

统计不相关最佳鉴别矢量集的本质研究

吴小俊^{1,2,3}, 杨静宇², 王士同^{1,2}, 刘同明¹, Josef Kittler⁴

(1. 华东船舶工业学院, 江苏镇江 212003; 2. 南京理工大学信息学院, 南京 210094;
3. 中国科学院沈阳自动化研究所机器人学重点实验室, 沈阳 110015;
4. CVSSP, University of Surrey, Surrey GU2 7XH, UK)

[摘要] 对统计不相关最佳鉴别矢量集的本质进行研究, 在基于总体散布矩阵特征分解的基础上, 构造了一种白化变换, 使得变换后的样本空间中的总体散布矩阵为单位矩阵, 这样使得传统的最佳鉴别矢量集算法得到的均是具有统计不相关的最佳鉴别矢量集, 从而揭示了统计不相关最佳鉴别变换的本质——白化变换加普通的线性鉴别变换。该方法的优点在于所获得的最优鉴别矢量同时具有正交性和统计不相关性。该方法对代数特征抽取具有普遍适用性。用 ORL 人脸数据库的数值实验, 验证了该方法的有效性。

[关键词] 模式识别; 特征抽取; 鉴别分析; 广义最佳鉴别矢量集; 人脸识别

[中图分类号] TP391.4; O235 **[文献标识码]** A **[文章编号]** 1009-1742(2004)02-0044-04

1 引言

在模式识别领域中, Fisher 线性判别方法有着重大的影响, 其基本思想是在 Fisher 鉴别准则函数取极值的条件下, 求得一个最佳鉴别方向, 然后再将模式高维特征向量投影到该最佳鉴别方向上, 构成一维鉴别特征空间, 于是模式鉴别分析就在一维空间中进行^[1~15]。

Foley 和 Sammon 在 1970 年发展了 Fisher 线性判别方法, 提出了 Sammon 最佳鉴别平面的技术, 并将它用于解决两类问题^[2]。Sammon 最佳鉴别平面的技术在模式识别领域中得到广泛的应用与发展, Duchene 和 Leclercq 给出了对多类问题的 Foley-Sammon 最佳鉴别矢量集的计算公式^[3], Longstaff 提出了分别基于 Fukunaga-Koontz 变换和基于矢径 (radius vector) 的最佳鉴别平面, Turk 和 Pentland 提出了特征脸的方法^[1], Hong 和 Yang 提出了基于 SVD 的特征抽取方法^[4, 5], Cheng 和 Yang 提出了一种新的相似鉴别准则^[7], Liu 提出

了广义最佳鉴别平面和广义最佳鉴别矢量集的一系列方法^[8, 9], 郭提出了广义最佳鉴别矢量的改进算法^[10~12], 笔者最近提出了广义最佳鉴别矢量集的解析算法^[13]。在实际问题中, 模式原始特征的维数一般比较高, 特征分量可能是相关的, 为此, 金忠提出了一种具有统计不相关性的最佳鉴别平面和统计不相关最佳鉴别矢量集的算法^[14, 15]。笔者对统计不相关最佳鉴别矢量集的本质进行了研究, 并将其应用于人脸识别问题的研究上。

2 统计不相关最佳鉴别矢量集

设 w_1, w_2, \dots, w_m 为 m 个模式类, $\mathbf{X} = \{x_i\}$, $i = 1, 2, \dots, N$ 为 n 维训练样本集, \mathbf{X} 中的每一个 x_i 属于 w_j 类, 即 $x_i \in w_j, i = 1, 2, \dots, N, j = 1, 2, \dots, m$ 。设 w_i 类的平均矢量、协方差矩阵与先验概率分别为 $m_i, C_i, P(w_i)$, 则类间散布矩阵 S_b 、类内散布矩阵 S_w 和总散布矩阵 S_t 分别为:

$$S_b = \sum_{i=1}^m P(w_i)(m_i - m_0)(m_i - m_0)^T, \quad (1)$$

[收稿日期] 2003-07-02; 修回日期 2003-10-23

[基金项目] 江苏省自然科学基金资助项目 (BK2002001); 江苏省高校自然科学基金研究计划资助项目 (01KJB520002); 中国科学院机器人学开放实验室基金资助项目 (RL200108); 图像处理与图像通信实验室开放基金资助项目 (IPICL0304)

