

基于散度差准则的隐空间特征抽取方法^{*}

陈才扣^{1,2} 宋枫溪³ 刘永俊¹ 杨静宇²

(扬州大学计算机科学与工程系 扬州 225009)¹ (南京理工大学计算机科学与工程系 南京 210094)²
(炮兵学院 1 系 合肥 230031)³

摘要 本文提出了一种新的非线性特征抽取方法——基于散度差准则的隐空间特征抽取方法。该方法的主要思想就是首先利用一核函数将原始输入空间非线性变换到隐空间,然后,在该隐空间中,利用类间离散度与类内离散度之差作为鉴别准则进行特征抽取。与现有的核特征抽取方法不同,该方法不需要核函数满足 Mercer 定理,从而增加了核函数的选择范围。更为重要的是,由于采用了散度差作为鉴别准则,从根本上避免了传统的 Fisher 线性鉴别分析所遇到的小样本问题。在 ORL 人脸数据库和 AR 标准人脸库上的试验结果验证了本文方法的有效性。

关键词 隐空间,散度差鉴别准则,特征抽取,人脸识别

Feature Extraction Method Based on Scatter Difference Criterion in Hidden Space

CHEN Cai-Kou^{1,2} SONG Feng-Xi³ LIU Yong-Jun¹ YANG Jing-Yu²

(Department of Computer Science and Engineering, Yangzhou University, Yangzhou 225001)¹

(Department of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094)²

(Department One, Artillery Academy, Hefei 230031)³

Abstract In this paper, a novel feature extraction method based on scatter difference criterion in hidden space is developed. Its main idea is that the original input space is first mapped into a hidden space through a prespecified kernel function, in which space the feature extraction is conducted using the difference of between-class scatter and within-class scatter as the discriminant criterion. Different from the existing kernel feature extraction methods, the kernel function used in the proposed one is not required to satisfy Mercer's theorem so that they can be chosen from a wide range. It is more important that due to adoption of the scatter difference as the discriminant criterion for feature extraction, the proposed method essentially avoids the small size samples problem usually occurred in the traditional Fisher linear discriminant analysis. Finally, extensive experiments are performed on ORL face database and AR face database. The experimental results indicate that the proposed method outperforms the traditional scatter difference discriminant analysis in recognition performance.

Keywords Hidden space, Scatter difference discriminant criterion, Feature extraction, Face recognition

1 引言

核方法^[6~10]是当前模式识别领域中一个迅猛发展的新方向。它的主要思想最初是由 V. Vapnik^[10]提出并应用于支持向量机(SVM)中。近年来,人们在核方法的基础上,做了不断的完善和扩展。尽管核方法在图像等模式识别领域应用得非常广泛和成功,但在核方法中使用的核函数要求必须满足 Mercer 定理,即要求核函数是正定的,而这类核函数的数量是有限的,如 Gauss 核函数、多项式核函数、样条核函数等。因此,这一正定条件成为核方法应用的一大限制。同时这些有限的函数未必符合所有的实际应用问题。针对这一问题, Zhang 等人^[6]提出了隐空间思想,它通过一簇隐函数 $\varphi(x) = [k(x_1, x), k(x_2, x), \dots, k(x_m, x)]^T$ (其中, $k(\cdot, \cdot)$ 可以是任意对称的实值函数)将输入空间中的样本映射到隐空间,然后在隐空间中使用最优超平面算法。该方法由于不要求核函数正定,使得基于核的算法可以采用更多的核函数。因此,隐空间

映射方法是对传统核方法的重要推广。

线性鉴别分析是公认的特征抽取的最为重要的工具之一,目前仍然被广泛地应用在人脸等图像识别领域^[1,2]。近年来人们在 Fisher 思想的基础上,做了不断的完善和扩展^[3~6]。虽然 Fisher 线性鉴别函数巧妙地应用广义 Rayleigh 商为我们提供了一种有效的准则函数,但是在诸如人脸识别等典型的高维小样本问题的求解过程中,类内散布矩阵常常是奇异的。目前人们往往采用的是一种“近似最优”的方法^[3],想方设法来回避这个问题,而都没有从根本上解决它。虽然修正交补空间法^[4]之后,杨建、杨静宇等^[6]给出了一整套较为完善的寻找最优鉴别矢量的算法,为这一问题的解决作出了重要贡献,但是其算法的复杂程度也是显而易见的。最近,宋枫溪等^[9]提出了一种最大散度差分类器,与 Fisher 鉴别准则不同的是,它是利用类间散度与类内散度之差作为分类器准则,这样,从理论上根本地消除了 Fisher 鉴别准则中存在的因类内散布矩阵奇异而无法求解的问题。而且在这种

