第4卷(A版) 第2期
1999年2月

中国图象图形学报
Journal of Image and Graphics

Vol. 4(A), No. 2
Feb. 1999

一种求解Fisher最佳鉴别矢量的新算法及人脸识别*

郭跃飞 黄修武 杨静宇
(南京理工大学计算机系，南京 210094)

摘 要 Fisher最佳鉴别矢量是高维模式分析中的有效方法。当训练样本数相对于特征空间的维数较小时，就成了小样本问题。为了求解小样本问题，人们提出了一系列方法并取得了良好的效果。但在类内距离为零的情况下已有的方法均得不到最佳解，该文从理论上说明了这一点，并给出了一种在任何情况下都能得到最佳解的新算法，试验结果证实了这样的推断。

关键词 模式分析 人脸识别 最佳鉴别矢量 最小距离分类器

0 引言

Fisher最佳鉴别矢量方法的基本思想是将原来高维的模式样本投影到最佳鉴别矢量空间以达到维数压缩的效果，投影后保证模式样本在新的空间中有最大的类间距离和最小的类内距离。即模式在该空间中有最佳的可分离性。因此它是一种有效的模式分析技术。这种方法的关键是如何求解最佳鉴别矢量。J. W. Sammon[1]提出了求最佳鉴别平面的技术。D. H. Foley与J. W. Sammon[2]进一步求出了最佳鉴别矢量集。求解最佳鉴别矢量可归结为求解下列广义特征值问题：$A\varphi=\lambda B\varphi$。其中$B$等于类内散布矩阵$S_w$。在大样本的情况下，$S_w$非奇异，故可精确求解。但在小样本情况下，$S_w$是奇异的，这种情况下不能直接求解。

为了在小样本集的情况下求得解，Q. Tian在文献[3]中提出用S_w的广义逆矩阵S_w^+代替S_w的逆矩阵S_w^-，Hong和Yang在文献[4]中提出在类内散布矩阵S_w中加入奇异值扰动，使S_w变为非奇异矩阵$\tilde{S}_w$。Cheng、Zhuang和Yang提出用矩阵秩分解的方法求解最佳鉴别矢量[5,6]，上述方法[3~6]都取得了一定成效，但求得的解都为近似解。Liu、Cheng和Yang给出了一种即使在小样本情况下也能求得精确解的方法[7,8]。但在类内距离为零的情况下用这种方法仍得不到最佳解。本文从理论上说明了这一点并提出了一个可在任意情况下求得最佳解的新算法。最后，我们用从人脸图象中提取的模式向量作了一系列的实验，证实了本文提出的方法的有效性。

1 基本方法

设S_w为类内散布矩阵，S_b为类间散布矩阵，$S_t=S_w+S_b$为总散布矩阵，Fisher鉴别准则为：

$$J(\alpha)=\frac{\alpha^{T}S_b\alpha}{\alpha^{T}S_w\alpha}=\sup_{\gamma}\frac{\gamma^{T}S_b\gamma}{\gamma^{T}S_w\gamma} \tag{1}$$

由文献[7]，Fisher鉴别准则与下列问题等价

$$\mathcal{J}(\alpha)=\frac{\alpha^{T}S_b\alpha}{\alpha^{T}S_t\alpha}=\sup_{\gamma}\frac{\gamma^{T}S_b\gamma}{\gamma^{T}S_t\gamma} \tag{2}$$

因为$\alpha^{T}S_w\alpha\geq0$，$\alpha^{T}S_b\alpha\geq0$，所以$0\leq\mathcal{J}(\alpha)\leq1$。即$\mathcal{J}(\alpha)$的绝对最大值是1，而$\mathcal{J}(\alpha)=1$等价于$\alpha^{T}S_w\alpha=0$且$\alpha^{T}S_b\alpha\neq0$，显然，当$\mathcal{J}(\alpha)=1$时，$\alpha^{T}S_b\alpha$越大可分性越好，文献[7]未单独考虑$\alpha^{T}S_b\alpha$的大小，因此得不到最佳解，基于此，我们设计了一个新的算法求解最佳鉴别矢量集。

