

doi:10.11835/j.issn.1000-582X.2014.04.007

带权稀疏 PCA 算法及其应用

宣士斌

(1. 四川大学 计算机学院, 成都 640005; 2. 广西民族大学 信息科学与工程学院, 南宁 530006)

摘要:主成份分析(PCA)算法是特征提取的重要方法之一,由于其本身没有提供更多的分类信息,直接在其上进行识别效果往往并不理想。为了提取 PCA 特征值中有利于识别的特征信息,提出一种带权稀疏 PCA 算法。它利用基本 PCA 算法实现去噪功能,利用 Lagrange 乘子方法求得使 PCA 特征空间中类内距离最小,类间距离最大的一组权值,并利用稀疏 PCA (SPCA) 算法解决维数约简和保留小特征值对应的特征向量所含的分类信息。在公开人脸数据库上对该算法进行测试,实验结果表明该算法不仅运行速度快,而且具有较高的正确识别率。

关键词:成份分析;线性判别分析;套索;带权稀疏主成份分析

中图分类号: TP391.4

文献标志码: A

文章编号: 1000-582X(2014)04-046-06

Weighted sparse principal component analysis and it's application

XUAN Shibin

(1. College of Computer, Sichuan University, Chengdu 640005, China; 2. College of Information Science and Engineering, Guangxi University for Nationalities, Nanning 530006, China)

Abstract: The principal component analysis(PCA) is one of the important methods for feature extraction, but it can't provided more classification information by itself. In order to pick up feature information in favor of recognition from PCA eigenvector, a weight sparse principal component analysis is proposed in the paper. It achieves image de-noising function by using primitive PCA algorithm, acquires the group of weight values which are able to maximize within-class distance and minimize between-class distance in PCA feature space by utilizing Lagrange multiplier, and finishes dimension reduction by using sparse PCA (SPCA) to retain effectively some classification information of eigenvectors with little eigenvalue. In the end, the proposed algorithm is tested on an all-known public face database. The experiment results indicate the proposed algorithm has not only faster running speed but also better rate of recognition.

Key words: PCA; linear discriminat analysis; lasso; weighted spatse PCA

基于概率的子空间方法是人脸识别算法中的一种重要方法,尽管出现了很多其他识别方法如 SVM,但子空间方法仍然受到研究者的关注,并有着广泛的应用前景。典型代表有 PCA^[1-2], LDA^[3-5] 等,以及将它们与核方法结合而得到的 KPCA^[6], KLDA^[7] 等,也有人将其与小波、神经网络等相结合

得到一些较好的结果。但这其中最基本的仍然是 PCA 算法,其他算法如 LDA、Bayes 由于自身计算复杂度等原因,一般都是先运行 PCA 得到相应的特征空间,然后在此特征空间上运行这些算法。目前关于 PCA 研究主要集中在 2 个方面,一方面是 PCA 理论研究^[8-9],针对特征向量的生成方法,特别是奇

收稿日期:2013-11-12

基金项目:国家自然科学基金委员会与中国民用航空局联合资助重点项目(60736046);广西自然科学基金(2012GXNSFAA053227)

作者简介:宣士斌(1964-),男,四川大学博士研究生,主要从事数字图像处理与识别及智能计算方向研究,

(E-mail)shibin1997@126.com。

异值分解(SVD),希望能用尽可能少的特征向量代表更多的数据信息。另一方面是利用 PCA 进行维数约简和特征抽取,在模式识别中,由于识别对象维数往往很高(图像),为了提高识别速度必须进行维数约简。PCA 进行维数约简时,它将作用对象(图像)从高维数据空间(如 $256 \times 256 = 65536$)压缩到指定维数(20),虽然在压缩过程中会丢失一些信息,但一般认为这些信息是不太重要的较小数据。但一些研究表明这部分丢失的数据仍然含有部分分类信息,由于数据过度集中且在产生 PCA 特征空间时没有考虑分类信息,使得直接在 PCA 特征空间进行分类识别效果并不理想。在 PCA 空间上的 LDA、Bayes 等方法就是希望结合分类信息来提高识别质量。而 KPCA 则是利用核技术将低维映射到高维空间以此来提高特征空间所含的分类信息。提出一种新的融合分类信息和维数约简的带权稀疏 PCA (WSPCA)方法,该方法根据分类信息对 PCA 特征空间中的特征值进行修正,并利用 SPCA 方法进行维数约简,使其含有更多的分类信息。在人脸数据库上的实验结果表明该方法能有效提高识别的正确率,并且有较快的计算速度。

