{2} \times \delta(\boldsymbol{\alpha})$$

且

$$\frac{\partial \xi(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} + \frac{\partial \delta(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$$

其中, $\gamma(\boldsymbol{\alpha})$ 针对 $\boldsymbol{\alpha}$ 的梯度为

$$\begin{aligned} \frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} &= \frac{\partial(\text{tr}((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I)^{\frac{k}{2}}))}{\partial \boldsymbol{\alpha}} = \\ &= \frac{\frac{k}{2}((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I)^{\frac{k}{2}-1} \times \\ &\quad \partial(\text{tr}((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I)))}{\partial \boldsymbol{\alpha}} = \\ &= \frac{k(X^T X \boldsymbol{\alpha} - X^T Y)}{((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I)^{1-\frac{k}{2}}} \end{aligned} \quad (14)$$

令对应 $\gamma(\boldsymbol{\alpha})$ 的加权矩阵 $R = ((Y - X(\boldsymbol{\alpha}))^T(Y - X(\boldsymbol{\alpha})) + \mu^2 I)^{(k-2)/2}$, 则上式可简化为

$$\frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = k(X^T X \boldsymbol{\alpha} - X^T Y)R \quad (15)$$

其次, $\delta(\boldsymbol{\alpha})$ 针对 $\boldsymbol{\alpha}$ 的梯度为

$$\begin{aligned} \frac{\partial \delta(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} &= \frac{\partial(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2)^{\frac{1}{2}}}{\partial \boldsymbol{\alpha}} = \\ &= \frac{\frac{l}{2} \left(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2 \right)^{\frac{1}{2}-1} \frac{\partial(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2)}{\partial \boldsymbol{\alpha}}}{\frac{\partial(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2)^{1-\frac{1}{2}}}{\partial \boldsymbol{\alpha}}} = \\ &= \frac{l \boldsymbol{\alpha}}{(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2)^{1-\frac{1}{2}}} \end{aligned} \quad (16)$$

令 $T = (\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2)^{(l-2)/2}$, 则上式可简化为

$$\frac{\partial \delta(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = l \boldsymbol{\alpha} T \quad (17)$$

通过 $\frac{\partial \xi(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{\partial \gamma(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} + \frac{\partial \delta(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}$, 并令结果为 0, 可得

$$\begin{aligned} k(X^T X \boldsymbol{\alpha} - X^T Y)R + \frac{1}{2} \lambda l \boldsymbol{\alpha} T &= 0 \rightarrow \\ kX^T X \boldsymbol{\alpha} R + \frac{1}{2} \lambda l \boldsymbol{\alpha} T &= kX^T Y R \rightarrow \\ kX^T X \boldsymbol{\alpha} + \frac{1}{2} \lambda l \boldsymbol{\alpha} T R^{-1} &= kX^T Y \end{aligned} \quad (18)$$

式 (18) 是经典的 Sylvester 方程形式 ($AX + XB = C$), 方程求解的算法复杂度是 $O(n^3)^{[39-40]}$. 实验过程中可利用 MATLAB 的 *lyap* 命令求解式 (18).

综上所述, 完整的 SMED 描述如算法 1 所示. 其中, $\text{Vec}(\cdot)$ 代表将矩阵通过列转化为向量的操作, $\text{Mat}(\cdot)$ 代表将向量转化为矩阵的操作.

算法 1. SMED 描述

输入. n 个 $p \times q$ 维图像矩阵 $X_1^1, X_2^1, \dots, X_{n_1}^1, X_1^2, X_2^2, \dots, X_{n_2}^2, \dots, X_1^C, X_2^C, \dots, X_{n_C}^C$ ($n = n_1 + n_2 + \dots + n_C$); 图像矩阵 $Y \in \mathbf{R}^{p \times q}$; 模型参数 $\lambda > 0, \mu > 0$; 终止条件参数 $\varepsilon > 0$.

输出. 重构编码系数向量 $\boldsymbol{\alpha}$.

1) 令

$$\begin{aligned} X &= [\text{Vec}(X_1^1), \text{Vec}(X_2^1), \dots, \text{Vec}(X_{n_1}^1), \\ &\quad \text{Vec}(X_1^2), \text{Vec}(X_2^2), \dots, \text{Vec}(X_{n_2}^2), \\ &\quad \vdots \\ &\quad \text{Vec}(X_1^C), \text{Vec}(X_2^C), \dots, \text{Vec}(X_{n_C}^C)] \end{aligned}$$

初始化加权矩阵 $R = I$.

2) 设迭代次数 $t = 1$.

3) 固定加权矩阵 R_t, T_t , 依照式 (19) 计算更新重构编码系数向量 $\boldsymbol{\alpha}_{t+1}$

$$kX^T X \boldsymbol{\alpha} + \frac{1}{2} \lambda \boldsymbol{\alpha} T_t R_t^{-1} = kX^T \text{Vec}(Y) \quad (19)$$

4) 固定 $\boldsymbol{\alpha}_{t+1}$, 依照式 (20) 和式 (21) 分别更新加权矩阵 R_{t+1}, T_{t+1}

$$\begin{aligned} R_{t+1} &= ((Y - \text{Mat}(X(\boldsymbol{\alpha}_{t+1})))^T \times \\ &\quad (Y - \text{Mat}(X(\boldsymbol{\alpha}_{t+1}))) + \mu^2 I)^{\frac{k}{2}-1} \end{aligned} \quad (20)$$

$$T_{t+1} = \left(\sum_{i=1}^n |\boldsymbol{\alpha}_i|^2 + \mu^2 \right)^{\frac{1}{2}-1} \quad (21)$$

5) 如满足收敛条件 $\|\boldsymbol{\alpha}_{t+1} - \boldsymbol{\alpha}_t\|_\infty \leq \varepsilon$, 则输出结果, 算法中止; 反之令 $t = t + 1$, 转至步骤 3.

3 收敛性和复杂度分析

SMED 采用迭代加权最小二乘法进行模型迭代求解, 在固定部分变量的前提下优化余下未知变量. 根据算法 1 描述, 每次迭代的关键步骤 (20) 和 (21) 都是闭式解, 因此其单个变量更新都存在唯一解. 定理 1. 说明所提算法在迭代过程中使目标函数 (11) 的值逐步下降, 并最终收敛.

引理 1. 给定任意矩阵 $X \in \mathbf{R}^{m \times n}$ 和 $Y \in \mathbf{R}^{m \times n}$, 其中, 矩阵 X 和 Y 的列向量为非零向量. 令 $f_i(x), i = 1, 2, \dots, n$ 为可微凹函数, 可得

$$\begin{aligned} \sum_{i=1}^n f_i(\|Y_i\|_2^2) - f_i(\|X_i\|_2^2) &\geq \\ \text{tr}((Y^T Y - X^T X)M) &\end{aligned} \quad (22)$$

其中, 对角矩阵 $M \in \mathbf{R}^{n \times n}$ 的对角元素为 $M_{ii} = \nabla f_i(\|Y_i\|_2^2)$.

设定 $f_i(x) = x^{q/2}$, $0 < q < 2$, 将其代入上式可得式 (22) 的一个特例

$$\|Y\|_q^q - \|X\|_q^q \geq \frac{q}{2} \text{tr}((Y^T Y - X^T X)M) \quad (23)$$

其中, $M_{ii} = (\|Y_i\|_2^2)^{q/2-1}$.

证明. 根据凹函数定义文献 [41] 可知, 在定义域 $(0, \infty)$ 中, 当 $p \in (0, 1]$ 时, 凹函数 x^p 满足 $y^p - x^p + py^{p-1}(x - y) \geq 0$. 对矩阵 $X \in \mathbf{R}^{m \times n}$ 和 $Y \in \mathbf{R}^{m \times n}$, 则有

$$\begin{aligned} f_i(\|Y_i\|_2^2) - f_i(\|X_i\|_2^2) &\geq \\ \nabla f_i(\|Y_i\|_2^2)(\|Y_i\|_2^2 - \|X_i\|_2^2) &= \\ \nabla f_i(\|Y_i\|_2^2)(Y_i^T Y_i - X_i^T X_i) &= \\ \text{tr}((Y^T Y - X^T X)M) &\quad (24) \end{aligned}$$

将式 (24) 中所有的 $i \in \{1, 2, \dots, n\}$ 相加, 则式 (22) 得证. \square

引理 2. 给定任意矩阵 $X, Y \in S_{++}^n$, 其中, S_{++}^n 为对称正定矩阵. 令 X, Y 的特征值 λ_i 分别满足 $\lambda_1(X) \geq \lambda_2(X) \geq \dots \geq \lambda_n(X) \geq 0$ 及 $\lambda_1(Y) \geq \lambda_2(Y) \geq \dots \geq \lambda_n(Y) \geq 0$. 则可得

$$\text{tr}(XY) \geq \sum_{i=1}^n \lambda_i(X)\lambda_{n-i+1}(Y) \quad (25)$$

证明. 见文献 [42] 中的引理 2. \square

引理 3. 给定任意矩阵 $X \in S_{++}^n$, 当 $p \in (0, 1]$ 时, $\text{tr}(X^p)$ 为凹函数.

