

特征加权组稀疏判别投影分析算法

郑建炜¹ 黄琼芳¹ 杨平¹ 王万良¹ 马文龙²

摘要 近来,稀疏表示分类算法已经在模式识别和特征提取领域获得了广泛的关注.受最近提出的稀疏表示判别投影算法启发,本文提出了一种新的特征加权组稀疏判别投影算法 (Feature weighted group sparse classification steered discriminative projection, FWGSDP).首先,提出特征加权组稀疏分类算法 (Feature weighted group sparsebased classification, FWGSC) 进行稀疏系数编码,该算法采用带特征加权约束的保局性信息,能够鲁棒地重构给定的输入数据;其次,通过类内重构散度最小、类间重构散度最大为目标计算最优投影判别矩阵,使得输入数据具有最佳的模式分类效果;最后,提出迭代重约束稀疏编码方法并结合特征分解操作进行 FWGSDP 模型高效求解.在 ExYaleB, PIE 和 AR 三个人脸数据库的实验验证了所提算法在普通数据和带噪数据中的分类效果都优于现存的算法.

关键词 稀疏表示,保局性,组稀疏,判别投影

引用格式 郑建炜,黄琼芳,杨平,王万良,马文龙.特征加权组稀疏判别投影分析算法.自动化学报,2016,42(5):746-759

DOI 10.16383/j.aas.2016.c150364

Feature Weighted Group Sparse Discriminative Projection Algorithm

ZHENG Jian-Wei¹ HUANG Qiong-Fang¹ YANG Ping¹ WANG Wan-Liang¹ MA Wen-Long²

Abstract Recently, sparse representation classification (SRC) has attracted more and more attention in pattern recognition and feature extraction. Motivated by the recent developed SRC steered discriminative projection algorithm, a new feature weighted group sparse discriminative projection algorithm (FWGSDP) is proposed in this paper. First, sparse coefficients are produced by a new proposed feature weighted group sparse representation classification algorithm (FWGSC), which can robustly regress a given signal with regularized regression coefficients by introducing the feature weighted locality structure of the data. Second, FWGSDP maximizes the subtraction of inter-class reconstruction residual and intra-class reconstruction residual, and thus enables data-in to achieve better separation. Finally, a sequentially iteratively re-restrained sparse coding and eigen-decomposition strategy is developed to solve the FWGSDP model efficiently. Experimental results on the ExYaleB, the PIE, and the AR database demonstrate that the proposed algorithm is more effective than other feature extraction methods.

Key words Sparse representation, locality-constraint, group sparse, discriminative projection

Citation Zheng Jian-Wei, Huang Qiong-Fang, Yang Ping, Wang Wan-Liang, Ma Wen-Long. Feature weighted group sparse discriminative projection algorithm. *Acta Automatica Sinica*, 2016, 42(5): 746-759

近十年,压缩感知和稀疏表示技术已成为模式识别和机器视觉领域最热门的研究方向之一.在大数据应用快速发展的今天,样本的高维性和冗余性已成为困扰研究者的关键科学难题.大量相关的探索性算法正在不断涌现.2009年,Wright等提出的基于稀疏表示的分类器 (Sparse representation

based classifier, SRC)^[1]成功应用于计算机视觉领域,面对高维以及变化丰富的图像数据, SRC 算法采用全体训练样本作为超完备字典,以最小化 l_1 范数为目标求解测试样本在完备字典上的稀疏表示系数矢量,然后通过各类稀疏系数矢量重构测试样本并求解其误差,根据重构误差的大小对测试样本进行分类.以 SRC 为基础, Xu 等^[2]和 He 等^[3]分别提出两阶段操作策略进行鲁棒的大数据量识别. Yang 等^[4]则提出结合稀疏表示与 Fisher 判别思想进行字典信息更新.上述 SRC 型算法通过稀疏性质的引入,最大程度地保留了测试样本的全局信息,在人脸识别^[5-6]、信号去噪^[7]、时间序列分类^[8]等应用中获得了广泛的关注.然而, Zhang 等^[9-10]通过理论分析和实验验证表明 SRC 型算法的成功并非源自于其稀疏特性,而是通过样本间的协作关系获得优秀的分类性能,并提出了协作表示分类器 (Collaborative representation classifier, CRC),以

收稿日期 2015-06-23 录用日期 2015-12-22
Manuscript received June 23, 2015; accepted December 22, 2015

国家自然科学基金 (61502424, 61379123), 浙江省自然科学基金 (LY15E050007, LY15F030014, LQ14F030003) 资助

Supported by National Natural Science Foundation of China (61502424, 61379123) and Zhejiang Provincial Natural Science Foundation (LY15E050007, LY15F030014, LQ14F030003)

本文责任编辑 杨健

Recommended by Associate Editor YANG Jian

1. 浙江工业大学计算机科学与技术学院 杭州 310023 2. 衢州职业技术学院信息工程学院 衢州 324000

1. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023 2. School of Information Engineering, Quzhou College of Technology, Quzhou 324000

l_2 范数替换 l_1 范数, 在不降低识别率的前提下获得了运行效率的大幅度提升.

SRC 和 CRC 算法在依据全体训练数据作为测试字典进行查询分类的过程中忽视了样本的类别标签信息, 其判别性能有待进一步提升. 针对此问题, Majumdar 等^[11] 在 SRC 和 CRC 的基础上提出了组稀疏分类算法 (Group sparse classification, GSC), 将训练样本根据类别信息进行分组, 假设测试样本由多个子空间组合近似表示的情况下, GSC 以最小化 l_{12} 组合范数为目标选择最少的子空间组数表达测试样本. 类似地, Elhamifar 等^[12] 提出了基于结构化稀疏表达的人脸识别算法, 其核心思想是从超完备字典中寻找最少组数来表达测试样本数据. GSC 的识别性能优于常规的 SRC 和 CRC 算法, 鉴于此优势, 已有大量基于组稀疏表示的改进型和应用型算法不断涌现^[13-15].

GSC、SRC 和 CRC 等算法都通过正则项范数的调整进行系数计算, 以全局样本表示的方式实现待测数据判定, 但没有添加局部特性约束. 基于此, Lu 等^[16] 和 Timofte 等^[17-18] 分别将测试样本和训练数据间的欧氏距离以权值形式添加至 l_1 与 l_2 范数约束上, 提出加权稀疏表示分类器 (Weighted sparse representation based classification, WSRC) 和加权协作表示分类器 (Weighted collaborative representation based classification, WCRC). 类似地, Chao 等^[19] 提出一种添加局部约束的组稀疏算法 (Locality-constrained group sparse representation, LGSR), 通过结合组稀疏惩罚项和数据局部约束, 增加算法局部特性. 然而, 局部约束破坏了组稀疏结构, 意味着组间差异效果的缺失. 因此, Tang 等^[20] 通过重构因素和局部约束联合加权的形式提出了加权组稀疏分类器 (Weighted group sparse classifier, WGSC), 获得了更好的结构化信息和鉴别效果. 上述加权算法都通过样本的整体特征进行权值系数计算, Yang 等^[21] 则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), 具有更高的抗噪能力, 却丢失了样本的局部分布特性引导.

稀疏表示型算法在模式分类中获得了广泛的应用, 但是在面向高维输入数据时, 其运行过程存在计算量过大的问题. 维数约减 (包括特征提取和特征选择)^[22-25] 成为其关键的预处理步骤. 典型的降维技术包括主成分分析法 (Principal component analysis, PCA)^[22] 和线性判别分析法 (Linear discriminative analysis, LDA)^[23] 以及各自的改进版本^[26-27], 两者都从全局散度出发, 分别通过无监督和有监督的形式构建投影目标函数. 对应地, He

等^[28] 提出局部保持投影算法 (Locality preserving projection, LPP), 从样本局部结构出发, 引入明确的变换矩阵, 以近邻关系保持为准则构建目标函数, 挖掘数据的本质子流形. 此外, 局部 Fisher 判别分析 (Local Fisher discriminant analysis, LFDA)^[29] 则兼顾了 LPP 和 LDA 的优势, 在 LDA 的散度矩阵中添加局部特性, 不仅增加了 LDA 算法的效果, 而且弥补了 LDA 的类内散度矩阵秩约束缺陷.

上述经典维数约减算法可以归结至图嵌入框架^[30], 其关键步骤是输入样本的几何邻域构建, 常见方法包括 k 近邻选择和 ε 半径选择, 算法性能受参数 k 和 ε 的干扰较大. 鉴于 SRC 型算法具有优越的编码能力和重构特性, Qiao 等^[31] 和 Cheng 等^[32] 分别开发了稀疏保持投影算法 (Sparsity preserving projections, SPP) 用于维数约减, SPP 采用 SRC 算法构建自适应的邻域表示系数, 免除了参数调整对算法性能的影响困扰. 受 SPP 的启发, 部分研究者将稀疏系数保持思想引入到经典 LPP^[33] 和 LFDA^[34], 提出了兼顾保局性和稀疏性的降维投影算法. 此外, Ly 等^[35] 结合 SPP 与无向有权图思想, 并引入类别标签, 提出了有监督的稀疏图判别分析算法, 具有更好的鉴别性能. Yang 等^[36] 和 Lu 等^[37] 分别提出了稀疏判别投影分类算法 (Sparse representation classifier steered discriminative projection, SRCDP) 和基于最佳投影的稀疏表示分类器, 以同类重构散度最小、异类重构散度最大为目标函数, 在保留 SRC 分类规则的基础上进行降维, 获得了理想的识别效果. 受 l_1 范数和迭代操作影响, SRCDP 算法的计算耗时较大, 近邻图构建效率低下. 鉴于此, 常见的解决方法是采用 l_2 范数^[38-39] 或者最小二乘法^[40] 替换 l_1 图构建, 以识别率降低为代价提升降维效率.

受 SRCDP 和 WGSC 等算法的启发, 本文提出特征加权组稀疏判别投影分析算法 (Feature weighted group sparse classification steered discriminative projection, FWGSDP). 首先, 针对 WGSC 和 RRC 存在的问题, 提出特征加权组稀疏分类器 (Feature weighted group sparsebased classification, FWGSC), 结合各类重构冗余以及样本距离测度逼近样本分布结构, 并兼顾特征加权因子进行奇异点剔除, 从样本和特征两方面减少重构表示误差; 其次, 以 FWGSC 为基础进行编码系数求解, 并计算特征加权约束的类内重构散度矩阵和类间重构散度矩阵; 最后, 依据类 Fisher 准则构建目标函数获得最优投影矩阵, 提升模式分类性能.

