中图法分类号：TP391　　文献标志码：A　　文章编号：1006-8961(2011)04-0618-07
论文索引信息：卢桂馥，林忠，金忠．具有统计不相关性的核化图嵌入算法[J]．中国图象图形学报，2011，16(4)：618-624

具有统计不相关性的核化图嵌入算法

卢桂馥[1),2)]，林忠[1)]，金忠[1)]
[1)]（南京理工大学计算机科学与技术学院，南京　210094）　[2)]（安徽工程科技学院计算机科学与工程系，芜湖　241000）

摘　要：提出统计不相关的核化图嵌入算法，为求解各种统计不相关的核化降维算法提供了一种统一方法。与已有核化降维算法相比，新的特征提取方法降低甚至消除了最佳鉴别矢量间的统计相关性，提高了识别率。通过在 ORL，YALE 和 FERET 人脸库上的实验结果表明，提出的具有统计不相关的核化图嵌入算法在识别率方面好于已有的核算法。另外，揭示了统计不相关的核化图嵌入与已有的核化图嵌入的内在关系。

关键词：最佳鉴别矢量；统计不相关；核化图嵌入

Uncorrelated kernel extension of graph embedding

Lu Guifu[1),2)], Lin Zhong[1)], Jin Zhong[1)]
[1)] (*School of Computer Science and Technology*, *Nanjing University of Science and Technology*, *Nanjing* 210094, *China*)
[2)] (*Department of Computer Science and Engineering*, *Anhui University of Technology and Science*, *Wuhu* 241000 *China*)

Abstract: An uncorrelated kernel extension of graph embedding which provides a unified method for computing all kinds of uncorrelated kernel dimensionality reduction algorithms is proposed. Compared with kernel dimensionality reduction methods, the proposed method is better in terms of reducing or eliminating the statistical correlation between features and improving the recognition rate. The experimental results on ORL, YALE and FERET face databases show that the proposed uncorrelated kernel extension of graph embedding method is better than other methods in terms of recognition rate. Besides, the relation between uncorrelated kernel extension of graph embedding and kernel extension of graph embedding is revealed.

Keywords: optimal discriminant vectors; statistically uncorrelation; kernel extension of graph embedding (KGE)

0　引　言

人脸识别由于在科学上的挑战性及其潜在的应用已成为计算机视觉和模式识别领域中的热门课题。对于人脸识别，其样本维数往往非常高。当样本维数较高时，往往会导致"维数灾难"，因此对高维样本进行降维显得十分必要。在降维时，如何最大程度地保留（或加强）样本之间的可分性一直是模式识别领域的一个研究热点。基于此，很多降维方法被提出。主分量分析（PCA）[1]和基于 Fisher 准则线性的鉴别分析（LDA）[2-3]是应用最广泛的两个线性降维方法。PCA 是一种无监督学习算法，其目的是寻找在最小均方差意义下最能代表原始数据的投影方向。LDA 作为一种有监督学习算法，其原理是使得降维后同类样本之间的距离最小，而异类样本之间的距离最大。受流形学习算法的启发，很多基于局部二阶统计量的线性降维方法被提出，它们不是以类中心来衡量异类或同类样本的距离，而是用每个样本与它邻近的样本之间的距离来衡量的，

收稿日期：2009-06-15；**修回日期：**2009-11-27
基金项目：国家高技术研究发展计划（863）项目（2006AA01Z119）；国家自然科学基金项目（60873151，60632050）；高等学校博士学科点专项科研基金项目（20060288013）；江苏省 2010 年度普通高校研究生科研创新计划项目（178）。
第一作者简介：卢桂馥（1976—　），男，讲师。模式识别与人工智能博士研究生，主要从事人工智能，模式识别的研究。
E-mail：luguifu_jsj@163.com。

较为典型的有局部保持投影(LPP)[4]、无监督鉴别投影(UDP)[5]、邻域保持嵌入(NPE)[6]。LPP 算法能保持数据的局部结构信息,而 UDP 算法是 LPP 算法在数据为均匀分布时的一种特殊情况[7]。NPE 算法能够保持数据的局部流形结构。

以上的各种降维方法都是基于线性变换的特征抽取方法,最终抽取得到的是线性特征。当人脸特征存在于受光照、姿态、表情变化等影响的复杂非线性结构空间中时,这些线性算法就有其不可避免的缺陷,它们并没有办法更好地发掘这些非线性因素。

当前,核方法已经成为模式识别领域的一个迅速发展方向,上述线性降维算法都有其相对应的核算法(PCA 和 KPCA,LDA 和 KDA,LPP 和 KLPP,NPE 和 KNPE),并且这些核算法都可以纳入核化图嵌入(KGE)框架[8-11]。

