

基于矢量表示组合的字典学习性能分析^①



焦健雄¹, 孙利雷^{2,3}, 徐 勇⁴

¹(电子科技大学 信息与通信工程学院, 成都 610054)

²(贵州大学 贵州省公共大数据重点实验室, 贵阳 550025)

³(贵州大学 计算机科学与技术学院, 贵阳 550025)

⁴(哈尔滨工业大学 深圳研究生院, 深圳 518055)

通讯作者: 焦健雄, E-mail: jjx_uestc@hotmail.com

摘 要: 在字典学习算法中, 使用图像的多矢量表示相比单一矢量表示, 可以获得分类精度更高且更具有鲁棒性的分类模型. 本文中我们采用多种矢量表示的组合以及合理的加权对数和方案, 来提升字典算法的性能. 通过在公共人脸数据集上进行实验, 验证了我们的方法应用于字典学习具有更高的准确度和鲁棒性. 充分挖掘和利用表示多样性可以获得被观察对象的各种潜在外观以及图像高分类精度.

关键词: 矢量表示组合; 字典学习; 图像识别

引用格式: 焦健雄, 孙利雷, 徐勇. 基于矢量表示组合的字典学习性能分析. 计算机系统应用, 2021, 30(1): 174-179. <http://www.c-s-a.org.cn/1003-3254/7752.html>

Dictionary Learning Performance Analysis Based on Combination of Vector Representations

JIAO Jian-Xiong¹, SUN Li-Lei^{2,3}, XU Yong⁴

¹(School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China)

²(Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guiyang, 550025, China)

³(College of Computer Science & Technology, Guizhou University, Guiyang 550025, China)

⁴(Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen 518055, China)

Abstract: In the dictionary learning algorithms, the model by the multi-vector representation can obtain better classification performance and more robustness than that by the single vector representation. In this study, we use the combined representation fused multiple vector representations and reasonable weighted logarithms sum schemes to improve the performance of the dictionary algorithm. Experiments on public face datasets verify that dictionary learning algorithms applied with proposed method has higher accuracy and robustness. It illustrates that the various potential appearances of observed objects generated by fully mining and utilizing the diversity of representations are beneficial to improve the performance of images classification.

Key words: combinations of vector representations; dictionary learning; image recognition

字典学习的目标, 是学习一个过完备字典, 从字典中选择少数的字典原子, 通过选择的字典原子的线性组合来近似表示给定的原始信号. 字典学习在图像降噪、复原、分类、超分辨率图像、图像压缩等图像领

域有非常好的应用效果^[1,2].

字典学习的鲁棒性非常重要^[3], 其不仅意味着学习获得的字典和特征可以有效抵抗噪声和奇异值, 而且可以为模式分类任务带来令人满意的识别结果. 利用

① 基金项目: 贵州省科技计划项目 (黔科合重大专项字 [2018]3001); 贵州省公共大数据重点实验室开放课题 (2018BDKFJJ001)

Foundation item: Science and Technology Plan of Guizhou Province ([2018]3001); Open Fund of Guizhou Provincial Key Laboratory of Public Big Data (2018BDKFJJ001)

收稿时间: 2020-06-03; 修改时间: 2020-06-30; 采用时间: 2020-07-10; csa 在线出版时间: 2020-12-31

适当的约束可以克服噪声和奇异值,相关标签和原子相关约束可以被用来达到这个目的^[4].当字典学习算法或是其他算法应用于图像数据时,首先需要将每个图像转换为矢量表示形式^[5],而转换方式则有多种,例如将图像矩阵的第一行到最后一行按照顺序连接为一个矢量,或者将图像矩阵的第一列到最后一列按照顺序连接为一个矢量.

字典学习的研究者大多数把精力用于寻找良好的分类算法和表示算法.然而,在固定算法条件下,寻找如何最好利用原始数据以达到最佳性能的方法很少被研究者注意.对于图像数据,原始图像的多重表示可以提供图像样本的不同观察视角,从而提高了数据的多样性^[6],当相同字典学习算法应用于它们时,由于多重表示的互补性,很有可能获得更好的性能.多矢量表示和多视图^[7]部分相似,但多矢量表示直接从原始图像生成,消耗的计算成本很小.