本文后续内容结构安排如下: 第 1 节描述了相关工作, 包括 SRC、GSC、WGSC 和 SRCDP. 第 2 节给出了特征加权组稀疏分类器的算法思想和求

解策略. 第 3 节在 FWGSC 的基础上进行了邻域构建, 并通过散度约束目标函数进行投影矩阵计算, 描述了完整的 FWGSDP 算法. 第 4 节分别通过有遮挡和无遮挡的人脸识别环境验证了 FWGSC 和 FWGSDP 的性能. 最后第 5 节为全文总结.

1 相关工作

本节主要回顾以 SRC 为基础的相关稀疏表示型分类及投影算法, 在此之前, 给出所需的数学符号定义. 设存在 c 类目标数据, 训练样本表示为 $X = [X_1, \dots, X_c] \in \mathbf{R}^{m \times n}$, 其中 $X_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{in_i}] \in \mathbf{R}^{m \times n_i}$ 是第 i 类数据的训练样本子集, $\mathbf{x}_{ij} \in \mathbf{R}^m$ 表示第 i 类目标的第 j 个样本, n_i 为第 i 类训练样本数, 即 $\sum_{i=1}^c n_i = n$ 为训练样本总数, $\mathbf{y} \in \mathbf{R}^m$ 为测试样本.

1.1 稀疏表示分类器

稀疏表示法具有明确的解释性, 其基本思想是假设各类输入样本具有过完备性^[1], 则第 i 类测试样本 \mathbf{y} 可通过该类训练样本进行稀疏线性表示:

$$\mathbf{y} = \mathbf{x}_{i1}\theta_{i1} + \mathbf{x}_{i2}\theta_{i2} + \dots + \mathbf{x}_{in_i}\theta_{in_i} = X_i\theta_i \quad (1)$$

其中, $\theta_i = [\theta_{i1}; \theta_{i2}; \dots; \theta_{in_i}] \in \mathbf{R}_i^n$ 是 X_i 对应的稀疏编码系数 (分号分隔符表明元素依列展开). 由于测试样本 \mathbf{y} 的所属类别 i 是未知的. 因此, SRC 采用完整输入数据作为表示样本, 即 \mathbf{y} 由 X 表示为

$$\mathbf{y} = X_1\theta_1 + X_2\theta_2 + \dots + X_c\theta_c = X\theta \quad (2)$$

依据基本思想假设, 式 (2) 中除第 i 类元素外, 系数矢量 $\theta \in \mathbf{R}^n$ 余下的元素都为 0^[1]. 因此, 计算矢量 θ 的过程是一个稀疏逼近的过程, 通过最小化重构误差以及 l_1 范数标准化求解, 即

$$\min_{\theta} \|\mathbf{y} - X\theta\|_2^2 + \lambda \|\theta\|_1 \quad (3)$$

其中, 参数 λ 用于平衡重构误差和系数值之间的贡献度. 得到系数 θ 后, 计算测试样本由各类样本重构的冗余值, 并将 \mathbf{y} 归类为最小重构值的类别标签, 即

$$k = \arg \min_i \|\mathbf{y} - X\delta_i(\theta)\|_2, \quad i = 1, \dots, c \quad (4)$$

其中, $\delta_i(\theta)$ 是一个选择算子, 表示将 θ 中除第 i 类外元素设为 0 值的矢量.

1.2 组稀疏分类器

针对模式分类问题, SRC 的一个主要缺点是在求解 l_1 范数最小化的过程中忽略了对训练样本类别标签的考虑. 因此, 在面向复杂高维输入样本时 (完备性缺失), SRC 很有可能存在通过异类数据表示测试样本的情形, 影响其分类效果. 组稀疏分类器尝试

从类组级别构建稀疏系数 θ , 意味着通过 GSC 所计算得到的系数 θ 中, 部分数据类组所对应的系数值全为 0, 而特定的类组系数则为非 0 值^[11]. GSC 采用 l_{12} 组合范数进行目标系数计算, 即

$$\min_{\theta} \sum_{i=1}^c \|\theta_i\|_2 \quad \text{s.t.} \quad \mathbf{y} = X\theta \quad (5)$$

式 (5) 等价于最小化重构误差及 l_{12} 组合范数标准化的解

$$\min_{\theta} \|\mathbf{y} - X\theta\|_2^2 + \lambda \sum_{i=1}^c \|\theta_i\|_2 \quad (6)$$

与 SRC 算法相比, GSC 利用训练样本的类别信息增强组稀疏性. 从式 (6) 中可以看出测试样本 \mathbf{y} 被尽可能少的训练样本类重构, 具有明确的类别指示性.

1.3 加权组稀疏分类算法

受到局部结构和组稀疏思想的启发, Tang 等^[20] 提出加权组稀疏分类算法 WGSC. WGSC 保持数据局部性的同时结合组稀疏优势, 目标是找到一个更优的稀疏表示策略. 对于任意测试样本 \mathbf{y} , WGSC 通过距离加权 \mathbf{d} 和重构冗余加权 \mathbf{r} 的形式将其表示为训练样本按类线性组合, 即

$$\min_{\theta} \|\mathbf{y} - X\theta\|_2^2 + \lambda \sum_{i=1}^c r_i \|\mathbf{d}_i \odot \theta_i\|_2 \quad (7)$$

其中, \odot 代表矢量 \mathbf{d}_i 和 θ_i 依元素相乘. 从式 (7) 可见, WGSC 采用 \mathbf{r} 评估各类样本表示测试数据的重要程度, 其计算形式为

$$r_i = \|\mathbf{y} - X_i\theta_i^*\|_2^2, \theta_i^* = \arg \min \|\mathbf{y} - X_i\theta_i\|_2^2 \quad (8)$$

其中, r_i 是 \mathbf{y} 到单类重构子空间的距离, 表示 \mathbf{y} 由 X_i 表示的准确度. 为规范化 r_i 取值, 其最终定义为 $r_i = \exp((r_i - r_{\min})/\sigma_1)$, 其中 $r_{\min} = \min\{r_1, r_2, \dots, r_c\}$, σ_1 是带宽参数, 用于调整重构权值 r_i 的衰减速度. 此外, 为避免选择远距离训练数据成为表示样本, WGSC 在式 (7) 中集成了局部距离约束 $\mathbf{d}_i = [d_{i1}; d_{i2}; \dots; d_{in_i}]$, 其元素定义为 $d_{ik} = \exp(\|\mathbf{y} - \mathbf{x}_{ik}\|_2^2/\sigma_2)$, σ_2 的意义与 σ_1 一致. 同 GSC 算法相比, WGSC 不仅考虑了训练数据的组结构特性, 而且兼顾数据局部分布状态以及测试样本和训练数据间的相似度因子, 包含了更多的鉴别性信息.

1.4 稀疏判别投影分类算法

除忽略类别标签引导的缺陷外, SRC 型算法的另一个劣势是计算复杂度高. 此外, 获取稀疏性的潜在条件是过完备训练字典, 即输入样本的维数 $m \ll$

个数 n . 然而, 现实应用数据的维度一般都很高. 因此, SRC 的实现过程往往伴随着维数约简过程, 稀疏判别投影分类算法 SRCDP^[36] 是最近提出的一个稀疏表示型降维算法代表.

定义矩阵 $P \in \mathbf{R}^{m \times d}$ 是输入特征至降维子空间的线性投影变换, 即任意数据 \mathbf{x}_{ij} 从 m 维原空间通过 $\mathbf{z}_{ij} = P^T \mathbf{x}_{ij}$ 映射到 d 维子空间. 同理, 将整个训练集投影到子空间表示为 $Z = P^T X$. SRCDP 的基本思想是要求降维后的子空间样本不仅能够提升测试效率, 且子空间的样本更适于 SRC 分类任务. 基于此目标, 定义 $\mathbf{v}_{ij}^s = Z \delta_s(\boldsymbol{\theta}_{ij})$ 为子空间数据 \mathbf{z}_{ij} 由第 s 类系数重构而成的矢量, \mathbf{z}_{ij} 和第 s 类的距离 (或称为重构冗余) 则为 $d_s(\mathbf{z}_{ij}) = \|\mathbf{z}_{ij} - \mathbf{v}_{ij}^s\|$. 在此基础上, SRCDP 引入与 LDA 类似的类内散度矩阵 B_w 和类间散度矩阵 B_b .

$$B_w = \frac{1}{n} \sum_{ij} (\mathbf{z}_{ij} - \mathbf{v}_{ij}^i)(\mathbf{z}_{ij} - \mathbf{v}_{ij}^i)^T \quad (9)$$

$$B_b = \frac{1}{n(c-1)} \sum_{ij} \sum_{s \neq i} (\mathbf{z}_{ij} - \mathbf{v}_{ij}^s)(\mathbf{z}_{ij} - \mathbf{v}_{ij}^s)^T \quad (10)$$

将 $\mathbf{z}_{ij} = P^T \mathbf{x}_{ij}$, $Z = P^T X$ 以及 $\mathbf{v}_{ij}^i = Z \delta_i(\boldsymbol{\theta}_{ij})$, $\mathbf{v}_{ij}^s = Z \delta_s(\boldsymbol{\theta}_{ij})$ 代入式 (9) 和式 (10) 中, 则有

$$B_w = \frac{1}{n} \sum_{ij} [P^T \mathbf{x}_{ij} - P^T X \delta_i(\boldsymbol{\theta}_{ij})] \times [P^T \mathbf{x}_{ij} - P^T X \delta_i(\boldsymbol{\theta}_{ij})]^T \quad (11)$$

$$B_b = \frac{1}{n(c-1)} \sum_{ij} \sum_{s \neq i} [P^T \mathbf{x}_{ij} - P^T X \delta_i(\boldsymbol{\theta}_{ij})] \times [P^T \mathbf{x}_{ij} - P^T X \delta_i(\boldsymbol{\theta}_{ij})]^T \quad (12)$$

最终, 为实现最优的 SRC 分类, 以类内散度最小、类间散度最大为准则构建目标函数为

$$J(P) = \max \frac{\text{tr}(B_b)}{\text{tr}(B_w)} \quad (13)$$

依据上述构建过程可知, SRCDP 的重构系数 $\boldsymbol{\theta}_{ij}$ 由子空间数据 Z 计算得到, 而 Z 需要通过投影矩阵 P 计算得到. 因此, SRCDP 的求解过程需要迭代更新投影矩阵 P 和稀疏编码系数 $\boldsymbol{\theta}_{ij}$, 其收敛性并不明确, 且时间开销较大.

