

L1 范数约束正交子空间非负矩阵分解^①

韩 东, 盖 杉

(南昌航空大学 信息工程学院, 南昌 330063)

摘 要: 针对非负矩阵分解 (NMF) 相对稀疏或局部化描述原数据时导致的稀疏能力和程度比较弱的问题, 提出了 L1 范数约束正交子空间非负矩阵分解方法. 通过将 L1 范数约束引入到正交子空间非负矩阵分解的目标函数中, 提升了分解结果的稀疏性. 同时给出累乘迭代规则. 在 UCI、ORL 和 Yale 三个数据库上进行的实验结果表明, 该算法在聚类效果以及稀疏表达方面优于其他算法.

关键词: 非负矩阵分解; 正交性; L1 范数; 稀疏性

引用格式: 韩东, 盖杉. L1 范数约束正交子空间非负矩阵分解. 计算机系统应用, 2018, 27(9): 205-209. <http://www.c-s-a.org.cn/1003-3254/6570.html>

Non-Negative Matrix Factorization on Orthogonal Subspace with L1 Norm Constrains

HAN Dong, GAI Shan

(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

Abstract: In order to solve the problem of unstable sparseness of Non-negative Matrix Factorization (NMF), an improved NMF on orthogonal subspace with L1 norm constraints was proposed. L1 norm constrained was introduced into the objective function of NMF on Orthogonal Subspace (NMFOS), which enhanced the sparsity of the decomposition results. The multiplicative updating procedure was also produced. Experiments on UCI, ORL, and Yale show that this algorithm is superior to other algorithms in clustering and sparse representation.

Key words: Non-negative Matrix Factorization (NMF); orthogonality; L1 norm; sparsity

1 引言

非负矩阵分解 (Non-negative Matrix Factorization, NMF)^[1] 算法因其收敛速度快以及分解后的稀疏分量能够清晰直观地描述原始数据等特点, 在计算机视觉, 文本聚类, 模式识别等领域受到了广泛关注. NMF 本质上是一种基于部分的矩阵分解方法, 能够以非负形式表示原始数据的局部特征.

NMF 将原始非负矩阵 X 分解为两个非负矩阵 WH 的乘积. 分解后的矩阵仅包含非负元素, 并且基向量 W 具有一定的数据局部表示能力, 这使得 NMF 在诸多领域得到广泛运用. 在文献[2]中, Park 等通过人眼过滤和最小化基于 NMF 的重构图像错误率来进行人眼

检测. 考虑到数据集合的内部几何结构, 文献[3]通过最近邻图来刻画数据集中相邻数据点的关系, 提出了图正则化非负矩阵分解. 为了充分利用判别信息, 同时考虑到数据中的几何结构, 文献[4]提出了 K 近邻非负矩阵分解 (NMF-K-NN). 在此基础上, Jun Ye 等使用模糊集来处理模式识别中的不确定因素, 提出了模糊 K 近邻非负矩阵分解 (NMF-FK-NN) 方法^[5]. Zhang 等^[6] 通过最小化约束梯度距离, 提出保持拓扑性非负矩阵分解 (TPNMF), 该方法能够保持脸部空间的局部内在拓扑结构. 在研究聚类问题的过程中, Yang 等指出^[7], 正交性的约束能在很大程度上优化聚类效果, 其本质是施加正交性约束后的 NMF 结果更加稀疏, 从而使原始

① 基金项目: 国家自然科学基金 (61563037); 江西省自然科学基金 (20171BAB202018)

Foundation item: National Natural Science Foundation of China (61563037); Natural Science Foundation of Jiangxi Province (20171BAB202018)

收稿时间: 2018-01-29; 修改时间: 2018-02-27; 采用时间: 2018-03-27; csa 在线出版时间: 2018-08-16

数据的基之间区别性增强,进而提升聚类效果. Li 等^[8]提出基于正交子空间的非负矩阵分解 (Non-negative Matrix Factorization on Orthogonal Subspace, NMFOS), 将对 \mathbf{W} (或 \mathbf{H}) 的正交性约束作为 NMF 目标函数中的一部分直接进行优化, 减少因施加正交性约束而带来的巨大计算量, 同时还能在一定程度上提升基矩阵 \mathbf{W} (系数矩阵 \mathbf{H}) 的稀疏性.

