

基于核化图嵌入的最佳鉴别分析与人脸识别*

卢桂馥^{1,2}, 林忠¹, 金忠¹⁺

¹(南京理工大学 计算机科学与技术学院, 江苏 南京 210094)

²(安徽工程大学 计算机与信息学院, 安徽 芜湖 241000)

Optimal Discriminant Analysis Based on Kernel Extension of Graph Embedding and Face Recognition

LU Gui-Fu^{1,2}, LIN Zhong¹, JIN Zhong¹⁺

¹(School of Computer Science and Technology, Nanjing University of Science and Technology, Nanjing 210094, China)

²(School of Computer Science and Information, Anhui Polytechnic University, Wuhu 241000, China)

+ Corresponding author: E-mail: zhongjin@mail.njust.edu.cn

Lu GF, Lin Z, Jin Z. Optimal discriminant analysis based on kernel extension of graph embedding and face recognition. Journal of Software, 2011, 22(7): 1561-1570. <http://www.jos.org.cn/1000-9825/3843.htm>

Abstract: By making use of compressive mapping and isomorphic mapping in the kernel extension of graph embedding, this paper proves that the essence of kernel extension of graph embedding (KGE) is KPCA (kernel principal component analysis) plus all kinds of linear dimension reduction approaches interpreted in a linear extension of graph embedding (LGE). Based on the theory framework, a combined framework, which takes advantage of the discriminant feature in both null and non-null spaces, is developed. Furthermore, every kernel dimensionality reduction algorithm has its own corresponding combined algorithm. The experimental results from ORL, Yale, FERET and PIE face databases show that the proposed methods are better than the original methods in terms of recognition rate.

Key words: kernel extension of graph embedding; optimal discriminant vector; kernel principal component analysis (KPCA); feature extraction; face recognition

摘要: 将压缩映射和同构映射引入核化图嵌入框架(kernel extension of graph embedding, 简称 KGE), 从理论上证明了 KGE 框架内的各种核算法其实质是 KPCA(kernel principal component analysis)+LGE(linear extension of graph embedding, 简称 LGE)框架内的线性降维算法, 并且基于所给出的理论框架提出了一种综合利用零空间和非零空间鉴别信息的组合方法. 任何一种可以用核化图嵌入框架描述的核算法, 都可以有相应的组合方法. 在 ORL, Yale, FERET 和 PIE 人脸数据库上验证了所提出的理论和方法的有效性.

关键词: 核化图嵌入; 最优鉴别矢量; 核主成分分析; 特征抽取; 人脸识别

中图法分类号: TP391 文献标识码: A

人脸识别由于在科学上的挑战性及其潜在的应用, 已成为计算机视觉和模式识别领域中的热门课题. 对于

* 基金项目: 国家自然科学基金(60632050, 60873151, 60973098); 国家高技术研究发展计划(863)(2006AA01Z119)

收稿时间: 2009-10-14; 修改时间: 2010-01-20; 定稿时间: 2010-03-11

人脸识别,其样本维数往往非常高.当样本维数较高时,常常会导致“维数灾难”^[1]的发生.因此,对高维样本进行降维就显得十分必要.基于此,很多降维方法被提了出来.主分量分析(principal component analysis,简称 PCA)^[1]和基于 Fisher 准则的线性鉴别分析(linear discriminant analysis,简称 LDA)^[1,2]是应用最广泛的两种线性降维方法.PCA 是一种无监督学习算法,其目的是寻找在最小均方差意义下最能代表原始数据的投影方向.LDA 作为一种有监督的学习算法,其原理是使得降维后同类样本之间的距离最小,而异类样本之间的距离最大.受流形学习算法的启发,很多基于局部二阶统计量的线性降维方法被提了出来,它们不是以类中心来衡量异类或同类样本的距离,而是用每个样本与其邻近的样本之间的距离来衡量.较为典型的有局部保持投影(locality preserving projection,简称 LPP)^[3-5]、无监督鉴别投影(unsupervised discriminant projection,简称 UDP)^[6]、邻域保持嵌入(neighborhood preserving embedding,简称 NPE)^[7].LPP 算法能够保持数据的局部结构信息,而 UDP 算法是 LPP 算法在数据为均匀分布时的一种特殊情况^[8].NPE 算法能够保持数据的局部流形结构.虽然上述各种线性降维方法的动机不同,但可以用线性图嵌入(linear extension of graph embedding,简称 LGE)的框架把这些算法统一起来^[9,10].

上述各种降维方法都是基于线性变换的特征抽取方法,最终抽取得到的是线性特征.当人脸特征存在于受光照、姿态、表情变化等影响的复杂非线性结构空间中时,这些线性算法就有其不可避免的缺陷,它们并没有办法更好地发掘这些非线性因素.

当前,核方法已经成为模式识别邻域的一个迅速发展的方向,上述线性降维算法都有其相对应的核算法(PKA 和 KPCA,LDA 和 KFD,LPP 和 KLPP,NPE 和 KNPE),并且这些核算法都可以纳入核化图嵌入(kernel extension of graph embedding,简称 KGE)框架^[9,10].

本文首先在从理论上分析了 KGE 框架内的各种算法,证明了其实质是 KPCA+LGE 框架内的各种线性降维算法.这样,我们可以从一个新的角度来理解核算法,并且使各种核算法更易于理解和实现.另外,在所提出的相关理论的基础上,又给出了一种综合利用零空间信息和非零空间信息的组合鉴别分析方法.由于本文的组合鉴别方法完整地利用了零空间和非零空间的鉴别信息,使我们的算法具有更高的识别率.最后,在 ORL, Yale, FERET 和 PIE 人脸数据库上验证了所提出的理论和方法的有效性.