在降维方法中,除了选择一个合适的可分性准则外,另外一个关键点就是如何对所要求解的基向量之间加一个合适的约束。Jin 等人从统计不相关的角度考察了 Fisher 线性鉴别变换,进而导出了相应的基于统计不相关的 Fisher 鉴别分析[12],并证实了算法的有效性。Lu 等人将统计不相关引入线性图嵌入,并提出了基于统计不相关的线性图嵌入算法[13]。

本文把统计不相关引入核化图嵌入算法,为各种核化降维算法提供一个求解其相应的统计不相关降维算法的统一框架,同时也为求解具有正交鉴别矢量的核算法提供了一个统一框架。仿真实验表明,引入统计不相关可以提高已有的核化降维算法的识别率,另外,还揭示了统计不相关核化图嵌入算法和已有的核化图嵌入算法之间的关系。

1　核化图嵌入框架[8-11]

设 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}\in\mathbf{R}^{m\times n}$ 为训练样本集,其中 m 为样本的维数,n 为样本数,$G=\{\boldsymbol{X},\boldsymbol{W}\}$ 为无向有权图,每个样本点 $\boldsymbol{x}_i$ 为图中的一个顶点,$\boldsymbol{W}$ 为相似度矩阵,$\boldsymbol{W}_{ij}$ 表示样本 i 和 j 的相似度。一个图的拉普拉斯矩阵 $\boldsymbol{L}$ 和对角矩阵 $\boldsymbol{D}$ 定义为

$$\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W},\quad \boldsymbol{D}_{ii}=\sum_j \boldsymbol{W}_{ij},\quad \forall i \tag{1}$$

为简单起见,先讨论 1 维的情形,很容易从 1 维推广到多维。图嵌入的目标是找到 1 个原数据的低维表示 $\boldsymbol{y}=\{y_1,y_2,\cdots,y_n\}^{\mathrm{T}}$,其中 y_i 样本点是 $\boldsymbol{x}_i$ 的低维表示。为了保持原高维空间中顶点间的相似性,$\boldsymbol{y}$ 可通过最小化下式得到

$$\sum_{i,j}(y_i-y_j)^2\boldsymbol{W}_{ij}=2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{y} \tag{2}$$

为了防止求出没意义的解,给式(2)加入约束:$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y}=1$,则式(2)变为

$$\begin{cases}\min & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{y}\\ \text{s.t.} & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y}=1\end{cases} \tag{3}$$

易知,可把式(3)写为

$$\boldsymbol{y}^*=\arg\min_{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y}=1}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{y}=\arg\min\frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y}} \tag{4}$$

由于 $\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W}$,则式(4)可以变为

$$\boldsymbol{y}^*=\arg\max_{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y}=1}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y}=\arg\max\frac{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y}}{\boldsymbol{y}^{\mathrm{T}}\boldsymbol{D}\boldsymbol{y}} \tag{5}$$

假设从低维空间到高维空间是一个线性映射,即 $\boldsymbol{y}=\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}$,则式(5)可变为

$$\boldsymbol{a}^*=\arg\max\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}} \tag{6}$$

$\boldsymbol{a}^*$ 可通过求解下式最大特征值所对应的特征向量得到。

$$\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a}=\lambda\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{a} \tag{7}$$

如果需要 d 个投影矢量 $\boldsymbol{A}=[\boldsymbol{a}_1,\cdots,\boldsymbol{a}_d]$,则式(6)变为[11]

$$\boldsymbol{A}^*=\arg\max tr[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{A})] \tag{8}$$

$\boldsymbol{A}^*$ 可通过求解式(7)各个最大特征值所对应的特征向量得到。上述方法就称为线性图嵌入(LGE),它为各种线性降维方法提供了一个统一的框架。通过选择不同的 $\boldsymbol{W}$ 和 $\boldsymbol{D}$,就可以得到常见的线性降维算法如 LDA,LPP,NPE 等。

核算法的基本思想是:首先将原始训练样本通过一个满足 Mercer 条件的非线性映射 ϕ 变换到某一高维特征空间(可能是无限维)特征空间 H 中,然后在高维特征空间 H 中进行线性降维。设训练样本 $\boldsymbol{x}$ 在高维特征空间 H 中的映射为 $\phi(\boldsymbol{x})$,因此,式(6)可以变为

$$\boldsymbol{\varphi}^*=\arg\max\frac{\boldsymbol{\varphi}^{\mathrm{T}}(\phi(\boldsymbol{x}_1),\cdots,\phi(\boldsymbol{x}_n))\boldsymbol{W}(\phi(\boldsymbol{x}_1),\cdots,\phi(\boldsymbol{x}_n))^{\mathrm{T}}\boldsymbol{\varphi}}{\boldsymbol{\varphi}^{\mathrm{T}}(\phi(\boldsymbol{x}_1),\cdots,\phi(\boldsymbol{x}_n))\boldsymbol{D}(\phi(\boldsymbol{x}_1),\cdots,\phi(\boldsymbol{x}_n))^{\mathrm{T}}\boldsymbol{\varphi}} \tag{9}$$

