

一种新的基于 Schur 分解的 Fisher 鉴别分析方法^{*}

林宇生 杨静宇

(南京理工大学计算机系 南京 210094)

摘要 特征抽取是模式识别研究领域的一个热点。本文提出了一种新的基于 Schur 分解的 Fisher 鉴别分析的特征抽取方法。此方法引入了零空间的概念,指出了前人算法中的不足之处;给出了一个完整的基于 Schur 分解的 Fisher 鉴别分析方法。ORL 人脸数据库上的实验表明了该方法的有效性。

关键词 特征抽取, Fisher 鉴别分析, Schur 分解, 零空间

A New Method of Fisher Discriminant Analysis with Schur Decomposition

LIN Yu-Sheng YANG Jing-Yu

(Department of Computer Science, Nanjing University of Science & Technology, Nanjing 210094)

Abstract Feature extraction is one of the hot topics in the field of pattern recognition. In this paper, we point out the weakness of the previous methods and a new method of Fisher discriminant analysis with Schur decomposition is proposed. Conception of null space is introduced in this paper. Experimental results on ORL face database indicate that the proposed method is valid.

Keywords Feature extraction, Fisher discriminant analysis, Schur decomposition, Null space

1 引言

在模式识别领域中,特征抽取是重要的研究内容之一,其基本任务是找出分类时最有效的特征^[1,2]。Fisher 鉴别分析(Fisher discriminant analysis, FDA)是公认的特征抽取最重要的工具之一,目前仍然被广泛地应用在人脸等图像识别领域^[3,4],其研究最早可追溯到 1936 年 Fisher 发表的论文^[5]。其基本思想为选择使得 Fisher 准则函数达到最优的向量作为最优投影方向,样本在该方向上的投影能达到类间散度最大和类内散度最小。但是, Fisher 鉴别矢量通常是统计相关的。许多试验已经表明,通过使鉴别矢量彼此正交可以减弱这种相关性。基于这一思想, Sammon 提出了计算鉴别平面的算法^[6]。Sammon 鉴别平面的第一个鉴别矢量为具有最大类间距离与类内距离比值的矢量,第二个鉴别矢量为与第一个鉴别矢量正交且使 Fisher 准则函数(即类间距离与类内距离比值)具有极大值的矢量。在此基础上, Foley 与 Sammon 在 2 类问题的识别中引入了最优鉴别矢量集的概念,使用多个正交鉴别矢量进行鉴别分析^[7]。此外,多类问题的正交鉴别矢量集的解也被求出^[8],一般称这类正交的矢量集为 F-S 鉴别矢量集。

F-S 鉴别分析方法(Foley-Sammon discriminant analysis, FSD)也是有效的特征抽取方法,但是该方法计算耗时。针对这一问题,文^[9]提出了一种基于 Schur 分解的 Fisher 鉴别分析方法(Fisher discriminant analysis with Schur decomposition, FDS)。此方法首先采用 PCA^[10]降维方法来避免小样本问题,然后利用 Schur 定理,简便地求出了同时满足正交条件和 Fisher 准则的最优鉴别矢量。然而,该方法并没有考虑在经过 PCA 变换之后样本的类内散度矩阵存在零空间^[11~13]的

问题。本文中关注此零空间的信息,并提出了一种完整的基于 Schur 分解的 Fisher 鉴别分析方法(Complete Fisher discriminant analysis with Schur decomposition, CFDS)。

2 Fisher 鉴别分析方法

Fisher 鉴别分析的目的是从高维特征空间中提取出最具有鉴别能力的低维特征,这些特征可以将同一类别的样本聚集在一起,而使不同类别的样本尽可能地分开,即选择使得样本类间散度和样本类内散度的比值最大的特征。