1 相关的算法

1.1 核化图嵌入框架^[9,10]

设 $X=\{x_1, x_2, \dots, x_M\} \in R^{n \times M}$ 为训练样本集,其中, n 为样本的维数, M 为样本数. $G=\{X, W\}$ 为无向有权图,每个样本点 x_i 为图中的一个顶点, W 为相似度矩阵, W_{ij} 表示样本 i 和 j 的相似度.一个图的拉普拉斯矩阵 L 和对角矩阵 D 定义为

$$L = D - W, D_{ii} = \sum_j W_{ij}, \forall j \quad (1)$$

图嵌入的目标是找到原数据的低维表示 $y=\{y_1, y_2, \dots, y_M\}^T$, 保持原高维空间中顶点间的相似性. y 可以通过最小化公式(2)得到:

$$\sum_{i,j} (y_i - y_j)^2 W_{ij} = 2y^T Ly \quad (2)$$

为了防止求出没有意义的解,给公式(2)加入约束: $y^T Dy=1$, 则公式(2)变为

$$\begin{cases} \min, y^T Ly \\ \text{subject to } y^T Dy = 1 \end{cases} \quad (3)$$

假设从低维空间到高维空间是一个线性映射,即 $y=X^T a$, a 为单位投影矢量,则公式(3)可变为

$$a^* = \arg \min \frac{a^T X L X^T a}{a^T X D X^T a} = \arg \max \frac{a^T X D X^T a}{a^T X L X^T a} \quad (4)$$

其中, a^* 为最优投影矢量.如果需要多个投影矢量 $A=\{a_1, \dots, a_h\}$, 则公式(4)变为

$$A^* = \arg \max \text{tr}[(A^T X D X^T A)^{-1} (A^T X W X^T A)].$$

其中, A^* 为最优投影矩阵. A^* 可通过求解公式(5)得到:

$$XDX^T a = \lambda XLX^T a \tag{5}$$

上述方法就称为线性图嵌入(LGE)^[9,10], 它为各种线性降维方法提供了一个统一的框架. 通过选择不同的 L 和 D , 就可以得到常见的线性降维算法, 如 LDA, LPP, NPE 等.

核算法的基本思想是: 首先将原始训练样本通过一个满足 Mercer 条件的非线性映射 Φ 变换到某一高维特征空间(可能是无限维) H 中, 然后在高维特征空间 H 中进行线性降维, 则公式(4)可写为

$$\varphi^* = \arg \max_{\varphi} \frac{\varphi^T (\Phi(x_1), \dots, \Phi(x_M)) D (\Phi(x_1), \dots, \Phi(x_M))^T \varphi}{\varphi^T (\Phi(x_1), \dots, \Phi(x_M)) L (\Phi(x_1), \dots, \Phi(x_M))^T \varphi} \tag{6}$$

其中, φ 是位于特征空间 H 中的投影矢量. φ^* 为特征空间 H 中的最优投影矢量. 一般地, 应减去样本在特征空间 H 中的均值, 则公式(6)可写为

$$\begin{aligned} \varphi^* &= \arg \max_{\varphi} J^{\Phi}(\varphi) \\ &= \arg \max_{\varphi} \frac{\varphi^T (\Phi(x_1) - m_0^{\Phi}, \dots, \Phi(x_M) - m_0^{\Phi}) D (\Phi(x_1) - m_0^{\Phi}, \dots, \Phi(x_M) - m_0^{\Phi})^T \varphi}{\varphi^T (\Phi(x_1) - m_0^{\Phi}, \dots, \Phi(x_M) - m_0^{\Phi}) L (\Phi(x_1) - m_0^{\Phi}, \dots, \Phi(x_M) - m_0^{\Phi})^T \varphi} \end{aligned} \tag{7}$$

其中, $m_0^{\Phi} = \frac{1}{M} \sum_{i=1}^M \Phi(x_i)$. 上述方法就称为核化图嵌入(kernel extension of graph embedding, 简称 KGE), 它为各种核算法提供了一个统一的框架. 通过选择不同的 L 和 D , 就可以得到常见的核算法, 如 KFD, KLPP, KNPE 等. 在后面的叙述中, 我们用记号 $GE(L, D)$ 来表示公式(4), 用记号 $KGE(L, D)$ 来表示公式(7).

LDA and KFD:

设数据为 c 类并且第 t 类有 m_t 个样本, $m_1 + \dots + m_c = M$. 定义

$$W_{ij}^{LDA} = \begin{cases} 1/m_t, & \text{if } x_i \text{ and } x_j \text{ both belong to the } t\text{-th class} \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

$D^{LDA} = I$, 则 LDA 算法可表示为 $GE(I - W^{LDA}, D)$, KFD 算法可表示为 $KGE(I - W^{LDA}, D)$ ^[10].

LPP and KLPP:

设 $N_k(x_i)$ 表示 x_i 的 k 近邻集合, 定义

$$W_{ij}^{LPP} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}, & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

LPP 算法可以表示为 $GE(D^{LPP} - W^{LPP}, D^{LPP})$, KLPP 算法可以表示为 $KGE(D^{LPP} - W^{LPP}, D^{LPP})$ ^[10].

