第 27 卷第 1 期　　电 子 与 信 息 学 报　　Vol.27No.1
2005 年 1 月　　Journal of Electronics & Information Technology　　Jan. 2005

改进的统计不相关最优鉴别矢量集

吴小俊[*,**,***]　杨静宇[****]　王士同[****]　Josef Kittler[***]　陆介平[*]

[*](江苏科技大学计算机系　镇江 212003)

[**](中国科学院沈阳自动化所机器人学重点实验室 沈阳 110015)

[***](CVSSP, Dept. of Electrical Engineering, University of Surrey GU2 7XH,UK)

[****](南京理工大学信息学院 南京 210094)

摘　要：该文对统计不相关最优鉴别矢量集算法进行研究，在分析统计不相关最优鉴别矢量集算法的基础上提出了一种改进的方法。该方法在类内散布矩阵的特征空间中求解统计不相关最优鉴别矢量集。为了加快特征抽取速度，利用基于图像鉴别分析的维数压缩方法，对图像数据进行了压缩。在 ORL 和 Yale 人脸数据库的数值实验，验证本文所提出的方法的有效性。

关键词：模式识别，特征抽取，鉴别分析，最佳鉴别矢量集，人脸识别

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1009-5896(2005)01-0047-04

An Improved Optimal Set of Statistical Uncorrelated Discriminant Vectors

Wu Xiao-Jun[*,**,***]　Yang Jing-Yu[****]　Wang Shi-Tong[****]　Josef Kittler[***]　Lu Jie-ping[*]

[*](*Dept of Computer Science, Jiangsu Univ. of Sci. and Tech., Zhenjiang* 212003, *China*)

[**](*Shenyang Institute of Automation, Chinese Academy of Sciences*, Shenyang 110015, *China*)

[***](*CVSSP, Dept of Electrical Engineering, University of Surrey GU2 7XH,UK*)

[****](*School of Information, Nanjing University of Science & Technology, Nanjing* 210094, *China*)

Abstract　This paper presents a research on the algorithm of optimal set of statistically uncorrelated discriminant vectors. An improved algorithm has been proposed on the basis of the analysis of the conventional algorithm of statistical uncorrelated discriminant vectors, which solves the optimal set of statistically uncorrelated discriminant vectors in the eigen space of the within-class scatter matrix S_w. The dimension of images has been reduced using the dimension reduction method based on image discriminant analysis in order to speed the process of feature extraction. The numerical experiments on facial databases of ORL and Yale show the effectiveness of the proposed method.

Key words　Pattern recognition, Feature extraction, Discriminant analysis, Optimal set of discriminant vectors, Face recognition

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]，Kittler 提出了基于 Fisher 鉴别准则的特征选择方法[4]，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[5,6]，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]，我们最近又提

2003-05-28 收到，2004-03-09 改回
国家自然科学基金(60072034)，中国科学院沈阳自动化研究所机器人学重点实验室基金(RL200108)，江苏省高校自然科学研究计划项目(01KJB52002)，江苏省自然科学基金(BK2002001，BK2004058)，图像处理与通信实验室开放基金(KJS03038)和校博士(后)基金资助课题

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，因此，统计不相关最优鉴别矢量集算法引起了人们的关注[14-18]，金忠提出了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]，文献[16]提出了广义统计不相关最优鉴别准则和广义统计不相关最优鉴别矢量集算法，文献[17]给出了一种等价形式的统计不相关最优鉴别矢量集算法，文献[18]给出了基于广义DKL变换的统计不相关最优鉴别矢量集算法。本文第2节介绍统计不相关最佳鉴别准则和统计不相关最优鉴别矢量集，第3节给出改进的统计不相关最佳鉴别矢量集，为了加快特征抽取速度，第4节给出基于图像鉴别分析的维数压缩方法，第5节给出实验结果及比较，最后给出结论。

