基于支持向量的Kernel判别分析

张宝昌[1)] 陈熙霖[2)] 山世光[2)] 高 文[1),2)]

[1)](哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

[2)](中国科学院计算技术研究所银晨科技面像识别联合实验室 北京 100080)

摘 要 提出了一种新的基于支持向量的核化判别分析方法(SV-KFD).首先深入地分析了支持向量机(SVM)以及核化费舍尔判别分析(Kernel Fisher)方法的相互关系.基于作者证明的SVM本身所固有的零空间性质:SVM分类面的法向量在基于支持向量的类内散度矩阵条件下,具有零空间特性,提出了利用SVM的法向量定义核化的决策边界特征矩阵(Kernelized Decision Boundary Feature Matrix,KDBFM)的方法.进一步结合均值向量的差向量构建扩展决策边界特征矩阵(Ex-KDBFM).最后以支持向量为训练集合,结合零空间方法来计算投影空间,该投影空间被用来从原始图像中提取判别特征.以人脸识别为例,作者在FERET和CAS-PEAL-R1大规模人脸图像数据库上对所提出的方法进行了实验验证,测试结果表明该方法具有比传统核判别分析方法更好的识别性能.

关键词 人脸识别;支持向量机;核分析;判别分析;零空间

中图法分类号 TP391

Kernel Discriminant Analysis Based on Support Vectors

ZHANG Bao-Chang[1)] CHEN Xi-Lin[2)] SHAN Shi-Guang[2)] GAO Wen[1),2)]

[1)](*School of Computer Science and Technology*, *Harbin Institute of Technology*, *Harbin* 150001)

[2)](*ICT-ISVISION Joint R&D Laboratory for Face Recognition*, *Institute of Computing Technology*, *Chinese Academy of Sciences*, *Beijing* 100080)

Abstract Discriminant analysis is one of crucial issues for the statistic-based face recognition methods. This paper proposes a novel Support Vectors based Kernel Fisher Discriminant analysis method (SV-KFD) for face recognition, which has combined the idea of the Support Vector Machine(SVM) and kernel Fiser analysis. The authors first discuss the relationship between SVM and kernel Fisher analysis. Based on the intrinsic nullspace property of the SVM proven by the authors, which shows that the normal vector of the SVM decision plane is of the nullspace property in terms of the support vectors-based within-class scatter matrix, a support vectors-based method is presented to construct the Kernelized Decision Boundary Feature Matrix (KDBFM) by using the SVM normal vectors. Furthermore the difference vector of the mean support vectors is combined to construct the Extended Kernelized Decision Boundary Feature Matrix (Ex-KDBFM). Finally, the nullspace Kernel Fisher method is exploited to seek the projection space, which is used to extract the discriminant features from the original face images. In addition, the proposed discriminant method includes two steps, and the tedious sigularity problem can be avoided. The proposed method is applied on face recognition, and the experimental results on the FERET and CAS-PEAL-R1 databases show that it performs much better than the traditional kernel discriminant analysis methods in terms of the recognition rate.

收稿日期:2005-05-08;修改稿收到日期:2006-06-06.本课题得到国家自然科学基金重点项目“基于生物特征的身份识别研究”(60332010)、中国科学院百人计划、新世纪优秀人才支持计划(NCET-04-0320)以及上海银晨智能识别科技有限公司资助.**张宝昌**,男,1976年生,博士研究生,研究方向为人脸识别、模式分析、机器学习、计算机视觉.E-mail:bczhang@jdl.ac.cn.**陈熙霖**,男,1965年生,博士,研究员,主要研究领域为计算机视觉、模式识别、机器学习、智能人机交互等.**山世光**,男,1975年生,博士,副研究员,研究方向为模式分析、机器学习、计算机视觉、尤其专注于人脸识别问题.**高 文**,男,1956年生,博士,教授,博士生导师,主要研究领域为多媒体数据压缩、图像处理、计算机视觉、多模式接口、人工智能、虚拟现实等.

Keywords face recognition; support vector machine; kernel analysis; discriminant analysis; nullspace

1 引 言

近年来，核方法已成为模式识别领域一个迅速发展的研究方向. 它的主要思想是由 Vapnik 提出并应用于 SVM[1,2]. 其后，Mika[3] 和 Baudat[4] 把核方法应用到判别分析领域，将 Fisher 线性判别分析方法进一步拓广到非线性情形，提出了 Kernel Fisher 方法. Kernel Fisher 在模式识别领域取得了成功的应用，它主要利用类均值向量的差进行判别分析[5]. 根据结构风险理论[1,2]，由于 SVM 分类器是线性形式的，分类时等价于样本在 SVM 的法向量(即分类面的法线方向，图 1 中的 $\boldsymbol{h}'(\boldsymbol{x})$)上进行投影，所以 $\boldsymbol{h}'(\boldsymbol{x})$ 是具有强判别分析能力的投影方向，该投影方向不依赖于原始类均值向量的差. 一般情况下，原始空间的样本投影到 SVM 的法向量形成一维空间，具有非常好的可区分性.

