

DOI:10.3969/j.issn.1673-4785.201205025

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20121116.1702.010.html>

## 从 Parzen 窗核密度估计到 特征提取方法:新的研究视角

刘忠宝<sup>1,2</sup>, 王士同<sup>1</sup>

(1. 江南大学 数字媒体学院, 江苏 无锡 214122; 2. 中北大学 电子与计算机科学技术学院, 山西 太原 030051)

**摘要:**当前主流特征提取方法大致有2种研究思路:1)从高维数据的几何性质出发,根据某种寻优准则得到基于原始空间特征的一组特征数更少的新特征;2)从降维误差角度出发,保证降维前后数据所呈现的某种偏差达到最小.试图从降维过程中数据分布特征的变化入手,基于广泛使用的 Parzen 窗核密度估计方法,来审视和揭示 Parzen 窗估计与典型特征提取方法 LPP、LDA 和 PCA 之间的关系,从而说明这些特征提取方法可统一在 Parzen 窗框架下进行研究,为特征提取方法的研究提供了一个新的视角.

**关键词:**特征提取; Parzen 窗; 密度估计; 数据分布特征; 新视角

**中图分类号:** TP392 **文献标志码:** A **文章编号:** 1673-4785(2012)06-0471-10

## From Parzen window estimation to feature extraction: a new perspective

LIU Zhongbao<sup>1,2</sup>, WANG Shitong<sup>1</sup>

(1. School of Digital Media, Jiangnan University, Wuxi 214122, China; 2. School of Electronics and Computer Science Technology, North University of China, Taiyuan 030051, China)

**Abstract:** Researches on current feature extraction methods are mainly based on two ways. One originates from geometric properties of high-dimensional datasets and attempt to extract fewer features from the original data space according to a certain criterion. The other originates from dimension reduction deviation and try to make the deviation between data before and after dimension reduction be as small as possible. However, there exists almost no study about them from the perspective of the scatter change of a dataset. Based on Parzen window density estimator, we thoroughly revisit the relevant feature extraction methods from a new perspective and the relationships between Parzen window and classical feature extraction methods, ie length of perpendiculars (LPP), linear discriminant analysis (LDA) and principal component analysis (PCA) are built in this paper. Therefore, these feature extraction methods can be researched in the same Parzen window, which provides a new perspective for the research of feature extraction.

**Keywords:** feature extraction; Parzen window; density estimation; data distribution characteristics; new perspective

在实际应用中常常遇到高维数据,如多媒体数据、文本数据、生物数据等.尽管高维数据比低维数据拥有更多的信息量,但在实际操作中非常不便.一种有效的解决方法就是对其进行特征提取.所谓特征提取是指原始特征空间根据某种准则变换得到低维投影空间的过程.近年来,研究人员提出众多特征提取方法(feature extraction methods, FEM):非负矩阵分解(non-negative matrix factorization, NMF)<sup>[1]</sup>通

过将原始特征空间低秩近似,保证降维后的特征非负;主成分分析(principal component analysis, PCA)<sup>[2]</sup>通过对原始特征空间方差的研究得到一组正交的主成分;奇异值分解(singular value decomposition, SVD)<sup>[3]</sup>通过考察奇异值的贡献率实现降维;独立成分分析(independent component analysis, ICA)<sup>[4]</sup>同时利用原始特征空间的二阶和高阶统计信息,进一步提高了 PCA 的降维效率;线性判别分析(linear discriminant analysis, LDA)<sup>[5]</sup>通过最大化类间离散度和类内离散度的广义 Rayleigh 熵实现特征变换;多维缩放(multi-dimensional scaling,

收稿日期:2012-05-14. 网络出版日期:2012-11-16.

基金项目:国家自然科学基金资助项目(61170122, 61272210).

通信作者:刘忠宝. E-mail:liu\_zhongbao@hotmail.com.

MDS)<sup>[6]</sup>通过保持数据点间的相似性实现降维;局部线性嵌入(locally linear embedding, LLE)<sup>[7]</sup>利用稀疏矩阵算法实现降维;拉普拉斯特征映射(Laplacian eigenmap)<sup>[8]</sup>利用谱技术实现降维;典型相关分析(canonical correlation analysis, CCA)<sup>[9]</sup>利用综合变量对之间的关系来反映2组指标之间的整体相关性;局部保持投影(locally preserving projections, LPP)<sup>[10]</sup>通过保持原始特征空间内在的局部几何结构实现降维。

在实际应用中,针对不同的应用场合,本课题组提出若干改进算法:皋军等<sup>[11]</sup>利用 Fisher 准则中的类内散度重新构造势支撑向量机的目标函数,提出广义势支撑特征选择方法(generalized potential support features selection method, GPSFM),该方法具有冗余度低、选择速度快等优点;王晓明等<sup>[12]</sup>针对监督的局部保持投影算法在小样本情况下矩阵的奇异性问题,提出广义的局部保留投影算法(generalized supervised locality preserving projection, GSLPP),该方法有效地解决了小样本问题;王超等<sup>[13]</sup>提出有局部保持的最大间距准则特征提取方法,该方法能使投影后的样本具有最大类内散度和最小类内散度,在解决小样本问题的同时更好地保持人脸图像的局部流形结构。

上述特征降维方法主要从空间几何和降维误差角度进行研究。从空间几何角度看,特征提取实质上是根据一定规则最优化缩小特征空间的过程;从降维误差角度看,特征提取的目的是保证降维前后数据的偏差尽可能小。目前大多数特征提取方法重点关注几何性质和降维误差,而对于降维过程中数据分布特征的变化往往重视不够。鉴于此,本文引入 Parzen 窗核密度估计表征数据的分布特征,通过分析降维过程中数据分布特征的变化,揭示 Parzen 窗估计与典型特征提取方法 LPP、LDA 和 PCA 之间的关系,从而说明这些特征提取方法可统一在 Parzen 窗框架下进行研究,为特征提取方法的研究提供了一个新的视角。

## 1 Parzen 窗

在统计学和其他相关领域,核密度估计备受关注。Parzen、Silverman、Rosenble 等是当前主流的核密度估计方法<sup>[14]</sup>。Parzen 窗法具有优良性能<sup>[15]</sup>,因而被广泛使用。本文采用 Parzen 窗表征数据的分布形状。

Parzen 窗定义为

$$p(\mathbf{x}) = \sum_{i=1}^N \alpha_i K_{\delta}(\mathbf{x}, \mathbf{x}_i),$$

$$\text{s. t. } \sum_{i=1}^N \alpha_i = 1,$$

$$\alpha_i \geq 0, i = 1, 2, \dots, N.$$

式中: $K_{\delta}(\mathbf{x}, \mathbf{x}_i)$ 为窗宽为 $\delta$ 的核函数, $\alpha_i (i = 1, 2, \dots, N)$ 为系数。其中,核函数 $K_{\delta}$ 必须满足:

$$1) K_{\delta}(t) \geq 0;$$

$$2) \int K_{\delta}(t) dt = 1.$$

常用的核函数有:

1) Gaussian 核函数  $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / 2\delta^2)$ , 参数  $\delta > 0$ 。

2) 多项式核函数  $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + c)^d$ ,  $c$  和  $d$  为参数。当  $c = 1$  且  $d = 0$  时,  $K(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$ 。此时多项式核为线性核。

3) Sigmoid 核函数  $K(\mathbf{x}, \mathbf{y}) = \tanh(b(\mathbf{x}^T \mathbf{y}) + c)^d$ ,  $b(\mathbf{x}^T \mathbf{y}) > 0$ , 且  $c < 0$ 。

4) Epanechnikov 核函数  $K(\mathbf{x}) = \frac{3}{4d} (1 - \frac{\|\mathbf{x}\|^2}{d^2})$ ,  $\|\mathbf{x}\|^2 / d^2 \leq 1$ ,  $d$  为参数。

Sigmoid 核函数在特定参数下与 Gaussian 核函数具有近似性能;多项式核函数参数较多,不易确定,且当阶数  $d$  较高时运算可能出现溢出。因此本文重点关注 Gaussian 核函数和 Epanechnikov 核函数,它们均为最小均方差意义下的最优核函数。通过合理的参数选择,它们可用于任意分布的数据。

## 2 Parzen 窗与 LPP

大多数特征提取方法主要从空间几何和降维误差角度进行研究,而对于降维过程中数据分布特征的变化则关注不够。本文从降维过程中数据分布特征的变化入手,基于广泛使用的 Parzen 窗核密度估计方法,通过探讨 Parzen 窗估计与典型特征提取方法 LPP、LDA 和 PCA 之间的关系,进而说明这些特征提取方法可统一在 Parzen 窗框架下进行研究。本节首先讨论 Parzen 窗与 LPP 之间的关系。

设有样本  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N] \in \mathbf{R}^{N \times N}$ , 其中  $\mathbf{x}_i (i = 1, 2, \dots, N) \in \mathbf{R}^N$  表示第  $i$  个样本,  $N$  表示样本总数。

### 2.1 LPP

LPP<sup>[16]</sup>在保持样本局部流形结构不变的基础上实现特征提取,即在高维空间的邻近样本点在低维空间的相对关系保持不变。算法流程如下:

1) 创建邻接图。如果样本点  $\mathbf{x}_i$  和  $\mathbf{x}_j$  满足  $\|\mathbf{x}_i - \mathbf{x}_j\|^2 < \varepsilon$ , 则将两点相连;

2) 计算权值。权值计算公式为

$$S_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t).$$

3) 求解投影矩阵  $\mathbf{W}$ .  $\mathbf{W}$  可通过最小化如下目标函数得到:

$$\min_{\mathbf{W}} \sum_{i,j} (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j)^2 S_{ij}.$$

对上式进行代数变换有

$$\begin{aligned} & \frac{1}{2} \sum_{i,j} (\mathbf{W}^T \mathbf{x}_i - \mathbf{W}^T \mathbf{x}_j)^2 S_{ij} = \\ & \sum_{i,j} \mathbf{W}^T \mathbf{x}_i \mathbf{D}_{ii} \mathbf{x}_i^T \mathbf{W} - \sum_{i,j} \mathbf{W}^T \mathbf{x}_i \mathbf{S}_{ij} \mathbf{x}_j^T \mathbf{W} = \\ & \mathbf{W}^T \mathbf{X}(\mathbf{D} - \mathbf{S})\mathbf{X}\mathbf{W} = \mathbf{W}^T \mathbf{X}\mathbf{L}\mathbf{X}\mathbf{W}. \end{aligned}$$

式中:  $\mathbf{D}$  为对角阵, 且  $\mathbf{D}_{ii} = \sum_j \mathbf{S}_{ij}$ ,  $\mathbf{L} = \mathbf{D} - \mathbf{S}$ .

约束条件为

$$\mathbf{W}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W} = 1.$$

上述优化问题可转化为求解如下广义特征值问题:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{W} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{W}.$$

式中: 前  $d$  个最小非零特征值对应的特征向量构成的投影矩阵  $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \cdots \ \mathbf{w}_d]$ .

## 2.2 Parzen 窗与 LPP 的关系

从概率密度角度看, 特征提取过程可以看作是数据分布特征保持的过程. 用 Parzen 窗表征数据的分布特征并保证降维前后数据分布尽量保持一致.

Parzen 窗可定义为

$$p_{\delta}(\mathbf{x}) = \frac{1}{N} \sum_{k=1}^c \sum_{j=1}^{N_k} \frac{1}{\sqrt{2\pi}\delta_k} \exp(-\|\mathbf{x} - \mathbf{x}_{kj}\|^2 / 2\delta_k^2). \quad (1)$$

式中: 方差  $\delta_k$  表明各类样本偏离其类中心的程度, 方差  $\delta_k$  也称为散度. 在降维过程中, 方差  $\delta_k$  的变化速度越小, 则数据的分布特征变化越小. 若数据服从高斯分布, 图 1 中的  $G_1$  和  $G_2$  表明降维前后数据分布的变化情况. 降维后的数据在一定程度上进行了缩减, 因此散度有变小的趋势.

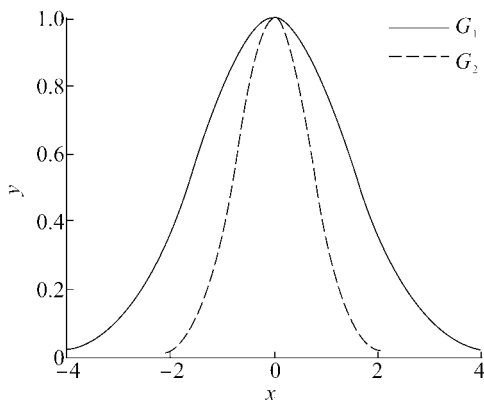


图 1 降维前后数据分布变化

Fig. 1 The transformation of data distribution before and after dimension reduction

降维过程中散度的变化速度用  $\frac{\partial p_{\delta}}{\partial \delta_k}$  表示. 对

式(1)求导可得

$$\frac{\partial p_{\delta}}{\partial \delta_k} = \frac{1}{N} \sum_{j=1}^{N_k} \frac{1}{\sqrt{2\pi}\delta_k^2} \exp(-\|\mathbf{x} - \mathbf{x}_{kj}\|^2 / 2\delta_k^2) (\|\mathbf{x}_i - \mathbf{x}_{kj}\|^2 / \delta_k^2 - 1). \quad (2)$$

