

基于非线性特征提取和 SVM 的人脸识别算法¹

孙大瑞 吴乐南

(东南大学无线电工程系 南京 210096)

摘要: 传统的 PCA 或 LDA 都是从像素的二阶依赖上考虑的, 对于多像素之间的依赖性 or 像素的高阶关系不敏感。该文利用核函数方法提取像素高阶相关, 并与线性 SVM 相结合来进行人脸识别。从 Yale 人脸库上的实验结果可以看出, 非线性特征提取是很有效的, 并且 SVM 分类器的性能优于最近邻分类器。

关键词: 非线性特征提取, KPCA, KDA, 线性 SVM

中图分类号: TP391.42, TN911.73 **文献标识码:** A **文章编号:** 1009-5896(2004)02-0307-05

Face Recognition Based on Nonlinear Feature Extraction and SVM

Sun Da-rui Wu Le-nan

(Department of Radio Engineering, Southeast University, Nanjing 210096, China)

Abstract Both PCA and LDA are performed by only using the second-order statistics among image pixels, and not sensitive to high order statistics in the data. In this paper, the kernel function method is used to extract the high order relations, and the Linear Support Vector Machines (LSVM) is selected to perform the face classification. The experiment on Yale face database shows that the nonlinear feature extraction method is effective, and SVM is better than nearest neighbor classifier.

Key words Nonlinear feature extraction, KPCA, KDA, Linear SVM

1 引言

子空间法在人脸检测、人脸识别、三维目标识别和跟踪等视觉识别领域获得了巨大的成功, 主分量分析 (PCA) 和线性判别分析 (LDA) 是典型的子空间方法。无论 PCA, 还是 LDA, 它们都是从像素的二阶依赖上考虑的, 对于多像素之间的依赖性 or 像素的高阶关系不敏感。图像的高阶统计量 (HOS) 包含了像素间的非线性关系, 比如边缘、曲线等结构信息, 显然这有利于识别。

Scholkopf 将 PCA 推广到非线性领域, 他直接利用像素的高阶依赖, 通过核函数使得非线性问题转化为普通的特征值问题^[1]。相对于传统的 PCA, 这种方法称为核 PCA(KPCA) 或非线性 PCA。Bandat 根据同样的处理方法, 将 LDA 推广到非线性领域, 称为核判别分析 (KDA)^[2]。核方法中的核函数只要满足 Mercer 条件即可, 具体应用中大多采用多项式核或径向基核。Kim 采用“ One-against-all ”方式的线性支撑向量机 (SVM) 作为分类器, 在 ORL 人脸库上应用“ Hold-out ”验证策略得出结论认为, KPCA 性能优于线性的 PCA^[3]。Moghaddam 比较了几种子空间方法的性能^[4], 认为虽然 KPCA 优于 PCA, ICA, 但却不如概率子空间法。Yang 将 KDA 和 KPCA 在人脸识别上进行了比较^[5], 他又称这两种方法为核特征脸和核 Fisher 脸, 在 Yale 和 ORL 库上根据“ Leave-one-out ”交叉验证策略得到的实验结果可知, KDA 性能优于 KPCA, SVM, 并且 SVM 不一定优于 KPCA。核方法的分类器为最近邻分类器 (NN), Yang 未将 SVM 和核方法相结合。

¹ 2002-10-07 收到, 2003-02-28 改回

SVM 是 Vapnik 统计学习理论的直接实现^[6]，它是用来解决两类分类问题的。SVM 的目的是要寻找一个超平面使两类可以无错误地分开，并且分类间隙最大。这个超平面称为最优分离超平面 (OSH)。SVM 分类器在视觉处理领域有着重要的应用^[7]。

本文第 1 节为引言，第 2 节分别介绍 KPCA 和 KDA 的特征提取原理，第 3 节给出多类 SVM 分类方法，第 4 节为实验结果，第 5 节为结论。

2 非线性特征提取

2.1 核主分量分析 (KPCA)