NPE and KNPE:

设 M 为 $m \times m$ 矩阵, 其定义为: 对于 M 的第 i 行, $M_{ij} = 0$, 如果 $x_j \notin N_k(x_i)$; 其他的 M_{ij} 可以通过最小化下式得到:

$$\min \|x_i - \sum_{j \in N_k(x_i)} M_{ij} x_j\|, \quad \sum_{j \in N_k(x_i)} M_{ij} = 1 \tag{10}$$

定义 $W^{NPE} = M + M^T - M^T M$, $D^{NPE} = I$, 则 NPE 算法可表示为 $GE(I - W^{NPE}, D)$, KNPE 可以表示为 $KGE(I - W^{NPE}, D)$ ^[10].

1.2 核主成分分析(kernel principal component analysis, 简称 KPCA)

KPCA^[11,12] 是 PCA 在核空间的非线性推广, 在特征空间 H 中的协方差 S_i^{Φ} 定义为

$$S_i^{\Phi} = \frac{1}{M} \sum_{j=1}^M (\Phi(x_j) - m_0^{\Phi})(\Phi(x_j) - m_0^{\Phi})^T \tag{11}$$

设 $Q = [\Phi(x_1), \dots, \Phi(x_M)]$, 则可以得到一个 Gram 矩阵 $\tilde{R} = Q^T Q$, 其元素为

$$\tilde{R}_{ij} = \Phi(x_i) \Phi(x_j)^T \tag{12}$$

对于 \tilde{R}_{ij} 可以通过下面的公式(13)进行中心化:

$$R = \tilde{R} - \mathbf{1}_M \tilde{R} - \tilde{R} \mathbf{1}_M + \mathbf{1}_M \tilde{R} \mathbf{1}_M, \quad \text{其中, } \mathbf{1}_M = (1/M)_{M \times M} \tag{13}$$

设正交向量 $\gamma_1, \gamma_2, \dots, \gamma_m$ 为 R 的 m 个最大的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 所对应的特征向量, 则 S_i^{Φ} 的 m 个最大的

特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 所对应的特征向量 $\beta_1, \beta_2, \dots, \beta_m$ 可由公式(14)求出:

$$\beta_j = \frac{1}{\sqrt{\lambda_j}} Q \gamma_j, j = 1, \dots, m \tag{14}$$

而一个样本 x 经过 KPCA 后提取的特征 $y = \{y_1, y_2, \dots, y_m\}^T$ 为

$$y = P^T (\Phi(x) - m_0^\phi), \text{ 其中, } P = (\beta_1, \beta_2, \dots, \beta_m) \tag{15}$$

如何在特征空间中减去均值 m_0^ϕ , 即如何对数据进行中心化, 可参考文献[13].

2 KGE 算法的本质:KPCA+LGE

设 $\Psi_t = \text{span}\{\beta_1, \beta_2, \dots, \beta_m\}$ 为特征空间 H 的子空间, 其中, $m = \text{rank}(S_t^\phi)$; Ψ_t 的正交补空间用 Ψ_t^\perp 表示. 易知, Ψ_t^\perp 为 S_t^ϕ 的零空间.

定理 1^[11]. $H = \Psi_t + \Psi_t^\perp$, 即对于任何 $\phi \in H$, ϕ 可唯一地表示为 $\phi = \phi + \zeta$, 其中, $\phi \in \Psi_t$ 和 $\zeta \in \Psi_t^\perp$.

定理 2. 设 Ψ_t^\perp 为 S_t^ϕ 的零空间, 则对于任意的 $\zeta \in \Psi_t^\perp$, 有 $\zeta^T (\Phi(x_i) - m_0^\phi) = 0, i = 1, 2, \dots, M$.

证明: 因为 Ψ_t^\perp 为 S_t^ϕ 的零空间, 则对于任意的 $\zeta \in \Psi_t^\perp$, 有 $S_t^\phi \zeta = 0$, 故 $\zeta^T S_t^\phi \zeta = 0$.

而 $\zeta^T S_t^\phi \zeta = \frac{1}{M} \sum_{i=1}^M (\zeta^T (\Phi(x_i) - m_0^\phi))^2$, 故有 $\zeta^T (\Phi(x_i) - m_0^\phi) = 0, i = 1, 2, \dots, M$. □

映射 $L: H \rightarrow \Psi_t$ 定义如下: $\phi = \phi + \zeta \rightarrow \phi$, 易知 L 是线性映射, 称其为压缩映射.

定理 3. 在压缩映射 $\phi = \phi + \zeta \rightarrow \phi$ 下, 有 $J^\phi(\phi) = J^\phi(\phi)$.

证明:

$$\begin{aligned} J^\phi(\phi) &= \frac{\phi^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) D(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T \phi}{\phi^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) L(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T \phi} \\ &= \frac{(\phi + \zeta)^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) D(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T (\phi + \zeta)}{(\phi + \zeta)^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) L(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T (\phi + \zeta)} \end{aligned} \tag{16}$$

由定理 2 可知, $(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T \zeta = 0, \zeta^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) = 0$, 故公式(16)可写为

$$J^\phi(\phi) = \frac{\phi^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) D(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T \phi}{\phi^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) L(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T \phi} \tag{17}$$

即 $J^\phi(\phi) = J^\phi(\phi)$. □

由定理 3 可知, 对于 KGE 的准则函数公式(7)而言, 如果所有的鉴别矢量都可以从 Ψ_t 中选取, 那么这样做并不会损失任何鉴别信息.