针对某类数据降维,  $\delta_k$  已知. 为了表示方便, 将  $N_k$  表示为  $N$ ,  $\delta_k$  表示为  $\delta$ , 去掉式(2)中常数项有

$$\frac{\partial p_{\delta}}{\partial \delta} = \sum_{j=1}^N \|\mathbf{x} - \mathbf{x}_j\|^2 \exp(-\|\mathbf{x} - \mathbf{x}_j\|^2 / 2\delta^2). \quad (3)$$

令式(3)中  $\mathbf{x} \in \{\text{训练样本}\}$ , 则式(3)变为

$$\frac{\partial p_{\delta}}{\partial \delta} = \sum_{i,j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|^2 \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\delta^2). \quad (4)$$

为了使降维后的数据尽量保持原始空间的分布特征, 因此投影方向  $\mathbf{W}$  应满足散度的变化速度  $\frac{\partial p_{\delta}}{\partial \delta_k}$  最小, 可得目标函数:

$$\min_{\mathbf{W}} \mathbf{I}^T \mathbf{W}^T \mathbf{T} \mathbf{W} \mathbf{I}. \quad (5)$$

式中:  $\mathbf{T} = \sum_{i,j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|^2 \exp(\mathbf{W}^T (-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\delta^2) \mathbf{W})$ ,  $\mathbf{I} = [1 \ 1 \ \cdots \ 1]^T$ . 向量  $\mathbf{I}$  保证式(4)投影后仍为标量. 由于  $\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\delta^2)$  可以反映样本点  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间关系, 因此为了计算方便, 式(5)中  $\mathbf{T}$  可简化为

$$\mathbf{T}^* = \sum_{i,j=1}^N \|\mathbf{x}_i - \mathbf{x}_j\|^2 \exp(\mathbf{W}^T (-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\delta^2) \mathbf{W}).$$

式(5)可转化为

$$\min_{\mathbf{W}} \mathbf{I}^T \mathbf{W}^T \mathbf{T}^* \mathbf{W} \mathbf{I}. \quad (6)$$

由于  $\mathbf{I}$  与  $\mathbf{W}$  无关, 因此式(6)可表示为

$$\min_{\mathbf{W}} \mathbf{W}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{W}.$$

式中:  $\mathbf{S}_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\delta^2)$ ,  $\mathbf{D}_{ii} = \sum_{j=1}^N \mathbf{S}_{ij}$ ,  $\mathbf{L} = \mathbf{D} - \mathbf{S}$ .

$\mathbf{D}$  反映样本点  $\mathbf{x}_i$  附近的局部密度.  $\mathbf{D}$  越大, 则  $\mathbf{x}_i$  局部密度越大. 因此对其约束有

$$\mathbf{W}^T \mathbf{X}^T \mathbf{D} \mathbf{X} \mathbf{W} = 1.$$

基于上述分析, 上述最优化问题可归纳为

$$\begin{aligned} & \min_{\mathbf{W}} \mathbf{W}^T \mathbf{X}^T \mathbf{L} \mathbf{X} \mathbf{W}, \\ & \text{s. t. } \mathbf{W}^T \mathbf{X}^T \mathbf{D} \mathbf{X} \mathbf{W} = 1. \end{aligned} \quad (7)$$

为了表述方便, 将上述优化问题称为基于 Parzen 窗的特征提取方法 (feature extraction method based on Parzen window, FEMPW).

由 LPP 算法不难看出:当样本点  $\mathbf{x}_i$  和  $\mathbf{x}_j$  满足  $\|\mathbf{x}_i - \mathbf{x}_j\|^2 < \varepsilon$  时,将 2 点相连并计算相应权重.当  $\mathbf{x} \in \{\text{训练样本}\}$  时,令 LPP 的权重函数为

$$S = \frac{1}{\sqrt{2\pi\delta}} \exp(-\|\mathbf{x} - \mathbf{x}_j\|^2 / 2\delta^2).$$

且满足

$$\int_{-\varepsilon}^{\varepsilon} \sum_{k=1}^c \sum_{j=1}^{N_k} \frac{C}{\sqrt{2\pi\delta_k}} \exp(-\|\mathbf{x} - \mathbf{x}_{kj}\|^2 / 2\delta_k^2) d\mathbf{x} = 1. \quad (8)$$

式中:  $C$  为常数.

求解式(8)可得

$$C = 1/N(2\Phi(\varepsilon) - 1).$$

$$\text{式中: } \Phi(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx.$$

对于任意给定的  $\varepsilon$  值,通过查表可求得  $\Phi(\varepsilon)$ , 进而得到  $C$  值.当  $\varepsilon \rightarrow \infty$  时,  $C = 1/N$ , 此时 FEMPW 等价于 LPP. 也就是说, LPP 是 FEMPW 的特例.

### 3 Parzen 窗与 LDA 和 PCA

LDA<sup>[5]</sup> 和 PCA<sup>[17]</sup> 是特征提取中最为经典和广泛使用的方法. 针对于 LDA 和 PCA, 科学家们进行了广泛研究并取得了众多成果. 与已有研究思路不同, 本文主要贡献在于: 1) 出发点不同, 目前已有研究主要从空间几何和最小误差角度展开, 而本文从降维过程中数据分布特征的变化入手对其进行探讨; 2) 解释模型不同, 目前已有研究主要基于空间几何和最小误差 2 种解释模型, 而本文通过 Parzen 窗建立概率密度估计与 LDA 和 PCA 的关系, 从概率密度角度对其进行解释.

#### 3.1 LDA

##### 3.1.1 几何解释

从几何角度看, LDA 的目的是将高维特征降到低维空间, 保证在低维空间异类样本尽量分开, 同类样本尽量紧密.

样本的类间离散度矩阵和类内离散度矩阵分别定义为

$$S_B = \sum_{i=1}^c N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T, \quad (9)$$

$$S_W = \sum_{i=1}^c \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i) (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)^T. \quad (10)$$

式中:  $\mathbf{x}_{ij} \in \mathbf{R}^n$  ( $i=1, 2, \dots, c; j=1, 2, \dots, N_i$ ) 表示第  $i$  类中的第  $j$  个样本,  $N_i$  表示第  $i$  类样本个数,  $c$  表示样本类别总数. 所有样本的均值  $\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$ ; 设第  $i$

类的样本均值为  $\bar{\mathbf{x}}_i$  ( $i=1, 2, \dots, c$ ), 则有  $\bar{\mathbf{x}} = \sum_{i=1}^c \frac{N_i}{N} \bar{\mathbf{x}}_i$ .