证明. 由引理 2 可知, 对 $X, Y \in S_{++}^n$ 有

$$\begin{aligned} \text{tr}(X^T Y^{p-1}) &\geq \sum_{i=1}^n \lambda_i(X)\lambda_{n-i+1}(Y^{p-1}) = \\ &\sum_{i=1}^n \lambda_i(X)\lambda_i^{p-1}(Y) \end{aligned}$$

则通过推论和引理 1 可得

$$\begin{aligned} \text{tr}(Y^p) - \text{tr}(X^p) + \text{tr}(p(X - Y)^T Y^{p-1}) &\geq \\ \sum_{i=1}^n [\lambda_i(Y^p) - \lambda_i(X^p) + & \\ p\lambda_i(X)\lambda_i^{p-1}(Y) - p\lambda_i(Y^p)] &= \\ \sum_{i=1}^n [\lambda_i^p(Y) - \lambda_i^p(X) + & \\ p\lambda_i^{p-1}(Y)(\lambda_i(X) - \lambda_i(Y))] &\geq 0 \quad (26) \end{aligned}$$

由式 (26) 可知, 当 $p \in (0, 1]$ 时, $\text{tr}(X^p)$ 为凹函数. \square

通过上述引理 1 可知, 假设 $g(X)$ 为可微凹函数, 那么对于 $X, Y \in S_{++}^n$ 可得

$$g(Y) - g(X) \geq \text{tr}((Y - X)^T \nabla g(X)) \quad (27)$$

此外, 由引理 3 可知, 当 $l \in (0, 2]$ 时, $\text{tr}(X^{l/2})$ 为凹函数. 则设定 $g(X) = \text{tr}(X^{l/2})$, 将 $g(X)$ 代入上式可得

$$\begin{aligned} \left\| \begin{bmatrix} Y \\ \mu I \end{bmatrix} \right\|_*^l - \left\| \begin{bmatrix} X \\ \mu I \end{bmatrix} \right\|_*^l &\geq \\ \frac{l}{2} \text{tr}((Y^T Y - X^T X)^T (Y^T Y + \mu^2 I)^{\frac{l}{2}-1}) &\quad (28) \end{aligned}$$

定理 1. 通过算法 1 迭代计算, 目标函数式 (11) 的值逐次下降直至收敛.

证明. 定义第 t 次迭代时的误差矩阵 $E_t = Y - X(\alpha_t)$, 由于 α_{t+1} 是式 (18) 的解, 则有

$$k(X^T X \alpha_{t+1} - X^T Y)R_t + \frac{1}{2} \lambda \alpha_{t+1}^T T_t = 0 \quad (29)$$

通过对式 (29) 两边分别点乘 $\alpha_t - \alpha_{t+1}$, 可得

$$\begin{aligned} \frac{1}{2} \lambda (\alpha_t - \alpha_{t+1})^T \alpha_{t+1} T_t &= \\ -k(\alpha_t - \alpha_{t+1})^T (X^T X \alpha_{t+1} - X^T Y)R_t &= \\ -k(\alpha_{t+1} - \alpha_t)^T X^T (Y - X \alpha_{t+1})R_t &= \\ -k(X \alpha_{t+1} - X \alpha_t)^T (Y - X \alpha_{t+1})R_t &= \\ -k(X \alpha_{t+1} - Y + Y - X \alpha_t)^T \times & \\ (Y - X \alpha_{t+1})R_t &= \\ -k(E_t - E_{t+1})^T E_{t+1} R_t &\quad (30) \end{aligned}$$

已知误差矩阵 $E = Y - X(\alpha) \in S_{++}^n$, 通过式 (28) 可得

$$\begin{aligned} \left\| \begin{bmatrix} Y - X(\alpha_t) \\ \mu I \end{bmatrix} \right\|_*^k - \left\| \begin{bmatrix} Y - X(\alpha_{t+1}) \\ \mu I \end{bmatrix} \right\|_*^k &\geq \\ \frac{k}{2} \text{tr}(((Y - X(\alpha_t))^T (Y - X(\alpha_t)) - & \\ (Y - X(\alpha_{t+1}))^T (Y - X(\alpha_{t+1})))^T \times & \\ ((Y - X(\alpha_t))^T (Y - X(\alpha_t)) + \mu^2 I)^{\frac{k}{2}-1}) &= \\ \frac{k}{2} \text{tr}((Y - X(\alpha_t) - Y + X(\alpha_{t+1}))^T & \\ (Y - X(\alpha_t) - Y + X(\alpha_{t+1}))R_t) + & \\ k \text{tr}((Y - X(\alpha_t) - Y + X(\alpha_{t+1}))^T & \\ (Y - X(\alpha_{t+1}))R_t) &= \\ \frac{k}{2} \text{tr}((E_t - E_{t+1})^T (E_t - E_{t+1})R_t) + & \\ k \text{tr}((E_t - E_{t+1})^T E_{t+1} R_t) &\quad (31) \end{aligned}$$

此外, 结合式 (23) 和式 (30) 有

$$\begin{aligned} & \frac{1}{2}\lambda \left\| \left[\begin{array}{c} \boldsymbol{\alpha}_t \\ \mu \mathbf{1} \end{array} \right] \right\|_l^l - \frac{1}{2}\lambda \left\| \left[\begin{array}{c} \boldsymbol{\alpha}_{t+1} \\ \mu \mathbf{1} \end{array} \right] \right\|_l^l \geq \\ & \frac{\lambda l}{2} \text{tr}((\boldsymbol{\alpha}_t^T \boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}^T \boldsymbol{\alpha}_{t+1}) T_t) = \\ & \frac{\lambda l}{2} \text{tr}((\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1})^T (\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}) T_t) + \\ & \frac{1}{2} \lambda \text{tr}((\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1})^T \boldsymbol{\alpha}_{t+1} T_t) = \\ & \frac{\lambda l}{2} \text{tr}((\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1})^T (\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}) T_t) - \\ & k \text{tr}((E_t - E_{t+1})^T E_{t+1} R_t) \end{aligned} \quad (32)$$

将式 (31) 与式 (32) 相加可以得到

$$\begin{aligned} & \xi(\boldsymbol{\alpha}_t, \mu) - \xi(\boldsymbol{\alpha}_{t+1}, \mu) = \\ & \left\| \left[\begin{array}{c} Y - X(\boldsymbol{\alpha}_t) \\ \mu I \end{array} \right] \right\|_*^k - \left\| \left[\begin{array}{c} Y - X(\boldsymbol{\alpha}_{t+1}) \\ \mu I \end{array} \right] \right\|_*^k + \\ & \frac{1}{2}\lambda \left\| \left[\begin{array}{c} \boldsymbol{\alpha}_t \\ \mu \mathbf{1} \end{array} \right] \right\|_l^l - \frac{1}{2}\lambda \left\| \left[\begin{array}{c} \boldsymbol{\alpha}_{t+1} \\ \mu \mathbf{1} \end{array} \right] \right\|_l^l = \\ & \frac{k}{2} \text{tr}((E_t - E_{t+1})^T (E_t - E_{t+1}) R_t) + \\ & \frac{\lambda l}{2} \text{tr}((\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1})^T (\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}) T_t) \geq 0 \end{aligned} \quad (33)$$

即目标函数 (11) 的值在迭代更新的过程中逐次下降, 并具有明确的下确界 0, 则该算法收敛性得证. \square

定理 2. 通过算法 1 产生的序列 $\boldsymbol{\alpha}_t$ 收敛到最优解 $\boldsymbol{\alpha}^*$.

证明. 结合式 (25) 和式 (33) 有

$$\begin{aligned} & \xi(\boldsymbol{\alpha}_t, \mu) - \xi(\boldsymbol{\alpha}_{t+1}, \mu) \geq \\ & \frac{k}{2} \sum_{i=1}^n \lambda_{n-i+1} (R_t) \lambda_i ((E_t - E_{t+1})^T (E_t - E_{t+1})) + \\ & \frac{\lambda l}{2} \sum_{i=1}^n T_t \lambda_i ((\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1})^T (\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1})) \geq \\ & \frac{\theta}{4} (2k \|E_t - E_{t+1}\|_F^2 + \lambda l \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}\|_2^2) \end{aligned}$$

上式通过 $t \geq 1$ 次迭代求和得

$$\begin{aligned} \xi(\boldsymbol{\alpha}_1, \mu) & \geq \frac{\theta}{4} \sum_{t=1}^{\infty} (2k \|E_t - E_{t+1}\|_F^2 + \\ & \lambda l \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}\|_2^2) \end{aligned} \quad (34)$$

由式 (34) 可知, $\lim_{t \rightarrow \infty} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}\|_2 = 0$.

