

一种新的基于 Fisher 准则的线性特征提取方法

黄国宏¹, 刘 刚²

(1. 广东工业大学 信息工程学院, 广东 广州 510006;

2. 上海电力学院 电力与自动化工程学院, 上海 200090)

摘 要:针对现有的基于 Fisher 准则的线性特征提取方法存在的不足,提出了一种新的加权 Fisher 特征提取方法。该方法通过引入一个加权函数来削弱边缘类别的影响,减少投影空间中相邻类别间的重叠,提高了识别正确率。针对小样本问题,也给出了该算法的一个可行的最优判别矢量集的求解方法。分别对 COIL 图像数据库以及 ORL 人脸数据库进行实验,结果表明,就识别率而言,该方法得到的最优判别矢量具有更好的特征提取能力。

关键词:特征提取;小样本问题;Fisher 准则;人脸识别

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2008)05-0227-04

A New Linear Feature Extraction Method Based on Fisher Criterion

HUANG Guo-hong¹, LIU Gang²

(1. Sch. of Information & Eng., Guangdong University of Technology, Guangzhou 510006, China;

2. Sch. of Electric Power and Automation Eng., Shanghai University of Electric Power, Shanghai 200090, China)

Abstract: A novel weighted Fisher discriminant analysis is developed in this paper. The approach aims at overcoming the drawback of the previous FDA-based methods by introducing a weighting function to weaken the influence of outlier classes and reduce the large overlapping of neighboring classes in projection space. And a feasible solution to this new approach is suggested. Experimental results on the COIL and ORL database show that new optimal discriminant vectors have more powerful ability of feature extraction in terms of rates of classification.

Key words: feature extraction; small sample problem; Fisher criterion; face recognition

0 引言

众所周知,基于 Fisher 准则的判别分析已被公认为是最有效的线性特征提取方法之一。到目前为止,众多学者已经给出了许多基于 Fisher 准则的线性特征提取方法,且被广泛应用于图像分类^[1,2]及人脸识别等问题^[3-5]。Wilks 等首先提出了经典的 Fisher 准则^[6,7],随后, Foley 和 Sammon 提出了著名的 F-S 线性判别分析^[8],并应用于两类模式分类问题。后来, Duchene 等^[9]对这一方法作了进一步拓展,实现了多类模式的分类,给出了最优判别向量集的结果。近来,又由 Jin 和 Yang 等提出了具有统计不相关性的 J-Y 线性判别方法^[10],取得了很好的效果。

但是,以上各种基于 Fisher 准则的判别分析方法都没有考虑样本在模式空间的分布情况,以及边缘类

别对于分类正确率的影响,因此,这些方法得到的最优判别向量可能不是最优的。这是因为 Fisher 准则在优化过程中过分强调了那些具有较大类间距离的模式类别,使得投影方向较好地分开这些类,却造成了其它类别的大量重叠。为此,文中提出一种新的加权 Fisher 判别分析方法。对 COIL 图像库和 ORL 人脸数据库进行实验,结果表明,该方法有效削弱了边缘类别的影响,提高了识别的正确率。

1 加权 Fisher 判别分析

1.1 加权 Fisher 准则

假定 $\omega_1, \omega_2, \dots, \omega_L$ 为 L 个已知模式类, X 为一 n 维训练样本。设 m_0 为全体训练样本的均值向量, $m_i, S_i, P_i (i=1, 2, 3, \dots, L)$ 分别为 ω_i 类的均值向量、协方差矩阵和先验概率,则类间散度矩阵 S_b , 类内散度矩阵 S_w 和总体散度矩阵 S_t 分别为:

$$S_b = \sum_{i=1}^L P_i (m_i - m_0)(m_i - m_0)^T \quad (1)$$

$$S_w = \sum_{i=1}^L P_i E[(X - m_i)(X - m_i)^T | \omega_i] = \sum_{i=1}^L P_i S_i \quad (2)$$

收稿日期:2007-08-30

作者简介:黄国宏(1975-)男,上海交大博士毕业,讲师,研究方向为图像处理与分析、机器视觉、模式识别等;刘 刚,上海交大博士毕业,讲师,研究方向为图像处理与分析、机器视觉、模式识别等。

$$S_t = E[(X - m_0)(X - m_0)^T] = S_b + S_w \quad (3)$$

Fisher 准则函数可定义为:

$$J(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T S_w \varphi} \quad (4)$$

