

# 稀疏约束下快速低秩共享的字典学习方法及其人脸识别

田泽杨明李爱师

(南京师范大学计算机科学与技术学院 南京 210023)

(zetian\_edu@126.com)

## Fast Low-Rank Shared Dictionary Learning with Sparsity Constraints on Face Recognition

Tian Ze, Yang Ming, and Li Aishi

(College of Computer Science and Technology, Nanjing Normal University, Nanjing 210023)

**Abstract** Dictionary learning is one of the most important feature representation methods. It has a wide range of applications in face recognition and other aspects. It is particularly suitable for solving face recognition problems under the change of pose, and has attracted much attention from many researchers. In order to enhance the discriminative ability of dictionary, researchers have put forward a large number of dictionary learning models in combination with domain knowledge and anti-noise strategies, including the recently proposed methods for simultaneous dimensionality reduction and dictionary learning, but these methods focus on the specific-class samples and fail to consider the sharing information between training samples. Therefore, we propose a fast low-rank shared dictionary learning with sparsity constraints approach. The method learns dimensionality reduction and dictionary jointly, and embeds Fisher discriminant criteria to obtain specific-class dictionary and coding coefficients. At the same time, we enforce a low-rank constraint to obtain the low-rank shared dictionary to enhance the discriminative ability of dictionary and coding coefficients. In addition, the Cayley transform is used to protect the orthogonality of the projection matrix to catch a compact feature set. Face recognition experiments on AR, Extended Yale B, CMU PIE, and FERET datasets demonstrate the superiority of our approach. The experimental results show that the proposed method has strong robustness to face recognition under facial expression changes, and plays an inhibitory role in lighting. It is especially suitable for solving small sample problems under illumination and expression changes.

**Key words** face recognition; dictionary learning; sparsity constraints; low-rank models; shared features

**摘要** 字典学习是重要的特征表示方法之一,在人脸识别等方面有广泛的应用,特别适合解决姿态变化下的人脸识别问题,因而倍受研究者的关注.为有效增强字典的判别能力,研究者结合领域知识和抗噪等策略提出大量的字典学习模型,其中包括最近提出的同时进行降维和字典学习的方法,但这些方法侧重考虑样本中特定类的信息,未能有效考虑训练样本间的共享信息.因此,提出了一种稀疏约束下

收稿日期:2018-05-18;修回日期:2018-06-06

基金项目:国家自然科学基金重点项目(61432008);国家自然科学基金项目(61272222)

This work was supported by the Key Program of the National Natural Science Foundation of China (61432008) and the National Natural Science Foundation of China (61272222).

通信作者:杨明(m.yang@njnu.edu.cn)

快速低秩共享的字典学习方法. 该方法采用降维和字典联合进行学习的方式, 并嵌入 Fisher 判别准则获得特定类字典和编码系数, 同时施加低秩约束获得低秩共享字典, 以此增强字典和编码系数的判别能力. 此外, 运用 Cayley 变换保护投影矩阵的正交性来获得紧凑的特征集合. 在 AR, Extended Yale B, CMU PIE 和 FERET 四个数据集上的人脸识别实验验证所提方法的优越性. 实验结果表明所提方法在表情变化下的人脸识别具有很强的鲁棒性, 并对光照起到了抑制作用, 尤其适合解决光照、表情变化下的小样本问题.

**关键词** 人脸识别; 字典学习; 稀疏约束; 低秩模型; 共享特征

**中图法分类号** TP181

人脸识别存在光照、表情、姿态、小样本等挑战性问题, 同时, 人脸识别的高维问题也增加了人脸识别的难度, 通常使用经典的降维方法主要有 LDA<sup>[1]</sup>, LPP<sup>[2]</sup> 和 PCA<sup>[3]</sup>, 降维得到的子空间能够提高人脸识别的准确率. 但是这些方法对人脸识别存在的光照、表情等问题不具有良好的鲁棒性. 稀疏表示可以较好地解决这些问题.

稀疏表示<sup>[4]</sup>已经成为信号处理领域的强大工具. 应用包括压缩感知<sup>[5]</sup>、稀疏信号恢复<sup>[6]</sup>、图像分割<sup>[7]</sup>以及信号分类. 在这些应用领域中, 信号通常可以由固定的字典来表示. 基于这一理论, 稀疏表示分类器被提出并应用于人脸识别. 然而, 固定的字典对于分类任务并没有足够的鉴别力, 这就促进基于稀疏约束的字典学习研究.

字典学习可以分为无监督字典学习和有监督字典学习. KSVD<sup>[8]</sup>算法为无监督字典学习的典型代表, 通过奇异值分解对字典原子进行更新. 文献[9]提出 DKSVD 算法, 将线性分类器加入 KSVD 字典学习模型, 促使 KSVD 算法具有分类能力. 文献[10]将鉴别性稀疏编码误差项加入 DKSVD 算法促使稀疏编码的判别能力进一步加强. 文献[11]将 Fisher 判别准则嵌入到字典和稀疏编码中, 这促使稀疏编码和字典都具有判别能力, 但其仅考虑样本中特有的信息, 没有考虑样本间的共享信息. 针对这一问题, 文献[12]提出利用低秩约束字典来获得样本间的共享信息, 这增强了字典和稀疏编码的判别能力.

此外, 大部分有监督字典学习<sup>[9-12]</sup>都是先对数据进行降维再进行字典学习, 这些方法不能从原数据集中获得更为重要的特征, 从而降低字典的学习能力. 针对这一问题, 稀疏嵌入框架<sup>[13]</sup>被提出, 通过同时降维和字典学习的策略来增强字典和稀疏编码的判别能力. 文献[14]通过约束投影矩阵正交来获得紧凑的特征, 但其投影矩阵的求解是不恰当的且

收敛性不能被保证, 因此降低了字典的学习能力. 针对此问题, SEDL<sup>[15]</sup>通过 Cayley<sup>[16]</sup>变换来保护投影矩阵的正交性, 从而获得重要和紧凑的特征.

针对上述所提字典学习的不足, 本文提出一种稀疏约束下快速低秩共享的字典学习 (FLRSDLSC) 方法, 并将其用于人脸图像分类. 本文的主要贡献有 3 个方面:

- 1) 字典学习框架由特定类的字典和共享的子字典所组成. 对于特定类的字典, 嵌入 Fisher 判别准则; 对于共享的子字典, 嵌入低秩约束. 因此, 该方法能从样本中获得共享和特定类的特征, 以此增强字典和稀疏编码的判别能力.
- 2) 通过 Cayley 变换保护投影矩阵的正交性来获得紧凑的特征.
- 3) 采用降维和字典学习同时进行的方法, 增强字典对降维后样本的表示能力.

## 1 相关工作

### 1.1 稀疏嵌入字典学习 (SEDL)

Chen 等人于 2017 年提出一种同时降维和字典学习的方法 (SEDL) 应用于人脸识别. 其目标函数如下:

$$J(\mathbf{P}, \mathbf{Y}, \mathbf{D}, \mathbf{X}) = \frac{1}{2} f(\mathbf{P}, \mathbf{Y}, \mathbf{D}, \mathbf{X}) + \lambda_1 \|\mathbf{X}\|_1 + \frac{\lambda_2}{2} g(\mathbf{X}) + \lambda_3 \|\mathbf{P}^T \mathbf{P} \mathbf{Y} - \mathbf{Y}\|_F^2, \quad (1)$$

s. t.  $\mathbf{P}^T \mathbf{P} = \mathbf{I},$

其中,  $f(\mathbf{P}, \mathbf{Y}, \mathbf{D}, \mathbf{X})$  为变量  $\mathbf{P}, \mathbf{Y}, \mathbf{D}, \mathbf{X}$  的函数,

$$f(\mathbf{P}, \mathbf{Y}, \mathbf{D}, \mathbf{X}) = \sum_{c=1}^C \gamma(\mathbf{P}, \mathbf{Y}_c, \mathbf{D}, \mathbf{X}_c), \quad (2)$$

$$\gamma(\mathbf{P}, \mathbf{Y}_c, \mathbf{D}, \mathbf{X}_c) = \|\mathbf{P} \mathbf{Y}_c - \mathbf{D} \mathbf{X}_c\|_F^2 + \|\mathbf{P} \mathbf{Y}_c - \mathbf{D}_c \mathbf{X}_c^c\|_F^2 + \sum_{j=1, j \neq c}^C \|\mathbf{D}_j \mathbf{X}_c^j\|_F^2; \quad (3)$$

$$g(\mathbf{X}) = \sum_{c=1}^C (\|\mathbf{X}_c - \mathbf{M}_c\|_F^2 - \|\mathbf{M}_c - \mathbf{M}\|_F^2) + \|\mathbf{X}\|_F^2; \quad (4)$$

$\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_c, \dots, \mathbf{Y}_C) \in \mathbb{R}^{d \times N}$  表示大小为  $N$  的训练数据集,  $\mathbf{Y}_c \in \mathbb{R}^{d \times n_c}$  表示第  $c$  类训练样本, 其中  $N = n_1 + \dots + n_c + \dots + n_C$ . 降维的目标是学习投影矩阵  $\mathbf{P} \in \mathbb{R}^{p \times d}$  ( $p < d$ ),  $p$  表示数据降维后的维度大小.  $\mathbf{D} = (\mathbf{D}_1, \dots, \mathbf{D}_c, \dots, \mathbf{D}_C) \in \mathbb{R}^{p \times k}$  表示结构化的字典, 其中  $\mathbf{D}_c \in \mathbb{R}^{p \times k_c}$ ,  $k = k_1 + \dots + k_c + \dots + k_C$ ,  $\mathbf{D}_c$  为第  $c$  类子字典. 让  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_c, \dots, \mathbf{X}_C) \in \mathbb{R}^{k \times N}$  表示  $\mathbf{Y}$  在字典  $\mathbf{D}$  下的编码系数, 其中  $\mathbf{X}_c$  表示  $\mathbf{Y}_c$  在字典  $\mathbf{D}$  下的编码系数. 令  $\mathbf{x}, \mathbf{x}_c$  列的平均向量分别为  $\mathbf{m}$  与  $\mathbf{m}_c$ , 则  $\mathbf{X}_c, \mathbf{X}$  平均稀疏编码矩阵分别为  $\mathbf{M}_c = (\mathbf{m}_c, \dots, \mathbf{m}_c)$  与  $\mathbf{M} = (\mathbf{m}, \dots, \mathbf{m})$ .

