

一种基于 Fisher 鉴别极小准则的特征提取方法

郑宇杰 杨静宇 徐 勇 於东军

(南京理工大学计算机科学与技术学院 南京 210094)

(yjzheng13@yahoo.com.cn)

A New Feature Extraction Method Based on Fisher Discriminant Minimal Criterion

Zheng Yujie, Yang Jingyu, Xu Yong, and Yu Dongjun

(School of Computer Science & Technology, 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, a new feature extraction method called Fisher discriminant minimal criterion is proposed to improve the performance of feature extraction. Conventional Fisher discriminant criterion is inversed and null space of between-class scatter matrix is defined in this algorithm. Therefore, limitation of final eigenvectors' dimensions determined by class number is overcome and more effective classification information can be achieved. Experimental results on face databases demonstrate the effectiveness of the proposed algorithm.

Key words feature extraction; face recognition; null space; Fisher discriminant criterion; principal component analysis

摘 要 特征提取是模式识别研究领域的一个热点. 为了更好地解决人脸识别中的特征提取问题, 定义了一种新的基于 Fisher 鉴别极小准则的特征提取方法, 并且提出了类间散布矩阵零空间的概念, 解决了先前 Fisher 线性变换方法中的最终特征维数受类别数的限制. 在人脸数据库上的实验结果验证了该算法的有效性.

关键词 特征提取; 人脸识别; 零空间; Fisher 鉴别准则; 主成分分析

中图法分类号 TP391.41

1 引 言

人脸识别是目前模式识别研究领域一项极富挑战性的研究课题, 其研究不但对于模式识别理论的发展具有重要的学术意义, 而且在商业、法律和其他领域都有着广泛的实用价值和巨大的市场应用前景. 到目前为止, 已经提出了很多算法来处理人脸识别问题^[1], 其中基于代数的人脸识别方法是

当前人脸识别方法的主流, 并且取得了较好的实验效果. 在此类人脸识别方法中, 基于 KL 变换的特征脸识别方法^[2,3]和基于 Fisher 鉴别准则的 Fisherface 人脸识别方法^[4,5]是使用较广的方法. 基于 Fisher 鉴别准则的人脸识别方法, 由于可得到有助于分类的最佳鉴别投影信息, 因此成为代数方法中的一个研究热点. 然而, Fisher 鉴别准则在实际使用中经常会遇到小样本问题^[6,7] (small sample size problem, SSSP). 此时在类内散布矩阵奇异的情况

下,就不能直接应用相应的鉴别准则,因此有很多学者提出了相应的算法,基于 PCA+LDA 的组合方法是此类方法的主流. Yang 等人^[8]提出了一个完整的 PCA+LDA 的人脸识别方法,但是该方法当 PCA 降维之后的样本类内散布矩阵为满秩^[9,10]时,就不存在相应零空间的信息,而且最终得到的特征数据维数受原始图像类别数的限制. 因此, Xu 等人^[11]提出了一种基于 Fisher 鉴别极小准则的算法,但是 Xu 的方法始终假设取得特征数据的类间散布矩阵是正定的,这个是必须的吗? 本文提出了一种新的基于样本类间散布矩阵零空间概念的 Fisher 鉴别极小准则新算法,解决了前人在应用 Fisher 鉴别准则上的不足之处. 在人脸数据库上的实验验证了该算法的有效性.

2 相关知识

2.1 Fisher 鉴别准则

线性鉴别分析方法的目的是从高维特征空间中提取出最具有分类能力的低维特征,我们希望投影后的特征数据,在低维空间里不同类别的样本尽可能分得开些,同时希望每个类别内部样本尽量密集,也就是说,样本类间离散度越大越好,而样本类内离散度越小越好.

现假设有 c 类已知的模式 $w_1, w_2, \dots, w_c, X = \{x_i\}, i=1, 2, \dots, N$ 为一个具有 n 维的样本集. 任何一个在样本集 X 中的样本 x_i 均属于某个已知类别 w_j , 即 $x_i \in w_j$, 其中 $i=1, 2, \dots, N, j=1, 2, \dots, c$.