2 特征加权组稀疏分类算法

有效的稀疏系数编码是表示型投影分析算法的核心步骤. 因此, 本节针对现有表示型分类算法所存在的问题提出一种新的特征加权组稀疏分类算法 FWGSC. 在实际模式识别应用中, 训练样本往往具有复杂的分布结构或者存在污损及遮挡情形, 不同

的样本特征贡献度各异. 因此, 稀疏表示型算法的关键问题描述为: 选择哪些训练样本以及哪部分样本特征用于精确的重构测试样本. SRC 和 CRC 分别利用 l_1 和 l_2 范数求解稀疏系数 $\boldsymbol{\theta}$, 但是缺乏明确的类别标签指示. GSC 利用训练数据类别信息增加鉴别力, 但是其相同类别的训练样本在稀疏表示的过程中同步选择和舍弃, 会造成信息的丢失. WSRC 和 WGSC 算法通过数据间的欧氏距离引导同类近邻样本精确表示测试数据, 但是缺乏特征优选方案. RRC 和松弛协作表示算法^[41] 对不同的特征进行分部加权, 有效地缓解了噪声污染下的目标识别问题, 但是并没有考虑输入数据的局部结构信息. 基于此, 本节所提的 FWGSC 算法兼顾了稀疏性、标签信息、特征贡献因素以及局部结构信息, 目标是在使用训练样本字典集表示测试样本的过程中, 摒除样本中的无效特征分量, 使选择的训练样本不仅是与测试样本邻近的样本数据, 还包括与测试样本强关联的样本特征.

2.1 FWGSC 算法描述

组稀疏表示分类算法 GSC 已经验证为较 SRC 和 CRC 具有更佳的鉴别能力^[12, 42]. WGSC 进一步在 GSC 的基础上添加了样本加权约束, 使得算法兼顾稀疏性、监督性和保局性. 然而, WGSC 并没有考虑样本不同特征的影响因子. 本节以 WGSC 为基础, 假设测试样本 $\mathbf{y} \in \mathbf{R}^m$ 中的不同特征具有不同的贡献度, 以 \mathbf{s} 表示特征加权矢量, 其中 $s_i \in [0, 1)$, $i = 1, \dots, m$, 所构建的目标函数为

$$\min_{\boldsymbol{\theta}} \|\mathbf{s} \odot (X\boldsymbol{\theta} - \mathbf{y})\|_2^2 + \lambda \sum_{i=1}^c r_i^s \|\mathbf{d}_i^s \odot \boldsymbol{\theta}_i\|_2 \quad (14)$$

其中, 第 1 部分 $\|\mathbf{s} \odot (X\boldsymbol{\theta} - \mathbf{y})\|_2^2$ 是重构误差, 第 2 部分则是系数 $\boldsymbol{\theta}$ 的加权 l_{12} 混合范数正则化项, 两者都包含特征加权约束. 式 (14) 中 r_i^s 表示带特征约束的类组权值, 用于评估各类在表示测试样本时的相对重要性, 借鉴线性回归分类算法 (Linear regression classifier, LRC)^[43] 思想, 其形式为

$$\begin{aligned} r_i^s &= \|\mathbf{s} \odot (X_i \boldsymbol{\theta}_i^* - \mathbf{y})\|_2 \\ \boldsymbol{\theta}_i^* &= \arg \min_{\boldsymbol{\theta}_i} \|\mathbf{s} \odot (X_i \boldsymbol{\theta}_i - \mathbf{y})\|_2 \end{aligned} \quad (15)$$

r_i^s 值越大, 说明第 i 类数据无法精确的重构测试样本, 则第 i 类系数 $\boldsymbol{\theta}_i$ 趋向于收缩至 0 值. 为避免选择相距较远的训练数据来表示测试样本, 需要在 l_{12} 范数中增加局部约束, 式 (14) 采用 $\mathbf{d}_i^s = [d_{i1}^s; d_{i2}^s; \dots; d_{im_i}^s]$ 惩罚测试样本与第 i 类训练数据的距离值, 其定义为

$$d_{ik}^s = \mathbf{s} \odot \|\mathbf{x}_{ik} - \mathbf{y}\|_2 \quad (16)$$

此外, 由于特征约束矢量 \mathbf{s} 的引入, FWGSC 对每个特征点进行了加权. 设 $\mathbf{e} = \mathbf{y} - X\boldsymbol{\theta}$ 是测试样本 \mathbf{y} 的重构冗余, 则 $e_i = y_i - r_i\theta_i$ 是第 i 个特征点的重构冗余, $i = 1, \dots, m$, 其中 r_i 表示 X 中第 i 行. 当测试样本中第 i 个特征点受到噪声污损时, 通过特征权值 s_i 削弱甚至剔除该点的重构贡献 (同样的思想亦存在于 d_i 和 r_i). 因此, 为精确表示测试样本, FWGSC 不仅兼顾了 WGSC 和 RRC 算法的优势, 而且将特征加权思想融入了距离、重构计算过程中, 其解具有更丰富的鉴别信息. 为简化算法描述, 定义 $\boldsymbol{\eta}_i = [\eta_{i1}; \eta_{i2}; \dots; \eta_{in_i}]$, $i = 1, 2, \dots, c$, 其中 $\eta_{ik} = r_i^s d_{ik}^s$, $k = 1, 2, \dots, n_i$. 最优化目标函数式 (14) 改写为

$$\min_{\boldsymbol{\theta}} \|\mathbf{s} \odot (X\boldsymbol{\theta} - \mathbf{y})\|_2^2 + \lambda \sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|_2 \quad (17)$$

$\boldsymbol{\eta}_i$ 是测试样本与第 i 类数据之间的距离矢量. 很明显, $\boldsymbol{\eta}_i$ 不同于 WSRC、WCRC 和 WGSC 中的加权形式, 包含了更多的数据分布信息和噪声信息.

通过最优化式 (17), 所得的系数 $\boldsymbol{\theta}^*$ 能够精确表示测试样本, 可用于后续 FWGSDP 算法计算重构散度. 此外, 与 SRC 类似, FWGSC 是一个高效分类器, 其判别规则为

$$k = \arg \min_i \|\mathbf{s} \odot (\mathbf{y} - X\delta_i(\boldsymbol{\theta}^*))\|_2 \quad (18)$$

即在特征权值 \mathbf{s} 的引导下, 选择重构误差最小的类为测试样本 \mathbf{y} 的归属模式.

2.2 特征权值矢量 \mathbf{s}

假设所采集的训练数据是无损的普通样本, 而测试数据 \mathbf{y} 是不确定的. 以人脸识别为例, 训练样本是无遮挡无坏点的图像集, 测试样本则形式多样, 可以包含墨镜、围巾等穿戴物, 也可能是图像损坏带来的高斯噪声等. 由于常规特征元素可以通过训练样本正确表示. 因此, 奇异特征点 (遮挡、带噪) 所对应的重构冗余 e_i 具有更高值, 需要将其对应的 s_i 赋较小的值 (最小为 0). 相反, 低重构冗余值 e_i 所对应的矢量元素是常规特征点, 其对应的 s_i 权值应该较大 (接近于 1). 换言之, 目标函数式 (17) 中的 \mathbf{s} 本质上反映了样本特征的重构冗余状态描述.

从式 (17) 第 1 部分 $\|\mathbf{s} \odot (X\boldsymbol{\theta} - \mathbf{y})\|_2^2$ 可以看出, 当 $s_i = 1$ (\mathbf{s} 中所有元素为 1) 时, 对应于 l_2 范数, 即假设冗余状态呈高斯分布^[1]; 当 $s_i = (1/\|e_i\|)^{1/2}$ 时, 对应于 l_1 范数, 即假设冗余状态呈拉普拉斯分布^[1]; 此外, 还可以选择更为丰富的权值函数形式, 当 $s_i = \exp(-e_i^2/2\sigma^2)$ 时, 冗余状态呈高斯核逼近分布^[5]; 当 $s_i = 1/(1 + 1/(\exp(-\mu e_i^2 + \mu\sigma)))$ 时, 冗余状态呈 logistic 逼近分布^[22]. 图 1 所示为上述四

种 s_i 函数的示意图.

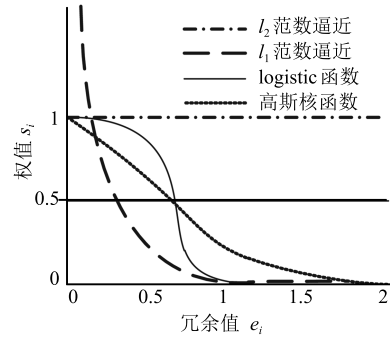


图 1 不同约束函数下的权值 s_i 变化

Fig. 1 Weight functions for different fidelity constraints

从图 1 可见, l_2 范数逼近对所有的特征点并无差别, 忽略了奇异点扰动和噪声影响等因素; l_1 范数逼近赋更高的权值于小冗余特征, 符合抗噪本质, 然而, 当冗余值逼近于 0 时, 其权值趋向于无穷, 为算法运行带来了不稳定性; 高斯核逼近函数和 logistic 权值函数的取值范围为 $[0, 1]$, 符合权值约束要求, 其中 logistic 函数对规范特征 (冗余值较小的特征) 赋更高的值而对奇异点 (冗余值较大的特征) 赋更小的值, 更符合抗噪保真的本质. 因此, 本文选用 logistic 作为权值矢量 \mathbf{s} 的特征约束函数.

2.3 FWGSC 优化求解与分析

FWGSC 算法兼顾样本和特征加权约束, 且其特征权值 s_i 会进一步影响样本权值. 因此, 常规的稀疏型求解算法, 如梯度投影法 SLEP^[44]、同伦分析法^[45]、近端梯度法^[46] 等, 都不能直接应用于求解 FWGSC 的最优系数 $\boldsymbol{\theta}$. 针对此问题, 本节提出迭代重约束方法进行模型求解. 为方便微分计算, 首先引入矩阵 $S = \text{diag}\{s_1, s_2, \dots, s_m\} \in \mathbf{R}^{m \times m}$, $\Pi = \text{diag}\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_c\} \in \mathbf{R}^{n \times n}$, 同时定义 $\sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|$ 简写为 $\|\Pi\boldsymbol{\theta}\|$, 将特征加权组稀疏的目标模型表示为

$$\min_{\boldsymbol{\theta}} \|S(X\boldsymbol{\theta} - \mathbf{y})\|_2^2 + \lambda \|\Pi\boldsymbol{\theta}\| \quad (19)$$

将式 (19) 依 $\boldsymbol{\theta}$ 求矢量微分并设为 0, 得

$$X^T S^T S X \boldsymbol{\theta} - X^T S^T S \mathbf{y} + \lambda \Pi^T D \Pi \boldsymbol{\theta} = 0 \quad (20)$$

其中, D 是块对角矩阵

$$\begin{bmatrix} \frac{1}{\|\boldsymbol{\eta}_1 \odot \boldsymbol{\theta}_1\|_2} I_{n_1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\|\boldsymbol{\eta}_c \odot \boldsymbol{\theta}_c\|_2} I_{n_c} \end{bmatrix}$$

则可得

$$\boldsymbol{\theta} = (X^T S^T S X + \lambda \Pi^T D \Pi)^{-1} X^T S^T S \mathbf{y} \quad (21)$$

其中, D 的取值依赖于系数 $\boldsymbol{\theta}$, 且样本权值对角矩阵 Π 的计算依赖于特征权值矢量 \mathbf{s} , 因此所提的迭代重约束求解策略细节如算法 1 所示.