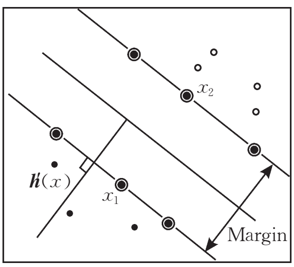


图 1 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$ 是支持向量，$\boldsymbol{h}'(\boldsymbol{x})$ 是法向量，Margin 是类间距

除了经典的 Fisher 等判别分析方法之外，还有一些重要的判别分析方法，比如 Lee[5] 提出 Decision Boundary 理论，它的一个主要贡献是通过构建决策边界矩阵(Decision Boundary Feature Matrix，DBFM)，然后对 DBFM 进行主成分分析来获取和原始特征空间具有同样判别能力的最小维数的子空间[5,6].

本文的主要贡献即是在高维空间，利用 SVM 的法向量来将 DBFM 扩展到非线性情形，从而构建核化的 DBFM，即 KDBFM，然后结合 Kernel Fisher 方法进行判别分析. 基于 SVM 的法向量具有零空间性质的这一特性，本文提出了在多类问题情况下利用 SVM 的法向量构建 KDBFM 的方法，最后利用基于零空间核化费舍尔判别分析方法来计算投影空间. 为了检验这种方法的有效性，本文把它应用到人脸识别领域. 人脸识别问题在最近几十年来成为一个热点问题，有很广阔应用前景，例如视频监控，安全访问以及考勤系统等等[7]. 研究者提出许多解决办法，如 Eigenface、Fisherface[8]、弹性图匹配法[9]等，近年来 Kernel 方法受到越来越多人脸识别研究者的重视[10～12]，逐渐成为一种重要的人脸识别方法.

本文第 2 节简单介绍核化费舍尔判别分析方法；第 3 节讨论了 SVM 法向量的一些特性；第 4 节提出基于 SVM 的非线性判别分析方法；第 5 节给出两组对比实验结果；最后一节对本文工作进行总结并对未来工作的作了展望.

2 背景知识

本节简要叙述核方法在支持向量机和 Kernel Fisher 判别分析中的应用，它们一方面是本文方法的基础，另一方面也将作为本文实验中主要的对比方法.

2.1 支持向量机

SVM 是统计学习理论中最年轻的部分，1995 年才完成其主要的理论内容，现在仍然在不断地发展. SVM 的基本思想是根据 Vapnik 提出的结构风险最小化(structure risk minimization)原理，通过最大化分类间隔或边缘(Margin)尽量提高学习机的泛化能力. 它是从线性可分情况下的最优分类面(optimal hyperplane)提出的，而在其定义空间线性不可分的情况下，SVM 则将其映射到一个更高维的空间变成一个线性可分的问题. 考虑到求解最优分类面算法的性质，在这个高维空间中，我们只需进行内积运算即可，所以可以利用 kernel 技术，这样即使高维空间的维数增加很多，在其中求解最优分类面的问题并没有增加多少计算复杂度[1,2].

SVM 本质上是一种二分类方法，而在现实世界中的大多数分类问题都是多类问题，如本文讨论的人脸识别问题就是一个典型的多类问题，因此 SVM 具有很大的局限性，必须寻找一种 SVM 多分类方法，才能使 SVM 具有实用价值. 应用较多的是一对一(One-against-One)方法和一对多(One-against-Rest)方法[1]. 假定样本集合包含 k 个类别，对于 One-against-One 方法而言，需要建立 $k(k-1)/2$ 个 SVM 分类器，对于 One-against-Rest 方法而言，

需要建立 k 个 SVM 分类器. 本文主要采用了复杂度较小的 One-against-Rest 方法计算支持向量集合和法向量. 支持向量集的性质是 SVM 理论非常重要的一部分，在文献[13,14]中，Bernhard 等人仅仅利用支持向量集来训练分类器，就取得了与基于全部训练样本得到的分类器同样或略好的性能. Fransens[6]等人提出优化支持向量机类间距（图 1 中的 Margin）的梯度的方法，并且成功地把这种方法应用到人脸检测中，取得了很好的实验结果.

2.2 Kernel Fisher 判别分析方法

Kernel Fisher 的目标是把样本映射到一个更高维空间然后进行 Fisher 判别分析[10~12]. 与 SVM 一样，Kernel Fisher 用一个隐函数 Φ 来实现从原始空间到高维空间的映射. 在实现的时候，无需知道 Φ 的具体表示形式，只要进行内积运算即可，则有

$$k(\boldsymbol{x},\boldsymbol{y})=(\Phi(\boldsymbol{x})\cdot\Phi(\boldsymbol{y})) \tag{1}$$