因序列 $\{\boldsymbol{\alpha}_t\}$ 有界, 故存在 $\hat{\boldsymbol{\alpha}}$ 和子序列 $\{\boldsymbol{\alpha}_{t_j}\}$ 使得 $\lim_{j \rightarrow \infty} \boldsymbol{\alpha}_{t_j} \rightarrow \hat{\boldsymbol{\alpha}}$. 而 $\boldsymbol{\alpha}_{t_j+1}$ 作为式 (18) 迭代过程的其中一解可表示为

$$k(X^T X \boldsymbol{\alpha}_{t_j+1} - X^T Y) R_{t_j} + \frac{1}{2} \lambda \boldsymbol{\alpha}_{t_j+1} T_{t_j} = 0 \quad (35)$$

式 (35) 意味着在 $j \rightarrow \infty$ 的过程中, $\boldsymbol{\alpha}_{t_j+1}$ 能收敛到某个 $\tilde{\boldsymbol{\alpha}}$.

因此, 通过 $\lim_{t \rightarrow \infty} \|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+1}\|_2 = 0$, 可得

$$\|\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}}\|_2 = \lim_{j \rightarrow \infty} \|\boldsymbol{\alpha}_{t_j} - \boldsymbol{\alpha}_{t_j+1}\|_2 = 0$$

则通过上式可知 $\hat{\boldsymbol{\alpha}} = \tilde{\boldsymbol{\alpha}}$. 将 $\hat{\boldsymbol{\alpha}}$ 表示为 $\boldsymbol{\alpha}^*$ 且使得 $j \rightarrow \infty$, 可将式 (35) 调整为

$$k(X^T X \boldsymbol{\alpha}^* - X^T Y) R^* + \frac{1}{2} \lambda \boldsymbol{\alpha}^* T^* = 0$$

其中, R^* 和 T^* 为对应 $\boldsymbol{\alpha}^*$ 下的值. 至此, 收敛到最优解得证. \square

算法 1 的关键耗时步骤是 3 个未知变量的更新操作, 包括 R_t , T_t 和 $\boldsymbol{\alpha}_t$. 其中 R_t 依式 (20) 计算, 其运算复杂度是 $O(d)$, T_t 依式 (21) 计算, 其运算复杂度是 $O(n)$, 重构编码系数 $\boldsymbol{\alpha}_t$ 通过求解式 (19) 获得, 式 (19) 是一个标准的 Sylvester 公式, 其求解的典型算法是 Bartel-Stewart 法, 先通过 QR 分解将系数矩阵变换成 Schur 形式, 再采用回代法求解三角系统, 整体运算复杂度是 $O(n^3)$. 一般情况下, $d \ll n^3$, 因此, SMED 每次迭代的复杂度是 $O(n^3)$. 假设算法的实际迭代次数是 t_m , 可得 SMED 的综合运算复杂度是 $O(n^3 t_m)$. 通过多次实验发现, 迭代次数 t_m 一般满足 $t_m \leq 30$.

4 实验分析

为验证所提算法的有效性, 以人脸识别为例, 分别采用 AR^[13]、ExYaleB^[43] 和 PubFig^[44] 三个公开人脸数据库进行实验分析. 对 SMED 来说, 因其模型中 k, l 的取值范围为 $k > 0, k \neq 2, l = 1$ 或 2 , 且当 $k = 1$ 时, 式 (10) 第 1 项为核范数, 第 2 项为 l_1 范数或 l_2 范数的平方; 当 $0 < k < 1$ 时, 式 (10) 第 1 项为非光滑凹函数; 当 $k > 1$ 时, 两项均为非光滑凸函数, 故实验中分别依据上述范围取 $k = 1, l = 1; k = 1/2, l = 1; k = 3/2, l = 2$. 即 SMED₁₁, SMED_{1/21} 和 SMED_{3/22} 与对比算法进行实验来观察模型对于不同特点目标函数优化求解后的算法性能情况. 第 4.1 节讨论了模型的参数敏感性并验证了所提算法的可视化重构性能; 第 4.2~4.4 节均为识别性能测试, 分别在真实影响下 (光照及伪装影响和像素污损影响)、随机遮

挡下(方形白色、黑色、baboon 图像遮挡和非方形鲜花图像遮挡)以及非限制场景下将 SMED 与经典算法的性能优劣进行了实验对比;最后,第 4.5 节对比了不同算法的模型运行效率. 实验中所有人脸图像都经过统一的裁剪,所有训练样本和测试样本都通过 l_2 范数进行规范化处理. 实验平台为 32 位 Win7 操作系统, Intel Core i5 处理器, 双核主频 2.80 GHz, 4 GB 内存以及 MATLAB2014 运行环境.

4.1 人脸重构

在本小节中,选用 AR 人脸数据库可视化所提算法并对比各算法的重构性能. AR 人脸库包含 126 人(70 位男性和 56 位女性)的 4000 多张彩色图像,图像通过两个不同时段拍摄获取,每个时段每人共拍摄 13 张包含在不同表情、光照条件和现实遮挡(墨镜和围巾)下的大小为 768 像素 \times 576 像素的彩色图像. 与文献 [13] 类似,本文抽选其中一个子集(50 位男性和 50 位女性)作为实验测试. 选取每人每时段 4 张无遮挡正面人脸图像,即 100 人每人 8 张共 800 张图像作为训练样本. 此外,构建三个测试样本: 1) 选取每人每时段 3 张在不同光照条件下的人脸图像; 2) 选取每人每时段 3 张在墨镜遮挡下的人脸图像; 3) 选取每人每时段 3 张在围巾遮挡下的人脸图像. 3 个测试样本均包含 600 张图像(100 人每人 6 张). 本文将 AR 人脸库中的图像统一裁剪至 83 像素 \times 60 像素. 对比算法选用经典的分类器 SRC 和 CRC, 抗噪性较强的 RRC, 能快速提取特征的 FFS^[45] 以及基于半二次相加框架的 HQ_A 和基于半二次相乘框架的 HQ_M^[15], 实验中 SRC, CRC, RRC, FFS, HQ_A 和 HQ_M 的参数设置分别参考

了文献 [3]、[5]、[13]、[15] 和 [45], 且 RRC 算法采用识别率较高的 l_1 范数约束.

4.1.1 参数 μ_c 和 ρ 的敏感度分析

参数的选择对算法整体运行具有重要的意义. 选择适当的参数,有利于加快目标函数的收敛速度并能精确地求解模型. 本小节对 SMED 算法中参数 μ 值选择的敏感度进行了分析. 实验中通过 $\mu_{t+1} = \mu_t/\rho$, $\rho > 1$ 来递减参数 μ , 并将 μ 的初始值设置为 $\mu_0 = \mu_c \|X\|_2$. 因此,参数 μ 的值取决于 μ_c 和 ρ . 本小节构建两个实验来评估 μ_c 和 ρ 的值对 SMED 算法性能的敏感度,其中选择具有代表性的 SMED₁₁ 作为实验对象,测试样本选取第二类包含墨镜遮挡的人脸图像,且统一设置参数 λ 的值为 0.1. 图 2 所示即为不同参数值及不同迭代次数下 SMED₁₁ 算法的目标函数值. 图 2(a) 固定 $\rho = 1.1$, 从而评估不同 μ_c 值对算法目标函数值的影响,图 2(b) 固定 $\mu_c = 0.01$, 从而评估不同 ρ 值对算法目标函数值的影响.

