中国图象图形学报
Journal of Image and Graphics

2 维不相关鉴别矢量集算法

林玉娥　顾国昌　刘海波
（哈尔滨工程大学计算机科学与技术学院,哈尔滨　150001）

摘　要　在人脸识别算法中,已有的计算不相关鉴别矢量集的算法均是基于图像向量模型的,因而将遇到所谓的小样本问题,而且由于采用迭代求解方式,算法运算速度缓慢,为此提出了一种新的求取不相关鉴别矢量集的算法,即一种基于图像矩阵模型的 2 维不相关鉴别矢量集算法。算法由于采用了图像矩阵模型,解决了小样本问题,通过对类内散布矩阵的白化变换,使得推广的 2 维线性鉴别分析模型具有类似的 2 维主成分分析模型的形式,从而将两种算法的模型有效地联系起来,进而可以非迭代地求得 2 维不相关鉴别矢量集,不但求解速度快且数值解稳定。在 ORL 和 Yale 人脸库上的实验结果表明,该算法不但减少了计算时间,同时也提高了识别率,为求解不相关鉴别矢量集提供了一个新的思路。
关键词　2 维不相关鉴别矢量集　图像矩阵模型　白化变换　散布矩阵　非迭代
中图法分类号: TP391.4　　**文献标识码**: A　　**文章编号**: 1006-8961(2009)05-944-06

An Algorithm of Two-dimensional Uncorrelated Discriminant Vectors

LIN Yu-e, GU Guo-chang, LIU Hai-bo
(*School of Computer Science and Technology, Harbin Engineering University, Harbin* 150001)

Abstract　All of the existed algorithms in face recognition, which can obtain uncorrelated discriminant vectors, are based on image vector model, so they encounter so called “small sample size” problem. These algorithms, which are solved using recursive methods, require much computation time. So a new algorithm is proposed in this paper, which is called Two-dimensional Uncorrelated Discriminant Vectors based on an image matrix model. The new algorithm solves small sample size problem through whitening transform of within-class scatter matrix, which makes the model of extended Two-dimensional Linear Discriminant Analysis have similar form of Two-dimensional Principal Component Analysis model. Thus two algorithms were combined effectively, uncorrelated discriminant vectors can be obtained non-recursively. The new method computes fast while maintaining numerical stability. The numerical experiments on facial databases of ORL and Yale show that the proposed method has not only reduced the computation complexity but also achieved higher recognition accuracy, providing new thought on how to obtain uncorrelated discriminant vectors.
Keywords　two-dimensional uncorrelated discriminant vectors, image matrix model, whitening transform, scatter matrix, non-recursively

1　引　言

在人脸识别方法中,基于 Fisher 准则的线性鉴别分析(LDA)[1]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主流。但是在求取鉴别矢量集时,由于矩阵 $\boldsymbol{S}_{\mathrm{w}}^{-1}\boldsymbol{S}_{\mathrm{b}}$ 不再是对称矩阵,因此所求得的鉴别矢量两两之间不再正交。为了获得基于 Fisher 线性判别准则的相互正交的鉴别矢量集,1975 年 Foley 和 Sammon 提出了 Foley-Sammon 正交鉴别矢量集[2],虽然 Foley-Sammon 正交鉴别矢量集的鉴别矢量是相互正交

收稿日期:2007-10-10;**改回日期**:2008-01-08
第一作者简介:林玉娥(1979～　),女。哈尔滨工程大学计算机科学与技术学院计算机应用技术专业博士研究生。主要从事图像处理、模式识别和人工智能方面的研究。E-mail:linyu_e@126.com

