

基于核鉴别共同矢量的小样本脸像鉴别方法

贺云辉^{①②} 赵力^① 邹采荣^①

^①(东南大学无线电工程系 南京 210096)

^②(南京信息工程大学通信工程系 南京 210044)

摘要 人脸识别中通常存在小样本问题,使得基于 Fisher 线性鉴别分析的特征抽取方法存在病态奇异问题。近年来针对此问题提出了不同的解决方法,其中基于共同鉴别矢量(DCV)的方法成功克服了已有各种方法存在的缺点,有较好的数值稳定性和较低的计算复杂度。该文将 DCV 方法推广到非线性领域,将两次 Gram-Schmidt 正交化过程,转化为只需计算两个核矩阵和进行一次 Cholesky 分解完成,且得到的非线性 Fisher 鉴别矢量有标准正交的性质。实验验证了所得 KDCV 方法的识别性能优于 DCV 方法。

关键词 人脸识别, 鉴别共同矢量, 核方法, 小样本问题, Fisher 线性鉴别分析

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2006)12-2296-05

Face Recognition Based on Kernel Discriminative Common Vectors

He Yun-hui^{①②} Zhao Li^① Zou Cai-rong^①

^①(Department of Radio Engineering, Southeast University, Nanjing 210096, China)

^②(Dept. of Communications Eng., Nanjing Univ. of Information Science & Tech., Nanjing 210044, China)

Abstract Face recognition tasks always encounter Small Sample Size (SSS) problem, which leads to the ill-posed problem in Fisher Linear Discriminant Analysis (FLDA). The Discriminative Common Vector (DCV) successfully overcomes this problem for FLDA. In this paper, the DCV is extended to nonlinear case, by performing the Gram-Schmidt orthogonalization twice in feature space, which involving computing two kernel matrices and performing a Cholesky decomposition of a kernel matrix. The experimental results demonstrate that the proposed KDCV achieve better performance than the DCV method.

Key words Face Recognition, Discriminative Common Vectors, Kernel method, Small Sample Size (SSS) problem, Fisher Linear Discriminant Analysis (FLDA)

1 引言

人脸识别中通常存在小样本问题,导致求解Fisher线性鉴别分析(FLDA)的广义特征方程存在病态奇异问题。对此问题近年来提出了不同的解决方法,如PCA+LDA方法^[1]首先利用主分量分析将人脸图像样本降维,使降维后的类内散度阵非奇异,再进行通常的FLDA。DLDA方法^[2]则首先将人脸样本投影到类间散度阵的秩空间,然后求解最佳鉴别矢量使类内散度最小。文献[3]证明了最佳鉴别矢量应位于类内散度阵的零空间中。PCA+零空间方法^[4]首先将所有样本映射到总散度阵的秩空间中,然后再进行零空间方法。然而这些方法均存在各自的问题,例如PCA+LDA和DLDA方法会丢失有较强鉴别力的鉴别矢量;零空间方法计算零空间投影阵的计算复杂度较大;当总样本数较多时,PCA+零空间方法首先用PCA降维也存在计算量较大的问题。

文献[5]新近提出的鉴别共同矢量(DCV)方法成功克服了上述各种方法存在的缺点,计算量小且数值稳定。其基本思想仍在类内散度阵的零空间中求解鉴别矢量,但整个计算过程分两步进行:首先在每类人脸中任选一个样本投影到类内散度阵的零空间中得到该类别的共同矢量,然后利用共同

矢量最优化类间散度得到最佳 Fisher 鉴别矢量,并提出了两种计算方法。DCV 算法 1 需要两次求解特征方程,计算量大且数值不稳定,难以得到较好的识别率,为此 DCV 算法 2 利用两次 Gram-Schmidt 正交化过程代替两次求解特征方程,数值稳定性较好且计算复杂度较小。