Scholkopf 利用核方法将 PCA 推广到非线性领域，从而达到提取像素高阶相关的目的^[1]。给定一组均值为 0 的样本 $x_k, k = 1, \dots, M, x_k \in R^N, \sum_{k=1}^M x_k = 0$ ，则协方差矩阵为 $C = \frac{1}{M} \sum_{i=1}^M x_i x_i^T$ ，对角化 C 即为特征值问题 $\lambda v = Cv$ ，对于所有特征值 $\lambda \geq 0$ ，且 $v \in R^N \setminus \{0\}$ ，由于 $Cv = \frac{1}{M} \sum_{i=1}^M \langle x_i, v \rangle x_i$ ，因此所有的特征矢量都是 x_1, \dots, x_M 的线性组合，特征值问题可以写为

$$\lambda \langle x_k, v \rangle = \langle x_k, Cv \rangle, \quad k = 1, \dots, M \quad (1)$$

存在非线性映射 $\phi: R^N \mapsto R^F, x \mapsto \phi(x), F > N$ ，协方差矩阵为

$$C^\phi = \frac{1}{M} \sum_{i=1}^M \phi(x_i) \phi(x_i)^T \quad (2)$$

对角化协方差矩阵等效于特征值问题：

$$\lambda v^\phi = C^\phi v^\phi \quad (3)$$

考虑到所有 v^ϕ 都是 $\phi(x)$ 的线性组合，即 $v^\phi = \sum_{i=1}^M \alpha_i \phi(x_i)$ 。式 (3) 两边对 $\phi(x)$ 做内积，得

$$\lambda \langle \phi(x_k), v^\phi \rangle = \langle \phi(x_k), C^\phi v^\phi \rangle, \quad k = 1, \dots, M \quad (4)$$

将 v^ϕ 代入上式

$$\lambda \sum_{i=1}^M \alpha_i \langle \phi(x_k), \phi(x_i) \rangle = \frac{1}{M} \sum_{i=1}^M \alpha_i \left\langle \phi(x_k), \sum_{j=1}^M \phi(x_j) \right\rangle \langle \phi(x_i), \phi(x_i) \rangle, \quad k = 1, \dots, M \quad (5)$$

定义 $K_{ij} = \langle \phi(x_i), \phi(x_j) \rangle$ ，式 (5) 可简化为

$$M \lambda K \alpha = K^2 \alpha \quad \text{或} \quad M \lambda \alpha = K \alpha \quad (6)$$

2.2 核判别分析

Mika 将 LDA 采用相同的处理方式推广到非线性领域，建立了两类核 Fisher 判别分析 (KFD)^[8]。Baudat 进一步将两类问题推广到多类问题，称为推广的判别分析 (GDA)^[2]，其他的称谓含义是一样的，如 KDA，KFLD。

假定类别数为 L ，第 l 类的样本数为 n_l ，有 $\sum_{l=1}^L n_l = M$ 。存在映射 $\phi: R^N \mapsto R^F, x \mapsto \phi(x), F > N$ 。协方差矩阵如式 (2)。

类间散布矩阵

$$B^\phi = \frac{1}{M} \sum_{l=1}^L n_l \bar{\phi}_l \bar{\phi}_l^T \quad (7)$$

其中 $\bar{\phi}_l = \frac{1}{n_l} \sum_{k=1}^{n_l} \phi(x_{lk})$ ，GDA 等效于如下的特征值问题：

$$\lambda C^\phi v^\phi = B^\phi v^\phi \quad (8)$$

与 KPCA 中处理方式同理, 特征矢量 \mathbf{v}^ϕ 也可表示为 $\phi(x_{pq}), p = 1, \dots, L, q = 1, \dots, n_p$ 的线性叠加:

$$\mathbf{v}^\phi = \sum_{p=1}^L \sum_{q=1}^{n_p} \alpha_{pq} \phi(x_{pq}) \quad (9)$$