LDA 的最优化表达式为

$$J(\mathbf{W}) = \max_{\mathbf{W}} \frac{\mathbf{W}^T S_B \mathbf{W}}{\mathbf{W}^T S_W \mathbf{W}}. \quad (11)$$

由拉格朗日乘子法可得: 求解式(11)等价于求解一般矩阵  $S_W^{-1} S_B$  的特征值问题.

##### 3.1.2 最小误差解释

以二分类问题为例, 从最小误差角度解释 LDA.

1) 用 Bhattacharyya 系数确定二分类问题的错误率上界.

由文献[18]可知, 为了使二分类问题错误率上界尽可能小, 必须保证 Bhattacharyya 系数尽可能大. 用  $J_B$  表示 Bhattacharyya 系数, 有

$$J_B = -\ln \int \sqrt{p(\mathbf{x} | \omega_1) p(\mathbf{x} | \omega_2)} d\mathbf{x}.$$

假设 2 类样本服从正态分布:  $\mathbf{x} | \mathbf{y} = 0 \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ ,  $\mathbf{x} | \mathbf{y} = 1 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$  则有

$$J_B = \frac{1}{8} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T (\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2})^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) + \frac{1}{2} \ln \frac{|\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2|/2}{|\boldsymbol{\Sigma}_1|^{\frac{1}{2}} + |\boldsymbol{\Sigma}_2|^{\frac{1}{2}}}. \quad (12)$$

当  $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$  时, 式(12)可转化为

$$J_B = \frac{1}{8} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1).$$

特别地, 当  $\boldsymbol{\Sigma} = \mathbf{E}$  时,

$$J_B = \frac{1}{8} (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1).$$

2) 样本的最大似然估计.

最大似然估计函数可表示为

$$\begin{aligned} L(\phi, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &= \\ \log \prod_{i=1}^N p(\mathbf{x}_i, \mathbf{y}_i; \phi, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &= \\ \log \prod_{i=1}^N p(\mathbf{x}_i | \mathbf{y}_i; \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) p(\mathbf{y}_i; \phi). \end{aligned}$$

求导并令导数为零可得

$$\begin{aligned} \boldsymbol{\Sigma} &= \sum_{i=1}^{N_1} (\mathbf{x}_i - \boldsymbol{\mu}_1) (\mathbf{x}_i - \boldsymbol{\mu}_1)^T + \\ &\quad \sum_{i=1}^{N_2} (\mathbf{x}_i - \boldsymbol{\mu}_2) (\mathbf{x}_i - \boldsymbol{\mu}_2)^T. \end{aligned}$$

式中:  $\boldsymbol{\Sigma}$  是样本特征方差均值.

为了使错误率的上界和样本散度尽可能小, 投影方向  $\mathbf{W}$  应满足

$$J(\mathbf{W}) = \max_{\mathbf{W}} \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}},$$

式中:

$$\begin{aligned} \mathbf{S}_B &= (\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1)^T, \\ \mathbf{S}_W &= \sum_{i=1}^{N_1} (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^T + \\ &\quad \sum_{i=1}^{N_2} (\mathbf{x}_i - \boldsymbol{\mu}_2)(\mathbf{x}_i - \boldsymbol{\mu}_2)^T. \end{aligned}$$

上述最优化问题可转化为求解  $\mathbf{S}_W^{-1} \mathbf{S}_B$  的特征值问题. 同理, 对于  $c (c > 2)$  类问题, 类间离散度矩阵和类内离散度矩阵的定义分别同式(9)和式(10).

### 3.2 Parzen 窗与 LDA 的关系

本节从概率密度角度讨论 Parzen 窗与 LDA 的关系. Epanechnikov 核函数<sup>[19]</sup>是在最小均方差意义下最优的核函数之一. 受 Epanechnikov 核函数的启发, 采用如下核函数进行密度估计:

$$p_\delta(\mathbf{x}) = 1 - \frac{\sum_{i=1}^c \|\mathbf{x} - \bar{\mathbf{x}}_i\|^2}{\delta^2}, \frac{\sum_{i=1}^c \|\mathbf{x} - \bar{\mathbf{x}}_i\|^2}{\delta^2} \leq 1. \quad (13)$$

为了计算散度的变化速度, 对式(13)求导可得

$$\frac{\partial p_\delta}{\partial \delta} = \frac{2}{\delta^3} \|\mathbf{x} - \bar{\mathbf{x}}_i\|^2, \quad i = 1, 2, \dots, c.$$

针对某类数据降维, 则  $\delta$  已知. 当  $\mathbf{x} \in \{\text{训练样本}\}$  时, 则  $\bar{\mathbf{x}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \mathbf{x}_{ij}, i = 1, 2, \dots, c.$

为了保证降维前后数据分布的稳定性, 因此投影方向  $\mathbf{W}$  应满足使得  $\frac{\partial p_\delta}{\partial \delta}$  最小. 去掉与  $\delta$  有关的常数项, 则有

$$\min_{\mathbf{W}} \mathbf{I}^T \mathbf{W}^T \sum_{j=1}^{N_i} \|\mathbf{x}_{ij} - \bar{\mathbf{x}}_i\|^2 \mathbf{W} \mathbf{I}. \quad (14)$$

式中: 向量  $\mathbf{I}$  保证式(14)为标量. 式(14)中  $\mathbf{I}$  与  $\mathbf{W}$  无关, 并综合考虑各类样本, 可得

$$\min_{\mathbf{W}} \mathbf{W}^T \mathbf{S}_W \mathbf{W}. \quad (15)$$

式中:  $\mathbf{S}_W = \sum_{i=1}^c \sum_{j=1}^{N_i} \|\mathbf{x}_{ij} - \bar{\mathbf{x}}_i\|^2.$

下面求类间离散度矩阵  $\mathbf{S}_B$ . 对式(13)进行如下数学变换:

$$\begin{aligned} p_\delta(\mathbf{x}) &= 1 - \frac{\sum_{i=1}^c \sum_{j=1}^{N_i} \|\mathbf{x}_{ij} - \bar{\mathbf{x}}_i\|^2}{\delta^2} = \\ &= 1 - \frac{\sum_{i=1}^c \|N_i \bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|^2}{\delta^2} \leq \end{aligned}$$

$$1 - \frac{\sum_{i=1}^c N_i^2 \|\bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|^2}{\delta^2} \leq 1 - \frac{\sum_{i=1}^c N_i \|\bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|^2}{\delta^2}.$$

可得

$$p_\delta(\mathbf{x}) \leq 1 - \frac{\sum_{i=1}^c N_i \|\bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|^2}{\delta^2}. \quad (16)$$

为了表示方便, 取式(16)的上界并求导可得

$$\frac{\partial p_\delta}{\partial \delta} = \frac{2}{\delta^3} \|\bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|^2, \quad i = 1, 2, \dots, c.$$