为抽取鉴别力更强的非线性 Fisher 鉴别特征,本文将 DCV 算法推广为基于核的 DCV(KDCV)方法。然而由于 DCV 算法是通过 Gram-Schmidt 正交化过程迭代计算最佳线性鉴别矢量,而核方法中的非线性映射未知,因此无法在特征空间直接进行 Gram-Schmidt 正交化过程。为此,本文首先利用样本数据矩阵在特征空间的 QR 分解,建立特征空间中样本和其正交化后基矢量的关系。然后将 DCV 算法中需要进行两次 Gram-Schmidt 正交化的过程,转化为只需计算两个核矩阵和对核矩阵进行一次 Cholesky 分解来完成。得到的 KDCV 方法无需求解特征方程,因此数值稳定性较好且计算复杂度也较低。此外所得非线性鉴别矢量还具有标准正交的性质,使得非线性鉴别矢量间的冗余度最小,使所有的非线性鉴别特征均得到了有效利用。

2 鉴别共同矢量方法简要回顾

设有 C 个类别,每类 N 个样本,总样本数为 $M = CN$,

$\mathbf{x}_m^i \in \mathbf{R}^d$ 为第 i 类的第 m 个样本。类间散度、类内散度和总散度阵分别定义为

$$\mathbf{S}_b = \sum_{i=1}^C N(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \quad (1)$$

$$\mathbf{S}_w = \sum_{i=1}^C \sum_{m=1}^N (\mathbf{x}_m^i - \boldsymbol{\mu}_i)(\mathbf{x}_m^i - \boldsymbol{\mu}_i)^T \quad (2)$$

$$\mathbf{S}_t = \sum_{i=1}^C \sum_{m=1}^N (\mathbf{x}_m^i - \boldsymbol{\mu})(\mathbf{x}_m^i - \boldsymbol{\mu})^T = \mathbf{S}_b + \mathbf{S}_w \quad (3)$$

最佳鉴别矢量应位于 \mathbf{S}_w 的零空间中, 因此 DCV 方法在 \mathbf{S}_w 的零空间中求解最佳鉴别矢量, 由其构成最佳变换阵 \mathbf{W}_{opt} 最大化类间散度, 其最优化准则为

$$J(\mathbf{W}_{\text{opt}}) = \arg \max_{\|\mathbf{W}^T \mathbf{S}_b \mathbf{W}\|} \left| \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\|\mathbf{W}^T \mathbf{S}_b \mathbf{W}\|} \right| = \arg \max_{\|\mathbf{W}^T \mathbf{S}_t \mathbf{W}\|} \left| \frac{\mathbf{W}^T \mathbf{S}_t \mathbf{W}}{\|\mathbf{W}^T \mathbf{S}_t \mathbf{W}\|} \right| \quad (4)$$

DCV 方法首先将样本映射到 \mathbf{S}_w 的零空间, 然后在 \mathbf{S}_w 的零空间最大化类间散度得到 Fisher 最佳鉴别矢量。可证明第 i 类的任意一个样本矢量 \mathbf{x}_m^i 映射到 \mathbf{S}_w 的零空间后均为相同矢量 $\mathbf{x}_{\text{com}}^i$ (共同矢量), 其独立于任意的样本, 这样第二阶段在 \mathbf{S}_w 的零空间求解鉴别矢量的计算复杂度仅和类别数有关。DCV 算法 1 在第一阶段首先计算 \mathbf{S}_w 的非零特征值对应的特征矢量 $\boldsymbol{\beta}_k$, 由其构成投影阵将样本投影到 \mathbf{S}_w 的零空间中得到每个类别的共同矢量

$$\mathbf{x}_{\text{com}}^i = \mathbf{x}_m^i - \mathbf{Q}\mathbf{Q}^T \mathbf{x}_m^i \quad (5)$$

其中 $\mathbf{Q} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_{M-C}]$ 。然后在第二阶段求解由共同矢量构成的散度阵 \mathbf{S}_{com} 的非零特征值对应的特征矢量 \mathbf{w}_k , 由其构成最佳变换阵 $\mathbf{W}_{\text{opt}}^T = [\mathbf{w}_1, \dots, \mathbf{w}_{C-1}]$ 。在每类中任选一个样本 \mathbf{x}_m^i , 利用最佳变化阵将其变换到低维空间得到第 i 类的特征矢量

$$\boldsymbol{\Omega}_i = \mathbf{W}_{\text{opt}}^T \mathbf{x}_m^i, \quad i=1, 2, \dots, C, \quad m=1, 2, \dots, N \quad (6)$$