[作者简介] 吴小俊 (1967-), 男, 江苏丹阳市人, 博士, 华东船舶工业学院副教授

$$S_w = \sum_{i=1}^m P(w_i) E\{(x - m_i)(x - m_i)^T / W_i\} = \sum_{i=1}^m P(w_i) C_i, \quad (2)$$

$$C_i = E\{(x - m_i)(x - m_i)^T / w_i\}, \quad (3)$$

$$S_t = S_b + S_w = E\{(x - m_0)(x - m_0)^T\}, \quad (4)$$

$$m_0 = E\{x\} = \sum_{i=1}^m P(w_i) m_i, \quad (5)$$

其中 m_0 为全体训练样本的平均矢量, E 表示数学期望。

由散布矩阵, Fisher 鉴别函数可定义为

$$J(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T S_w \varphi}, \quad (6)$$

其中 φ 为任一 n 维列矢量。使函数 $J(\varphi)$ 达到最大值的矢量 φ_1^* 为 Fisher 最佳鉴别方向, 训练样本在方向 φ_1^* 上的投影集, 在一维子空间 $\text{Span}\{(\varphi_1^*)\}$ 中有最小的类内距离和最大的类间距离。

设 $\varphi_1 = \varphi_1^* / \|\varphi_1^*\|$, 则 φ_1 是 Foley-Sammon 最佳鉴别矢量集的第一个矢量, Foley-Sammon 最佳鉴别矢量集的第 i 个鉴别矢量 φ_i ($1 < i \leq r$) 可以由解下列问题计算得到:

$$\max_{\substack{\varphi_i^T \varphi_j = 0 \\ \|\varphi_i\| = 1}} (J(\varphi_i)) \quad j = 1, \dots, i-1. \quad (7)$$

设 $S = \{\varphi_i\}, i = 1, 2, \dots, r$ 。由 Foley-Sammon 最佳鉴别矢量集可以构成线性变换:

$$y = \Phi^T x, \quad (8)$$

其中 $\Phi = [\varphi_1 \varphi_2 \dots \varphi_r]$ 。

当 $r=2$ 时, Foley-Sammon 最佳鉴别矢量集等价于 Sammon 最佳鉴别平面。

首先求出 Fisher 最佳鉴别方向 φ_1 。在求出 r ($r \geq 1$) 个最佳鉴别方向 $\varphi_1, \varphi_2, \dots, \varphi_r$ 后, 第 $r+1$ 个最佳鉴别方向在满足共轭正交条件式 (9) 下, 使 Fisher 鉴别准则函数式 (6) 取得最大值的向量 φ_{r+1} :

$$\varphi_{r+1}^T S_t \varphi_i = 0 \quad (i = 1, 2, \dots, r). \quad (9)$$

关于具有统计不相关性的最佳鉴别方向的求解有如下的定理:

定理 (文献 [14, 15] 引理 2) 具有统计不相关性的最佳鉴别矢量集的第 $r+1$ 个最佳鉴别方向 φ_{r+1} 是下列广义本征方程中最大的本征值对应的本征向量