在人脸数据库中,原始样本的潜在样本对于计算机视觉和分类任务很有用,因为它们和真实面部图像非常像.比较典型的潜在样本包括虚拟面部样本和合成面部样本^[8],例如,对称人脸图像获得真实人脸的对称结构^[9],也可以通过复杂算法生成虚拟或合成面部样本^[10].潜在的多个样本和多矢量表示都可以提供更多外观表示,而它们的不同点在于,潜在的多个样本意味着从原始样本中生成了新的替代样本,多矢量表示只是原始样本的不同排列结果.

2019年,徐勇将相同的字典学习算法应用于原始图像的多矢量表示,并在此基础上设计了一种鲁棒的字典学习方法^[6].

通过对以上文献可以发现,虽然现有方法在人脸数据集上已经获得了较好的性能,但仍然存在没有充分利用原始图像信息的问题.为了解决这一问题,我们采用组合更多种原始图像的矢量表示应用到字典学习算法中,这种方法能够充分挖掘和利用表示多样性,提高字典学习算法的精度和鲁棒性.我们在本文分析不同组合矢量表示对字典学习算法性能的影响.在第2节中阐述研究方法.在第3节中展示研究方法的合理性.第4节描述实验结果.第5节提出一个简要的结论.

1 研究方法

为了更加充分提取原始图像的判别特征,我们通过不同扫描方式提取原始图像的矢量表示,让机器能

够从不同观察角度提取原始图像判别特征,配合合理的加权对数和方案,对分类准确率和鲁棒性进行改进.

1.1 研究方法的主要步骤

本文研究方法的主要步骤如下:

步骤1.由原始图像直接生成 N 种矢量表示,每种矢量表示都是使用特定的转换方案生成.例如,将原始图像逐像素转换为矢量表示就是一种矢量表示可行方案^[11,12].在下一节中,设置4类基础矢量表示,每类基础矢量表示有4种矢量表示, $N=16$.

步骤2.将相同的字典学习算法应用于原始图像的每种矢量表示.因为原始图像有 N 种矢量表示,我们有 N 种训练样本和测试样本,从而由相同字典学习算法生成 N 种训练样本和测试样本的特征.

步骤3.对这 N 种训练样本和测试样本的特征使用相同的分类算法,并且针对每种特征,获得所有测试样本相对不同类别的分类分数.

步骤4.对于测试样本,采用求和融合方案来添加该测试样本相对同一类别的分类分数.求和融合结果被称为该类别的测试样本的最终分数.求和融合方案以4类基础矢量表示为基础并选择组合,获得15种矢量表示组合方案,得到15种分类分数.

研究方法的流程图如图1所示.



图1 研究方法的流程图

1.2 方法的细节说明

我们将在本小节介绍设计方法的详细信息.对于第 r 种样本,我们用 y_r^i 表示矢量形式的第 i 个训练样本,用 p_r 表示矢量形式的测试样本. y_r^i 和 p_r 是由原始图像生成,使用字典学习算法获得的 y_r^i 和 p_r 的特征分别是 x_r^i 和 q_r ,对应的字典用 D_r 表示.

下面的文字解释步骤1.用于产生矢量形式的训练样本 y_r^i 和矢量形式的测试样本 p_r 的原始图像矩阵为 Y_r^i 和 P_r .图2是原始图像的示意图,其中, $row_x, Rrow_x$ ($x=1, \dots, m$)分别表示原始图像第 x 行的顺序表示和逆序表示; $col_y, Rcol_y$ ($y=1, \dots, n$)分别表示原始图像第 y 行的顺序表示和逆序表示.

本文设计的矢量表示共有4类基础矢量表示形式,每类基础矢量表示形式有4种矢量表示产生方式,共16种矢量表示产生方式.

