

DOI: 10.7652/xjtuxb201602006

基于 L2 范数的局部保持投影算法

丁 铭¹, 贾维敏¹, 姚敏立²

(1. 第二炮兵工程大学 907 教研室, 710025, 西安; 2. 第二炮兵工程大学 403 教研室, 710025, 西安)

摘要: 针对传统局部保持投影算法对外点敏感的问题,提出了一种基于 L2 范数的局部保持投影算法。该算法通过采用 L2 范数定义目标函数并重新定义了权值矩阵,多次迭代计算投影矩阵得到局部最小值,直至达到收敛条件,进而获得最终的最优投影矩阵;通过利用最优投影矩阵将原始数据投影到最优的投影子空间,降低高维数据维度,同时能够保持原有数据特征。合成数据实验结果表明,与传统局部保持投影算法相比,所提基于 L2 范数的局部保持投影算法能够有效地降低数据维度,改善了算法对外点的敏感问题,提高了算法的鲁棒性。人脸识别实验结果表明,该算法能够取得较高且较为稳定的人脸识别率,人脸识别率可达 80%。

关键词: 降维;局部保持投影;L2 范数

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 0253-987X(2016)02-0033-05

A Projection Algorithm with Local Preservation Based on L2 Norm

DING Ming¹, JIA Weimin¹, YAO Minli²

(1. Staff Room 907, The Second Artillery Engineering University, Xi'an 710025, China;
2. Staff Room 403, The Second Artillery Engineering University, Xi'an 710025, China)

Abstract: A projection algorithm with locality preserving projections based on the L2 norm (LPP-L2) is proposed to solve the problem that objective functions of traditional projection algorithms with local preservation (LPP) are based on the squared L2 norm and very sensitive to outliers. The weight matrix of the algorithm is recalculated using an iterative method and the objective function is also simplified, as a result, the optimized projection matrix is obtained. The algorithm converges to a local optimum in each iteration. The optimized projection matrix is used to project the original data into an optimal projection subspace with reduced dimension, while the characters of original data are preserved. Experimental results on synthesized data and a comparison with the LPP algorithm show that the LPP-L2 algorithm effectively reduces the data dimension and makes the algorithm more robust to outliers, it gains higher accurate and steady classification rate in face recognition, and a recognition rate of 80% is obtained.

Keywords: dimensionality reduction; locality preserving projections; L2 norm

随着信息爆炸时代的来临,信息获取的数据量越来越大,这些信息、数据,特别是高维数据的有效处理,成为了迫切需要解决的问题和巨大的挑战。因此,如何降低数据维度^[1-3]成为处理高维数据的必

要过程。降维算法旨在寻找一个低维子空间来表示高维数据固有的内在结构,通过将高维数据投影到低维子空间,剔除高维数据中不相关的部分,使固有的内在结构得到更清晰的表示。

经典的数据降维方法有主成分分析(principal component analysis, PCA)^[3]、线性判别分析(linear discriminant analysis, LDA)^[4]以及局部保持投影算法(locality preserving projections, LPP)^[5-7],然而 PCA、LDA 均将原始高维数据假定为线性结构继而进行线性降维。近期研究表明,当高维数据存在于一个低维流形时,不能将高维数据假定为线性结构,因此,采用流形学习的方式对数据进行降维是当前的趋势。LPP 作为流形学习中的一种算法,具有更为广泛的应用,当原始数据非线性时,LPP 可以使数据得到较为准确的表示,并且提供从高维空间到低维空间的显式映射;同时,PCA 与 LDA 均是注重保持数据的全局结构,而在人脸识别^[8-10]、信息检索等领域,数据的局部结构对于识别更为重要,LPP 注重保持数据的局部结构有利于进行识别。