类间散度矩阵 $\boldsymbol{S}_b$ 和类内散度矩阵 $\boldsymbol{S}_w$ 的定义如下所示：

$$\begin{aligned}\boldsymbol{S}_b&=\sum_{i=1}^{C}\sum_{j=1,j\neq i}^{C}p(\varpi_i)p(\varpi_j)(\boldsymbol{u}_i-\boldsymbol{u}_j)(\boldsymbol{u}_i-\boldsymbol{u}_j)^{\mathrm{T}}\\&=\sum_{i=1}^{C}p(\varpi_i)(\boldsymbol{u}_i-\boldsymbol{u})(\boldsymbol{u}_i-\boldsymbol{u})^{\mathrm{T}}\end{aligned} \tag{2}$$

$$\boldsymbol{S}_w=\sum_{i=1}^{C}p(\varpi_i)E\{((\Phi(\boldsymbol{x}_i)-\boldsymbol{u}_i)(\Phi(\boldsymbol{x}_i)-\boldsymbol{u}_i)^{\mathrm{T}})|\varpi_i\} \tag{3}$$

这里 $\boldsymbol{u}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}\boldsymbol{\phi}(x_{ij})$ 表示类别 i 的均值向量，$\boldsymbol{u}$ 表示所有样本的均值向量，$p(\varpi_i)$ 是先验概率. 为了实现在高维空间 F 中判别分析，等价于最大化下面的表达式[2]：

$$J(\boldsymbol{w})=\frac{|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{w}|}{|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}|} \tag{4}$$

一般情况下，$\boldsymbol{w}\in F$ 定义在整个训练样本空间，有如下表示形式[10]：

$$\boldsymbol{w}=\sum_{i=1}^{n}\alpha_i\Phi(\boldsymbol{x}_i) \tag{5}$$

则 Kernel Fisher 的优化准则变成如下表达式[10,12]：

$$J(\boldsymbol{\alpha})=\frac{|\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}_b\boldsymbol{\alpha}|}{|\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}_w\boldsymbol{\alpha}|} \tag{6}$$

其中核矩阵 $\boldsymbol{K}_w$ 和 $\boldsymbol{K}_b$[10] 的定义如下所示：

$$\boldsymbol{K}_w=\sum_{i=1}^{C}p(\varpi_i)E(\boldsymbol{\eta}_j-\boldsymbol{m}_i)(\boldsymbol{\eta}_j-\boldsymbol{m}_i)^{\mathrm{T}} \tag{7}$$

$$\boldsymbol{K}_b=\sum_{i=1}^{C}p(\varpi_i)(\boldsymbol{m}_i-\bar{\boldsymbol{m}})(\boldsymbol{m}_i-\bar{\boldsymbol{m}})^{\mathrm{T}} \tag{8}$$

其中 $\boldsymbol{\eta}_j=(k(\boldsymbol{x}_1,\boldsymbol{x}_j),k(\boldsymbol{x}_2,\boldsymbol{x}_j),\cdots,k(\boldsymbol{x}_n,\boldsymbol{x}_j))^{\mathrm{T}}$，$\boldsymbol{\eta}_j$ 的类别标号与 $\boldsymbol{x}_j$ 相同，$\boldsymbol{m}_i$ 表示类别 i 的 $\boldsymbol{\eta}_j$ 的均值向量，$\bar{\boldsymbol{m}}$ 是全部 $\boldsymbol{\eta}_j$ 的均值向量.

类似于线性判别分析，这个问题能够通过求解矩阵 $\boldsymbol{K}_w^{-1}\boldsymbol{K}_b$ 的特征向量得到变换空间（本文称之为 GKFD）. 本文利用求伪逆的方法来计算 $\boldsymbol{K}_w^{-1}\boldsymbol{K}_b$，然后求解它的主成分 $\boldsymbol{\alpha}$，这样我们可以用下面的表达式来计算 Kernel 判别特征：

$$\boldsymbol{v}=(\boldsymbol{w}.\Phi(\boldsymbol{x}))=\sum_{i=1}^{n}\alpha_i k(\boldsymbol{x}_i,\boldsymbol{x}) \tag{9}$$

对于投影方向 $\boldsymbol{w}^*$，如果：

$$(\boldsymbol{w}^*)^{\mathrm{T}}\boldsymbol{S}_w\boldsymbol{w}^*=0 \tag{10}$$

即 $\boldsymbol{w}^*$ 具有零空间特性[12]. 如果 $(\boldsymbol{w}^*)^{\mathrm{T}}\boldsymbol{S}_b\boldsymbol{w}^*>0$，那么 $J(\boldsymbol{w}^*)=\infty$，则该投影空间具有最好的判别分析能力.

3 支持向量机的零空间特性

为了简化问题，本节假定考虑的是两类问题，样本集合为 $\{(\boldsymbol{x}_i,\boldsymbol{y}_i)|\boldsymbol{x}_i\in R^n,y_i\in\{-1,1\}\}$，该集合分成正例集合和反例集合，SVM 分类面函数表示为

$$h(\boldsymbol{x})=\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}+b \tag{11}$$

其中 $\boldsymbol{w}=\sum_{i}^{n}\alpha_i y_i\boldsymbol{x}_i$，$\sum_{i}^{n}\alpha_i y_i=0$，$\alpha_i>0$，$\boldsymbol{x}_i$ 是支持向量，$\boldsymbol{b}$ 是阈值. 则最大化两类之间的类间距等价于最小化如下表达式[1,2]：

$$\min_{\boldsymbol{w}}\quad\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} \tag{12}$$

$$\text{约束条件：}y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}_i+b)-1\geq 0 \tag{13}$$