并对待识别样本 \mathbf{x}_{test} 计算特征矢量

$$\boldsymbol{\Omega}_{\text{test}} = \mathbf{W}_{\text{opt}}^T \mathbf{x}_{\text{test}} \quad (7)$$

识别阶段利用最近距离分类准则, 将待识别样本 \mathbf{x}_{test} 归入类别 c 中

$$c = \min_i \|\boldsymbol{\Omega}_{\text{test}} - \boldsymbol{\Omega}_i\| \quad i=1, 2, \dots, C \quad (8)$$

DCV 算法 1 在两个阶段均需要求解特征方程, 计算量大且数值很不稳定, 难以获得较好的识别性能。为此在文献[6]所得结论的基础上, DCV 算法 2 将两次求解特征方程的问题转化为只需进行两次 Gram-Schmidt 正交化过程, 较大增强了数值稳定性。DCV 算法 2 的具体推导过程见文献[5], 现将其计算过程总结如下:

(1) 由样本构造子空间 $B = B_1 + \dots + B_C = \{\mathbf{b}_1^1, \dots, \mathbf{b}_{N-1}^1, \mathbf{b}_1^2, \dots, \mathbf{b}_{N-1}^C\}$, 其中 $\mathbf{b}_k^i = \mathbf{x}_{k+1}^i - \mathbf{x}_1^i$, $k=1, \dots, N-1$ 。

(2) 对 \mathbf{b}_k^i 进行 Gram-Schmidt 正交化得到标准正交的矢量 $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{M-C}$ 。

(3) 由式(5)在第 i 类中任选一个样本 \mathbf{x}_m^i 投影到 \mathbf{S}_w 的零空间得到第 i 类的共同矢量 $\mathbf{x}_{\text{com}}^i$ 。

(4) 计算矢量 $\mathbf{b}_{\text{com}}^k = \mathbf{x}_{\text{com}}^{k+1} - \mathbf{x}_{\text{com}}^1$, $k=1, 2, \dots, C-1$, 并对其进行 Gram-Schmidt 正交化得到矢量 \mathbf{w}_k , 由其构成最佳变换阵 $\mathbf{W}_{\text{opt}}^T = [\mathbf{w}_1, \dots, \mathbf{w}_{C-1}]$ 。

(5) 由式(6)和式(7)得到第 i 类的特征矢量 $\boldsymbol{\Omega}_i$ 和待识别样本 \mathbf{x}_{test} 的特征矢量 $\boldsymbol{\Omega}_{\text{test}}$ 。

(6) 由式(8)计算 $\boldsymbol{\Omega}_{\text{test}}$ 和 $\boldsymbol{\Omega}_i$ ($i=1, 2, \dots, C$) 的最小距离对待识别样本分类。

3 特征空间中的 Gram-Schmidt 正交化过程

核方法利用一非线性映射 $\Phi(\mathbf{x})$ 将原始图像样本 \mathbf{x} 映射到特征空间 F 中, 在 F 中进行通常的线性分析得到在输入空间的非线性算法。 F 中的内积运算可用核函数 $k(\mathbf{x}, \mathbf{y})$ 完成, $k(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$ 。本文使用常用的多项式核函数 $k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^d$ 。Gram-Schmidt 正交化过程计算每个正交矢量均需要预先知道前面的各个正交矢量, 而核方法中的非线性映射函数未知, 因此无法在特征空间直接进行 Gram-Schmidt 正交化。这是推导 KDCV 方法面临的主要困难, 为此本节首先得到特征空间中 Gram-Schmidt 正交化过程的表达形式。

设映射到特征空间后的样本数据矩阵为 $\mathbf{X}_\phi = [\Phi(\mathbf{x}_1), \dots, \Phi(\mathbf{x}_n)]$, 第一个基矢量标准化后为

$$\mathbf{q}_1 = \Phi(\mathbf{x}_1) / \|\Phi(\mathbf{x}_1)\| \quad (9)$$

根据 Gram-Schmidt 正交化过程^[7,8], 为得到第 i 个标准正交的基矢量 \mathbf{q}_i , 第 i 个矢量 $\Phi(\mathbf{x}_i)$ 应减去其在前 $i-1$ 个基矢量 $\mathbf{q}_1, \dots, \mathbf{q}_{i-1}$ 上的投影

$$\tilde{\mathbf{q}}_i = \Phi(\mathbf{x}_i) - \sum_{j=1}^{i-1} \langle \Phi(\mathbf{x}_i), \mathbf{q}_j \rangle \mathbf{q}_j = (\mathbf{I} - \mathbf{Q}_{i-1} \mathbf{Q}_{i-1}^T) \Phi(\mathbf{x}_i) \quad (10)$$