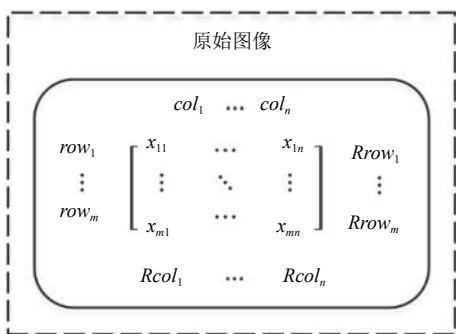


图2 原始图像示意图

第1类基础矢量表示形式: 行/列顺序排列表示.

- 1) 行顺序排列: $row_1 \rightarrow row_2 \rightarrow \dots \rightarrow row_{m-1} \rightarrow row_m$.
- 2) 列顺序排列: $col_1 \rightarrow col_2 \rightarrow \dots \rightarrow col_{n-1} \rightarrow col_n$.
- 3) 行顺序倒排列: $row_m \rightarrow row_{m-1} \rightarrow \dots \rightarrow row_2 \rightarrow row_1$.
- 4) 列顺序倒排列: $col_n \rightarrow col_{n-1} \rightarrow \dots \rightarrow col_2 \rightarrow col_1$.

第2类基础矢量表示形式: 行/列倒序排列表示.

- 1) 行倒序排列: $Rrow_1 \rightarrow Rrow_2 \rightarrow \dots \rightarrow Rrow_{m-1} \rightarrow Rrow_m$.
- 2) 列倒序排列: $Rcol_1 \rightarrow Rcol_2 \rightarrow \dots \rightarrow Rcol_{n-1} \rightarrow Rcol_n$.
- 3) 行倒序倒排列: $Rrow_m \rightarrow Rrow_{m-1} \rightarrow \dots \rightarrow Rrow_2 \rightarrow Rrow_1$.
- 4) 列倒序倒排列: $Rcol_n \rightarrow Rcol_{n-1} \rightarrow \dots \rightarrow Rcol_2 \rightarrow Rcol_1$.

第3类基础矢量表示形式: 顺序行/列之字形排列表示.

- 1) 顺序行之字形顺排列: $row_1 \rightarrow Rrow_2 \rightarrow \dots \rightarrow row_{m-1} \rightarrow Rrow_m$.
- 2) 顺序列之字形顺排列: $col_1 \rightarrow Rcol_2 \rightarrow \dots \rightarrow col_{n-1} \rightarrow Rcol_n$.
- 3) 顺序行之字形倒排列: $row_m \rightarrow Rrow_{m-1} \rightarrow \dots \rightarrow row_2 \rightarrow Rrow_1$.
- 4) 顺序列之字形倒排列: $col_n \rightarrow Rcol_{n-1} \rightarrow \dots \rightarrow col_2 \rightarrow Rcol_1$.

第4类基础矢量表示形式: 倒序行/列之字形排列表示.

- 1) 倒序行之字形顺排列: $Rrow_1 \rightarrow row_2 \rightarrow \dots \rightarrow Rrow_{m-1} \rightarrow row_m$.
- 2) 倒序列之字形顺排列: $Rcol_1 \rightarrow col_2 \rightarrow \dots \rightarrow$

$Rcol_{n-1} \rightarrow col_n$.

3) 倒序行之字形倒排列: $Rrow_m \rightarrow row_{m-1} \rightarrow \dots \rightarrow Rrow_2 \rightarrow row_1$.

4) 倒序列之字形倒排列: $Rcol_n \rightarrow col_{n-1} \rightarrow \dots \rightarrow Rcol_2 \rightarrow col_1$.

在步骤3)中,通过以下方式使用分类算法获得分数.训练样本的标签矩阵用 L 表示, L 的第 j 行行向量 L_j 表示第 j 个训练样本的类别标签.如果第 j 个训练样本属于第 j 类,则 L_j 的第 k 个条目为1,所有其他条目为零.对于一种样本,所有样本的特征形成矩阵 X .假设矩阵 W 可以将 X 近似转换为 L ,则 $XW=L$.我们要求 W 具有最小范数,因此我们有 $W=(X^T X + \gamma I)^{-1} X^T L$ 的解,其中 γ 是一个小的正常数.对于测试样本,使用 $l_r = q_r W$ 计算其所需标签,然后我们用 $d_r^g = \|l_r - label_g\|$ 表示测试样本相对于第 g 类的得分. $label_g$ 是第 g 类的类标签,使用 one-hot 编码形式.