的，但样本在正交鉴别矢量集上的投影系数却是统计相关的，而在模式识别理论中，通常希望所获得的鉴别矢量集相关性越小越好，最好是具有不相关的特征，从而提高算法的识别率。为此，金忠等人提出了一种不相关鉴别矢量集算法，并且通过实验证明了不相关最佳鉴别变换优于 Foley-Sammon 最佳鉴别变换[3]。在此基础上，杨静宇等人进一步说明了对含有 L 个类别的模式识别问题，可最有效地抽取 $L-1$ 不相关鉴别特征[4]。吴小俊等人[5]又对不相关鉴别矢量集算法进行了进一步研究，并提出了一种广义上的不相关鉴别矢量集算法，同时又证明了广义上的不相关鉴别矢量集与文献[3]的方法是等价的[5]，从而丰富了不相关鉴别矢量集算法的理论与应用。不相关鉴别矢量集算法实际上是一种迭代算法，在求取不相关鉴别矢量集时，要反复求逆矩阵，因而算法的计算量是相当大的，对此文献[6]提出了一种加快不相关鉴别矢量集提取的方法，即先对图像进行压缩变换，然后再求取不相关鉴别矢量集，但仍是采用文献[3]的迭代算法，因而并没有真正减少算法的运算量。文献[7]则提出一种非迭代求取不相关鉴别矢量集的不相关空间算法，算法首先求得不相关空间，然后在不相关空间中寻找不相关鉴别矢量集，由于可以非迭代的求得不相关鉴别矢量集，因此降低了运算量，但算法为了避免小样本问题，首先对人脸图像进行了降低分辨率的处理，因而在一定程度上可能损失了具有鉴别意义的信息。

另外上述人脸识别方法均是基于向量模型的，即对于 2 维的图像矩阵首先要转换成 1 维的向量，然后将所有图像向量按次序以列或者行的形式排列，形成训练样本矩阵，再使用相应的计算方法，求取最佳投影矩阵。然而当采用由向量构成的矩阵做线性鉴别分析进行特征提取时，由于实际应用中图像的像素个数往往远大于训练图像样本的个数，由此将遇到小样本问题，从而导致相应的类内散布矩阵奇异。最近，Yang 等人提出基于图像矩阵模型的 2 维线性鉴别分析（2DLDA）方法[8]，是直接对 2 维图像矩阵做线性鉴别分析，这就是说图像矩阵不需要转换成向量，图像的散布矩阵可以由原始图像矩阵直接计算，不但减少了计算相应的鉴别矢量所花费的时间，同时也避免了类内散布矩阵奇异的问题（除非只有一个训练样本）。

受到文献[7]的启发，以 2DLDA 算法为基础，针对已有求取不相关鉴别矢量集算法的求解复杂、计算量大进行了研究，提出一种新的求取不相关鉴别矢量集的算法，即 2 维不相关鉴别矢量集算法（2DUDV）。新算法由于采用了图像模型，从而避免了类内散布矩阵奇异的问题，解决了所谓的小样本问题，通过对类内散布矩阵的白化变换，使得推广的 2DLDA 具有类似的 2 维主成分分析（2DPCA）模型的形式，从而将两种算法的模型有效地联系起来，进而可以非迭代地求得一组不相关鉴别矢量，同时这种类似 2DPCA 模型也避免了由于求解过程中出现非对称阵所引起的数值解不稳定现象。最后，在 ORL 和 Yale 人脸库上的实验结果表明，2DUDV 方法，不但减少了计算不相关鉴别矢量集的时间，同时也提高了识别率。

2　相关内容介绍

假设有 C 个样本类，共有 n 个图像样本，第 i 类有 n_i 个样本，$\boldsymbol{X}_j^i=[\boldsymbol{x}_1^i,\boldsymbol{x}_2^i,\cdots,\boldsymbol{x}_l^i]$ 表示人脸图像集中第 i 类的第 j 个样本，$i=1,2,\cdots,C$，$j=1,2,\cdots,n_i$，$\boldsymbol{X}_j^i$ 为 $m\times l$ 的矩阵，$\boldsymbol{x}_l^i$ 代表列向量，$\overline{\boldsymbol{X}}^i=[\boldsymbol{m}_1^i,\boldsymbol{m}_2^i,\cdots,\boldsymbol{m}_l^i]$ 为第 i 类人脸图像的均值矩阵，$\overline{\boldsymbol{X}}=[\boldsymbol{m}_1,\boldsymbol{m}_2,\cdots,\boldsymbol{m}_l]$ 是人脸图像的总体均值矩阵，则基于图像模型的类间散布矩阵和类内散布矩阵分别为

$$\boldsymbol{S}_{\mathrm{b}}=\frac{1}{n}\sum_{i=1}^{C}n_i(\overline{\boldsymbol{X}}^i-\overline{\boldsymbol{X}})(\overline{\boldsymbol{X}}^i-\overline{\boldsymbol{X}})^{\mathrm{T}}\tag{1}$$

$$\boldsymbol{S}_{\mathrm{w}}=\frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\boldsymbol{X}_j^i-\overline{\boldsymbol{X}}^i)(\boldsymbol{X}_j^i-\overline{\boldsymbol{X}}^i)^{\mathrm{T}}\tag{2}$$