其中 $\mathbf{Q}_{i-1} = [\mathbf{q}_1, \dots, \mathbf{q}_{i-1}]$, 矩阵 $(\mathbf{I} - \mathbf{Q}_{i-1} \mathbf{Q}_{i-1}^T)$ 为 $\mathbf{q}_1, \dots, \mathbf{q}_{i-1}$ 张成空间的正交补空间的投影阵。

对 $\tilde{\mathbf{q}}_i$ 标准化后得到第 i 个基矢量 \mathbf{q}_i

$$\mathbf{q}_i = \mu_i^{-1} (\mathbf{I} - \mathbf{Q}_{i-1} \mathbf{Q}_{i-1}^T) \Phi(\mathbf{x}_i) \quad (11)$$

其中 $\mu_i = \|\mathbf{I} - \mathbf{Q}_{i-1} \mathbf{Q}_{i-1}^T\| \Phi(\mathbf{x}_i)\|$ 。由式(11)进一步得到

$$\begin{aligned} \Phi(\mathbf{x}_i) &= \mathbf{Q}_{i-1} \mathbf{Q}_{i-1}^T \Phi(\mathbf{x}_i) + \mu_i \mathbf{q}_i = \mathbf{Q}_i \begin{pmatrix} \mathbf{Q}_{i-1}^T \Phi(\mathbf{x}_i) \\ \mu_i \end{pmatrix} \\ &= \mathbf{Q} \begin{pmatrix} \mathbf{Q}_{i-1}^T \Phi(\mathbf{x}_i) \\ \mu_i \\ \mathbf{0}_{n-i} \end{pmatrix} = \mathbf{Q} \mathbf{r}_i \end{aligned} \quad (12)$$

其中 $\mathbf{Q}_i = [\mathbf{Q}_{i-1}, \mathbf{q}_i] = [\mathbf{q}_1, \dots, \mathbf{q}_i]$, $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_n]$, $\mathbf{0}_{n-i}$ 为所有元素均为 0 的 $(n-i) \times 1$ 矢量。

将式(12)写为矩阵形式得到

$$\mathbf{X}_\phi = \mathbf{Q} \mathbf{R} \quad (13)$$

其中 \mathbf{R} 为上三角矩阵, 其第 i 列为

$$\mathbf{r}_i = \begin{pmatrix} \mathbf{Q}_{i-1}^T \Phi(\mathbf{x}_i) \\ \mu_i \\ \mathbf{0}_{n-i} \end{pmatrix}$$

$\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_i]$ 的列矢量标准正交, \mathbf{R} 为上三角矩阵, 因

此式(13)即为特征空间中样本矩阵 X_ϕ 的 QR 分解。

由于有 $Q^T Q = I$ ，利用式(13)计算核矩阵 K 得到

$$K = X_\phi^T X_\phi = R^T Q^T Q R = R^T R \quad (14)$$

其中核矩阵 K 可由核函数计算得到 $(K)_{ij} = k(x_i, x_j)$ 。对相互独立的人脸图像样本，核矩阵 K 为对称正定阵， R 为上三角阵，因此根据 Cholesky 分解定理^[8]，式(14)就是 K 的 Cholesky 分解，对 K 进行 Cholesky 分解就可得到上三角阵 R 。 K 为满秩阵，由 Cholesky 分解得到的 R 可逆，因此由式(13)得到

$$X_\phi R^{-1} = Q \quad (15)$$

由式(13)和式(15)可看出，对非线性映射到特征空间中的样本 $\Phi(x_1), \dots, \Phi(x_n)$ 进行 Gram-Schmidt 正交化过程计算标准正交矢量 q_1, \dots, q_n ，等价于对 X_ϕ 进行 QR 分解， Q 的各列即为需要计算的标准正交矢量，因此式(13)和式(15)建立起了特征空间中样本和标准正交基之间的关系。