1 算法基础

以二维人脸识别为例,来说明算法的基本理论。2D 人脸图像可被看成是图像空间的一个向量,人脸图像样本 $\{x_i\}$ 集合可以用一个 $N \times M$ 的矩阵 $\mathbf{X} = [x_1, \dots, x_M]$, 这里 N 是图像的象素点数, M 是样本个数。每幅图像属于某一人脸类 $\{X_1, \dots, X_L\}$, $l(x_i)$ 是 x_i 的类标志。当要识别的对象被输入时,人脸识别的任务就是在数据库中找它的类标志。为了引出研究方法,这里首先介绍 LDA 算法及其导出方法。

1.1 LDA 简介

LDA 希望寻找这样的一个子空间,它通过最大化类间分散矩阵 S_b , 最小化类内分散矩阵 S_w , 来最大化不同类的差异度。 S_b 和 S_w 定义如下:

$$S_w = \sum_{i=1}^L \sum_{x_{ki}} (x_k - m_i)(x_k - x_i)^T$$

$$S_b = \sum_{i=1}^L n_i (m_i - m)(m_i - m)^T$$

这里, m_i 是个体 X_i 类的平均脸, n_i 是类 X_i 的样本数。

LDA 子空间由满足下列条件的向量 W 所张成的空间。

$$W = \operatorname{argmax} \left| \frac{W^T S_b W}{W^T S_w W} \right|, \quad (1)$$

所以 W 能由 $S_w^{-1} S_b$ 的特征向量组成。计算的特征向量,等于同时对角化 S_w 和 S_b 。首先计算 S_w 的特征向量矩阵 Φ 和特征值矩阵 Θ 。然后,投影类中心到 $\Phi \Theta^{-1/2}$ 上。所以 S_b 转换为 $K_b = \Theta^{-1/2} \Phi^T S_b \Phi \Theta^{-1/2}$ 。然后,计算 K_b 的特征向量矩阵 Ψ 和特征值矩阵 Λ , LDA 的投影向量为: $W = \Phi \Theta^{-1/2} \Psi$ 。识别时,线性判别函数定义为

$$d(T) = W^T (T - P).$$

选择使 $\|d\|$ 最小的类作为待识别人脸所属的类。

从上面的 LDA 特征向量的计算过程可以看出, LDA 它的本意是针对原始训练样本空间的,希望得到一个与 PCA 算法一样的一组特征向量,然后用这组得特向量将原始数据转换到特征子空间来,从而达到在维数约简的同时使得类内间的距离尽可能近,而类间距离尽可能远目的。但由于 S_w 和 S_b 的维数一般都比较高,使得直接用在原始人脸数据集上比较困难,且可能遇到对奇异矩阵求逆的困难,因此一般先通过 PCA 算法获取 PCA 特征空间,然后在此空间上运行 LDA 算法。由于 PCA 子空间在获取时已经略去了部分信息,因此,人们指出零子空间的存在并研究出了一些消去零子空间的方法,也有些研究者,直接作用于原始人脸数据的 LDA 算法^[10],并且取得了较好结果。

1.2 PCA 特征空间维数约简

在具体应用中,由于 PCA 算法将学习对象所携带的信息 80% 以上集中到对应的特征值大的几个特征向量上,特征值较小的部分由一些对识别有用的信息和噪声组成,一般情况下,可以认为 5% 最小特征值对应噪声,在去除 5% 最小特征值所对应的特征向量后,剩下的 95% 的相对较大的特征值对应的特征向量组成的矩阵作为映射,这样可充分利用原始学习样本所携带的有利于识别的信息。但带来的问题是用来作为映射的特征向量的个数变得较多。特别是,当用于学习的样本较多时,对识别速度有较大影响,为此有必要进一步降低维数,即减少用于作为映射的特征向量个数。这方面工作也有不少研究成果,如 ScotLass^[11]法、SPCA^[9]法、最新的成果是文献中提出的非负装载方法来求解主成份^[12]。这里选择 SPCA(sparse principal component analysis)法来实现 PCA 子空间的维数约简。为此在下面简要介绍 SPCA 技术。