由泛函分析的知识可知^[14], Ψ_t 同构于 m 维欧氏空间 \mathbb{R}^m , 其同构映射为

$$\phi = P \eta \tag{18}$$

其中, $P = (\beta_1, \beta_2, \dots, \beta_m), \eta \in \mathbb{R}^m$. 在此同构映射下, $J^\phi(\phi)$ 变为

$$J^\phi(\phi) = \frac{\eta^T (P^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) D(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T P) \eta}{\eta^T (P^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi) L(\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)^T P) \eta} \tag{19}$$

定义 $\tilde{X} = P^T (\Phi(x_1) - m_0^\phi, \dots, \Phi(x_M) - m_0^\phi)$, 则公式(19)变为

$$\tilde{J}_m(\eta) = \frac{\eta^T \tilde{X} D \tilde{X}^T \eta}{\eta^T \tilde{X} L \tilde{X}^T \eta} \tag{20}$$

易知有如下定理:

定理 4. 在同构映射 $\phi = P \eta$ 下, $\phi^* = P \eta^*$ 是 $J^\phi(\phi)$ 的极值点当且仅当 η^* 是 $\tilde{J}_m(\eta)$ 的极值点.

给定一个样本点 x 及其在特征空间 H 中的映射 $\Phi(x)$, 则可以通过公式(21)得到此样本的鉴别特征 z :

$$z = W^T (\Phi(x) - m_0^\phi) \tag{21}$$

其中, $W^T = (\varphi_1, \varphi_2, \dots, \varphi_d)^T = (P \eta_1, P \eta_2, \dots, P \eta_d)^T = (\eta_1, \eta_2, \dots, \eta_d)^T P^T$, 则公式(21)可分解为两部分:

$$y = P^T(\Phi(x) - m_0^\phi), \text{其中}, P = (\beta_1, \beta_2, \dots, \beta_m) \quad (22)$$

和

$$z = G^T y, \text{其中}, G = (\eta_1, \eta_2, \dots, \eta_d) \quad (23)$$

比较公式(22)和公式(15)可知,公式(22)就是进行了 KPCA 变换.比较公式(20)和公式(4)可知,公式(20)是在 KPCA 的变换空间 \mathbb{R}^m 内进行了 LGE 算法,而公式(23)是在 KPCA 的变换空间 \mathbb{R}^m 内利用 LGE 算法得到的鉴别矢量进行了特征抽取.至此,我们揭示了核算法的实质,即:KGE 框架内的任何一种核算法其实质是 KPCA+LGE 框架内的线性降维算法.

3 组合鉴别分析

3.1 有效鉴别矢量的分类

根据第 2 节的分析,只需在 KPCA 变换空间 \mathbb{R}^m 内讨论问题,也即如何通过公式(20)来获得最佳鉴别矢量.故接下来讨论在 KPCA 变换空间 \mathbb{R}^m 内鉴别矢量的构成.注意到 KFD 算法是有监督算法,而 KLPP,KNPE 等算法既可以以无监督的方式实现,也可以以有监督的方式实现,但一般而言,有监督算法比无监督算法的分类性能要好.因此,本文在后面讨论如何求得在有监督情形下的最佳鉴别矢量.

一般地, $m = \text{rank}(S_i^\phi) = M - 1$,因此 $\tilde{X}L\tilde{X}^T$ 是一个 $(M-1) \times (M-1)$ 的矩阵,易知 $\text{rank}(\tilde{X}L\tilde{X}^T) \leq \text{rank}(L)$.对于矩阵 L ,有如下定理:

定理 5. 假设拉普拉斯矩阵 L 以有监督的方式创建, $\text{rank}(L) \leq M - c$, c 表示类别数.

证明:不失一般性,假设 $\{x_1, \dots, x_{n_1}\}$ 为第 1 类, $\{x_{n_1+1}, \dots, x_{n_1+n_2}\}$ 为第 2 类,依次类推.考虑有监督的情形,则 L 可

写为 $L = \begin{bmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_c \end{bmatrix}$, L_i 是对称方阵,且可表示为

$$L_i = \begin{bmatrix} D_{N+1,N+1} - w_{N+1,N+1} & -w_{N+1,N+2} & \dots & -w_{N+1,N+n_i} \\ -w_{N+2,N+1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -w_{N+n_i,N+1} & \dots & \dots & D_{N+n_i,N+n_i} - w_{N+n_i,N+n_i} \end{bmatrix},$$

其中, $N = n_1 + n_2 + \dots + n_{i-1}$,把第 2, ..., n_i 行加到第 1 行,并且由 D 的定义(见公式(1))可知,

$$L_i = \begin{bmatrix} 0 & 0 & \dots & 0 \\ -w_{N+2,N+1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ -w_{N+n_i,N+1} & \dots & \dots & D_{N+n_i,N+n_i} - w_{N+n_i,N+n_i} \end{bmatrix}.$$

故 $\text{rank}(L_i) \leq n_i - 1, \text{rank}(L) \leq M - c$. □

由定理 5 可知, $\text{rank}(\tilde{X}L\tilde{X}^T) \leq \text{rank}(L) \leq M - c$, 而类别数 c 一般大于 1,故在 KPCA 变换的空间 \mathbb{R}^m 内,矩阵 $\tilde{X}L\tilde{X}^T$ 往往是奇异的.

设 $\alpha_1, \dots, \alpha_m$ 是矩阵 $\tilde{X}L\tilde{X}^T$ 的特征向量,其秩 $q = \text{rank}(\tilde{X}L\tilde{X}^T)$, 定义子空间 $\Theta_{\tilde{x}} = \text{span}\{\alpha_{q+1}, \dots, \alpha_m\}$, 其正交补空间为 $\Theta_{\tilde{x}}^\perp = \text{span}\{\alpha_1, \dots, \alpha_q\}$.