式(1)由 4 部分组成: 重构误差项  $f(\mathbf{P}, \mathbf{Y}, \mathbf{D}, \mathbf{X})$ 、 $l_1$  范数正则化项  $\|\mathbf{X}\|_1$ 、鉴别性稀疏编码项  $g(\mathbf{X})$  和 F 范数正则化项. 式(1)等号右边第 1 项希望降维后的训练样本  $\mathbf{Y}_c$  能由字典  $\mathbf{D}$  表示, 也可以由子字典  $\mathbf{D}_c$  表示, 但不能由其他类子字典表示, 从而增强字典的判别能力; 第 2 项促使编码系数尽可能稀疏; 第 3 项最小化稀疏编码的类内离散度, 最大化稀疏编码的类间离散度, 因此稀疏编码系数具有判别能力,  $\|\mathbf{X}\|_F^2$  确保式(4)是凸和稳定的; 第 4 项最小化由降维引起的源数据域的重构误差, 防止降维时引起病态投影. 约束项希望降维矩阵  $\mathbf{P}$  尽可能正交, 以此获得更为紧凑的特征信息. SEDL 只考虑特定类的样本信息, 忽略样本间的相关性, 导致字典和稀疏编码判别能力的降低.

### 1.2 矩阵秩最小化理论

Natarajan 于 2006 年提出用稀疏近似解<sup>[17]</sup>去求解矩阵最小化问题, 其目标函数如下:

$$\min_{\mathbf{H}} \text{rank}(\mathbf{H}), \quad (5)$$

s. t.  $L(\mathbf{H}) = b,$

其中  $\mathbf{H} \in \mathbb{R}^{s \times d}$  是决策变量,  $L$  是线性映射,  $L$  将  $\mathbf{H}$  从  $s$  维映射到  $p$  维. 然而式(5)的求解是 NP 难问题. 当

$\|\mathbf{H}\| \leq 1$ , 式(5)的解就是核范数  $\|\mathbf{H}\|_* = \sum_{k=1}^n \sigma_k(\mathbf{H})$  (即所有奇异值之和)的解. 人们转而求解如下凸优化问题:

$$\min_{\mathbf{H}} \|\mathbf{H}\|_*, \quad (6)$$

s. t.  $L(\mathbf{H}) = b,$

式(6)通过奇异值阈值<sup>[18]</sup>算法近似求解.

## 2 稀疏约束下快速低秩共享的字典学习 (FLRSDLSC)

### 2.1 FLRSDLSC 模型构建

针对 SEDL 方法的不足, 本文提出一种稀疏约束下快速低秩共享的字典学习方法. 该方法学习特定类字典和共享子字典, 以此增强字典和稀疏编码的判别能力, 同时采用字典学习和降维同时进行的方式. 其模型定义如下:

$$J(\mathbf{P}, \mathbf{Y}, \bar{\mathbf{D}}, \bar{\mathbf{X}}) = \frac{1}{2} f(\mathbf{P}, \mathbf{Y}, \bar{\mathbf{D}}, \bar{\mathbf{X}}) + \lambda_1 \|\bar{\mathbf{X}}\|_1 + \frac{\lambda_2}{2} g(\bar{\mathbf{X}}) + \lambda_3 \|\mathbf{P}^T \mathbf{P} \mathbf{Y} - \mathbf{Y}\|_F^2 + \eta \|\mathbf{D}_0\|_* \quad (7)$$

s. t.  $\mathbf{P}^T \mathbf{P} = \mathbf{I},$

其中,

$$f(\mathbf{P}, \mathbf{Y}, \bar{\mathbf{D}}, \bar{\mathbf{X}}) = \|\mathbf{P} \mathbf{Y} - \bar{\mathbf{D}} \bar{\mathbf{X}}\|_F^2 + \sum_{c=1}^C (\|\mathbf{P} \mathbf{Y}_c - \mathbf{D}_c \mathbf{X}_c - \mathbf{D}_0 \mathbf{X}_c^0\|_F^2 + \sum_{j=1, j \neq c}^C \|\mathbf{D}_j \mathbf{X}_c^j\|_F^2); \quad (8)$$

$$g(\bar{\mathbf{X}}) = g(\mathbf{X}) + \|\mathbf{X}^0 - \mathbf{M}^0\|_F^2; \quad (9)$$

$\lambda_1, \lambda_2, \eta, \lambda_3$  为正则化参数;  $\mathbf{X}^0$  列的均值向量为  $\mathbf{m}^0$ ,  $\mathbf{X}^0$  的均值矩阵为  $\mathbf{M}^0 = (\mathbf{m}^0, \dots, \mathbf{m}^0)$ ;  $\mathbf{D}_0 \in \mathbb{R}^{p \times k_0}$  为共享子字典,  $\bar{\mathbf{D}} = (\mathbf{D}, \mathbf{D}_0)$  为总字典,  $\|\cdot\|_*$  为核范数正则化促使子字典  $\mathbf{D}_0$  具有低秩结构.  $\mathbf{X}^0$  代表  $\mathbf{Y}$  在共享字典  $\mathbf{D}_0$  下的编码系数,  $\bar{\mathbf{X}} = (\mathbf{X}^T, (\mathbf{X}^0)^T)^T$  代表  $\mathbf{Y}$  在总字典  $\bar{\mathbf{D}}$  下的编码系数.

式(7)中的目标函数的第 2 项、第 4 项以及约束项与式(1)一致, 不同处在于第 1 项、第 3 项, 还增加了第 5 项, 这源于将共享字典加入到 SEDL 模型. 第 1 项不同处在于希望样本  $\mathbf{Y}$  可以由总字典  $\bar{\mathbf{D}}$  表示, 也可以由第  $c$  类子字典  $\mathbf{D}_c$  和共享子字典  $\mathbf{D}_0$  表示, 以此增强字典的判别能力(图 1 解释了式(8)中变量间的关系). 第 3 项不同处在于嵌入共享字典所对应的稀疏编码, 由于该字典是低秩的, 其对应的稀疏编码应该是相似的(图 2 解释了式(9)变量间的关系), 以此增强稀疏编码的判别能力.

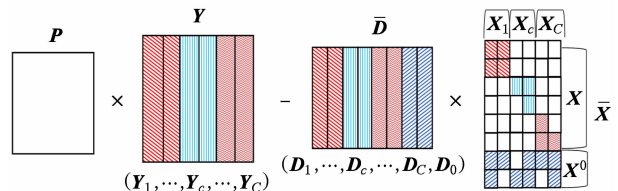


Fig. 1 The relationship between variables in term (8)

图 1 式(8)变量间的关系

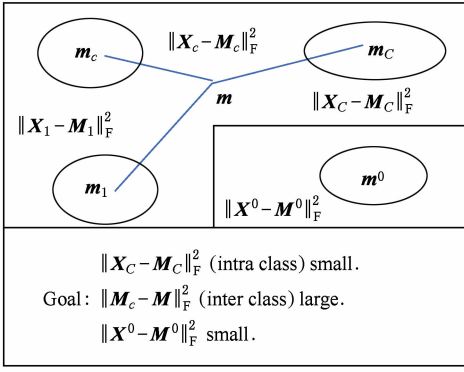


Fig. 2 The relationship between variables in term (9)

图2 式(9)变量间的关系

## 2.2 FLRSDLSC 模型求解

式(7)是一个非凸的问题,其求解通过反复更新稀疏编码  $\bar{X}$ 、字典  $\bar{D}$ 、降维矩阵  $P$  这 3 个过程,直至收敛。

### 2.2.1 更新稀疏编码 $\bar{X}$

固定  $P, \bar{D}$  更新  $\bar{X}$ , 式(7)重新定义如下:

$$J_{\bar{X}} = \arg \min_{\bar{X}} \bar{h}(\bar{X}) + \lambda_1 \|\bar{X}\|_1, \quad (10)$$

其中

$$\bar{h}(\bar{X}) = \frac{1}{2} f_{P,Y,\bar{D}}(\bar{X}) + \frac{\lambda_2}{2} g(\bar{X}), \quad (11)$$

$f_P(\mathbf{D}) = f(P, \mathbf{D})$  代表  $P$  固定  $\mathbf{D}$  的函数,用式(11)求导后的结果和 FISTA<sup>[19]</sup> 算法对式(10)进行求解,式(11)求导如下:

$$\frac{\partial \bar{h}(\bar{X})}{\partial \bar{X}} = \begin{pmatrix} \frac{\partial \bar{h}_{X^0}(\mathbf{X})}{\partial \mathbf{X}} \\ \frac{\partial \bar{h}_X(\mathbf{X}^0)}{\partial \mathbf{X}^0} \end{pmatrix}, \quad (12)$$

式(12)的第 1 分量推导如下:

$$\begin{aligned} \bar{h}_{X^0}(\mathbf{X}) &= \frac{1}{2} f_{P,Y,\bar{D},X^0}(\mathbf{X}) + \frac{\lambda_2}{2} \bar{g}_{X^0}(\mathbf{X}) = \\ &= \frac{1}{2} f_{P,Y,\bar{D}}(\mathbf{X}) + \frac{\lambda_2}{2} g(\mathbf{X}). \end{aligned} \quad (13)$$

引理 1. 
$$\frac{\partial \bar{h}_{X^0}(\mathbf{X})}{\partial \mathbf{X}} = (\hat{R}(\mathbf{D}^T \mathbf{D} + 2\lambda_2 \mathbf{I})) \mathbf{X} - \hat{R}(\mathbf{D}^T \bar{Y}) + \lambda_2 (\mathbf{M} - 2\hat{\mathbf{M}}), \quad (14)$$

其中,  $\hat{\mathbf{M}} = (\mathbf{M}_1, \dots, \mathbf{M}_c, \dots, \mathbf{M}_C)$ ;  $\mathbf{I}$  代表单位矩阵;  $\bar{Y} = \mathbf{P}\mathbf{Y} - \mathbf{D}_0 \mathbf{X}^0$ ;  $\hat{R}$  定义为函数,对于矩阵

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \cdots & \mathbf{A}_{1C} \\ \mathbf{A}_{21} & \cdots & \mathbf{A}_{2C} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{C1} & \cdots & \mathbf{A}_{CC} \end{pmatrix},$$

$\hat{R}(\mathbf{A})$  定义为  $\mathbf{A}$  的 2 倍对角矩阵,即:

$$\hat{R}(\mathbf{A}) = \mathbf{A} + \begin{pmatrix} \mathbf{A}_{11} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{A}_{CC} \end{pmatrix}. \quad (15)$$

证明. 见附录 A.