LPP 虽然能够对数据进行相对准确的表征,保持局部信息,但是分析 LPP 算法的目标函数可以发现,其目标函数采用的是 L2 范数的平方,对外点敏感。为此,本文提出了一种基于 L2 范数的局部保持投影算法(locality preserving projections based on L2 norm, LPP-L2),该算法在有效降低数据维度的基础上,能够保证较高的目标识别正确率,同时降低对外点的敏感度,提高算法的鲁棒性。

1 基于 L2 范数局部保持投影算法

1.1 算法原理

LPP 算法能够克服 PCA 算法与 LDA 算法的不足,并且很好地保持了高维数据固有的几何结构和局部结构。LPP 算法中的目标函数为

$$\min \sum_{ij} \|y_i - y_j\|^2 S_{ij} \quad (1)$$

式中:输入数据为 $\mathbf{X} = \{x_1, x_2, x_3, \dots, x_n\}$, $\mathbf{X} \in R^{d \times n}$; $y_i = \mathbf{W}^T x_i$ 表示 x_i 在投影矩阵 $\mathbf{W} \in R^{d \times m}$ 构成的子空间上的投影; $\mathbf{S}$ 是权值矩阵。

式(1)中运用了 L2 范数的平方,将会使高维数据中的外点呈平方倍放大,这样会导致 LPP 算法对于外点十分敏感,导致无法对高维数据进行准确的降维,同时进行分类识别时识别准确率降低。

为了克服上述 LPP 算法存在的问题,借鉴文献[11-13]聚类方法的思想,将其引入到降维算法中,提出了 LPP-L2 算法。文献[11]中对传统的归一化割聚类模型进行改进,设有 n 个数据点 $\{x_1, \dots, x_n\} \in R^{d \times 1}$, $\mathbf{W} \in R^{n \times n}$ 为构造图所需的权值矩阵, $\mathbf{Q} = [q_1, q_2, \dots, q_c] \in R^{n \times c}$ 为聚类指标矩阵,其中 $q_k \in$

$R^{n \times 1}$ 为矩阵 $\mathbf{Q}$ 的第 k 列,传统的归一化割聚类模型表示为

$$\min_{\mathbf{Q}^T \mathbf{DQ} = \mathbf{I}} \sum_{i,j=1}^n \mathbf{W}_{ij} \|q^i - q^j\|_2^2 \quad (2)$$

当 x_i 与 x_j 属于同一聚类时,此时式(7)中 $\mathbf{Q}$ 的最优解为 $q^i = q^j$,对于一些数据对 (i, j) 存在 $\|q^i - q^j\|_2 = 0$,因此使用范数作为目标函数,将其改进为

$$\min_{\mathbf{Q}^T \mathbf{DQ} = \mathbf{I}} \sum_{i,j=1}^n \mathbf{W}_{ij} \|q^i - q^j\|_2$$

根据文献[13]中的改进方式,改进 LPP 算法的目标函数,即 LPP-L2 算法的目标函数为

$$\min_{\mathbf{W}^T \mathbf{XDX}^T \mathbf{W} = \mathbf{I}_m} \sum_{i,j} S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2 \quad (3)$$

由于本文中外点即数值较大的点,采用范数作为目标函数与采用范数的平方作为目标函数相比,外点的值将不会造成平方倍的增大,即运用 L2 范数代替原有的 L2 范数的平方,可以避免在存在外点的情况下,大部分局部近邻的数据之间的距离最小化,而个别近邻数据的距离却相距很远,但是使用范数作为目标函数抑制外点由此带来的问题就是计算复杂度的提高,2.2 节对计算复杂度进行了理论分析与实验说明。

运用拉格朗日函数解式(3)得到

$$J(\mathbf{W}) = \sum_{i,j} S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2 - \text{tr}(\mathbf{A}(\mathbf{W}^T \mathbf{XDX}^T \mathbf{W} - \mathbf{I}_m)) \quad (4)$$

为简化式(4),重新定义权值矩阵,新的权值矩阵 $\tilde{\mathbf{S}}$ 定义为

$$\tilde{S}_{ij} = \frac{S_{ij}}{2 \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2} \quad (5)$$