2 统计不相关最优鉴别准则与统计不相关最优鉴别矢量集

定义1[15] 如果两个随机变量ξ和η的协方差满足：

$$E[\xi-E\xi][\eta-E\eta]=0 \tag{1}$$

则称ξ和η是统计不相关的。

设$w_1,w_2,\cdots,w_m$为m个模式类，$X=\{x_i\}$，$i=1,2,\cdots,N$为n维训练样本集，X中的每一个x_i属于w_j类，即$x_i\in w_j$，$i=1,2,\cdots,N$，$j=1,2,\cdots,m$。设w_i类的的平均矢量、协方差矩阵和先验概率分别为$\boldsymbol{m}_i,\boldsymbol{c}_i,P(w_i)$，类间散布矩阵为$\boldsymbol{S}_b$，类内散布矩阵为$\boldsymbol{S}_w$和总散布矩阵为$\boldsymbol{S}_t$。Fisher鉴别函数可定义为

$$J(\boldsymbol{\varphi})=\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{\varphi}/(\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}) \tag{2}$$

其中$\boldsymbol{\varphi}$为任一n维列矢量。使函数$J(\boldsymbol{\varphi})$达到最大值的矢量$\boldsymbol{\varphi}_1^*$为Fisher最佳鉴别方向，训练样本在方向$\boldsymbol{\varphi}_1^*$上的投影集在一维子空间$\mathrm{Span}\{\boldsymbol{\varphi}_1^*\}$中有最小的类内距离和最大的类间距离。

引理1[15] 样本矢量投影到两个最佳鉴别矢量$\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j$得到的两个分量统计不相关的充要条件为$\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{S}_t\boldsymbol{\varphi}_j=0$，则称$\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j$为关于$\boldsymbol{S}_t$共轭正交。

定义2[15] 若两个最佳鉴别矢量$\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j$满足$\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{S}_t\boldsymbol{\varphi}_j=0$，即$\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j$关于$\boldsymbol{S}_t$共轭正交，则称$\boldsymbol{\varphi}_i,\boldsymbol{\varphi}_j$为统计不相关的。

首先求出Fisher最佳鉴别方向$\boldsymbol{\varphi}_1$。在求出$\mathrm{r}\,(r\ge1)$个最佳鉴别方向$\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_r$后，第r+1个最佳鉴别方向在满足共轭正交条件式(3)下的使Fisher鉴别准则函数式(1)取到最大值的向量$\boldsymbol{\varphi}_{r+1}$：

$$\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{S}_t\boldsymbol{\varphi}_i=0,\qquad i=1,2,\cdots,r \tag{3}$$

引理2[14,15] 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$\boldsymbol{\varphi}_{r+1}$是下列广义特征方程中最大的特征值对应的特征向量：

$$\boldsymbol{P}\boldsymbol{S}_b\boldsymbol{\varphi}_{r+1}=\lambda\boldsymbol{S}_w\boldsymbol{\varphi}_{r+1} \tag{4}$$

式中

$$\boldsymbol{P}=\boldsymbol{I}-\boldsymbol{S}_t\boldsymbol{D}^{\mathrm{T}}\left(\boldsymbol{D}\boldsymbol{S}_t\boldsymbol{S}_w^{-1}\boldsymbol{S}_t\boldsymbol{D}^{\mathrm{T}}\right)^{-1}\boldsymbol{D}\boldsymbol{S}_t\boldsymbol{S}_w^{-1} \tag{5}$$

这里，$\boldsymbol{I}$为单位矩阵。

$$\boldsymbol{D}=[\boldsymbol{\varphi}_1\ \boldsymbol{\varphi}_2\cdots\boldsymbol{\varphi}_r]^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{\varphi}_1^{\mathrm{T}}\\ \boldsymbol{\varphi}_2^{\mathrm{T}}\\ \vdots\\ \boldsymbol{\varphi}_r^{\mathrm{T}}\end{bmatrix} \tag{6}$$

3 改进的统计不相关最优鉴别矢量集

定理1 $\boldsymbol{\varphi}$是$\boldsymbol{S}_w$的对应于特征值为k的特征矢量，当且仅当下式成立:

$$\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}=k\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\varphi} \tag{7}$$

证明 （1）必要性

设$\boldsymbol{\varphi}$是$\boldsymbol{S}_w$的对应于特征值为k的特征矢量，则$\boldsymbol{S}_w\boldsymbol{\varphi}=k\boldsymbol{\varphi}$。

在上式两边同时左乘$\boldsymbol{\varphi}^{\mathrm{T}}$，则有$\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}=k\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\varphi}$。