分别将式 (2), (7), (9) 代入式 (8), 化简, 得

$$\lambda \mathbf{K} \mathbf{K} \boldsymbol{\alpha} = \mathbf{K} \mathbf{W} \mathbf{K} \boldsymbol{\alpha} \quad (10)$$

其中 $\mathbf{K} = (K_{pq})_{p=1, \dots, L, q=1, \dots, L}$, $K_{pq} = (k_{ij})_{i=1, \dots, n_p, j=1, \dots, n_q}$, $(k_{ij})_{pq} = \langle \phi(x_{pi}), \phi(x_{qj}) \rangle$, $\mathbf{W} = (\mathbf{W}_l)_{l=1, \dots, L}$,

\mathbf{W}_l 是 $n_l \times n_l$ 矩阵, 每个元素均为 $1/n_l$ 。

考虑到算法稳定性问题, 没有采用类内散布矩阵, 特征值问题的求解需要一定的技巧, 可参见文献 [2]。为了保证 \mathbf{v}^ϕ 归一化, 最终 $\boldsymbol{\alpha}$ 需要除以 $\sqrt{\boldsymbol{\alpha}^\top \mathbf{K} \boldsymbol{\alpha}}$ 。前述的 KPCA 以非线性方式提供了样本在低维空间的最优表征, 但最优表征并不意味着具有最优的辨别能力。GDA 是 LDA 的非线性推广, 是在最优辨别思想下得到的特征提取方法, 由样本的 GDA 得到的特征表示是辨别特征, 因此, 对比于 KPCA, GDA 空间应具有更好的分类性能。

3 支撑向量机 (SVM) 分类器

设可分样本集为 $(x_i, y_i), i = 1, \dots, n, x \in R^N, y \in \{+1, -1\}$ 是类别标号。N 维空间中线性判别函数的一般形式为 $g(x) = \mathbf{w}^\top x + b$, 归一化 x , 使得 $|g(x)| \geq 1$ 。分类面方程为

$$\mathbf{w}^\top x + b = 0 \quad (11)$$

如果分类面对所有样本正确分类, 则满足

$$y_i(\mathbf{w}^\top x_i + b) - 1 \geq 0, \quad i = 1, \dots, n \quad (12)$$

分类间隔最大等价于 $\|\mathbf{w}\|^2$ 最小, 因此最优分类问题可转化为约束优化问题, 即在式 (12) 的约束下, 求解函数式 $(1/2)\mathbf{w}^\top \mathbf{w}$ 的最小值。由 Lagrange 乘子法可以变为对偶问题:

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \boldsymbol{\alpha}^\top \mathbf{Q} \boldsymbol{\alpha} - \mathbf{e}^\top \boldsymbol{\alpha}, \quad \alpha_i \geq 0, \quad i = 1, \dots, n, \quad \mathbf{y}^\top \boldsymbol{\alpha} = 0 \quad (13)$$

其中 $\mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{n \times 1}$, $Q_{ij} = y_i y_j \mathbf{K}(x_i, x_j)$, $\mathbf{K}(x_i, x_j) = \langle x_i, x_j \rangle$ 。

在线性不可分的情况下, 需要在条件式 (12) 中增加一个松弛项 $\xi_i^{[9]}$, 则约束优化问题变为

$$\min_{\mathbf{w}, b, \xi_i} \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^n \xi_i$$

$$y_i(\mathbf{w}^\top x_i + b) - \xi_i \geq 0, \quad \xi_i \geq 0, \quad i = 1, \dots, n \quad (14)$$

由于 SVM 是面向两类问题的, 对于多类问题可以通过建立多个 SVM 分类器的方式加以解决。一种方法就是 “One-against-one”, 如果有 K 个类别, 则每两类都需要建立一个 SVM, 共有 $K(K-1)/2$ 个 SVM, 采用投票的方式进行分类。对于输入的样本, 使之通过所有 SVM 分类器, 最后票数最多的类即为样本类别。