算法 1. 迭代重约束稀疏表示构建算法

输入. 训练样本 X , 测试样本 \mathbf{y} , 平衡参数 λ , 初始化系数 $\boldsymbol{\theta}^{(1)}$.

输出. 最优系数 $\boldsymbol{\theta}^*$ 和最终特征加权系数 \mathbf{s} .

步骤 1. 设迭代数 $t = 1$, 计算冗余矢量 $\mathbf{e}^{(t)} = \mathbf{y} - X\boldsymbol{\theta}^{(t)}$.

步骤 2. 按照 logistic 函数获得特征权值矢量 \mathbf{s} , 并计算 $\boldsymbol{\eta}_i = [\eta_{i1}; \eta_{i2}; \dots; \eta_{in_i}]$, 其中 $\eta_{ik} = r_i^s d_{ik}^s$.

步骤 3. 构建特征加权矩阵 S 、样本加权矩阵 Π 以及块对角矩阵 D .

步骤 4. 依式 (21) 求解 $\boldsymbol{\theta}^{(t+1)}$.

步骤 5. 验证收敛条件, 如满足则输出 $\boldsymbol{\theta}^*$, 算法中止, 否则转至步骤 1.

算法 1 中值得强调以下几点:

1) 式 (21) 是一个凸优化问题, 因此算法 1 中输入初始化系数 $\boldsymbol{\theta}^{(1)}$ 的选择较为灵活, 本文采用 $\mathbf{1}/n$ 作为初始化系数, 其中 $\mathbf{1}$ 是所有元素为 1 的 n 维矢量.

2) 系数 $\boldsymbol{\theta}$ 的升级过程是一个闭式求解公式, 形式简单, 易于实现.

3) 算法 1 的收敛条件包含三个:

a) 目标函数式 (19) 满足 $\|obj^{(t+1)} - obj^{(t)}\|_2 / \|obj^{(t)}\|_2 < \varepsilon$;

b) 特征约束满足 $\|\mathbf{s}^{(t+1)} - \mathbf{s}^{(t)}\|_2 / \|\mathbf{s}^{(t)}\|_2 < \varepsilon$ (本文实验部分将之设定为 $\varepsilon = 5 \times 10^{-2}$);

c) 迭代步数 t 达到最大设定值 (本文为 20). 通过实验发现, 20 步的迭代次数多于实际所需, 一般在 10 次迭代内前两个算法收敛条件已经满足.

4) 通过定理 1 可知, 算法 1 具有全局最优解. 此外, 对比压缩感知理论中经典的迭代重加权方法 (Iteratively reweighted, IR)^[47-48], 算法 1 求解策略虽然基本思想迥异, 但其求解步骤却与 IR 具有相似性, 仅将 IR 中的非负加和因子收缩替换为特征加权系数更新. 因此, 所提的求解算法具有指数级收敛速率^[47].

定理 1. 目标函数式 (19) 通过算法 1 更新求解, 其值在每次迭代中逐步减少, 直至收敛.

证明. 通过算法 1 步骤 4 可知

$$\boldsymbol{\theta}^{(t+1)} = \min_{\boldsymbol{\theta}} \|S(X\boldsymbol{\theta} - \mathbf{y})\|_2^2 + \lambda \boldsymbol{\theta}^T \Pi^T D^{(t)} \Pi \boldsymbol{\theta}$$

因此

$$\|S(X\boldsymbol{\theta}^{(t+1)} - \mathbf{y})\|_2^2 + \lambda \boldsymbol{\theta}^{(t+1)T} \Pi^T D^{(t)} \Pi \boldsymbol{\theta}^{(t+1)} \leq$$

$$\|S(X\boldsymbol{\theta} - \mathbf{y})\|_2^2 + \lambda \boldsymbol{\theta}^T \Pi^T D^{(t)} \Pi \boldsymbol{\theta} \quad (22)$$

类似于 $-(a-b)^2 \leq 0$, 可以证明

$$\begin{aligned} \|\Pi \boldsymbol{\theta}^{(t+1)}\|_2 - \frac{\|\Pi \boldsymbol{\theta}^{(t+1)}\|_2^2}{2\|\Pi \boldsymbol{\theta}\|_2} &\leq \\ \frac{\|\Pi \boldsymbol{\theta}\|_2}{2} = \|\Pi \boldsymbol{\theta}\|_2 - \frac{\|\Pi \boldsymbol{\theta}\|_2^2}{2\|\Pi \boldsymbol{\theta}\|_2} &\quad (23) \end{aligned}$$

将式 (23) 改写为

$$\begin{aligned} \lambda \sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i^{(t+1)}\|_2 - \lambda \sum_{i=1}^c \frac{\|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i^{(t+1)}\|_2^2}{2\|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|_2} &\leq \\ \lambda \sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|_2 - \lambda \sum_{i=1}^c \frac{\|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|_2^2}{2\|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|_2} &\quad (24) \end{aligned}$$

即

$$\begin{aligned} \lambda \sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i^{(t+1)}\|_2 - \lambda \boldsymbol{\theta}^{(t+1)T} \Pi^T D^{(t)} \Pi \boldsymbol{\theta}^{(t+1)} &\leq \\ \lambda \sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|_2 - \lambda \boldsymbol{\theta}^T \Pi^T D^{(t)} \Pi \boldsymbol{\theta} &\quad (24) \end{aligned}$$

结合不等式 (22) 和 (24), 得到

$$\begin{aligned} \|S(X\boldsymbol{\theta}^{(t+1)} - \mathbf{y})\|_2^2 + \lambda \sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i^{(t+1)}\|_2 &\leq \\ \|S(X\boldsymbol{\theta} - \mathbf{y})\|_2^2 + \lambda \sum_{i=1}^c \|\boldsymbol{\eta}_i \odot \boldsymbol{\theta}_i\|_2 &\quad (25) \end{aligned}$$

由于目标函数通过迭代会逐次下降, 且式 (19) 具有下确界 (大于 0), 因此算法 1 收敛. \square

FWGSC 求解的关键耗时步骤为式 (21), 即步骤 4. 考虑到矩阵 S 和 Π 都是对角矩阵, 式 (21) 的计算复杂度为 $O(n^2m)$, 设算法 1 的迭代次数为 t , 则 FWGSC 的算法复杂度为 $O(tn^2m)$. 通过多次实验发现, 一般 t 的取值 ≤ 7 , 远远小于 m 和 n . 对比 SRC 的求解复杂度 $O(m^2n^{1.5})$ ^[36] 可知, FWGSC 的模型复杂度低于经典 SRC 算法, 尤其在高维度小样本应用中 (如图像分类、基因序列诊断等), 其求解效率更具有明显的优越性.

3 特征加权组稀疏判别投影分析算法

综上所述, FWGSC 算法兼顾特征和样本加权, 所得稀疏系数能够精确地重构未知样本, 本节以此作为系数计算关键步骤, 并借鉴 SRCDP 思想, 提出特征加权组稀疏判别投影分析算法 (Feature weighted group sparse discriminative projection, FWGSDP), 实现输入数据特征降维, 进一步提升模式分类效果.

3.1 FWGSDP 算法描述

依次选择第 i 类目标数据的第 j 个样本 \mathbf{x}_{ij} , 采用 FWGSC 算法, 通过余下的训练样本 X_{-i} 进行线性表示构建得到稀疏系数 $\boldsymbol{\theta}_{ij}$ 和特征加权系数 $S_{ij} = \text{diag}\{s_{ij}\}$. 定义 $\delta_k(\boldsymbol{\theta}_{ij})$ 为 R^m 列矢量, 其中仅保留第 k 类数据所对应的稀疏系数, 余下的置为 0. 基于 FWGSC 的决策规则, 分别定义类内冗余散度矩阵式 (25) 和类间冗余散度矩阵式 (26) 为

$$\begin{aligned} B'_w &= \frac{1}{n} \sum_{ij} [P^T S_{ij} \mathbf{x}_{ij} - P^T S_{ij} X \delta_i(\boldsymbol{\theta}_{ij})] \times \\ & [P^T S_{ij} \mathbf{x}_{ij} - P^T S_{ij} X \delta_i(\boldsymbol{\theta}_{ij})]^T = \\ & P^T \frac{1}{n} \sum_{ij} [S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_i(\boldsymbol{\theta}_{ij})] \times \\ & [S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_i(\boldsymbol{\theta}_{ij})]^T P = P^T B_w P \quad (25) \end{aligned}$$

$$\begin{aligned} B'_b &= \frac{1}{n(c-1)} \sum_{ij} \sum_{k \neq i} [P^T S_{ij} \mathbf{x}_{ij} - P^T S_{ij} X \delta_k(\boldsymbol{\theta}_{ij})] \times \\ & [P^T S_{ij} \mathbf{x}_{ij} - P^T S_{ij} X \delta_k(\boldsymbol{\theta}_{ij})]^T = \\ & P^T \frac{1}{n(c-1)} \sum_{ij} [S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_k(\boldsymbol{\theta}_{ij})] \times \\ & [S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_k(\boldsymbol{\theta}_{ij})]^T P = P^T B_b P \quad (26) \end{aligned}$$

其中, $B_w = (1/n) \sum_{ij} [S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_i(\boldsymbol{\theta}_{ij})][S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_i(\boldsymbol{\theta}_{ij})]^T$, $B_b = (1/n(c-1)) \sum_{ij} \sum_{k \neq i} [S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_s(\boldsymbol{\theta}_{ij})][S_{ij} \mathbf{x}_{ij} - S_{ij} X \delta_s(\boldsymbol{\theta}_{ij})]^T$ 分别是原空间的类内冗余散度矩阵和类间冗余散度矩阵. 与 SRCDP 类似, 以类内冗余散度最小、类间冗余散度最大为目标, 得到

$$J(P) = \text{tr}(\beta B'_b - B'_w) \quad (27)$$

其中, β 是平衡参数, 需人工设定, 一般范围为 (0, 2), 通过本文实验发现, 其具体选值对结果影响较小. 式 (27) 的基本思想是在子空间投影数据中, 同类数据重构散度缩小而异类数据重构散度扩大, 提升 FWGSC 分类器的鉴别性能.