（2）充分性（反证法）

设$\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}=k\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\varphi}$，又假设$\boldsymbol{\varphi}$不是$\boldsymbol{S}_w$的对应于特征值为$k$的特征矢量，则有$\boldsymbol{S}_w\boldsymbol{\varphi}\ne k\boldsymbol{\varphi}$。在上式两边同时左乘$\boldsymbol{\varphi}^{\mathrm{T}}$，则有$\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}\ne k\boldsymbol{\varphi}^{\mathrm{T}}\boldsymbol{\varphi}$。但这与前提矛盾，故假设不真，所以$\boldsymbol{\varphi}$是$\boldsymbol{S}_w$的对应于特征值为$k$的特征矢量。

在文献[15]中，假设前r个最优鉴别矢量$\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_r$已经求出后，求第r+1个最优鉴别矢量$\boldsymbol{\varphi}_{r+1}$时，一方面使得最优鉴别矢量$\boldsymbol{\varphi}_{r+1}$满足归一化条件，即

$$\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{\varphi}_{r+1}=1 \tag{8}$$

另一方面又让$\boldsymbol{\varphi}_{r+1}$满足如下条件:

$$\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}_{r+1}=1 \tag{9}$$

所以$\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}_{r+1}=\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{\varphi}_{r+1}$，则$\boldsymbol{\varphi}_{r+1}$应是$\boldsymbol{S}_w$的对应于特征值为1的特征矢量。

但一般来说，很难保证$\boldsymbol{S}_w$存在特征值1。这说明文献[15]式(9)的条件过于苛刻。故不失一般性，假设$\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}_{r+1}=k$，因$\boldsymbol{S}_w$为非负定矩阵,所以$k\ge0$,故有$\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{\varphi}_{r+1}=k\boldsymbol{\varphi}_{r+1}^{\mathrm{T}}\boldsymbol{\varphi}_{r+1}$。由定理1，上式当且仅当$\boldsymbol{\varphi}_{r+1}$是$\boldsymbol{S}_w$的对应于特征值为$k$的特征矢量时成立，即$\boldsymbol{S}_w\boldsymbol{\varphi}_{r+1}=k\boldsymbol{\varphi}_{r+1}$。上述结果表明，$\boldsymbol{\varphi}_{r+1}$可在$\boldsymbol{S}_w$的特征空间中求得。

令$\boldsymbol{T}=[t_1,t_2,\cdots,t_n]$，其中$t_i(i=1,\cdots,n)$是$\boldsymbol{S}_w$的特征矢量。为了在$\boldsymbol{S}_w$的特征空间中求$\boldsymbol{\varphi}_{r+1}$，将训练样本$\boldsymbol{X}$向$\boldsymbol{T}$投影，故有

$$\boldsymbol{Y}=\boldsymbol{T}^{\mathrm{T}}\boldsymbol{X} \tag{10}$$

在变换后的空间$\boldsymbol{Y}$中的类内散布矩阵、类间散布矩阵和总体散布矩阵分别为

$$\tilde{\boldsymbol{S}}_w=\boldsymbol{T}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{T} \tag{11}$$

$$\tilde{S}_b = T^{\mathrm{T}} S_b T \tag{12}$$

$$\tilde{S}_t = T^{\mathrm{T}} S_t T \tag{13}$$

仿文献[15]可得 φ_{r+1} 是下列方程中最大特征值对应的特征矢量：

$$P\tilde{S}_b \varphi_{r+1} = \lambda \tilde{S}_w \varphi_{r+1} \tag{14}$$

其中，

$$P = I - \tilde{S}_t D^{\mathrm{T}} (D\tilde{S}_t \tilde{S}_w^{-1} \tilde{S}_t D^{\mathrm{T}})^{-1} D\tilde{S}_t \tilde{S}_w^{-1} \tag{15}$$