式(12)的第 2 分量,使用引理 3,推导如下:

$$\bar{h}_X(\mathbf{X}^0) = \|\mathbf{V} - \mathbf{D}_0 \mathbf{X}^0\|_F^2 + \frac{\lambda_2}{2} \|\mathbf{X}^0 - \mathbf{M}^0\|_F^2, \quad (16)$$

$$\begin{aligned} \frac{\partial \bar{h}_X(\mathbf{X}^0)}{\partial \mathbf{X}^0} &= 2\mathbf{D}_0^T \mathbf{D}_0 \mathbf{X}^0 - 2\mathbf{D}_0^T \mathbf{V} + \lambda_2 (\mathbf{X}^0 - \mathbf{M}^0) = \\ &= (2\mathbf{D}_0^T \mathbf{D}_0 + \lambda_2 \mathbf{I}) \mathbf{X}^0 - 2\mathbf{D}_0^T \mathbf{V} - \lambda_2 \mathbf{M}^0, \end{aligned} \quad (17)$$

其中  $\mathbf{V} = \mathbf{P}\mathbf{Y} - \frac{1}{2} \mathbf{D} \hat{R}(\mathbf{X})$ .

式(14)和式(17)组合求解式(12).

### 2.2.2 更新字典 $\bar{D}$

引理 2. 固定  $P, \bar{X}$  更新  $\bar{D} = [\mathbf{D}, \mathbf{D}_0]$ , 提出如下方法分别解决  $\mathbf{D}, \mathbf{D}_0$ . 更新  $\mathbf{D}$ , 式(7)简化如下:

$$\mathbf{D} = \arg \min_{\mathbf{D}} -2\text{tr}(\mathbf{E}\mathbf{D}^T) + \text{tr}(\mathbf{F}\mathbf{D}^T \mathbf{D}), \quad (18)$$

其中  $\mathbf{E} = \bar{Y}\hat{R}(\mathbf{X}^T)$  和  $\mathbf{F} = \hat{R}(\mathbf{X}\mathbf{X}^T)$ .

证明. 见附录 B.

式(18)运用 ODL<sup>[20]</sup> 算法进行求解.

引理 3. 固定  $\mathbf{D}, \bar{X}, P$  更新  $\mathbf{D}_0$ , 式(7)简化如下:

$$\begin{aligned} J_{\mathbf{D}_0} &= \arg \min_{\mathbf{D}_0} \|\mathbf{V} - \mathbf{D}_0 \mathbf{X}^0\|_F^2 + \frac{\lambda_2}{2} \|\mathbf{X}^0 - \mathbf{M}^0\|_F^2 + \\ &= \eta \|\mathbf{D}_0\|_* + \lambda_1 \|\mathbf{X}^0\|_1, \end{aligned} \quad (19)$$

其中  $\mathbf{V} = \mathbf{P}\mathbf{Y} - \frac{1}{2} \mathbf{D} \hat{R}(\mathbf{X})$ .

证明. 见附录 C.

基于引理 3, 式(19)简化如下:

$$\begin{aligned} J_{\mathbf{D}_0} &= \arg \min_{\mathbf{D}_0} -2\text{tr}(\hat{\mathbf{E}}\mathbf{D}_0^T) + \\ &= \text{tr}(\hat{\mathbf{F}}\mathbf{D}_0^T \mathbf{D}_0) + \eta \|\mathbf{D}_0\|_*, \end{aligned} \quad (20)$$

其中  $\hat{\mathbf{E}} = \mathbf{V}(\mathbf{X}^0)^T$ ,  $\hat{\mathbf{F}} = \mathbf{X}^0(\mathbf{X}^0)^T$ .

式(20)使用 ADMM<sup>[21]</sup> 模型和奇异值阈值算法. ADMM 过程如下: 首先选择一个正数  $\rho$ , 初始化  $\mathbf{Z} = \mathbf{U} = \mathbf{D}_0$ , 然后迭代求解如下子问题直到收敛:

$$\begin{aligned} J_{\mathbf{D}_0} &= \arg \min_{\mathbf{D}_0} -2\text{tr}(\bar{\mathbf{E}}\mathbf{D}_0^T) + \\ &= \text{tr}(\bar{\mathbf{F}}\mathbf{D}_0^T \mathbf{D}_0) + \eta \|\mathbf{D}_0\|_*, \end{aligned} \quad (21)$$

其中

$$\bar{\mathbf{E}} = \hat{\mathbf{E}} + \frac{\rho}{2} (\mathbf{Z} - \mathbf{U}), \quad (22)$$

$$\bar{\mathbf{F}} = \hat{\mathbf{F}} + \frac{\rho}{2} \mathbf{I}, \quad (23)$$

$$\mathbf{Z} = K_{\eta/\rho}(\mathbf{D}_c + \mathbf{U}), \quad (24)$$

$$\mathbf{U} = \mathbf{U} + \mathbf{D}_0 - \mathbf{Z}, \quad (25)$$

其中  $K$  为软阈值运算符. 式(21)的优化问题运用 ODL 算法进行求解.

### 2.2.3 更新降维矩阵 $\mathbf{P}$

固定  $\bar{\mathbf{X}}, \bar{\mathbf{D}}$  更新  $\mathbf{P}$ , 式(7)重新定义如下:

$$J_{\mathbf{P}} = \arg \min_{\mathbf{P}} \frac{1}{2} f(\mathbf{P}, \mathbf{Y}, \bar{\mathbf{D}}, \bar{\mathbf{X}}) + \lambda_3 \|\mathbf{P}^T \mathbf{P} \mathbf{Y} - \mathbf{Y}\|_{\mathbf{F}}^2, \quad (26)$$

s. t.  $\mathbf{P}^T \mathbf{P} = \mathbf{I},$

式(26)使用 Cayley 变换求解  $\mathbf{P}$ , 其求解简化为

$$\mathbf{P}^{i+1} = \mathbf{P}^i - \frac{\tau}{2} \hat{\mathbf{A}}(\mathbf{P}^i + \mathbf{P}^{i+1}), \quad (27)$$

其中  $i$  初始值为 1;  $\mathbf{P}^1$  为训练样本的 PCA 降维矩阵;  $\mathbf{P}^i$  代表第  $i$  次迭代的降维矩阵;  $\tau$  是一个恰当的步伐;  $\mathbf{P}^{i+1} = \mathbf{Q}\mathbf{P}^i, \mathbf{Q} = \left(\mathbf{I} + \frac{\tau}{2} \hat{\mathbf{A}}\right)^{-1} \left(\mathbf{I} - \frac{\tau}{2} \hat{\mathbf{A}}\right), \hat{\mathbf{A}}$  为斜对称矩阵,  $\hat{\mathbf{A}} = \nabla J \mathbf{P}^T - \mathbf{P} \nabla J^T, \nabla J$  为式(26)的导数, 求导简化成如下形式:

$$\nabla J(\mathbf{P}) = -\mathbf{D} \hat{\mathbf{R}}(\mathbf{X}) \mathbf{Y}^T + 2\mathbf{P} \mathbf{Y} \mathbf{Y}^T - 2\mathbf{D}^0 \mathbf{X}^0 \mathbf{Y}^T + \lambda_3 (-4\mathbf{P} \mathbf{Y} \mathbf{Y}^T + 2\mathbf{P} \mathbf{Y} \mathbf{Y}^T \mathbf{P}^T \mathbf{P} + 2\mathbf{P} \mathbf{P}^T \mathbf{P} \mathbf{Y} \mathbf{Y}^T), \quad (28)$$

通过对上述变量的更新, 输出  $\mathbf{P}, \mathbf{D}, \mathbf{D}_0, \mathbf{X}, \mathbf{X}^0$ .

### 2.3 FLRSDLSC 模型的分类

基于 2.2 节, 利用学习得到降维矩阵  $\mathbf{P}$ 、总字典  $\bar{\mathbf{D}}$  和其对应的稀疏编码  $\bar{\mathbf{X}}$ , 则 FLRSDLSC 分类方法定义如下:

$$\bar{\mathbf{x}} = \arg \min_{\bar{\mathbf{x}}} \frac{1}{2} \|\mathbf{P} \mathbf{y} - \bar{\mathbf{D}} \bar{\mathbf{x}}\|_{\mathbf{F}}^2 + \frac{\lambda_2}{2} \|\mathbf{x}^0 - \mathbf{m}^0\|_{\mathbf{F}}^2 + \lambda_1 \|\bar{\mathbf{x}}\|_1, \quad (29)$$

其中  $\mathbf{y}$  为测试样本;  $\mathbf{m}^0$  为稀疏编码系数  $\mathbf{x}^0$  的列平均向量. 式(29)运用 FISTA 算法可求得  $\bar{\mathbf{x}} = (\mathbf{x}^T, (\mathbf{x}^0)^T)^T$ , 从  $\bar{\mathbf{x}}$  提取  $\mathbf{x}^0$  和  $\mathbf{x}_c$ , 通过如下最小化重构误差来预测  $\mathbf{y}$  的类别:

$$\text{identity}(\mathbf{y}) = \arg \min_{1 \leq c \leq C} \|\mathbf{P} \mathbf{y} - \mathbf{D}_0 \mathbf{x}^0 - \mathbf{D}_c \mathbf{x}_c\|_2^2 + (1-\omega) \|\mathbf{x} - \mathbf{m}_c\|_2^2, \quad (30)$$

其中,  $\omega$  为平衡参数,  $\mathbf{m}_c$  为  $\mathbf{x}_c$  的列平均向量.