4 核鉴别共同矢量

基于上述所得特征空间中的 Gram-Schmidt 正交化过程，本节将 DCV 算法 2 推广为非线性领域。设特征空间中 DCV 算法 2 中子空间 B 对应的矩阵为 $B_\phi = [\Phi(b_1^1), \dots, \Phi(b_{N-1}^1), \Phi(b_1^2), \dots, \Phi(b_{N-1}^C)]$ ，其中 $\Phi(b_k^i) = \Phi(x_{k+1}^i) - \Phi(x_k^i)$ 。对应 DCV 算法 2 的第 2 步，设 B_ϕ 的秩空间的标准正交基矢量构成矩阵 Q_B 的各列，则由式(13)和式(15)得到

$$B_\phi = Q_B R_B \quad \text{和} \quad B_\phi R_B^{-1} = Q_B \quad (16)$$

对应 DCV 算法 2 的第 3 步，在特征空间每类任选一个样本投影到 B_ϕ 的秩空间的正交补空间，得到共同矢量

$$\Phi(x_{\text{com}}^i) = \Phi(x_m^i) - Q_B Q_B^T \Phi(x_m^i) \quad (17)$$

$\Phi(x_{\text{com}}^i)$ 独立于每类所选样本，本文选每类中的第一个样本，即 $m=1$ 。将式(16)代入式(17)得到共同矢量为

$$\Phi(x_{\text{com}}^i) = \Phi(x_1^i) - B_\phi R_B^{-1} (R_B^{-1})^T B_\phi^T \Phi(x_1^i) = \Phi(x_1^i) - B_\phi K_B^{-1} v_i$$

其中 $M-C \times M-C$ 矩阵 K_B 和列矢量 v_i 为

$$K_B = B_\phi^T B_\phi = R_B^T R_B \quad (18)$$

$$v_i = B_\phi^T \Phi(x_m^i) \quad (19)$$

其中 K_B 和 v_i 可由核函数计算得到。为简化计算，定义

$$u_i = K_B^{-1} v_i \quad (20)$$

因此共同矢量表示为

$$\Phi(x_{\text{com}}^i) = \Phi(x_1^i) - B_\phi u_i \quad (21)$$

在特征空间和 b_{com}^k 对应的矢量为 $\Phi(b_{\text{com}}^k)$

$$\begin{aligned} \Phi(b_{\text{com}}^k) &= \Phi(x_{\text{com}}^{k+1}) - \Phi(x_{\text{com}}^1) \\ &= \Phi(x_1^{k+1}) - B_\phi u_{k+1} - (\Phi(x_1^1) - B_\phi u_1) \\ &= \Phi(x_1^{k+1}) - \Phi(x_1^1) - B_\phi (u_{k+1} - u_1) \\ &= g_k - B_\phi h_k = [g_k, B_\phi] \begin{bmatrix} 1 \\ -h_k \end{bmatrix}, \quad k=1, 2, \dots, C-1 \end{aligned} \quad (22)$$

其中为简化，定义

$$g_k = \Phi(x_1^{k+1}) - \Phi(x_1^1), \quad h_k = (u_{k+1} - u_1) \quad (23)$$

由 $\Phi(b_{\text{com}}^k)$ 构成矩阵 $B_{\text{com}}^\phi = [\Phi(b_{\text{com}}^1), \dots, \Phi(b_{\text{com}}^{C-1})]$ ，对应

DCV 算法 2 的第 4 步，对 $\Phi(b_{\text{com}}^k)$ 进行 Gram-Schmidt 正交化首先需要计算核矩阵 $K_{\text{com}} = (B_{\text{com}}^\phi)^T B_{\text{com}}^\phi$ ，其第 i 行第 j 列元素计算为

$$\begin{aligned} (K_{\text{com}})_{ij} &= \langle \Phi(b_{\text{com}}^i), \Phi(b_{\text{com}}^j) \rangle \\ &= [1, -h_i^T] [g_j, B_\phi]^T [g_j, B_\phi] \begin{bmatrix} 1 \\ -h_j \end{bmatrix} \\ &= [1, -h_i^T] \begin{bmatrix} g_j^T g_j & g_j^T B_\phi \\ B_\phi^T g_j & B_\phi^T B_\phi \end{bmatrix} \begin{bmatrix} 1 \\ -h_j \end{bmatrix} \end{aligned} \quad (24)$$