基于正交子空间的非负矩阵分解虽然能在一定程度上提升分解矩阵的稀疏性, 但是它导致的稀疏程度是难以控制的. 本文为了在分解过程中进一步提升分解矩阵的稀疏性, 在分解过程中引入了 L1 范数约束, 将 L1 范数约束转换成目标函数的正则部分进行求解, 提出了 L1 范数约束正交子空间非负矩阵分解 (Non-negative Matrix Factorization on Orthogonal Subspace with L1 norm constrains, NMFOS-L1). 本文方法不仅能提升聚类效果, 同时还提升了分解结果的稀疏表达能力, 具有实用价值.

2 非负矩阵分解

给定非负矩阵 $\mathbf{X} = [x_1, x_2, \dots, x_n] \in R_+^{m \times n}$, NMF 将原始矩阵分解为两个非负低秩矩阵 \mathbf{W} 和 \mathbf{H} , 即:

$$\mathbf{X} \approx \mathbf{W}_{m \times r} \mathbf{H}_{r \times n} \quad (1)$$

其中, $r \ll \min\{m, n\}$. NMF 常采用欧氏距离衡量 \mathbf{WH} 对 \mathbf{X} 的逼近程度, 目标函数如下:

$$\begin{aligned} \min f(\mathbf{W}, \mathbf{H}) &= \|\mathbf{X} - \mathbf{WH}\|_F^2 \\ \text{s.t. } \mathbf{W}, \mathbf{H} &\geq 0 \end{aligned} \quad (2)$$

式中, $\|\cdot\|_F$ 为 Frobenius 范数, 矩阵 \mathbf{W} 的每一列称作基向量, 矩阵 \mathbf{H} 每一列为系数向量, 将基向量进行线性组合来表示原始数据矩阵. Lee 和 Seung^[9] 给出如下乘性迭代规则:

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{(\mathbf{W}^T \mathbf{X})}{(\mathbf{W}^T \mathbf{WH})} \quad (3)$$

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{(\mathbf{XH}^T)}{(\mathbf{WHH}^T)} \quad (4)$$

式中, \otimes 为矩阵元素的乘积运算符号, 交替进行式 (3) 和式 (4), 可以求得式 (2) 的系数矩阵和基矩阵.

3 基于正交子空间的非负矩阵分解

NMFOS 将分解所得矩阵的正交性约束通过拉格朗日乘子引入到矩阵分解的目标函数中进行优化, 从

而使分解结果的正交性不必通过正交性约束完成, 减少计算量. NMFOS 的目标函数如下: 对矩阵 \mathbf{W} 加入正交性约束, 目标函数为:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} J_{\text{NMFOS}} &= \|\mathbf{X} - \mathbf{WH}^T\|_F^2 + \lambda \|\mathbf{W}^T \mathbf{W} - \mathbf{I}\|_F^2 \\ \text{s.t. } \mathbf{W}, \mathbf{H} &\geq 0 \end{aligned} \quad (5)$$

对矩阵 \mathbf{H} 加入正交性约束, 目标函数为:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} J_{\text{NMFOS}} &= \|\mathbf{X} - \mathbf{WH}^T\|_F^2 + \lambda \|\mathbf{HH}^T - \mathbf{I}\|_F^2 \\ \text{s.t. } \mathbf{W}, \mathbf{H} &\geq 0 \end{aligned} \quad (6)$$

其中, $\lambda \geq 0$ 为正则参数, \mathbf{I} 是全 1 矩阵. 对于式 (5) 和式 (6), Li 等^[8] 给出了如下的乘性迭代规则:

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{\mathbf{XH} + 2\lambda \mathbf{W}}{\mathbf{WH}^T \mathbf{H} + 2\lambda \mathbf{W}^T \mathbf{W}} \quad (7)$$

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T \mathbf{X}}{\mathbf{W}^T \mathbf{WH}^T} \quad (8)$$

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{\mathbf{XH}}{\mathbf{WH}^T \mathbf{H}} \quad (9)$$

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T \mathbf{X} + 2\lambda \mathbf{H}^T}{\mathbf{W}^T \mathbf{WH}^T + 2\lambda \mathbf{H}^T \mathbf{HH}^T} \quad (10)$$

4 L1 范数约束正交子空间非负矩阵分解

NMF 算法的分解结果在一定程度上呈现稀疏性, 但是稀疏程度难以控制. Hoyer 于 2004 年提出稀疏性非负矩阵分解^[10], 在目标函数上添加 L1 正则化的稀疏约束. 如果对 NMFOS 加上正则化的稀疏约束, 那么就可以得到更加稀疏的分解矩阵, 从而提高分解质量.