由图 2(a) 可知,当 $\mu_c = 0.001$ 和 $\mu_c = 0.0001$ 时, μ_c 的值极小,对应的算法收敛值曲线显示经过不到 5 次的迭代 SMED₁₁ 就达到了收敛,这说明 μ_c 的值越小越将导致算法的不精确求解. 但是随着 μ_c 值的增大, SMED₁₁ 就需要更多次的迭代才能收敛算法,如图所示,当 $\mu_c = 0.01$ 时,算法仅需约 15 次的迭代就能达到收敛;当 $\mu_c = 0.1$ 时,需要约 40 次;而当 $\mu_c = 1$ 时,却需要将约 60 次的迭代才能达到收敛. 由图 2(b) 可知,参数 ρ 值的选择对算法敏感度的情况与参数 μ_c 类似. ρ 的值越大越将导致算法的快速收敛,使得模型求解不精确,而 ρ 的值越小越能提升模型求解的精确度. 因此,为了兼顾算法对模型求解的效率和精确度,综合图 2 的显示情况可

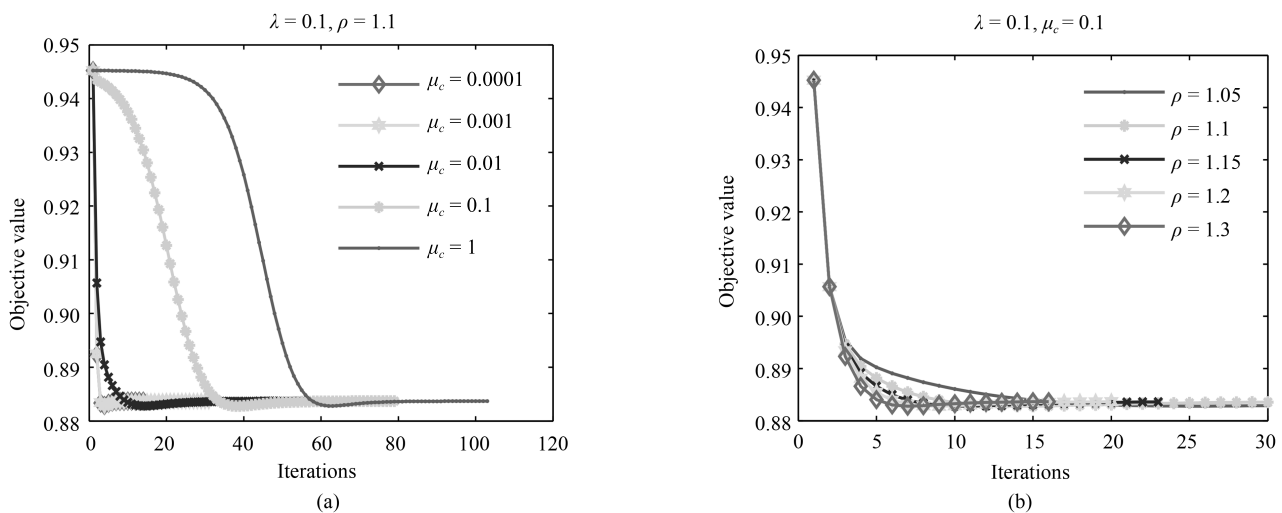


图 2 不同参数 μ_c 和 ρ 值下的 SMED₁₁ 算法收敛值曲线

Fig. 2 Convergence curves of SMED₁₁ algorithm on AR with different regularization parameters μ_c and ρ

知所选参数 μ_c 的值不应太小, 而 ρ 的值不应太大. 综上所述, 本节选取 $\mu_c = 0.01, \rho = 1.1$ 作为实验中的参数值输入.

4.1.2 可视化重构性能

本小节评估所提算法的重构性能, 通过可视化所构建的 AR 人脸库的 3 个测试样本中的图像来比较 SMED₁₁, SMED_{1/21}, SMED_{3/22} 和对比算法的人脸重构能力. 部分实验结果如图 3 所示. 图 3(a) 为重构前的原始图像, 从上至下依次为光照影响的人脸图像、墨镜遮挡的人脸图像、围巾遮挡的人脸图像. 图 3(b) 为通过 SMED₁₁ 重构后的图像与对应每类图像的重构误差值 $e_i(Y) = \|\hat{Y} - \hat{Y}_i\| = \|X\alpha^* - X\delta_i(\alpha^*)\|_*$, 其中 α^* 代表重构编码系数的最优解, 则 $X\alpha^*$ 为 Y 的重构图像. $\delta_i(\alpha^*)$ 代表 α^* 向量中除关联第 i 类的元素含值外其他元素均为零, 则

$X\delta_i(\alpha^*)$ 为关联第 i 类的 Y 的重构图像. 而分类识别的原理是 $e_i(Y) = \min_i e_i(Y)$, 即第 i 类的重构误差值最小, 则图像 Y 就归属于第 i 类. 从图 3(b) 中可见, 圆点为重构误差的最小值, 横坐标代表图像类别, 从上至下依次可得在光照影响、墨镜遮挡、围巾遮挡下的人脸图像通过 SMED₁₁ 重构后均归属于第 10 人的图像类别, 符合实际情况. 此外, 图 3(b) 的这三张结果图还说明了光照影响对 SMED₁₁ 算法的分类性能干扰最小, 遮挡类的干扰明显大于光照, 特别地, 围巾遮挡的干扰最大. 图 3(c)~3(k) 为所提算法与对比算法基于图 3(a) 的重构图像 (左侧) 和冗余图像 (右侧), $E = Y - \hat{Y}$.

由图 3(c)~3(k) 可知, SMED 算法在重构性能上较其他对比算法突出, 这揭示了重构编码系数的稀疏性对人脸图像的重构具有一定程度的有效性. 而对于 SMED₁₁, SMED_{1/21} 和 SMED_{3/22} 这三者

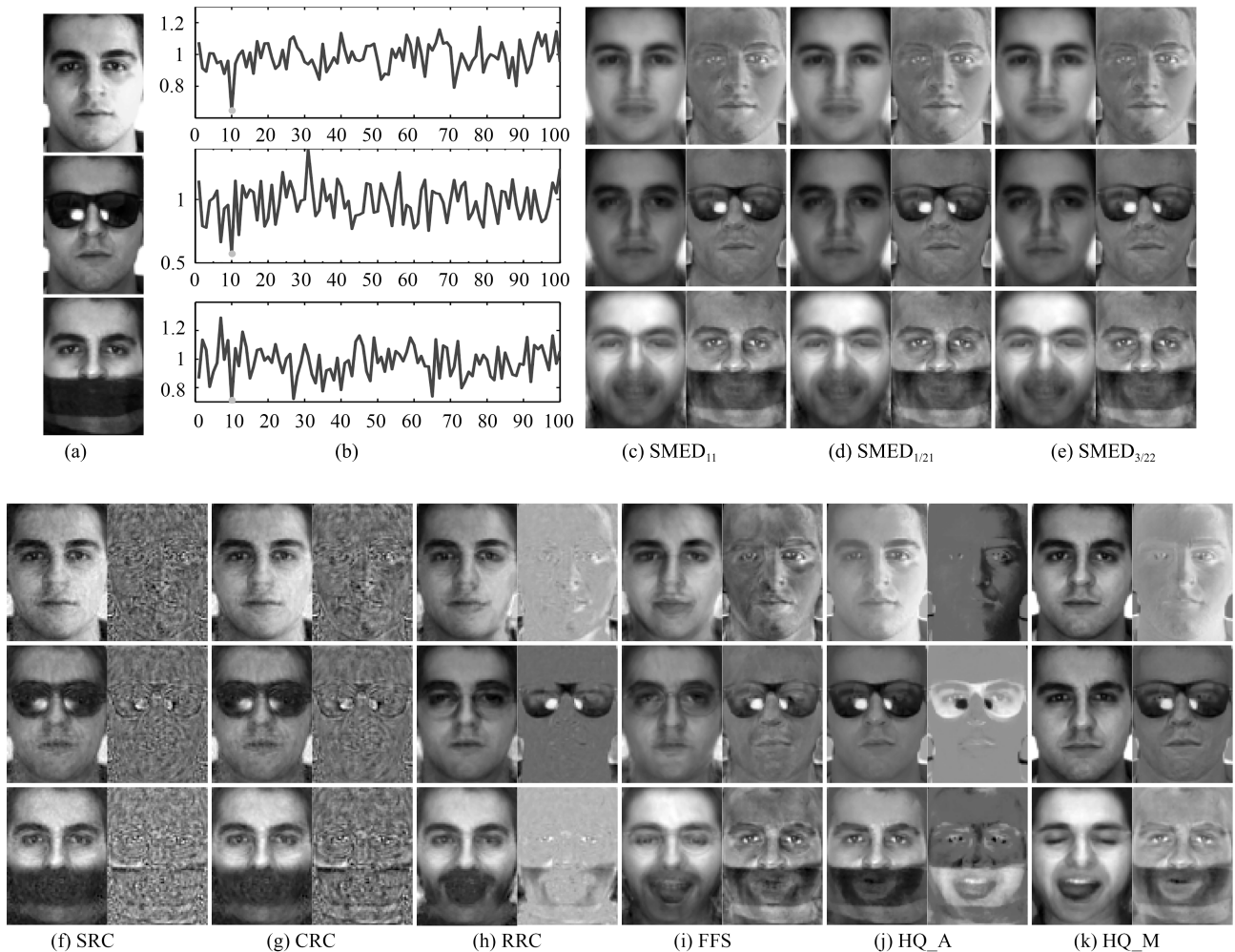


图 3 不同光照影响及墨镜和围巾遮挡下的人脸重构性能对比
 Fig. 3 The reconfiguration performance contrast of different algorithms under illumination changes and sunglasses or scarf occlusion