定义 $\tilde{\mathbf{L}} = \tilde{\mathbf{D}} - \tilde{\mathbf{S}}$, $\tilde{\mathbf{D}}$ 是权值矩阵 $\tilde{\mathbf{S}}$ 的列和(或行和),即 $\tilde{D}_{ii} = \sum_j \tilde{S}_{ji}$ 。将式(5)代入式(4)并求导可得

$$\frac{\partial J(\mathbf{W})}{\partial (\mathbf{W})} = \tilde{\mathbf{X}} \tilde{\mathbf{L}} \mathbf{X}^T \mathbf{W} - \mathbf{XDX}^T \mathbf{W} \mathbf{A} = 0 \quad (6)$$

根据式(6),式(4)可以简化为

$$J(\mathbf{W}) = \sum_{i,j=1}^n \tilde{S}_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^2 - \text{tr}(\mathbf{A}(\mathbf{W}^T \mathbf{XDX}^T \mathbf{W} - \mathbf{I}_m)) = \text{tr}(\mathbf{W}^T \tilde{\mathbf{X}} \tilde{\mathbf{L}} \mathbf{X}^T \mathbf{W}) - \text{tr}(\mathbf{A}(\mathbf{W}^T \mathbf{XDX}^T \mathbf{W} - \mathbf{I}_m)) \quad (7)$$

通过计算式(7),可以得到 LPP-L2 的投影矩阵 $\mathbf{W}$ 即 $d \times d$ 维矩阵 $(\mathbf{XDX}^T)^{-1} \tilde{\mathbf{X}} \tilde{\mathbf{L}} \mathbf{X}^T$ 的 m 个最小特征值对应的特征向量,由于 $\tilde{\mathbf{L}}$ 是根据已有的 $\mathbf{W}$ 得到的,而 $\mathbf{W}$ 的更新依据 $\tilde{\mathbf{L}}$,所以 LPP-L2 运用迭代的方式得到最优投影矩阵,并且保证算法达到局部最优。

根据式(5)与式(7)可以得到若经过 t 次迭代 LPP-L2 算法取得局部最优,则求得最优投影矩阵 $\mathbf{W}$ 需要的计算复杂度为 $O(d^2mt + 2dmn^2t)$ 。

1.2 收敛性证明

要证明收敛性,首先需要证明以下两个不等式成立。

(1)对于标量 x ,存在

$$2|x| - x^2 - 1 \leq 0 \quad (8)$$

证明 令 $f(x) = 2x^{1/2} - x - 1$,可以得到 $f'(x) = x^{-1/2} - 1$, $f''(x) = -1/2x^{-3/2}$ 。显然,当 $x > 0$ 时, $f''(x) < 0$ 且 $f''(x)$ 当且仅当 $x = 1$ 时,存在 $f'(x) = 0$ 。由于 $x = 1$ 时, $f(x) = 0$,即当 $x > 0$ 时, $f(x)$ 单调递减且 $f(x) \leq 0$,则 $f(x^2) \leq 0$,即 $2|x| - x^2 - 1 \leq 0$ 。

(2)对于任意非零向量 $\mathbf{v}$ 和 $\mathbf{v}_0$,存在

$$\|\mathbf{v}\|_2 - \frac{1}{2} \|\mathbf{v}_0\|_2^{-1} \|\mathbf{v}\|_2^2 \leq \|\mathbf{v}_0\|_2 - \frac{1}{2} \|\mathbf{v}_0\|_2^{-1} \|\mathbf{v}_0\|_2^2 \quad (9)$$