通过引入稀疏约束条件到 NMFOS 的目标函数, 将稀疏约束正交子空间非负矩阵分解归结为下列优化问题: 对矩阵 \mathbf{W} 而言, 目标函数为:

$$\begin{aligned} J_{\text{NMFOS SC}} &= \|\mathbf{X} - \mathbf{WH}\|_F^2 + \lambda \|\mathbf{W}^T \mathbf{W} - \mathbf{I}\|_F^2 + \alpha \|\mathbf{W}\|_1 \\ \text{s.t. } \mathbf{W} &\geq 0, \mathbf{H} \geq 0 \end{aligned} \quad (11)$$

对矩阵 \mathbf{H} 而言, 目标函数为:

$$\begin{aligned} J_{\text{NMFOS SC}} &= \|\mathbf{X} - \mathbf{WH}\|_F^2 + \lambda \|\mathbf{HH}^T - \mathbf{I}\|_F^2 + \beta \|\mathbf{H}\|_1 \\ \text{s.t. } \mathbf{W} &\geq 0, \mathbf{H} \geq 0 \end{aligned} \quad (12)$$

式中, λ, α, β 均为大于 0 的常数. 利用最速下降法和粒子迭代法, 推导出上式的乘性迭代规则; 首先新的目标函数可表示为:

$$J_{\text{NMFOSSC}} = \text{Tr} \left((\mathbf{X} - \mathbf{W}\mathbf{H}^T)(\mathbf{X} - \mathbf{W}\mathbf{H}^T)^T \right) + \lambda \text{Tr} \left((\mathbf{W}^T\mathbf{W} - \mathbf{I})(\mathbf{W}^T\mathbf{W} - \mathbf{I})^T \right) + \alpha \|\mathbf{W}\|_1 \quad (13)$$

由 Lagrange 定理:

$$J_{\text{NMFOSSC}} = \text{Tr} \left((\mathbf{X} - \mathbf{W}\mathbf{H}^T)(\mathbf{X} - \mathbf{W}\mathbf{H}^T)^T \right) + \lambda \text{Tr} \left((\mathbf{W}^T\mathbf{W} - \mathbf{I})(\mathbf{W}^T\mathbf{W} - \mathbf{I})^T \right) + \alpha \|\mathbf{W}\|_1 + \text{Tr}(\Psi\mathbf{W}^T) + \text{Tr}(\Phi\mathbf{H}^T) \quad (14)$$

式中, $\Psi = [\psi_{ik}]$ 和 $\Phi = [\varphi_{jk}]$ 分别为约束 $\mathbf{W} \geq 0$ 和 $\mathbf{H} \geq 0$ 的拉格朗日乘子. 对式 (13) 分别关于 \mathbf{W} 和 \mathbf{H} 进行偏微分, 结果为:

$$\frac{\partial J}{\partial \mathbf{W}} = \mathbf{X}\mathbf{H}^T + \mathbf{W}\mathbf{H}\mathbf{H}^T - 2\lambda\mathbf{W} + 2\lambda\mathbf{W}\mathbf{W}^T\mathbf{W} + \alpha\mathbf{I} \quad (15)$$

$$\frac{\partial J}{\partial \mathbf{H}} = \Delta\mathbf{W}^T\mathbf{X} + \mathbf{W}^T\mathbf{W}\mathbf{H} \quad (16)$$

式中, \mathbf{I} 是全 1 矩阵. 由 KKT 最优条件 $\varphi_{ik}\mathbf{W}_{ik} = 0$ 和 $\psi_{jk}\mathbf{H}_{jk} = 0$, 得 \mathbf{W} 和 \mathbf{H} 的乘性迭代规则为:

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{\mathbf{X}\mathbf{H}^T + 2\lambda\mathbf{W}}{\mathbf{W}\mathbf{H}^T\mathbf{H} + 2\lambda\mathbf{W}\mathbf{W}^T\mathbf{W} + \alpha\mathbf{I}} \quad (17)$$

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T\mathbf{X}}{\mathbf{W}^T\mathbf{W}\mathbf{H}} \quad (18)$$

同理可得式 (12) 的矩阵更新规则为:

$$\mathbf{W} \leftarrow \mathbf{W} \otimes \frac{\mathbf{X}\mathbf{H}^T}{\mathbf{W}\mathbf{H}\mathbf{H}^T} \quad (19)$$

$$\mathbf{H} \leftarrow \mathbf{H} \otimes \frac{\mathbf{W}^T\mathbf{X} + 2\lambda\mathbf{H}}{\mathbf{W}^T\mathbf{W}\mathbf{H} + 2\lambda\mathbf{H}\mathbf{H}^T\mathbf{H} + \beta\mathbf{I}} \quad (20)$$

基于上述的乘性迭代规则, 可将 NMFOSS-L1 算法归结为:

步骤 1. 初始化: 输入矩阵 \mathbf{X} , 初始化矩阵 \mathbf{W} 和 \mathbf{H} 为非负矩阵;

步骤 2. 迭代更新: 将矩阵 \mathbf{W} 和 \mathbf{H} 带入式 (17)~(20), 解得 \mathbf{W} 和 \mathbf{H} ;

步骤 3. 终止条件: 若 $\|\mathbf{X} - \mathbf{W}\mathbf{H}\|_F^2$ 小于阈值或超出预定的迭代次数, 则算法终止; 否则返回步骤 3.

5 实验与结果分析

为了验证 NMFOSS-L1 算法有效性, 本文在手写体数字光学识别数据集 (Optical Recognition of Handwriting Digits)^[11]、ORL 人脸数据库^[12]和 Yale 人脸数据库^[13]进行了聚类的对比实验. 同时, 为了验证本文算法所得

到的基矩阵的稀疏性, 在 ORL 和 Yale 人脸数据库进行实验, 比较了几种不同算法的稀疏表达能力.

手写体数字光学识别数据集: 该数据集从 UCI 数据库中选取 0, 2, 4, 6 几个数字, 构成 2237 个样本, 每个样本有 62 特征, 分为 4 个类. ORL 人脸数据库^[12]是由 40 个人, 每人 10 幅图像构成. 每幅图像为 256 个灰度级, 分辨率为 112×92 . 该库的人脸图像表情变化, 面部细节, 以及拍摄角度变化较大. 图 1 为 ORL 人脸库同一个人的 10 张图像.



图 1 ORL 人脸数据库

Yale 人脸库^[13]包含 15 个人每人 11 幅共 165 幅人脸图像, 这些照片在不同的光照条件和角度下拍摄, 人脸表情也有较大变化. 每幅图像均为 100×100 像素. 图 2 为 Yale 同一个人的 10 张图像.



图 2 Yale 人脸数据库

5.1 聚类实验

在聚类问题中, 常见的评测指标是纯度和 F 值. 本文在已知类标签情况下, 将不同算法的聚类结果进行对比, 利用纯度来评价不同算法产生的分类效果. 纯度: 所有簇的纯净度的均值. 范围为 $[0, 1]$, 数值越大, 纯净度越高, 效果越好. 定义式为:

$$\text{Purity} = \frac{1}{n} \sum_{k=1}^P \max_{1 \leq l \leq q} n_k^l \quad (21)$$

式中, q 为总的类数, n_k^l 是簇 k 中标记为类 l 的个数. 聚类熵: 度量各簇中所有类的分布情况. 取值范围为 $[0, 1]$, 取值越小, 聚类效果越好. 定义如下:

$$\text{Entropy} = \Delta \frac{1}{n \log_2 q} \sum_{k=1}^P \sum_{l=1}^q n_k^l \log_2 \frac{n_k^l}{n_k} \quad (22)$$

式中, $n_k = \sum_l n_k^l$, q 为总得类数.

在本节实验中设定 $P = q$, 在每个数据库独立地重复实验 200 次, 并设定迭代次数的最大值为 2000. 在实验时, 选取参数 $\lambda = 5$, $\beta = 1$. 实验结果如表 1 所示.