而言,从图 3(c)、图 3(d) 和图 3(e) 的比较中可以看出它们的重构效果非常接近,仅存有细微差别,如对于光照影响,SMED₁₁ 和 SMED_{3/22} 的处理效果优于 SMED_{1/21},具体表现在眼睛这块区域的重构;对于墨镜遮挡,SMED_{1/21} 重构的图像棱角更为明确;而对于围巾遮挡,虽然三者眼睛区域的重构都表现不佳,但是对围巾遮挡处嘴巴的修复却表现不错,基本还原了遮挡处的嘴型,而且 SMED_{1/21} 比 SMED₁₁ 和 SMED_{3/22} 修复将更清晰.总体来说,对于 AR 数据库而言,属于特例的 SMED₁₁ 对光照影响的处理较优,而 SMED_{1/21} 对大面积遮挡情况的处理表现突出.此外,对于对比算法,从图 3(f) 和图 3(g) 可得 SRC 和 CRC 不太适合遮挡类图像的修复工作,重构的人脸图像都带有明显的光照影响以及墨镜和围巾遮挡.而图 3(j) 基于半二次相加框架的 HQ_A 虽然它的重构效果优于 SRC 和 CRC,但是同样存在墨镜和围巾的痕迹,且光照影响去除不佳.图 3(h) 和图 3(i) 中的 RRC 和 FFS,前者引入了特征加权特性,后者通过快速提取特征重构图像,其重构人脸有一定程度的光照去除效果,且对于墨镜遮挡,其消除了墨镜效果,但存留明显的眼镜框架,围巾遮挡上 RRC 在眼睛区域的重构效果优于 FFS,但对嘴巴的修复 FFS 比 RRC 表现更佳.对于对比算法中表现最佳的基于半二次相乘框架的 HQ_M 来说,如图 3(k),虽然 HQ_M 重构图像的清晰度高于 SMED 算法且对光照影响的处理上 HQ_M 优于 SMED,但是对于遮挡图像,SMED 的修复较 HQ_M 更加接近于正面无遮挡无光照影响的原图像,特别是眼睛区域和嘴巴区域.

4.2 真实影响下的人脸识别

通过 AR 数据库进行真实影响下的人脸识别实验,训练样本和测试样本的选择与第 4.1 节一致.实验步骤为:首先采用训练样本和测试样本迭代求解模型中的最优重构编码系数 α^* ;然后通过 α^* 和测试样本数据得到重构数据;最后利用分类原理获得每类测试样本的整体识别率.本小节仍然选择第 4.1 节中的对比算法进行实验且参数设置保持一致.

4.2.1 光照及伪装影响

表 1 给出了所提算法和对比算法在真实影响下的识别率对比,其中粗体数据值即为同等条件下的最优识别率.通过表 1 可以发现:除在围巾遮挡的测试实验中 RRC 算法共享了 SMED_{1/21} 的最优识别率外,其他情况下 SMED 都具有最优的测试值.而对于 SMED 来说,SMED_{1/21} 在墨镜、围巾遮挡情形下的效果均优于 SMED₁₁ 和 SMED_{3/22},在光照影响下,SMED₁₁ 表现更突出.这一现象与第 4.1.1 节表现一致.值得注意的是, HQ_M 在墨镜遮挡情

形下效果较优却在围巾遮挡情形下效果较弱,而 CRC 则正处于相反的情况.总体而言,在光照影响的样本测试实验中,最优的 SMED₁₁ 较 SRC, CRC, RRC, FFS, HQ_A 和 HQ_M 分别提升了 5.33%, 0.50%, 7.66%, 12.83%, 8.00%, 2.32% 的识别率.在墨镜遮挡的样本测试实验中,最优的 SMED_{1/21} 较 SRC, CRC, RRC, FFS, HQ_A 和 HQ_M 分别提升了 8.34%, 7.50%, 6.01%, 11.67%, 7.17%, 1.34% 的识别率.在围巾遮挡的样本测试实验中,最优的 SMED_{1/21} 较 SRC, CRC, FFS, HQ_A 和 HQ_M 分别提升了 14.00%, 6.33%, 19.00%, 25.33%, 20.90% 的识别率.

表 1 AR 人脸库真实影响下不同算法的识别率对比 (%)
Table 1 Comparison of recognition rates on AR database (%)

算法	光照影响	墨镜遮挡	围巾遮挡
SMED ₁₁	95.83	74.66	67.64
SMED _{1/21}	91.93	77.17	69.83
SMED _{3/22}	93.16	72.67	60.33
SRC	90.50	68.83	55.83
CRC	95.33	69.67	63.50
RRC	88.17	71.16	69.83
FFS	83.00	65.50	50.83
HQ_A	87.83	70.00	44.50
HQ_M	93.51	75.83	48.93

4.2.2 像素污损影响

本小节对 AR 数据库进行更具挑战性的实验对比.通过对墨镜遮挡及围巾遮挡样本图像进行不同程度的随机像素污损来比较所提算法和对比算法的识别性能.图 4 为实验结果图,其中各子图上方显示的即为 0%~50% 像素污损程度下的测试样本图像.由图 4 可知,对比 SRC, CRC, FFS 和 HQ_A, SMED 表现突出,而其与 RRC 和 HQ_M 的识别率比较呈现竞争状态.总体来说,各算法对墨镜遮挡像素污损图像的识别率均高于围巾遮挡的像素污损图像,符合第 4.2.1 节的一般规律,且 SMED_{1/21} 在不同像素污损程度下较 SMED₁₁ 和 SMED_{3/22} 突出.此外,从图 4(a) 和图 4(b) 可得,随着污损程度的增加, CRC 比 SRC 呈现更快的下降趋势,这源于 CRC 对于数据间的协作关系描述不适用于处理这类稀疏型噪声.因此,对于像素污损这类稀疏型噪声, SRC 较 CRC 更具鲁棒性.

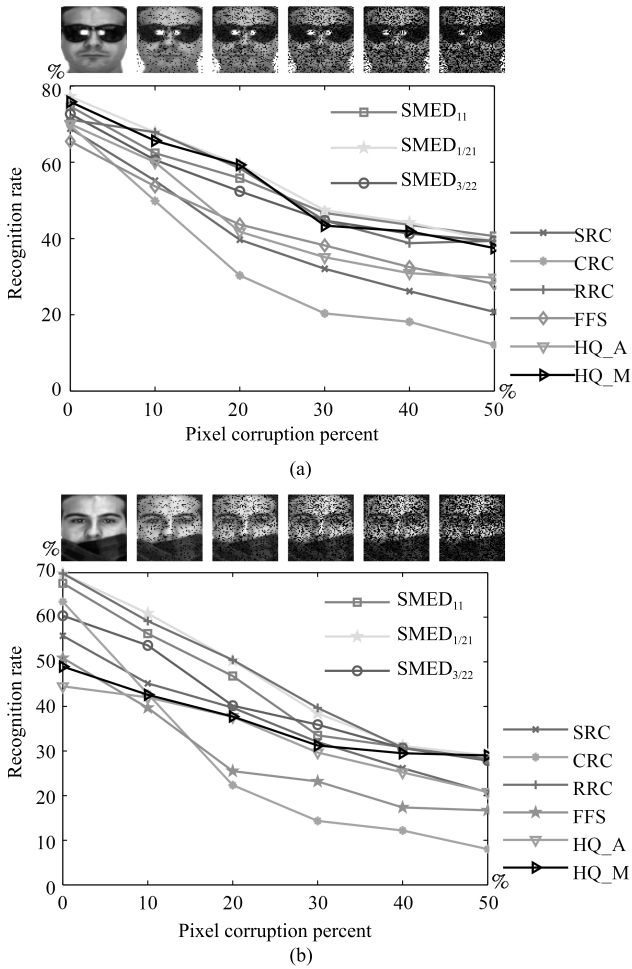


图 4 不同像素污损程度下的人脸识别率对比
 Fig. 4 Comparison of recognition rates under the different pixel corruption levels

4.3 随机遮挡下的人脸识别

在本节中, 采用 ExYaleB 数据库测试 SMED 在随机遮挡下的人脸识别应用中的算法性能. ExYaleB 包含 38 人每人 9 种姿态形式、64 种光照变化, 总共 21888 张人脸图像. 与文献 [13] 一致, 本文抽选其中接近正面, 与光照变化有关的 2414 幅图像用于实验. 每张图像大小都统一调整至 32 像素 \times 32 像素. 任意选取其中每个人物 20 幅图像共 760 张作为训练样本, 其余为测试样本. 对比算法选择在第 4.2.1 节表现不错的 CRC, RRC 和 HQ_M, 此外, 添加抗噪性较强的 CESR^[14] 算法. CRC, RRC 和 HQ_M 的实验参数与第 4.1 节一致, CESR 的参数设置参考文献 [14]. 此外, 本小节通过参数敏感度分析设置 SMED 算法中的参数 $\mu_c = 0.1, \rho = 1.1$.