$$PS_b \varphi_{r+1} = \lambda S_w \varphi_{r+1}, \quad (10)$$

式中

$$P = I - S_t D^T (D S_t S_w^{-1} S_t D^T)^{-1} D S_t S_w^{-1}, \quad (11)$$

I 为单位矩阵, 而且,

$$D = [\varphi_1 \varphi_2 \dots \varphi_r]^T. \quad (12)$$

3 统计不相关最佳鉴别矢量集的本质

事实上, 若统计不相关性最佳鉴别矢量集为 $\varphi_1, \varphi_2, \dots, \varphi_n$, 则有

$$\varphi_i^T S_t \varphi_j = 0,$$

$$(i, j = 1, 2, \dots, n, i \neq j).$$

因此, 若 S_t 退化为单位矩阵时, 统计不相关性就退化为正交性, 那么, 能否找到一个线性变换使得 S_t 变换为单位矩阵呢? 这是下面将要解决的问题。

对 S_t 进行特征分解,

$$S_t = V D V^T.$$

由于 V 是正交矩阵, 所以 $V V^T = I$, 即 $V^T = V^{-1}$, 因此白化变换为

$$Q = V D^{-1/2} V^T.$$

样本 x 经白化变换后的样本为 \tilde{x} , 同样, 原始样本空间为 X , 变换后的空间为 \tilde{X} , 则在空间 \tilde{X} 中有

$$\tilde{x} = Qx.$$

同样,

$$\tilde{m}_0 = Qm_0,$$

$$\tilde{S}_t = E\{(\tilde{x} - \tilde{m}_0)(\tilde{x} - \tilde{m}_0)^T\} =$$

$$E\{(Qx - Qm_0)(Qx - Qm_0)^T\} =$$

$$QE\{(x - m_0)(x - m_0)^T\}Q^T =$$

$$QS_t Q^T = QVDV^{-1}Q^T =$$

$$VD^{-1/2}V^TVDV^{-1}(VD^{-1/2}V^T)^T =$$

$$VD^{-1/2}V^TVDV^{-1}VD^{-1/2}V^T =$$

$$VD^{-1/2}DD^{-1/2}V^T = VV^T = I,$$

$$\tilde{S}_b = QS_b Q^T,$$

$$\tilde{S}_w = QS_w Q^T.$$

这样, 用任何一种方法求得的 2 个最优鉴别矢量 φ_i 和 φ_j ($i \neq j$) 都是正交的, 即 $\varphi_i^T \varphi_j = 0$ 。在空间 \tilde{X} 中, 由于 $\tilde{S}_t = I$, 则有 $\varphi_i^T S_t \varphi_j = 0$, 即 φ_i 和 φ_j ($i \neq j$) 也是统计不相关的。

根据上述分析, 可以得到统计不相关最优鉴别矢量集的本质, 即统计不相关最优鉴别是白化变换加普通线性鉴别变换。

4 实验与结果分析

从 ORL 人脸图像库中分别取出若干个人的脸

部图像 (92×112), 每人 10 幅图像。计算中, 取每人的 4 幅图像训练, 其余 6 幅图像作为检验样本。分别用文献 [9]、文献 [10] 和文献 [13] 的方法以及它们的改进算法抽取最佳鉴别矢量。改进算法是指, 首先对原始数据在基于总体散布矩阵的白化变换后, 再用文献 [9]、文献 [10] 和文献 [13] 的方法进行特征抽取, 并在鉴别矢量空间构造最小距离分类器进行分类。图 1 为用于实验的部分人脸图像, 表 1 至表 3 为各种算法与它们相应的改进算法的实验结果比较, 可见改进的方法比原方法的识别性能要好, 但是由于改进方法需要先进行白化变换, 因此计算时间略有增加。