此外, 为获得规范化解, 可添加相关性、正交性等约束^[49] 于式 (27). 本文将投影矩阵 P 约束为由单位矢量组成, 即 $P = [\mathbf{p}_1, \dots, \mathbf{p}_d]$ 且 $\mathbf{p}_k^T \mathbf{p}_k = 1$, $k = 1, \dots, d$, 用以更好地保持输入数据的分布形态^[49]. 因此, 最终的目标函数可以描述为

$$\begin{aligned} \max & \sum_{k=1}^d \mathbf{p}_k^T (\beta B_b - B_w) \mathbf{p}_k \\ \text{s.t.} & \mathbf{p}_k^T \mathbf{p}_k = 1, k = 1, \dots, d \quad (28) \end{aligned}$$

3.2 FWGSDP 优化求解与分析

使用拉格朗日乘子将 FWGSDP 的目标函数进行变换得到

$$L(\mathbf{p}_k, \lambda_k) = \sum_{k=1}^d \mathbf{p}_k^T (\beta B_b - B_w) \mathbf{p}_k - \lambda_k (\mathbf{p}_k^T \mathbf{p}_k - 1) \quad (29)$$

通过对式 (29) 求 \mathbf{p}_k 的导数, 且令求导结果为 0 求取最优解

$$\frac{\partial L}{\partial \mathbf{p}_k} = (\beta B_b - B_w - \lambda_k) \mathbf{p}_k = 0 \quad (30)$$

可得

$$(\beta B_b - B_w) \mathbf{p}_k = \lambda_k \mathbf{p}_k \quad (31)$$

从上式可知 λ_k 是矩阵 $\beta B_b - B_w$ 的特征值, \mathbf{p}_k 则是对应的特征矢量. 因此

$$\begin{aligned} J(P) &= \sum_{k=1}^d \mathbf{p}_k^T (\beta B_b - B_w) \mathbf{p}_k = \\ & \sum_{k=1}^d \lambda_k \mathbf{p}_k^T \mathbf{p}_k = \sum_{k=1}^d \lambda_k \quad (32) \end{aligned}$$

综上所述, 最优解 P^* 是由矩阵 $\beta B_b - B_w$ 前 d 个最大特征值对应的特征矢量组成. 此外, 由于 $\beta B_b - B_w$ 是实对称矩阵, 则命题 1 成立, FWGSDP 是一个有监督正交线性投影算法, 能够保留更多的鉴别性信息, 尤其适合于 FWGSC 算法.

命题 1. 式 (28) 的最优化解 P 问题是正交的, 即当 $i \neq j$ 时, $\mathbf{p}_i^T \mathbf{p}_j = 0$, 且 $\mathbf{p}_i^T \mathbf{p}_i = 1$.

结合算法 1 求解步骤, 特征加权组稀疏判别投影算法的求解步骤如算法 2 所示. 对比 SRCDP 和 FWGSDP, 前者需要迭代更新投影矩阵 P , 而后者采用一次计算作为最优化 P 解. 以一次计算为例, 两者在投影判别部分的计算复杂度基本一致, 包含散度矩阵构建和特征分解, 分别需要 $O(n^2(cm+n))$ 和 $O(m^3)$ 的计算量. 根据上一节所述, 在稀疏编码阶段, SRCDP 采用 SRC 进行系数求解, 复杂度高于 FWGSC, 因此 FWGSDP 的单次训练复杂度低于 SRCDP. 进一步考虑 SRCDP 是迭代型算法, 因此它的总体复杂度明显高于 FWGSDP 算法.

算法 2. 特征加权组稀疏判别投影算法

输入. 训练样本 X , 测试样本 \mathbf{y} , 平衡参数 λ , 初始化系数 $\boldsymbol{\theta}^{(1)}$.

输出. 投影矩阵 P .

步骤 1. 针对完整训练样本 X , 依次选择 \mathbf{x}_{ij} , 采用算法 1 构建稀疏系数 $\boldsymbol{\theta}_{ij}$.

步骤 2. 按式 (25) 和式 (28) 构建类内冗余散度矩阵 S_w 和类间冗余散度矩阵 S_b .

步骤 3. 计算矩阵 $\beta B_b - B_w$ 最大的 d 个特征矢量, 并构建投影矩阵 P .

4 实验分析

为验证所提算法的有效性, 以人脸识别为例, 分别通过原空间与降维子空间两方面对 FWGSC 和 FWGSDP 进行实验分析. 第 4.1 节为原空间性能测试, 验证了所提算法在无遮挡和有遮挡情形下的人脸识别率; 第 4.2 节为降维子空间中的性能测试, 将 FWGSDP 与经典降维分析算法的性能优劣进行了实验对比; 第 4.3 节详细讨论了 FWGSC 和 FWGSDP 的参数优选过程. 实验中采用的所有人脸图像都经过统一地中心配准和边缘裁剪. 所有样本包括训练数据和测试数据都通过 l_2 范数进行规范化处理. 实验平台为 32 位 Win7 操作系统, i5 处理器, 4 GB 内存以及 Matlab2014 运行环境.

4.1 原空间人脸识别性能分析

4.1.1 无遮挡人脸识别

首先在光照、表情变化却不包括噪声遮挡的环境中验证 FWGSC 算法的性能, 对比算法包括 WGSC^[20]、GSC^[11]、WSRC^[16]、SRC^[1] 和 LRC^[43]. 实验过程中不采用任何降维算法. 选用的人脸数据库为 ExYaleB^[50] 和 PIE^[51]. ExYaleB 包括 38 个人物对象的 21 888 张人脸图像, 其中含 9 个姿态调整和 64 种光照变化. 与文献 [21] 一致, 本文抽选其中 2 414 幅接近正面, 光照各异的图像用于测试, 每幅图像大小都统一调整至 32 像素 \times 32 像素. 任意选择其中每个人物 $n_{tr} = \{5, 10, 20, 30, 40, 50\}$ 幅图像作为训练样本, 余下的为测试样本. PIE 包括 68 人的 41 368 幅面部图像, 其中含 13 个姿态调整和

43 个光照变化. 本文选取其中 11 554 幅正面、不同光照的图像用于实验测试, 每幅图像大小都统一调整至 32 像素 \times 32 像素. 任意选择其中每个人物 $n_{tr} = \{5, 10, 20, 30, 40\}$ 幅图像作为训练样本, 余下的为测试样本.

图 2 显示了两个数据库中不同算法在不同训练样本量下的识别率结果, 具体数值由 10 次随机样本选择实验取平均得到. 从中可见, 六种表示型算法都具有优秀的识别性能, 在 $n_{tr} = 20$ 的实验条件下基本达到了大于 90% 的精度. 在所有算法中, LRC 分类器仅采用单类样本进行重构表示, 因此在训练样本量较少的情形中, 如 $n_{tr} = 5$ 和 $n_{tr} = 10$ 时, 其识别率明显低于其他几种算法, 而 n_{tr} 足够大时, 其识别率与 SRC 算法较为接近. WSRC 算法的综合识别率较 SRC 算法略有提升, 但是在 PIE 数据库的少样本实验中 ($n_{tr} = 5$ 和 $n_{tr} = 10$), 其精度反而低于 SRC, 说明单纯的样本距离加权不能准确地反映数据分布结构, 部分姿态接近的异类图像会比表情夸张和姿势变化的同类图像具有更小的欧氏距离值, 反而弱化分类器的判别能力. GSC 较 WSRC 算法具有明显的识别率提升, 说明 l_{12} 组范数的鉴别性能优于单纯的 l_1 范数或 l_2 范数. WGSC 算法在样本距离加权的基础上, 又添加了重构冗余加权, 其性能优于 GSC 算法, 尤其在 ExYaleB 数据库中, WGSC 较 GSC 的识别率平均提升了 2%. 最后, FWGSC 算法集合了特征加权、重构冗余加权、样本距离加权、组范数等优势特性, 其识别率全面优于其他几种表示型算法, 在 $n_{tr} = 20$ 时, 分别在 ExYaleB 和 PIE 数据库中获得了 97.34% 和 95.69% 的识别率.

4.1.2 有遮挡人脸识别

SRC 型算法的一个重要特征是遮挡鲁棒性^[1], 在本小节中, 选用 AR 人脸数据库^[21] 验证 FWGSC

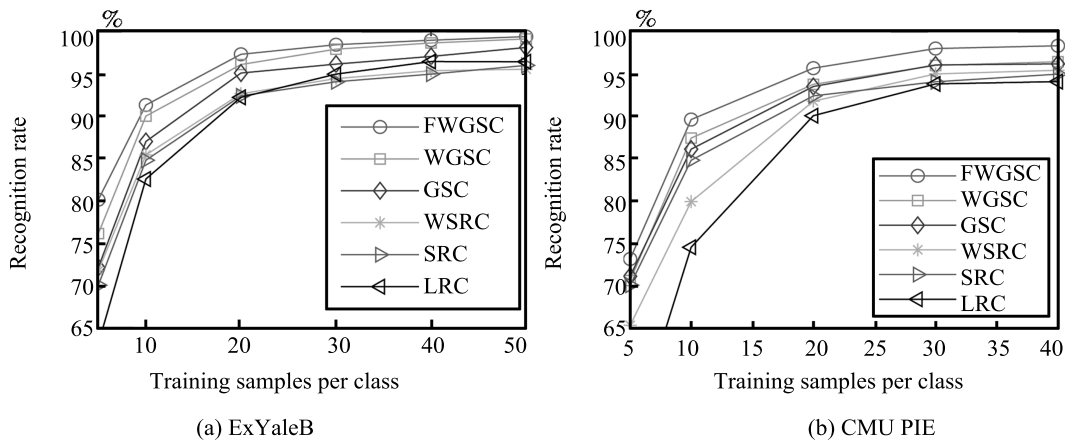


图 2 稀疏型分类器在不同数据库中识别率对比

Fig. 2 The recognition rates of different sparse representation classifiers in ExYaleB and PIE databases

分类器的抗遮挡能力. AR 人脸库包含 126 人的 4000 多幅人脸图像, 除不同的光照条件和面部表情外, 其中部分图像还含有局部遮挡(眼镜、头罩、围巾等). 与文献 [21] 类似, 本文抽选其中一个子集作实验测试, 包含 50 位男性和 50 位女性, 仅随机选取每个人物同时期 4 张无遮挡图片作为训练样本, 在不同时期中选取各人物不同遮挡情形下(眼镜或围巾)的 3 幅图像作为测试样本, 每幅图像都统一裁剪至 42 像素 \times 30 像素. 对比算法选用抗噪性较强的 RRC^[21]、CESR^[5]、GSRC^[6] 以及基准分类器 SRC, 其中 RRC 算法采用识别率较高的 l_1 范数约束.