在步骤4)中,求和融合方案的公式为:

$$score_g = \sum_{r=1}^k w_r \ln(d_r^g) \quad (1)$$

$\ln(d_r^g)$ 表示 d_r^g 的自然对数.如果 $h = \arg \min score_g$,则将测试样本分类为第 h 类. w_r 计算如下:⁸按升序对 d_r^1, \dots, d_r^C 进行排序,并将排序结果记为 e_r^1, \dots, e_r^C ,则将 w_r 定义为 $e_r^1 - e_r^2$.

求和融合方案以4类基础矢量表示作为基础并对其进行选择组合,获得15种矢量表示组合方案,得到15种分类分数.

每一类矢量表示作为基础表示,用A1、A2、A3、A4表示,将4类基础表示进行组合,具体矢量表示组合方案为:

- 仅基础表示: A1, A2, A3, A4;
- 选取两种基础表示组合:
 - B1(A1+A2), B2(A1+A3), B3(A1+A4),
 - B4(A2+A3), B5(A2+A4), B6(A3+A4);
- 选取3种基础表示组合:
 - C1(A1+A2+A3), C2(A1+A2+A4),
 - C3(A1+A3+A4), C2(A2+A3+A4);
- 选取4种基础表示组合:
 - D1(A1+A2+A3+A4).

2 研究方法合理性分析

本文充分利用原始图像数据的表示多样性,为同

一对象提供了更多的外观,原始图像更多的矢量表示意味着获得对同一对象更全面的观察,更充分地利用了原始图像的信息.不同的矢量表示组合则意味着从同一对象的不同外观组合中进行字典学习,可以通过观察不同矢量表示组合下的性能,进而选择更好的矢量表示组合.

由于字典学习算法中的矩阵分解也进一步增加了原始图像多矢量表示的差异.具体来说,字典学习算法将原始样本矩阵 P 分解为字典矩阵 D 和特征矩阵 X ,不同 D 的差异和不同 X 的差异通常比不同 P 的差异大.基于字典的特征表示:

$$\|Y\| \approx \|DX\| \leq \|D\| \cdot \|X\| \quad (2)$$

从式(2)可以直观看出来在矩阵分解之后,范数具有更大的范围,从而数据具有更大的差异.更多矢量表示组合意味着更多的训练样本以及更高的分类精确度.

采用的求和融合方案也是合理的.有两个方面原因:

1) 将第二小的得分减去最小的得分作为权值是合理的,如果第二小的得分比最小的得分大很多,则分类决策具有很高的置信度^[13].

2) 本文中的得分表示相异程度,采用最小的得分确定测试样本的分类标签,得分越小意味着属于相应类别的可能性越高.并且求和融合公式采用得分的对数,不仅可以保持得分的数字顺序,而且扩大了得分的差异.

所以本文采用的求和融合方案可以保持小得分的主导作用,有利于正确分类.

本文采用多种原始图像不同的矢量表示意味着在实际中从不同观测角度观察对象.例如,行倒序排序是在镜子里观察观测对象;列倒序排序是从旋转180°的角度观察观测对象;之字形矢量表示则考虑人脸对称性,在不同时间左脸和右脸的信息会有细微变化,之字形矢量表示可以观察不同时间观测对象的面部信息.采用更多不同原始图像的矢量表示意味着拥有更大的观测对象数据集,也就意味着字典学习算法将获得更高的识别准确度.而随着数据集的扩大,训练时间也将变长,且与数据集大小成正比.

3 实验

本文在 K-SVD^[14]、D-KSVD^[15]、LC-KSVD^[16] 算法上实现不同矢量表示组合,并比较不同矢量表示组合下相同字典学习算法的性能,实验中采用扩展 Yale B 人脸数据库^[17], AR 人脸数据库^[18] 和 COIL 20 数据库^[19],每种算法执行 10 次,并记录平均分类精确度.