PCA 寻求的是获得最大方差的原始变量的线性组合,PCA 可以通过已中心化的数据矩阵 X 的奇异值分解(SVD)实现。 X 的 SVD 分解可表示为

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T. \quad (3)$$

\mathbf{U} 为单位长度的主成份(PCs), \mathbf{V} 的列是主成份的相应的装载, 第 i 个 PC 的方差是 $D_{i,i}^2$, $\mathbf{D} = \text{diag}\{d_1, d_2, \dots\}$ 。

一种直接 sparse PCA 方法可表述如下

$\forall_i, Y_i = U_i d_i$ 考虑下列优化问题

$$\hat{\beta} = \arg \min_{\beta} \{ |Y_i - X\beta|^2 + \lambda |\beta|^2 + \lambda_1 |\beta_1| \}, \quad (4)$$

$\hat{v}_i = \frac{\hat{\beta}}{|\hat{\beta}|}$ 是对 v_i 的近似, $\mathbf{X}\hat{v}_i$ 是第 i 个近似主成份。

上式也称为 naive elastic net^[9]。很清楚, 足够大的 λ_i 可以得出一个 sparse \hat{v}_i 。给定一个固定的 λ , 利用 LARS-EN 算法^[9] 对于所有 λ_1 上式能有效求解。所以我们可以很容易选择一个第 i 个主成份的 sparse 逼近。所以直接 sparse PCA 方法可以分 2 步实现: 首先执行标准 PCA; 再利用式(4) 求出各主成份相应的 sparse 逼近。

文献[9]同时提出了一种自含式 sparsePCA 算法, 记为 SPCA。该算法初始化 $\alpha = V[1:k]$; 对每个 $j=1, 2, \dots, k$, 求解下式

$$\beta_j = \arg \min_{\beta^*} \{ \mathbf{X}^T \mathbf{X} + \lambda \} \lambda^* - 2\alpha_j^T \mathbf{X}^T \mathbf{X} \beta^* + \lambda_{1,j} |\beta^*|_1, \quad (5)$$

然后作 $\mathbf{X}^T \mathbf{X} \beta$ 的 SVD 分解, 记为 $\mathbf{U}\mathbf{D}\mathbf{V}^T$, 再让 $\alpha = \mathbf{U}\mathbf{V}^T$, 重复这个过程直至 β 收敛。最后令: $\hat{v}_j = \beta_j / |\beta_j|$, $j=1, \dots, k$ 。

另外从(3)式, 得 $\mathbf{X}^T \mathbf{X}$ 的特征分解:

$$\mathbf{X}^T \mathbf{X} = \mathbf{V}\mathbf{D}^2\mathbf{V}^T, \quad (6)$$

称特征向量 v_j (\mathbf{V} 的列) 为 \mathbf{X} 的主成份 (或 Karhunen-Loeve) 方向。第一个主成份方向 v_1 满足: 在所有的 \mathbf{X} 列向量的规范化线性组合中, $z_1 = \mathbf{X}v_1$ 有最大的样本方差, 即

$$\text{Var}(z_1) = \text{Var}(\mathbf{X}v_1) = \frac{d_1^2}{N}. \quad (7)$$

事实上

$$z_1 = \mathbf{X}v_1 = u_1 d_1, \quad (8)$$

由此导出变量 z_1 称为 \mathbf{X} 的第一个主成份, 所以 u_1 是第一个规范化主成份。

事实上, 在利用 PCA 算法进行特征抽取时, 所用的抽取矩阵 (或转换矩阵) 即为此处的 \mathbf{V} , 它利用 \mathbf{V} 作为转换算子将样本空间转换到 PCA 特征子空间。因此, 在模式识别的特征抽取中, 感兴趣的是 PCA 算法所得的 \mathbf{V} , 上面介绍的 sparse PCA 算法在约简主成份空间维度的同时, 也实现了对 \mathbf{V} 的维度