其中, φ 为任一使 $J(\varphi)$ 达到最大值的单位向量, 即模式样本在 φ 方向上的投影具有最大的类间散度和最小的类内散度, 亦即可分性最好。

作为文中所提算法的主要依据, 类间散度矩阵 S_b 可改写为如下形式^[11]:

$$S_b = \sum_{i=1}^{L-1} \sum_{j=i+1}^L P_i P_j (m_i - m_j)(m_i - m_j)^T = \sum_{i=1}^{L-1} \sum_{j=i+1}^L P_i P_j S_{ij} \quad (5)$$

可以看出, 类间散度矩阵可被表示成为两模式类别均值向量差的形式。事实上, $S_{ij} = (m_i - m_j)(m_i - m_j)^T$ 就是类别 i 与类别 j 的类间散度矩阵。因此, Fisher 判别准则函数可以写为:

$$J(\varphi) = \sum_{i=1}^{L-1} \sum_{j=i+1}^L P_i P_j (\varphi^T S_w \varphi)^{-1} [\varphi^T S_{ij} \varphi] \quad (6)$$

假设有一个类别与其它类别分隔很远, 称之为边缘类。在这种情况下, 通过最大化 Fisher 准则函数得到的最优投影方向 φ 就是尽可能地分开边缘类和其它类别, 因为在该方向上类别的方差最大。在最大化 Fisher 函数值的过程中, 那些具有较大类间距离的类别起了主导作用, 显然, 这样的方向并不能帮助分开除边缘类之外的其它类别, 甚至造成它们彼此重叠, 从而导致判别性能下降。因此, 传统的基于 Fisher 准则函数的判别分析并不是最优的。

针对这个问题, 通过引入一个加权函数 ω , 重新定义 Fisher 准则函数如下:

$$J(\varphi) = \sum_{i=1}^{L-1} \sum_{j=i+1}^L P_i P_j \omega(\Delta_{ij}) \frac{\varphi^T S_{ij} \varphi}{\varphi^T S_w \varphi} = \frac{\varphi^T S_B \varphi}{\varphi^T S_w \varphi} \quad (7)$$

其中, $\Delta_{ij} = \sqrt{(m_i - m_j)^T (m_i - m_j)}$,

$S_B = \sum_{i=1}^{L-1} \sum_{j=i+1}^L P_i P_j \omega(\Delta_{ij}) (m_i - m_j)(m_i - m_j)^T$ 为重新定义的加权类间散度矩阵, $\omega(\cdot)$ 为加权函数, Δ_{ij} 为两个类别之间的欧氏距离。显然, 当选择加权函数 $\omega(\cdot)$ 为常量时, 则式(7)与式(4)具有相同的投影方向。

既然要削弱边缘类在优化过程中的主导作用, 直觉的考虑是 $\|m_i - m_j\|$ 越大, 给它们的权值就越小。因此, 定义加权函数为:

$$\omega(\Delta_{ij}) = \frac{1}{2\Delta_{ij}^2} \operatorname{erf}\left(\frac{\Delta_{ij}}{2\sqrt{2}}\right) \quad (8)$$

其中, $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$ 为误差函数。这样, 式(7)就称之为加权 Fisher 准则。

1.2 最优判别向量集

一般来说, 式(7)的最优解可以通过特征值分解 $S_w^{-1} S_B$ 得到。然而, 对于高维小样本问题, 例如, 图像分类和人脸识别等问题, 最优解却不能直接获得, 因为此时类内散度矩阵 S_w 通常是奇异的。对于最优判别向量来说, 使得判别准则取值越大, 且提取的特征之间的统计相关性越小, 则判别向量的特征提取能力就越强^[10]。因此, 这里取极值的情形, 在类内散度矩阵 S_w 的零空间里求解具有统计不相关性的最优判别向量集。