式中,φ 是位于特征空间 H 中的鉴别矢量。按照再生核的理论,鉴别矢量一定位于所有训练样本 $\phi(\boldsymbol{x}_1),\cdots,\phi(\boldsymbol{x}_n)$ 张成的空间内,即

$$\boldsymbol{\varphi}=\sum_{i=1}^{n}\alpha_i\phi(\boldsymbol{x}_i)=\boldsymbol{Q}\boldsymbol{\alpha} \tag{10}$$

式中,$\boldsymbol{Q}=(\phi(\boldsymbol{x}_1),\cdots,\phi(\boldsymbol{x}_n))$,$\boldsymbol{\alpha}=(\alpha_1,\cdots,\alpha_n)^{\mathrm{T}}$。把式(10)代入式(9)可得

$$\boldsymbol{\alpha}^*=\arg\max J^K(\boldsymbol{\alpha})=\arg\max\frac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{KWK\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{KDK\alpha}}\tag{11}$$

式中,$J^K(\boldsymbol{\alpha})=\dfrac{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{KWK\alpha}}{\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{KDK\alpha}}$,$\boldsymbol{K}=\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}$ 是 $n\times n$ 的 Gram 核对称矩阵且

$$\boldsymbol{K}_{ij}=\phi(\boldsymbol{x}_i)^{\mathrm{T}}\phi(\boldsymbol{x}_j)=(\phi(\boldsymbol{x}_i)\cdot\phi(\boldsymbol{x}_j))=k(\boldsymbol{x}_i,\boldsymbol{x}_j)\tag{12}$$

$k(\boldsymbol{x}_i,\boldsymbol{x}_j)$为一给定的核函数。最优的 $\boldsymbol{\alpha}^*$ 可通过求解式(13)最大特征值所对应的特征向量得到:

$$\boldsymbol{KWK\alpha}=\lambda\boldsymbol{KDK\alpha}\tag{13}$$

类似的,如果需要 d 个投影矢量 $\boldsymbol{A}=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_d]$,则式(11)变为[11]

$$\boldsymbol{A}^*=\arg\max tr[(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{KWK}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{KDK}^{\mathrm{T}}\boldsymbol{A})]\tag{14}$$

$\boldsymbol{A}^*$ 可通过求解式(13)各个最大特征值所对应的特征向量得到。上述方法就称为核化图嵌入(KGE),它为各种核算法提供了一个统一的框架。通过选择不同的 $\boldsymbol{W}$ 和 $\boldsymbol{D}$,就可以得到常见的核算法如 KDA,KLPP,KNPE 等。

在后面的叙述中,我们用记号 GE($\boldsymbol{W},\boldsymbol{D}$)来表示式(8),用记号 KGE($\boldsymbol{W},\boldsymbol{D}$)来表示式(14)。

1) LDA 和 KDA 算法

设数据为 c 类,并且第 t 类有 m_t个样本,$m_1+\cdots+m_c=n$。定义

$$W_{ij}^{\mathrm{LDA}}=\begin{cases}1/m_t, & \boldsymbol{x}_i \text{ 和 } \boldsymbol{x}_j \text{ 同属 } t \text{ 类}\\ 0, & \boldsymbol{x}_i \text{ 和 } \boldsymbol{x}_j \text{ 不属于同类}\end{cases}$$

$\boldsymbol{W}^{\mathrm{LDA}}=\boldsymbol{I}$,则 LDA 算法可表示为 GE($\boldsymbol{W}^{\mathrm{LDA}},\boldsymbol{I}$),KDA 算法可表示为 KGE($\boldsymbol{W}^{\mathrm{LDA}},\boldsymbol{I}$)。

2) LPP 和 KLPP 算法

设 $\boldsymbol{N}_k(\boldsymbol{x}_i)$表示 $\boldsymbol{x}_i$ 的 k 近邻集合,定义:

$$W_{ij}^{\mathrm{LPP}}=\begin{cases}\mathrm{e}^{-\frac{\|x_i-x_j\|^2}{\sigma^2}} & \boldsymbol{x}_i\in\boldsymbol{N}_k(\boldsymbol{x}_j)\text{ 或 }\boldsymbol{x}_j\in\boldsymbol{N}_k(\boldsymbol{x}_i)\\ 0 & \text{其他}\end{cases}$$

LPP 算法可以表示为 GE($\boldsymbol{W}^{\mathrm{LPP}},\boldsymbol{D}^{\mathrm{LPP}}$),KLPP 算法可以表示为 KGE($\boldsymbol{W}^{\mathrm{LPP}},\boldsymbol{D}^{\mathrm{LPP}}$)。