那么,样本类间散布矩阵定义为

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

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

样本类内散布矩阵定义为

$$S_w = \sum_{i=1}^c p_i e_i, \quad (2)$$

其中,

$$e_i = E[(x - u_i)(x - u_i)^T | x \in w_i]. \quad (3)$$

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,如果类内散布矩阵 S_w 是非奇异矩阵,最优的投影方向 W_{opt} 就是使得样本类间散布矩阵和样本类内散布矩阵的行列式比值最大的那些正交特征向量. 因此 Fisher 准则函数定义为

$$f(W_{opt}) = \arg \max_w \frac{W^T S_b W}{W^T S_w W}. \quad (4)$$

通过线性代数理论,我们可知 W_{opt} 就是满足如下等式的解:

$$S_b W_i = \lambda_i S_w W_i. \quad (5)$$

通常的 Fisher 准则函数通过求得对应矩阵 $S_w^{-1} S_b$ 大于零的特征值 λ_i 所对应的特征向量以得到投影空间,此时我们注意到该矩阵最多只有 $c-1$ 个非零特征值,其中 c 为原始模式数据的类别数.

当类内散布矩阵奇异的时候, Fisher 准则函数也可以采用如下的形式:

$$F(W_{opt}) = \arg \max_w \frac{W^T S_b W}{W^T S_t W}, \quad (6)$$

其中的 S_t 为样本的总体散布矩阵,并且满足

$$S_t = S_b + S_w.$$

2.2 PCA+LDA 的特征提取

基于 Fisher 鉴别准则的线性变换方法给我们提供了一个取得低维分类数据的有效方法,但是在人脸识别等小样本问题的模式识别问题中,经常会出现不能直接应用 Fisher 鉴别准则的情形,因此有很多学者提供了相应的解决方法. Yang 等人提出了一个完整的线性变换过程: PCA+LDA. 该方法中,首先通过 PCA 变换将原始图像空间从 n 维降到 m ($m = \text{rank}(S_t)$) 维空间 R^m , 其中 ($n > m$). 此时可以保证所有的分类信息都存在 m 维的子空间中. 重新定义在新的 m 维空间中相应的样本散布矩阵.

样本类间散布矩阵为

$$\tilde{S}_b = P^T S_b P, \quad (7)$$

样本类内散布矩阵为

$$\tilde{S}_w = P^T S_w P, \quad (8)$$

其中的 P 为 PCA 变换矩阵, S_b, S_w 分别为原始图像对应的样本类间散布矩阵和类内散布矩阵.

对于 R^m 空间的类内散布矩阵, Yang 提出将此矩阵分为相应的非零空间和零空间,分别计算得到相应空间的鉴别向量构成最终的投影向量.

此方法虽然给出了一个完整的线性特征提取过程,但是如何保证降维之后的类内散布矩阵始终存在零空间的,而且如此特征向量的求解过程中最终特征向量的维数还受到样本类别数的限制. 因此 Xu 提出了一种基于 Fisher 鉴别极小准则的特征提取方法,该方法同样采用了基于 PCA+LDA 的特征提取过程. 首先采用 PCA 方法将原始图像从 n 维降到 m ($m = \text{rank}(S_t)$) 维空间中. 在降维后的子空间中,采用基于 Fisher 鉴别极小准则的线性变换方

法得到相应的特征向量. 在 X_u 的方法中, 始终假设抽取得到的是处于样本类间散布矩阵 \hat{S}_b 大于零的空间内的信息. 当 \hat{S}_b 位于零空间内时, 我们能否从 \hat{S}_w 空间得到相应的鉴别信息呢? 本文结合 \hat{S}_b 零空间内的信息, 提出了一个完整的基于 PCA + LDA 变换的 Fisher 鉴别极小准则的算法.

2.3 理论基础

本文中我们定义 Fisher 鉴别极小准则如下:

定义 1. Fisher 鉴别极小准则函数为

$$j_{\min}(W_{\text{opt}}) = \arg \min_w \frac{W^T S_w W}{W^T S_b W} \text{ 或}$$

$$J_{\min}(W_{\text{opt}}) = \arg \min_w \frac{W^T S_w W}{W^T S_t W}.$$

从 Fisher 鉴别极小准则的意义来分析, 我们可以得出, 当准则函数 $j_{\min}(J_{\min})$ 取得的值越小, 将原始模式投影到 W 方向后的模式可分性越好.