实验结果如表 1 所示, 其中加粗数据值即为同等条件下的最优识别率. 很明显, 除在墨镜遮挡的不同时期测试实验中 RRC 算法共享了最优识别率外, 其他环境下 FWGSC 都具有最优的测试值. 值得注意的是, CESR 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 GSRC 则正处于相反的情况. 平均而言, 在同时期样本测试实验中, FWGSC 较 RRC、CESR、GSRC 和 SRC 分别提升了 0.65%、30.85%、11.35%、36.70% 的识别率. 在不同时期样本测试实验中, FWGSC 则分别提升了 1.0%、34.85%、28.3%、48.8% 的识别率.

表 1 AR 数据库有遮挡情形下不同算法的识别率对比
Table 1 Recognition rate by competing algorithms on AR database with occlusion

测试样本	算法	墨镜遮挡 (%)	围巾遮挡 (%)
同时期	FWGSC	99.3	95.7
	RRC	99.0	94.7
	CESR	95.3	38.0
	GSRC	87.3	85.0
	SRC	89.3	32.3
不同时期	FWGSC	89.3	78.3
	RRC	89.3	76.3
	CESR	79.0	20.7
	GSRC	45.0	66.0
	SRC	57.3	12.7

图 3 选择 AR 数据库中第一个人物目标的墨镜遮挡样例进行了分类过程展示, 其中, 图 3(a) 是测试样本; 图 3(b) 是算法收敛后的特征权值展示, 即将优化后的 s 向量展开为等像素图像, 其中黑色部分表示特征权值趋向于 0, 可见其正好反映了测试人脸的墨镜区域, 消除了墨镜遮挡部分对分类结果的影响; 图 3(c)~(g) 分别是经 FWGSC、RRC、CESR、GSRC、SRC 算法重构后

人脸图像. 其中, CESR、GSRC 和 SRC 三个算法缺乏特征加权考虑, 其重构人脸都带有明显的墨镜痕迹, 降低了识别精度; RRC 和 FWGSC 都引入了特征加权特性, 其重构人脸消除了墨镜效果, 相比较而言, FWGSC 算法的重构图像棱角更为明确, 而 RRC 则显得比较模糊; 图 3(h) 是算法的重构冗余值, 其中横坐标为不同类别标签, 可见 FWGSC 和 RRC 分类器的目标类(第一类)重构冗余小于其他类别的重构冗余, 其中, FWGSC 的目标类重构冗余值小于 RRC 的相应值, 说明其更为优秀的鉴别能力. CESR、GSRC 和 SRC 三个算法则将目标类别误认为冗余值最小的异类样本(第 60 类目标). 图 3(i) 是算法 1 的收敛过程展示, 从中可见特征权值在第 9 次迭代达到收敛条件 ($5E-2$), 而目标函数值则在第 5 次迭代即达到收敛条件, 两者都明显少于最高迭代次数 20. 同时, 图 3(i) 中展示了目标函数值变化的拟合曲线 $f(x) = 0.67 \exp(-0.68x)$, 说明本文所提迭代重约束求解算法具有快速收敛性.

4.2 降维子空间人脸识别性能分析

4.2.1 无遮挡人脸识别

通过 ExYaleB 和 PIE 两个数据进行降维空间中的无遮挡人脸识别实验, 数据库描述与上一节一致, 任意选择每类 20 个样本作为训练数据, 余下的所有样本作为测试数据.

实验步骤为: 首先采用训练样本构建不同子空间维数的投影矩阵 P ; 然后将训练数据和测试数据都依 P 投影至子空间; 最后在子空间中通过分类器得到整体测试识别率. 对比特征提取算法包括 PCA、LPP、LFDA、SRCDP 以及本文所提的 FWGSDP, 其中 FWGSDP 和 PCA、LPP、LFDA 的子空间分类器选用上一节实验结果中识别性能最优的 FWGSC, SRCDP 的子空间分类器则保持为文献 [36] 所述的 SRC (SRCDP 以 SRC 分类性能最优为目标进行降维投影目标泛函构建). LPP 和 LFDA 的近邻样本数选为 19, 即每类样本数减 1. 图 4 显示了上述算法在不同子空间维数下的识别率和方差对比. 通过图 4 可以发现: 1) 当子空间维数达到 300 后, 所有降维算法本身的识别率都不再存在明显变化, 说明对于一般的数据集来说, 300 维的特征描述已经足以反映样本间的鉴别性, 高于 300 的特征维数即存在表达冗余度; 2) 几何图嵌入降维的算法包括 PCA、LPP 和 LFDA 的识别率梯度较难确定, 如在 ExYaleB 300 维实验中 PCA 的识别率高于 LPP 和 LFDA, 在 PIE 100 和 200 维实验中 LPP 的识别率高于 PCA 和 LFDA, 而在 PIE 300 至 500 维实验中则是 LFDA 的识别率优于 PCA 和

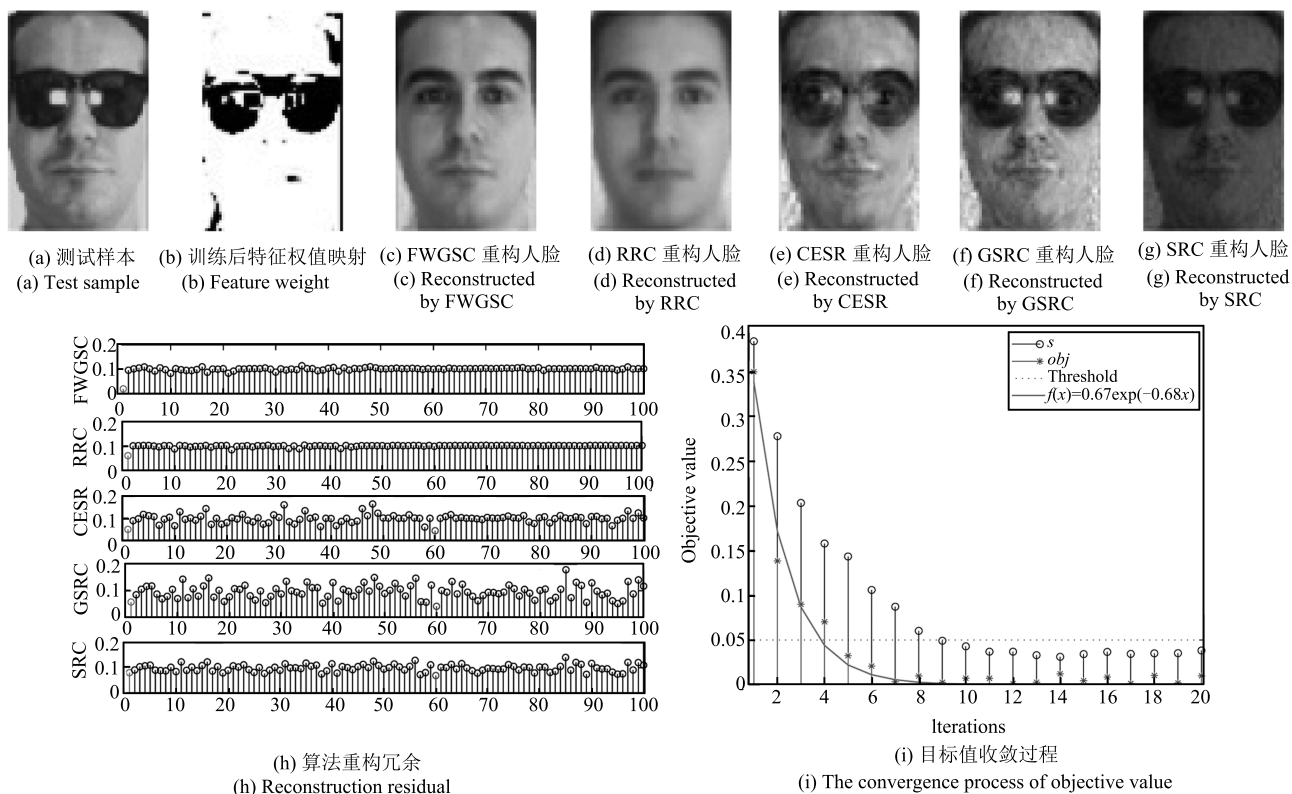


图 3 带墨镜遮挡的人脸识别效果展示

Fig. 3 Example of face recognition with sunglasses disguise

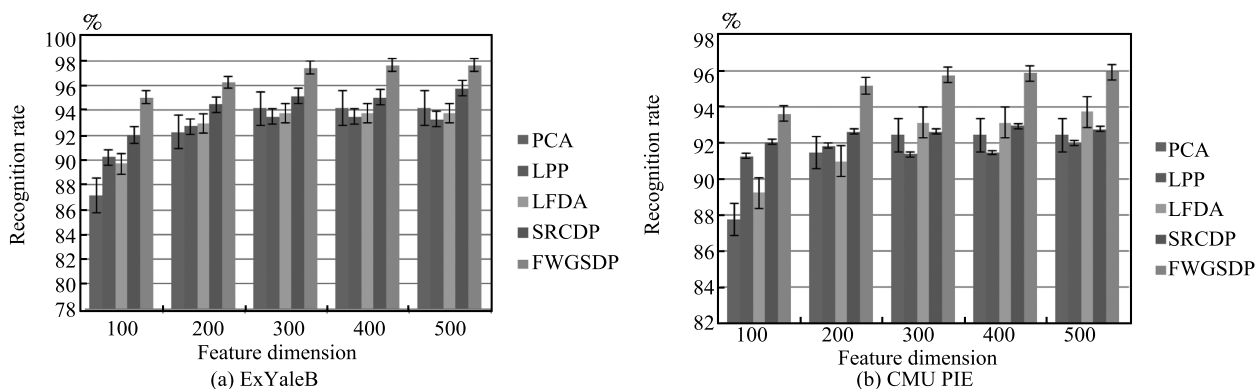


图 4 不同子空间维数的人脸识别率和方差对比

Fig. 4 Comparison of recognition rates and variance under different feature dimensions

LPP, 这主要是由于图嵌入降维算法并未与后续分类器进行有效融合, 实现为独立的降维模块和判别模块, 缺乏稳定的分类性能; 3) 在两个数据库中, FWGSDP 的人脸识别性能在不同子空间维数下都明显优于其他几种算法, 包括 SRCDP, 说明稀疏邻域模式设计与稀疏分类器融合的重要性以及 FWGSC 较 SRC 在分类性能上的优越性 (从上一节实验已经得到验证).