若将每个样本矩阵中每一列进行头尾相连的顺序排列，即让样本矩阵变为一个维数为 $m\times l$ 的列向量，则对于样本矩阵 $\boldsymbol{X}_j^i$ 的列向量表示为 $\tilde{\boldsymbol{X}}_j^i=[(\boldsymbol{x}_1^i)^{\mathrm{T}},\cdots,(\boldsymbol{x}_l^i)^{\mathrm{T}}]^{\mathrm{T}}$，同理第 i 类人脸图像的均值向量应为 $\tilde{\boldsymbol{m}}^i=\frac{1}{n_i}\sum_{j=1}^{n_i}\tilde{\boldsymbol{X}}_j^i$，总体均值向量为 $\tilde{\boldsymbol{m}}=\frac{1}{C}\sum_{i=1}^{C}\tilde{\boldsymbol{m}}^i$，则基于向量模型的类间散布矩阵和类内散布矩阵分别为

$$\boldsymbol{S}_{\mathrm{b}}=\frac{1}{C}\sum_{i=1}^{C}(\tilde{\boldsymbol{m}}^i-\tilde{\boldsymbol{m}})(\tilde{\boldsymbol{m}}^i-\tilde{\boldsymbol{m}})^{\mathrm{T}}\tag{3}$$

$$\boldsymbol{S}_{\mathrm{w}}=\frac{1}{n}\sum_{i=1}^{C}\sum_{j=1}^{n_i}(\tilde{\boldsymbol{X}}_j^i-\tilde{\boldsymbol{m}}^i)(\tilde{\boldsymbol{X}}_j^i-\tilde{\boldsymbol{m}}^i)^{\mathrm{T}}\tag{4}$$

对于上述不同模型下，总体散布矩阵 $\boldsymbol{S}_{\mathrm{t}}$ 均为

$$\boldsymbol{S}_{\mathrm{t}}=\boldsymbol{S}_{\mathrm{b}}+\boldsymbol{S}_{\mathrm{w}}\tag{5}$$

在人脸识别中，首先要找到一个投影矩阵 $\boldsymbol{W}$，

该投影矩阵不但要使投影后的人脸图像具有较低的维数,而且也要具有较好的可分性。设 $\boldsymbol{X}$ 为人脸图像,将人脸图像 $\boldsymbol{X}$ 投影到 $\boldsymbol{W}$ 上

$$\boldsymbol{Y}=\boldsymbol{W}^{\mathrm{T}}\boldsymbol{X} \tag{6}$$

式中,$\boldsymbol{Y}$ 代表图像 $\boldsymbol{X}$ 的投影特征系数矩阵,$\boldsymbol{W}$ 是按照某种准则函数求得的最佳投影矩阵,下面将分别介绍两种准则函数。

LDA 是以样本的可分性为目标,寻找一组最佳鉴别矢量作为投影轴进行投影变换,使得特征空间的类间散度与类内散度之比最大,其准则函数为

$$J(\boldsymbol{W})=\max_{w_i}\sum_{i=1}^{d}\frac{\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{S}_{\mathrm{b}}\boldsymbol{w}_i}{\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{S}_{\mathrm{w}}\boldsymbol{w}_i} \tag{7}$$

按上式求得的最优鉴别矢量集,实际上就是求解下式广义线性方程特征值与特征向量的问题

$$\boldsymbol{S}_{\mathrm{b}}\boldsymbol{w}_i=\lambda_i\boldsymbol{S}_{\mathrm{w}}\boldsymbol{w}_i(i=1,2,\cdots,l) \tag{8}$$

假设把特征值按降序排列 $\lambda_i\geqslant\lambda_{i+1}$,选择对应前 d(通常 $d<l$)个非零特征值的特征向量 $\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_d$,那么 $\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_d$ 就构成了最优的投影方向矩阵 $\boldsymbol{W}$。若式(7)中的类间散布矩阵与类内散布矩阵按式(1)与式(2)的方法计算,则式(5)即是所谓的基于图像模型的 2DLDA[8],而按式(3)与式(4)的方法计算,则式(7)即是传统的 LDA 方法。