我们首先证明定义 1 中的两个等式在求解最优特征向量集时可以等价使用.

定理 1. 对于 $\forall x$, 函数 $r(x) \geq 0, s(x) > 0, y_1$

$(x) = \frac{r(x)}{s(x)}, y_2(x) = \frac{r(x)}{r(x) + s(x)}$, 若存在 x_0 , 使得 $y_1(x_0) = \min y_1(x)$, 则有 $y_2(x_0) = \min y_2(x)$.

证明. 根据代数基本理论, 显然有 $y_1(x) \geq 0, y_2(x) \geq 0$, 所以

$$y_2(x) = \frac{\frac{r(x)}{s(x)}}{1 + \frac{r(x)}{s(x)}} = \frac{y_1(x)}{1 + y_1(x)} = 1 - \frac{1}{1 + y_1(x)}.$$

由于 $y_1(x_0) = \min y_1(x)$, 所以

$$\frac{1}{1 + y_1(x_0)} = \max \frac{1}{1 + y_1(x)},$$

因此

$$y_2(x_0) = 1 - \frac{1}{1 + y_1(x_0)} = \min y_2(x).$$

证毕.

通过定理 1 我们可以得到, 当样本类间散布矩阵正定时, 定义 1 中的 j_{\min} 和 J_{\min} 是可以等价使用的, 因此本文的后续部分都以 J_{\min} 作为我们的鉴别准则. 此外, 对于处于样本类间散布矩阵 S_b 零空间内的最优特征向量求解我们将给出新的方法.

定理 2. Fisher 鉴别极小准则可转化成如下广义特征问题:

$$S_w W = \lambda S_b W \text{ 或者 } S_w W = \lambda S_t W.$$

引理 1^[8]. 当且仅当 $X^T S_w X = 0$ 和 $X^T S_b X = 0$ 时, 才有 $X^T S_t X = 0$.

定义 2^[8]. 若 $\beta_1, \beta_2, \dots, \beta_n$ 为总体散布矩阵 S_t 的正交特征向量, 并且按照对应特征值从大到小的次序排列. 假设 $\text{rank}(S_t) = m$, 由此可知 $\beta_1, \beta_2, \dots, \beta_m$ 为非零特征值所对应的特征向量, $\beta_{m+1}, \beta_{m+2}, \dots, \beta_n$ 为零特征值所对应的特征向量, 因此可定义相应的特征向量空间 $R_m = \text{span}\{\beta_1, \dots, \beta_m\}$ 和 $R_m^\perp = \text{span}\{\beta_{m+1}, \dots, \beta_n\}$.

定理 3. 对于任意的 $\vartheta \in R^n, \vartheta$ 都能表述成 $\vartheta = X + \zeta$ 的形式, 其中的 $X \in R_m, \zeta \in R_m^\perp$, 那么我们可以得到 $J_{\min}(\vartheta) = J_{\min}(X)$.

由定理 3 可知所有的最优鉴别特征向量都可以从原始样本的总体散布矩阵非零空间中得到. 这个变换过程可通过 PCA 方法实现, 经过该方法之后, 我们可以得到降维之后的 m 维数据. PCA 变换过程如下:

$$Y = P^T X,$$

其中 $P = (\beta_1, \beta_2, \dots, \beta_m), X \in R^n, Y \in R^m$.

经过 PCA 降维之后, 此时 R^m 空间内的 Fisher 鉴别极小准则就变为如下的形式:

$$J_{\min}(X) = \frac{Y^T (P^T S_w P) Y}{Y^T (P^T S_t P) Y} = \frac{Y^T \tilde{S}_w Y}{Y^T \tilde{S}_t Y} = \tilde{J}_{\min}(Y).$$

由此, 我们就能通过 Fisher 鉴别极小准则在 PCA 降维之后的新空间内得到最终的特征向量集.

3 如何执行我们的算法

Yang 给出了 PCA + LDA 的完整算法, 但是该算法并没有考虑如果当 PCA 降维之后的样本类内散布矩阵满秩时, 此时不存在零空间的概念, 并且受样本类间散布矩阵秩的限制, 最终得到的特征向量维数始终在一个有限的范围之内. 此外 X_u 的基于 Fisher 鉴别极小准则的特征抽取方法, 虽然也提出了一个基于 PCA + LDA 的方法, 但是该方法始终假设是在降维之后的样本类间散布矩阵非零空间中求取最终特征向量, 因此, 被认为丢失了原始数据的部分鉴别信息. 本文克服了这些局限, 提出了一个完整的基于 Fisher 鉴别极小准则的特征提取方法.