样本类间散度矩阵定义为

$$S_b = \sum_{i=1}^C p_i (u_i - u)(u_i - u)^T \quad (1)$$

其中, u_i 是 C_i 类的均值, u 是所有样本的均值, p_i 是先验概率。

样本类内散度矩阵定义为

$$S_w = \sum_{i=1}^C p_i S_i \quad (2)$$

其中 $S_i = E\{(x - u_i)(x - u_i)^T | x \in C_i\}$

我们希望原始数据经过投影后,在低维空间内得到的就是最有助于模式分类的信息。因此,如果类内散度矩阵 S_w 是非奇异矩阵,最优的投影方向 W 就是使得样本类间散度矩阵 S_b 和样本类内散度矩阵的行列式比值最大的那些特征向量。因此, Fisher 准则函数定义为

$$J(W) = \arg \max_w \frac{W^T S_b W}{W^T S_w W} \quad (3)$$

通过线性代数理论,我们可知 W 就是满足如下等式的解:

$$S_b W_i = \lambda_i S_w W_i \quad (4)$$

通常的 Fisher 准则函数通过求得对应矩阵大于零的特征值所对应的特征向量以得到投影空间,此时我们注意到该

^{*} 国家自然科学基金资助(编号:60472060, 60473039, 60503026)。林宇生 博士研究生,主要从事图像处理、模式识别方面研究;杨静宇 教授,博导,研究方向为模式识别、计算机视觉、图像处理。

矩阵 $S_w^{-1}S_b$ 最多只有 $C-1$ 个非零特征值,其中 C 为类别数目。

当类内散度矩阵奇异的时候,Fisher 准则函数也可以采用如下的形式:

$$J(W) = \arg \max_w \frac{W^T S_b W}{W^T S_i W} \quad (5)$$

其中的 S_i 为样本的总体散度矩阵。

3 FDS 方法

为了解决 FSD 方法的计算耗时问题,Song^[9] 等人利用 Schur 定理提出了 FDS 方法,下面给出 Schur 定理及 FDS 方法的描述。

定理 1^[14](Schur) 设 A 为 n 阶方阵, $\lambda_1, \lambda_2, \dots, \lambda_n$ 是 A 的特征值,不论它们是实数还是复数,总存在相似酉矩阵 U ,使得 $A=UTU^T$,其中 T 为三角矩阵,对角线上的元素是 $\lambda_1, \lambda_2, \dots, \lambda_n$ 。

依据定理 1,当 A 为实 n 阶方阵时,它的 Schur 分解是 $A=UTU^T$,这里 U 是正交矩阵。Fisher 鉴别分析是对 $S_w^{-1}S_b$ (当 S_w 可逆时)或 $S_i^{-1}S_b$ (当 S_w 不可逆时)进行特征分析,而在 FDS 中是对 $S_w^{-1}S_b$ 或 $S_i^{-1}S_b$ 进行 Schur 分解。设 $S_w^{-1}S_b$ 的 Schur 分解是 UTU^T , u_1, u_2, \dots, u_n 是矩阵 U 的所有的列向量,显然 u_1, u_2, \dots, u_n 是相互正交的。假设 u_1, u_2, \dots, u_d 是 $S_w^{-1}S_b$ 前 d 个最大特征值所对应的 Schur 分解向量,这样通过映射: $X \rightarrow V^T X$ 可以把高维的输入空间压缩成一个低维的特征空间。这里, $V = u_1, u_2, \dots, u_d$ 是 FDS 的鉴别矩阵, u_1, u_2, \dots, u_d 是鉴别向量。

4 我们的方法

文[9]在采用 FDS 方法之前,为了能够避免小样本问题,先采用 PCA 对样本进行降维,然后对 $S_w^{-1}S_b$ 进行 Schur 分解。但是该方法并没有考虑经过 PCA 变换后的空间中会出现新的类内散度矩阵奇异的情况,此时利用 FDS 方法是否丢弃了相应的分类信息呢?那么如何有效地得到更多的鉴别信息,以提高该方法的识别率呢?在本文中,我们改进了文[9]的方法,利用零空间的概念,提出了一个完整的基于 Schur 分解的 Fisher 鉴别分析方法。