约简。但上述算法是针对整个 \mathbf{V} 的维度约简, 没有考虑去噪, 为达到去噪作用, 在执行上述 SPCA 算法之前, 需要对 \mathbf{X} 进行去噪预处理。即先对 \mathbf{V} 执行 PCA 算法, 得到其主成份装载向量 \mathbf{V} , 按 95% 选取大特征值所对应的装载向量, 其结果仍然记为 \mathbf{V} 。利用式(6)和(8)可令式(4)中的 Y_i 为: $Y_i = \mathbf{X}v_i$, 然后利用直接 sparse PCA 方法求得相应的 sparse \mathbf{V} 。但在式(4)、(5)中的 \mathbf{X} 还是原有的含噪声数据, 为了去除噪声对进一步估计的影响, 需用去噪以后的 \mathbf{V} 还原数据矩阵 \mathbf{X} 。由前面的讨论可知

$$u_i = \mathbf{X}u_i / d_i, \quad (9)$$

从而可求出相应的主成份矩阵 \mathbf{U} , 最后利用式 $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ 求出 \mathbf{X} , 此时的 \mathbf{X} 可看成是已对原数据经过去噪处理的数据。最后用此 \mathbf{X} 通过 SPCA 算法求得 \mathbf{X} 即得特征抽取中所需要的转换矩阵 \mathbf{V} 。

2 带权主成分分析

分析 LDA 过程会发现, 最初构造的类间分散矩阵 \mathbf{S}_b 和类内分散矩阵 \mathbf{S}_w , 完全是为了获最 LDA 特征向量 \mathbf{W} , 但如果不是为了获取向量, 构造这样的矩阵也就没有意义了。PCA 特征空间已经是通过维数约简得到的, 原始人脸数据信息已被集中在几个特征值上。再通过 LDA 进行二次维数约简反而使得特征值更加集中, 从而使分类更加困难, 这可能就是一些文献中提到的在某些情况下 LDA 的识别率还没有 PCA 高的原因所在。其实在得到 PCA 子空间后, 只需将各特征值中有利于增大类间距离而缩小类内距离的有用信息提取出来即可。即希望获得每个特征值中最有价值的部分, 也即希望计算出每个特征值中最有价值的信息占总信息量的百分比。记这个百分比向量为 $\mathbf{u} = \{u_1, \dots, u_K\}$, 满足: $\sum_{k=1}^K u_k = 1$ 。这里 K 为 PCA 子空间的维度, 即 PCA 特征向量个数。为此, 用类内分散度和类间分散度来代替在 LDA 中定义类间分散矩阵 \mathbf{S}_b 和类内分散矩阵 \mathbf{S}_w , 并定义

$$\mathbf{S}_w(\mathbf{u}) = \sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^K (x_{i,j,k} - m_{i,k})^2 u_k^2; \quad (10)$$

$$\mathbf{S}_b(\mathbf{u}) = \sum_{i \neq j} \sum_{k=1}^K (m_{j,k} - m_{i,k})^2 u_k^2. \quad (11)$$

目标是寻找使 $\mathbf{S}_w(\mathbf{u})$ 尽可能小, 而 $\mathbf{S}_b(\mathbf{u})$ 尽可能大的 \mathbf{u} 。为此, 利用 Lagrange 乘子方法构造函数如下

$$J(\mathbf{u}) = \mathbf{S}_w(\mathbf{u}) - \lambda_1 \mathbf{S}_b(\mathbf{u}) + \lambda_2 \left(\sum_{k=1}^K u_k - 1 \right). \quad (12)$$

将式(10)、(11)代入式(12)可得

$$\begin{aligned}
 J(\mathbf{u}) &= \sum_{i=1}^M \sum_{j=1}^N \sum_{k=1}^K (x_{i,j,k} - m_{i,k})^2 \mathbf{u}_k^2 - \lambda_1 \sum_{i \neq j} \sum_{k=1}^K (m_{j,k} - m_{i,k})^2 \mathbf{u}_k^2 + \lambda_2 \left(\sum_{k=1}^K \mathbf{u}_k - 1 \right) = \\
 &= \sum_{k=1}^K \sum_{i=1}^M \sum_{j=1}^N (x_{i,j,k} - m_{i,k})^2 \mathbf{u}_k^2 - \lambda_1 \sum_{k=1}^K \sum_{i \neq j} (m_{j,k} - m_{i,k})^2 \mathbf{u}_k^2 + \lambda_2 \left(\sum_{k=1}^K \mathbf{u}_k - 1 \right) = \\
 &= \sum_{k=1}^K \left(\sum_{i=1}^M \sum_{j=1}^N (x_{i,j,k} - m_{i,k})^2 - \lambda_1 \sum_{i \neq j} (m_{j,k} - m_{i,k})^2 \right) \mathbf{u}_k^2 + \lambda_2 \left(\sum_{k=1}^K \mathbf{u}_k - 1 \right). \quad (13)
 \end{aligned}$$