首先, 将原始图像通过 PCA 方法从 n 维变换到 m ($m = \text{rank}(S_t)$) 维子空间 R^m 中, 然后在该子空间中应用 Fisher 鉴别极小准则. 在 PCA 变换之后的子空间 R^m 中, 样本的类间散布矩阵 \tilde{S}_b 可以分为

$\tilde{\phi}_b = (\tau_1, \tau_2, \dots, \tau_q)$ 和它的正交补零空间 $\tilde{\phi}_b^\perp = (\tau_{q+1}, \tau_{q+2}, \dots, \tau_m)$. 其中 $\tau_1, \tau_2, \dots, \tau_m$ 是 \tilde{S}_b 的正交特征向量集, 并且按特征值从大到小的次序排列相应的特征向量, 前 q 个正交特征向量对应正的特征值. 因此, 基于 Fisher 鉴别极小准则的特征提取过程可以分 2 个空间来实现: 当求取 $\tilde{\phi}_b$ 空间内的特征向量时, 此时相应的类内散布矩阵为 $\hat{S}_w = Y^T \tilde{S}_w Y$, 总体散布矩阵 $\hat{S}_t = Y^T \tilde{S}_t Y$, 根据 Fisher 鉴别极小准则, 可得到相应的特征向量集; 当求取 $\tilde{\phi}_b^\perp$ 空间内的特征向量时, 此时相应的类内散布矩阵 $\tilde{S}_b = Y^T \tilde{S}_b Y = 0$, 类内散布矩阵 $\tilde{S}_w = Y^T \tilde{S}_w Y$, 因此此时的 Fisher 鉴别极小准则可定义为 $\tilde{J}_w = Y^T \tilde{S}_w Y$, 求解相应的最小特征值所对应的特征向量构成最优鉴别特征向量集. 因此, 完整的基于 Fisher 鉴别极小准则的算法描述如下:

1) 首先对原始图像做 PCA 变换, 将原始图像降到 m 维空间 R^m , m 的值由原始图像样本总体散布矩阵 S_t 的秩所决定.

2) 取 $P_1 = (\tau_1, \tau_2, \dots, \tau_q)$, 重新计算得到相应 $\tilde{\phi}_b$ 空间内的类内散布矩阵 $\hat{S}_w = P_1^T \tilde{S}_w P_1$ 和总体散布矩阵 $\hat{S}_t = P_1^T \tilde{S}_t P_1$. 根据 Fisher 鉴别极小准则, 得到相应的按特征值从小到大排列的特征向量 Z_1, Z_2, \dots, Z_k . 那么相应的在 $\tilde{\phi}_b$ 空间中的最优特征向量为 $Y_j = P_1 Z_j, j = 1, 2, \dots, l$. 其中 $l = \text{rank}(\tilde{S}_b)$. 一般 $l \leq k$.

3) 取 $P_2 = (\tau_{q+1}, \tau_{q+2}, \dots, \tau_m)$, 重新得到相应 $\tilde{\phi}_b^\perp$ 空间内的类内散布矩阵 $\tilde{S}_w = P_2^T \tilde{S}_w P_2$, 计算得到 \tilde{S}_w 的正交特征向量集, 并且将这些特征向量按对应特征值从小到大的顺序排列, 取前 $d-l$ 个特征向量, 由此我们可以得到在 $\tilde{\phi}_b^\perp$ 空间中的最优特征向量为 $Y_j = P_2 Z_j, j = l+1, l+2, \dots, d$.

4) 由 $Y_j = P_1 Z_j, j = 1, 2, \dots, l$ 和 $Y_j = P_2 Z_j, j = l+1, l+2, \dots, d$ 组成特征向量集 $(Y_1, \dots, Y_l, Y_{l+1}, \dots, Y_d)$.

4 实验结果

本文采用 ORL^[12]人脸数据库和 NUST603 人脸数据库来验证该算法. 最终经特征提取之后的图像数据采用 KNN 分类器实现分类, 其中 $K = 3$.