3) NPE 和 KNPE 算法

设 $\boldsymbol{M}$ 为 $m\times m$ 矩阵,其定义为:对于 $\boldsymbol{M}$ 的第 i 行,$\boldsymbol{M}_{ij}=0$,如果 $\boldsymbol{x}_j\notin\boldsymbol{N}_k(\boldsymbol{x}_i)$;其他 $\boldsymbol{M}_{ij}$可通过最小化下式得到

$$\min\Big\|\boldsymbol{x}_i-\sum_{j\in N_k(x_i)}\boldsymbol{M}_{ij}\boldsymbol{x}_j\Big\|\qquad\sum_{j\in N_k(x_i)}\boldsymbol{M}_{ij}=1$$

定义$\boldsymbol{W}^{\mathrm{NPE}}=\boldsymbol{M}+\boldsymbol{M}^{\mathrm{T}}-\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}$,$\boldsymbol{D}^{\mathrm{NPE}}=\boldsymbol{I}$。则 NPE 算法可表示为 GE($\boldsymbol{W}^{\mathrm{NPE}},\boldsymbol{I}$),KNPE 算法可表示为 KGE($\boldsymbol{W}^{\mathrm{NPE}},\boldsymbol{I}$)。

2　统计不相关的核化图嵌入

设 $\boldsymbol{S}_t^{\phi}=\dfrac{1}{n}\sum_{i=1}^{n}(\phi(\boldsymbol{x}_i)-\boldsymbol{m}_0^{\phi})(\phi(\boldsymbol{x}_i)-\boldsymbol{m}_0^{\phi})^{\mathrm{T}}$ 为高维特征空间 H 的协方差阵,其中$\boldsymbol{m}_0^{\phi}=\dfrac{1}{n}\sum_{i=1}^{n}\phi(\boldsymbol{x}_i)$。因此,要想使抽取的特征统计不相关,则应满足下式

$$\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{S}_t^{\phi}\boldsymbol{\varphi}_j=0\qquad i\neq j\tag{15}$$

为方便起见,假设数据已是中心化了的数据。把式(10)代入式(15)可得

$$\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{S}_t^{\phi}\boldsymbol{\varphi}_j=\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{S}_t^{\phi}\boldsymbol{Q}\boldsymbol{\alpha}_j=\frac{1}{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\alpha}_j=\frac{1}{n}\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_j\tag{16}$$

因此,如要想使抽取的特征统计不相关,需满足条件

$$\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_j=0\tag{17}$$

因此,具有统计不相关的核化图嵌入的求解目标是得到这样一组鉴别矢量 $\boldsymbol{A}=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_d]$,使训练样本在各鉴别矢量上的投影统计不相关,即满足式(17),且满足式(11)。假设已求出前 $k-1$ 个鉴别矢量,则第 k 个矢量可通过求解式(18)得到:

$$\begin{cases}\boldsymbol{\alpha}*=\arg\max\dfrac{\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KWK}\boldsymbol{\alpha}_k}{\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KDK}\boldsymbol{\alpha}_k}\\ \text{s. t.}\quad \boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KDK}\boldsymbol{\alpha}_k=1,\ \boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_1=\\ \qquad\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_2=\cdots=\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_{k-1}=0\end{cases}\tag{18}$$

定理 1　具有统计不相关的核化图嵌入算法的最佳鉴别矢量集的第一个鉴别矢量 $\boldsymbol{\alpha}_1$ 取为式(13)的最大的非零特征值所对应的单位特征矢量;当已求出前 $k-1$ 个鉴别矢量 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_{k-1}$后,第 k 个矢量为式(19)的最大的非零特征值 λ 所对应的单位特征矢量

$$\boldsymbol{M}^{(k)}\boldsymbol{\alpha}_k=\lambda\boldsymbol{\alpha}_k\tag{19}$$

式中

$$\boldsymbol{M}^{(k)}=(\boldsymbol{I}-(\boldsymbol{KDK})^{-1}\boldsymbol{KKA}^{(k-1)}[\boldsymbol{B}^{(k-1)}]^{-1}[\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{KK})(\boldsymbol{KDK}^{-1}\boldsymbol{KWK}$$

$$\boldsymbol{A}^{(k-1)}=[\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_{k-1}]$$

$$\boldsymbol{B}^{(k-1)}=[\boldsymbol{B}_{ij}^{(k-1)}]=[\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KKA}^{(k-1)}$$

$$\boldsymbol{B}_{ij}^{(k-1)}=\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{XDX}^{\mathrm{T}})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_j$$

证明　用拉格朗日乘子法求解式(18)的最优

解,构造拉格朗日辅助函数:

$$C^{(k)} = \boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KWK}\boldsymbol{\alpha}_k - \lambda(\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KDK}\boldsymbol{\alpha}_k - 1) - \mu_1\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_1 - \cdots - \mu_{k-1}\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_{k-1}$$