基于主成分分析(PCA)[9]的人脸识别算法,目的是寻找一组最优的单位正交投影轴 $\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_d$,使得由这组投影轴 $\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_d$ 的线性组合所重建的样本和原样本的均方误差最小,即满足如下条件

$$\begin{cases}J(\boldsymbol{W})=\max\sum_{i=1}^{d}\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{S}_t\boldsymbol{w}_i\\ \text{s.t.}\quad \boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{w}_j=\delta_{ij},i,j=1,2,\cdots,d\end{cases} \tag{9}$$

式中,δ_{ij}是 Kronecker 记号,实际上这组最佳投影轴 $\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_d$ 即是总体散布矩阵的前 d 最大特征值所对应的特征矢量。若式(9)中的总体散布矩阵是按式(1)与式(2)的方法计算出的散布矩阵之和,则式(9)即是所谓的基于图像模型的 2DPCA[10],而按式(3)与式(4)的方法计算,则式(9)即是传统的 PCA 方法。

3 2 维不相关鉴别矢量集算法

在模式识别理论中,通常希望所获得的一组鉴别矢量相关性越小越好,最好是具有不相关的特征,从而提高算法的识别率。下面给出不相关性鉴别矢量集的概念。

定义 1 若训练样本为 $\boldsymbol{X}$,一组鉴别矢量为 $\boldsymbol{\Phi}=[\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_r]$,则训练样本在该组鉴别矢量集上的投影为 $\boldsymbol{y}=[\boldsymbol{y}_1,\cdots,\boldsymbol{y}_r]^{\mathrm{T}}=[\boldsymbol{\varphi}_1,\cdots,\boldsymbol{\varphi}_r]^{\mathrm{T}}\boldsymbol{X}$,在投影变换后的所得的特征矢量集 $\boldsymbol{y}$ 中,任意两个分量的统计相关性为 $E\{[\boldsymbol{y}_j-E\boldsymbol{y}_j]^{\mathrm{T}}[\boldsymbol{y}_i-E\boldsymbol{y}_i]\}=\boldsymbol{\varphi}_j^{\mathrm{T}}\boldsymbol{S}_t\boldsymbol{\varphi}_i$,若满足 $\boldsymbol{\varphi}_j^{\mathrm{T}}\boldsymbol{S}_t\boldsymbol{\varphi}_i=0(j\neq i)$,则称这组鉴别矢量 $\boldsymbol{\Phi}=[\boldsymbol{\varphi}_1,\boldsymbol{\varphi}_2,\cdots,\boldsymbol{\varphi}_r]$具有不相关性。

基于图像模型的 2DLDA 是寻找一组最佳鉴别矢量作为投影轴进行投影变换,这组最佳鉴别矢量使得特征空间的类间散度与类内散度之比最大,这组最佳鉴别矢量可按式(8)求得,虽然文献[11]指出广义特征方程的特征值各不相同时,求解不相关鉴别矢量集与经典 Fisher 鉴别分析等价,即如果由式(8)求得的特征值各不相同,那么所得的鉴别矢量之间具有不相关性,但是在实际应用中却难以满足这一条件,另外由于 2DLDA 算法是直接对广义线性方程进行求解特征向量,而 $\boldsymbol{S}_{\mathrm{w}}^{-1}\boldsymbol{S}_{\mathrm{b}}$ 不再是对称矩阵,因此将造成数值解的不稳定。而基于图像模型的 2DPCA 能够非迭代地求得一组不相关且相互正交的最佳投影轴,然而这组最佳投影轴,与分类的目的并不直接相关。因此,本文的 2DUDV 方法就是将 2DLDA 和 2DPCA 的模型联系到一起,利用二者各自的优点,既可以快速地求得一组不相关鉴别矢量,又可以避免 2DLDA 直接对广义特征方程求解引起数值解不稳定的现象。