文[15]中指出,假设原始图像数据空间为 R^n ,那么所有的最优鉴别矢量都能从低维空间 R^m ($m = \text{rank}(S_i)$)中得到而不丢失任何的信息。这种低维空间通过 PCA 方法获得。我们给出以下定义和定理。

定义 1 设 $m = \text{rank}(S_i)$,矩阵 $P = (P_1, P_2, \dots, P_m)$,其中 P_1, P_2, \dots, P_m 为 S_i 的 m 个非零特征值所对应的 m 个标准的特征向量,定义:

$$\tilde{J}(Y) = \frac{Y^T \tilde{S}_b Y}{Y^T \tilde{S}_i Y} \quad (6)$$

其中 $\tilde{S}_b = P^T S_b P, \tilde{S}_i = P^T S_i P, Y = Y_1, Y_2, \dots, Y_d$ 。

公式(6)就是在线性变换之后在 R^m 空间的 Fisher 鉴别准则函数。

定理 2 设 $Y = Y_1, Y_2, \dots, Y_d$ 是在 PCA 变换空间 R^m 中采用 FDS 方法所求的鉴别向量集,那么 $W = PY = (W_1 = PY_1, W_2 = PY_2, \dots, W_d = PY_d)$ 就是同时满足 Fisher 准则函数 $J(W)$ 与正交条件的最优鉴别矢量集。

证明:因 Y_1, Y_2, \dots, Y_d 是 FDS 的鉴别矢量,所以由 Schur 定理知向量满足正交条件,即

$$Y_i^T Y_j = \delta_{ij} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases}$$

故 $W_i^T W_j = Y_i^T P^T P Y_j = \delta_{ij}$

又 Y_1, Y_2, \dots, Y_d 是矩阵 $\tilde{S}_i^{-1} \tilde{S}_b$ 前 d 个最大特征值所对应的 Schur 分解向量,所以有

$$\tilde{S}_i^{-1} \tilde{S}_b Y_j = \lambda_j Y_j \quad j=1, 2, \dots, d$$

$$[Y_1, Y_2, \dots, Y_d]^T \tilde{S}_b [Y_1, Y_2, \dots, Y_d] = [Y_1, Y_2, \dots, Y_d]^T \tilde{S}_i [Y_1, Y_2, \dots, Y_d] \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$$

$$J(Y_1, Y_2, \dots, Y_d) = \frac{[Y_1, Y_2, \dots, Y_d]^T \tilde{S}_b [Y_1, Y_2, \dots, Y_d]}{[Y_1, Y_2, \dots, Y_d]^T \tilde{S}_i [Y_1, Y_2, \dots, Y_d]} = \prod_{j=1}^d \lambda_j = \max \tilde{J}(Y)$$

$$J(W) = \frac{(PY)^T S_b (PY)}{(PY)^T S_i (PY)} = \frac{Y^T (P^T S_b P) Y}{Y^T (P^T S_i P) Y} = \frac{Y^T \tilde{S}_b Y}{Y^T \tilde{S}_i Y} = \tilde{J}(Y)$$

证毕。

总的线性变换可以如下定义:

$$Z = W^T X \quad (7)$$

因此,我们现在的问题就是如何求经过 PCA 变换的子空间中的最优鉴别矢量。因为经过 PCA 降维的子空间 R^m 中,仍然会存在 $m > \text{rank}(\tilde{S}_w)$ 的问题,所以对此空间作特征提取的时候就应当考虑存在于 \tilde{S}_w 对应的零空间中的特征信息。

假设 $\gamma_1, \gamma_2, \dots, \gamma_m$ 是降维之后空间类内散度矩阵的 m 个特征向量,并且其对应的特征值按照从大到小的顺序排列 $\lambda_1 > \lambda_2 > \dots > \lambda_m$ 。如果存在零空间,我们假定其中的前 m_1 个特征值是正值。因此,对应的特征矢量可以分为 2 部分:特征值为 0 的值对应的特征向量 $Q_1 = (\gamma_{m_1+1}, \gamma_{m_1+2}, \dots, \gamma_m)$ 和特征值不为 0 的值对应的特征向量 $Q_2 = (\gamma_1, \gamma_2, \dots, \gamma_{m_1})$ 。此时利用 FDS 做特征抽取的时候必须同时考虑在这两个不同的子空间的信息。基于以上关于零空间信息,我们可以得到以下完整的算法描述。