为了使降维后的各类样本尽可能隔开, 因此投影方向  $\mathbf{W}$  应满足使得各类  $\frac{\partial p_\delta}{\partial \delta}$  最大. 去掉与  $\delta$  有关的常数项, 则有

$$\max_{\mathbf{W}} \mathbf{I}^T \mathbf{W}^T \mathbf{S}_B \mathbf{W} \mathbf{I}. \quad (17)$$

式中:  $\mathbf{S}_B = \sum_{i=1}^c N_i \|\bar{\mathbf{x}} - \bar{\mathbf{x}}_i\|^2.$

由于式(17)中  $\mathbf{I}$  与  $\mathbf{W}$  无关, 则有

$$\max_{\mathbf{W}} \mathbf{W}^T \mathbf{S}_B \mathbf{W}, \quad (18)$$

综合式(15)和式(18)可得

$$J(\mathbf{W}) = \max_{\mathbf{W}} \frac{\mathbf{W}^T \mathbf{S}_B \mathbf{W}}{\mathbf{W}^T \mathbf{S}_W \mathbf{W}}.$$

### 3.3 PCA

#### 3.3.1 几何解释

从几何角度看, PCA 通过计算方差来衡量样本包含信息量的大小. PCA 称方差较大的分量为主成分, 主成分反映了原始变量的主要信息.

设投影方向为  $\mathbf{W}$ , 为了使投影后样本的总体离散度最大, 由方差的定义可得

$$\begin{aligned} J(\mathbf{W}) &= \max_{\mathbf{W}} \frac{1}{N} \sum_{i=1}^N ((\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{W})^2 = \\ &= \max_{\mathbf{W}} \frac{1}{N} \sum_{i=1}^N \mathbf{W}^T (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{W} = \\ &= \max_{\mathbf{W}} \mathbf{W}^T \left( \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \right) \mathbf{W}. \end{aligned}$$

令  $\mathbf{S}_T = \frac{1}{N} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T$ , 则上式变为

$$J(\mathbf{W}) = \max_{\mathbf{W}} \mathbf{W}^T \mathbf{S}_T \mathbf{W}.$$

$\mathbf{W}$  可通过求解如下特征值问题得到:

$$\mathbf{S}_T \mathbf{W} = \lambda \mathbf{W}.$$

#### 3.3.2 最小误差解释

设样本点  $\mathbf{x}_i$  降维后变为  $\mathbf{x}_i'$ , 希望降维后的样本点与原样本点之间的误差尽可能小, 则引入最小平方误差:

$$\sum_{i=1}^N \| \mathbf{x}_i' - \mathbf{x}_i \|^2.$$

假设降维后的样本点  $\mathbf{x}_0$  代表所有降维后的样本点:

$$\begin{aligned} J_0(\mathbf{x}_0) &= \sum_{i=1}^N \| \mathbf{x}_0 - \mathbf{x}_i \|^2 = \\ &= \sum_{i=1}^N \| (\mathbf{x}_0 - \bar{\mathbf{x}}) - (\mathbf{x}_i - \bar{\mathbf{x}}) \|^2 = \\ &= \sum_{i=1}^N \| \mathbf{x}_0 - \bar{\mathbf{x}} \|^2 - 2 \sum_{i=1}^N (\mathbf{x}_0 - \bar{\mathbf{x}})^T (\mathbf{x}_i - \bar{\mathbf{x}}) + \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2 = \\ &= \sum_{i=1}^N \| \mathbf{x}_0 - \bar{\mathbf{x}} \|^2 - 2 (\mathbf{x}_0 - \bar{\mathbf{x}})^T \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}}) + \\ &= \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2 = \sum_{i=1}^N \| \mathbf{x}_0 - \bar{\mathbf{x}} \|^2 + \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2. \end{aligned} \quad (19)$$

式(19)第2项与  $\mathbf{x}_0$  无关,可看作常数,因此最小化  $J_0(\mathbf{x}_0)$  有  $\mathbf{x}_0 = \bar{\mathbf{x}}$ , 即  $\mathbf{x}_0$  为样本均值.

设降维后的样本点可表示为

$$\mathbf{x}_i' = a_i \mathbf{W} + \bar{\mathbf{x}}. \quad (20)$$

式中:  $a_i$  是  $\mathbf{x}_i'$  到  $\bar{\mathbf{x}}$  的距离.

将式(20)代入式(19)有

$$\begin{aligned} J_1(a_1, a_2, \dots, a_N, \mathbf{W}) &= \sum_{i=1}^N \| \mathbf{x}_i' - \mathbf{x}_i \|^2 = \\ &= \sum_{i=1}^N \| (a_i \mathbf{W} + \bar{\mathbf{x}}) - \mathbf{x}_i \|^2 = \sum_{i=1}^N \| a_i \mathbf{W} - (\mathbf{x}_i - \bar{\mathbf{x}}) \|^2 = \\ &= \sum_{i=1}^N a_i^2 \| \mathbf{W} \|^2 - 2 \sum_{i=1}^N a_i \mathbf{W}^T (\mathbf{x}_i - \bar{\mathbf{x}}) + \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2. \end{aligned}$$

由  $\| \mathbf{W} \|^2 = 1$ , 对  $a_i$  求导并令导数为零, 有

$$a_i = \mathbf{W}^T (\mathbf{x}_i - \bar{\mathbf{x}}).$$

固定  $a_i$ , 对  $\mathbf{W}$  求偏导, 并将上式代入  $J_1$  有

$$\begin{aligned} J_1(a_1, a_2, \dots, a_N, \mathbf{W}) &= \\ &= \sum_{i=1}^N a_i^2 \| \mathbf{W} \|^2 - 2 \sum_{i=1}^N a_i \mathbf{W}^T + \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2 = \\ &= - \sum_{i=1}^N (\mathbf{W}^T (\mathbf{x}_i - \bar{\mathbf{x}}))^2 + \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2 = \\ &= - \sum_{i=1}^N (\mathbf{W}^T (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{W}) + \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2 = \\ &= - \mathbf{W}^T \mathbf{S}_T \mathbf{W} + \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2. \end{aligned}$$

式中:  $\mathbf{S}_T = (\mathbf{x}_i - \bar{\mathbf{x}}) (\mathbf{x}_i - \bar{\mathbf{x}})^T$ .

由于  $\| \mathbf{W} \|^2 = 1$ , 引入拉格朗日乘子法可得

$$\mathbf{S}_T \mathbf{W} = \lambda \mathbf{W}. \quad (21)$$

$\mathbf{W}$  可由式(21)求得.