### 2.4 FLRSDLSC 算法

输入: 训练集  $\mathbf{Y}$ , 参数  $\lambda_1, \lambda_2, \lambda_3, \eta$  和最大迭代次数  $T$ ;

步骤 1. 初始化  $\mathbf{P}$  为训练样本的 PCA 降维矩阵,  $\mathbf{X} = \mathbf{0}, \bar{\mathbf{D}}$  为随机矩阵;

步骤 2. 固定  $\mathbf{P}, \bar{\mathbf{D}}, \mathbf{X}^0$ , 更新  $\mathbf{X}$ , 使用 FISTA 算法及式(14)对式(10)求解编码系数  $\mathbf{X}$ ;

步骤 3. 固定  $\mathbf{P}, \bar{\mathbf{D}}, \mathbf{X}$ , 更新  $\mathbf{X}^0$ , 使用 FISTA 算法及式(17)对式(10)求解编码系数  $\mathbf{X}^0$ ;

步骤 4. 固定  $\mathbf{P}, \bar{\mathbf{X}}, \mathbf{D}_0$ , 更新  $\mathbf{D}$ , 使用 ODL 算法对式(18)求解字典  $\mathbf{D}$ ;

步骤 5. 固定  $\mathbf{P}, \bar{\mathbf{X}}, \mathbf{D}$ , 更新  $\mathbf{D}_0$ , 使用 ODL 算法对式(21)求解共享字典  $\mathbf{D}_0$ ;

步骤 6. 固定  $\bar{\mathbf{D}}, \bar{\mathbf{X}}$ , 更新  $\mathbf{P}$ , 使用式(27)求解降维矩阵  $\mathbf{P}$ ;

7) 重复步骤 2 到步骤 6, 直到达到收敛条件或者满足最大迭代次数;

输出: 降维矩阵  $\mathbf{P}$ 、字典  $\bar{\mathbf{D}}$ 、稀疏编码系数  $\bar{\mathbf{X}}$ . 结合 2.3 节, 利用式(30)对测试样本进行分类.

## 3 复杂度分析

我们的算法快速主要体现在更新稀疏编码  $\mathbf{X}$  和结构化(特定类)字典  $\mathbf{D}$  上, 本节比较了原始和高效的 FDDL 字典学习方法的复杂度. 每个算法的复杂度被估计为一次迭代所需乘法的数量. 假设: 1) 每类样本数量是特定类字典数量的 2 倍, 特定类字典数量与共享类字典数量是相同的. 令  $n_c = n, k_c = k_0 = k$ ; 2) 每类字典的个数、训练样本的数量都远小于样本维数, 即  $k < d, n < d$ ; 3) 每个算法需要  $q$  次迭代才能收敛.

在本节分析中使用以下事实: 1) 如果  $\mathbf{A}_1 \in \mathbb{R}^{m_1 \times n_1}, \mathbf{B}_1 \in \mathbb{R}^{n_1 \times p_1}$ , 则矩阵乘法  $\mathbf{A}_1 \mathbf{B}_1$  具有复杂度  $O(m_1 n_1 p_1)$ ; 2) 如果  $\mathbf{A}_1 \in \mathbb{R}^{n \times n}$  是非奇异的, 则矩阵求逆  $\mathbf{A}_1^{-1}$  的复杂度为  $O(n^3)$ ; 3) 矩阵的奇异值分解  $\mathbf{A}_1 \in \mathbb{R}^{p \times q}, p > q$ , 其复杂度  $O(pq^2)$ .

### 3.1 Fisher 判别准则下的字典学习 (FDDL)

#### 1) 更新 $\mathbf{X}$ (O-FDDL-X)

参考文献[12], 原始的稀疏编码更新被分为  $C$  个子问题 O-FDDL-X 的复杂度:

$$C^2 k(dn + qCkn + Cdk). \quad (31)$$

#### 2) 更新 $\mathbf{X}$ (E-FDDL-X)

由引理 1 可知,

$$\frac{\partial \frac{1}{2} f_{\mathbf{Y}, \mathbf{D}}(\mathbf{X})}{\partial \mathbf{X}} = \hat{\mathbf{R}}(\mathbf{D}^T \mathbf{D}) - \hat{\mathbf{R}}(\mathbf{D}^T \mathbf{Y}), \quad (32)$$

$$\lambda_2 \left[ \frac{\partial \frac{1}{2} g(\mathbf{X})}{\partial \mathbf{X}} \right] = \lambda_2 (2\mathbf{X} + \mathbf{M} - 2\hat{\mathbf{M}}), \quad (33)$$

其中,  $\hat{\mathbf{R}}(\cdot)$  计算代价不是很高,  $\mathbf{M}$  与  $\mathbf{M}_c$  的计算可以忽略, 因为都是列运算. 将式(32)和式(33)组合起来得到该算法的复杂度:

$$\frac{(Ck)d(Ck)}{M(\mathbf{D}^T \mathbf{D} + \lambda_2 \mathbf{D})} + \frac{(Ck)d(Cn)}{M(\mathbf{D}^T \mathbf{Y})} + \frac{q(Ck)(Ck)(Cn)}{M(\mathbf{D}^T \mathbf{D} + \lambda_2 \mathbf{D} \mathbf{X})} = C^2 k(dk + dn + qCnk). \quad (34)$$

3) 更新  $\mathbf{D}$ (O-FDDL-D)

参考文献[12],原始的字典更新被分为  $C$  个子问题,O-FDDL-D的复杂度:

$$Cdk(qkn+C^2n). \quad (35)$$

4) 更新  $\mathbf{D}$ (E-FDDL-D)

参考文献[12],E-FDDL-D的复杂度:

$$Cdk(Cqk+Cn)+C^3k^2n. \quad (36)$$

表 1,2 分别展示原始的 FDDL 算法和高效的 FDDL 算法的复杂度分析与不同字典学习方法的总体复杂度分析.

**Table 1 Complexity Analysis for the Proposed Efficient Algorithms and Their Original Versions**

表 1 高效算法和原始算法的复杂度分析

Method	Complexity	Plugging Numbers
O-FDDL-X	$C^2k(dn+qCkn+Cdk)$	$1.51 \times 10^{11}$
E-FDDL-X	$C^2k(dk+dn+qCnk)$	$1.01 \times 10^{11}$
O-FDDL-D	$Cdk(qkn+C^2n)$	$10^{11}$
E-FDDL-D	$Cdk(Cqk+Cn)+C^3k^2n$	$2.8 \times 10^{10}$

**Table 2 Complexity Analysis for Different Dictionary Learning Methods**

表 2 不同字典学习方法的复杂度分析

Method	Complexity	Plugging Numbers
O-FDDL	$C^2dk(n+Ck+Cn)+Ck^2q(d+C^2n)$	$2.51 \times 10^{11}$
E-FDDL	$C^2k((q+1)k(d+Cn)+2dn)$	$1.29 \times 10^{11}$

## 3.2 复杂度计算

选择一系列参数集合:类别数为 100、数据降维后的维度为 50、数据维度为 500、迭代次数为 50、特定类与共享类字典数为 10、每类训练样本为 20,也就是  $C=100, n=20, d=500, q=50, k=10$ . 假设  $q_2=50$ . 从表 1,2 可以得出高效 FDDL 算法相比原始 FDDL 算法实现低复杂度.

## 4 实验与结果

本文在 4 个公开的数据集上进行实验:AR<sup>[22]</sup>, Extended Yale B<sup>[23]</sup>, CMU PIE<sup>[24]</sup>, FERET<sup>[25]</sup> 人脸数据集. 对比的方法主要有 SRC<sup>[4]</sup>, LCKSVD<sup>[10]</sup>, FDDL<sup>[11]</sup>, LRS DL<sup>[12]</sup>, SEDL<sup>[15]</sup>, FLRS DLSC\*. 为了解释联合降维和字典学习的能力,固定降维矩阵的 FLRS DLSC\* 的方式被提出. 在所有实验中,主成分分析的方法应用于数据的降维或者初始化 SEDL, FLRS DLSC\* 的降维矩阵. 在 AR, the Extended

Yale B, CMU PIE, FERET 和 AR 性别数据集下, FLRS DLSC 模型的训练参数设置如下: $\lambda_1, \eta$  统一设置为 0.001 与 0.003; $\lambda_2$  分别设置为 0.2, 0.2, 0.2, 0.001, 0.2; $\lambda_4$  分别设置为 0.5, 0.3, 0.4, 0.3, 0.5. 其分类参数设置如下: $\omega$  分别设置为 1.2, 0.1, 1, 1, 0.1; $\lambda_1$  分别设置为 0.001, 0.3, 0.001, 0.01, 0.01; $\lambda_2$  分别设置为 0.1, 10, 1, 1, 1.

由实验参数的设置可以看出,式(7)的第 1 项与第 4 项对 FLRS DLSC 的贡献几乎同等重要,说明获得结构化(特定类)字典和防止病态降维的重要性;式(7)的第 2,3,5 项对 FLRS DLSC 的贡献相对较小.

## 4.1 AR 人脸数据集

AR 数据库由 126 个个体的 4 000 多幅图像组成,这些图像在照明、表情和配件方面各不相同. 如文献[15]中所述,使用包含 100 个个体中的 1 400 张图像,其中 50 名男性,50 名女性,其人脸图像只存在表情和光照问题. 图 3 展示 AR 数据集中第 1 个人的样本图像. 每一个个体随机选择 7 张图片用于训练,剩下的 7 张图片用于测试. 重复实验 10 次以计算识别率的平均值和相应的标准偏差. 所有实验人脸图像大小调整为  $60 \times 43$ . 在 AR 数据集上的对比实验中,特征的维度通过 PCA 降至 300.