4　基于图像鉴别分析的维数压缩方法

$A_j^{(i)}$ 为训练图像样本，其中 $i=1,2,\cdots,C$,　$j=1,2,\cdots,M_i$，C 为总的类别数，M_i 为第 i 类 ω_i 的总训练图像样本数，$A_j^{(i)} \in R^{m\times n}$，第 i 类训练图像的平均图像为 $\overline{A^{(i)}}$：

$$\overline{A^{(i)}} = \frac{1}{M_i}\sum_{j=1}^{M_i} A_j^{(i)} \tag{16}$$

所有 C 类训练图像的平均图像为 $\overline{A} = \sum_{i=1}^{C} P_i \overline{A^{(i)}}$，其中 P_i，$i=1,2,\cdots,C$ 是第 i 类的先验概率。

设 $D_b = \sum_{i=1}^{C} P_i \left(\overline{A^{(i)}} - \overline{A}\right)^{\mathrm{T}} \left(\overline{A^{(i)}} - \overline{A}\right)$，

$D_w = \sum_{i=1}^{C} P_i \frac{1}{M_i} \sum_{l=1}^{M_i} \left(A_l^{(i)} - \overline{A^{(i)}}\right)^{\mathrm{T}} \left(A_l^{(i)} - \overline{A^{(i)}}\right)$，则

$$db_x = X^{\mathrm{T}} D_b X \tag{17}$$

$$dw_x = X^{\mathrm{T}} D_w X \tag{18}$$

我们称 D_b，D_w 及 $D_t = D_b + D_w$ 分别为训练图像样本的广义类间散布矩阵、广义类内散布矩阵与广义总散布矩阵。因此，基于图像的 Fisher 准则函数为

$$J(X) = X^{\mathrm{T}} D_b X / X^{\mathrm{T}} D_w X \tag{19}$$

其中 X 为任一 n 维列矢量，使函数 $J(X)$ 达到最大的矢量 u 称为最佳鉴别投影矢量。

最佳鉴别投影矢量 u 的物理意义是训练图像样本在方向 u 上的投影特征矢量在 m 维空间 R^m 中具有最小的类内散布，与最大的类间散布，即具有最大的可分性。

求出最佳鉴别投影矢量 u 后，将训练样本图像向 u 投影，得 $Y_j^{(i)} = A_j^{(i)} u$，$i=1,2,\cdots,K$,　$j=1,2,\cdots,M_i$，这样，对任一训练样本由原来的 $m\times n$ 维压缩成 m 维矢量。

5　实验结果

为了测试本文提出的方法的有效性，我们分别在 ORL 和 Yale 大学两个图像库上进行人脸识别实验。在两个实验中，我们首先用基于图像鉴别分析的维数压缩方法，将原来 $m\times n$ 的图像压缩成 m 维矢量。然后用本文方法和文献[15]的方法抽取统计不相关最优鉴别矢量，形成特征空间，在特征空间中设计最小距离分类器进行模式分类。

（1）ORL 上的实验结果

ORL 库由 40 人的脸部图像组成，每人 10 幅 92×112 的图像，其中有些图像拍摄于不同的时期；人脸脸部表情和脸部细节有变化，如笑或不笑、眼睛睁着或闭着、戴或不戴眼镜；人脸姿态有变化，深度旋转与平面旋转可达到 20°；人脸的尺度也有最多 10%的变化。从 ORL 人脸图像库中分别取出若干个人的脸部图像（92×112），计算中取每人的 4 幅图像训练，全部图像作为检验样本。表 1 给出了在不同大小的图像下两种方法的识别率的比较。从表 1 可以看出本文方法的识别率比文[15]的方法的识别率高。

表 1　ORL 上的识别率比较

图像大小	文献[15] (%)	本文的方法 (%)
112×92	72.50	89.25
56×46	86.75	90.25
39×30	86.75	92.25
28×23	88.25	92.50
14×12	87.50	88.75
7×6	75.00	76.25
平均识别率	82.79	88.21

（2）Yale 大学人脸图像库上的实验结果

在 Yale 大学人脸图像库中，由 15 人的脸部图像组成，每人 11 幅图像，计 165 幅图像，每幅图像的分辨率为 121×160。每人的每幅图像均是从不同的视角获得的，而且有较大的表情变化、光照条件变化以及部分缺损。从 Yale 大学人脸图像库中分别取出若干个人的脸部图像（121×160），计算中任意取每人的 5 幅图像训练，全部图像作为检验样本。表 2 给出了两种算法在该图像库上 15 个人计 165 幅图像的识别性能的比较，包括识别率与识别时间的比较。