步骤 1:对原始图像做 PCA 变换,将原始图像变换到 m ($m = \text{rank}(S_i)$)维空间。计算得到 R^m 空间类内散度矩阵的特征向量 $\gamma_1, \gamma_2, \dots, \gamma_m$ 。所有的特征向量按照对应特征值从大到小的顺序排列,其中前 m_1 个特征向量对应大于 0 的特征值。

步骤 2:得到特征值 0 所对应的特征向量 $Q_1 = (\gamma_{m_1+1}, \gamma_{m_1+2}, \dots, \gamma_m)$ 。那么我们可以得到新的类间散度矩阵 $\tilde{S}_b = Q_1^T \tilde{S}_b Q_1$,由于此时位于 \tilde{S}_w 的零空间中,对 \tilde{S}_b 进行 Schur 分解,得到相应的 Schur 向量 Z_1, Z_2, \dots, Z_l (通常 $l = C-1, C$ 为样本数目)。因此我们可以得到当相应的类内散度矩阵奇异时的最优鉴别矢量集 $Y_j = Q_1 Z_l$ 。

步骤 3:得到特征值不为 0 所对应的特征向量 $Q_2 = (\gamma_1, \gamma_2, \dots, \gamma_{m_1})$ 。那么我们可以得到在新的子空间内对应的类间散度矩阵 $\tilde{S}_b = Q_2^T \tilde{S}_b Q_2$ 和总体散度矩阵 $\tilde{S}_i = Q_2^T \tilde{S}_i Q_2$ 。因此,可以对 $\tilde{S}_i^{-1} \tilde{S}_b$ 进行 Schur 分解,得到余下的 $d-l$ 个最优鉴别矢量 Z_j ($j=l+1, l+2, \dots, d$)。由此,我们可以得到当相应的类内散度矩阵非奇异时的最优鉴别矢量集 $Y_j = Q_2 Z_j$ ($j=l+1, l+2, \dots, d$)。

步骤 4:通过 $Y_j = Q_1 Z_l$ 和 $Y_j = Q_2 Z_j$ 得到所有在经过 PCA 变换之后子空间内的最优鉴别矢量集。

$$\text{vects} = (Y_1, Y_2, \dots, Y_l, Y_{l+1}, \dots, Y_d)$$

步骤 5:得到测试样本在最优鉴别矢量集上的投影数据进行分类。

5 实验结果

我们的实验数据采用 ORL^[16] 人脸数据库。ORL 人脸数

据库包括从 1992 年 4 月到 1994 年 4 月剑桥大学实验室拍摄的一系列人脸图像,具体为 40 个人,每个人有不同表情或不

同视点的 10 幅图像所构成,倾斜角度不超过 20 度。人脸库中的部分人脸图像如图 1 所示。



图 1 ORL 人脸数据库部分人脸图像

实验中的训练数据集和测试数据集均随机生成,分别从每类中取 $\theta=4,5,6$ 构成训练样本集。每次实验中的训练样本集均随机产生,数据库中训练样本集之外的数据构成测试样本集。最后采用最近邻分类器进行分类。我们的实验中,在每个不同训练样本数目下均做 10 次不同的实验。表 1 显示了在不同的实验方法中 10 次不同结果的平均识别率和方差比较。

表 1 ORL 人脸数据库识别率(均值和方差)

训练样本	4	5	6
FDA	90.11±3.40	92.20±2.04	94.37±2.56
FSD	92.75±1.52	94.10±1.22	95.50±2.14
FDS	94.67±1.66	95.40±1.85	96.20±2.10
CFDS	95.45±1.11	96.55±1.50	97.75±1.42