定理 6. 对于任意非零 $\eta \in \mathbb{R}^m$ 有 $\eta^T (\tilde{X}D\tilde{X}^T)\eta > 0$.

证明:设在空间 \mathbb{R}^m 内的总体散布矩阵为 \tilde{S}_i , 则 $\tilde{S}_i = P^T S_i^\phi P$.

对于任意的非零 $\eta \in \mathbb{R}^m$, 有 $\eta^T \tilde{S}_i \eta > 0, \eta^T \tilde{S}_i \eta = \frac{1}{M} \sum_{i=1}^M ((\Phi(x_i) - m_0^\phi)^T P \eta)^2 > 0$, 故至少存在一个 $(\Phi(x_i) - m_0^\phi)$, 使

得 $(\Phi(x_i) - m_0^\phi)^T P \eta > 0$. 而 $\eta^T (\tilde{X} D \tilde{X}^T) \eta = \sum_{i=1}^M D_{ii} ((\Phi(x_i) - m_0^\phi)^T P \eta)^2$, 由于 $\forall D_{ii} > 0$, 故 $\eta^T (\tilde{X} D \tilde{X}^T) \eta > 0$. \square

由定理 5、定理 6 和 KPCA 后的鉴别准则公式(20)可知,我们可以把最优鉴别矢量分为两类:第 1 类满足 $\eta^T (\tilde{X} L \tilde{X}^T) \eta = 0$ 和 $\eta^T (\tilde{X} D \tilde{X}^T) \eta > 0$, 这一类最优鉴别矢量主要利用矩阵 $\tilde{X} L \tilde{X}^T$ 的零空间鉴别信息;第 2 类满足 $\eta^T (\tilde{X} L \tilde{X}^T) \eta > 0$ 和 $\eta^T (\tilde{X} D \tilde{X}^T) \eta > 0$, 这一类鉴别矢量主要利用矩阵 $\tilde{X} L \tilde{X}^T$ 的非零空间的信息.

3.2 鉴别准则的优化

对于第 1 类鉴别矢量,若按公式(20)的准则函数来衡量,分母为 0,故需重新定义准则函数.设 $P_1=(\alpha_{q+1}, \dots, \alpha_m)$,则有同构映射:

$$\eta = P_1 \xi \quad (24)$$

对于第 1 类鉴别矢量,由于可以把 $X D X^T$ 看作是加权的 S_i ,故将公式(20)重新定义为

$$\xi^* = \arg \max_{\xi} \frac{\xi^T (\tilde{X} D \tilde{X}^T) \xi}{\xi^T \xi} \quad (25)$$

其中, $\tilde{X} = P_1^T \tilde{X}$.

对于第 2 类鉴别矢量,设 $P_2=(\alpha_1, \dots, \alpha_q)$,有同构映射:

$$\eta = P_2 \xi \quad (26)$$

则将公式(20)重新定义为

$$\xi^* = \arg \max_{\xi} \frac{\xi^T (\tilde{X} D \tilde{X}^T) \xi}{\xi^T (\tilde{X} L \tilde{X}^T) \xi} \quad (27)$$

其中, $\tilde{X} = P_2^T \tilde{X}$.

3.3 本文算法的流程

综上所述,综合利用零空间信息和非零空间信息的组合鉴别分析方法的流程如下:

- ① 通过公式(1)、公式(8)~公式(10)来构造不同算法所需的 L 和 D ;
- ② 使用 KPCA 算法把原始样本数据转换到空间 \mathbb{R}^m 内,设 \mathbf{y} 为原始数据经 KPCA 后得到的特征矢量;
- ③ 根据公式(25)计算第 1 类最佳鉴别矢量(零空间内的鉴别信息),设 $U^1=(\xi_1, \xi_2, \dots, \xi_d)$ 为矩阵 $\tilde{X} D \tilde{X}^T$ 最大的 d 个最大特征值所对应的特征向量,则用式 $(P_1 U^1)^T \mathbf{y}$ 来进行特征的抽取;
- ④ 根据公式(27)计算第 2 类的最佳鉴别矢量(非零空间内的鉴别信息),设 $U^2=(\xi_1, \xi_2, \dots, \xi_p)$ 为式 $(\tilde{X} D \tilde{X}^T) \xi = \lambda (\tilde{X} L \tilde{X}^T) \xi$ 最大的 p 个特征值所对应的特征向量,则可用式 $(P_2 U^2)^T \mathbf{y}$ 来进行特征的提取;
- ⑤ 对得到的第 1 类和第 2 类鉴别特征进行融合.这里,我们直接进行特征层的融合,即将得到的两类特征串联起来得到新的特征 $\begin{bmatrix} (P_1 U^1)^T \mathbf{y} \\ (P_2 U^2)^T \mathbf{y} \end{bmatrix}$.

因此,对于 KGE 框架中的每种核算法,都可以先对数据作 KPCA 变换,然后再利用 LGE 算法来求解最佳鉴别矢量;而对于每一种 LGE 算法,又有其相对应的综合利用零空间(第 1 类鉴别矢量)和非零空间信息(第 2 类鉴别矢量)的组合算法.因此,对于 KFD 算法有其对应的 KPCA+CLDA(combined LDA,简称 CLDA)算法;KNPE 算法有其相对应的 KPCA+CNPE(combined NPE,简称 CNPE)算法;KLPP 算法有其对应的 KPCA+CLPP(combined LPP,简称 CLPP)算法.在文献[11]中, Yang 等人证明了核 Fisher 鉴别分析(KFD)的实质是 KPCA+LDA,因此文献[11]中的算法只是本文算法框架内的一个特例.由于在文献[11]中已经比较了 KPCA+CLDA 算法和常规的 KFD 算法的性能,因此在后面的实验中,我们不再比较这两种算法的性能.