表 2　Yale 上的识别性能比较

图像大小	文献[15] (%)		本文的方法	
	识别率(%)	时间 (s)	识别率(%)	时间 (s)
160×121	69.70	6.75	78.79	6.21
80×60	65.45	2.97	67.88	3.08
40×30	86.06	2.36	85.45	2.47
20×15	84.85	2.47	81.21	2.30
10×7	76.97	2.20	80.00	2.25
5×4	76.97	2.14	78.79	2.19
平均识别率和时间	76.67	3.15	78.69	3.08

6　结论

本文对统计不相关最优鉴别矢量集算法进行研究，获得如下结论：(1)在分析文献[15]的统计不相关最优鉴别矢量集算法的基础上，提出了一种改进的求解统计不相关最优鉴别矢量集的算法。(2)利用图像鉴别分析方法能有效地压缩图像

数据的维数。(3)在ORL和Yale人脸数据库的数值实验中，验证了本文所提出方法的有效性。(4)该算法对手写体数字识别、汉字识别等模式识别领域的研究也有一定的意义。

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, *IEEE Trans. on Computers*, 1975, 24(3): 281 – 289.

[3] Duchene J, Leclercq S. An optimal transformation for discriminant and principal component analysis. *IEEE Trans. on PAMI*, 1988, 10(6): 978 – 983.

[4] Kittler J. On the discriminant vector method of feature selection. *IEEE Trans. on Computers*, 1977, 26(6): 604 – 606.

[5] Hong Z Q. Algebraic feature extraction of image for recognition. *Pattern Recognition*, 1991, 24(3): 211 – 219.

[6] 洪子泉, 杨静宇. 用于图像识别的图像代数特征抽取. 自动化学报, 1992, 18(2): 232 – 238.

[7] Cheng Y Q, Yang J Y, *et al.*. A novel feature extraction method for image recognition based on similar discriminant function. *Pattern Recognition*, 1993, 26(1): 115 – 125.

[8] Liu K, Cheng Y Q,Yang J Y, *et al.*. Algebraic feature extraction for image recognition based on an optimal discriminant criterion. *Pattern Recognition*, 1993, 26(6): 903 – 911.

[9] Liu K, Yang J Y, *et al.*. A generalized optimal set of discriminant vectors. *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. 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. [博士论文], 南京: 南京理工大学, 2000.

[12] Guo Yue-Fei, Yang Jing-Yu, *et al.*. Feature extraction method based on the generalized fisher discriminant criterion and facial recognition. *Pattern Analysis & Application*, 2001, 4(1): 61 – 66.

[13] 吴小俊, 杨静宇, 王士同等. A new algorithm for solving optimal discriminant vectors. *Journal of Computer Science and Technology*, 2001, 3(17): 324 – 331.

[14] 金忠. 人脸图像特征抽取与维数研究. [博士论文], 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. 计算机学报, 1999, 22(10): 1105 – 1108.

[16] 吴小俊. 图像特征抽取与识别理论及其在人脸识别中的应用研究. [博士论文], 南京: 南京理工大学, 2002.

[17] 吴小俊, 杨静宇, 王士同等. A new optimal set of uncorrelated discriminant vectors and its application. Proceedings of WCICA'02 (IEEE 02EX527), Shanghai, China, 2002: 340 – 344.

[18] 吴小俊, 杨静宇, 王士同等. 广义 DKL 变换及其在人脸识别中的应用研究. 计算机科学, 2003, 30(1): 87 – 89.

吴小俊： 男，1967年生，博士，副教授，主要从事神经网络、模式识别和人工智能的研究.

杨静宇： 男，1941年生，教授，博士生导师，主要研究领域为计算机视觉、信息融合、模式识别和智能机器人等.

王士同： 男，1964年生，教授，博士生导师，主要从事神经网络、模糊系统和模糊人工智能的研究.

Josef Kittler： 男，1946年生，博士，教授，模式识别与计算机视觉领域国际知名学者，国际模式识别协会(IAPR)前主席，英国皇家工程院院士

陆介平： 男，1963年生，教授，主要从事模式识别和人工智能的研究.