定理1[7]设A为$n\times n$非负定矩阵，x为n维矢量。

* 本项研究获国家自然科学基金(No. 69062013)资助和博士点基金资助
收稿日期：1998-01-24；收到修改稿日期：1998-06-11

则，$x^TAx=0$ 当且仅当 $Ax=0$。

显然，S_w 与 S_b 都是非负定矩阵，我们有：

$S_w\alpha=0$ 当且仅当 $\alpha^TS_w\alpha=0$

$S_t\alpha=0$ 当且仅当 $\alpha^TS_t\alpha=0$

定理2　设 $S_t^{-1}(0)=\{\alpha|S_t\alpha=0\}$，$\overline{S_t^{-1}(0)}$为 $S_t^{-1}(0)$ 的正交补空间，则 $S_w\alpha=0$ 的解可在空间$\overline{S_t^{-1}(0)}$中选取。

证明：设 $S_w\alpha=0$　　令 $\alpha=\beta+\gamma$

其中 $\beta\in S_t^{-1}(0)$，$\gamma\in S_t^{-1}(0)$。

因为 $\beta\in S_t^{-1}(0)$，所以 $S_t\beta=0$

所以 $\beta^TS_t\beta=0$，

因此 $\beta^TS_w\beta\le\beta^TS_t\beta=0$

由定理1　　$S_w\beta=0$

所以 $S_w\alpha=S_w\gamma=0$

结论成立。

显然从$\overline{S_t^{-1}(0)}$中选取的 $S_w\alpha=0$ 的非零解一定满足 $\alpha^TS_b\alpha\neq0$

定理3　设 $J_b(\alpha)=\alpha^TS_b\alpha$，则使 $J_b(\alpha)$ 达到最大值的单位向量即为 S_b 的最大特征值对应的单位特征向量。

证明：求使 $J_b(\alpha)$ 达到最大值的单位向量相当于求解下列问题

$$J_b(\alpha)=\alpha^TS_b\alpha=\sup_{\gamma}\frac{\gamma^TS_b\gamma}{\|\gamma\|^2}$$

因为 $\dfrac{\gamma^TS_b\gamma}{\|\gamma\|^2}=\dfrac{\gamma^TS_b\gamma}{\gamma^TI\gamma}$

其中 I 为单位矩阵

所以 α 为 $I^{-1}S_b=S_b$ 的最大特征值对应的单位特征向量

定理4　设 $\beta_1,\cdots,\beta_k$ 为 n 维单位正交向量，$P=(\beta_1,\cdots,\beta_k)$，$Z$ 为 k 维向量，则 $\|PZ\|=\|Z\|$。

证明：

$$\|PZ\|^2=(PZ)^T(PZ)=Z^TP^TPZ=\|Z\|^2$$

所以 $\|PZ\|=\|Z\|$

由上述定理，我们可以用下列新方法计算最佳鉴别矢量的最佳解：

(1)　在$\overline{S_t^{-1}(0)}$中求取 $S_w\alpha=0$ 的解空间 V，并从中选取一组使 $\alpha^TS_b\alpha$ 达到最大值的标准正交基 $u_1,\cdots,u_k$，即为前 k 个最佳鉴别矢量。

(2)　在空间 $V+S_t^{-1}(0)$ 的正交补空间中选取其余的最佳鉴别矢量。

2　算　法

计算最佳鉴别矢量最佳解的具体方法：

2.1　计算使 $\tilde{J}(\alpha)=1$ 的最佳鉴别矢量

$$u_1,\cdots,u_k\quad(k\ge0)$$

情形1　S_t 非奇异

设：

$$S_w^{-1}(0)=\{\alpha|S_w\alpha=0\}=\mathrm{Span}\{\beta_1,\cdots,\beta_k\}$$

(若 $S_w\alpha=0$ 无非零解，则 $k=0$)

其中 $\beta_1,\cdots,\beta_k$ 为单位正交向量，我们可在此空间中选取一组标准正交基使类间距离 $\alpha^TS_b\alpha$ 达到最大。