4 仿真实验及分析

本文的实验中,采用高斯核函数 $K_G(x, y) = \varphi(x) \cdot \varphi(y) = \exp(-\|x - y\|^2 / \delta^2)$, 对于 δ 和 LPP 算法中的 σ , 其取值为

$2^{(b-10)/25} \delta_0, b=0,1,\dots,20, \delta_0$ 为训练样本的标准偏差.在实验中使用的分类器是最近邻分类器.

4.1 实验中所选用的数据库

ORL 标准人脸库由 40 人、每人 10 幅图像组成,其中有些图像是拍摄于不同时期的,人的脸部表情和脸部细节有着不同程度的变化,比如笑或不笑、眼睛或睁或闭、戴或不戴眼镜;人脸姿态也有相当程度的变化,深度旋转和平面旋转可达 20° ;人脸的尺度也有多达 10% 的变化.实验中,图像被处理成 32×32 的形式.

Yale 人脸库包括 15 个人的 165 幅灰度人脸图像,每个人由 11 幅照片构成.这些照片在不同的表情和光照等条件下拍摄.实验中,图像被处理成 32×32 的形式.

FERET 人脸库由取自 1 199 个人的 14 000 多幅不同姿态、表情、光照和时期的人脸图像组成.我们选择其中的一个子库进行实验,该子库包含 200 个人的 1 400 幅图像(每人 7 幅).这些图像的文件名均包含了标识字符串:“ba”,“bj”,“bk”,“be”,“bf”,“bd”,“bg”,这些字符串标识了每幅图像的变化情况,如表情、光照等.该子库包含了人脸表情的变化、光照的变化、姿态的变化($\pm 15^\circ$ 和 $\pm 25^\circ$).

PIE 人脸数据库由 68 人的 41 368 幅不同姿态、不同光照、不同表情的人脸图像组成.我们从中选择了包含姿态、光照、表情变化的 5 个子集(C05,C07,C09,C27,C29)共 68 人的正面人脸图像,每人 170 幅图像.实验中,图像处理成 32×32 的形式.图 1 显示了预处理后各人脸图像库的图像示例.

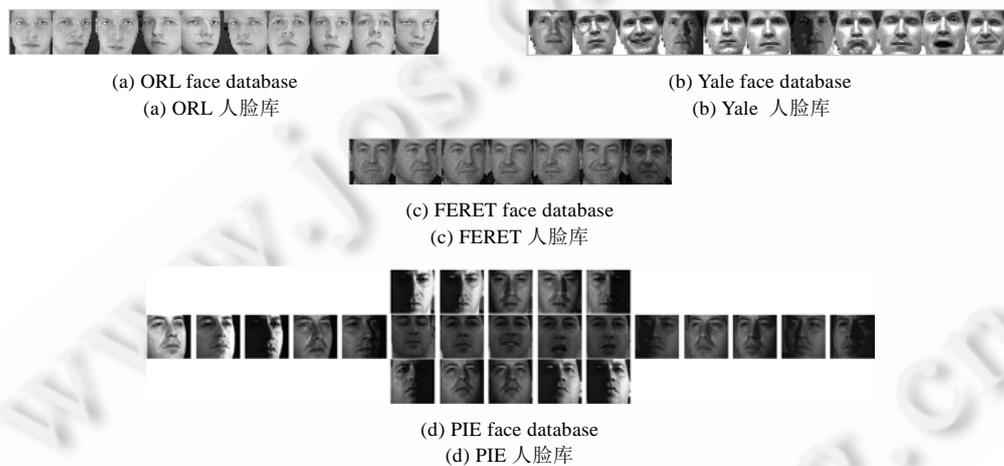


Fig.1 Preprocessed face images

图 1 预处理后的图像示例

4.2 实验结果与分析

首先考察识别精度与特征维数之间的关系,在 ORL, Yale, FERET 和 PIE 库上进行实验.对于 ORL 和 Yale 人脸库选择 4 幅图像作为训练样本,剩余的其他图像作为测试样本;对于 FERET 人脸库选择 3 幅图像作为训练样本,剩余的其他图像作为测试样本;对于 PIE 人脸库选择 5 幅图像作为训练样本,剩余的其他图像作为测试样本.实验结果如图 2、图 3 所示.

从图中可以发现,总体上看,随着特征维数的增加,本文算法的识别精度逐渐提高,并且本文算法识别精度随特征维数提高的速度比原始算法要快,且识别率基本上始终高于原始算法.

接下来考察在不同训练样本情况下各种算法的识别率.随机地在 ORL, Yale 库中选择 $i(i=2,3,4)$ 、在 FERET 库中选择 $i(i=2,3)$ 、在 PIE 库中选择 $i(i=5,10)$ 幅图像作为训练样本幅图像作为训练样本,剩余的图像作为测试样本.每组实验均重复了 20 次,实验结果见表 1~表 4,表中给出了 20 次实验的平均识别率和标准方差.

由实验结果可以看出,本文的方法在 ORL, Yale, FERET 和 PIE 库上都取得了非常好的识别效果.由于本文算法在 KPCA 进行降维时并没有损失鉴别信息,同时又综合利用了零空间和非零空间的鉴别信息,因此使得识别

率有了较大的提高.