证明 由步骤1可以得到

$$\begin{aligned} & \frac{1}{2} \left(\frac{\|\mathbf{v}\|_2}{\|\mathbf{v}_0\|_2} \right) - \left(\frac{\|\mathbf{v}\|_2}{\|\mathbf{v}_0\|_2} \right)^2 - 1 \leq 0 \Rightarrow \\ & 2\|\mathbf{v}\|_2 - \|\mathbf{v}_0\|_2^{-1} \|\mathbf{v}\|_2^2 \leq 2\|\mathbf{v}_0\|_2 \Rightarrow \\ & \|\mathbf{v}\|_2 - \frac{1}{2} \|\mathbf{v}_0\|_2^{-1} \|\mathbf{v}\|_2^2 \leq \|\mathbf{v}_0\|_2 - \\ & \frac{1}{2} \|\mathbf{v}_0\|_2^{-1} \|\mathbf{v}_0\|_2^2 \end{aligned}$$

根据上述不等式可以得到收敛性定理,即目标函数式(3)为单调递减函数,经过每一次迭代均会收敛到一个局部最小值。

证明 根 LPP-L2 据算法的源理,将迭代计算得到的更新投影矩阵定义为

$$\begin{aligned} \tilde{\mathbf{W}} &= \arg \min_{\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W} = \mathbf{I}_m} \text{tr}(\mathbf{W}^T \mathbf{X} \tilde{\mathbf{L}} \mathbf{X}^T \mathbf{W}) = \\ & \arg \min_{\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W} = \mathbf{I}_m} \sum_{i,j=1}^n \tilde{S}_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^2 \end{aligned}$$

因此可以得到

$$\begin{aligned} & \sum_{i,j=1}^n \frac{1}{2} S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^{-1} \|\tilde{\mathbf{W}}^T x_i - \tilde{\mathbf{W}}^T x_j\|_2^2 \leq \\ & \sum_{i,j=1}^n \frac{1}{2} S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^{-1} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^2 \end{aligned} \quad (10)$$

根据步骤2可以得到

$$\begin{aligned} & \sum_{i,j=1}^n S_{ij} \|\tilde{\mathbf{W}}^T x_i - \tilde{\mathbf{W}}^T x_j\|_2 - \\ & \sum_{i,j=1}^n \frac{1}{2} S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^{-1} \|\tilde{\mathbf{W}}^T x_i - \tilde{\mathbf{W}}^T x_j\|_2^2 \leq \end{aligned}$$

$$\begin{aligned} & \sum_{i,j=1}^n S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2 - \\ & \sum_{i,j=1}^n \frac{1}{2} S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^{-1} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2^2 \end{aligned} \quad (11)$$

将式(10)、式(11)相加得到

$$\begin{aligned} & \sum_{i,j=1}^n S_{ij} \|\tilde{\mathbf{W}}^T x_i - \tilde{\mathbf{W}}^T x_j\|_2 \leq \\ & \sum_{i,j=1}^n S_{ij} \|\mathbf{W}^T x_i - \mathbf{W}^T x_j\|_2 \end{aligned} \quad (12)$$

因此,本文提出的算法目标函数呈单调递减,并且在每次迭代计算过程中均可以取得局部最小值直到收敛。在收敛的过程中,若不等式(12)成立, $\mathbf{W}$ 和 $\tilde{\mathbf{L}}$ 将会满足式(6)和式(3),即目标函数的 KKT 最优化条件。因此,本文提出的 LPP-L2 算法将会收敛到一个局部最优值,即取得局部投影的最优解。

2 实验与结果

本节通过合成数据实验以及人脸识别实验来验证算法的鲁棒性和有效性。

2.1 S curve 曲线合成数据实验

合成的“S curve”曲线数据中包含 500 个数据点,同时加入不同数目的外点以验证算法的鲁棒性。分别在合成数据中加入 100 个和 200 个随机外点,“S curve”曲线由以下方式合成: $x = 12.5[\cos t, -\cos t]$, $y = h$, $x = 6[\sin t, 2 - \sin t]$, 其中 $t \in [-\pi, \pi/2]$, $h \in [0, 41]$ 。随机的外点满足以下条件: $x \in [0, 30]$, $y \in [0, 40]$, $z \in [0, 30]$ 。之后运用传统的 LPP 算法以及 LPP-L2 算法对已加外点的三维合成数据进行降维,将其投影到二维空间。图 1~图 3 分别为加入 100 个外点时的“S curve”曲线和运用 LPP 算法及 LPP-L2 算法降维的结果,图 4~图 6 分别为加入 200 个外点时的“S curve”曲线和运用 LPP 算法及 LPP-L2 算法降维的结果,其中 X、Y、Z 轴均为 Matlab R2014a 自动生成,坐标轴的数值没有实际意义,对实验结果不会产生不良影响。