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, 每个人由 10 幅图像所构成. 这些人

脸图像为分辨率是 92×112 的灰度图像. 实验中, 图像被预处理成 31×40 维的形式. 预处理之后的部分人脸图像如图 1 所示:



Fig. 1 Some images from ORL face database.

图 1 ORL 人脸数据库部分人脸数据

我们的实验中, 分别采用 ORL 人脸库中每类人脸的前几幅图像构成整个训练样本集, 并且把人脸库中的所有样本均作为检测样本(下同), 得到的最终识别性能表如表 1 所示:

Table 1 Recognition Rate Comparison on ORL Face Database
表 1 在 ORL 人脸数据库上的识别率比较 %

Training Sample/Class	Eigenface	Fisherface	Proposed Method
3	83.25	88.25	88.50
4	90.50	91.25	91.75
5	93.00	93.50	95.75

NUST603 人脸数据库中包含 96 个人, 每人由 10 幅图像构成. 每幅图像的分辨率均为 256×256 . 为了便于识别, 我们首先将图像中的脸像切割出来^[13], 并归一化为 32×32 的标准图像. 图 2 和图 3 分别为 NUST603 人脸数据库的部分原始人脸图像和标准化后的人脸图像.



Fig. 2 Some original images from NUST603 face database.

图 2 NUST603 人脸数据库部分原始人脸图像



Fig. 3 Some normalized images from NUST603 face database.

图 3 NUST603 人脸数据库部分归一化人脸图像

在与 ORL 人脸数据库实验相同的环境下, 我们得到的最终识别率如表 2 所示:

Table 2 Recognition Rate Comparison on NUST603 Face Database

表 2 在 NUST603 人脸数据库上的识别率比较 %

Training Sample/Class	Eigenface	Fisherface	Proposed Method
3	93.75	95.46	96.67
4	95.73	96.65	97.50
5	96.19	97.25	99.06
6	97.60	98.48	99.17

从表 1 和表 2 的实验结果中,可以看出基于 Fisher 鉴别极小准则的人脸识别方法性能优于其他的代数特征提取方法. 基于 PCA 思想的 Eigenface 得到的特征向量是有效重构数据的向量,并不是最有助于分类的特征向量. 而 Fisherface 虽然也采用了 PCA + LDA 的变换过程,但是该组合方法在变换过程中损失了有助于分类的特征信息. 本文提出的基于 Fisher 鉴别极小准则的 PCA + LDA 特征提取方法,有效地得到了降低了原始图像的维数,并且保留了原始图像的分类信息. 特别地,本文的特征提取方法克服了 Yang 和 Xu 的方法中得到的最终特征数据维数必须受类别数限制的不足之处,有效地得到了最终的特征鉴别向量集.

5 结 论

本文在 Fisher 鉴别准则的基础上,提出了一种基于 Fisher 鉴别极小准则的 PCA + LDA 人脸识别方法. Fisher 鉴别极小准则利用了投影数据的最大可分性,同时克服了先前算法中最终得到的特征数据维数必须受类别数限制的局限,从而能取得更多的有助于分类的低维信息,并且给出了一个完整的实现步骤. 在人脸数据库上的实验结果也验证了该算法的有效性.

利用 Fisher 鉴别极小准则提取人脸图像的有效分类数据,当前在线性的情况下取得了良好的分类效果,如何在非线性的情况下取得好的性能,这个是有待我们进一步研究的工作.