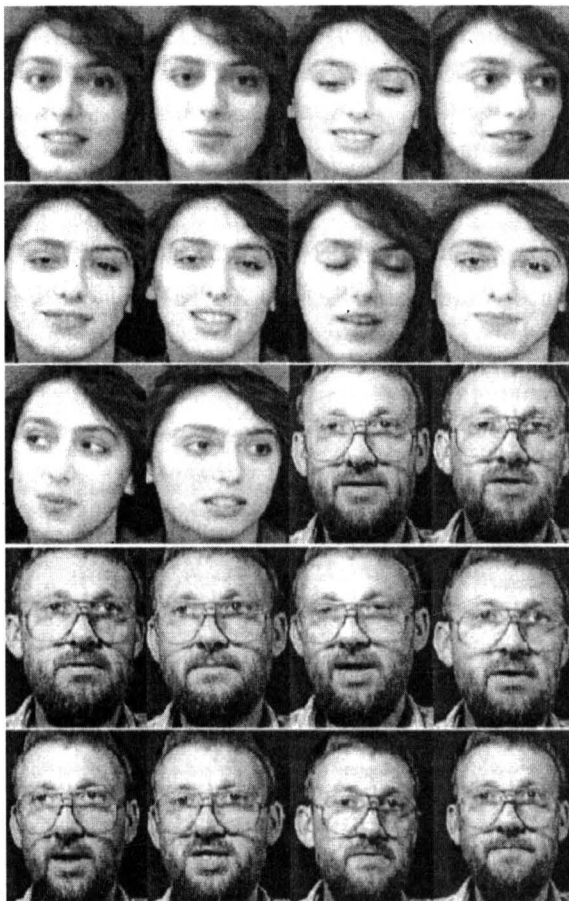


图 1 ORL 图像库部分图像

Fig.1 Part of the images of ORL

5 结论

笔者对统计不相关最佳鉴别矢量集的求解方法进行的研究, 获得如下结论:

1) 对统计不相关最佳鉴别矢量集的本质进行研究表明, 统计不相关最优鉴别是白化变换加普通

线性鉴别变换。这丰富了模式识别中有关特征提取的理论。

表 1 文献 [9] 的算法及其改进算法的性能比较

Table 1 Performance comparison of the algorithm in [9] and its improvement algorithm

类数	鉴别 矢量数	训练 样本数	文献[9]的方法		改进后的方法	
			错误	计算	错误	计算
			识别数	时间/s	识别数	时间/s
5	4	4	1	13.73	1	73.32
10	9	4	4	21.26	2	100.68
15	14	4	8	25.93	2	106.72
20	19	4	33	25.49	19	83.93
25	24	4	57	21.03	23	61.63
30	29	4	64	15.16	16	15.27
35	34	4	92	20.21	17	20.05
40	39	4	92	26.04	33	25.71

表 2 文献 [10] 的算法及其改进算法的性能比较

Table 2 Performance comparison of the algorithm in [10] and its improvement algorithm

类数	鉴别 矢量数	训练 样本数	文献[10]的方法		改进后的方法	
			错误	计算	错误	计算
			识别数	时间/s	识别数	时间/s
25	24	4	57	4.34	15	12.03
30	29	4	73	1.04	16	1.48
35	34	4	80	1.43	17	1.59
40	39	4	93	1.59	33	1.87

表 3 文献 [13] 的算法及其改进算法的性能比较

Table 3 Performance comparison of the algorithm in [13] and its improvement algorithm

类数	鉴别 矢量数	训练 样本数	文献[13]的方法		改进后的方法	
			错误	计算	错误	计算
			识别数	时间/s	识别数	时间/s
5	4	4	1	3.84	1	16.04
10	9	4	3	3.46	2	12.41
15	14	4	3	3.25	1	9.67
20	19	4	16	2.97	13	5.27
25	24	4	15	2.91	14	5.88
30	29	4	24	0.61	16	1.16
35	34	4	28	0.71	17	1.26
40	39	4	41	0.88	33	1.49