3.1 在扩展 Yale B 人脸数据库上的实验

扩展 Yale B 人脸数据库包含 38 个人的 2,414 张正面图像,这些图像是在各种光照条件和表情下拍摄的.每个人有 59 到 64 张图像,每张图像被转换为 32×32 像素的大小.来自该数据库的一些面部图像在图 3 中示出.随机选择每个人的 10 张图像作为训练样本,其余图像用于测试.对于 K-SVD、D-KSVD、LC-KSVD 算法,字典原子数从 38 变为 380,间隔为 38.不同矢量表示组合在扩展 Yale B 人脸数据库上 K-SVD、LC-KSVD、D-KSVD 算法识别准确度分别如图 4、图 5 和图 6 所示.其中, a1 表示行顺序排列矢量表示.



图 3 扩展 Yale B 人脸数据库的部分示例图像

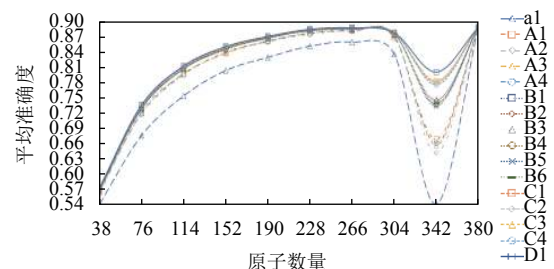


图 4 不同矢量表示组合在扩展 Yale B 人脸数据库上 K-SVD 算法识别准确度

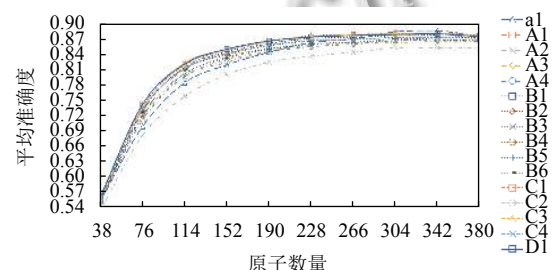


图 5 不同矢量表示组合在扩展 Yale B 人脸数据库上 LC-KSVD 算法识别准确度

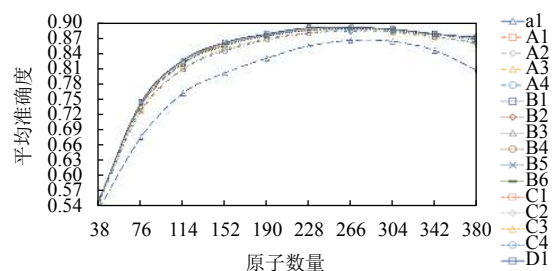


图 6 不同矢量表示组合在扩展 Yale B 人脸数据库上 D-KSVD 算法识别准确度

3.2 在 AR 人脸数据库上的实验

AR 人脸数据库包含 126 个人超 4000 张图像, 这些图片分两次拍摄而得. 采集环境中的摄像机参数、光照环境、摄像机距离等都是严格控制的. 本文采用 AR 人脸数据库的子集进行实验^[1], 包含 120 人的 3120 张图像. 每张 AR 人脸图像的尺寸为 40×50 像素. 来自该数据库的一些面部图像在图 7 中示出. 随机选择每个人的 10 张图像作为训练样本, 其余图像用于测试. 对于 K-SVD、D-KSVD、LC-KSVD 算法, 字典原子数从 120 变为 1200, 间隔为 120. 不同矢量表示组合在 AR 人脸数据库上 K-SVD、LC-KSVD、D-KSVD 算法识别准确度分别如图 8、图 9 和图 10 所示.