其中 $g_i^T g_j$ 、 $g_i^T B_\phi$ 、 $B_\phi^T g_j$ 可由核函数计算， $K_B = B_\phi^T B_\phi$ 和 $h_k = (u_{k+1} - u_1)$ 已在式(18)和式(22)得到。

为实现 DCV 算法 2 的第 4 步，需要对 K_{com} 进行 Cholesky 分解获得上三角阵 R_{com}

$$K_{\text{com}} = (B_{\text{com}}^\phi)^T B_{\text{com}}^\phi = R_{\text{com}}^T R_{\text{com}} \quad (25)$$

设特征空间中的鉴别矢量为 w_{ϕ_i} ($i=1, \dots, C-1$)，由其构成非线性最佳变换阵 W_ϕ ，由式(15)建立的关系得到

$$W_\phi = [w_{\phi_1}, \dots, w_{\phi_{C-1}}] = B_{\text{com}}^\phi R_{\text{com}}^{-1} \quad (26)$$

由于非线性映射未知，特征空间中的最佳变换阵 W_ϕ 不能直接求出，但可计算得到第 i 类的特征矢量 Ω_i^ϕ

$$\Omega_i^\phi = W_\phi^T \Phi(x_1^i) = (R_{\text{com}}^{-1})^T (B_{\text{com}}^\phi)^T \Phi(x_1^i) = (R_{\text{com}}^{-1})^T p_i \quad (27)$$

其中 $p_i = (B_{\text{com}}^\phi)^T \Phi(x_1^i)$ 。 p_i 由下式计算出

$$\begin{aligned} (p_i)_j &= \Phi(b_{\text{com}}^j)^T \Phi(x_1^i) = g_j^T \Phi(x_1^i) - h_j^T B_\phi^T \Phi(x_1^i), \\ & \quad i=1, 2, \dots, C, \quad j=1, \dots, C-1 \end{aligned} \quad (28)$$

其中 $g_j^T \Phi(x_1^i) = k(x_1^{j+1}, x_1^i) - k(x_1^1, x_1^i)$ 和 $B_\phi^T \Phi(x_1^i)$ 可由核函数计算得到。待识别样本 x_{test} 的特征矢量为

$$\Omega_{\text{test}}^\phi = W_\phi^T \Phi(x_{\text{test}}) = (R_{\text{com}}^{-1})^T (B_{\text{com}}^\phi)^T \Phi(x_{\text{test}}) = (R_{\text{com}}^{-1})^T p_{\text{test}} \quad (29)$$

其中 $(p_{\text{test}})_j = k(x_1^{j+1}, x_{\text{test}}) - k(x_1^1, x_{\text{test}}) - h_j^T B_\phi^T \Phi(x_{\text{test}})$ ， $i=1, 2, \dots, C$ ， $j=1, \dots, C-1$ ，由核函数计算出。

分类阶段，将待识别样本归入 $\Omega_{\text{test}}^\phi$ 和 Ω_i^ϕ 的最小距离对应的类别 c 中

$$c = \min_i \|\Omega_i^\phi - \Omega_{\text{test}}^\phi\|, \quad i=1, 2, \dots, C \quad (30)$$

由此得到了 KDCV 方法。由于 KDCV 方法本质上是在特征空间用 Gram-Schmidt 正交化过程得到非线性鉴别矢量 w_{ϕ_i} ，因此最佳变换阵 W_ϕ 的列矢量 w_{ϕ_i} 具有标准正交的性质，此性质可用式(25)和式(26)得到验证

$$\begin{aligned} W_\phi^T W_\phi &= (R_{\text{com}}^{-1})^T (B_{\text{com}}^\phi)^T B_{\text{com}}^\phi R_{\text{com}}^{-1} = (R_{\text{com}}^{-1})^T K_{\text{com}} R_{\text{com}}^{-1} \\ &= (R_{\text{com}}^{-1})^T R_{\text{com}}^T R_{\text{com}} R_{\text{com}}^{-1} = I \end{aligned} \quad (31)$$