本文假定 $\Phi(\boldsymbol{x}_i)$ 是 $\boldsymbol{x}_i$ 的高维空间表示，Φ 是一个隐函数，无需知道函数具体形式，则有如下目标函数：

$$\min_{\boldsymbol{w}}\quad\frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} \tag{14}$$

$$\text{约束条件：}y_i(\boldsymbol{w}^{\mathrm{T}}\Phi(\boldsymbol{x}_i)+b)-1\geq 0 \tag{15}$$

其中 $\boldsymbol{w}=\sum_{i=1}^{n}\alpha_i y_i\Phi(\boldsymbol{x}_i)$.

若 $\Phi(\boldsymbol{x}_i)$ 不是支持向量，则 $\alpha_i=0$，否则 $\alpha_i\neq 0$，且式(15)的等号成立[2]，即

$$y_i(\boldsymbol{w}^{\mathrm{T}}\Phi(\boldsymbol{x}_i)+b)-1=0 \tag{16}$$

下面定义两个由支持向量构成的集合：

$$\boldsymbol{S}_1=\{\Phi(\boldsymbol{x}_i)|y_i=1,\Phi(\boldsymbol{x}_i)\text{是支持向量}\} \tag{17}$$

$$\boldsymbol{S}_2=\{\Phi(\boldsymbol{x}_i)|y_i=-1,\Phi(\boldsymbol{x}_i)\text{是支持向量}\} \tag{18}$$

对任意 $\Phi(\boldsymbol{x}_i)\in\boldsymbol{S}_1$，有

$$\boldsymbol{U}_1^{\Phi}=\frac{1}{|s_1|}\sum_{\Phi(\boldsymbol{x})\in\boldsymbol{S}_1}\Phi(\boldsymbol{x}) \tag{19}$$

同理对任意 $\Phi(\boldsymbol{x}_i)\in\boldsymbol{S}_2$，有

$$U_2^{\Phi}=\frac{1}{|s_2|}\sum_{\Phi(x)\in S_2}\Phi(x) \tag{20}$$

在两类的情况下，基于支持向量的类内散度矩阵定义如下所示：

$$S'_w=\sum_{\Phi(x_i)\in S_1}(\Phi(x_i)-U_1^{\Phi})(\Phi(x_i)-U_1^{\Phi})^{\mathrm{T}}+\sum_{\Phi(x_i)\in S_2}(\Phi(x_i)-U_2^{\Phi})(\Phi(x_i)-U_2^{\Phi})^{\mathrm{T}} \tag{21}$$

我们先前工作证明如下性质[15]：

$$w^{\mathrm{T}}S'_w w=0 \tag{22}$$

根据式(10)，可以知道 SVM 的法向量在基于支持向量的类内散度矩阵前提下，具有零空间特性.

在高维空间，SVM 的分类面函数表示为

$$h(\Phi(x))=w^{\mathrm{T}}\Phi(x)+b \tag{23}$$

则它对 $\Phi(x)$的导数 $h'(\Phi(x))$是分类面的法向量即 $h'(\Phi(x))=w$，则

$$h'(\Phi(x))=\sum_{i=1}^{n}\alpha_i y_i\Phi(x_i) \tag{24}$$

其中 $\Phi(x_i)$是支持向量. 由于 $y_i\in\{-1,1\}$，故 $h'(\Phi(x))$可以理解为以 α_i为权重计算支持向量集中属于不同类别的两个子集 S_1和 S_2的均值向量，然后做差. 假定两类样本的原始均值向量为 u_1 和 u_2，Fisher 的投影方向依赖于 u_1-u_2[5]，且它被用来构建类间散度矩阵，所以，采用以差向量为表示形式的 $h'(\Phi(x))$做判别分析是很自然的选择.

4 基于 SVM 的 Kernel 判别分析

本节首先简述 Decision Boundary 理论，该理论的具体细节部分可参考文献[5]. 然后提出构建 KDBFM 和基于支持向量定义类内散度矩阵的方法，最后利用零空间判别分析方法计算投影空间.

4.1 决策边界特征矩阵

Decision Boundary 理论的一个最主要贡献是给出决策边界特征矩阵(DBFM)的定义和性质[5].

DBFM 定义. S 是 Decision Boundary 或分类面，$N(x)$是 Decision Boundary 上 x 点的法向量，$p(x)$是概率密度函数. 则 DBFM 的定义如下所示：

$$\sum_{\mathrm{DBFM}}=\int_S N(x)N^{\mathrm{T}}(x)p(x)\mathrm{d}x \tag{25}$$

DBFM 性质. DBFM 的大于零的特征值对应的特征向量构成最本质的投影子空间，该子空间能够保留原始空间全部判别信息并且维数是最小的空间.