进一步, 表 2 给出了在 ExYaleB 和 PIE 数据库

中 FWGSDP 降维前后 FWGSC 分类算法的人脸识别测试实验对比, 包括识别率和每个样本识别时间 (单位为秒). 从表 2 可见, 当子空间维数达到 300 后, 人脸识别率都较原空间有所提升. 在 ExYaleB 中, 降维至 300 后人脸识别率较原空间提升了 0.1%, 在 PIE 中则是提升了 0.11%, 虽然提升幅度并不显著, 但是考虑到 FWGSC 算法在原空间中已经达到了约 97% 的识别性能, 能实现进一步提升已属难得. 此外, 在测试时间方面, 由于维数约简后的人脸数据特

征量大大减少,使得算法的运行效率得到了成倍提升.在 ExYaleB 和 PIE 数据库中分别获得了较原空间运行效率约 8 倍和 3 倍的测试时间缩减.

4.2.2 有遮挡人脸识别

与第 4.1.2 节实验环境一致,采用 AR 数据库测试 FWGSDP 和 FWGSC 在有遮挡人脸识别应用中的算法性能,对比算法仍然选取 PCA、LPP、LFDA、SRCDP 和 FWGSDP.表 3 为该实验测试结果,其中包括识别率和投影矩阵 P 的构建耗时(单位为秒),SRCDP 和 FWGSDP 的稀疏邻域构建采用了 PCA 预投影数据.依据第 4.2 节实验结果,表 3 中各算法统一选择子空间维数为 300.从表 3 可见,在四个不同遮挡情形中,通过 FWGSDP 维数约简后的子空间人脸识别率分别为 99.5%、96.0%、90.8% 和 79.6%,较原空间的 99.3%、95.7%、89.3% 和 78.3% 分别提升了 0.2%、0.3%、1.5% 和 1.3%;类似地,通过 SRCDP 维数约简后的子空间识别率较原空间中的 SRC 分

类精度分别提升了 3.5%、26.1%、21.4% 和 39.7%;而 PCA、LPP 和 LFDA 三种几何图嵌入型降维算法虽然也采用了 FWGSC 分类器,但其识别率却对比原空间不增反降,说明针对性的对稀疏系数采用类内紧缩类间离散策略构建子空间投影目标函数对于提升或保持 SRC 型算法性能是极其重要的.此外,表 3 反映了 FWGSDP 在模型构建效率上优于 SRCDP,符合文中算法复杂度分析.虽然经典图嵌入模型的构建效率更高,但考虑 FWGSDP 突出的识别性能以及模型构建效率在重要性上弱于运行效率的特点,GWGSDP 算法是值得推广应用的.

4.3 最优参数选择

本文实验所对比的算法 CESR、GSRC、WSRC、WGSRC、SRC、SRCDP、LFDA、RRC 都包含人工设定参数,在具体实验过程中采用各自文献所公开的源码进行性能验证,模型参数以原著设定为基本准则,通过网格式搜索确定最佳值.本节重点讨论

表 2 降维前后 FWGSC 分类算法的人脸识别对比

Table 2 Comparison of FWGSC performance with and without dimension reduction

维数	ExYaleB		PIE	
	识别率 (%)	测试时间 (s)	识别率 (%)	测试时间 (s)
原始样本 1024	97.34	0.9983	95.69	1.4002
FWGSDP 100	95.04	0.1103	93.62	0.4683
FWGSDP 200	96.25	0.1170	95.20	0.4896
FWGSDP 300	97.44	0.1215	95.80	0.5046
FWGSDP 400	97.61	0.1286	95.89	0.5175
FWGSDP 500	97.61	0.1332	95.95	0.5296

表 3 AR 数据库有遮挡情形下不同算法的识别率对比

Table 3 Comparison of competing dimension reduction algorithms on AR database with occlusion

测试样本	算法	墨镜遮挡识别率 (%)	墨镜遮挡耗时 (s)	围巾遮挡识别率 (%)	围巾遮挡耗时 (s)
同时期	PCA	91.0	1.083	85.7	1.184
	LPP	92.5	2.810	89.6	2.672
	LFDA	93.4	4.380	91.5	4.253
	SRCDP	92.8	21.27	58.4	21.96
	FWGSDP	99.5	14.72	96.0	14.03
不同时期	PCA	84.6	1.079	70.6	1.067
	LPP	87.3	2.714	75.1	2.705
	LFDA	86.5	4.525	75.4	4.361
	SRCDP	78.7	22.15	52.4	22.06
	FWGSDP	90.8	15.18	79.6	14.98

所提算法 FWGSDP 的最优参数设定思路, 包括 FWGSC 系数求解算法中的特征选择参数 σ 和 μ 、重构误差与系数平衡参数 λ , 以及类间散度和类内散度平衡参数 β . 参数 σ 和 μ 是 logistic 函数计算的前提, σ 称为分界点参数, 即 logistic 函数开始下降的位置, 用于选择有效特征数量. 一种直观的参数设定方案是取值为平方误差向量 e_i^2 , $i = 1, \dots, m$ 的第 $\lceil \tau m \rceil$ 大元素, 其中 m 是特征维数, $\lceil \cdot \rceil$ 是取上整数运算, $\tau \in (0, 1)$ 用于确定有效特征百分比. 设定 $\mu = 10/\sigma$, 确保低误差特征权值的 logistic 函数映射接近于 1. 因此具体实验中的待定参数为 τ 、 λ 和 β , 采取先确定 β 和 λ 再在投影过程中确定 β 的思路.

以 PIE 为例, 选取每类数据 20 个样本、子空间维数为 300 作为参数优选过程演示. 考虑到有效特征数量不可能过少, 因此设定 β 的选取范围为 $\{0.95, 0.9, 0.85, 0.8, 0.75\}$, 而 λ 则以文献 [17] 的默认值 $1E-3$ 展开设定, 即 $\{1E-5, 1E-4, 5E-4, 1E-3, 5E-3, 1E-2\}$. 图 5 展示了所述两个参数变化下的人脸识别结果, 从中可见, 当取 $\tau = 0.9$ 且 $\lambda = 1E-4$ 时, FWGSC 具有最佳识别性能. 以此为基础, 图 6 在不同的 β 值下进行了识别率展示. 综合图 5 和图 6 的参数选择过程可知, 虽然 FWGSDP 模型具有多个人工选择参数, 但其确定过程却比较直观. 首先, 如图 6 所示, 不同的 β 值下的人脸识别率曲线波动很小, 一般选择 $0.5 \sim 1.5$ 之间时, 可以达到近似最优的算法性能; 在图 5 中随着 τ 和 λ 值变化时, 识别率波动相对较大, 在具体应用中可依据经验进行参数值设定, 其中 λ 的取值在 $1E-4$ 和 $1E-3$ 之间较为稳定, 而 τ 值在有遮挡和无遮挡应用中一般可分别设定为 0.7 和 0.9 进行算法实现.

5 总结

本文提出了一种特征加权组稀疏判别投影算法

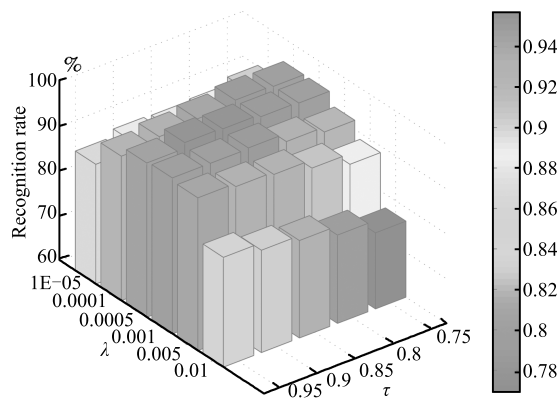


图 5 PIE 中 FWGSC 参数 λ 和 τ 选择

Fig. 5 The selection of parameter λ and τ of FWGSC in PIE database

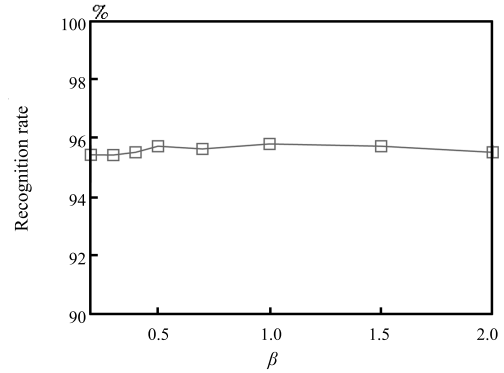


图 6 PIE 中 FWGSDP 参数 β 选择

Fig. 6 The selection of parameter β of FWGSDP in PIE database

FWGSDP, 包含两个关键操作步骤, 即特征加权组稀疏分类和最优化投影判别分析. 特征加权组稀疏分类以加权组稀疏模型 WGSC 为基础, 通过引入特征加权约束, 改进了样本距离结构和类组重构冗余的权值信息, 使得 FWGSC 在带噪和普通环境中都能够获得精确的表示系数, 其独立分类性能优于经典的稀疏表示型算法, 并且是 FWGSDP 算法中重构系数构建的关键步骤. 最优化投影判别分析以 FWGSC 分类效果最优化为目标计算类内冗余散度矩阵和类间冗余散度矩阵, 并以类 Fisher 准则制定约束函数, 计算最优的判别投影矩阵, 使得 FWGSC 在降维后的数据中具有更高的识别率, 且算法运行效率大大提升. 以人脸数据为例, 采用 ExYaleB、PIE 和 AR 数据库验证了所提算法的性能优于经典模式分类和特征提取算法.

通过实验发现, 所提算法 FWGSDP 虽然易于实现并性能出众, 但应用过程中需要人工设定参数 σ 、 μ 、 λ 和 β , 虽然可以通过经验方式进行指引设置, 且 β 的取值对最终结果的影响较小, 但仍然会削弱所提算法的应用扩展能力. 因此, 后续将集中进行参数的自适应确定或参数削减工作.