根据图 1~图 3 可以发现,即使在外点数仅有 100 个的情况下, LPP 算法的降维效果受到了外点极大的影响,但 LPP-L2 算法仍然能够提取有效的特征,使得投影后的结果仍然是“S curve”曲线。

根据图 4~图 6 可知,在外点数较多(200 个)的情况下, LPP-L2 算法能够很好地保持“S curve”的固有结构,克服了原有 LPP 算法对于外点十分敏感的问题。

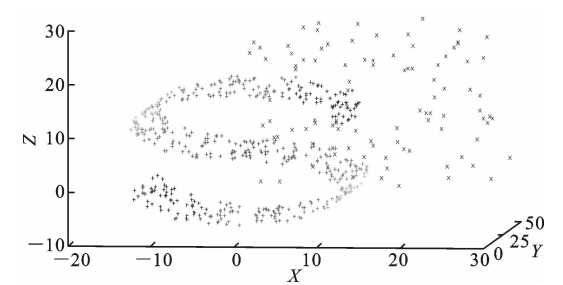


图 1 加入 100 个外点时的“S curve”曲线

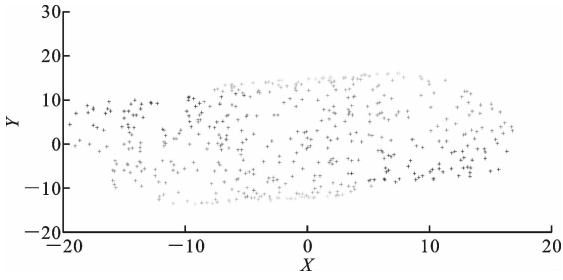


图 2 LPP 算法的降维结果

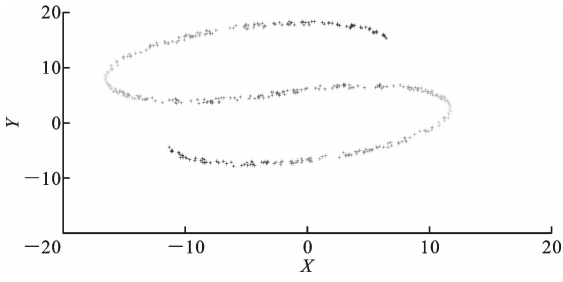


图 3 LPP-L2 算法的降维结果

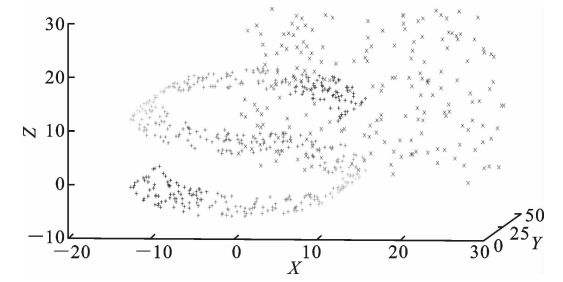


图 4 加入 200 个外点时的“S curve”曲线

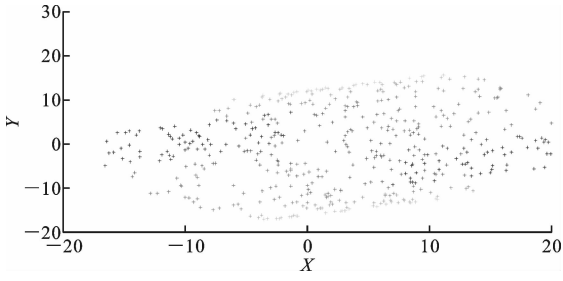


图 5 LPP 算法的降维结果

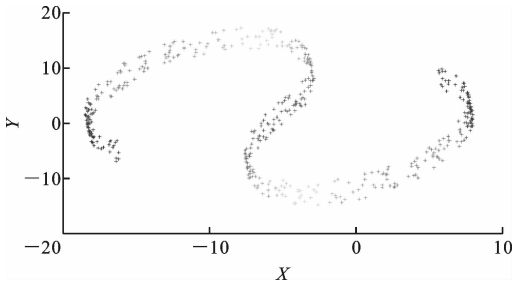


图 6 LPP-L2 算法的降维结果

2.2 AR 人脸数据库实验