在多类问题情形，DBFM 是基于 One-against-One 策略定义的[5]. u_i-u_j决定 Fisher 的投影方向，在 Fisher 原则中定义的类间散度矩阵(即式(2))可以认为是利用类均值向量(等权重)的差构建的 DBFM.

在高维空间 F 中，我们利用 SVM 计算分类面，则其任意一点的法向量都是相同的而且是可计算的，基于此，下一小节将提出构建核化的决策边界特征矩阵的方法.

4.2 基于 SVM 的零空间判别分析

本小节提出在 F 空间以及多类问题情形定义核化的决策边界特征矩阵和类内散度矩阵的方法. 为了计算上的方便，本文采用 One-against-Rest 方法.

$$\Sigma_{\mathrm{KDBFM}}=\sum_{i=1}^{C}p(\varpi_i)h'_i(\Phi(x))(h'_i(\Phi(x)))^{\mathrm{T}} \tag{26}$$

$$S'_w=\sum_{i=1}^{c}p(\varpi_i)\sum_{k=0}^{n_i}p(\varpi_k|\varpi_i)E((\Phi(x'_k)-u'_{ik})(\Phi(x'_k)-u'_{ik})^{\mathrm{T}}) \tag{27}$$

Σ_{KDBFM}即核化的决策边界特征矩阵. 假定 S_i表示在 C_i构成正例集以及 $C_{j,j\neq i}$构成反例集时，利用两类 SVM 算法计算得到的所有支持向量构成的集合，$h'_i(\Phi(x))$是分类面的法向量. S_i被分成两个子集合，分别代表来自不同 Margin 上的样本构成的集合，$S_{i1}\subset C_i$，这里 u'_{i0}表示 S_{i1}的均值向量，而剩下的支持向量构成集合 S_{i2}. 我们知道 S_{i2}中的样本属于多个类别，假定其类别数为 n_i，针对每一个类别计算它们的类均值向量分别表示为 u'_{ik}，$k=1,2,\cdots,n_i$(如果只有一个样本，它自身就是该类别的中心). 这样我们可以利用多个均值向量来表示 S_i 为(u'_{i0}，u'_{i1}，…，u'_{in_i})，x'_k代表以 u'_{ik}为中心的支持向量，$p(\varpi_i)$是先验概率密度. 因此，式(30)不同于传统的类内散度矩阵定义方法，每一类是由一个类中心向量来表示. 而本文考虑的是支持向量集合，我们利用多个支持向量均值来表示它.

需要说明的是：在 S_i集合中，如果某个类别的支持向量的个数大于 1，那么该类别对类内散度矩阵 S'_w的计算是有贡献的. 而且在实验中我们发现在一对多的策略下，C_i类中几乎全部的样本都包含在 S_i中，这样就保证式(27)的计算是基于整个样本空间的，而且这种基于多个均值向量定义类内散度矩阵的方法可以利用尽可能多的类内信息.

在 One-against-Rest 策略下，反例集合的样本偏多，为了解决这个问题，把反例集合分成 k_{subpart}，正例集合保持不变，则可以得到 k_{subpart} 个 SVM 的法

向量和支持向量集来计算 KDBFM 和基于支持向量类内散度矩阵.

根据式(2)以及第 3 节和 4.1 节的讨论可知，$(\boldsymbol{u}_i'-\boldsymbol{u}_{i0}')$ 也可用来构建特殊的 KDBFM. 由前文的阐述，我们知道据 $(\boldsymbol{u}_i'-\boldsymbol{u}_{i0}')$ 和 $\boldsymbol{h}_i'(\Phi(\boldsymbol{x}))$ 具有一致性，即它们都是均值向量的差向量，但是它们提供不同类型的判别信息. 据此结合式(26)定义扩展决策边界特征矩阵(Ex-KDBFM).

$$\boldsymbol{\Sigma}_{\text{Ex-KDBFM}}=\sum_{i=1}^{C}p(\varpi_i)((\boldsymbol{u}_i'-\boldsymbol{u}_{i0}')(\boldsymbol{u}_i'-\boldsymbol{u}_{i0}')^{\mathrm{T}}+\boldsymbol{h}_i'(\Phi(\boldsymbol{x}))(\boldsymbol{h}_i'(\Phi(\boldsymbol{x})))^{\mathrm{T}}) \tag{28}$$

需要说明的是，$(\boldsymbol{u}_i'-\boldsymbol{u}_{i0}')$ 和 $\boldsymbol{h}_i'(\Phi(\boldsymbol{x}))$ 都需要利用自身的模进行归一化.

根据 Decision Boundary 理论，KDBFM 的大于零特征值对应的特征向量 $\boldsymbol{w}^*$ 满足

$$(\boldsymbol{w}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\text{KDBFM}}\boldsymbol{w}^*>0 \tag{29}$$

结合式(25)，则 $J(\boldsymbol{w}^*)=\dfrac{(\boldsymbol{w}^*)^{\mathrm{T}}\boldsymbol{\Sigma}_{\text{KDBFM}}\boldsymbol{w}^*}{(\boldsymbol{w}^*)^{\mathrm{T}}\boldsymbol{S}_w'\boldsymbol{w}^*}=+\infty$，这些特征向量构成的投影空间具有最好的识别性能，所以可以利用基于零空间核化费舍尔判别分析方法来计算得到.