表1 三种数据库上的聚类纯度(均值±方差)

数据库	NMF	ONMF	NMFOS	NMFOS-L1
Digit	0.96±0.00	0.96±0.00	0.93±0.01	0.93±0.01
ORL	0.44±0.03	0.68±0.02	0.71±0.02	0.76±0.03
Yale	0.40±0.04	0.61±0.03	0.66±0.02	0.71±0.02

表2 三种数据库上的聚类熵(均值±方差)

数据库	NMF	ONMF	NMFOS	NMFOS-L1
Digit	0.10±0.00	0.10±0.00	0.18±0.01	0.17±0.01
ORL	0.36±0.03	0.20±0.01	0.18±0.01	0.16±0.02
Yale	0.33±0.02	0.16±0.01	0.15±0.01	0.14±0.02

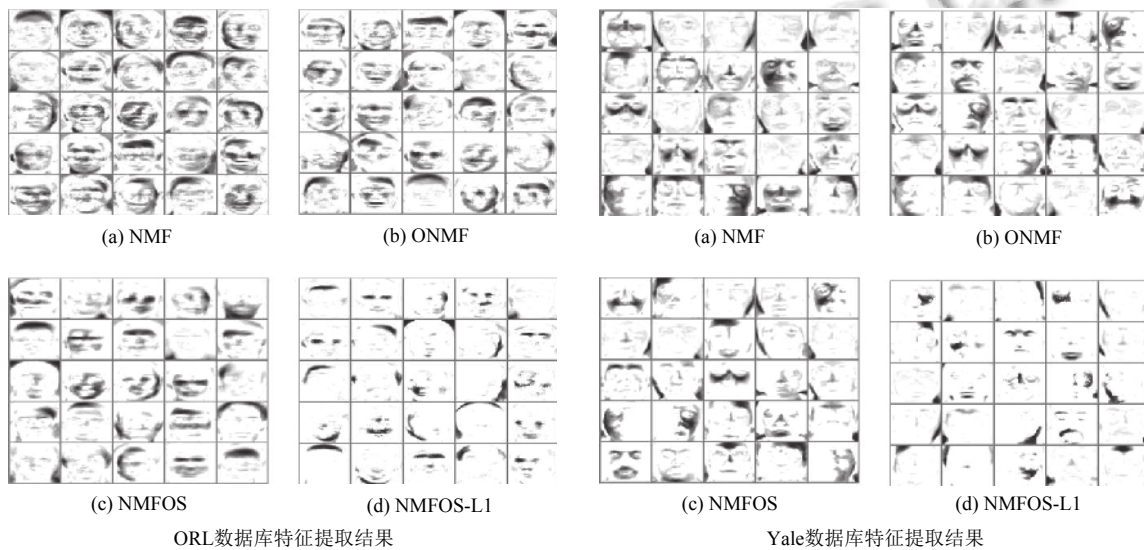


图3 ORL和Yale数据库不同算法人脸特征提取结果对比

Hoyer在文献[9]中给出了度量向量稀疏度的函数:

$$\text{sparse}(x) = \frac{\sqrt{n} - (\sum |x_i|) / \sqrt{\sum x_i^2}}{\sqrt{n} - 1} \quad (23)$$

其中, n 表示 x 的维数. 这个函数的值域为 $[0, 1]$, 取值越小, 向量 x 越稠密, 反之越稀疏, 当该函数取值为1时, 表明非负向量 x 只包含一个非零元素.

实验最后, 我们对矩阵分解结果的稀疏性进行对比. 从表3和表4中我们可以看到, 本文算法所得的基矩阵和稀疏矩阵更加稀疏, 本文算法的稀疏表达能力优于对比的几种算法.

表3 ORL数据库上不同算法的稀疏性

算法	W稀疏性	H稀疏性
NMF	0.35	0.32
ONMF	0.59	0.51
NMFOS	0.67	0.61
NMFOS-L1	0.71	0.62

5.2 稀疏性对比实验

本节我们在ORL和Yale人脸数据库上进行人脸特征提取, 对比了NMF、ONMF、NMFOS、和本文NMFOS-L1几种算法的局部表达能力. 图3给出了秩为25时, 不同算法得到的基矩阵图像.