(1)　计算第1个最佳鉴别矢量 u_1

设 $\alpha=\xi_1\beta_1+\cdots+\xi_k\beta_k=PZ$

其中 $P=(\beta_1,\cdots,\beta_k)$　　$Z=(\xi_1,\cdots,\xi_k)^T$

则 $\alpha^TS_b\alpha=Z^TP^TS_bPZ$

因此问题转化为求使 $Z^TP^TS_bPZ$ 达到最大的单位向量 Z。

由定理3　P^TS_bP 的最大特征值对应的特征向量 Z 即为所求。

由定理4　$\|PZ\|=1$ 等价于 $\|Z\|=1$

所以　$u_1=PZ$

(2)　计算第 i 个最佳鉴别矢量 u_i　　$(i<k)$

设已求得 $i-1$ 个最佳鉴别矢量 $u_1,\cdots,u_{i-1}$

$\overline{S_w^{-1}(0)}=\mathrm{Span}\{\varphi_1,\cdots,\varphi_{n-k}\}$为 $S_w^{-1}(0)$ 的正交补空间

$\overline{V}_i=\mathrm{Span}\{\varphi_1,\cdots,\varphi_{n-k},u_1,\cdots,u_{i-1}\}$

其中 $\varphi_1,\cdots,\varphi_{n-k}$为单位正交向量。

因为 $u_1,\cdots,u_{i-1}\in S_w^{-1}(0)$

所以 $\varphi_1,\cdots,\varphi_{n-k},u_1,\cdots,u_{i-1}$也相互正交。

令 $V_i=\mathrm{Span}\{\theta_1,\cdots,\theta_{k-i+1}\}$为 $\overline{V}_i$ 的正交补空间，其中 $\theta_1,\cdots,\theta_{k-i+1}$为单位正交向量，则 $V_i\subset S_w^{-1}(0)$

u_i 可在 V_i 中选取，与前面类似

设 $P_i=(\theta_1,\cdots,\theta_{k-i+1})$

求 $P_i^TS_bP_i$ 的最大特征值对应的单位特征向量 Z_i，则 $u_i=P_iZ_i$

(3)　计算第 k 个最佳鉴别矢量 u_k

由上一步 $V_k=\mathrm{Span}\{\theta_1\}$

θ_1 即为所求，即 $u_k=\theta_1$

情形2　S_t 奇异

设 $S_t^{-1}(0)=\mathrm{Span}\{\gamma_1,\cdots,\gamma_{n_1}\}$

$\overline{S_t^{-1}(0)}=\mathrm{Span}\{\varphi_1,\cdots,\varphi_{n-n_1}\}$

其中 $\gamma_1,\cdots,\gamma_{n_1}$ 与 $\varphi_1,\cdots,\varphi_{n-n_1}$ 均为单位正交向量组。

由定理2　$S_w\alpha=0$ 的解可在$\overline{S_t^{-1}(0)}$中选取。

设 $P=(\varphi_1,\cdots,\varphi_{n-n_1})Z=(\xi_1,\cdots,\xi_{n-n_1})^{\mathrm{T}}$

$\alpha=PZ$　　则 $S_w\alpha=0$ 等价于

$S_wPZ=0$　　解方程组 $S_wPZ=0$

得基础解系 $Z_1,\cdots,Z_k$

(若 $S_wPZ=0$ 无非零解，则 $k=0$)

$\alpha_1=PZ_1,\cdots,\alpha_k=PZ_k$

将 $\alpha_1,\cdots,\alpha_k$ 标准正交化得 $\beta_1,\cdots,\beta_k$

令 $V=\mathrm{Span}\{\beta_1,\cdots,\beta_k\}$

则 $V=S_t^{-1}(0)\cap S_w^{-1}(0)$

可在 V 中选取一组标准正交基使 $\alpha^{\mathrm{T}}S_b\alpha$ 达到最大值，这组基即为所求的最佳鉴别矢量 $u_1,\cdots,u_k$，以下的计算过程类似于情形 1。

2.2　计算第 K+1 个最佳鉴别矢量 u_{k+1}

情形 1　S_t 非奇异

设 $V=\mathrm{Span}\{u_1,\cdots,u_k\}$

$\bar{V}=\mathrm{Span}\{\varphi_1,\cdots,\varphi_{n-k}\}$ 为 V 的正交补空间。其中 $\varphi_1,\cdots,\varphi_{n-k}$ 为单位正交向量，我们可在 $\bar{V}$ 中选取 u_{k+1}

设　$\alpha=\xi_1\varphi_1+\cdots+\xi_{n-k}\varphi_{n-k}=PZ$

其中 $P=(\varphi_1,\cdots,\varphi_{n-k})$，$Z=(\xi_1,\cdots,\xi_{n-k})^{\mathrm{T}}$