对上式求偏导并令其为 0,即

$$\frac{\partial C^{(k)}}{\partial \boldsymbol{\alpha}_k} = 0 \Rightarrow$$

$$2\boldsymbol{KWK}\boldsymbol{\alpha}_k - 2\lambda\boldsymbol{KDK}\boldsymbol{\alpha}_k - \mu_1\boldsymbol{KK}\boldsymbol{\alpha}_1 - \cdots - \mu_{k-1}\boldsymbol{KK}\boldsymbol{\alpha}_{k-1} = 0 \tag{20}$$

式(20)两边同时左乘 $\boldsymbol{\alpha}_k^{\mathrm{T}}$,则

$$2\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KWK}\boldsymbol{\alpha}_k - 2\lambda\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KDK}\boldsymbol{\alpha}_k = 0 \Rightarrow$$

$$\lambda = \frac{\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KWK}\boldsymbol{\alpha}_k}{\boldsymbol{\alpha}_k^{\mathrm{T}}\boldsymbol{KDK}\boldsymbol{\alpha}_k} \tag{21}$$

对式(20)两边分别左乘 $\boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1},\cdots,\boldsymbol{\alpha}_{k-1}^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}$,则可得到 $k-1$ 个等式:

$$\mu_1\boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_1 + \cdots + \mu_{k-1}\boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_{k-1} = 2\boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k$$

$$\mu_1\boldsymbol{\alpha}_2^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_1 + \cdots + \mu_{k-1}\boldsymbol{\alpha}_2^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_{k-1} = 2\boldsymbol{\alpha}_2^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k$$

$$\cdots\cdots$$

$$\mu_1\boldsymbol{\alpha}_{k-1}^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_1 + \cdots + \mu_{k-1}\boldsymbol{\alpha}_{k-1}^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_{k-1} = 2\boldsymbol{\alpha}_{k-1}^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k$$

定义

$$\boldsymbol{\mu}^{(k-1)} = [\mu_1,\cdots,\mu_{k-1}]^{\mathrm{T}}, \quad \boldsymbol{A}^{(k-1)} = [\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_{k-1}]$$

$$\boldsymbol{B}^{(k-1)} = [\boldsymbol{B}_{ij}^{(k-1)}] = [\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{A}^{(k-1)}$$

$$\boldsymbol{B}_{ij}^{(k-1)} = \boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_j$$

则上述的 $k-1$ 个等式可写为矩阵形式

$$\boldsymbol{B}^{(k-1)}\boldsymbol{\mu}^{(k-1)} = 2[\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k$$

则

$$\boldsymbol{\mu}^{(k-1)} = 2[\boldsymbol{B}^{(k-1)}]^{-1}[\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k \tag{22}$$

式(20)两边左乘$(\boldsymbol{KDK})^{-1}$可得

$$2(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k - 2\lambda\boldsymbol{\alpha}_k - \mu_1(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_1 - \cdots - \mu_{k-1}(\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{\alpha}_{k-1} = 0$$

上式可表示为矩阵形式

$$2(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k - 2\lambda\boldsymbol{\alpha}_k - (\boldsymbol{KDK})^{-1}\boldsymbol{KKA}^{(k-1)}\boldsymbol{\mu}^{(k-1)} = 0$$

由式(22)可得

$$\{(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k - (\boldsymbol{KDK})^{-1}\boldsymbol{KKA}^{(k-1)}[\boldsymbol{B}^{(k-1)}]^{-1}[\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{KK}(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k\} = \{\lambda\boldsymbol{\alpha}_k(\boldsymbol{I} - (\boldsymbol{KDK})^{-1}\boldsymbol{KK}\boldsymbol{A}^{(k-1)}[\boldsymbol{B}^{(k-1)}]^{-1}[\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{KK})(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\boldsymbol{\alpha}_k\} = \lambda\boldsymbol{\alpha}_k$$

即 $\boldsymbol{M}^{(k)}\boldsymbol{\alpha}_k = \lambda\boldsymbol{\alpha}_k$。证毕。

定理 1　为求解各种统计不相关核化降维算法提供了一个统一框架。当 $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{LDA}}$ 和 $\boldsymbol{D} = \boldsymbol{I}$ 时,通过定理 1 可求得统计不相关 KDA(UKDA)的最佳鉴别矢量;当 $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{LPP}}$ 和 $\boldsymbol{D} = \boldsymbol{D}^{\mathrm{LPP}}$ 时,通过定理 1 可求得统计不相关核 LPP(UKLPP)的最佳鉴别矢量;当 $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{NPE}}$ 和 $\boldsymbol{D} = \boldsymbol{I}$ 时,通过定理 1 可求得统计不相关核 NPE(UKNPE)的最佳鉴别矢量。

实际上,若满足

$$\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{\varphi}_j = \boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{\alpha}_j = \boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}_j = 0,\ i \neq j$$

原共轭正交的条件也就变成了正交,易知有如下推论:

推论 1　具有正交性的核化图嵌入算法的最佳鉴别矢量集的鉴别矢量 $\boldsymbol{\alpha}_1$ 取为式(13)的最大的非零特征值所对应的单位特征矢量;当已求出前 $k-1$ 个鉴别矢量 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_{k-1}$ 后,第 k 个矢量为式(23)的最大的非零特征值 λ 所对应的单位特征矢量:

$$\boldsymbol{M}^{(k)}\boldsymbol{\alpha}_k = \lambda\boldsymbol{\alpha}_k \tag{23}$$

式中

$$\boldsymbol{M}^{(k)} = \{(\boldsymbol{I} - (\boldsymbol{KDK})^{-1}\boldsymbol{KA}^{(k-1)}[\boldsymbol{B}^{(k-1)}]^{-1}[\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{K})(\boldsymbol{KDK})^{-1}\boldsymbol{KWK}\}$$

$$\boldsymbol{A}^{(k-1)} = [\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_{k-1}]$$

$$\boldsymbol{B}^{(k-1)} = [\boldsymbol{B}_{ij}^{(k-1)}] = [\boldsymbol{A}^{(k-1)}]^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{KDK})^{-1}\boldsymbol{KA}^{(k-1)}$$

$$\boldsymbol{B}_{ij}^{(k-1)} = \boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{K}(\boldsymbol{XDX}^{\mathrm{T}})^{-1}\boldsymbol{K}\boldsymbol{\alpha}_j$$

推论 1 也为求解各种具有正交鉴别矢量的核算法提供了一个统一框架。当 $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{LDA}}$ 和 $\boldsymbol{D} = \boldsymbol{I}$ 时,通过推论 1 可求得正交 KDA(OKDA)的最佳鉴别矢量;当 $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{LPP}}$ 和 $\boldsymbol{D} = \boldsymbol{D}^{\mathrm{LPP}}$ 时,通过推论 1 可求得正交核 LPP(OKLPP)的最佳鉴别矢量;当 $\boldsymbol{W} = \boldsymbol{W}^{\mathrm{NPE}}$ 和 $\boldsymbol{D} = \boldsymbol{I}$时,通过推论 1 可求得正交核 NPE(OKNPE)的最佳鉴别矢量。

根据定理 1 求解统计不相关鉴别矢量时需要用递归方式来求解,时间较长。对于一些 $\boldsymbol{D} = \boldsymbol{I}$ 的核算法,如 KDA 和 KNPE 算法,有如下定理:

定理 2　当 $\boldsymbol{D} = \boldsymbol{I}$ 时,对于式(13)的任意两个不相等的非零广义特征值 λ_i 和 λ_j,其对应的鉴别矢量 $\boldsymbol{\alpha}_i$ 和 $\boldsymbol{\alpha}_j$ 是统计不相关的。

证明　设 λ_i 和 λ_j 的鉴别矢量为 $\boldsymbol{\alpha}_i$ 和 $\boldsymbol{\alpha}_j$。当 $\boldsymbol{D} = \boldsymbol{I}$时,则根据式(13)有

$$\boldsymbol{KWK}\boldsymbol{\alpha}_i = \lambda_i\boldsymbol{KDK}\boldsymbol{\alpha}_i = \lambda_i\boldsymbol{KK}\boldsymbol{\alpha}_i,$$

$$\boldsymbol{KWK}\boldsymbol{\alpha}_j = \lambda_j\boldsymbol{KDK}\boldsymbol{\alpha}_j = \lambda_j\boldsymbol{KK}\boldsymbol{\alpha}_j$$

因为 $\boldsymbol{W}$ 为对称阵,有

$$\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KWK}\boldsymbol{\alpha}_j = (\boldsymbol{KWK}\boldsymbol{\alpha}_i)^{\mathrm{T}}\boldsymbol{\alpha}_j = (\lambda_i\boldsymbol{KK}\boldsymbol{\alpha}_i)^{\mathrm{T}}\boldsymbol{\alpha}_j = \lambda_i\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_j\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KWK}\boldsymbol{\alpha}_j = \boldsymbol{\alpha}_i^{\mathrm{T}}(\boldsymbol{KWK}\boldsymbol{\alpha}_j) = \boldsymbol{\alpha}_i^{\mathrm{T}}\lambda_j\boldsymbol{KK}\boldsymbol{\alpha}_j = \lambda_j\boldsymbol{\alpha}_i^{\mathrm{T}}\boldsymbol{KK}\boldsymbol{\alpha}_j$$

即有

$$\lambda_i \boldsymbol{\alpha}_i^{\mathrm{T}} \boldsymbol{K}\boldsymbol{K}\boldsymbol{\alpha}_j - \lambda_j \boldsymbol{\alpha}_i^{\mathrm{T}} \boldsymbol{K}\boldsymbol{K}\boldsymbol{\alpha}_j = 0$$

因为 $\lambda_i \neq \lambda_j$，故 $\boldsymbol{\alpha}_i^{\mathrm{T}} \boldsymbol{K}\boldsymbol{K}\boldsymbol{\alpha}_j = 0$，即 $\boldsymbol{\alpha}_i$ 和 $\boldsymbol{\alpha}_j$ 是统计不相关的。证毕。