AR 人脸数据库由西班牙巴塞罗那计算机视觉中心建立,包含两组图像,分别是间隔 2 周所拍摄。每组 120 人,其中 65 人为男性,55 人为女性,每组每人 13 幅图片,前 4 幅为表情变化,5~7 幅为光照变化,8~10 幅为佩戴眼镜,11~13 幅为佩戴围巾。每幅图片大小为 50×40 像素。

在 AR 人脸数据库上进行实验时,选择特定图片作为训练样本,本文将 2 周拍摄的图像进行合并编号,即每人 26 幅图像,共进行了 3 种不同的选择。首先选择每人的第 5~第 7 幅图像作为训练样本,剩余为测试图像;其次选择每人前 7 幅图像作为训练样本,剩余为测试图像;最后选择前 13 幅图像作为训练样本,剩余为测试图像。实验结果如图 7~图 9 所示,横轴为特征向量的维数,纵轴为运用最近邻分类器的分类准确率。表 1 为 AR 人脸数据库中进行 3 种不同实验时 LPP 算法与 LPP-L2 算法所需要耗费的 CPU 时间,其中选择降低的维数为 100 维,算法均在主频为 3.30 GHz 和内存为 16 GB 的计算机上的 Matlab R2014a 上运行。

图 7~图 9 表明,AR 人脸数据库中采用 LPP-L2 算法的识别率优于采用 LPP 算法的识别率。在训练样本数较少的情况下,LPP-L2 算法仍然可以取得较高的人脸识别率;相较于 LPP 算法,LPP-L2 算法在训练样本数不同的条件下,能够保持相对稳定的识别率,即 LPP-L2 算法能够在很大程度上克服不同因素的影响,取得较高较稳定的人脸识别率。