Fig. 3 Sample images of the first subject on AR dataset

图 3 AR 数据集第 1 个人的样本图像

表 3 展示不同的方法的识别率. 在 SRC, LCKSVD, FDDL 和 LRS DL 字典学习方法中, LRS DL 达到最佳性能为 97.34%, 比 FLRS DLSC 方法低 0.5% 左右. 基于特征和字典联合学习的 SEDL 方法获得了第 3 高的识别率为 94.21%, 比 LRS DL 方法低 3.1% 左右, 可以得出样本间共享信息的重要性, 比 FLRS DLSC 方法低 3.7%, 这是因为未考虑样本中共享信息. 实验验证我们的方法在光照、表情变化下人脸识别具有鲁棒性.

图 4 展示了不同维度的识别率. PCA+FDDL, PCA+LRS DL, JDDRDL, SEDL 作为基准线进行比较. FLRS DLSC 方法与其他方法相比在不同维度下

都实现了最佳性能,并且在低维特征的人脸识别上仍然能获得高分性能,说明在表情和光照下的人脸图像上具有良好的鲁棒性. SEDL 和 FLRSDLSC 方法在特征维数为 450~600 时识别率缓慢增加,到 600 维时开始降低,这说明增加维数对字典学习方法可能无效.

**Table 3 Recognition Rate on AR Dataset**

**表 3 在 AR 数据集下的识别率**

Method	REC±STD/%
SRC	90.13±0.81
LCKSVD	89.02±0.75
FDDL	92.21±0.56
LRSDDL	97.34±0.73
SEDL	94.21±0.51
FLRSDLSC*	97.76±0.98
FLRSDLSC	97.91±0.57

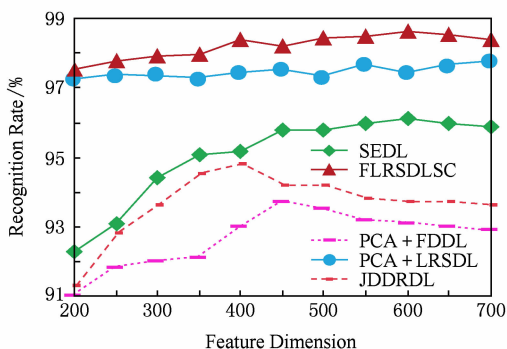


Fig. 4 Recognition rate under different dimensions of features on AR dataset

图 4 AR 数据集下不同特征维度的识别率

表 4 展示用不同数量的样本(4~7)进行训练获得的识别率. FLRSDLSC 方法与 3 种经典方法(FDDL, JDDRDL, SEDL)进行比较,在不同数量的训练样本下都实现了最好的性能. 当样本为 4 时,

**Table 4 Recognition Rate with Different Numbers of Training Samples on AR Dataset**

**表 4 AR 数据集下不同数量的训练样本的识别率**

Number	REC±STD/%			
	FDDL	JDDRDL	SEDL	FLRSDLSC
4	78.5±1.5	80.7±1.7	80.3±1.7	96.1±0.7
5	87.8±1.1	88.6±1.1	88.9±1.2	97.1±0.7
6	91.1±0.9	91.5±0.9	92.3±0.8	97.6±0.4
7	92.2±0.5	94.0±0.5	94.1±0.6	97.9±0.6

FLRSDLSC 方法在识别率上比 SEDL 高出 15% 左右,并且实现 96.1% 的准确率,说明对小样本的人脸识别也是有效的. 当样本为 5~7 时,识别率增长地缓慢,但仍然要比其他方法要高至少 3%.

图 5 绘制了 FLRSDLSC 的收敛图. 如图 5 所示,迭代次数从 1 增加到 100,目标函数值从 18.9 下降到 18.62 且函数值在每次迭代中都下降并最终收敛.

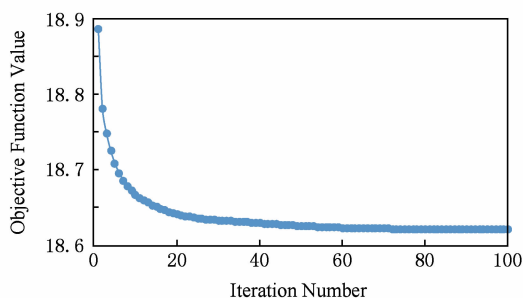


Fig. 5 The convergence of FLRSDLSC on AR dataset

图 5 在 AR 数据集下 FLRSDLSC 的收敛性

#### 4.2 Extended Yale B 数据集

Extended Yale B 数据库包含 38 个个体在各种光照条件下的 2414 张图像. 图 6 展示第 1 个个体在各种照明条件下的样本图像. 对于每个个体,随机选择 20 张图像进行训练,其余用于测试. 实验重复 10 次,所有人脸图像大小都调整为  $54 \times 48$ . 在 Extended Yale B 数据库的对比实验中,特征的维度通过 PCA 降至 500.



Fig. 6 Sample images of the first subject from Extended Yale B dataset

图 6 Extended Yale B 数据集中第 1 个人的样本图像

表 5 展示了不同算法下的识别率. 在 SRC, LCKSVD, FDDL 和 LRSDDL 字典学习方法中, LRSDDL 达到最佳性能为 95.23%, 但比基于联合学习的 SEDL 模型低了 1.3% 左右. FLRSDLSC\* 的识别率高于 SEDL 大约 0.2%, 说明样本间共享信息对字典学习的重要性. FLRSDLSC 方法比 FLRSDLSC\* 高出 0.3% 左右, 这是因为降维时能获得更为紧凑的特征.

图 7 展现了不同维度下 the Extended Yale B 数据的识别率, 对比方法有 PCA + FDDL, PCA +

LRSDDL, JDDRDL, SEDL. 在 200 ~ 400 维时, FLRSDLSC 方法比其他算法有显著的提升,再次说明其算法在低维特征人脸识别的有效性. 当特征维度为 700 时,识别率达到最高,说明提高特征维数对字典学习起到了促进作用. 与 AR 数据集不同的是 SEDL 在不同维度下的识别率比 LRSDDL 高,展示在该数据集下 SEDL 通过降维和字典学习同时进行的方法能获得更重要的特征,从而加强字典的判别能力.

Table 5 Recognition Rate on Extended Yale B Dataset

表 5 在 Extended Yale B 数据集下的识别率

Method	REC±STD/%
SRC	90.21±0.68
LCKSVD	88.44±0.97
FDDL	94.41±0.64
LRSDDL	95.23±0.67
SEDL	96.60±0.62
FLRSDLSC*	96.84±0.64
FLRSDLSC	97.15±0.64

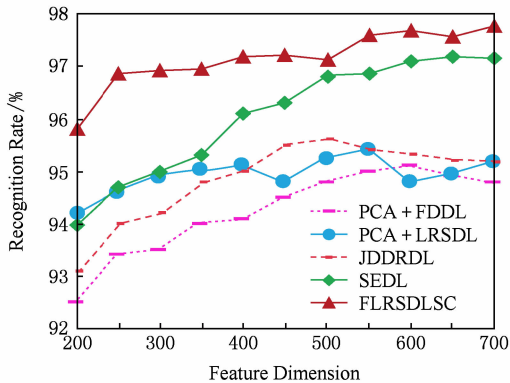


Fig. 7 Recognition rate under different dimensions of features on Extended Yale B dataset

图 7 Extended Yale B 数据集下不同特征维度的识别率

表 6 展示用不同数量的样本进行训练获得的识别率,FLRSDLSC 方法实现了最好的实验效果,对比方法有 PCA+FDDL, JDDRDL, SEDL. 当训练样本为 5 时, JDDRDL 获得第 2 高的性能,但比 FLRSDLSC 方法少 8%,说明了 FLRSDLSC 方法对于光照、小样本问题的人脸识别具有鲁棒性. 在不同数量的训练样本下, FLRSDLSC 都实现了最高的性能. 当训练样本为 20 时, FLRSDLSC 只比 SEDL 高出 0.6%,展示了从样本间提取共享信息的重要性.

Table 6 Recognition Rate with Different Number of Training Samples on Extended Yale B Dataset

表 6 Extended Yale B 数据集下不同数量的训练样本的识别率

Number	REC±STD/%			
	FDDL	JDDRDL	SEDL	FLRSDLSC
5	68.6±2.0	70.4±1.9	70.3±1.7	78.4±1.4
10	83.5±1.8	85.1±1.5	86.2±1.2	90.9±1.0
20	94.4±0.6	95.7±0.6	96.6±0.6	97.2±0.4

### 4.3 CMU PIE 数据集

CMU PIE 数据集包含 68 个个体的 41 368 张图像,图像中含有不同的姿态、光照、表情等问题. 如文献[15]所述,取 CMU PIE 数据集的子集(C05, C07, C09, C27, C29)并且每个个体有 170 张图片. 图 8 展示第 1 个个体不同姿态下的样本图像. 每个个体随机挑选 20 张图片进行训练,剩下的图片作为测试样本. 重复实验 10 次计算识别率的平均值和相应的标准偏差. 所有人脸图像大小调整成 60×45. 在 CMU PIE 数据库的对比实验中,特征的维度通过 PCA 降至 500.



Fig. 8 Sample images of the first subject from CMU PIE dataset

图 8 CMU PIE 数据集第 1 个人的人脸图像

表 7 显示我们的方法实现了最佳性能,识别率为 94.21%,比 SEDL 高出约 0.8%,说明样本间的共享信息增强了在姿态、光照、表情问题下的人脸识别. LRSDDL 与 SEDL 的性能基本相同,意味着样本间的共享信息与降维和字典学习同时进行的方式都能对鉴别性字典的学习起到促进作用.