实验中, 将测试样本做四种不同类型图像的随机遮挡处理: 1) 不同尺寸大小方形白色图像随机遮挡; 2) 不同尺寸大小方形黑色图像随机遮挡; 3) 不同尺寸大小方形 baboon 图像随机遮挡; 4) 不同尺

寸大小非方形鲜花图像随机遮挡. 遮挡图像的尺寸决定了测试样本图像被遮挡的程度. 本小节首先对四种经过遮挡处理的测试样本进行分类识别测验, 然后通过图 4 比较不同遮挡程度下所提算法与对比算法的识别率, 其中各子图上方显示的即为 0%~60% 遮挡程度下的测试样本图像.

由图 5 可知, 方形白色图像遮挡对各算法识别率的影响较大, 而非方形鲜花图像影响较小. 此外, 方形黑色及 baboon 图像对各算法识别率影响的程度相似. 具体来说, 图 5(a) 中 SMED₁₁ 表现最佳, 虽然在 20% 遮挡程度时 RRC 的识别率高于 SMED₁₁, 但当遮挡程度大于 30% 时, 其识别性能均优于 RRC, 特别在 60% 遮挡程度时, SMED₁₁ 较 RRC 的识别率提升了 4.3%. SMED_{1/21} 和 SMED_{3/22} 虽然没有 SMED₁₁ 表现突出, 但当遮挡大于 50% 时, 它们与 RRC 的识别率结果不分伯仲. 此外, 当遮挡低于 20% 时, HQ_M 表现不错, 识别率高于 SMED_{1/21} 和 SMED_{3/22}, 但随着遮挡程度的增加其优势不复存在; 同样地, 遮挡低于 30% 时, CESR 的识别率随遮挡程度的增加递减平稳, 基本保持在 85%~90%, 此后其识别率急剧下降, 特别当遮挡为 60% 时, 识别率低至 6.89%; 而对于 CRC, 随着遮挡程度的增加识别率就开始快速递减, 60% 遮挡时, 识别率为 21.58%, 高于 CESR. 综合图 5(a) 可得对方形白色图像遮挡表现最佳的 SMED₁₁ 算法在 60% 遮挡程度下比 CRC, RRC, CESR 和 HQ_M 分别提升了 22.56%, 4.3%, 37.25%, 16.81% 的识别率. 对于方形黑色图像遮挡, 图 5(b) 显示了各算法的识别率对比, 可以看出各算法对方形黑色遮挡的处理性能优于方形白色遮挡. 特别地, HQ_M 表现突出, 非常接近 SMED₁₁ 的识别结果, 其在 10%~30% 遮挡时, 识别率略优于 SMED₁₁, 40%~50% 时, 两者识别率基本一致, 到 60% 时, SMED₁₁ 显现出高于 HQ_M 的识别性能. 对于 SMED_{1/21} 和 SMED_{3/22}, 虽然它们在低遮挡程度下表现并不突出, 但随着遮挡程度的增加, 逐渐体现出抗遮挡的能力, 尤其在 60% 遮挡时, 识别结果高于表现不错的 SMED₁₁. 此外, RRC 并不像在图 5(a) 中表现的那么出色, 反而与 CRC, CESR 的识别效果类似, 表现平平. 总体而言, 对于方形黑色图像遮挡, 在 60% 遮挡程度下, 识别率最高的 SMED_{1/21} 比 CRC, RRC, CESR 和 HQ_M 分别提升了 19.52%, 12.92%, 14.56%, 10.22% 的识别率. 此外, 对于方形 baboon 图像及非方形鲜花图像遮挡, 图 5(c) 和图 5(d) 显示 SMED₁₁, SMED_{1/21} 和 SMED_{3/22} 均表现突出. 60% 遮挡下, 在图 5(c) 中识别性能最优的 SMED_{1/21} 比 CRC, RRC, CESR 和 HQ_M 分别提升了 33.12%, 17.95%, 28.02%,

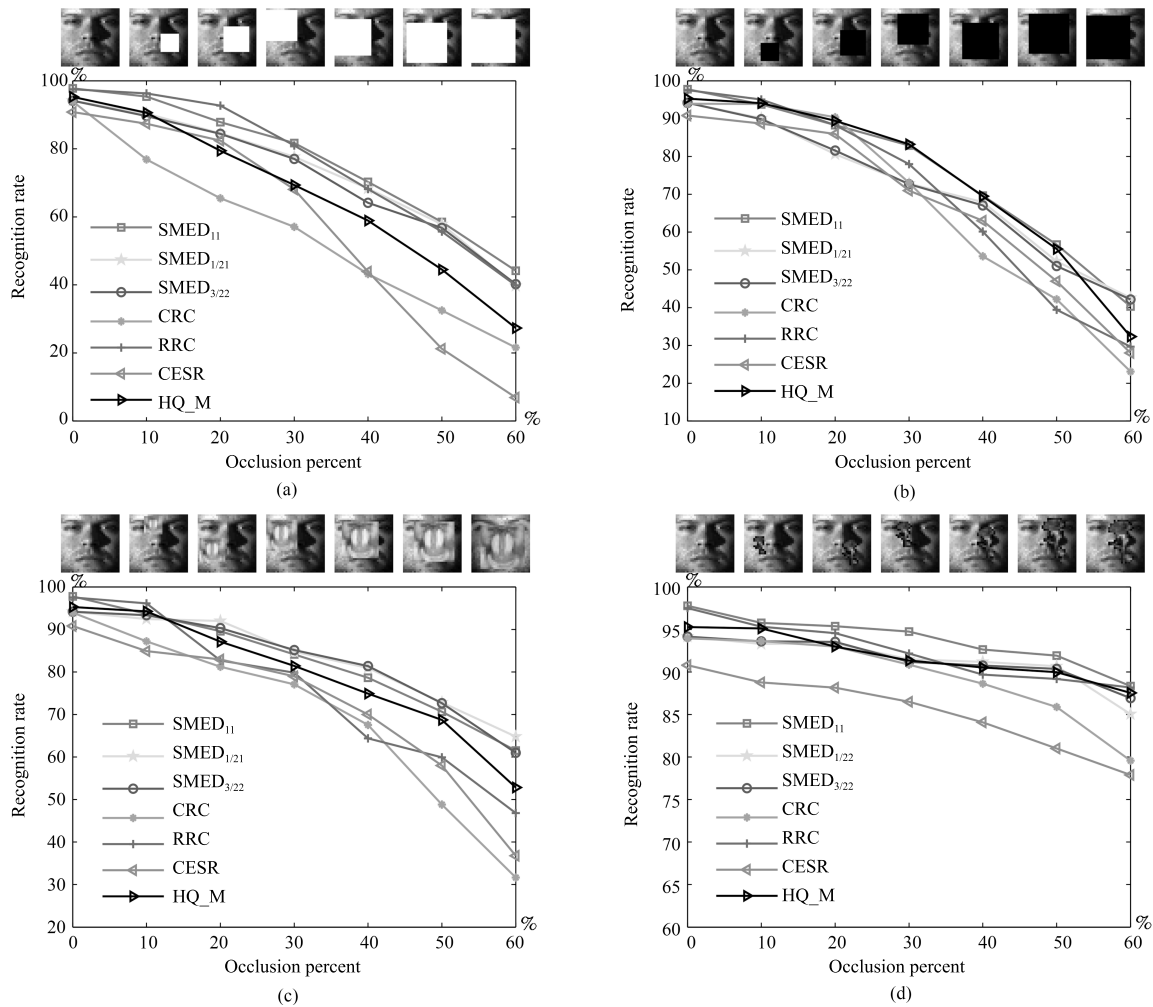


图 5 不同遮挡程度下的人脸识别率对比

Fig. 5 Comparison of recognition rates under the different occlusion levels

11.94% 的识别率. 图 5 (d) 中表现最佳的 SMED₁₁ 比 CRC, RRC, CESR 和 HQ_M 分别提升了 8.76%, 0.15%, 10.44%, 0.79% 的识别率.