^{*} 本文得到国家自然科学基金(60472060)、江苏省博士后科研资助计划项目和江苏省高校自然科学基金(05KJB520152)的资助。陈才扣 博士后,副教授,主要研究方向为模式识别理论与应用、生物特征识别;宋枫溪 教授,博士,主要研究方向为自动文本分类、人脸识别;刘永俊 硕士研究生,杨静宇 教授,博士生导师,主要研究方向为模式识别与智能系统。

基于散度差的鉴别准则中由于不需要再构造逆矩阵,使得整个算法的速度有了很大的提高。但是,最大散度差分类器只是抽取一个最优鉴别矢量作为样本模式的投影分类方向,而在样本类别数较多的情况下,单一的投影方向是不够的。

本文将散度差鉴别准则与隐空间思想相结合,提出了一种新的基于散度差准则的隐空间图像投影鉴别分析方法。该方法从根本上避免了由于小样本问题而引起的类内散布矩阵奇异的问题。与最大散度差分类器(只能抽取一个最优鉴别矢量)不同的是,本方法可以抽取多个最优鉴别特征,提高了分类性能。同时,由于采用了隐空间思想,能够提取数据的非线性特征,提高了算法的性能。通过在 ORL 人脸数据库和 AR 人脸数据库上试验比较验证了本文所提算法的有效性。

2 原理与方法

2.1 隐空间

设 $X = \{x_1, x_2, \dots, x_N\}$ 表示 N 个 n 维的独立同分布的样本组成的集合。定义一个由实值函数集合 $\{\varphi_i(x) | i=1, 2, \dots, n_1\}$ 构成的向量:

$$\varphi(x) = [\varphi_1(x), \varphi_2(x), \dots, \varphi_{n_1}(x)]^T \quad (1)$$

向量 $\varphi(x)$ 将 n 维输入空间中的样本映射到一个新的 n_1 维空间,即

$$x \xrightarrow{\varphi} y = [\varphi_1(x), \varphi_2(x), \dots, \varphi_{n_1}(x)]^T \quad (2)$$

由于函数集合 $\{\varphi_i(x)\}$ 所起的作用类似于径向基函数网络中的隐神经元,因此称之为隐函数。相应的空间 $Y = \{y | y = [\varphi_1(x), \varphi_2(x), \dots, \varphi_{n_1}(x), x \in X]^T\}$ 称为隐空间。

考虑一种特殊的隐函数:核函数 $k(x_i, x_j)$, 则核映射定义为

$$x \xrightarrow{k} y = [k(x_1, x), k(x_2, x), \dots, k(x_N, x)]^T \quad (3)$$

基于 X 的相应的隐空间可表示为 $Y = \{y | y = [k(x_1, x), k(x_2, x), \dots, k(x_N, x), x \in X]^T\}$ 。

在 SVMs 中经常使用的几类核函数主要有:多项式核函数 $(k(x, y) = (a(x \cdot y) + b)^d, a > 0, b \geq 0, d \text{ 为正整数})$ 、高斯核函数 $(k(x, y) = \exp(-\frac{\|x - y\|^2}{2\sigma^2}), \sigma^2 \text{ 代表方差})$ 和 Sigmoid 核函数 $(k(x, y) = S(v(x \cdot y) + c), S(u) \text{ 是 Sigmoid 函数, 如双曲正切函数 } \tanh \text{ 等})$ 。

下面定义一种直接基于图像矩阵的相似度。

定义 1 设 A_i, A_j 是 $m \times n$ 阶图像矩阵,实数 s 定义为

$$s(A_i, A_j) = \frac{\text{Tr}(A_i A_j^T + A_j A_i^T)}{\text{Tr}(A_i A_i^T + A_j A_j^T)} \quad (4)$$

则称 $s(A_i, A_j)$ 为矩阵 A_i, A_j 的相似度。

根据定义 1, 容易证明 $s(A_i, A_j)$ 具有下列性质:

- (1) $s(A_i, A_j) = s(A_j, A_i)$;
- (2) $s(A_i, A_j) = s(A_i^T, A_j^T)$;
- (3) $-1 \leq s(A_i, A_j) \leq 1$, 若 $(A_i, A_j) = 1$, 则 $A_i = A_j$ 。

由上述定义与性质,可以知道 $s(A_i, A_j)$ 描述了 A_i, A_j 之间的一种相似关系。若 $s(A_i, A_j)$ 趋向于 1, 则 A_i, A_j 之间的差距接近于零。

定义 2 定义一个核映射 $\varphi: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^N, \varphi = s(\cdot, A) = [s(A_1, A), s(A_2, A), \dots, s(A_N, A)]^T$ 。

2.2 基于散度差准则的特征抽取

设 $\omega_1, \omega_2, \dots, \omega_c$ 为 c 个模式类, $X = \{A_i\}, i=1, 2, \dots, N$ 为 $m \times n$ 阶训练样本图像集, X 中的每一个 A_i 属于 ω_j 类,即

$$A_i \in \omega_j, i=1, 2, \dots, N, j=1, 2, \dots, c.$$

根据定义 2, 将所有的 $m \times n$ 阶的训练样本图像 $A_i (i=1, \dots, N)$ 通过映射 φ 变换到隐空间 $\mathbb{R}^N$ 中, 对于每个原始输入空间中的训练样本 A_i , 令其对应的隐空间中的像为 z_i , 则有

$$z_i = [s(A_i, A), s(A_2, A), \dots, s(A_N, A)]^T, i=1, 2, \dots, N \quad (5)$$

令隐空间中 ω_i 类的平均矢量、协方差矩阵与先验概率分别为 $m_i, S_i, P(\omega_i)$, 则隐空间上样本模式 z_i 的类间散布矩阵 S_b 和类内散布矩阵 S_w 分别为

$$S_b = \sum_{i=1}^c P(\omega_i) (m_i - m)(m_i - m)^T \quad (6)$$

$$S_w = \sum_{i=1}^c P(\omega_i) E\{(z - m_i)(z - m_i)^T | \omega_i\} = \sum_{i=1}^c P(\omega_i) S_i \quad (7)$$

$$S_i = E\{(x - m_i)(x - m_i)^T | \omega_i\} \quad (8)$$

其中, $m_i = E\{z | \omega_i\} = \sum_{k=1}^{N_i} z_k, m = E\{z\} = \sum_{i=1}^c P(\omega_i) m_i$ 为隐空间上全体训练样本的平均矢量。

由散布矩阵, 散度差鉴别准则函数可定义为:

$$J_s(w) = w^T (S_b - S_w) w \quad (9)$$

其中, w 为隐空间上的 N 维列矢量。

使函数 $J_s(w)$ 达到最大值的矢量 w^* 为最大散度差鉴别分析的最优鉴别方向, 训练样本在方向上的投影集在一维子空间 $\text{Span}\{w^*\}$ 中有最小的类内距离和最大的类间距离。

定理 1 设 $J_s(w) = w^T (S_b - S_w) w$, 则使 $J_s(w)$ 达到最大值的单位矢量即为 $S_b - S_w$ 的最大特征值对应的单位特征矢量。

证明: 求使 $J_s(w)$ 达到最大值的单位矢量相当于求解下列问题

$$\text{Max}_w J_s(w) = (w^*)^T (S_b - S_w) w^* = \text{Max}_w \frac{w^T (S_b - S_w) w}{w^T w} \quad (10)$$

因为

$$\frac{w^T (S_b - S_w) w}{w^T w} = \frac{w^T (S_b - S_w) w}{w^T I w} \quad (11)$$

其中, I 为单位矩阵。

根据广义 Rayleigh 商的极值性质^[10], w^* 为 $I^{-1} (S_b - S_w) = S_b - S_w$ 的最大特征值对应的单位特征矢量。

一般说来, 在样本类别数较多的情况下, 单一的最优投影方向是不够的, 我们希望寻找一组满足标准正交条件且极大化准则函数式(9)的最优投影轴 $w_1^*, w_2^*, \dots, w_d^*$ 。很容易从定理 1 得到下面的推论:

推论 1 设 $J_s(w) = w^T (S_b - S_w) w$, 最大化散度差准则函数 $J_s(w)$ 的一组最优鉴别矢量 $w_1^*, w_2^*, \dots, w_d^*$ 一般取为特征方程 $(S_b - S_w) w = \lambda w$ 的 d 个最大的特征值所对应的特征向量。即 $w_1^*, w_2^*, \dots, w_d^*$ 满足以下条件:

$$(S_b - S_w) w_j^* = \lambda_j w_j^*, j=1, \dots, d \quad (12)$$

这样我们就能够得到一个与较大本征值对应的一组最优投影轴构成的投影矩阵 $W = [w_1^*, w_2^*, \dots, w_d^*]$ 进行特征抽取。

可以看到, 在进行式(12)的特征分解的时候, 与 Fisher 鉴别准则的特征分解 $(S_w^{-1} S_b)$ 相比, 把对 S_w 这样一个庞大矩阵的求逆过程变为了求一个散度差矩阵 $(S_b - S_w)$, 从而使得计算的复杂度大大降低。更重要的是, 由于避免了对 S_w 的求逆, 再也不用担心由 S_w 的奇异性带来的困扰了。因为由

定义知 S_b, S_w 均为 n 阶实对称矩阵, 所以散度差矩阵 $(S_b - S_w)$ 仍是一个 $n \times n$ (n 为样本向量的维数) 的实对称矩阵。因此, 不管 S_w 奇异与否, 我们总能够对 $(S_b - S_w)$ 进行特征分解求得所需要一组最优投影轴。这也是最大散度差线性鉴别分析较 Fisher 线性鉴别分析最大的优势所在。

2.3 特征抽取方法

设最优图像投影轴为 $w_1, \dots, w_d$, 对于一给定的图像 A , 令

$$y = W^T z = W^T [s(A_1, A), s(A_2, A), \dots, s(A_N, A)]^T$$
$$= \begin{bmatrix} w_1^T \\ \vdots \\ w_d^T \end{bmatrix} [s(A_1, A), s(A_2, A), \dots, s(A_N, A)]^T$$
$$= [y_1, y_2, \dots, y_d]^T \tag{13}$$

3 实验及分析

3.1 实验 1

ORL 标准人脸库由 40 人, 每人 10 幅 112×92 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。图 1 是 ORL 人脸库中某一人的 5 幅测试图像。

本试验中, 我们每次随机选取每人的五幅图像作为训练样本, 剩余的五幅作为测试样本, 这样训练样本和测试样本总数均为 200。分别选用多项式核函数、高斯核函数和基于相似度的核函数构造隐空间。为了描述方便, 称基于上面 3 种核函数的散度差特征抽取方法分别为 HSDD-P、HSDD-G 和 HSDD-S, 普通的散度差鉴别分析方法为 SDD。用上述四种方法对不同分辨率的 ORL 人脸图像进行特征抽取, 得到 3 组最优投影矢量, 将所有的样本分别投影到由上述 3 组投影矢量张成的子空间上, 然后, 在每个投影空间内, 采用最近邻分类器进行分类, 此外, 我们还做了线性鉴别分析 (LDA) 和核 Fisher 鉴别分析方法 (KFD) 的比较试验, 识别结果见表 1。表 2 给出了 6 种方法在 28×23 分辨率下的特征抽取时间和识别时间。

从表 1 中 6 种分类器在不同分辨率人脸图像下的识别结果可以看到, HSDDs 在各个分辨率图像上有着良好的表现, 其性能优于其它 5 个特征抽取方法, 并且, KFD、HSDDp 与 HSDDs 的结果不相上下。随着图像分辨率的降低 ($112 \times 92, 56 \times 46, 28 \times 23$), 6 种分类器的识别性能略有下降。从表 2 中 6 种方法的特征抽取时间不难发现, 基于散度差准则

的特征抽取方法在特征抽取速度上明显比基于 Fisher 准则的 LDA 和 KFD 方法快, 这也进一步说明散度差方法的良好速度优势。

3.2 实验 2

本试验在一个规模较大的 AR 人脸图像数据库上进行。它由 126 人的 4000 多幅彩色图像组成。其中 120 人每人有 26 张在不同时期、光照、姿势、表情、遮挡等条件下拍摄的人脸图像。这 26 张图像摄于两个时期, 前 13 张图像为一个时期, 后 13 张图像为另一个时期, 前后相差 14 天。本实验仅考察无遮挡的人脸图像, 即每个人的第 1~7 张和第 14~20 张图像。取每个人的第 1~7 张图像作为训练样本, 第 14~20 张图像作为测试样本, 训练样本和测试样本总数均为 840 个。为消除人脸位置、光照等因素的影响, 我们采用了分辨率为 50×40 的规范化人脸图像, 并进行了灰度均衡化处理。AR 人脸库中某人经过灰度均衡化处理后的 14 幅规范化人脸图像如图 2 所示。