图 9 展示了不同维度下 CMU PIE 数据集的识别率,对比方法有 PCA+FDDL, PCA+LRSDDL, JDDRDL, SEDL. 我们的方法在 200, 250 维下的识别率要小于 PCA+LRSDDL 的实验结果,但在其他维度上都要好于其他算法,展示了姿态问题对降维和字典同时进行学习的影响. PCA+LRSDDL 的性能整体上要优于 SEDL 和 JDDRDL,说明 LRSDDL 模型



中样本间的共享信息对于解决人脸识别的姿态问题有着重要作用。

Table 7 Recognition Rate on CMU PIE Dataset

表 7 CMU PIE 数据集下的识别率

Method	REC±STD/%
SRC	93.02±0.41
LCKSVD	81.46±0.22
FDDL	92.33±0.51
LRSDL	93.34±0.26
SEDL	93.40±0.42
FLRSDLSC*	93.72±0.34
FLRSDLSC	94.21±0.11

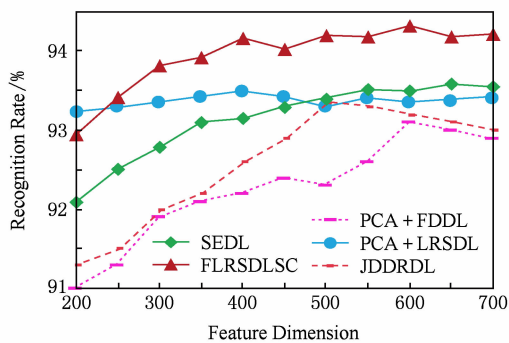


Fig. 9 Recognition rate under different dimensions of features on the CMU PIE dataset

图 9 CMU PIE 数据集下不同特征维度的识别率

#### 4.4 FERET 数据集

FERET 数据库由 14 051 张具有不同姿态、光照和表情的图像组成。如文献[15]所述,使用 200 个个体中带有“ba”,“bj”和“bk”的图像,即使用 600 张图像进行实验。图 10 展示了 FERET 数据集的部分样本,每个个体使用“ba”和“bj”的图像作为训练集,剩下的一张图片作为测试集。所有图像大小调整成  $70 \times 60$ 。在 FERET 数据库的对比实验中,特征的维度通过 PCA 降至 400。

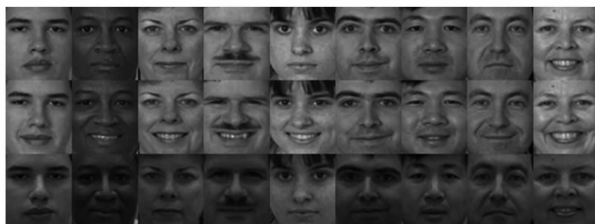


Fig. 10 Sample images from FERET dataset

图 10 FERET 数据集样本图像

表 8 展示了在 FERET 数据集下的识别率。SEDL 实现第二高性能比 FLRSDLSC\* 高出 0.5%,它保护了投影矩阵的正交性,获得了更紧凑的特征。FLRSDLSC 实现了最高的识别率,展示了该方法在光照、表情下人脸识别的鲁棒性。

Table 8 Recognition Rate with Different Number of Training Samples on FERET Dataset

表 8 FERET 数据集下不同数量的训练样本的识别率

Method	REC±STD/%
SRC	79.0±0.51
LCKSVD	74.0±0.23
FDDL	74.0±0.67
LRSDL	77.0±0.56
SEDL	82.5±1.21
FLRSDLSC*	82.0±0.16
FLRSDLSC	83.0±0.78

图 11 展示了不同维度下 FERET 数据集的识别率,对比方法有 PCA+FDDL, PCA+LRSDDL, JDDRDL, SEDL。与 AR, Extended Yale B 数据集相似,FLRSDLSC 在光照、表情下的人脸识别都要高于其他方法且在 200 维时高于其他方法 5%,说明该方法适合解决光照、表情下的小样本问题。

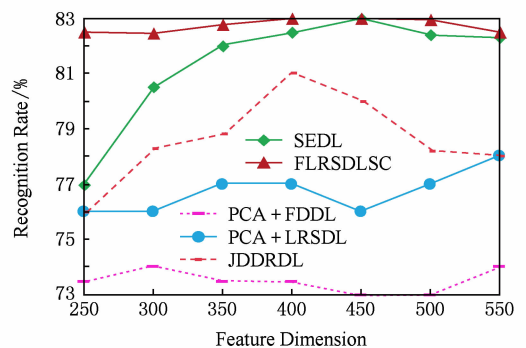


Fig. 11 Recognition rate under different dimensions of features on FERET dataset

图 11 FERET 数据集下不同特征维度的识别率

#### 4.5 AR 性别数据集

AR 数据集通常用于性别分类的研究。如文献[15]所述,使用 1 400 张图像进行实验,其中 50 名男性和 50 名女性。我们选择随机选择 25 名男性和 25 名女性的图像进行训练,其余图像进行测试,实验重复 10 次。所有图像大小调整为  $60 \times 43$ 。在 AR 性别数据集下的对比实验中,特征的维度通过 PCA 降至 300。

表 9 展示 FLRSDLSC 在 AR 性别数据集下取

得最高的识别率并且高于 SEDL 方法,说明了该方法对光照、表情问题下的人脸识别具有鲁棒性。

**Table 9 Recognition Rate with Different Numbers of Training Samples on AR Gender Dataset**

**表 9 AR 性别数据集下不同数量的训练样本的识别率**

Method	REC±STD/%
SRC	87.09±0.93
LCKSVD	88.44±0.97
FDDL	92.91±1.84
LRSDL	93.21±0.57
SEDL	93.90±2.11
FLRSDLSC*	93.80±1.21
FLRSDLSC	94.13±1.28

## 5 总 结

本文提出了一种稀疏约束下快速低秩共享的字典学习(FLRSDLSC)方法.在字典学习阶段,嵌入 Fisher 判别准则来获得结构化字典,同时嵌入低秩约束获得低秩共享字典;在特征学习阶段,利用 Cayley 变换保护投影矩阵的正交性来获得紧凑的特征信息.最后,特征和字典进行联合学习促使字典获得更为重要的特征信息,以此增强字典和稀疏编码的判别能力.实验结果表明该方法在表情变化下的人脸识别具有很强的鲁棒性,并对光照起到了抑制作用,尤其适合解决光照、表情变化下的小样本问题。

## 参 考 文 献

- [1] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2002, 19(7): 711-720
- [2] He Xiaofei, Niyogi P. Locality preserving projections [J]. *Advances in Neural Information Processing Systems*, 2004, 16(1): 186-197
- [3] Turk M, Pentland A. Eigenfaces for recognition [J]. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86
- [4] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition [J]. *Proceedings of the IEEE*, 2010, 98(6): 1031-1044
- [5] Peng Yigang, Suo Jinli, Dai Qionghai, et al. From compressed sensing to low rank matrix recovery: Theory and application [J]. *Acta Automatica Sinica*, 2013, 39(7): 981-994 (in Chinese)
- [6] Mousavi H S, Monga V, Tran T D. Iterative convex refinement for sparse recovery [J]. *IEEE Signal Processing Letters*, 2015, 22(11): 1903-1907
- [7] Spratling M W. Image segmentation using a sparse coding model of cortical area V1 [J]. *IEEE Trans on Image Processing*, 2013, 22(4): 1631-1643
- [8] Aharon M, Elad M, Bruckstein A. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation [M]. Piscataway, NJ: IEEE, 2006: 4311-4322
- [9] Zhang Qiang, Li Baoxin. Discriminative K-SVD for dictionary learning in face recognition [C]//*Proc of IEEE Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE, 2010: 2691-2698
- [10] Jiang Zhuolin, Lin Zhe, Davis L S. Label consistent K-SVD: Learning a discriminative dictionary for recognition [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2651-2664
- [11] Yang Meng, Zhang Lei, Feng Xiangchu, et al. Fisher discrimination dictionary learning for sparse representation [C]//*Proc of 2011 IEEE Int Conf on Computer Vision (ICCV2011)*. Piscataway, NJ: IEEE, 2011: 543-550
- [12] Vu T H, Monga V. Fast low-rank shared dictionary learning for image classification [J]. *IEEE Trans on Image Processing*, 2017, 26(11): 5160-5175
- [13] Nguyen H V, Patel V M, Nasrabadi N M, et al. Sparse Embedding: A Framework for Sparsity Promoting Dimensionality Reduction [C]//*Proc of European Conf on Computer Vision (ECCV 2012)*. Berlin: Springer, 2012: 414-427
- [14] Feng Zhizhao, Yang Meng, Zhang Lei, et al. Joint discriminative dimensionality reduction and dictionary learning for face recognition [J]. *Pattern Recognition*, 2013, 46(8): 2134-2143
- [15] Chen Yefei, Su Jianbo. Sparse embedded dictionary learning on face recognition [J]. *Pattern Recognition*, 2017 (64): 51-59
- [16] Wen Zaiwen, Yin Wotao. A feasible method for optimization with orthogonality constraints [J]. *Mathematical Programming*, 2013, 142(1/2): 397-434
- [17] Natarajan B K. Sparse approximate solutions to linear systems [J]. *SIAM Journal on Computing*, 2006, 24(2): 227-234
- [18] Cai Jianfeng, Candès E J, Shen Zuowei. A singular value thresholding algorithm for matrix completion [J]. *SIAM Journal on Optimization*, 2008, 20(4): 1956-1982
- [19] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems [J]. *SIAM Journal on Imaging Sciences*, 2009, 2(1): 183-202

[20] Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding [J]. Journal of Machine Learning Research, 2009, 11(1): 19-60

[21] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. Foundations & Trends in Machine Learning, 2010, 3(1): 1-122

[22] Martinez A M. The AR face database [J]. CVC Technical Report, 1998, 24

[23] Georgiades A S, Belhumeur P N, Kriegman D J. From few to many: Illumination cone models for face recognition under variable lighting and pose [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(6): 643-660

[24] Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression database [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(12): 1615-1618

[25] Phillips P J, Moon H, Rizvi S A, et al. The FERET evaluation methodology for face-recognition algorithms [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(10): 1090-1104



**Tian Ze**, born in 1994. Master candidate. His main research interests include machine learning, pattern recognition and computer vision.



**Yang Ming**, born in 1964. PhD, professor. Member of CCF. His main research interests include machine learning, pattern recognition, image processing and computer vision.



**Li Aishi**, born in 1994. Master candidate. His main research interests include machine learning and computer vision (liamgsal@gmail.com).

附录

附录 A 证明引理 1.