### 3.4 Parzen 窗与 PCA 的关系

针对某类数据降维, 令  $\mathbf{x}_i = \bar{\mathbf{x}}$ . 则式(13)可变为

$$p_\delta(\mathbf{x}_i) = 1 - \frac{\sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2}{\delta^2}. \quad (22)$$

对式(22)求导可得

$$\frac{\partial p_\delta}{\partial \delta} = \frac{2 \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2}{\delta^2}, i = 1, 2, \dots, N.$$

去掉与  $\delta$  有关的常数项有

$$\frac{\partial p_\delta}{\partial \delta} = \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2.$$

为了保证 PCA 提取出的主成分包含尽可能多的信息, 因此投影方向  $\mathbf{W}$  应保证所有样本的  $\frac{\partial p_\delta}{\partial \delta}$  尽可能大. 基于上述分析可得:

$$\max_{\mathbf{W}} \mathbf{I}^T \mathbf{W}^T \mathbf{S}_T \mathbf{W}. \quad (23)$$

式中:  $\mathbf{S} = \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2 (i = 1, 2, \dots, N)$ , 向量  $\mathbf{I}$  保证式(23)为标量,  $\mathbf{I}$  与  $\mathbf{W}$  无关. 综合考虑各类样本, 可得

$$\max_{\mathbf{W}} \mathbf{W}^T \mathbf{S}_T \mathbf{W}. \quad (24)$$

式中:  $\mathbf{S}_T = \frac{1}{N} \sum_{i=1}^N \| \mathbf{x}_i - \bar{\mathbf{x}} \|^2$ .

式(24)可转化为求解特征值问题, 则投影方向  $\mathbf{W}$  由  $\mathbf{S}_T$  的特征向量组成.

## 4 实验分析

为了进一步考察 Parzen 窗与主流特征提取方法的关系, 分别在人工数据集和人脸数据集上对算法进行测试. 实验的基本流程如图2所示.

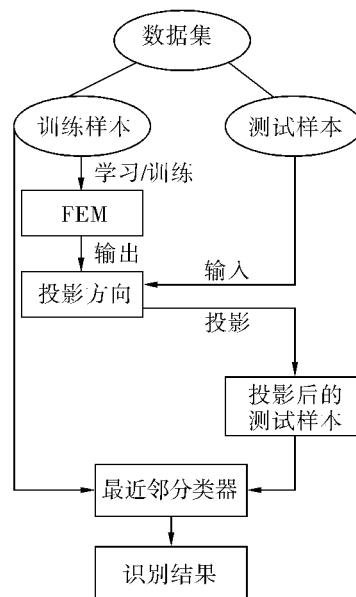
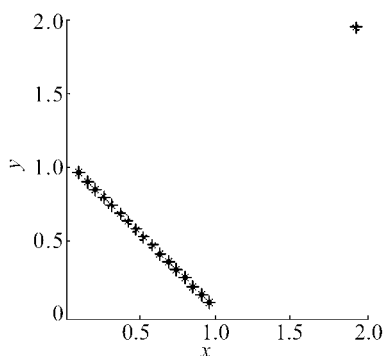


图2 实验流程

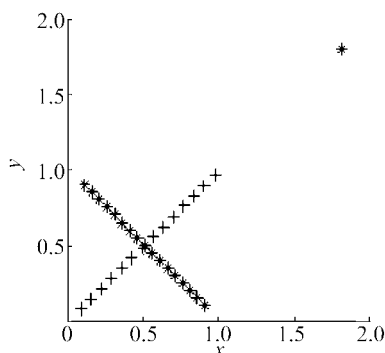
Fig.2 The process of experiments

#### 4.1 人工数据集

为了直观地考察 FEMPW 的有效性,人为构造一些数据集.图 3(a)为实验数据集,图 3(b)为实验结果.实验结果表明,FEMPW 很好地发现了数据的本质特征.图 3(b)中由“ $\times$ ”组成的线段表示 FEMPW 发现的主特征,它忽略了右上角的数据点.由此可见,FEMPW 能发现数据的局部特征,因而找到了更适合的特征方向.



(a) 人工数据集 1

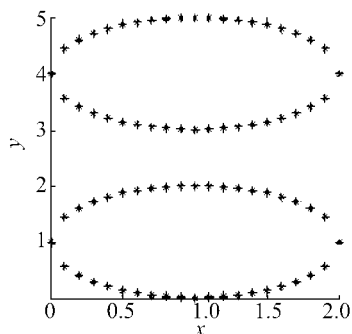


(b) 实验结果

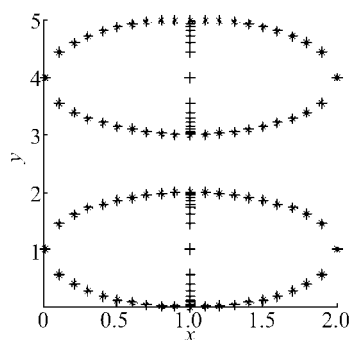
图 33 人工数据集 1 及实验结果

Fig. 3 Man-made dataset 1 and experimental results

图 3 中数据的本质特征是 2 个几何圆形,FEMPW 发现的主特征在竖直方向(投影点用“ $+$ ”表示),该方向上数据被很好地分开.实验结果表明,FEMPW 能识别数据的本质特征.



(a) 人工数据集 2



(b) 实验结果

图 4 人工数据集 2 及实验结果

Fig. 5 Man-made dataset 2 and experimental results

#### 4.2 人脸数据集

##### 4.2.1 ORL

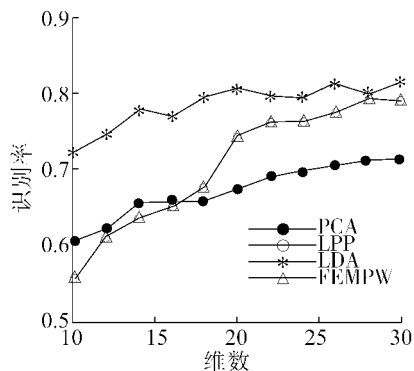
ORL 人脸数据集包含 40 个人的人脸灰度图像,每人 10 幅,共 400 幅.图 5 列举了 ORL 人脸数据集部分人脸图像.



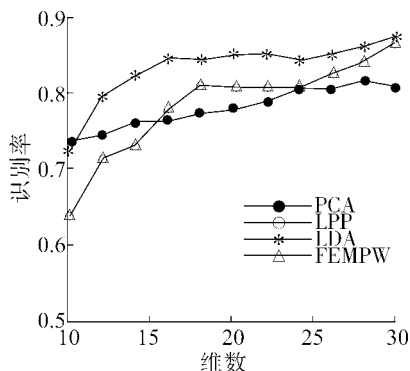
图 5 ORL 人脸数据集部分人脸图像

Fig. 5 Part images of ORL facial datasets

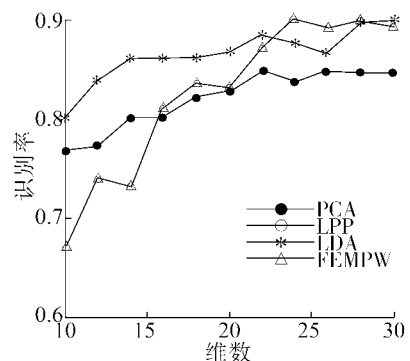
实验选取每人的前  $k$  幅图像作为训练样本,剩余图像作为测试样本. $k=3,4,\dots,8$ ,降维数分别取 10, 12,  $\dots$ , 30.  $k$  取各值时,PCA、LPP、LDA 和 FEMPW 的识别率与降维数的关系如图 6 所示,各算法最优识别率记录于表 1.表 1 中,括号中的维数表明取到最优识别率时的维数.实验中 LDA 采用 PCA + LDA 方法.