这说明得到的非线性鉴别矢量之间的冗余度最小，所得非线性鉴别特征之间的冗余度也最小，使所有的非线性鉴别特征均得到了最有效的利用。文献[9]加上鉴别矢量标准正交的约束条件，得到了计算标准正交 Fisher 线性鉴别矢量的迭代算法，因此本文的方法也可看作是对文献[9]结论的非线性推广。总结所得 KDCV 方法的计算步骤为：(1)由式(18)、式

(19)和式(20)计算 K_B , v_i 和 $u_i = K_B^{-1}v_i$; (2)由式(24)计算核矩阵 K_{com} ; (3)对 K_{com} 进行 Cholesky 分解得到上三角阵 R_{com} ; (4)由式(28)和式(30)得到每个类的特征矢量 Ω_i^{ϕ} 和待识别样本的特征矢量 Ω_{test}^{ϕ} ; (5)由式(31)计算最小距离对待识别样本分类。

由于每类的特征矢量独立于类别中所选样本, 对每个类只需计算一个特征矢量, 因此在识别阶段只需计算待识别样本特征和每个类别特征的距离, 计算复杂度只和类别数有关, 而和总样本数无关, 因此降低了 KDCV 算法在识别阶段的计算复杂度。

5 实验结果与分析

实验在 ORL 标准人脸数据库上进行。ORL 人脸库由 40 个人, 每人 10 幅 112×92 的灰度图像组成。其人脸表情和细节均有不同程度的变化。为降低计算量, 先对每幅人脸图像进行 2 级小波变换, 用低频分量作为原始图像, 每幅图像缩小为 28×23 。所有图像的像素值均归一化到 $[0, 1]$ 区间内, 再按列拉直成矢量, 得到 644 维的高维矢量。

实验将 40 个人共 400 幅图像随机分成两组, 一组作为训练样本, 剩下的一组为测试样本, 两组之间的样本没有重叠。每类别所选样本数均相同。为消除单次选择样本的随机性, 每次实验均独立重复进行 20 次取平均识别率。实验均用最近邻作为分类准则, 并用欧氏距离作为距离测度。取不同训练样本数和不同特征数, 将本文的 KDCV 方法和 DCV 以及 KPCA+LDA 方法^[10]进行比较。文献[10]证明了 KPCA+LDA 方法等价于 KFDA 方法, 可得到非线性 Fisher 最佳鉴别特征。每次实验中 3 种方法所得结果均是用相同训练样本和测试样本得到, KDCV 和 KPCA+LDA 方法使用 2 阶多项式核函数。总类数为 40, 因此 3 种方法最多可得 39 个鉴别特征。表 1 是 3 种方法的平均识别率, 可看出在每类选不同训练样本数时 KDCV 均明显优于 KPCA+LDA 方法, 当每类训练样本数取 2 个以上时, KDCV 方法优于 DCV 方法。由于 KPCA+LDA 方法中的 Fisher 准则用总散度阵取代了类内散度阵, 由文献[3]知其所得非线性鉴别特征仅能最优化类内散度, 而不能保证最优化类间散度, 因此虽然 DCV 方法抽取的是线性特征, 也能获得优于 KPCA+LDA 方法所得非线性特征的识别性能。

图 1 为 3 种方法在取不同最佳鉴别特征数时的平均识别率。每类选 5 个样本作为训练样本, 剩下的 5 个作为测试样本, 独立重复实验 20 次取平均值, 每次实验中 3 种方法所

的最佳特征时, KDCV 仍优于 DCV, 且两种方法均优于 KPCA+LDA 方法。

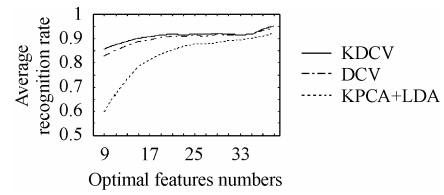


图 1 取不同最佳鉴别特征数时的平均识别率
Fig.1 Average recognition rate with different optimal discriminant features