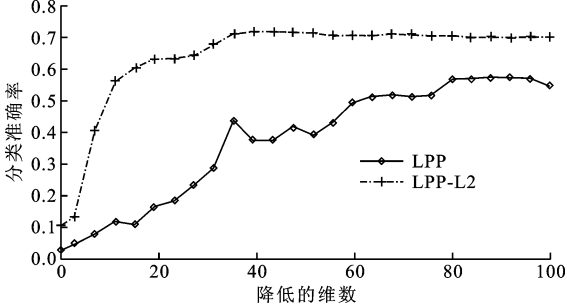


图 7 选定 3 幅图像作为训练样本时的分类准确率

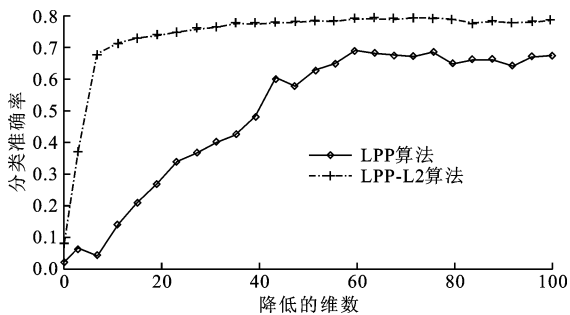


图 8 选定 7 幅图像作为训练样本时的分类准确率

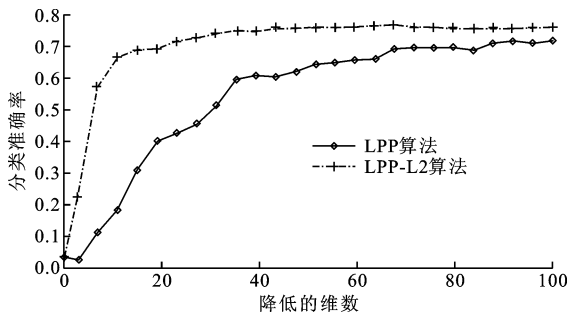


图 9 选定 13 幅图像作为训练样本时的分类准确率

表 1 不同算法的 CPU 耗时

实验	t/s	
	LPP 算法	LPP-L2 算法
实验 1	0.285	35.892
实验 2	0.336	363.380
实验 3	1.428	2 659.098

表 1 结果表明,由于 LPP-L2 算法采用了迭代的计算方法,在本次实验中迭代次数选择为 200 次,相较 LPP 算法将会消耗更多的 CPU 时间,但是,LPP-L2 算法能够更好地克服外点造成的不良影响,即用相对较多的时间获得了更准确的识别率。LPP-L2 算法的不足在于计算耗时较多,其算法鲁棒性的提高是以牺牲计算效率换来的。

在实际应用中,可以根据应用对鲁棒性和计算耗时的要求来选择算法。如果对鲁棒性要求高而对计算耗时要求不高时,可以选择 LPP-L2 算法;如果对计算耗时要求高而对鲁棒性要求不高时,可以选择 LPP 算法。

3 结 论

本文提出了一种基于 L2 范数的局部保持投影算法,采用迭代的方法,通过重新计算权值矩阵 $\tilde{S}$ 得到 $\tilde{L}$,继而得到投影矩阵 W ,根据 $\tilde{L}$ 更新投影矩阵 W ,最终得到最优投影矩阵。该算法通过运用 L2 范

数计算两点之间距离,克服了传统 LPP 算法对于外点敏感的问题。在两种合成数据的实验中,LPP-L2 算法能够克服加入的随机外点的影响,保存原始数据的局部特征结构,取得了优于传统 LPP 算法的结果,降低了算法对于外点的敏感程度,具有较好的鲁棒性;在人脸识别实验中,对不同的人脸数据库的实验结果表明,LPP-L2 算法能够克服各种不良因素的影响,取得较高的人脸识别率,能够在人脸识别领域具有较好的应用。对于 LPP-L2 算法计算耗时上的缺陷,需要在以后进一步的研究中改进和提高。

参考文献:

[1] 张丽梅. 面向降维的图学习研究及应用 [D]. 南京: 南京航空航天大学, 2012.

[2] YAN Shuicheng, XU Dong, ZHANG Benyu, et al. Graph embedding and extensions: a general framework for dimensionality reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51.

[3] JOLLIFFE I T. Principal component analysis [M]. Berlin, Germany: Springer, 2002: 2-5.

[4] MARTINEZ A M, KAK A C. PCA versus LDA [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2): 228-233.

[5] NIYOGI P, HE Xiaofei. Locality preserving projections [J]. Advances in Neural Information Processing Systems, 2003: 153-162.

[6] HE Xiaofei, YAN Shuicheng, HU Yuxiao, et al. Face recognition using Laplacian faces [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340.

[7] 何晓飞. 人脸识别的几何观点: 拉普拉斯脸 [J]. 科学观察, 2010, 5(6): 67-68.

HE Xiaofei. Face recognition: Laplasse's face [J]. Science Focus, 2010, 5(6): 67-68.

[8] PTUCHA R, TSAGKATAKIS G, SAVAKIS A. Manifold learning for simultaneous pose and facial expression recognition [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2011: 3021-3024.

[9] TURK M, PENTLAND A. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86.

[10] RAO A, NOUSHATH S. Subspace methods for face recognition [J]. Computer Science Review, 2010, 4 (1): 1-17.

(下转第 79 页)