假设 $\beta_1, \beta_2, \dots, \beta_m$ 为 m 个 S_w 的零特征值所对应的标准正交的特征向量, 其中 $n - m = \operatorname{rank}(S_w)$ 。定义 R^n 的子空间 $\Phi = \operatorname{span}\{\beta_1, \beta_2, \dots, \beta_m\}$, Φ 显然就是类内散度矩阵 S_w 的零空间。现在, 问题转化为如何在子空间 Φ 中求解最优判别向量的问题。依据代数定理, 子空间 Φ 与 m 维欧氏空间 R^m 同构, 且相应的同构映射为 $\varphi = PY$, 其中 $P = (\beta_1, \beta_2, \dots, \beta_m)$, $\varphi \in \Phi$, $Y \in R^m$ 。这样准则(7)可转化为:

$$J(Y) = Y^T (P^T S_B P) Y = Y^T \bar{S}_B Y \quad (9)$$

同时满足约束条件

$$\begin{cases} Y_n^T \bar{S}_t Y_t = 0 & t = 1, 2, \dots, n-1 \\ Y_n^T Y_n = 1 \end{cases}$$

其中, $\bar{S}_B = P^T S_B P$, $\bar{S}_t = P^T S_t P$ 。第一个最优判别向量 Y_1 可以通过直接最大化 J 得到, 即式(10)最大特征值对应的特征向量。

$$\bar{S}_B Y = \lambda Y \quad (10)$$

根据 Lagrange 乘子法, 求解第 n ($n > 1$) 个最优判别向量可转化为使下述函数取到最大值:

$$J = Y_n^T \bar{S}_B Y_n - \lambda (Y_n^T Y_n - 1) - \mu_1 Y_n^T \bar{S}_1 Y_1 - \dots - \mu_{n-1} Y_n^T \bar{S}_{n-1} Y_{n-1} \quad (11)$$

关于 Y_n 求导数, 并令导数为零, 可得到:

$$\frac{\partial J}{\partial Y_n} = 0$$

$$2\bar{S}_B Y_n - 2\lambda Y_n - \sum_{i=1}^{n-1} \mu_i \bar{S}_i Y_i = 0 \quad (12)$$

在式(12)两边左乘 Y_n^T , 得

$$2Y_n^T \bar{S}_B Y_n - 2\lambda Y_n^T Y_n = 0$$

$$\lambda = Y_n^T \bar{S}_B Y_n \quad (13)$$

问题就是使 λ 取到最大值。

在式(12)两边分别左乘 $Y_1^T \bar{S}_1$, $Y_2^T \bar{S}_2$, \dots , $Y_{n-1}^T \bar{S}_{n-1}$, 可得到:

$$D \bar{S}_t \bar{S}_t D^T \mu = 2 D \bar{S}_t \bar{S}_B Y_n \quad (14)$$

或记为:

$$\mu = 2 (D \bar{S}_t \bar{S}_t D^T)^{-1} D \bar{S}_t \bar{S}_B Y_n \quad (15)$$

其中, $D = [Y_1, Y_2, \dots, Y_{n-1}]^T$, $\mu = [\mu_1, \mu_2, \dots, \mu_{n-1}]^T$. 将式(15)代入式(12), 则得到:

$$2\bar{S}_B Y_n - 2\lambda Y_n - 2\bar{S}_t D^T (D \bar{S}_t D^T)^{-1} D \bar{S}_t \bar{S}_B Y_n = 0 \quad (16)$$

或写为:

$$[I - \bar{S}_t D^T (D \bar{S}_t D^T)^{-1} D \bar{S}_t] \bar{S}_B Y_n = \lambda Y_n \quad (17)$$

可知, Y_n 就是式(18)的最大特征值对应的特征向量。

$$M = [I - \bar{S}_t D^T (D \bar{S}_t D^T)^{-1} D \bar{S}_t] \bar{S}_B \quad (18)$$

2 实验结果与分析

为了对文中提出的加权 Fisher 判别分析(WFDA)和文献中同样基于 Fisher 准则的其它线性特征提取方法作比较, 分别采用了 COIL(Columbia Object Image Library)图像数据库和 ORL 人脸数据库进行识别实验。每个实验做 5 次, 每次实验的训练样本都是随机选取, 剩余样本作检测, 给出的实验结果都是 5 次实验的平均值。

2.1 实验 1

COIL 图像数据库由 100 个物体, 每个物体水平旋转一周采集 72 幅图像所组成, 每幅图像分辨率为 128×128 。图 1 所示为数据库中部分物体图像。