则　$$\frac{\alpha^{\mathrm{T}}S_b\alpha}{\alpha^{\mathrm{T}}S_t\alpha}=\frac{Z^{\mathrm{T}}(P^{\mathrm{T}}S_bP)Z}{Z^{\mathrm{T}}(P^{\mathrm{T}}S_tP)Z}$$

求 $(P^{\mathrm{T}}S_tP)^{-1}(P^{\mathrm{T}}S_bP)^{-1}$ 的最大特征值对应的特征向量 Z

则　$$u_{k+1}=\frac{PZ}{\|PZ\|}$$

情形 2　S_t 奇异

因为 $u_1,\cdots,u_k\in\overline{S_t^{-1}(0)}$

所以 $u_1,\cdots,u_k,\gamma_1,\cdots,\gamma_{n_1}$ 为单位正交向量组（其中 $S_t^{-1}(0)=\mathrm{Span}\{\gamma_1,\cdots,\gamma_{n_1}\}$）

令 $V=\mathrm{Span}\{u_1,\cdots,u_k,\gamma_1,\cdots,\gamma_{n_1}\}$

$\bar{V}=\mathrm{Span}\{\varphi_1,\cdots,\varphi_{n-n_1-k}\}$ 为 V 的正交补空间。我们可在 $\bar{V}$ 中选取 u_{k+1}

令 $P=(\varphi_1,\cdots,\varphi_{n-n_1-k})$

求 $(P^{\mathrm{T}}S_tP)^{-1}(P^{\mathrm{T}}S_bP)$ 的最大特征值对应的特征向量 Z，则 $u_{k+1}=\dfrac{PZ}{\|PZ\|}$

2.3　计算第 k+i 个最佳鉴别矢量

计算过程类似于 2.2。

3　实验结果

为了验证本算法的有效性，我们从南京理工大学计算机系模式识别与智能控制实验室建立的 FDB603 图象库中提取了 8 个人的 64 幅脸部图象（每人 8 幅），图 1 为 8 个人的图象。将它们处理成 32×32大小的图象，用文献[9]中的方法抽取每个图象的特征（为 32 维模式向量）。共得到 8×8 个模式向量，对这些数据我们分别用文献[7]中的方法与本文中的方法作了一系列的实验进行对比。首先，我们将每类的前 4 个样本作为训练样本，用两种方法分别计算第一个最佳鉴别矢量，结果如下，其中，前者为用文献[7]中的方法得到的结果，后者为用本文中的方法得到的结果，$Db(\alpha)$ 与 $Dw(\alpha)$ 分别为模式向量都投影到向量 α 上后得到的类间距离与类内距离。

(1) $\alpha=\{-0.03\ 0.04\ -0.15\ 0.44\ 0.04\ 0.02\ -0.33\ -0.20\ -0.12\ 0.07\ -0.07\ 0.27\ -0.24\ -0.01\ 0.26\ -0.07\ -0.04\ 0.32\ -0.02\ -0.19\ 0.11\ -0.09\ -0.25\ 0.16\ 0.18\ -0.12\ -0.02\ -0.02\ 0.03\ -0.09\ -0.27\ 0.17\}$

$Db(\alpha)=19\,880.71$　　$Dw(\alpha)=0.00$

(2) $\beta=\{-0.03\ 0.01\ 0.05\ 0.26\ 0.15\ 0.14\ -0.41\ -0.14\ -0.23\ 0.09\ -0.05\ -0.06\ -0.03\ 0.21\ 0.10\ -0.02\ -0.05\ 0.07\ -0.10\ 0.08\ 0.11\ -0.33\ 0.08\ 0.07\ 0.18\ -0.12\ -0.02\ -0.06\ -0.05\ -0.06\ -0.29\ 0.54\}$