图 2 显示了每类随机抽取 5 个训练样本的 10 次不同的识别率比较。

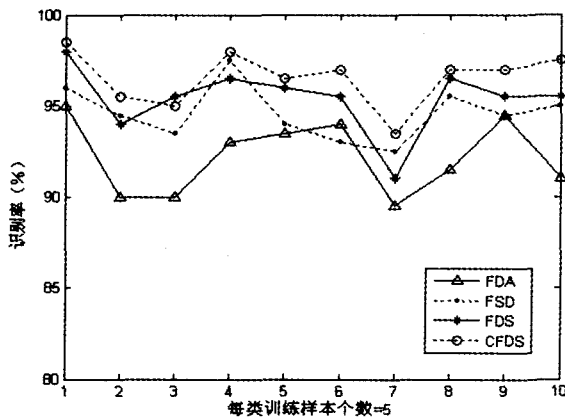


图 2 四种特征提取方法识别率比较

从表 1 可以得出,CFDS 的平均识别率优于其它的四种特征提取方法。图 2 的 10 次不同样本情况下的识别率比较图说明了 CFDS 方法有较好的鲁棒性。从以上的实验结果中可以得出,由于 CFDS 引入了零空间信息,得到了更多的分类信息,人脸的识别性能有了明显的提高,并且该方法的鲁棒性较好。

结束语 本文提出了一种完整的基于 Schur 分解的 Fisher 鉴别分析方法,该方法引入了零空间的概念,抽取得到经过 PCA 变换后的所有分类信息,从而得到了更多有助于分类的特征信息。同时所得到的鉴别矢量之间具有正交性,减

弱了鉴别矢量间的相关性。在 ORL 人脸数据库的实验结果表明所提出方法的识别率优于 FDS 方法。

参考文献

- 1 Fukunaga K. Introduction to statistical pattern recognition [M]. New York: Academic Press, 1990
- 2 边肇祺,张学工. 模式识别[M]. 第 2 版. 北京:清华大学出版社, 1999
- 3 Pentland A. Looking at people: sensing for ubiquitous and wearable computing [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22 (1): 107~119
- 4 Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs fisherfaces: recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19 (7): 711 ~720
- 5 Fisher R A. The use of multiple measurements in taxonomic problem [J]. Annals of Eugenics, 1936, 7 (1): 179~188
- 6 Sammon J W. An optimal discriminant plane [J]. IEEE Transaction on Computer, 1970, 19 (9): 826~829
- 7 Foley D H, Sammon J W. An optimal set of discriminant vectors [J]. IEEE Transaction on Computer, 1975, 24(3): 281~289
- 8 Duchene J, Leclercq S. An optimal transformation for discriminant and principal component analysis. IEEE Trans on pattern analysis and machine intelligence, 1988, 10 (6): 978~983
- 9 Song Fengxi, Zhang D, Yang Jingyu. A novel dimensionality-reduction approach for face recognition [J]. Neurocomputing, 2006, 39: 1683~1687
- 10 Turk M, Pentland A. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3 (1): 71~86
- 11 Yu H, Weng J. Using Discriminant Eigen features for Image Retrieval [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 1996, 18(8): 831~836
- 12 Chen L F, Liao H Y M, Lin J C, et al. A New LDA based Face Recognition System Which Can Solve the Small Sample Size Problem [J]. Pattern Recognition, 2000, 33(10): 1713~1726
- 13 Yang J, Frangi A F, Yang J Y, et al. KPCA plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature extraction and Recognition [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2005, 27(2): 230~244
- 14 刘慧,袁文燕. 矩阵论及其应用[M]. 化学工业出版社, 2003, 9
- 15 Yang J, Yang J Y. Why can LDA be performed in PCA transformed space? [J]. Pattern Recognition, 2003, 36(2): 563~566
- 16 ORL face database. <http://www.uk.research.att.com/facedatabase.html>