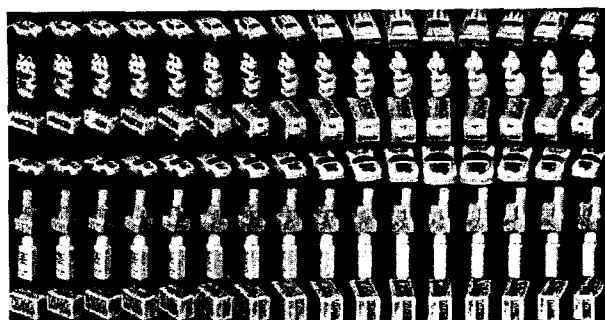


图 1 COIL 库部分图像

任意选取其中的 20 个物体进行 5 次实验, 每次实验随机选取每个物体的 36 幅图像作为训练样本, 其余的图像作为检测样本, 提取特征的维数为 20, 采用最近邻分类器。实验结果如表 1 所示。

表 1 识别率(%)结果比较

FST ^[9]	UFDA ^[10]	WFDA
90.56	94.16	97.84

由表 1 可知, 就识别正确率而言, 文中提出的加权 Fisher 判别分析(WFDA)在对 COIL 图像识别的效果要比著名的 F-S 判别分析(FST)和统计不相关判别分析(UFDA)好很多。这可以从直观上来解释, 由于 COIL 图像库的物体形状差异较大, 必然在空间分布上比较分散, 就有可能出现边缘类的情形。而文中方法对边缘类做了限制, 削弱了它在求解最优投影向量中的影响, 因此可以说, 文中提出的 WFDA 方法具有更

好的特征提取能力。

2.2 实验 2

实验采用 ORL 人脸图像数据库, 该数据库由 40 人, 每人 10 幅 92×112 图像所组成, 其中有些图像是拍摄于不同时期; 人脸脸部表情与脸部细节有变化, 例如: 笑或不笑, 眼睛睁着或闭着, 带或不带眼镜; 人脸姿态有变化, 深度旋转与平面旋转可达到 20° ; 人脸的尺度也有多达 10% 的变化。图 2 是 ORL 人脸数据库中的一些图像。



图 2 ORL 人脸图像数据库中的部分人脸

分别随机选取每人 $k(k=3, 4, 5)$ 幅图像作为训练样本, 剩余样本作为检测样本, 采用最近邻分类器, 对所提取的最优判别特征作识别实验, 特征维数为 39。结果如表 2 所示。

表 2 识别率(%)比较

k	Fisherface	UFDA	WFDA
3	87.5	87.5	87.6
4	88.7	93.8	94.0
5	89.5	95.0	96

通过与 Fisherface^[4]和 UFDA^[10]方法进行对比发现, 文中所提出的方法优于 Fisherface 方法, 而与 UFDA 方法相当。从直观上理解就是因为人脸图像之间的差别(包括光照、表情、姿态、背景等因素)不是很大, 因此在空间分布上相对集中, 类别间的距离相差不大, 导致权系数也相差很小; 而本质上, 如果文中算法(WFDA)的权值取为常数, 则与 UFDA 具有几乎相同的最优判别矢量。

当最优判别矢量的个数由 1 变化到 39 时, 采用最近邻和最小距离分类器, 文中算法分类错误率如图 3 所示。由图 3 可知, 随着特征维数的增加, 分类错误率迅速下降, 最小可达 4%。当特征维数大于 20 时, 分类错误率趋于稳定, 这一结果同时也说明所提取的特征满足最佳维数定理^[12], 即一般认为, 对一个模式识别问题来说, 最优模式特征的维数不大于模式类别数。

3 讨论

由以上两个实验可以看出, 如果待识别样本图像存在较大的差异, 在空间分布上就有可能出现“边缘

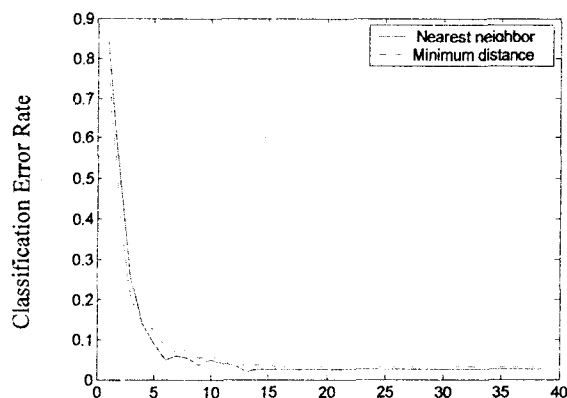


图 3 识别错误率随维数变化曲线

类”的情形,在利用基于 Fisher 准则的判别分析方法提取模式特征时,就要考虑边缘类别的影响,而文中提出的算法很好地解决了这个问题。对于待识别样本图像之间差异不大的情形,文中算法与文献[10]的算法识别结果几乎一致,因此,可以认为文中提出的特征提取方法更具有通用性。