对式(13)两边关于 $\mathbf{u}_k, k=1, \dots, K$, 求导并令其为零得

$$2 \left(\sum_{i=1}^M \sum_{j=1}^N (x_{i,j,k} - m_{i,k})^2 - \lambda_1 \sum_{i \neq j} (m_{j,k} - m_{i,k})^2 \right) \mathbf{u}_k + \lambda_2 = 0, k = 1, \dots, K. \quad (14)$$

从而有

$$\mathbf{u}_k = \frac{-\lambda_2}{2 \left(\sum_{i=1}^M \sum_{j=1}^N (x_{i,j,k} - m_{i,k})^2 - \lambda_1 \sum_{i \neq j} (m_{j,k} - m_{i,k})^2 \right)}, k = 1, \dots, K. \quad (15)$$

考虑到 λ_2 与 k 值无关,且在识别过程中可以将公因子提取出来,其取值不影响识别,为此不妨取 $\lambda_2 = -2$. 所以式(15)可简化为

$$\mathbf{u}_k = \frac{1}{\sum_{i=1}^M \sum_{j=1}^N (x_{i,j,k} - m_{i,k})^2 - \lambda_1 \sum_{i \neq j} (m_{j,k} - m_{i,k})^2}, k = 1, \dots, K. \quad (16)$$

由于 $0 < \mathbf{u}_k < 1, k=1, \dots, K$, 不妨设

$$\mathbf{u}_k = \left| \frac{1}{\sum_{i=1}^M \sum_{j=1}^N (x_{i,j,k} - m_{i,k})^2 - \lambda_1 \sum_{i \neq j} (m_{j,k} - m_{i,k})^2} \right|, k = 1, \dots, K. \quad (17)$$

再根据

$$\sum_{k=1}^K \mathbf{u}_k = 1, \quad (18)$$

可求得 λ_1 , 并代入式(17)即可求得所有 $\mathbf{u}_k, k=1, \dots, K$.

求出 $\mathbf{u}_k, k=1, \dots, K$ 后,可以利用它从 PCA 特征子空间中分离出所需要的分类信息.由于 \mathbf{u} 要作用于 PCA 特征子空间的所有元素,所以不妨将 \mathbf{u} 整合到 PCA 的转换矩阵中.设 PCA 的转换矩阵为 $\mathbf{G} = \{g_1, \dots, g_K\}$, 与 \mathbf{u} 整合后得: $\mathbf{G}' = \{u_1 g_1, \dots, u_K g_K\}$, 为方便起见,仍记为 \mathbf{G} . 在进行具体识别时式(2)就变成了

$$d(\mathbf{T}) = \mathbf{G}^T (\mathbf{T} - \mathbf{P}), \quad (19)$$

因此,结合所介绍的维数约简方法,可得稀疏带权 PCA 算法(WSPCA),该算法对给定的训练集,可按给定的维数要求求出相应的转换矩阵 \mathbf{G} . 具体算法描述如下

算法 weighted sparse principal component analysis(WSPCA)

Input:

\mathbf{X} // training data set

\mathbf{M} // the number of individual

\mathbf{N} // the number of training sample per individual

\mathbf{K} // the number of eigenvectors

in sparsePCA

Output:

\mathbf{G} // the transformation matrix of WSPCA

1) $(\mathbf{V}, \mathbf{D}) = \text{PCA}(\mathbf{X}, \mathbf{M}, \mathbf{N}, 95\%)$; // calculating the 95% eigenvectors and eigenvalues of $\mathbf{X}^T \mathbf{X}$

2) $u_i = Xv_i/d_i$ // computing the principal components of matrix \mathbf{X}

3) $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ // denoising

4) $\mathbf{G}' = \text{SPCA}(\mathbf{X}, \mathbf{K})$ // computing the transformation matrix of SPCA

5) $\mathbf{Y} = \mathbf{G}'\mathbf{X}$; // computing PCA subspace

6) $m_i = \text{mean}(Y_i), i=1, \dots, M$ // computing the mean of every class

7) compute $\mathbf{u}_k, k=1, \dots, K$ // where \mathbf{K} is columns of \mathbf{G}

8) $g_i = u_i \times g_i, i=1, \dots, K$ // where $\mathbf{G} = [g_1, \dots, g_K]$