4 实验结果

本实验是在 Yale 人脸库上进行的。库中含有 15 个人, 每人 11 个图像, 具有表情 (眨眼、微笑、恐惧等)、光照 (侧向光) 和附属物 (如眼镜) 的变化, 如图 1 所示。

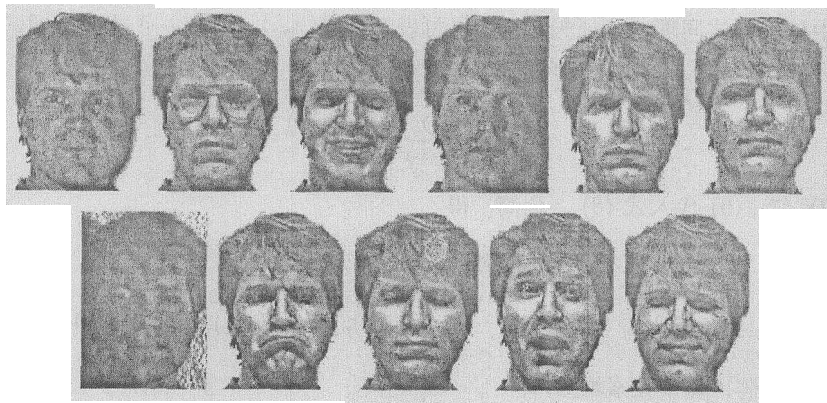


图1 Yale库中人脸图像示例

实验中主要考察了非线性特征提取的问题,即KPCA和KDA在余弦分类器和SVM分类器下的性能。我们采用线性SVM分类器,核方法特征提取中采用多项式核函数 $K_{ij} = \langle \phi(x_i), \phi(x_j) \rangle = ((x_i, x_j))^d$ 。表1给出的是每人训练样本为4,检测样本为7时的实验结果。

在表1中, $d=1$ 对应的是线性情况。由表1可以看出,非线性特征提取的性能优于线性特征提取。在多项式核 d 的指数较小时,如果特征提取方法(KPCA或KDA)确定,一般来说,SVM分类器优于cos分类器。如果分类器确定,KDA优于KPCA。但当 d 较大时(如 d 为10),SVM分类器反而不如cos分类器。这说明SVM分类器并不总是最优的。由表1还可以看出,随多项式指数的增加,SVM分类器性能有所下降。实际上,从更大的指数范围看,cos分类器性能也是下降的,这是因为并非所有的非线性特征都有利于识别分类^[5],并且高次非线性特征的辨别能力有可能低于较低次的非线性特征。

表2给出的是不同非线性特征提取方法和相应SVM的典型训练时间,与表1识别性能比较,增加的核方法训练时间和SVM的附加训练时间是值得的。图2给出的是随特征数目的不同,两种特征提取方法的识别率变化。由图中可以看出,如果特征维数较低,则SVM分类器性能不如cos分类器,随特征维数的增大(对KPCA须大于20,对KDA须大于12),SVM分类器性能优于cos分类器。

表1 训练样本为4, KPCA(特征数目为40)和KDA的识别性能(%)

多项式指数 d	KPCA		KDA	
	cos 分类器	SVM 分类器	cos 分类器	SVM 分类器
1	92.38	97.14	94.29	97.14
2	93.33	97.14	96.19	98.10
3	93.33	95.24	96.19	97.14
4	94.29	95.24	96.19	96.19
5	94.29	95.24	96.19	96.19
6	94.29	95.24	95.24	96.19
10	95.24	93.33	95.24	94.29

表2 KPCA、KDA特征提取和SVM分类器的训练时间(s)

	$d=1$	$d=2$	$d=3$	SVM
KPCA	0.99	1.05	1.15	0.22
KDA	2.53	2.58	2.75	0.11

一般地,对于给定的样本,我们总能找到足够的特征建立一个合适的SVM分类器,其性能优于最近邻(NN)分类器。这是因为NN严重地依赖于训练样本所包容的类内变化和类间变化,在小训练集情况下,NN分类器的推广能力较差;而SVM是以结构风险最小化为基础的,它不仅减小了训练集的错分风险,也减小了未见样本的推广的错分风险^[6]。因此,SVM可以