3.1 2 维不相关鉴别矢量集算法的分析与实现

定理 1 若 $\mathbf{R}^n$ 为 n 维空间,对于任一 $\boldsymbol{x}\in\mathbf{R}^n$,若 $f(\boldsymbol{x})\geqslant 0$,$g(\boldsymbol{x})>0$,令 $A(\boldsymbol{x})=f(\boldsymbol{x})/g(\boldsymbol{x})$,$B(\boldsymbol{x})=[f(\boldsymbol{x})+g(\boldsymbol{x})]/g(\boldsymbol{x})$,则 $A(\boldsymbol{x})$取最大值当且仅当$B(\boldsymbol{x})$取最大值。

证明 因为 $f(\boldsymbol{x})\geqslant 0$,$g(\boldsymbol{x})>0$,所以有 $0\leqslant A(\boldsymbol{x})<+\infty$,而 $B(\boldsymbol{x})=[f(\boldsymbol{x})+g(\boldsymbol{x})]/g(\boldsymbol{x})=f(\boldsymbol{x})/g(\boldsymbol{x})+g(\boldsymbol{x})/g(\boldsymbol{x})=A(\boldsymbol{x})+1$,显然 $\mathrm{d}B/\mathrm{d}A=1>0$,即$B(\boldsymbol{x})$是关于 $A(\boldsymbol{x})$的递增函数,所以当 $A(\boldsymbol{x})$取最大值时,则 $B(\boldsymbol{x})$取最大值。同理,当 $B(\boldsymbol{x})$取最大值时,$A(\boldsymbol{x})$亦取最大值。证毕。

因此,根据定理 1,式(7)与下式等价

$$\begin{aligned}J(\boldsymbol{W})&=\max_{w_i}\sum_{i=1}^{d}\frac{\boldsymbol{w}_i^{\mathrm{T}}S_{\mathrm{b}}\boldsymbol{w}_i+\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{S}_{\mathrm{w}}\boldsymbol{w}_i}{\boldsymbol{w}_i^{\mathrm{T}}\boldsymbol{S}_{\mathrm{w}}\boldsymbol{w}_i}\\&=\max_{w_i}\sum_{i=1}^{d}\frac{\boldsymbol{w}_i^{\mathrm{T}}S_t\boldsymbol{w}_i}{\boldsymbol{w}_i^{T}S_{\mathrm{w}}\boldsymbol{w}_i}\end{aligned} \tag{10}$$

从式(10)的形式上分析,若有变换矩阵 $\boldsymbol{Q}$ 使

$Q^{\mathrm{T}}S_wQ=\mathbf{I}$，则上式可变成 $J(W)=\sum_{i=1}^{d}\frac{w_i^{\mathrm{T}}Q^{\mathrm{T}}S_tQw_i}{w_i^{\mathrm{T}}w_i}$，令 $S_t^Q=Q^{\mathrm{T}}S_tQ$，而此时的 S_t^Q 仍然满足对称性，即 $(S_t^Q)^{\mathrm{T}}=(Q^{\mathrm{T}}S_tQ)^{\mathrm{T}}=S_t^Q$，给定一组限定条件，则可得下面一个新的准则函数模型

$$\begin{cases}J(W)=\max\sum_{i=1}^{d}w_i^{\mathrm{T}}S_t^Qw_i\\ \text{s.t.}\quad w_i^{\mathrm{T}}w_j=\delta_{ij},i\neq j;i,j=1,\cdots,d\end{cases}\tag{11}$$

上面模型的意义就是在使类内散度投影变换为单位矩阵的超球面上，寻找一组鉴别矢量，使得总体散布矩阵的投影值最大，而这组鉴别矢量又将类内散布矩阵的投影值限制为一个常数。对式(11)采用 Lagrange 乘数法，构造函数

$$f(w_1,\cdots,w_d;\lambda_1,\cdots,\lambda_d)=\sum_{i=1}^{d}w_i^{\mathrm{T}}S_t^Qw_i-\sum_{i=1}^{d}\lambda_i(w_i^{\mathrm{T}}w_i-1)\tag{12}$$