图 1 ORL 人脸库中的 5 幅图像

表 1 6 种特征抽取方法在不同分辨率 ORL 图像下的识别率 (%)

分辨率	112(92	56 × 46	28 × 23	14 × 12	7 × 6
LDA	0.885	0.880	0.865	0.850	0.810
KFD	0.965	0.960	0.940	0.925	0.915
SDD	0.950	0.950	0.945	0.930	0.925
HSDD-P	0.965	0.965	0.960	0.945	0.920
HSDD-G	0.965	0.965	0.960	0.945	0.920
HSDD-S	0.970	0.970	0.965	0.960	0.945

表 2 6 种方法在最近邻分类器下的特征抽取和识别的时间对照表

时间(秒)	特征抽取时间	分类时间	总时间
LDA	69.24	6.36	76.60
KFD	80.18	7.24	87.42
SDD	46.89	6.78	53.67
HSDD-P	61.67	7.21	68.88
HSDD-G	61.58	7.25	68.83
HSDD-S	61.69	7.19	68.88



图 2 AR 人脸库中某人经过灰度均衡化处理后的 14 幅规范化人脸图像

(下转第 199 页)

果,怎样使约简数达到最少,以及在此基础上开发增量式的挖掘算法、并行挖掘算法等都是今后有用的研究方向。

参考文献

- 1 王晓峰,王天然.一种自顶向下挖掘频繁项的有效方法.计算机研究与发展,2004(1)
- 2 Lin D I, Kedem Z M, Pincer-search: A New Algorithm for Discovering the Maximum Frequent Set [C]. In: Proc. of the 6th European Conf. on extending database technology, 1998
- 3 Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD Conference on Management of data, 1993. 207~216

- 4 Lin J L, Dunham M H. Mining association rules: Anti-skew algorithms. In: Proceedings of the International Conference on Data Engineering, Orlando, Florida, February 1998
- 5 Brin S, Motwani R, Ullman J D, et al. Dynamic Itemset counting and implication rules for market basket data. In: ACM-SIGMOD International Conference on the Management of Data, 1997
- 6 冯玉才,冯剑琳.关联规则的增量式更新算法[J].软件学报,1998(4):3010
- 7 Roberto J, Bayardo Jr. Efficiently mining long patterns from databases, [C]. In: Proc. of the '98 ACM-SIGMOD int. Conf. On Management of Data (SIGMOD'98), 1998. 85~93
- 8 王晓峰,王天然.相关测度与增量式支持度和信任度的计算[J].软件学报,2002(11):2208~2214

(上接第 163 页)

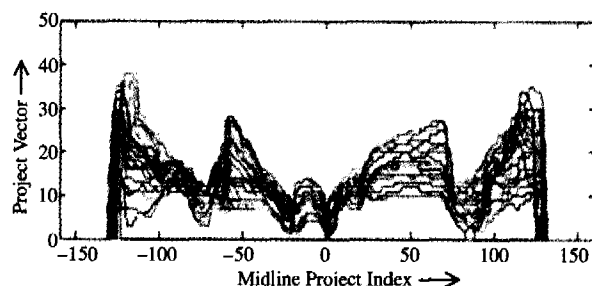


图5 轮廓沿中线投影叠加效果

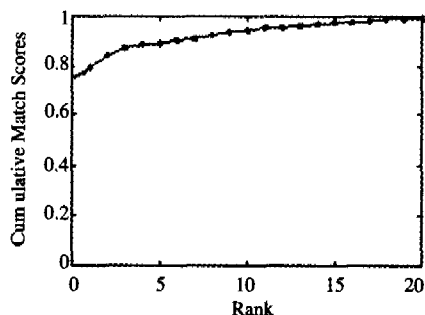


图6 累积匹配分值

4.3 实验结果分析和总结

本文讨论了将人体轮廓沿中线投影的新颖的步态特征提取方法,此方法包括了更多的人体特征,如手的摆动情况。与文[6]相比,本文的特征提取方法更加新颖,方法更易实现,且计算量低,从识别性能中,相比指纹、虹膜等生理特征,步态、笔迹等行为特征的识别能力偏弱的情况下,该方法的正确分类率 CCR(Correct Classification Rate)达到如图 6 所示,实验结果比较令人满意。实验表明该算法不仅获得了令人鼓舞的识别性能,而且拥有相对较低的计算代价,是一种有效的步态特征提取与识别方法。