参 考 文 献

- 1 W. Zhao, R. Chellappa, A. Rosenfeld, *et al.* Face recognition: A literature survey [R]. National Institute of Standards and Technology, Tech. Rep.: CAR-TR-948, UMD CS-TR-4167R, 2002
- 2 M. Kirby, L. Sirovich. Application of the KL procedure for the characterization of human faces[J]. IEEE Trans. Pattern Anal. Machine Intell., 1990, 12(1): 103~108
- 3 M. Turk, A. Pentland. Eigenfaces for recognition [J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71~86
- 4 P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman. Eigenfaces vs Fisherfaces: Recognition using class specific linear projection [J]. IEEE Trans. Pattern Anal. Machine Intell., 1997, 19(7): 711~720
- 5 D. L. Swets, J. Weng. Using discriminant eigenfeatures for image retrieval [J]. IEEE Trans. Pattern Anal. Machine Intell., 1996, 18(8): 831~836
- 6 L. F. Chen, H. Y. M. Liao, J. C. Lin, *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
- 7 R. Huang, Q. S. Liu, H. Q. Lu, *et al.* Solving the small sample size problem of LDA [C]. The Proc. Int'l Conf. Pattern Recognition, Quebec, Canada, 2002. 29~32
- 8 J. Yang, J. Y. Yang. Why can LDA be performed in PCA transformed space [J]. Pattern Recognition, 2003, 36(2): 563~566
- 9 J. C. Bezdek, J. Keller, R. Krishnapuram. Fuzzy Models and Algorithms for Pattern Recognition and Images Processing [M]. San Francisco, CA: Kluwer Academic Publishers, 1999
- 10 J. M. Keller, M. R. Gray, J. A. Givens. A fuzzy k-nearest neighbor algorithm [J]. IEEE Trans. Syst., Man Cybernet, 1985, 15(4): 580~585
- 11 Xu Yong, Lu Jianfeng, Yang Jingyu. A novel method on face recognition [J]. Journal of System Simulation, 2004, 16(1): 14~16 (in Chinese)
(徐勇, 陆建峰, 杨静宇. 一种人脸图像自动识别方法及实验 [J]. 系统仿真学报, 2004, 16(1): 14~16)
- 12 ORL face database [OL]. <http://www.uk.research.att.com/facedatabase.html>, 2005-08-25
- 13 Z. Jin, J. Y. Yang, Z. S. Hu, *et al.* Face recognition based on the uncorrelated discriminant transformation [J]. Pattern Recognition, 2001, 34(7): 1405~1416

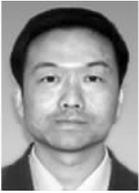


Zheng Yujie, born in 1977. Ph. D. candidate in Nanjing University of Science and Technology, Nanjing, China. His current research interests include pattern recognition, artificial intelligence and so on.
郑宇杰, 1977 年生, 博士研究生, 主要研究方向为模式识别、人工智能等.



Yang Jingyu, born in 1941. Professor and Ph. D. supervisor in Nanjing University of Science and Technology, Nanjing, China, senior member of CCF. His current research interests include pattern recognition, robot vision, image processing, data fusion, artificial intelligence and so on.

杨静宇, 1941 年生, 教授, 博士生导师, 中国计算机学会高级会员, 主要研究方向为模式识别、机器视觉、图像处理、数据融合、人工智能等.



Xu Yong, born in 1972, Ph. D. and lecturer. His current research interests include pattern recognition, face recognition and detection, kernel trick application and so on.

徐勇, 1972年生, 博士, 讲师, 主要研究方

向为模式识别、人脸识别与检测、核方法应用等。



Yu Dongjun, born in 1975. Ph. D. and associate professor in Nanjing University of Science and Technology, Nanjing, China. His current research interests include pattern recognition, fuzzy theory, neural network and so on.

於东军, 1975年生, 博士, 副教授, 主要研究方向为模式识别、模糊理论、神经网络等。

Research Background

This research work is supported by the National Natural Science Foundation (60503026, 60472060, 60572034) and the Scientific Research Development Foundation of Nanjing University of Science and Technology (AB96125).

Feature extraction is one of the key steps in face recognition. Within the past two decades, numerous feature extraction algorithms have been proposed. Among these feature extraction algorithms, linear discriminant analysis (LDA) is widely used. However, statistical learning methods including the LDA-based algorithms often suffer from the "small sample size (SS) problem". Up to now, principal component analysis (PCA) plus LDA is a popular method to address this problem. Unfortunately, number of final features is limited by the rank of between-class scatter matrix. In this paper, a new feature extraction method called Fisher discriminant minimal criterion is proposed to improve the performance of feature extraction. Conventional Fisher discriminant criterion is inversed and null space of between-class scatter matrix is defined in this algorithm. Therefore, limitation of final eigenvectors' dimensions determined by class number is overcome and more effective classification information can be achieved. Experimental results on ORL face database and NUST603 face database demonstrate the effectiveness of the proposed algorithm.