由定理 2 可知，对于 $\boldsymbol{D} = \boldsymbol{I}$ 的核算法(如 KDA 和 KNPE)，如果核算法的广义特征值是互不相等的，则其鉴别矢量是统计不相关的，也就是与统计不相关的核算法等价。

3 仿真实验及分析

在本文实验中，用 MATLAB 7.0 来实现各种算法，所用计算机的内存为 896 MB，CPU 为 Intel Pentium 双核处理器，主频为 1.67 GHz。采用高斯核函数 $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2/\delta^2)$，$\delta$ 和 LPP，KLPP，ULPP(Uncorrelated LPP)[13]，正交 KLPP (OKLPP)，统计不相关 KLPP(UKLPP)算法中的 σ，其取值参照文献[10]，即为 $2^{(b-10)/2.5}\delta_0$，$b = 0, 1, \cdots, 20$，δ_0 为训练样本的标准偏差。对于 LDA 算法先用 PCA 进行降维，PCA 时保留 98% 的能量。在实验中使用的分类器是最近邻分类器。因为由定理 2 可知，UKDA 和 KDA，UKNPE 和 KNPE 在广义特征值互不相等时是等价的，因此在下面的实验中并没有比较 UKDA 和 KDA，UKNPE 和 KNPE 算法的性能。

3.1 实验中所选用的数据库

ORL 标准人脸库由 40 人，每人 10 幅图像组成，其中有些图像是拍摄于不同时期的，人的脸部表情和脸部细节有着不同程度的变化，比如，笑或不笑，眼睛或睁或闭，戴或不戴眼镜；人脸姿态也有相当程度的变化，深度旋转和平面旋转可达 20°；人脸的尺度也有多达 10% 的变化。实验中，图像被处理成 32×32 的形式，图 1 为 ORL 人脸数据库中某个人的 10 幅人脸图像。

Yale 人脸库包括 15 个人的 165 幅灰度人脸图像。每个人由 11 幅照片构成。这些照片在不同的表情和光照等条件下拍摄。实验中，图像被处理成 32×32 的形式，图 2 为 Yale 人脸数据库中某个人的 11 幅人脸图像。

FERET 是一个大规模人脸库，由取自 1 199 个人的 14 000 多幅不同姿态、表情、光照和时期的人脸图像组成。从中选择 200 个人的 1 400 幅人脸图像(每个人 7 幅图像)构成一个 FERET 子库进行实验。实验中，图像处理成 80×80 的形式。图 3 为 FERET 人脸数据库中某个人的 7 幅图像。

图 1 ORL 人脸库中的 10 幅图像

Fig. 1 Ten samples cropped from the ORL face database

图 2 Yale 人脸库中的 11 幅图像

Fig. 2 Eleven samples cropped from the Yale face database

图 3 FERET 人脸库中的 7 幅图像

Fig. 3 Seven samples cropped from the FERET face database

3.2 实验结果与分析

首先考察在不同训练样本下各种算法的识别率。随机在 ORL，Yale 库中每人选择 $i(i = 2, 3, 4)$，在 FERET 库中选择 $i(i = 2, 3)$ 幅图像作为训练样本，剩余的图像作为测试样本。每组实验均重复了 20 次，实验结果见表 1，表中给出了 20 次实验的平均识别率和标准方差。

表 1 在 ORL、Yale、FERET 人脸库中不同方法的识别率对比

Tab. 1 Recognition rate comparison on ORL、Yale、FERET face database with different methods

	样本数	LDA	LPP	ULPP	KLPP	OKLPP	UKLPP
ORL	2	75.8 ±3.3	75.9 ±3.1	79.8 ±3.8	83.0 ±2.5	86.9 ±2.3	87.5 ±2.4
	3	85.1 ±1.9	86.0 ±2.1	89.0 ±1.8	89.6 ±1.4	93.3 ±1.6	94.0 ±1.6
	4	91.3 ±1.9	92.7 ±1.7	94.1 ±1.7	94.1 ±1.6	96.8 ±1.1	97.2 ±1.1
Yale	2	52.1 ±5.6	53.2 ±5.6	54.3 ±5.1	57.7 ±4.1	59.3 ±4.2	59.4 ±2.4
	3	65.4 ±4.6	65.6 ±4.7	67.9 ±3.9	67.9 ±4.0	70.5 ±4.9	70.6 ±4.9
	4	72.1 ±5.4	74.3 ±5.4	75.7 ±5.2	74.4 ±5.8	78.0 ±4.7	78.1 ±4.4
FERET	2	42.9 ±7.4	46.3 ±6.8	55.3 ±6.5	51.9 ±6.8	61.0 ±6.5	61.2 ±6.6
	3	63.0 ±7.9	63.3 ±7.8	68.3 ±7.2	68.5 ±7.9	75.3 ±7.7	75.8 ±7.8