式中，λ_i 为 Lagrange 乘子，分别对式(12)中 w_i 和 λ_i 求偏导数，可得：$\partial f/\partial w_i=2S_t^Qw_i-2\lambda_iw_i$，$\partial f/\partial\lambda_i=w_i^{\mathrm{T}}w_i-1$，令 $\partial f/\partial w_i=0$，$\partial f/\partial\lambda_i=0$，得：$S_t^Qw_i=\lambda_iw_i$，$w_i^{\mathrm{T}}w_i=1$。因此，$\lambda_i$ 和 w_i 分别为 S_t^Q 的特征值以及相对应的特征向量，因此所要求的最佳鉴别矢量集 $\{w_1,\cdots,w_d\}$ 就是 S_t^Q 的前 d 个最大特征值所对应的特征向量。不难发现，新的准则函数模型实际上就是 2DPCA 的准则函数模型，只不过，总体散布矩阵发生了改变，S_t^Q 是由样本经 Q 变换后所得的总体散布矩阵，而且 Q 使 $Q^{\mathrm{T}}S_wQ=\mathbf{I}$，也就是对类内散布矩阵 S_w 进行了白化变化，从而限制了类内的散布矩阵的大小。下面给出关于 S_w 的白化矩阵 Q 的求法。由于本文是基于图像模型的，所以类内散布矩阵是非奇异的，而且 S_w 是对称的，因此存在正交阵 U，使得 $S_w=U\Lambda U^{\mathrm{T}}$，其中 Λ 为 S_w 的特征值构成的对角阵，令

$$Q=U\Lambda^{-1/2}\tag{13}$$

则有

$$Q^{\mathrm{T}}S_wQ=\mathbf{I}\tag{14}$$

因此 2DUDV 实现步骤如下：

(1)按式(1)~式(3)计算出训练样本集的类内散布矩阵、类间散布矩阵和总体散布矩阵；

(2)求取类内散布矩阵的特征值和相应的特征矢量，然后按式(13)计算白化矩阵 Q；

(3)计算新的总体散布矩阵 S_t^Q，按式(11)计算 S_t^Q 前 d 个最大特征值所对应的特征矢量，记为 $W=\{w_1,\cdots,w_d\}$；

(4)令 $\varphi=QW$，$\varphi=\{\varphi_1,\cdots,\varphi_d\}$，$\varphi$ 即为不相关最佳鉴别矢量集。

按上述步骤即可非迭代地求得不相关最佳鉴别矢量集。

3.2　关于 2 维不相关鉴别矢量集算法几点讨论

定理 2　由式(11)所得鉴别矢量集，相对于原总体散布矩阵 S_t 具有不相关性。

证明　由上一节讨论可知，在式(11)中新的总体散布矩阵是由原样本经 Q 变换后计算得到的，所以相对于原空间的总体散布矩阵，最终求得的鉴别矢量应为 $\varphi_i=Qw_i(i=1,\cdots,d)$，则对于任意两个鉴别矢量相对于原总体散布矩阵有 $\varphi_i^{\mathrm{T}}S_t\varphi_j=(Qw_i)^{\mathrm{T}}S_t(Qw_j)=w_i^{\mathrm{T}}(Q^{\mathrm{T}}S_tQ)w_j$ 由对式(11)分析可知，有 $(Q^{\mathrm{T}}S_tQ)w_j=\lambda_jw_j$，将两边同时乘以 w_i^{T} 得，$w_i^{\mathrm{T}}(Q^{\mathrm{T}}S_tQ)w_j=\lambda_jw_i^{\mathrm{T}}w_j$，因而有 $\varphi_i^{\mathrm{T}}S_t\varphi_j=w_i^{\mathrm{T}}(Q^{\mathrm{T}}S_tQ)w_j=\begin{cases}\lambda_j,i=j\\0,i\neq j\end{cases}$。证毕。

因此由定理 2 可知，按照 2DUDV 方法求得的鉴别矢量集具有不相关性，因而能够消除投影变换后各特征矢量之间的相关性，也就是减少投影变换后特征分量之间信息冗余，从而将提高算法的识别率。