$Db(\beta)=25\,727.19$　　$Dw(\beta)=0.00$

从上面可以看到，用两种方法求得的 2 个最佳鉴别矢量对应的类内距离均为 0，但用本文中的方法得到的最佳鉴别矢量对应的类间距离比用文献[7]中的方法得到的明显大。

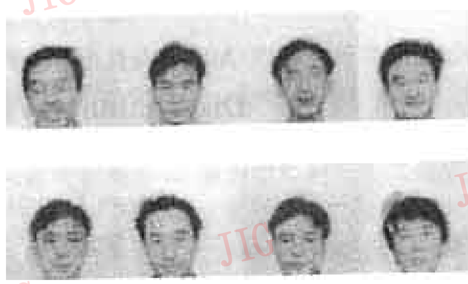


图 1

表 1 给出了在多种情况下分别用两种方法得到的识别结果。从表 1 中可以看出，在所有情况下，用本文中的方法得到的识别率都高于或等于文献[7]中的方法所得到的识别率。

表 1

每类的训练样本数	鉴别向量数	识别错误数		识别准确率	
		老方法	新方法	老方法	新方法
3	3	15	14	76.6%	78.1%
3	4	8	7	87.5%	89.1%
3	5	6	6	90.6%	90.6%
4	2	17	9	73.4%	85.9%
4	3	10	7	84.4%	89.1%
4	4	9	3	85.9%	95.3%
4	5	6	3	90.6%	95.3%

参考文献

1 Sammon J W. An Optimal discriminant plane. IEEE Trans. Computer, 1970, C-19: 826～829.

2 Foley D H, Sammon J W. An optimal set of discriminant vectors. IEEE Trans. Computer, 1975, C-24(3): 281～289.

3 Tian Q. Comparison of statistical pattern-recognition algorithms for hybrid processing, II: eigenvector-based algorithm, J. Opt. Soc. Am. A5, 1988, 1670～1672.

4 Hong Zi quan, Yang Jing yu. Optimal discriminant plane for a small number of samples and design method of classifier on the plane. Pattern Recognition, 1991, 24(4): 314～317.

5 Cheng Y Q, Zhuang Y M, Yang J Y. Optimal Fisher discriminant analysis using the rank decomposition. Pattern Recognition, 1992, 25(1): 101～111.

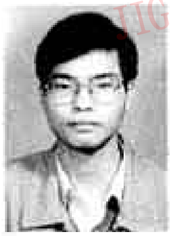
6 陈永清,庄永明,杨静宇. 一种改进的 Fisher 判别准则. 计算机研究与发展, 1991(6).

7 Liu K, Cheng Y Q, Yang J Y. An efficient algorithm for Foley-Sammon optimal set of discriminant vectors by algebraic method. International Journal of Pattern Recognition and Artificial intelligence, 1982, 6(5): 817～829.

8 Liu K, Cheng Y Q, Yang J Y. A generalized optimal set of discriminant vectors. Pattern Recognition, 1992, 25(1): 731～739.

9 郭跃飞,姜志华,杨静宇等. 一种新的代数特征抽取方法及人脸识别. 南京:南京理工大学学报, 1997(5).

黄修武 1986年获华东师范大学数学专业学位,1993年获南京理工大学计算机应用专业硕士学位并留校任教,1998年获南京理工大学计算机系模式识别与智能系统专业博士。主要研究方向为计算机视觉、模式识别和人工智能。



郭跃飞 1985年毕业于华东师范大学数学系获学士学位,1993年毕业于南京理工大学计算机系获硕士学位,现为南京理工大学计算机系模式识别与智能控制专业博士生。主要研究方向为计算机视觉、模式识别和人工智能。

杨静宇 1965年毕业于炮兵工程学院,现任南京理工大学信息学院院长,教授,博士生导师,现主要研究方向为计算机视觉、图象处理、模式识别和人工智能。

A Novel Algorithm Solving Fisher Optimal Discriminant Vector and Facial Recognition

Guo Yuefei, Huang Xiouwu, Yang Jingyu

(*Department of Computer, NUST, Nanjing* 210094)

Abstract Fisher optimal discriminant vectors method is an effective method in pattern analysis of high dimension. When the number of the training samples is small compared with the dimensionality of the feature space, the problem becomes the case of a small number of samples. People have proposed many methods for solving the problem of a small number of samples, and made great progress. However, none of the previous methods can obtain the optimal solution when within－class distance equals zero. This point is illustrated in terms of theory, and a new solving method that can obtain the optimal solution in any situations is presented in this paper. The experimental result confirms our inference.

Keywords Pattern analysis, Facial recognition, Optimal discriminant vectou, Minimum distance classifier