4 结束语

提出了一种新的基于 Fisher 准则的特征提取方法——加权 Fisher 判别分析。首先,针对现有的基于 Fisher 准则的判别分析方法存在的不足,通过引入一个加权函数来削弱边缘类的影响,减少相邻类别间在投影空间中的大量重叠,从而有效提高了正确识别率。同时,针对小样本情形,给出了该算法的一种可行的最优判别向量集求解方法。通过实验证明了该方法性能良好,具有较强的特征提取能力,可以进一步应用于其它的图像识别问题。

参考文献:

- [1] Tian Q. Comparison of statistical pattern recognition algo-

rithms for hybrid processing[J]. I. Linear - mapping algorithm. Opt. Soc. Am., 1988, A5(10):1655 - 1669.

- [2] Tian Q. Comparison of statistical pattern recognition algorithms for hybrid processing[J]. II. Eigenvector - based algorithms. Opt. Soc. Am., 1988, A5(10):1670 - 1682.

- [3] Swets L, Weng J. Using discriminant eigenfeatures for image retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(8):831 - 836.

- [4] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces versus fisherfaces: Recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19:711 - 720.

- [5] Cheng J L, Wechsler H. A shape - and texture - based enhanced Fisher classifier for face recognition[J]. IEEE Transactions on Image Processing, 2001, 10(4):598 - 608.

- [6] Wilks S S. Mathematical Statistics[M]. New York: Wiley Press, 1962:577 - 578.

- [7] Duda R, Hart P. Pattern Classification and Scene Analysis [M]. New York: Wiley Press, 1973.

- [8] Foley D H, Sammon J W. An optimal set of discriminant vectors[J]. IEEE Trans. Comput., 1975, 24:281 - 289.

- [9] Duchene J, Leclercq S. An optimal Transformation for discriminant and principal component analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1988, 10(6):978 - 983.

- [10] Zhong J, Yang J Y, Hu Z S, et al. Face Recognition based on uncorrelated discriminant transformation[J]. Pattern Recognition, 2001, 33(7):1405 - 1416.

- [11] Long M, Duin R P W, Haeb - Umbach R. Multiclass linear dimension reduction by weighted Pairwise Fisher criteria[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(7):762 - 766.

- [12] Fukunaga K. Introduction to statistical pattern recognition [M]. New York: Academic Press, 1990.

(上接第 226 页)

4 结束语

(1)网上选课系统构建后,首先在本科生中投入使用,经过四届学生的选课操作,不断修改系统功能,目前系统功能已经完善,用户操作更加人性化。

(2)政策先行,制度做保障。在系统整个开发阶段,围绕学院学分制的政策进行了多次讨论,由于政策没有事先落实和制定,导致系统模块进行大范围调整,影响了系统的开发进度。

(3)网上选课系统的数据处理算法有待进一步研究和探讨。目前系统采用的是优选级和先来先服务的办法,不能保证选课的公平、公正。

参考文献:

- [1] 胡世军,卫春燕,芮执元.学分制的实施现状及其研究[J]. 甘肃科技, 2004, 20(2):151 - 153.

- [2] 萨师煊,王 珊.数据库系统概论[M]. 北京:高等教育出版社, 2002.

- [3] 唐 扬,熊 伟,陈宏盛,等.数据库触发器机制的设计与实现[J]. 电子技术应用, 2005(2):16 - 18.

- [4] 刘德春.数据库系统原理与应用[M]. 武汉:湖北人民出版社, 2003.

- [5] 郭海峰,阳国贵. Oracle 数据库性能调优技术与实现[J]. 计算机工程, 2006, 32(19):82 - 94.