参考文献

- 1 Abdelkader C B, Cutler R, Nanda H, et al. EigenGait: Motion-Based Recognition Using Image Self-Similarity. LNCS, 2001, 2091:289~294
- 2 Hayfron-Acquah J B, Nixon M S, Carter J N. Automatic Gait Recognition by Symmetry Analysis. Pattern Recog Letters, 2003, 24: 2175~2183
- 3 Lee L, Grimson WEL. Gait Analysis for Recognition and classification. In: Proceeding of the IEEE Conference on Face and Gesture Recognition, 2002. 155~161
- 4 Collins R, Gross R, Shi J. Silhouette-based Human Identification from Body Shape and Gait. In: Proc IEEE Conf, 2002
- 5 Wang L(王亮), Tan T N, Hu W M, et al. Silhouette analysis-based gait recognition for human identification. IEEE, 2003. 1505~1518
- 6 王亮,胡卫明,谭铁牛.基于步态的身份识别.计算机学报, 2006, 26(3):353~360
- 7 <http://www.sinobiometrics.com>

(上接第 176 页)

实验结果如表 3 所示,不难看出,HSDD-S 的正确识别率均分别优于其他 5 种特征抽取方法,基于散度差准则的特征抽取方法的特征抽取速度也优于基于 Fisher 准则的 LDA 和 KFD 方法。

表 3 6 种方法在最近邻分类器下的识别率(%),特征抽取和识别时间(s)

特征抽取方法	识别率	特征抽取时间	分类时间	总时间
LDA	68.0	569.24	76.36	645.60
KFD	72.5	744.14	70.47	814.61
SDD	68.4	351.27	75.45	426.72
HSDD-P	74.5	514.89	72.39	587.28
HSDD-G	74.5	517.12	72.19	589.31
HSDD-S	75.1	516.46	72.08	588.54

结论 本文提出了一种新的非线性鉴别分析方法——基于散度差准则的隐空间图像特征抽取方法。对于图像的特征抽取来说,该方法具有以下优点:一是计算更高效,特征抽取的速度有了显著的提高;二是避免了高维小样本问题所引起的类内散布矩阵奇异的问题;三是不仅能够抽取非线性特征,同时,与现有的其他核方法相比,本文的方法不需要核函数满足正定条件,扩大了核函数的选取范围,这样可以选择更

加适合具体问题的核函数。在 ORL 人脸数据库和 AR 人脸数据库上的实验结果验证了以上的结论。

参考文献

- 1 Fisher R A. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 1936, 7:178~188
- 2 Belhumeur P N, et al. Eigenfaces vs Fisherfaces: Recognition using class specific linear projections. IEEE Trans Pattern Anal Machine Intell, 1997, 19(7): 711~720
- 3 Hong Z Q, Yang J Y, et al. Optimal discriminant plane for a small number of samples and design method of classifier on the plane. Pattern Recognition, 1991, 24(4):317~324
- 4 Liu K, Yang J Y, et al. An efficient algorithm for Foley-sammon optimal set of discriminant vectors by algebraic method. International Journal of Pattern Recognition and Artificial Intelligence, 1992, 6(5):817~829
- 5 金忠,杨静宇,陆建峰.一种具有统计不相关性的最优鉴别向量集.计算机学报,1999,22(10):1105~1108
- 6 杨健.线性投影分析的理论及算法及其在特征抽取中的应用研究.[学位论文].南京:南京理工大学,2002
- 7 Yang J, Zhang D, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Machine Intell, 2004, 26(1):131~137
- 8 Yang J, Yang J Y. From image vector to matrix: a straightforward image projection IMPCA vs PCA. Pattern Recognition, 2002, 35(9):1997~1999
- 9 宋枫溪,程科,杨静宇.最大散度差和大间距线性投影与支持向量.自动化学报,2004, 30(11):890~896
- 10 程云鹏.矩阵论.西安:西北工业大学出版社,1999, 294~302