在 F 空间，$\boldsymbol{w}$ 的定义形式同式(5)，则很容易计算核化的决策边界特征矩阵以及 $\boldsymbol{S}_w'$ 的核矩阵 $\boldsymbol{K}_w'$. 为了表述方便，设 $\boldsymbol{K}_b'$ 为 KDBFM 或者 Ex-KDBFM，则利用下一小节的方法计算投影空间. 另外，当 $\boldsymbol{K}_b'$ 为 KDBFM 时候，称之为 SV-KFD，而 $\boldsymbol{K}_b'$ 为 Ex-KDBFM 时候，称之为 SV-KFD1.

4.3　算法实现

本小节将给出实现基于 SVM 的零空间判别分析方法，从文献[2]中，可以得到等价于式(6)的最大化目标函数 $J(\boldsymbol{\alpha})=\dfrac{\boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{K}_b+\boldsymbol{K}_w)\boldsymbol{\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{K}_w\boldsymbol{\alpha}}$. 类似地，我们首先对 $\boldsymbol{K}_t'=\boldsymbol{K}_w'+\boldsymbol{K}_b'$ 进行主成分分析.

$$\boldsymbol{K}_t'\boldsymbol{\Xi}=\boldsymbol{\Xi}\boldsymbol{\Gamma}\text{ 和 }\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{\Xi}=\boldsymbol{I} \tag{30}$$

$$\boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{K}_t'\boldsymbol{\Xi}\boldsymbol{\Gamma}^{-1/2}=\boldsymbol{I} \tag{31}$$

其中，$\boldsymbol{\Xi}$，$\boldsymbol{\Gamma}$ 是矩阵 $\boldsymbol{K}_t'$ 的特征向量和特征值(对角阵)对应的矩阵，而 $\boldsymbol{\Xi}'$ 对应的是特征值大于 0 的那些特征向量，即去掉那些特征值接近于 0 的特征向量[12]之后得到的由剩余特征向量构成的变换矩阵，利用该矩阵修改类内散度矩阵，则有

$$\boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Xi}^{\mathrm{T}}\boldsymbol{K}_w'\boldsymbol{\Xi}\boldsymbol{\Gamma}^{-1/2}=\boldsymbol{\Xi}_w \tag{32}$$

对新的类内散度矩阵 $\boldsymbol{\Xi}_w$ 进行主成分分析：

$$\boldsymbol{\Xi}_w\boldsymbol{\theta}=\boldsymbol{\theta}\boldsymbol{\gamma}\text{ 和 }\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}=\boldsymbol{I} \tag{33}$$

其中，$\boldsymbol{\theta}$，$\boldsymbol{\gamma}$ 是 $\boldsymbol{\Xi}_w$ 的特征向量和特征值(对角阵)对应的矩阵，特征值是按着增序排列的，去掉那些特征值远大于 0 的向量，剩余那部分特征向量构成 $\boldsymbol{\theta}^*$，从而得到如下变换矩阵：

$$\boldsymbol{\alpha}=\boldsymbol{\Xi}'\boldsymbol{\Gamma}^{-1/2}\boldsymbol{\theta}^* \tag{34}$$

因为 $\boldsymbol{\alpha}_i^{\mathrm{T}}(\boldsymbol{K}_b'+\boldsymbol{K}_w')\boldsymbol{\alpha}_i=1$，$\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{K}_w'\boldsymbol{\alpha}_i=r_i\approx 0$，$\boldsymbol{\alpha}_i^{\mathrm{T}}(\boldsymbol{K}_b')\boldsymbol{\alpha}_i\approx 1>0$，满足零空间判别分析要求，所以可以知道该投影空间具有强判别分析能力. 假定图像 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$ 在该投影空间抽取的特征为 v_1，v_2，在识别的过程中，本文将利用如下相似度：

$$S(\boldsymbol{x}_1,\boldsymbol{x}_2)=\frac{v_1\cdot v_2}{\|v_1\|\cdot\|v_2\|} \tag{35}$$

5　实验分析

在实验中，首先根据手工标定的人眼睛的位置把数据抠成 64×64 大小的图片，利用一个 Mask 去掉背景信息以及头发的影响，然后进行主成分分析来降低原始图像的维数(保留 98%的能量)，从而减少训练时间. 实验时 $k_{\text{subpart}}=10$，本文利用的是多项式核函数，$k(x,y)=\left(\dfrac{x\cdot y}{|x|\cdot|y|}+1\right)^r$，$r$ 是一个常整数.