9) output \mathbf{G}

上述算法中,除第 4 步外,每步都可以通过直接计算得到所需要的结果,因此算法是收敛的,由于算法第 4 步所得 \mathbf{G}' 中的列向量是正交向量序列,由第 7 步计算出的 u_k 是一常量,因此最后输出的 \mathbf{G} 的列向量也是正交向量序列。

3 应用举例

3.1 在 UIC^[13] 数据库上新算法的测试情况

将待测样本集分为 7 个待测集,分别是不带墨
镜下的除训练样本外的 3 种表情和带墨镜下的所有
4 种表情,分别用算法 PCA、LDA、KPCA、WSPCA
在总特征空间大小为 15、20、25、30 的情况下进行测
试,测试结果如图 1 所示。

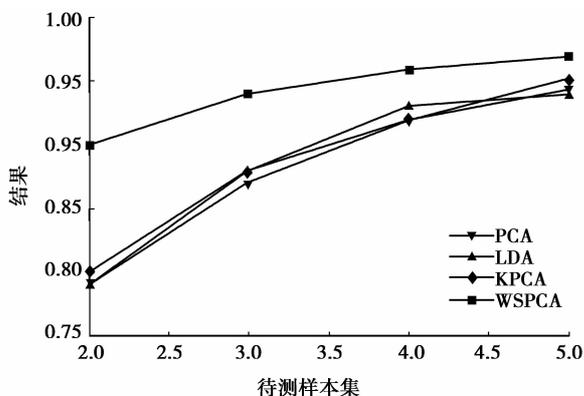


图 1 UIC 数据库上新算法的测试

由于 F_JDLDA 算法的特征空间大小由算法自
身根据学习情况设定,在本例中该算法自动设定特
征空间大小为 15,其识别率仅为 70%。从上图可以
看出,在总特征空间大小分别为 15、20、25、30 4 种
情况下,所提出的算法与另外 4 种算法相比有更好的
识别率,特别是当有较小的特征空间时,越显出所
提出算法的优越性。当特征空间逐渐变大时各方法
的识别率趋于一致,但所提算法还是比其他算法的
识别率高。

3.2 在 AR 数据库上新算法的测试情况

为了进一步说明 WSPCA 算法的可行性,在
AR 数据库上进行实验。为此先对 AR 数据库中的
图片进行简单的预处理。将原 RGB 图片转换成
YcBcR 彩色模型,利用肤色确定人脸位置,再从
上面和左右 3 个方向向中间搜索按同一大小确定脸
部位置矩形,再将该位置矩形中的脸部取出转换
成灰度值图像并按 2:1 压缩存贮。没有按眼睛位
置进行对准及旋转处理。在本次实验中对上述 4
种算法按不同的训练样本数和不同的个体数进行
比较,所得结果如图 2 和图 3 所示,从实验结果
来看提出的算法是有效的。