先求  $f(\mathbf{Y}, \mathbf{D}, \mathbf{X})$  和  $g(\mathbf{X})$  的梯度.

$$f(\mathbf{Y}, \mathbf{D}, \mathbf{X}) = \sum_{c=1}^C r(\mathbf{Y}_c, \mathbf{D}, \mathbf{X}_c) = \sum_{c=1}^C (\|\mathbf{Y}_c - \mathbf{D}\mathbf{X}_c\|_F^2 + \|\mathbf{Y}_c - \mathbf{D}_c\mathbf{X}_c\|_F^2 + \sum_{j \neq c} \|\mathbf{D}_j\mathbf{X}_c\|_F^2)$$

的梯度,其函数重写如下:

$$f(\mathbf{Y}, \mathbf{D}, \mathbf{X}) = \sum_{c=1}^C r(\mathbf{Y}_c, \mathbf{D}, \mathbf{X}_c) = \|\hat{\mathbf{Y}} - \hat{\mathbf{D}}\mathbf{X}\|_F^2.$$

其中,

$$\hat{\mathbf{Y}} = \begin{pmatrix} \mathbf{Y}_1 & \mathbf{Y}_2 & \cdots & \mathbf{Y}_C \\ \mathbf{Y}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Y}_C \end{pmatrix}, \quad \hat{\mathbf{D}} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_C \\ \mathbf{D}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{D}_C \end{pmatrix}. \quad (\text{A1})$$

然后我们获得

$$\frac{\partial \frac{1}{2} f_{\mathbf{Y}, \mathbf{D}}(\mathbf{X})}{\partial \mathbf{X}} = \hat{\mathbf{D}}^T \hat{\mathbf{D}} - \hat{\mathbf{D}}^T \hat{\mathbf{Y}} = \hat{\mathbf{R}}(\mathbf{D}^T \mathbf{D}) - \hat{\mathbf{R}}(\mathbf{D}^T \mathbf{Y}), \quad (\text{A2})$$

当目标函数定义如下:

$$f_{\mathbf{P}, \mathbf{Y}, \mathbf{D}}(\bar{\mathbf{X}}) = \sum_{c=1}^C r(\mathbf{P}, \mathbf{Y}_c, \bar{\mathbf{D}}, \bar{\mathbf{X}}),$$

$f_{\mathbf{P}, \mathbf{Y}, \mathbf{D}}(\bar{\mathbf{X}})$  的求导公式如下:

$$\frac{\partial \frac{1}{2} f_{\mathbf{P}, \mathbf{Y}, \mathbf{D}}(\bar{\mathbf{X}})}{\partial \bar{\mathbf{X}}} = \hat{\mathbf{R}}(\mathbf{D}^T \mathbf{D})\bar{\mathbf{X}} - \hat{\mathbf{R}}(\mathbf{D}^T \bar{\mathbf{Y}}), \quad (\text{A3})$$

其中  $\bar{\mathbf{Y}} = \mathbf{P}\mathbf{Y} - \mathbf{D}_0\mathbf{X}^0$ .

$g$  的梯度  $\mathbf{E}_p^q \in \mathbb{R}^{p \times q}$  为全 1 矩阵,很容易验证如下公式:

$$(\mathbf{E}_p^q)^T = \mathbf{E}_q^p, \quad (\text{A4})$$

$$\mathbf{M}_c = \mathbf{m}_c \mathbf{E}_1^{n_c} = \frac{1}{n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c}, \quad (\text{A5})$$

$$\mathbf{E}_p^q \mathbf{E}_p^r = q \mathbf{E}_p^r, \quad (\text{A6})$$

$$(\mathbf{I} - \frac{1}{p} \mathbf{E}_p^p)(\mathbf{I} - \frac{1}{p} \mathbf{E}_p^p)^T = (\mathbf{I} - \frac{1}{p} \mathbf{E}_p^p). \quad (\text{A7})$$

我们推导出:

$$\mathbf{X}_c - \mathbf{M}_c = \mathbf{X}_c - \frac{1}{n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c} = \mathbf{X}_c (\mathbf{I} - \frac{1}{n_c} \mathbf{E}_{n_c}^{n_c}), \quad (\text{A8})$$

$$\frac{\partial}{\partial \mathbf{X}_c} \frac{1}{2} \|\mathbf{X}_c - \mathbf{M}_c\|_F^2 = \mathbf{X}_c (\mathbf{I} - \frac{1}{n_c} \mathbf{E}_{n_c}^{n_c}) (\mathbf{I} - \frac{1}{n_c} \mathbf{E}_{n_c}^{n_c})^T =$$

$$\mathbf{X}_c (\mathbf{I} - \frac{1}{n_c} \mathbf{E}_{n_c}^{n_c}) = \mathbf{X}_c - \mathbf{M}_c. \quad (\text{A9})$$

因此我们获得:

$$\frac{\partial \frac{1}{2} \sum_{c=1}^C \|\mathbf{X}_c - \mathbf{M}_c\|_F^2}{\partial \mathbf{X}} = \mathbf{X} - \hat{\mathbf{M}}, \quad (\text{A10})$$

其中,  $\mathbf{X}=(\mathbf{X}_1, \dots, \mathbf{X}_c, \dots, \mathbf{X}_C)$ ,  $\hat{\mathbf{M}}=(\mathbf{M}_1, \dots, \mathbf{M}_c, \dots, \mathbf{M}_C)$ .

$\mathbf{M}_c - \mathbf{M}$  我们重写为如下 2 个式子:

$$\begin{aligned} \mathbf{M}_c - \mathbf{M} &= \frac{1}{n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c} - \frac{1}{N} \mathbf{X} \mathbf{E}_N^{n_c} = \\ &= \frac{1}{n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c} - \frac{1}{N} \sum_{j=1}^c \mathbf{X}_j \mathbf{E}_{n_j}^{n_c} = \\ &= \frac{N-n_c}{N n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c} - \frac{1}{N} \sum_{j \neq c} \mathbf{X}_j \mathbf{E}_{n_j}^{n_c} = \\ &= \frac{1}{n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c} - \frac{1}{N} \mathbf{X}_l \mathbf{E}_{n_l}^{n_c} - \frac{1}{N} \sum_{j \neq c} \mathbf{X}_j \mathbf{E}_{n_j}^{n_c} (l \neq c), \end{aligned} \quad (\text{A11})$$

然后推导如下:

$$\begin{aligned} \frac{N-n_c}{N n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c} - \frac{1}{N} \sum_{j \neq c} \mathbf{X}_j \mathbf{E}_{n_j}^{n_c} &\Rightarrow \\ \frac{\partial}{\partial \mathbf{X}_c} \frac{1}{2} \|\mathbf{M}_c - \mathbf{M}\|_{\text{F}}^2 &= \\ \left( \frac{1}{n_c} - \frac{1}{N} \right) (\mathbf{M}_c - \mathbf{M}) \mathbf{E}_{n_c}^{n_c} &= \\ (\mathbf{M}_c - \mathbf{M}) + \frac{1}{N} (\mathbf{M} - \mathbf{M}_c) \mathbf{E}_{n_c}^{n_c}, & \\ \frac{1}{n_c} \mathbf{X}_c \mathbf{E}_{n_c}^{n_c} - \frac{1}{N} \mathbf{X}_l \mathbf{E}_{n_l}^{n_c} - \frac{1}{N} \sum_{j \neq c} \mathbf{X}_j \mathbf{E}_{n_j}^{n_c} (l \neq c) &\Rightarrow \\ \frac{\partial}{\partial \mathbf{X}_l} \frac{1}{2} \|\mathbf{M}_c - \mathbf{M}\|_{\text{F}}^2 = \frac{1}{N} \sum_{c=1}^c (\mathbf{M} - \mathbf{M}_c) \mathbf{E}_{n_c}^{n_l} (l \neq c). & \\ \frac{\partial}{\partial \mathbf{X}_l} \frac{1}{2} \sum_{c=1}^c \|\mathbf{M}_c - \mathbf{M}\|_{\text{F}}^2 = & \\ \mathbf{M}_l - \mathbf{M} + \frac{1}{N} \sum_{c=1}^c (\mathbf{M} - \mathbf{M}_c) \mathbf{E}_{n_c}^{n_l} (l \neq c). & \end{aligned}$$

现在我们证明

$$\begin{aligned} \sum_{c=1}^c (\mathbf{M} - \mathbf{M}_c) \mathbf{E}_{n_c}^{n_l} &= \mathbf{0}, \\ \sum_{c=1}^c (\mathbf{M} - \mathbf{M}_c) \mathbf{E}_{n_c}^{n_l} &= \sum_{c=1}^c (\mathbf{m} \mathbf{E}_1^{n_c} - m_c \mathbf{E}_1^{n_c}) \mathbf{E}_{n_c}^{n_l} = \\ \sum_{c=1}^c (\mathbf{m} - m_c) \mathbf{E}_1^{n_c} \mathbf{E}_{n_c}^{n_l} &= \sum_{c=1}^c n_c (\mathbf{m} - m_c) \mathbf{E}_1^{n_l} = \\ \left( \sum_{c=1}^c n_c \mathbf{m} - \sum_{c=1}^c n_c m_c \right) \mathbf{E}_1^{n_l} &= \mathbf{0}, \quad (\text{A12}) \end{aligned}$$

其中,  $\mathbf{m} = \frac{\sum_{c=1}^c n_c \mathbf{m}_c}{\sum_{c=1}^c n_c}$ .

然后我们推导出:

$$\frac{\partial \frac{1}{2} \sum_{c=1}^c \|\mathbf{M}_c - \mathbf{M}\|_{\text{F}}^2}{\partial \mathbf{X}} = \hat{\mathbf{M}} - \mathbf{M}, \quad (\text{A13})$$

其中,  $\hat{\mathbf{M}}=(\mathbf{M}_1, \mathbf{M}_c, \dots, \mathbf{M}_C)$ .

组合(A10), (A13)和  $\frac{\partial \frac{1}{2} \|\mathbf{X}\|_{\text{F}}^2}{\partial \mathbf{X}} = \mathbf{X}$ , 得到:

$$\frac{\partial \frac{1}{2} g(\mathbf{X})}{\partial \mathbf{X}} = 2\mathbf{X} + \mathbf{M} - 2\hat{\mathbf{M}}. \quad (\text{A14})$$

证毕.