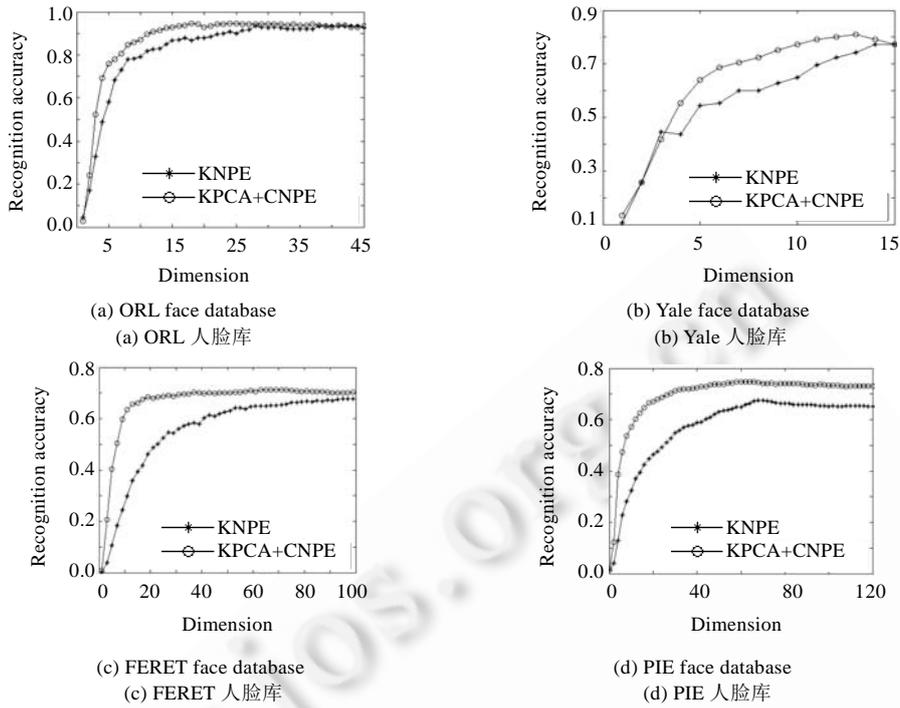


Fig.2 Variation of recognition accuracies of KNPE and KPCA+CNPE along with dimensionality

图 2 KNPE 和 KPCA+CNPE 算法识别精度随特征维度的变化情况

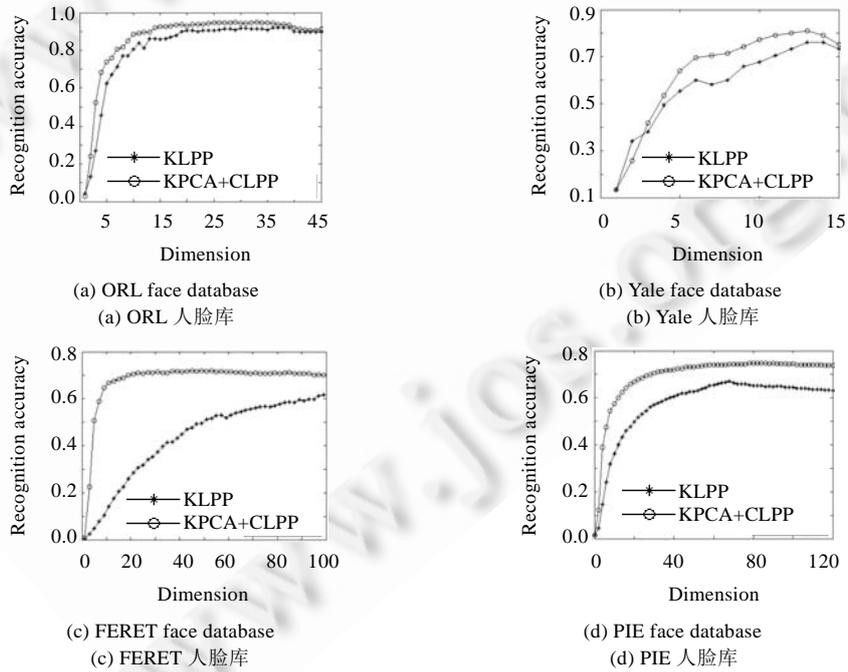


Fig.3 Variation of recognition accuracies of KLPP and KPCA+CLPP along with dimensionality

图 3 KLPP 和 KPCA+CLPP 算法识别精度随特征维度的变化情况

Table 1 Recognition rate comparison on ORL face database with different methods

表 1 在 ORL 人脸库中不同方法的识别率对比

Size	LDA	LPP	KLPP	KPCA+CLPP	NPE	KNPE	KPCA+CNPE
2	75.4±3.1	75.9±3.1	83.0±2.5	87.6±2.4	76.2±3.1	82.6±2.5	87.7±2.5
3	85.1±1.9	86.0±2.1	89.6±1.4	94.1±1.9	84.6±1.9	89.9±1.7	94.0±1.6
4	91.3±1.9	92.7±1.7	94.1±1.6	97.0±1.0	90.9±2.4	94.3±1.4	97.1±1.1