4.4 非限制场景下的人脸识别

与第 4.3 节实验环境一致, 本节采用 PubFig 数据库测试 SMED 在非限制场景下人脸识别应用中的算法性能, 对比算法仍然选取 CRC, RRC 和 HQ_M. PubFig 是一个大规模、非限制场景下的真实人脸数据库, 包含从网络搜集来的 200 人共 58 797 幅彩色图像. 图像可能受强光光照、自由姿态、障碍物遮挡、配饰遮挡、夸张表情等多种复杂因素的干扰. 本文抽选 100 人, 每人 20 张, 共 2 000 张图像用于实验. 每张图像都通过统一的人脸部位裁剪并调整至 64 像素 × 64 像素的灰度图像. 选取其中每个人前 10 张共 1 000 张图像作为训练样本, 其余为测试样本. 图 6 为其中 1 人的训练样本和测试样本示例, 图 7 为本节实验的测试结果.



图 6 PubFig 中 1 人的训练样本和测试样本示例图
Fig. 6 Sample images for one person from PubFig face database

图 7 中 SMED1 代表 SMED₁₁ 算法, SMED2 代表 SMED_{1/21} 算法, SMED3 代表 SMED_{3/22} 算法. 由图 7 可知, CESR 对 PubFig 数据库的识别效果最差, 这是因为它忽视了图像中噪声的空间结构, 对比来说, SMED 很好地利用了空间结构所提供的信息, 其中, SMED₁₁ 取得了最优的识别率, 即对于非限制场景下的人脸图像来说, 属于特例的凸函数模型求解表现突出. 此外, 由于 PubFig 中包含的是非限制场景下的人脸图像, 这使得抗噪性强的 RRC 算法因为训练样本数据的不稳定性, 不能有效揭示

测试样本数据间的差异而影响识别性能, 其识别结果仅略高于 CRC. 而对于 HQ_M, 它取得了 44.3% 的有竞争力的识别结果, 略高于 SMED_{1/21} 算法. 综上所述, SMED 由于不需要过多地依赖训练样本数据之间的变化, 而是通过考虑图像中噪声的内部结构, 采用切合实际的分布来描述误差矩阵, 故而能获得相对更佳的识别性能.

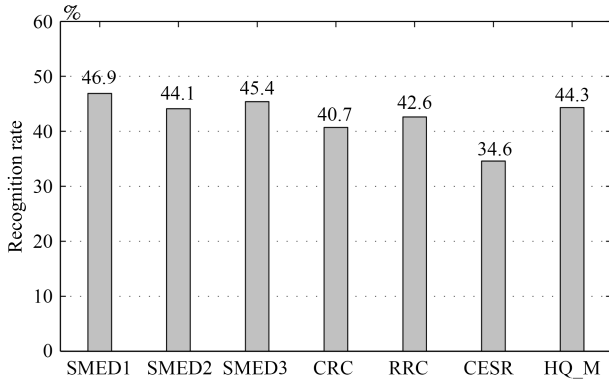


图 7 PubFig 人脸库中不同算法的识别率对比
Fig. 7 Comparison of recognition rates on PubFig database

4.5 运行效率分析

除识别性能外, 运行效率是影响分类算法实际应用能力的另一关键指标. 本节对 AR、ExYaleB 和 PubFig 数据库中的几种算法进行测试效率对比. 表 2 和表 3 显示了各算法在不同数据库中单个样本的测试时间.

表 2 为第 4.2.1 节实验状态下的测试时间对比; 表 3 中 ExYaleB 数据库里白色、黑色、baboon 和鲜花遮挡的测试时间均在 60% 遮挡程度下获得. 此外, 所有数据都通过 10 次实验运行并取平均值得到.

表 2 AR 数据库中各算法测试时间对比 (s)
Table 2 Running time of competing algorithms on AR database (s)

分类器	测试时间
SMED ₁₁	0.1703
SMED _{1/21}	0.1748
SMED _{3/22}	0.1648
SRC	0.0917
CRC	0.0041
RRC	3.2165
FFS	0.1738
HQ_A	4.3898
HQ_M	13.2725

结合表 1 和表 2 可知, CRC 在识别率上虽然与 SRC、FFS 以及 HQ_A 较为接近, 但是其运行效率却远远优于这些算法, 达到了 ms 级的反应速度. HQ_M 以运行效率减弱为代价提升鉴别性能, 从表 2 可见其运行时间为 CRC 的 3000 多倍, 因此实用性欠佳. RRC 在光照影响、墨镜遮挡和围巾遮挡这三种情形下的识别率表现突出, 接近 SMED 的识别性能, 但其在实现过程中需要迭代计算操作, 因此运行率较低. 最后, 在 AR 数据库中 SMED 的运行效率高于 RRC、FFS、HQ_A 和 HQ_M, 仅次于 CRC 和 SRC, 但其识别率却远高于 CRC 和 SRC. 此外, 结合表 3、图 5、图 7 中各算法在 ExYaleB、PubFig 数据库中的表现可知, 虽然 HQ_M 的鉴别性能在大多数情况下接近于 SMED 算法, 但在大部分数据集都需要消耗更多的运行时间, 不适宜于实际的分类应用. 而抗噪性不错的 RRC 算法的运行时间却随数据集的不同浮动较大, 稳定性不高. 综合而言, SMED 由于采用迭代加权最小二乘法进行模型求解, 能在很少的迭代次数下获得算法收敛, 因此其综

表 3 ExYaleB、PubFig 数据库中各算法测试时间对比 (s)
Table 3 Running time of competing algorithms on ExYaleB, PubFig database (s)

分类器	测试时间				PubFig
	ExYaleB (白色遮挡)	ExYaleB (黑色遮挡)	ExYaleB (baboon 遮挡)	ExYaleB (鲜花遮挡)	
SMED ₁₁	0.2429	0.2508	0.2191	0.2376	1.6705
SMED _{1/21}	0.3450	0.2452	0.2561	0.2445	1.7864
SMED _{3/22}	0.2638	0.2435	0.2318	0.2186	1.6425
CRC	0.0037	0.0045	0.0031	0.0035	0.0152
RRC	0.2510	0.2370	0.2426	0.2375	2.5819
CESR	0.3244	0.1276	0.2966	0.2240	2.3499
HQ_M	12.2049	11.7532	12.3732	11.6457	6.8705

合运行效率领先于除 CRC 和 SRC 的其他对比算法. 而且鉴于 SMED 具有较高较稳定的识别性能, 故而值得推广应用.

5 结束语

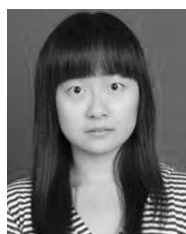
本文提出一种联合平滑矩阵多变量椭圆分布的稀疏表示算法, 该算法强调误差矩阵中各个像素间的依赖性并假定误差矩阵作为一个随机矩阵变量服从于矩阵多变量椭圆分布. 由于通过上述假定构建的目标函数具有非光滑性, 不利于问题的优化求解, 故先引入辅助变量光滑模型, 而后采用迭代加权最小二乘法优化求解模型. 此外, 文中对 SMED 算法的收敛性和复杂度进行了理论分析, 并讨论了模型的参数敏感性. 以人脸数据为例, 采用 AR、ExYaleB 和 PubFig 数据库验证了所提算法的性能优于经典算法. 综上所述, 本文的创新点包括: 1) 采用矩阵多变量椭圆分布描述误差矩阵, 增强算法抗噪性; 2) 引入辅助变量光滑模型, 易于获得全局最优解, 提升算法识别能力; 3) 采用迭代加权最小二乘法优化求解模型, 加快算法收敛速度.

分析研究发现, 所提算法 SMED 虽然易于实现且性能卓越, 但运行效率上仍受到限制, 无法满足视频跟踪等在线应用系统的需求, 且其与目前最先进算法的识别率仍有一定差距, 后续工作有待进一步提升算法的识别精度. 此外, SMED 是否适用于更加复杂的噪声特征样本和遮挡环境, 且能否将该算法扩展应用于一般的现实噪声^[46], 也是后续有待进一步研究探索的问题.