图 7 AR 人脸数据库的部分示例图像

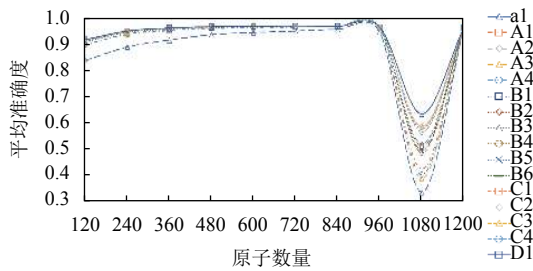


图 8 不同矢量表示组合在 AR 人脸数据库上 K-SVD 算法识别准确度

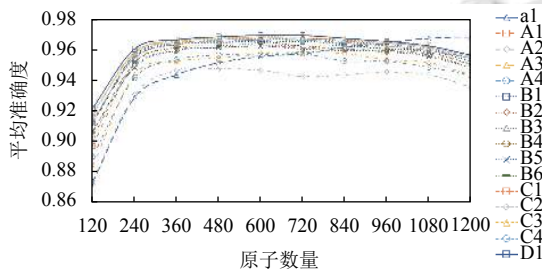


图 9 不同矢量表示组合在 AR 人脸数据库上 LC-KSVD 算法识别准确度

3.3 在 COIL20 数据库上的实验

COIL20 数据库包含不同光照条件下 20 个物体的不同视角, 每张图片尺寸为 32×32 像素. 来自该数据库的一些面部图像在图 11 中示出. 随机选择每个人的 10 张

图像作为训练样本, 其余图像用于测试. 对于 K-SVD、D-KSVD、LC-KSVD 算法, 字典原子数从 40 变为 200, 间隔为 20. 不同矢量表示组合在 COIL20 数据库上 K-SVD、LC-KSVD、D-KSVD 算法识别准确度分别如图 12、图 13 和图 14 所示.

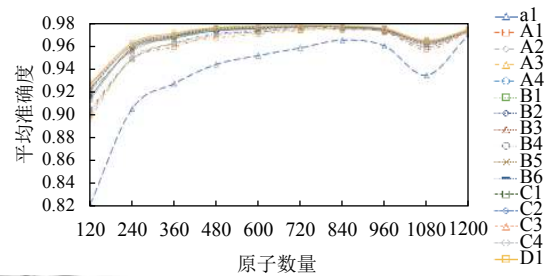


图 10 不同矢量表示组合在 AR 人脸数据库上 D-KSVD 算法识别准确度



图 11 COIL20 数据库的部分示例图像

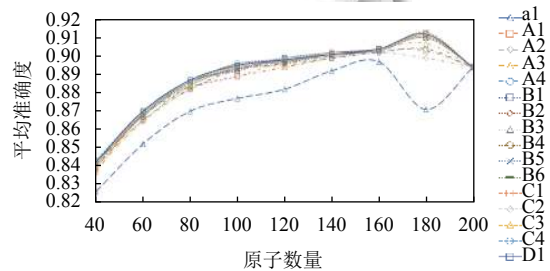


图 12 不同矢量表示组合在 COIL20 数据库上 K-SVD 算法识别准确度

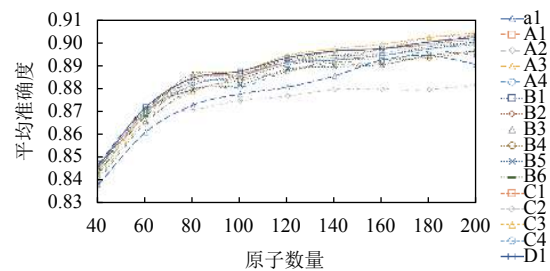


图 13 不同矢量表示组合在 COIL20 数据库上 LC-KSVD 算法识别准确度

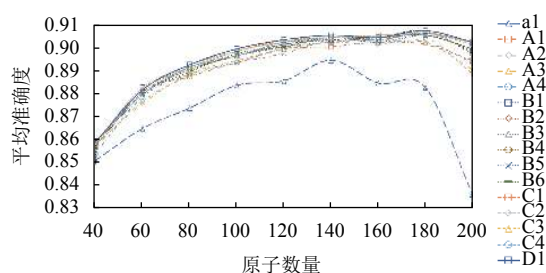


图14 不同矢量表示组合在 COIL20 数据库上 D-KSVD 算法识别准确度

4 结论

本文提出一种基于矢量表示组合的字典学习方法,该方法通过充分利用人脸图像的多矢量表示提升分类模型的准确率和鲁棒性.通过在多个人脸数据集进行实验对比可以得出:相同数量的基础表示组合的得分融合数据相差很小,在各识别准确度的图中呈曲线簇;组合中的矢量表示种类越多样,字典学习算法的识别精确度越好,具有更好的鲁棒性.本文通过组合更多的原始图像矢量表示,充分挖掘和利用表示多样性,提高字典学习算法的分类精度.