从算法的求解过程可以看出，2DUDV 是先对类内散布矩阵进行白化，再利用 2DPCA 的求解模型，但又不同于 2DPCA，这是因为 2DPCA 对类内散布矩阵的投影值大小是没有限制的，是一种无监督的算法，这一点使其所求得的特征矢量，从分类的角度上说并不是最优的，而 2DUDV 则是一有种监督的算法，由上一节的分析可知，式(11)求得的任意一个鉴别矢量都使类内散布矩阵满足 $w_i^{\mathrm{T}}Q^{\mathrm{T}}S_wQw_i=1$，因而 2DUDV 所求的任一鉴别矢量都使类内散布矩阵的投影值都是一个常数，也就是在限制类内散布矩阵的投影值的大小同时，使得总体散布矩阵投影值最大；同时从 2DUDV 的结构框架也可以看出，该方法实际上是 2DLDA 的一种变型，但 2DUDV 通过对类内散布矩阵的白化变换，将推广的 2DLDA 模型转化成另外一种类似于 2DPCA 模型的形式，从而避免了 2DLDA 直接对广义特征方程求解引起数值解不稳定的现象，而且也不需假设广义特征方程的特征值各不相同时，求得的鉴别矢量集才具有不相关性，这一点可由上一节对式(11)的分析及定理 2 得出，因此 2DUDV 求得的鉴别矢量集不但具有不

相关性，而且所得到的数值解也较 2DLDA 算法稳定。

4 实验结果及分析

为了验证提出算法的性能，分别在 ORL 和 Yale 人脸库上进行了两组不同目的的实验，采用最小距离分类器进行分类，因为最小距离分类器比较简单，能最大限度地排除分类器对算法优劣的影响。实验平台为 AMD3 600 +，1.90 GHz，512 MB 内存，程序用 Matlab 编写。

实验 1　在 ORL 人脸库上的实验。实验目的主要是为了验证 2DUDV 是否优于文献[3]方法，由于 2DUDV 采用图像模型，因此也将文献[3]方法推广到图像模型。

在 ORL 人脸库中，有 40 个人的 400 幅图像，每人 10 幅，是在不同时间、不同光照、不同头部角度、不同面部表情（睁/闭眼，笑/严肃）和不同人脸细节（有/无眼镜）条件下拍摄得到的，图像尺寸为 112 × 92 pixels。计算中取每人的 5 幅图像作为训练样本，剩余 5 幅图像作为检验样本。注意，由于 2DUDV 与文献[3]方法都是基于图像模型的，故此，若取 K 个鉴别矢量，则整体鉴别矢量的维数为 $K \times 112$。表 1 给出了 2DUDV 与文献[3]的识别性能的比较，包括识别率与执行时间，执行时间为抽取鉴别矢量和识别测试样本两个阶段所用的时间。

表 1　ORL 人脸库上的识别性能比较

Tab. 1　Performance comparisons on ORL face database

鉴别矢量数	文献[3]方法		2DUDV	
	识别率(%)	执行时间(s)	识别率(%)	执行时间(s)
7	85	9.156	88.5	4.062
8	85.5	10.438	89	4.5
9	87.5	12.047	90.5	4.87
10	88	13.641	91	5.047
11	88.5	15.093	92	5.875
12	88	17.016	92	6.062

从表 1 可看出，2DUDV 在识别性能及执行速度均优于文献[3]的算法，特别是 2DUDV 具有较快的执行速度，这是因为 2DUDV 是一种非迭代的算法，而文献[3]是一种迭代算法，因而随着鉴别矢量个数的增加，所需执行时间也将越来越长，当取得最好识别效果时，所需执行时间大约是 2DUDV 的 3 倍。

实验 2　在 Yale 人脸库上的实验。为了进一步验证本文提出的算法的性能，选择了另一个人脸库，本次实验主要是为了验证 2DUDV 是否优于 2DLDA 和 2DPCA 的识别率。在 Yale 人脸库中包含 15 个人的 165 幅图像，每人 11 幅图像，其中包括在不同光照条件下（如左逆光、右逆光）、不同表情的人脸图像，图像的尺寸为 320 × 243 pixels，为了便于识别，首先将 Yale 人脸库中所有图片进行裁剪，并归一化成为 100 × 100 pixels 的标准人脸图像。实验中，每人分别随机选取 3，4 和 5 幅图像作为训练样本，其余图像用于测试，每组实验重复进行 10 次，取平均值作为识别结果，3 种算法均取 10 个鉴别矢量，采用最小距离分类器进行分类，实验结果如表 2 所示。

表 2　Yale 人脸库上的识别结果

Tab. 2　Performance comparisons on Yale face database

样本的个数	识别率(%)		
	2DPCA	2DLDA	2DUDV
3	81.83	83.80	84.72
4	85.76	87.57	88.10
5	88.78	90.33	90.86