附录 B. 证明引理 2.

令  $\mathbf{w}_c \in \{0, 1\}^K$  为二值化向量, 其第  $j$  个元素为 1 当且仅当  $\mathbf{D}$  的第  $j$  列属于  $\mathbf{D}_c$ .  $\mathbf{W}_c = \text{diag}(\mathbf{w}_c)$ , 我们观察  $\mathbf{D}_c \mathbf{X}_c^i = \mathbf{D} \mathbf{W}_c \mathbf{X}_c^i$ . 重写  $f(\mathbf{Y}, \mathbf{X}, \mathbf{D})$  如下:

$$\begin{aligned} &\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_{\text{F}}^2 + \\ &\sum_{c=1}^c (\|\mathbf{Y}_c - \mathbf{D}_c \mathbf{X}_c^c\|_{\text{F}}^2 + \sum_{j \neq c} \|\mathbf{D}_j \mathbf{X}_c^j\|_{\text{F}}^2) = \\ &\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_{\text{F}}^2 + \\ &\sum_{c=1}^c (\|\mathbf{Y}_c - \mathbf{D} \mathbf{W}_c \mathbf{X}_c^c\|_{\text{F}}^2 + \sum_{j \neq c} \|\mathbf{D}_j \mathbf{X}_c^j\|_{\text{F}}^2) = \\ &\text{tr}(\mathbf{X}\mathbf{X}^{\text{T}} + \sum_{c=1}^c \sum_{j=1}^c \mathbf{W}_j \mathbf{X}_c \mathbf{X}_c^{\text{T}} \mathbf{W}_j^{\text{T}}) \mathbf{D}^{\text{T}} \mathbf{D} - \\ &2\text{tr}(\mathbf{Y}\mathbf{X}^{\text{T}} + \sum_{c=1}^c \mathbf{Y}_c \mathbf{X}_c^{\text{T}} \mathbf{W}_c) \mathbf{D}^{\text{T}}) + \text{constant} = \\ &-2\text{tr}(\mathbf{E}\mathbf{D}^{\text{T}}) + \text{tr}(\mathbf{F}\mathbf{D}^{\text{T}} \mathbf{D}) + \text{constant}, \quad (\text{B1}) \end{aligned}$$

其中,

$$\begin{aligned} \mathbf{E} &= \mathbf{Y}\mathbf{X}^{\text{T}} + \sum_{c=1}^c \mathbf{Y}_c \mathbf{X}_c^{\text{T}} \mathbf{W}_c = \\ \mathbf{Y}\mathbf{X}^{\text{T}} + (\mathbf{Y}_1 (\mathbf{X}_1^{\text{C}})^{\text{T}} \dots \mathbf{Y}_c (\mathbf{X}_c^{\text{C}})^{\text{T}}) &= \\ \mathbf{Y} \left[ \mathbf{X}^{\text{T}} + \begin{pmatrix} (\mathbf{X}_1^{\text{C}})^{\text{T}} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & (\mathbf{X}_c^{\text{C}})^{\text{T}} \end{pmatrix} \right] &= \\ \mathbf{Y} \hat{\mathbf{R}} (\mathbf{X})^{\text{T}}, & \quad (\text{B2}) \end{aligned}$$

$$\begin{aligned} \mathbf{F} &= \mathbf{X}\mathbf{X}^{\text{T}} + \sum_{c=1}^c \sum_{j=1}^c \mathbf{W}_j \mathbf{X}_c \mathbf{X}_c^{\text{T}} \mathbf{W}_j^{\text{T}} = \\ \mathbf{X}\mathbf{X}^{\text{T}} + \sum_{j=1}^c \mathbf{W}_j \left( \sum_{c=1}^c \mathbf{X}_c \mathbf{X}_c^{\text{T}} \right) \mathbf{W}_j^{\text{T}} &= \\ \mathbf{X}\mathbf{X}^{\text{T}} + \sum_{j=1}^c \mathbf{W}_j \mathbf{X} \mathbf{X}^{\text{T}} \mathbf{W}_j^{\text{T}}. & \quad (\text{B3}) \end{aligned}$$

令:

$$\mathbf{X}\mathbf{X}^{\text{T}} = \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \dots & \mathbf{A}_{1j} & \dots & \mathbf{A}_{1c} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{A}_{21} & \dots & \mathbf{A}_{jj} & \dots & \mathbf{A}_{2c} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{A}_{c1} & \dots & \mathbf{A}_{cj} & \dots & \mathbf{A}_{cc} \end{pmatrix}. \quad (\text{B4})$$

根据  $\mathbf{W}_j$  的定义, 我们观察到“左乘”矩阵  $\mathbf{W}_j$  迫使矩阵在除了第  $j$  个块行之外的任何地方都为零.

同样,“右乘”矩阵将只保留其  $j$  块列. 从而得到如下结果:

$$W_j A W_j^T = \begin{pmatrix} \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & A_{jj} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix}. \quad (\text{B5})$$

然后  $\mathbf{F} = \mathbf{X}\mathbf{X}^T + \sum_{j=1}^c W_j \mathbf{X}\mathbf{X}^T W_j^T =$

$$\mathbf{A} + \begin{pmatrix} A_{11} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & A_{cc} \end{pmatrix} = \hat{R}(\mathbf{A}) = \hat{R}(\mathbf{X}\mathbf{X}^T). \quad (\text{B6})$$

根据  $f(\mathbf{Y}, \mathbf{X}, \mathbf{D})$  公式的重写过程,  $f(\mathbf{P}, \mathbf{Y}, \bar{\mathbf{D}}, \bar{\mathbf{X}})$  简化如下:

$$f(\mathbf{P}, \mathbf{Y}, \bar{\mathbf{D}}, \bar{\mathbf{X}}) = \|\mathbf{P}\mathbf{Y} - \bar{\mathbf{D}}\bar{\mathbf{X}}\|_F^2 +$$

$$\sum_{c=1}^c (\|\mathbf{P}\mathbf{Y}_c - \mathbf{D}_c \mathbf{X}_c\|_F^2 + \sum_{j \neq c} \|\mathbf{D}_j \mathbf{X}_j\|_F^2) = -2\text{tr}(\mathbf{E}\mathbf{D}^T) + \text{tr}(\mathbf{F}\mathbf{D}^T \mathbf{D}), \quad (\text{B7})$$

其中,

$$\mathbf{E} = \bar{\mathbf{Y}}\hat{R}(\mathbf{X})^T, \quad \bar{\mathbf{Y}} = \mathbf{P}\mathbf{Y} - \mathbf{D}_0 \mathbf{X}^0, \quad (\text{B8})$$

$$\mathbf{F} = \hat{R}(\mathbf{A}) = \hat{R}(\mathbf{X}\mathbf{X}^T). \quad (\text{B9})$$

证毕.

附录 C. 证明引理 3.

当  $\mathbf{P}, \mathbf{Y}, \mathbf{D}, \mathbf{X}$  固定, 我们有

$$J_{P,Y,D,X}(\mathbf{D}_0, \mathbf{X}^0) = \frac{1}{2} \|\mathbf{P}\mathbf{Y} - \mathbf{D}_0 \mathbf{X}^0 - \mathbf{D}\mathbf{X}\|_F^2 +$$

$$\sum_{c=1}^c \frac{1}{2} \|\mathbf{P}\mathbf{Y}_c - \mathbf{D}_c \mathbf{X}_c - \mathbf{D}_0 \mathbf{X}^0\|_F^2 + \lambda_1 \|\mathbf{X}^0\|_1 + \eta \|\mathbf{D}_0\|_* + \text{constant}. \quad (\text{C1})$$

令  $\tilde{\mathbf{Y}} = \mathbf{P}\mathbf{Y} - \mathbf{D}\mathbf{X}$ ,  $\hat{\mathbf{Y}}_c = \mathbf{P}\mathbf{Y}_c - \mathbf{D}_c \mathbf{X}_c$  和  $\hat{\mathbf{Y}} = (\hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2, \dots, \hat{\mathbf{Y}}_c)$ , 式(C1)简化如下:

$$J_{P,Y,D,X}(\mathbf{D}_0, \mathbf{X}^0) = \frac{1}{2} \|\tilde{\mathbf{Y}} - \mathbf{D}_0 \mathbf{X}^0\|_F^2 +$$

$$\frac{1}{2} \|\hat{\mathbf{Y}} - \mathbf{D}_0 \mathbf{X}^0\|_F^2 + \lambda_1 \|\mathbf{X}^0\|_1 +$$

$$\eta \|\mathbf{D}_0\|_* + \text{constant} = \left\| \frac{\tilde{\mathbf{Y}} + \hat{\mathbf{Y}}}{2} - \mathbf{D}_0 \mathbf{X}^0 \right\|_F^2 + \lambda_1 \|\mathbf{X}^0\|_1 + \eta \|\mathbf{D}_0\|_* + \text{constant}. \quad (\text{C2})$$

我们得到:

$$\tilde{\mathbf{Y}} + \hat{\mathbf{Y}} = 2\mathbf{P}\mathbf{Y} - \mathbf{D}\mathbf{X} - (\mathbf{D}_1 \mathbf{X}_1^1, \dots, \mathbf{D}_c \mathbf{X}_c^c) = 2\mathbf{P}\mathbf{Y} - \mathbf{D}\hat{R}(\mathbf{X}). \quad (\text{C3})$$

现在令  $\mathbf{V} = \frac{\tilde{\mathbf{Y}} + \hat{\mathbf{Y}}}{2}$ , 式(C2)简化如下:

$$J_{P,Y,D,X}(\mathbf{D}_0, \mathbf{X}^0) = \|\mathbf{V} - \mathbf{D}_0 \mathbf{X}^0\|_F^2 + \lambda_1 \|\mathbf{X}^0\|_1 + \eta \|\mathbf{D}_0\|_* + \text{constant}. \quad (\text{C4})$$

证毕.