5.1　FERET 数据库

FERET 是一个基本的人脸识别算法测试数据库，本文选择其一个子集来测试各种算法，该子集包括 200 人[15]. 如图 2 所示，每个人 7 张照片，涉及到光照、姿态、表情的变化. 在实验中，Gallery 每人一张照片(ba 部分)，其余的作为 probe 中的图片. 首先按着人脸图像编号的大小把该数据库划分成两个子集合，每一个子集有 100 人，用其中一个子集来训练模型，而另外一个子集进行测试，这样可以得到

图 2　FERET 数据库的样本示例

两组实验结果,即如下所示结果是一个平均值.因为零空间核化费舍尔判别分析(即图 3 所示的 KNullspace)方法同样具有零空间特性,所以本文和这种方法进行了对比,该方法的具体细节可以参考文献[12].SVM 方法利用一对多策略,具体的细节可以参考文献[16],核函数与本文算法相同.从图 3 知道,本文提出的两个方法都取得了比其它传统的判别分析更好的性能.而且非线性的判别分析方法(核方法)比 Fisherface 和主成分分析(PCA)表现出更好的识别能力.

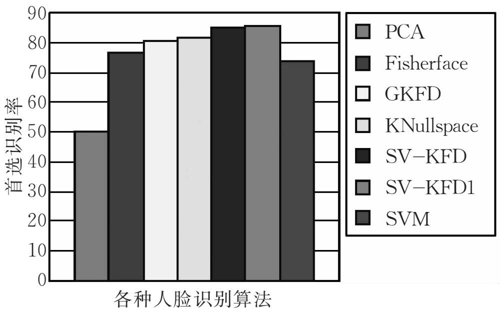


图 3 FERET 数据库上首选识别率对比($r=2$)

5.2 CAS-PEAL-R1 数据库

CAS-PEAL-R1 数据库是一个大型人脸数据库[17],包括 1040 人以及 99594 张图片,涉及到姿态、表情、饰物、老化等变化,主要是用来衡量基于东方人的人脸识别系统的性能.如图 4 所示,我们给出一些该数据库的示例图像.本文的实验中 Gallery 数据库中每人一张标准照片,并且选择如下测试子集合:饰物(Accessory)、背景(Background)、距离(Distance)、表情(Expression)、老化(Aging)来检验算法的识别性能.训练集合是通过随机选择 1200 张图片得到的.有关该数据库的详细信息可以参考文献[17].

从以上两个实验可以看出,本文提出的方法都好于其它算法,而且在数据量较大的 CAS-PEAL-R1 数据库上也有一定幅度的提高,可以说该方法具有很强的识别能力.在 CAS-PEAL-R1 数据库上,Fisherface 的识别率有时候高于 GKFD,可能的原因是 GKFD 利用求伪逆的方法保留一些噪声信息,所以使得识别性能不够稳定.从以上的两组实验也可以看到 KNullspace 方法取得了相对稳定的性能,该方法避免了伪逆运算.本文提出的两种基于支持向量的判别分析方法试图区分不同类中离得最近的样本.该思想最早起源于支持向量机结构风险最小化原理,我们把这一思想成功地应用到判别分析领域.由图 3 和表 1 可以看到该方法都取得比传统的判别分析方法更好的识别性能.除此以外,基于扩展决策边界特征矩阵方法能够提高系统的性能,这表明了基于均值向量的差向量对支持向量机的法向量是一种补充.

图 4 CAS-PEAL-R1 数据库的样本示例

表 1 CAS-PEAL-R1 数据库的识别结果($r=2$)

	PCA	Fisherface	KNullspace	GKFD	SVM	SV-KFD	SV-KFD1
Accessory	37.1	61	61.5	58.7	58	64	64.1
Aging	50	72.7	83.3	77.3	77.3	87.9	87.9
Background	80.5	94.4	94	91.7	89	94.4	95.2
Distance	74.2	93.5	93.8	94.9	94.4	96	96.4
Expression	53.7	71.3	77.5	78.2	74.3	80	81.2

6 结论和未来工作

本文基于 SVM 本身所具有的零空间特性,提出基于 SVM 的核判别分析方法,并把该方法用到人脸识别领域取得很好的识别效果.SV-KFD 的核心部分是用 SVM 的法向量和支持向量集合来构建 KDBFM 和类内散度矩阵,然后利用零空间方法来进行判别分析.本文的主要贡献包括如下几方面:(1)利用法向量定义的 KDBFM,根据结构风险最小化理论,在 SVM 的法向量上投影构成一维空间,具有最好的可分性.(2)尽管仅仅利用支持向量似乎减

少了样本,但是在实验中发现几乎全部正例集合中的样本都在支持向量集中,所以可以知道在一对多策略情况下,能保留尽可能多类内散度信息.(3)基于支持向量的均值向量的差向量进一步用来构建扩展的决策边界特征矩阵,实验表明它能提高系统的性能.大量的对比实验表明该方法比 Eigenface,Fisherface,GKFD 和 Kernel Nullspace 有更好的识别性能.未来的工作将集中在特征表示方面,比如尝试 Gabor 特征等.另外构建一个鲁棒的分类器是一个很重要的研究方向,我们希望构建的分类器具有更好的泛化能力.