2) 该方法的最大优点在于所获得的最优鉴别矢量, 同时具有正交性和统计不相关性。不仅如此, 该方法对代数特征抽取具有普遍适用性。

3) 该方法不仅对人脸的特征提取非常有效, 而且对手写体数字识别、汉字识别以及基于内容的检索等模式识别领域的研究都有一定的意义。

参考文献

- [1] Turk M, Pentland A. Eigenfaces for face recognition [J]. *Cognitive Neuroscience*, 1991, 3 (1): 71~86
- [2] Foley D H, Sammon J W. An optimal set of discriminant vectors [J]. *IEEE Trans Computers*, 1975, 24 (3): 281~289
- [3] Duchene J, Leclercq S. An optimal transformation for discriminant and principal component analysis [J]. *IEEE Trans PAMI*, 1988, 10 (6): 978~983
- [4] Hong Z Q. Algebraic feature extraction of image for recognition [J]. *Pattern Recognition*, 1991, 24 (3): 211~219
- [5] 洪子泉, 杨静宇. 用于图像识别的图像代数特征抽取 [J]. *自动化学报*, 1992, 18 (2): 232~238
- [6] 洪子泉, 杨静宇. 基于奇异值特征和统计模型的人像识别算法 [J]. *计算机研究与发展*, 1994, 31 (3): 60~65
- [7] Cheng Y Q, Yang J Y, Liu K, et al. A novel feature extraction method for image recognition based on similar discriminant function [J]. *Pattern Recognition*, 1993, 26 (1): 115~125
- [8] Liu K, Cheng Y Q, Yang J Y, et al. An efficient algorithm for Foley - Sammon optimal set of discriminant vectors by algebraic method [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 1992, 6 (5): 817~829
- [9] Liu K, Cheng Y Q, Yang J Y, et al. A generalized optimal set of discriminant vectors [J]. *Pattern Recognition*, 1992, 25 (7): 731~739
- [10] 郭跃飞, 杨静宇. 求解广义最佳鉴别矢量的一种迭代算法及人脸识别 [J]. *计算机学报*. 2000, 23 (11): 1189~1195
- [11] 郭跃飞. 人脸图像代数特征提取与最佳鉴别矢量的研究 [D]. 南京: 南京理工大学, 2000
- [12] Guo Yuefei, Li Shijin, Yang Jingyu, et al. A generalized Foley - Sammon transform based on generalized fisher discriminant criterion and its application to face recognition [J]. *Pattern Recognition Letters*, 2003, 24 (1-3): 147~158
- [13] Wu Xiaojun, Yang Jingyu, Wang Shitong, et al. A new algorithm for solving the optimal discriminant vectors [J]. *Journal of Computer Science and Technology*, 2002, 17 (3): 324~330
- [14] 金忠. 人脸图像特征抽取与维数研究 [D]. 南京: 南京理工大学, 1999
- [15] 金忠, 杨静宇. 一种具有统计不相关性的最佳鉴别矢量集 [J]. *计算机学报*, 1999, 22 (10): 1105~1108

A Study on the Essence of Optimal Statistical Uncorrelated Discriminant Vectors

Wu Xiaojun^{1, 2, 3}, Yang Jingyu², Wang Shitong^{1, 2}, Liu Tongming¹, Josef Kittler⁴

(1. *East China Shipbuilding Institute, Zhenjiang, Jiangsu 212003, China*;

2. *School of Information, Nanjing University of Science & Technology, Nanjing 210094, China*;

3. *Shenyang Institute of Automation, Chinese Academy of Science, Shenyang 110015, China*;

4. *CVSSP, University of Surrey, Surrey GU2 7XH, UK*)

[Abstract] A study has been made on the essence of optimal set of uncorrelated discriminant vectors in this paper. A whitening transform has been constructed on the basis of the eigen decomposition of population scatter matrix, which makes the population scatter matrix an identity matrix in the transformed sample space. Thus, the optimal discriminant vectors solved by conventional LDA methods are statistical uncorrelated. The research indicates that the essence of the statistical uncorrelated discriminant transform is the whitening transform plus conventional linear discriminant transform. The distinguished characteristic of the proposed method is that the obtained optimal discriminant vectors are orthogonal and statistical uncorrelated. The proposed method suits for all the problems of algebraic feature extraction. The numerical experiments on facial database of ORL show the effectiveness of the proposed method.

[Key words] pattern recognition; feature extraction; discriminant analysis; generalized optimal set of discriminant vectors; face recognition