References

- Jalali S, Poor H V. Universal compressed sensing for almost lossless recovery. *IEEE Transactions on Information Theory*, 2017, **63**(5): 2933–2953
- Hyder M, Mahata K. An approximate L_0 norm minimization algorithm for compressed sensing. In: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing. Taipei, China: IEEE, 2009. 3365–3368
- Wright J, Yang A Y, Ganesh A, Sastry S S, Ma Y. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(2): 210–227
- Yang J, Zhang L, Xu Y, Yang J Y. Beyond sparsity: the role of L_1 -optimizer in pattern classification. *Pattern Recognition*, 2012, **45**(3): 1104–1118
- Zhang L, Yang M, Feng X C. Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011. 471–478
- Zhang H M, Yang J, Xie J C, Qian J J, Zhang B. Weighted sparse coding regularized nonconvex matrix regression for robust face recognition. *Information Sciences*, 2017, **394–395**: 1–17
- Yang G, Wang Y, Guo L. A sparser reduced set density estimator by introducing weighted l_1 penalty term. *Pattern Recognition Letters*, 2015, **58**: 15–22
- Liu C L, Hsaio W H, Xiao B, Chen C Y, Wu W L. Maximum-margin sparse coding. *Neurocomputing*, 2017, **238**: 340–350
- Lu C W, Shi J P, Jia J Y. Online robust dictionary learning. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013. 415–422
- Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(11): 2106–2112
- Naseem I, Togneri R, Bennamoun M. Robust regression for face recognition. *Pattern Recognition*, 2012, **45**(1): 104–118
- Yang M, Zhang L, Yang J, Zhang D. Robust sparse coding for face recognition. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 625–632
- Yang M, Zhang L, Yang J, Zhang D. Regularized robust coding for face recognition. *IEEE Transactions on Image Processing*, 2013, **22**(5): 1753–1766
- He R, Zheng W S, Hu B G. Maximum correntropy criterion for robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(8): 1561–1576
- He R, Zheng W S, Tan T N, Sun Z N. Half-quadratic-based iterative minimization for robust sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(2): 261–275
- He R, Zheng W S, Hu B G, Kong X W. Two-stage nonnegative sparse representation for large-scale face recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, **24**(1): 35–46
- Kibria B M G, Haq M S. The multivariate linear model with matrix-T error variables. *Journal of Applied Statistical Sciences*, 2000, **9**(4): 277–290
- Basu S, Micchelli C A, Olsen P. Power exponential densities for the training and classification of acoustic feature vectors in speech recognition. *Journal of Computational and Graphical Statistics*, 2001, **10**(1): 158–184
- Liu M H. Multivariate Nonnormal Regression Models, Information Complexity, and Genetic Algorithms: A Three Way Hybrid for Intelligent Data Mining [Ph.D. dissertation]. University of Tennessee, USA, 2006.
- Balakrishnan N, Hashorva E. Scale mixtures of Kotz dirichlet distributions. *Journal of Multivariate Analysis*, 2013, **113**: 48–58
- Abdous B, Genest C, Rémillard B. Dependence properties of meta-elliptical distributions. *Statistical Modeling and Analysis for Complex Data Problems*. Boston, MA, USA: Springer, 2005. 1–15
- Long C, You D H, Hu J, Wang G, Dong M L. Quick and effective multiple contingency screening algorithm based on long-tailed distribution. *IET Generation, Transmission and Distribution*, 2016, **10**(1): 257–262
- Chong C. Stochastic PDEs with heavy-tailed noise. *Stochastic Processes and Their Applications*, 2017, **127**(7): 2262–2280
- Sun Y, Babu P, Palomar D P. Robust estimation of structured covariance matrix for heavy-tailed elliptical distributions. *IEEE Transactions on Signal Processing*, 2016, **64**(14): 3576–3590

- 25 Liu M H, Bozdogan H. Multivariate regression models with power exponential random errors and subset selection using genetic algorithms with information complexity. *European Journal of Pure and Applied Mathematics*, 2008, **1**(1): 4–37
- 26 Sun Y, Babu P, Palomar D P. Robust estimation of structured covariance matrix for heavy-tailed distributions. In: Proceedings of the 2015 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Brisbane, QLD, Australia: IEEE, 2015. 5693–5697
- 27 Fang K T, Anderson T W. *Statistical Inference in Elliptically Contoured and Related Distributions*. New York, USA: Allerton Press, 1990.
- 28 Zhao T, Liu H. Calibrated precision matrix estimation for high-dimensional elliptical distributions. *IEEE Transactions on Information Theory*, 2014, **60**(12): 7874–7887
- 29 Díaz-García J A, Gutiérrez-Jáimez R. Compound and scale mixture of matrix-variate and matrix-variate Kotz-type distributions. *Journal of the Korean Statistical Society*, 2010, **39**(1): 75–82
- 30 Wai P S, Kun S S, Ismail M T, Ariffin S, Karim A. Model performance between linear vector autoregressive and Markov switching vector autoregressive models on modelling structural change in time series data. In: Proceedings of the 2015 International Symposium on Mathematical Sciences and Computing Research (iSMSC). Ipon, Malaysia: IEEE, 2015. 379–383
- 31 Hsu D, Kakade S M, Zhang T. Robust matrix decomposition with sparse corruptions. *IEEE Transactions on Information Theory*, 2011, **57**(11): 7221–7234
- 32 Lai M J, Xu Y Y, Yin W T. Improved iteratively reweighted least squares for unconstrained smoothed ℓ_q minimization. *SIAM Journal on Numerical Analysis*, 2013, **51**(2): 927–957
- 33 Li W Y. Visualizing network communities with a semi-definite programming method. *Information Sciences*, 2015, **321**: 1–13
- 34 Lin Q H, Xiao L. An adaptive accelerated proximal gradient method and its homotopy continuation for sparse optimization. *Computational Optimization and Applications*, 2015, **60**(3): 633–674
- 35 Goldstein T, O’Donoghue B, Setzer S, Baraniuk R. Fast alternating direction optimization methods. *SIAM Journal on Imaging Sciences*, 2014, **7**(3): 1588–1623
- 36 Xu Kai, Wu Xiao-Jun, Yin He-Feng. Distributed low rank representation-based subspace clustering algorithm. *Journal of Computer Research and Development*, 2016, **53**(7): 1605–1611
(许凯, 吴小俊, 尹贺峰. 基于分布式低秩表示的子空间聚类算法. 计算机研究与发展, 2016, **53**(7): 1605–1611)
- 37 Barrah E M, Safi S, Malaoui A. New technique for face recognition based on singular value decomposition (SVD). In: Proceedings of the 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech). Marrakech, Morocco: IEEE, 2016. 96–99
- 38 Chen K, Lv Q, Lu Y, Dou Y. Robust regularized extreme learning machine for regression using iteratively reweighted least squares. *Neurocomputing*, 2017, **230**: 345–358
- 39 Li Bo, Lu Chun-Yuan, Leng Cheng-Cai, Jin Lian-Bao. Robust low rank subspace clustering based on local graph Laplace constraint. *Acta Automatica Sinica*, 2015, **41**(11): 1971–1980
(李波, 卢春园, 冷成财, 金连宝. 基于局部图拉普拉斯约束的鲁棒低秩表示聚类方法. 自动化学报, 2015, **41**(11): 1971–1980)
- 40 Benner P, Li R C, Truhar N. On the ADI method for Sylvester equations. *Journal of Computational and Applied Mathematics*, 2009, **233**(4): 1035–1045
- 41 Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2009.
- 42 Marshall A W, Olkin I, Arnold B C. *Inequalities: Theory of Majorization and Its Applications* (2nd edition). New York: Springer, 2011.
- 43 Lee K C, Ho J, Kriegman D J. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(5): 684–698
- 44 Kumar N, Berg A C, Belhumeur P N, Nayar S K. Attribute and simile classifiers for face verification. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 365–372
- 45 Wu J X, Brubaker S C, Mullin M D, Rehg J M. Fast asymmetric learning for cascade face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(3): 369–382
- 46 Gu J, Hu H, Li H. Local robust sparse representation for face recognition with single sample per person. *IEEE/CAA Journal of Automatica Sinica*, 2018, **5**(2): 547–554



邱虹 浙江万里学院大数据与软件工程学院讲师。主要研究方向为模式识别, 图像处理。

E-mail: qianzhihe.17@163.com

(**QIU Hong** Lecturer at the College of Big Data and Software Engineering, Zhejiang Wanli University. Her research interest covers pattern recognition and image processing.)



王万良 浙江工业大学计算机科学与技术学院教授。主要研究方向为人工智能, 模式识别。本文通信作者。

E-mail: zjutwwl@zjut.edu.cn

(**WANG Wan-Liang** Professor at the College of Computer Science and Technology, Zhejiang University of Technology. His research interest covers

artificial intelligence and pattern recognition. Corresponding author of this paper.)



郑建炜 浙江工业大学计算机科学与技术学院副教授。主要研究方向为机器学习, 模式识别。E-mail: zjw@zjut.edu.cn

(**ZHENG Jian-Wei** Associate professor at the College of Computer Science and Technology, Zhejiang University of Technology. His research interest covers machine learning and pattern

recognition.)