从表 2 可以看出，2DUDV 识别率优于 2DPCA 和 2DLDA 的识别率，这是因为 2DUDV 利用了类别信息，是一种有监督的算法因而优于 2DPCA，而且 2DUDV 避免直接对 $\boldsymbol{S}_{\mathrm{w}}^{-1}\boldsymbol{S}_{\mathrm{b}}$ 进行分解，因而其数值解较 2DLDA 方法稳定且具有不相关的特性，因而也优于 2DLDA。

5 结　论

理论分析与实验结果表明，本文提出的基于图像矩阵模型 2DUDV，相对于文献[3]的算法、2DLDA 和 2DPCA 的特征提取具有一定的优势。算法由于采用了图像矩阵模型，从而避免了类内散布矩阵奇异的问题，其次，通过对类内散布矩阵的白化变换，使得推广的 2DLDA 模型具有类似的 2DPCA 模型的形式，从而将两种算法的模型有效地联系起来，进而可以非迭代地求得具有统计不相关的最佳鉴别矢量集，不但降低了算法的复杂度，同时也大大降低了求解鉴别矢量所需的时间，为求取不相关鉴别矢量集提供了一种新的思路。但由于算法采用了图像矩阵

模型,因而在后期的识别阶段,造成图像的投影特征系数多,需要较多的存储空间,因此下一步的工作是如何在减少图像的投影特征系数的基础上进一步提高算法的识别率。

参考文献(References)

1 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.

2 Foloy D H, Sammon J W. An optimal set of vecors[J]. IEEE Transactions on Computers,1975,**24**(3):281-289.

3 Jin Zhong, Yang Jing-yu, Lu Jian-feng. An optimal set of uncorrelated discriminant features[J]. Chinese Journal of Computers, 1999, **22**(10):1105-1108.[金忠,杨静宇,陆建锋. 一种具有统计不相关性的最佳鉴别矢量集[J]. 计算机学报, 1999, **22**(10): 1105-1108.]

4 Yang Jing-yu, Jin Zhong, Hu Zhong-shan. A theorem on dimensionality of the uncorrelated optimal discriminant feature Space[J]. Chinese Journal of Computers,2003, **26**(1):110-115.[杨静宇,金忠,胡钟山. 具有统计不相关性的最佳鉴别特征空间的维数定理[J]. 计算机学报,2003,**26**(1):110-115.]

5 Wu Xiao-jun, Yang Jing-yu, Wang Shi-tong. A theoretical result on the generalized optimal set of statistically uncorrelated discriminant vectors[J]. Acta Electronica Sinica, 2005,**32**(10):1720-1724.[吴小俊,杨静宇,王士同. 广义统计不相关最优鉴别矢量集的一个理论结果[J]. 电子学报, 2005,**32**(10):1720-1724.]

6 Wu Xiao-jun, Yang Jing-yu, Wang Shi-tong, *et al.* An improved optimal set of statistical uncorrelated discriminant vectors[J]. Journal of Electronics & Information Technology, 2005,**27**(1):47-50.[吴小俊,杨静宇,王士同等. 改进的统计不相关最优鉴别矢量集[J]. 电子与信息学报,2005,**27**(1):47-50.]

7 Chen Mian-shu, Chen He-xin, Liu Wei. A new method for resolving the uncorrelated set of discriminant vectors[J]. Chinese Journal of Computers, 2004,**27**(7):913-917.[陈绵书,陈贺新,刘伟. 一种新的求解无相关鉴别矢量集方法[J]. 计算机学报,2004,**27**(7):913-917.]

8 Yang Jian, Zhang David, Xu Yong, *et al.* Two-dimensional discriminant transform for face recognition[J]. Pattern Recognition,2005,**38**(7):1125-1129.

9 Kirby M, Sirovich L. Application of the Karhunen-Loeve procedure for the charaterization of human faces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1990,**12**(1):103-108.

10 Yang Jian, Zhang David, Frangi Alejandro, *et al.* Two-dimensional PCA a new approach to appearance-based face representation and recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2004,**26**(1):131-137.

11 Jin Zhong, Yang Jing-yu, Tang Zhen-min, *et al.* A theorem on uncorrelated optimal discriminant vectors, Pattern Recognition, 2001,**34**(10): 2041-2047.