由表 1 可以看出,本文提出的 UKLPP 和 OKLPP 算法在 ORL,Yale 和 FERET 库中都取得了非常好的识别效果。由于 UKLPP 算法消除了抽取的特征之间的统计相关性,UKLPP 算法的识别率最高,而 OKLPP 算法虽然抽取的是正交的鉴别矢量,但是并没有消除抽取的特征之间的统计相关性,其识别率略低于 UKLPP 算法。

接下来比较 KLPP 和 UKLPP 算法的运行效率。表 2 为分别为在 ORL,Yale 库中选择训练样本为 $i(i=2,3,4)$ 和在 FERET 库中选择 $i(i=2,3)$ 时求解 KLPP 和 UKLPP 算法所需时间,单位为 s。

表 2　在 ORL、Yale、FERET 人脸库上求解 KLPP 和 UKLPP 所需时间对比

Tab. 2　Time comparison on ORL、Yale、FERET face database for implementing KLPP and UKLPP

/s

	算法	训练样本 2	训练样本 3	训练样本 4
ORL	KLPP	0.031 3	0.062 5	0.171 9
	UKLPP	0.156 3	0.203 1	0.281 3
Yale	KLPP	0.000 1	0.000 1	0.015 6
	UKLPP	0.000 1	0.015 6	0.031 3
FERET	KLPP	1.687 5	6.328 1	
	UKLPP	53.015 6	59	

由表 2 可以看出,在 ORL 和 Yale 人脸库中,本文的 UKLPP 算法效率略低于 KLPP 算法,但相差不多;而在 FERET 人脸库中,本文的 UKLPP 算法的效率同 KLPP 算法的效率相比相差较大,如何进一步提高本文算法的效率是下一步值得研究的问题。

4　结　论

基于核化图嵌入框架,提出统计不相关的核化图嵌入,为求解各种统计不相关核化降维算法提供了一种统一方法,同时也为求解具有正交鉴别矢量的核算法提供了一种统一算法。另外,还揭示了统计不相关核化图嵌入算法和已有核化图嵌入算法之间的关系。在 ORL,Yale 和 FERET 人脸库上的实验表明,在核化图嵌入中引入统计不相关可以提高识别率。但是本文算法由于是递归实现的,其算法的效率较低,如何进一步提高本文算法的效率是下一步值得研究的课题。

参考文献(References)

[1] Fukunaga K. Introduction to Statistical Pattern Recognition [M]. 2 nd Edition. Boston: Academic Press, 1990: 400-417.

[2] Duda R O, Hart P E, Stork D G. Pattern Classification[M]. 2nd Edition. New York: John Wiley & Sons, 2000: 117-124.

[3] Belhumeur P N, Hespanha Joao P, Kriegman David J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7):711-720.

[4] He Xiaofei, Yan Shuicheng, Hu Yuxiao, et al. Face recognition using Laplacianfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27 (3): 328-340.

[5] Yang Jian, Zhang D, Yang Jingyu, et al. Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29 (4): 650-664.

[6] He X, Cai D, Yan S, et al. Neighborhood preserving embedding [C]//Proc. Int. Conf. Computer Vision, USA: Curran Associates, Inc, 2005: 1208-1213.

[7] Deng Weihong, Hu Jiani, Guo Jun, et al. Comments on globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(8): 1503-1504.

[8] Cai D, He X, Hu Y, et al. Learning a spatially smooth subspace for face recognition [C]//Proc. IEEE Conf. Computer Vision and Pattern Recognition. Minneapolis, USA: Curran Associates, Inc, 2007: 1-7

[9] Cai Deng, He Xiaofei, Han Jiawei. Spectral Regression for Dimensionality Reduction, Department of Computer Science Technical Report No. 2856 [R]. Urbana, USA: University of Illinois at Urbana Champaign, 2007.

[10] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: a general framework for dimensionality reduction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1):40-51.
[11] Xu D, Yan S, Tao D, et al. Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval [J]. IEEE Transactions Image Processing, 2007, 16(11): 2811-2821.
[12] Jin Z, Yang J Y, Lu J F. An optimal set of uncorrelated discriminant features [J]. Chinese Journal of Computers, 1999, 22(10):1105-1108 [金忠,杨静宇,陆建峰. 一种具有统计不相关的最优鉴别向量集[J]. 计算机学报,1999,22(10): 1105-1108.]
[13] Lu Guifu, Lin Zhong, Jin Zhong. Face recognition based on uncorrelated linear extension of graph embedding [C]// Proceeding of the Chinese Conference on Pattern Recognition. Red Hook, NY: Curran Associates, Inc, 2009, 460-464. [卢桂馥,林忠,金忠. 基于统计不相关线性图嵌入的人脸识别[C]//全国模式识别会议. Red Hook, NY: Curran Associates 出版社,2009, 460-464.]