参考文献

- 1 孙利雷, 秦进. 稀疏字典学习方法综述. 贵州大学学报(自然科学版), 2018, 35(5): 81–86.
- 2 Xu Y, Li ZM, Yang J, *et al.* A survey of dictionary learning algorithms for face recognition. *IEEE Access*, 2017, 5: 8502–8514. [doi: 10.1109/ACCESS.2017.2695239]
- 3 Pan QH, Kong DG, Ding CHQ, *et al.* Robust non-negative dictionary learning. *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. Québec City, QC, Canada. 2014. 2027–2033.
- 4 Li ZM, Lai ZH, Xu Y, *et al.* A locality-constrained and label embedding dictionary learning algorithm for image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(2): 278–293. [doi: 10.1109/TNNLS.2015.2508025]
- 5 Xu Y, Zhong AN, Yang J, *et al.* LPP solution schemes for use with face recognition. *Pattern Recognition*, 2010, 43(12): 4165–4176. [doi: 10.1016/j.patcog.2010.06.016]
- 6 Xu Y, Li ZM, Tian CW, *et al.* Multiple vector representations of images and robust dictionary learning. *Pattern Recognition Letters*, 2019, 128: 131–136. [doi: 10.1016/j.patrec.2019.08.022]
- 7 Zhao CR, Wang XK, Miao DQ, *et al.* Maximal granularity

- structure and generalized multi-view discriminant analysis for person re-identification. *Pattern Recognition*, 2018, 79: 79–96. [doi: 10.1016/j.patcog.2018.01.033]
- 8 Xu Y, Fang XZ, Li XL, *et al.* Data uncertainty in face recognition. *IEEE Transactions on Cybernetics*, 2014, 44(10): 1950–1961. [doi: 10.1109/TCYB.2014.2300175]
- 9 Xu Y, Zhu XJ, Li ZM, *et al.* Using the original and ‘symmetrical face’ training samples to perform representation based two-step face recognition. *Pattern Recognition*, 2013, 46(4): 1151–1158. [doi: 10.1016/j.patcog.2012.11.003]
- 10 Xu Y, Zhang Z, Lu GM, *et al.* Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification. *Pattern Recognition*, 2016, 54: 68–82. [doi: 10.1016/j.patcog.2015.12.017]
- 11 Xu Y, Li ZM, Zhang B, *et al.* Sample diversity, representation effectiveness and robust dictionary learning for face recognition. *Information Sciences*, 2017, 375: 171–182. [doi: 10.1016/j.ins.2016.09.059]
- 12 Xu Y, Zhang B, Zhong ZF. Multiple representations and sparse representation for image classification. *Pattern Recognition Letters*, 2015, 68: 9–14. [doi: 10.1016/j.patrec.2015.07.032]
- 13 Xu Y, Lu YW. Adaptive weighted fusion: A novel fusion approach for image classification. *Neurocomputing*, 2015, 168: 566–574. [doi: 10.1016/j.neucom.2015.05.070]
- 14 Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 2006, 54(11): 4311–4322. [doi: 10.1109/TSP.2006.881199]
- 15 Zhang Q, Li BX. Discriminative K-SVD for dictionary learning in face recognition. *Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA. 2010. 2691–2698.
- 16 Jiang ZL, Lin Z, Davis LS. Learning a discriminative dictionary for sparse coding via label consistent K-SVD. *Proceedings of CVPR 2011*. Providence, RI, USA. 2011. 1697–1704.
- 17 Georghiadis 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, 23(6): 643–660. [doi: 10.1109/34.927464]
- 18 Martinez AM, Benavente R. *The AR face database*. Columbus: The Ohio State University, 1998.
- 19 Nene SA, Nayar SK, Murase H. *Columbia object image library (COIL-20)*. New York: Columbia University, 1996.