References

- 1 Wright J, Yang A Y, Ganesh A, Sastry S S, Yi M. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(2): 210–227
- 2 Xu Y, Zhang D, Yang J, Yang J Y. A two-phase test sample sparse representation method for use with face recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, **21**(9): 1255–1262
- 3 He R, Hu B G, Zheng W S, Guo Y Q. Two-stage sparse representation for robust recognition on large-scale database. In: *Proceedings of the 24th AAAI Conference on Artificial Intelligence*. Atlanta, Georgia, USA: AAAI, 2010. 475–480
- 4 Yang M, Zhang L, Feng X C, Zhang D. Sparse representation based fisher discrimination dictionary learning for im-

- age classification. *International Journal of Computer Vision*, 2014, **109**(3): 209–232
- 5 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): 1561–1577
- 6 Yang M, Zhang L. Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary. In: Proceedings of the 11th European Conference on Computer Vision. Heraklion, Crete, Greece: Springer, 2010. 448–461
- 7 Liu Jian-Wei, Cui Li-Peng, Liu Ze-Yu, Luo Xiong-Lin. Survey on the regularized sparse models. *Chinese Journal of Computers*, 2018, **38**(7): 1307–1325
(刘建伟, 崔立鹏, 刘泽宇, 罗雄麟. 正则化稀疏模型综述. 计算机学报, 2015, **38**(7): 1307–1325)
- 8 Chen Z H, Zuo W M, Hu Q H, Lin L. Kernel sparse representation for time series classification. *Information Sciences*, 2015, **292**: 15–26
- 9 Zhang L, Yang M, Feng X C. Sparse representation or collaborative representation: which helps face recognition? In: Proceedings of the 2011 International Conference on Computer Vision. Barcelona: IEEE, 2011. 471–478
- 10 Zhu P F, Zuo W M, Zhang L, Shiu S C K, Zhang D. Image set-based collaborative representation for face recognition. *IEEE Transactions on Information Forensics and Security*, 2014, **9**(7): 1120–1132
- 11 Majumdar A, Ward R K. Fast group sparse classification. *Canadian Journal of Electrical and Computer Engineering*, 2009, **34**(4): 136–144
- 12 Elhamifar E, Vidal R. Robust classification using structured sparse representation. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2011. 1873–1879
- 13 Huang J, Nie F P, Huang H, Ding C. Supervised and projected sparse coding for image classification. In: Proceedings of the 27th AAAI Conference on Artificial Intelligence. Bellevue, Washington, USA: AAAI, 2013. 438–444
- 14 Wang L J, Lu H C, Wang D. Visual tracking via structure constrained grouping. *IEEE Signal Processing Letters*, 2015, **22**(7): 794–798
- 15 Liu H C, Li S T, Yin H T. Infrared surveillance image super resolution via group sparse representation. *Optics Communications*, 2013, **289**: 45–52
- 16 Lu C Y, Min H, Gui J, Zhu L, Lei Y K. Face recognition via weighted sparse representation. *Journal of Visual Communication and Image Representation*, 2013, **24**(2): 111–116
- 17 Timofte R, van Gool L. Adaptive and weighted collaborative representations for image classification. *Pattern Recognition Letters*, 2014, **43**: 127–135
- 18 Wu J Q, Timofte R, van Gool L. Learned collaborative representations for image classification. In: Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, HI: IEEE, 2015. 456–463
- 19 Chao Y W, Yeh Y R, Chen Y W, Lee Y J, Wang Y C F. Locality-constrained group sparse representation for robust face recognition. In: Proceedings of the 18th IEEE International Conference on Image Processing. Brussels: IEEE, 2011. 761–764
- 20 Tang X, Feng G C, Cai J X. Weighted group sparse representation for undersampled face recognition. *Neurocomputing*, 2014, **145**: 402–415
- 21 Yang M, Zhang L, Yang J, Zhang D. Regularized robust coding for face recognition. *IEEE Transactions on Image Processing*, 2013, **22**(5): 1753–1766
- 22 Jerome F, Trevor H, Robert T. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2001. 534–553
- 23 Duda R O, Hart P E, Stork D G. *Pattern Classification* (2nd edition). New York, USA: Wiley, 2001.
- 24 Zheng Jian-Wei, Wang Wan-Liang, Yao Xiao-Min, Shi Hai-Yan. Face recognition using tensor local Fisher discriminant analysis. *Acta Automatica Sinica*, 2012, **38**(9): 1485–1495
(郑建伟, 王万良, 姚晓敏, 石海燕. 张量局部 Fisher 判别分析的人脸识别. 自动化学报, 2012, **38**(9): 1485–1495)
- 25 Zheng J W, Yang D, Chen S Y, Wang W L. Incremental min-max projection analysis for classification. *Neurocomputing*, 2014, **123**: 121–130
- 26 Yang Li-Ping, Gong Wei-Guo, Gu Xiao-Hua, Li Wei-Hong, Du Xing. Complete discriminant locality preserving projections for face recognition. *Journal of Software*, 2010, **21**(6): 1277–1286
(杨利平, 龚卫国, 辜小花, 李伟红, 杜兴. 完备鉴别保局投影人脸识别算法. 软件学报, 2010, **21**(6): 1277–1286)
- 27 Zhao Jia-Cheng, Cui Hui-Min, Feng Xiao-Bing. Analyzing cross-core performance interference on multi-core processors based on statistical learning. *Journal of Software*, 2013, **24**(11): 2558–2570
(赵家程, 崔慧敏, 冯晓兵. 基于统计学习分析多核间性能干扰. 软件学报, 2013, **24**(11): 2558–2570)
- 28 He X F, Yan S C, Hu Y X, Niyogi P, Zhang H J. Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(3): 328–340
- 29 Sugiyama M. Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis. *Journal of Machine Learning Research*, 2007, **8**(5): 1027–1061
- 30 Yan S C, Xu D, Zhang B Y, Zhang H J, Yang Q, Lin S. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(1): 40–51
- 31 Qiao L S, Chen S C, Tan X Y. Sparsity preserving projections with applications to face recognition. *Pattern Recognition*, 2010, **43**(1): 331–341
- 32 Cheng B, Yang J C, Yan S C, Fu Y, Huang T S. Learning with l_1 -graph for image analysis. *IEEE Transactions on Image Processing*, 2010, **19**(4): 858–866
- 33 Zheng Zhong-Long, Huang Xiao-Qiao, Jia Jiong, Yang Jie. Locality preserving projection with sparse penalty. *Chinese Journal of Computers*, 2014 **37**(9): 2038–2046
(郑忠龙, 黄小巧, 贾洞, 杨杰. 稀疏局部保持投影. 计算机学报, 2014, **37**(9): 2038–2046)
- 34 Shao Z F, Zhang L. Sparse dimensionality reduction of hyperspectral image based on semi-supervised local Fisher discriminant analysis. *International Journal of Applied Earth Observation and Geoinformation*, 2014, **31**: 122–129

- 35 Ly N H, Du Q, Fowler J E. Sparse graph-based discriminant analysis for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, **52**(7): 3872–3884
- 36 Yang J, Chu D L, Zhang L, Yu Y, Yang J Y. Sparse representation classifier steered discriminative projection with applications to face recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, **24**(7): 1023–1035
- 37 Lu C Y, Huang D S. Optimized projections for sparse representation based classification. *Neurocomputing*, 2013, **113**: 213–219
- 38 Ly N H, Du Q, Fowler J E. Collaborative graph-based discriminant analysis for hyperspectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, **7**(6): 2688–2696
- 39 Yang W K, Wang Z Y, Sun C Y. A collaborative representation based projections method for feature extraction. *Pattern Recognition*, 2015, **48**(1): 20–27
- 40 Ma Xiao-Hu, Tan Yan-Qi. Face recognition based on discriminant sparsity preserving embedding. *Acta Automatica Sinica*, 2014, **40**(1): 73–82
(马小虎, 谭延琪. 基于鉴别稀疏保持嵌入的人脸识别算法. 自动化学报, 2014, **40**(1): 73–82)
- 41 Yang M, Zhang D, Zhang D, Wang S L. Relaxed collaborative representation for pattern classification. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2012. 2224–2231
- 42 Liu Y, Li X M, Liu C Y, Tang Y F. Group sparsity in dimensionality reduction of sparse representation. In: Proceedings of the 2014 International Symposium on Wireless Personal Multimedia Communications. Sydney, NSW: IEEE, 2014. 541–546
- 43 Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(11): 2106–2112
- 44 Liu J, Ji S W, Ye J P. SLEP: Sparse Learning with Efficient Projections. Arizona State University, USA, 2009 [Online], available: <http://www.yelab.net/software/SLEP/>, May 4, 2016
- 45 Salman A M, Romberg J. Dynamic updating for l_1 minimization. *IEEE Journal of Selected Topics in Signal Processing*, 2010, **4**(2): 421–434
- 46 Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2009, **2**(1): 183–202
- 47 Lu C Y, Lin Z C, Yan S C. Smoothed low rank and sparse matrix recovery by iteratively reweighted least squares minimization. *IEEE Transactions on Image Processing*, 2015, **24**(2): 646–654
- 48 Lei Y, Song Z J. An improved IRLS algorithm for sparse recovery with intra-block correlation. *Optik*, 2015, **126**(7–8): 850–854
- 49 Li F, Wang J X, Tang B P, Tian D Q. Life grade recognition method based on supervised uncorrelated orthogonal locality preserving projection and K -nearest neighbor classifier. *Neurocomputing*, 2014, **138**: 271–282

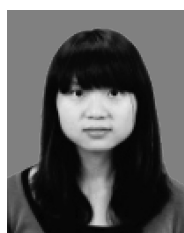
50 Lee K C, Ho J, Kriegman D. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(5): 684–698

51 Gross R, Matthews I, Cohn J, Kanade T, Baker S. Multi-PIE. *Image and Vision Computing*, 2010, **28**(5): 807–813



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, pattern recognition.)



黄琼芳 浙江工业大学计算机科学与技术学院硕士研究生. 主要研究方向为模式识别. E-mail: gdhqf@sina.cn

(**HUANG Qiong-Fang** Master student at the College of Computer Science and Technology, Zhejiang University of Technology. Her research interest covers pattern recognition.)



杨平 浙江工业大学计算机科学与技术学院硕士研究生. 主要研究方向为模式识别.

E-mail: 2111412076@zjut.edu.cn

(**YANG Ping** Master student at the College of Computer Science and Technology, Zhejiang University of Technology. His research interest covers pattern recognition.)



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. E-mail: wwl@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.)



马文龙 衢州职业技术学院信息工程学院副教授. 主要研究方向为机器学习, 智能推荐. 本文通信作者.

E-mail: qzmwlmwl@126.com

(**MA Wen-Long** Associate professor at the School of Information Engineering, Quzhou College of Technology. His research interest covers machine learning and intelligent recommendation. Corresponding author of this paper.)