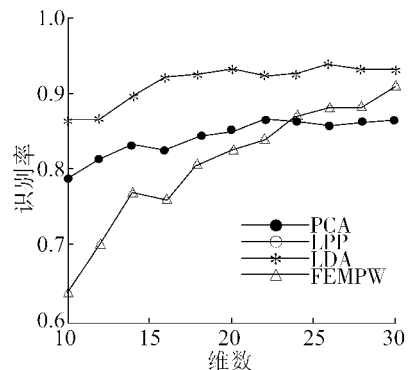
(a)  $k=3$  的训练



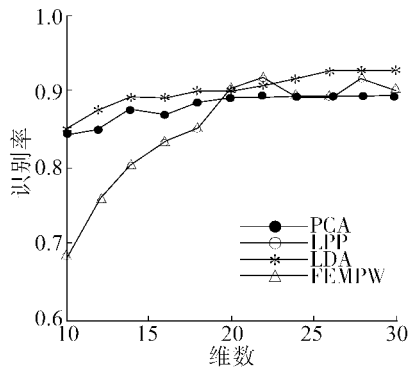
(b)  $k=4$  的训练



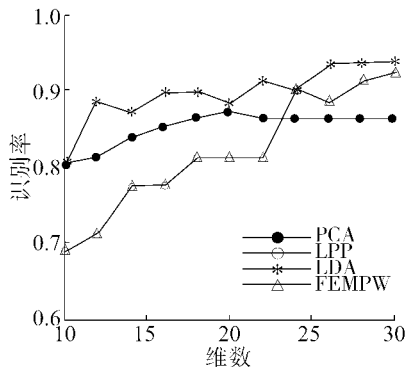
(c)  $k=5$  的训练



(d)  $k=6$  的训练



(e)  $k=7$  的训练



(f)  $k=8$  的训练

图6 ORL 人脸数据集上各算法识别率与降维数的关系

Fig.6 Relation between recognition accuracy and reduced dimension on ORL

表1 ORL 人脸数据集上各算法识别率

Table 1 Recognition accuracy of PCA, LPP, LDA and FEMPW on ORL

尺寸 $k$	PCA	LPP	LDA	FEMPW
3	0.711(28)	0.789(28)	0.814(30)	0.789(28)
4	0.808(28)	0.867(30)	0.875(30)	0.867(30)
5	0.845(22)	0.890(24)	0.905(30)	0.890(24)
6	0.863(22)	0.906(30)	0.950(30)	0.906(30)
7	0.892(20)	0.917(22)	0.925(26)	0.900(22)
8	0.873(20)	0.925(30)	0.938(26)	0.925(30)
平均	0.832	0.882	0.901	0.882

由图6和表1可以看出:随着训练样本数的增加,各算法的识别率基本呈上升趋势.当降维数较高时,FEMPW 具有优良的降维性能,甚至可以达到最优的识别率.从平均性能看,FEMPW 的识别率明显高于 PCA 且略低于 LDA,但亦能较好地完成降维任务.此外,FEMPW 的降维效果等同于 LPP,该实验现象与第2节理论分析得到的“当  $\varepsilon \rightarrow \infty$  时,FEMPW 等价于 LPP”结论一致.

#### 4.2.2 Yale

Yale 人脸数据集共有 15 个人的人脸灰度图像,每人 11 幅,共 165 幅.图7列举了部分 Yale 人脸数据集上的图像.Yale 数据集中人脸的表情、光照等变化比 ORL 数据集要快.



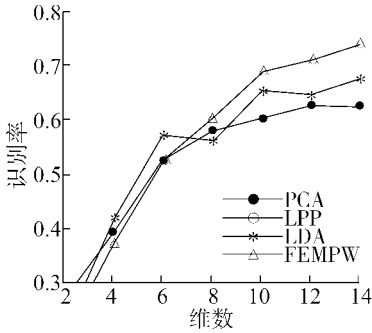
图7 Yale 人脸数据集部分人脸图像

Fig.7 Part of Yale facial datasets

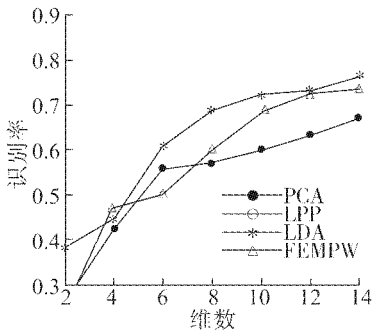
将 Yale 人脸数据集中每个人的前  $k$  幅图像作为训练样本,剩下的图像作为测试样本.实验中  $k = 4, 5, \dots, 8$ ,降维数分别取  $2, 4, \dots, 14$ .  $k$  取各值时,PCA、LPP、LDA 和 FEMPW 的识别率与降维数的关



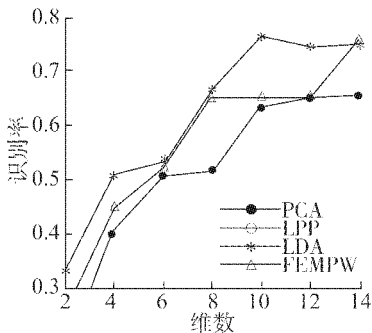
系如图 8 所示,各算法最优识别率记录于表 2. 表 2 中,括号中的维数是取到最优识别率时的维数.