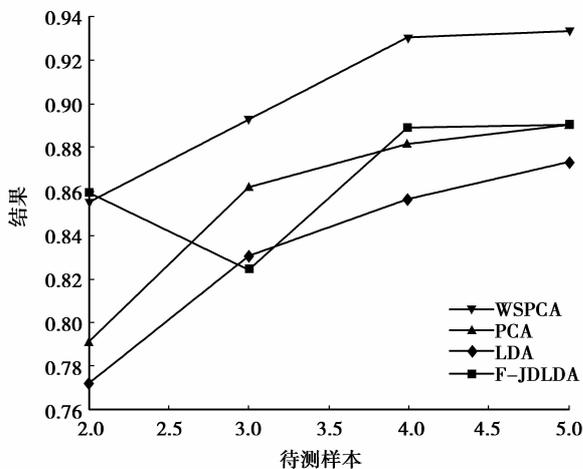


图 2 5 男 5 女个体随训练样本数的变化识别率

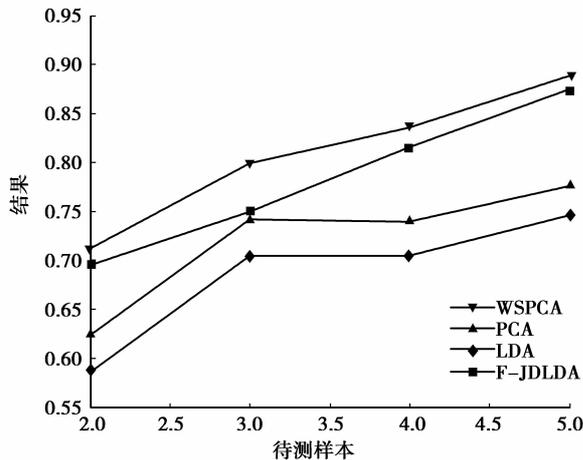


图 3 10 男 10 女随训练样本数的变化识别率

图 2 描述了针对 5 男 5 女共 10 个个体随各
个体训练样本数的变化识别率的变化情况,从图 2
中看出,WSPCA、PCA、LDA 3 个算法的识别率
随训练样本数的增加而稳步增加,但 F-JDLDA
算法却波动很大,当各个体训练样本只有 2 个
时它是最好的,但当各个体训练样本有 3 个时
它的识别率反而降低。但总的来说,当各个体
训练样本大于等于 3 时,可以看出所提新算法
相对于其他 3 种算法有较强的优势。图 3 描
述了针对 10 男 10 女共 20 个个体随各个体
训练样本数的变化识别率的变化情况,从图 3
中看出,在这种情况下,F-JDLDA 算法随个
体训练样本数增加稳步上升,反而 PCA、LDA
有小部分波动,同时可以看出所提新算法相
对于其他 3 种算法有较好的识别率。且从 2
种情况下,所提新算法 WSPCA 都有较好的
稳定性,其识别率随个体训练样本数增加
稳步上升。

4 结 论

模式识别中的 PCA 算法所得的特征子空间没有提供过多的分类信息,且为了约简维数,省略了小特征值所对应的特征向量,而这部分特征向量同样也含有相当多的分类信息。为了达到去噪、维数约简和增强特征子空间的分类信息,提出算法 WSPCA 将 PCA 算法、SPCA 算法、LDA 算法有效地揉合到一起,利用 PCA 算法达到去噪作用,利用 SPCA 算法达到维数约简效果,利用 LDA 算法的基本思想完成可分功能。从实验效果看,所提出的算法有较好的正确识别率和稳定性。下一步工作,可以从 2 个方面继续研究,一是探索更好的维数约简方法,二是将该算法推广到三维模式识别和其他具体应用中。

参考文献:

- [1] Turk, Pentland. Eigenfaces for recognition [J]. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86.
- [2] Zafeiriou, Petrou. Nonlinear non-negative component analysis algorithms [J]. *IEEE transactions on image processing*, 2010, 19(4): 1050-1066.
- [3] Zhao Y. Incremental linear discriminant analysis for face recognition [J]. *IEEE transactions on systems, man, and cybernetics part B: cybernetics*, 2008, 38(1): 210-221.
- [4] Cai. SRDA: an efficient algorithm for large-scale discriminant analysis [J]. *IEEE transactions on knowledge and data engineering*, 2008, 20(1): 1-12.
- [5] Zhu. Subspace discriminant analysis [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2006, 28(8): 1274-1286.
- [6] Suter. Incremental principal component analysis [J]. *IEEE transactions on image processing*, 2007, 16(6): 1662-1674.
- [7] Liu Q S, Huang R, Lu H Q. Face recognition using kernel based fisher discriminant analysis [C] // *Proceedings of International Conference on Automatic Face and Gesture Recognition*, Washington: IEEE, 2002: 197-201.
- [8] Zou, Hastie. Regularization and variable selection via the elastic net [J]. *Journal of Royal Statistical Society Series B*, 2005(67): 301-320.
- [9] Zou H, Hastie, Tibshirani. Sparse principal component analysis [J]. *Journal of Computational and Graphical Statistics*, 2006(15): 262-286.
- [10] Lu, Plataniotis, Venetsanopoulos. regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition [J]. *Pattern Recognition Letters*, 2005, 26(2): 181-191.
- [11] Trendafilov, Jolliffe. DALASS: Variable selection in discriminant analysis via the LASSO [J]. *Computational Statistics & Data Analysis*, 2007(51): 3718-3736.
- [12] Liipovetsky. PCA and SVD with nonnegative loading [J]. *Pattern Recognition*, 2009(42): 68-76.
- [13] Blake C, Keogh E, Merz C. UCI repository of machine learning database, 1998 [EB/OL]. <http://www.ics.uci.edu/mllearn/MLRepository.html>, 1998.

(编辑 侯 湘)