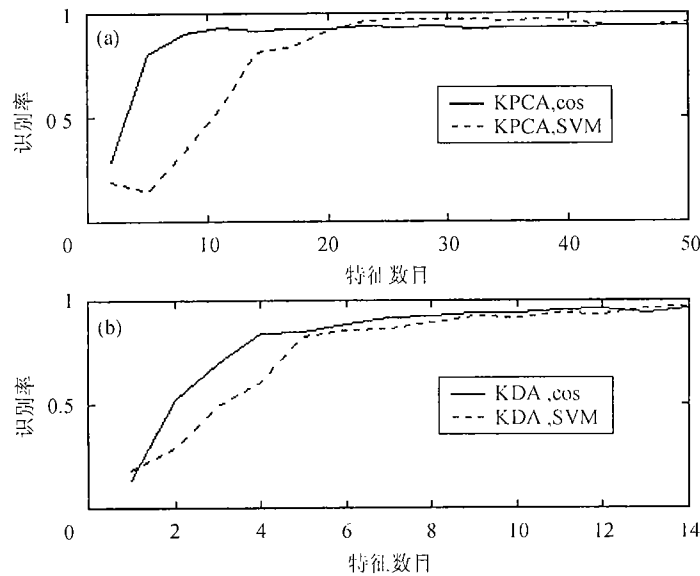


图 2 训练样本为 4, $d = 3$ 时, 不同特征数目下的识别率

获得比 NN 更优的性能。

5 结论

本文主要研究了基于核的非线性特征提取问题, 通过核函数的使用达到提取像素高维相关性的目的。由于传统的 PCA 和 LDA 只利用二阶相关, 因此可以称为线性 PCA 或线性判别分析。文中核函数为多项式形式, 高次多项式反映了高阶相关。我们也实验了核方法和线性 SVM 相结合的情况, 由 Yale 人脸库上实验结果可知, 非线性特征提取性能确实优于线性情况。KDA 的性能也优于 KPCA, 这和线性情况下是一致的。从实验中还可以看出线性 SVM 的优越性。我们工作的下一步将集中在研究非线性 SVM 的性能以及 SVM 分类器的优化上。

参 考 文 献

- [1] Scholkopf B, Smola A, Muller K. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 1998, 10(5): 1299-1319.
- [2] Baudat G, Anouar F. Generalized discriminant analysis using a kernel approach. *Neural Computation*, 2000, 12(10): 2385-2404.
- [3] Kim K I, Jung K, Kim H J. Face recognition using kernel principal component analysis. *IEEE Signal Processing Letters*, 2002, 9(2): 40-42.
- [4] Moghaddam B. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, PAMI-24(6): 780-788.
- [5] Yang M H. Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods. Proc. of 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition, Washington D. C., 2002: 215-220.
- [6] Vapnik V. *The Nature of Statistical Learning Theory*, New York, NY: Wiley, 1998, Chapter 5.
- [7] Guo G, Li S Z, Chan K L. Support vector machine for face recognition. *Image and Vision Computing*, 2001, 19(9-10): 631-638.
- [8] Mika S, Ratsch G, et al.. Fisher discriminant analysis with kernels. In Y. H. Hu, J. Larsen, E. Wilson, S. Douglas, ed., *Neural Networks for Signal Processing*, IEEE, 1999, IX: 41-48.
- [9] Scholkopf B, Smola A, et al.. New support vector algorithms. *Neural Computation*, 2000, 12(5): 1207-1245.

孙大瑞: 男, 1975 年生, 博士生, 研究方向为人脸识别、计算机视觉等。

吴乐南: 男, 1952 年生, 教授, 博士生导师, 研究方向为多媒体信息处理、智能信息处理等。