由图3可以看出, 在这两个数据库上对比这4种算法的基图像稀疏度, NMF稀疏度最低, NMFOS-L1的基图像最为稀疏, 换言之, 该算法具有最优的局部表达能力.

表4 Yale数据库上不同算法的稀疏性

算法	W稀疏性	H稀疏性
NMF	0.33	0.29
ONMF	0.54	0.46
NMFOS	0.61	0.55
NMFOS-L1	0.69	0.59

6 结束语

针对正交子空间非负矩阵分解相对稀疏或局部化描述原数据时导致的稀疏能力和程度比较弱的问题, 本文将稀疏约束引入正交子空间非负矩阵分解的目标函数中, 提出稀疏约束正交子空间非负矩阵分解. 同时给出了迭代公式. 实验证明该算法具有更好的聚类效果以及稀疏表达能力, 在人脸特征提取领域具有应用潜力. 进一步提升正交子空间非负矩阵分解算法效率, 以及将本文方法推广应用到计算机视觉中都是我们进一步要研究的内容.

参考文献

- 1 Lee D, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, 401(6755): 788–791. [doi: [10.1038/44565](https://doi.org/10.1038/44565)]
- 2 Park CW, Park KT, Moon YS. Eye detection using eye filter and minimisation of NMF-based reconstruction error in facial image. *Electronics Letters*, 2010, 46(2): 130–132. [doi: [10.1049/el.2010.3239](https://doi.org/10.1049/el.2010.3239)]
- 3 Cai D, He X, Han J, *et al.* Graph regularized nonnegative matrix factorization for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1548–1560. [doi: [10.1109/TPAMI.2010.231](https://doi.org/10.1109/TPAMI.2010.231)]
- 4 An SC, Yoo J, Choi S. Manifold-respecting discriminant nonnegative matrix factorization. *Pattern Recognition Letters*, 2011, 32(6): 832–837. [doi: [10.1016/j.patrec.2011.01.012](https://doi.org/10.1016/j.patrec.2011.01.012)]
- 5 Ye J, Jin Z. Non-negative matrix factorisation based on fuzzy K nearest neighbour graph and its applications. *Computer Vision*, 2013, 7(5): 346–353. [doi: [10.1049/iet-cvi.2013.0055](https://doi.org/10.1049/iet-cvi.2013.0055)]
- 6 Zhang T, Fang B, Tang YY, *et al.* Topology preserving non-negative matrix factorization for face recognition. *IEEE Transactions on Image Processing*, 2008, 17(4): 574–584. [doi: [10.1109/TIP.2008.918957](https://doi.org/10.1109/TIP.2008.918957)]
- 7 Yang ZR, Laaksonen J. Multi-plicative updates for non-negative projections. *Neurocomputing*, 2007, 71(1–3): 363–373. [doi: [10.1016/j.neucom.2006.11.023](https://doi.org/10.1016/j.neucom.2006.11.023)]
- 8 Li Z, Wu XD, Peng H. Nonnegative matrix factorization on orthogonal subspace. *Pattern Recognition Letters*, 2010, 31(9): 905–911. [doi: [10.1016/j.patrec.2009.12.023](https://doi.org/10.1016/j.patrec.2009.12.023)]
- 9 Lee DD, Seung HS. Algorithms for non-negative matrix factorization. *Advanced in Neural Information Processing Systems*. 2001. 556–562.
- 10 Hoyer PO. Non-negative sparse coding. *Proceedings of Neural Networks for Signal Processing*. 2002. 557–565.
- 11 Yang ZR, OJA E. Linear and nonlinear projective nonnegative matrix factorization. *Signal Processing*, 2007, 87(8): 1904–1916. [doi: [10.1016/j.sigpro.2007.01.024](https://doi.org/10.1016/j.sigpro.2007.01.024)]
- 12 Samaria FS, Harter AC. Parameterisation of a stochastic model for human face identification. *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*. Sarasota, FL, USA. 1994. 138–142. [doi: [10.1109/ACV.1994.341300](https://doi.org/10.1109/ACV.1994.341300)]
- 13 Georghiades AS, Belhumeur PN, Kriegman DJ. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 26(6): 643–660. [doi: [10.1109/34.927464](https://doi.org/10.1109/34.927464)]