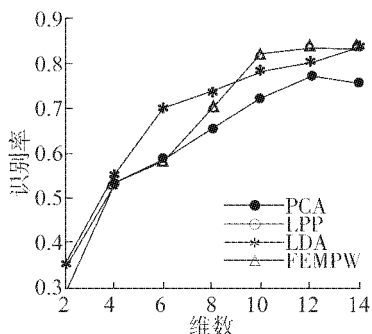
(a)  $k=4$  的训练



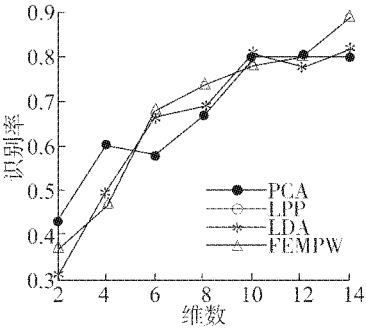
(b)  $k=5$  的训练



(c)  $k=6$  的训练



(d)  $k=7$  的训练



(e)  $k=8$  的训练

图 8 Yale 人脸数据集上各算法识别率与降维数的关系

Fig. 8 Relation between recognition accuracy and reduced dimension on Yale

表 2 Yale 人脸数据集上各算法的识别率

Table 2 Recognition accuracy of PCA, LPP, LDA, FEMPW on ORL

尺寸	PCA	LPP	LDA	FEMPW
4	0.619(12)	0.733(14)	0.667(14)	0.733(14)
5	0.667(14)	0.763(14)	0.767(14)	0.763(14)
6	0.653(12)	0.770(14)	0.747(10)	0.770(14)
7	0.750(12)	0.833(12)	0.833(14)	0.833(12)
8	0.800(10)	0.899(14)	0.822(14)	0.899(14)
平均	0.698	0.800	0.767	0.800

由图 8 和表 2 可以看出:由于 Yale 人脸数据集脸的表情、光照等变化比 ORL 人脸数据集大,因此上述算法的识别率与 ORL 数据集相比有所下降. 随着训练样本数的增加,各算法的识别率基本呈上升趋势. 在训练样本数相同的情况下,FEMPW 具有最优的降维性能. 从平均性能看,FEMPW 的识别率明显高于 PCA 和 LDA,有效地完成了降维任务. 由 ORL 和 Yale 人脸数据集的特点及结果可以看出, FEMPW 更适用于人脸表情与光照变化较大的数据集.

5 结束语

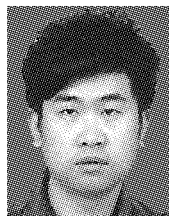
特征提取是模式识别的主要任务之一,它是根据某种准则从原始空间得到一组新特征. 目前大多数特征提取方法关注的是数据的几何性质和降维误差,而数据的分布特征往往被忽略. 鉴于此,本文引入 Parzen 窗核密度估计表征数据的分布特征,从数据散度变化出发,将特征提取过程看作是数据分布特征保持的过程. 通过探讨 Parzen 窗与 LPP、LDA 和 PCA 的关系,说明这 3 种特征提取方法可统一在 Parzen 窗框架下进行研究,为特征提取方法的研究提供了一个新的视角. 如何将其他特征提取方法统一在 Parzen 窗框架下进行研究是本文下一步的工作.

参考文献:

[1] WANG Qinghua, ZHANG Youyun, LEI Cai, et al. Fault diag-

- nosis for diesel value trains based on non-negative matrix factorization and neural network ensemble[J]. *Mechanical Systems and Signal Processing*, 2009, 23(5): 1683-1695.
- [2] CAMACHO J, PIC J, FERRER A. Data understanding with PCA: structural and variance information plots[J]. *Chemometrics and Intelligent Laboratory Systems*, 2010, 100(1): 48-56.
- [3] LIPOVETSKY S. PCA and SVD with nonnegative loadings[J]. *Pattern Recognition*, 2009, 42(1): 68-76.
- [4] RADULOVIC J, RANKOVIC V. Feedforward neural network and adaptive network-based fuzzy inference system in study of power lines[J]. *Expert Systems with Applications*, 2010, 37(1): 165-170.
- [5] PETER N B, JOAO P H, DAVID J K. Eigenfaces vs Fisherfaces: recognition using class specific linear projection[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711-720.
- [6] HU Y H, HE S H. Integrated evaluation method[M]. Beijing: Scientific Press, 2000: 27-31.
- [7] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding[J]. *Science*, 2000(290): 2323-2326.
- [8] BELKIN M, IYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C]// *Systems (NIPS)*. Cambridge: MIT NIIYOGI P Press, 2001: 585-591.
- [9] BORGA M, KNUTSSON H. Canonical correlation analysis in early vision[C]// *Proceedings of the 9th European Symposium on Artificial Neural Networks*. Bruges, Belgium, 2001: 309-314.
- [10] HE X F, NIYOGI P. Locality preserving projections[C]// *Advances in Neural Information Processing Systems (NIPS)*. Vancouver, Canada, 2003: 153-160.
- [11] 皋军, 王士同, 邓赵红. 广义的势支撑特征选择 GPS-FM[J]. *计算机研究与发展*, 2009, 46(1): 41-51.  
GAO Jun, WANG Shitong, DENG Zhaohong. Generalized potential support vector machine for feature selection[J]. *Computer Research and Development*, 2009, 46(1): 41-51.
- [12] 王晓明, 王士同. 广义的监督局部保留投影算法[J]. *电子与信息学报*, 2009, 31(8): 1840-1845.  
WANG Xiaoming, WANG Shitong. Generalized supervised locality preserving projection algorithm[J]. *Electronics and Information*, 2009, 31(8): 1840-1845.
- [13] 王超, 王士同. 有局部保持的最大间距准则特征提取方法[J]. *模式识别与人工智能*, 2009, 22(6): 898-902.
- WANG Chao, WANG Shitong. Maximum margin and locality preserving feature selection method[J]. *Artificial Intelligence and Pattern Recognition*, 2009, 22(6): 898-902.
- [14] ODIOWEI P P, YI Cao. Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations[J]. *IEEE Trans on Industrial Informatics*, 2010, 6(1): 36-45.
- [15] 吴葛铭, 霍剑青, 王晓蒲. 一种基于加权 Parzen 窗的聚类算法[J]. *中国科学技术大学学报*, 2002, 32(5): 546-551.  
WU Geming, HUO Jianqing, WANG Xiaopu. Weighted Parzen window based clustering[J]. *Journal of Chinese University of Sciences and Technologies*, 2002, 32(5): 546-551.
- [16] HE X F, YAN S C, HU Y X, et al. Face recognition using Laplacianfaces[J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2005, 27(3): 328-340.
- [17] CAMACHO J, PIC J, FERRER A. Data understanding with PCA: structural and variance information plots[J]. *Chemometrics and Intelligent Laboratory Systems*, 2010, 100(1): 48-56.
- [18] 边肇祺, 张学工. 模式识别[M]. 2版. 北京清华大学出版社, 2000: 1-30.
- [19] GUO Q C, CHANG X J, CHU H X. Mean-shift of variable window based on the Epanechnikov kernel[C]// *Proceedings of IEEE Int Conf on Mechatronics and Automation*. Piscataway, USA, 2007: 2314-2319.

## 作者简介:



刘忠宝,男,1981年生,博士,主要研究方向为模式识别、机器学习,发表学术论文数十篇。



王士同,男,1964年生,教授,博士生导师,主要研究方向为人工智能与机器学习。