Table 2 Recognition rate comparison on Yale face database with different methods

表 2 在 Yale 人脸库中不同方法的识别率对比

Size	LDA	LPP	KLPP	KPCA+CLPP	NPE	KNPE	KPCA+CNPE
2	52.1±5.6	53.2±5.6	57.7±4.1	59.6±4.4	53.3±4.3	56.8±3.9	59.6±4.4
3	65.4±4.6	65.6±4.7	67.9±4.0	70.2±5.0	65.3±5.0	67.5±3.9	70.7±4.8
4	72.1±5.4	74.3±5.4	74.4±5.8	78.3±4.7	73.4±4.1	74.7±4.8	78.7±4.4

Table 3 Recognition rate comparison on FERET face database with different methods

表 3 在 FERET 人脸库中不同方法的识别率对比

Size	LDA	LPP	KLPP	KPCA+CLPP	NPE	KNPE	KPCA+CNPE
2	42.9±7.4	46.3±6.8	51.9±6.8	61.1±6.5	41.8±7.4	49.4±7.1	61.3±7.0
3	63.0±7.9	63.3±7.8	68.5±7.9	76.1±7.8	63.0±7.5	66.3±5.9	75.5±7.7

Table 4 Recognition rate comparison on PIE face database with different methods

表 4 在 PIE 人脸库中不同方法的识别率对比

Size	LDA	LPP	KLPP	KPCA+CLPP	NPE	KNPE	KPCA+CNPE
5	67.1±1.4	67.4±1.4	68.1±1.2	76.4±1.2	67.4±1.4	68.0±1.0	74.8±1.0
10	81.5±1.0	82.9±1.1	82.6±0.9	87.9±0.7	81.2±1.1	82.1±0.9	87.1±0.8

5 结束语

本文分析了核化图嵌入框架内的各种核算法,同时将压缩映射和同构映射引入核化图嵌入框架,从理论上证明了 KGE 框架内的任何一种核算法其本质是 KPCA+LGE 框架内的线性降维算法;并且基于所给出理论框架,提出了一种综合利用零空间和非零空间鉴别信息的组合方法,任何一种可以用核化图嵌入框架描述的核算法都可以有相应的组合方法.在 ORL, Yale, FERET 和 PIE 人脸数据库的实验证实了本文所提出的组合方法的有效性.

References:

- [1] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed., New York: John Wiley & Sons, 2000.
- [2] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997,19(7):711-720. [doi: 10.1109/34.598228]
- [3] He XF, Yan SC, Hu YX, Niyogi P, Zhang HJ. Face recognition using Laplacianfaces. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005,27(3):328-340. [doi: 10.1109/TPAMI.2005.55]
- [4] Shen ZH, Pan YH, Wang ST. A supervised locality preserving projection algorithm for dimensionality reduction. Pattern Recognition and Artificial Intelligence, 2008,21(2):233-239 (in Chinese with English abstract).
- [5] Wang XM, Wang ST. Generalized supervised locality preserving projection. Journal of Electronics & Information Technology, 2009,31(8):1840-1845 (in Chinese with English abstract).
- [6] Yang J, Zhang D, Yang JY, Niu B. Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(4):650-664. [doi: 10.1109/TPAMI.2007.1008]
- [7] He XF, Cai D, Yan SC, Zhang HJ. Neighborhood preserving embedding. In: Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV). Beijing, 2005. 1208-1213. <http://www.zjucadcg.cn/dengcai/Publication/Conference/he-embedding2.pdf> [doi: 10.1109/ICCV.2005.167]

- [8] Deng WH, Hu JN, Guo J, Zhang HG, Zhang C. Comments on “Globally maximizing, locally minimizing: Unsupervised discriminant projection with application to face and palm biometrics”. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008,30(8):1503–1504. [doi: 10.1109/TPAMI.2007.70783]
- [9] Cai D, He XF, Hu YX, Han JW, Huang T. Learning a spatially smooth subspace for face recognition. In: Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR). Minneapolis, 2007. 1–7. <http://www.zjucadcg.cn/dengcai/Publication/Conference/CVPR07.pdf> [doi: 10.1109/CVPR.2007.383054]
- [10] Yan SC, Xu D, Zhang BY, Zhang HJ, Yang Q, Lin S. Graph embedding and extensions: A general framework for dimensionality reduction. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(1):40–51. [doi: 10.1109/TPAMI.2007.250598]
- [11] Yang J, Frangi AF, Yang JY, Zhang D, Jin Z. KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005,27(2):230–244. [doi: 10.1109/TPAMI.2005.33]
- [12] Shi WY, Guo YF, Xue XY. Efficient kernel principal component analysis algorithm for large-scale data set. Journal of Software, 2009,20(8):2153–2159 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3391.htm> [doi: 10.3724/SP.J.1001.2009.03391]
- [13] Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 1998,10(5):1299–1319. [doi:10.1162/089976698300017467]
- [14] Kreyszig E. Introductory Functional Analysis with Applications. New York: John Wiley & Sons, 1978.

附中文参考文献:

- [4] 申中华,潘永惠,王士同.有监督的局部保留投影降维算法.模式识别与人工智能,2008,21(2):233–239.
- [5] 王晓明,王士同.广义的监督局部保留投影算法.电子与信息学报,2009,31(8):1840–1845.
- [12] 史卫亚,郭跃飞,薛向阳.一种解决大规模数据集问题的核主成分分析算法.软件学报,2009,20(8):2153–2159. <http://www.jos.org.cn/1000-9825/3391.htm> [doi: 10.3724/SP.J.1001.2009.03391]



卢桂馥(1976—),男,浙江东阳人,博士生,副教授,主要研究领域为人工智能,模式识别.



金忠(1961—),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为模式识别,计算机视觉.



林忠(1983—),男,博士生,主要研究领域为模式识别,数字图像处理.