参　考　文　献

1 Burges C. J. C.. A tutorial on support vector machines for pattern recognition. Knowledge Discovery and Data Mining, 1998, 2(2): 121～167

2 Bian Zhao-Qi, Zhang Xue-Gong *et al*. Pattern Recognition. Beijing: Tsinghua University Press, 2000(in Chinese)
(边肇祺,张学工等.模式识别.北京:清华大学出版社,2000)

3 Mika S., Ratsch G., Weston J., Scholkopf B., Muller K. R.. Fisher discriminant analysis with kernels. In: Proceedings of the IEEE International Workshop on Neural Networks for Signal Processing, Madison, USA, 1999, 41～48

4 Baudat G., Anouar F.. Generalized discriminant analysis using a kernel approach. Neural Computation, 2000, 12(10): 2385～2404

5 Lee C., Landgrebe D. A.. Feature selection based on decision boundaries. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1993, 15(1): 388～400

6 Fransens R., Pris Jan De.. SVM-based nonparametric discriminant analysis, an application to face detection. In: Proceedings of the 9th International Conference on Computer Vision, Nice, France, 2003, 1289～1296

7 Daugman J. G.. Face and gesture recognition: Overview. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 675～676

8 Belhumeur Peter N., Hespanha John P., Kriegman David J.. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711～720

9 Wiskott L., Fellous J. M., Kruger N., von der Malsburg C.. Face recognition by elastic bunch graph matching. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 775～779

10 Liu Qing-Shan, Huang Rui. Face recognition using kernel based fisher discriminant analysis. In: Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition, Washington, DC, USA, 2002, 187～191

11 Guo G., Li S. Z., Kapluk C.. Face recognition by support vector machines. In: Proceedings of the 4th International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000, 196～201

12 Liu Wei, Wang Yun-Hong. Null space based kernel fisher discriminant analysis for face recognition. In: Proceedings of the 6th International Conference on Automatic Face and Gesture Recognition, Seoul, Korea, 2004, 369～374

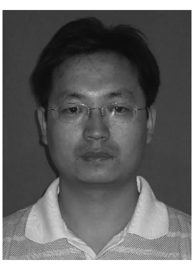
13 Scholkopf Bernhard, Burges Chris. Extracting support data for a given task. In: Proceedings of the 1st Conference on the Knowledge Discovery and Data Mining, Menlo Park, GA: AAAT Press, 1995, 252～257

14 Scholkopf Bernhard. Learning with Kernels. USA: MIT Press, 2002

15 Zhang B., Chen X., Shan S., Gao W.. Nonlinear face recognition based on maximum average margin criterion. In: Proceedings of the IEEE Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, 554～559

16 Cui G., Gao W.. Svms for few examples-based face recognition. In: Proceedings of the International Conference on Acoustic, Speech and Signal Processing, Hong Kong, 2003, 381～384

17 Gao W., Cao B., Shan S.. The cas-peal large-scale face database and evaluation protocols. JDL, CAS, Beijing: Technical Report No. JDL_TR_04_FR_001, 2004



ZHANG Bao-Chang, born in 1976, Ph. D. candidate. His research interests include face recognition, pattern analysis, machine learning, and computer vision.

CHEN Xi-Lin, born in 1965, Ph. D., professor. His research interests include computer vision, pattern recognition, machine learning, and intelligent human-computer interface.

SHAN Shi-Guang, born in 1975, Ph. D., associate researcher. His research interests include pattern analysis, machine learning, and computer vision. He is especially absorbed in face recognition related research topics.

GAO Wen, born in 1956, Ph. D., professor, Ph. D. supervisor. His research interests include multimedia data compression, image processing, computer vision, multimodal interface, artificial intelligence and virtual reality.

Background

This work is conducted in the ICT-ISVISION Joint R&D Lab for face recognition, Institute of Computing Technology Chinese Academy of Sciences, which has been working on face recognition related researches for more than 9 years as the first specialized lab on face recognition in China. This paper is partly supported by the National Natural Science Foundation of China under project "The Study of Identification Based on Biometrics" with grant No. 60332010, 100 Talents Program of CAS, Program of New Century Excellent Talents in University(NCET-04-0320), and ISVISION Technology Co. Ltd. The group has always been aiming at inventing fully innovative kernel technologies and dealing with key issues in Biometrics, especially in face recognition area, based on which, multiple disciplines such as pattern recognition, machine learning, and computer vision, are theoretically studied. In the last 9 years, plenty of research work has been done on the basic theories in face recognition, as well as the practical engineering techniques. About 100 research papers had been published by the members of the group. Many papers has been published or accepted by some well-known journals or proceedings, such as IEEE Transactions on Image Processing, IEEE international conference on Computer Vision and Pattern Recognition, IEEE international conference on Computer Vision, etc.. And many of the research fruits have been successfully applied to practical applications such as bank surveillance, access control, and time-card systems.

The discriminant analysis is one of critical issues for the statistic-based face recognition methods. The work presented in this paper is just motivated by the research about the relationship between Support Vector Machine and kernel Fisher analysis. The property of the SVM has been proven by them that it is of the nullspace property in terms of the support vector based within-class scatter matrix. The authors also find that SVM normal vectors can be used to make discriminant analysis, therefore, this paper proposes the support vector based discriminant analysis method for face recognition.