为测试所得鉴别特征的鉴别能力, 取 5 个类别, 每类 5 个测试样本, 图 2 为 3 种方法抽取测试样本的两个最佳鉴别特征的散度图, 可看出 KDCV 方法所得鉴别特征的类可分性最好, 类间没有重叠, 而 DCV 方法所得鉴别特征的两个类间有部分重叠, 且各类的类内散度大于 KDCV 方法, 这说明 KDCV 可得到类可分性较好的非线性鉴别特征。KPCA+LDA 所得非线性鉴别特征仅能最优化类内散度, 其特征的类间重叠较大, 类可分性较差。

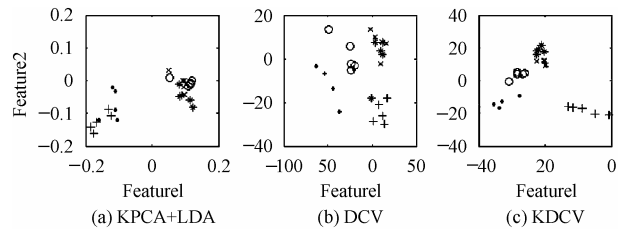


图 2 3 种方法抽取两个最佳鉴别特征的散度图
Fig.2 Scatter figure of three methods with two optimal discriminant features

6 结束语

本文利用核方法将 DCV 方法推广到非线性领域。首先建立特征空间中样本矢量和标准正交基矢量的关系, 然后将 DCV 算法 2 中需要进行两次 Gram-Schmidt 正交化的过程, 转化为只需计算两个核矩阵和一次 Cholesky 分解。由于无需计算特征方程, 得到的 KDCV 方法有较好的计算稳定性和较低的计算复杂度, 且所得非线性鉴别矢量具有标准正交的性质, 这使所得非线性鉴别矢量间的冗余度最小, 使所有鉴别特征均能被有效利用。实验显示了所得 KDCV 方法的非线性鉴别特征优于 DCV 方法的线性鉴别特征和 KPCA+LDA 方法的非线性特征。

表 1 平均识别率比较(%)

Tab.1 Comparison of average recognition rate (%)

方法	每类训练样本数							
	2	3	4	5	6	7	8	9
KPCA+LDA	79.13	86.93	89.85	93.05	93.31	94.29	96.19	96.75
DCV	83.81	90.43	92.29	94.85	95.06	96.08	95.94	96.50
KDCV	83.25	90.50	93.02	95.53	96.00	97.33	98.00	98.75

参 考 文 献

- [1] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711–720.
- [2] Yu H, Yang J. A Direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 2001, 34(10): 2067–2070.
- [3] Chen L F, Liao H YM, Ko M T, Lin JC, Yu G J. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, 2000, 33(10): 1713–1726.
- [4] Huang R, Liu Q, Lu H, Ma S. Solving the small size problem of LDA. Proc. 16th Int'l Conf. Pattern Recognition, Quebec City, Que., Canada, 2002, 3(8): 29–32.
- [5] Cevikalp H, Neamtu M, Wilkes M, Barkana A. Discriminative common vectors for face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2005, 27(1): 4–13.
- [6] Gülmezoglu M B, Dzhafarov V, Barkana A. The common vector approach and its relation to principal component analysis. *IEEE Trans. Speech and Audio Processing*, 2001, 9(6): 655–662.
- [7] Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. England: Cambridge Univ. Press, 2004, Part 2.
- [8] 程云鹏. 矩阵论[M]. 西安: 西北工业大学出版社, 2001, 第 4 章.
- [9] Foley D H, Sammon J W. An optimal set of discriminant vectors. *IEEE Trans. on Comput*, 1975, 24(3): 281–289.
- [10] Yang J, Jin Z, Yang JY. Essence of kernel Fisher discriminant: KPCA plus LDA. *Pattern Recognition*, 2004, 37(10): 2097–2100.
- 贺云辉: 男, 1975 年生, 博士, 研究方向为图像识别.
赵 力: 男, 1958 年生, 教授, 研究方向为语音信号处理.
邹采荣: 男, 1963 年生, 教授, 研究方向为图像处理.