

基于概率主成分分析的人脸识别

刘直芳, 游志胜, 王运琼

(四川大学 计算机学院 图形图像研究所, 四川 成都 610065)

摘要:人脸自动识别是模式识别和图像处理等学科的一大研究热点, 在身份鉴别、信用卡识别、护照核对以及监控系统等方面有着广泛的应用。提出一种基于概率主成分分析方法(PPCA)的人脸识别, 该方法与传统的主成分分析(PCA)相比, 克服了简单的“丢弃”其他非主成分因子, 在 PPCA 中将“丢弃”因子作为噪声成分进行估计, 同时 PPCA 方法是一种基于概率模型的方法, 因此很容易延伸为混合模型, 对于 PPCA 概率模型参数, 提出利用 EM 算法对其进行估计。用两个不同的数据集(姿势表情变化集和光照变化集), 将 PPCA 人脸识别算法和传统的 PCA 算法进行比较, 基于 PPCA 的人脸识别算法中的“丢弃”方差的收敛速度快于传统的 PCA 算法。实验结果表明, 无论是姿势表情变化集, 还是光照变化集, PPCA 算法的识别率都优于传统的 PCA 识别算法。

关键词: 主成分分析; 概率主成分分析; 人脸识别

中图分类号: TP391 **文献标识码:** A **文章编号:** 1007-2276(2004)01-0071-05

Face recognition based on PPCA

LIU Zhi-fang, YOU Zhi-sheng, WANG Yun-qiong

(Institute of Image & Graphic, Department of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract: Automatic face recognition was an active research area in the last decade. With the increased importance of security and organization, identification and authentication methods were developed into a key technology in various areas such as entrance control in building. Face recognition method based on probabilistic principle component analysis (PPCA) was proposed. However, a notable feature of traditional PCA was the absence of an associated probabilistic model for the observed data. A probabilistic formulation of PCA from a Gaussian latent variable model was obtained, which was closely related to statistical factor analysis. The parameters of PPCA could be determined using EM algorithm. In experiments, the proposed methods have been successfully evaluated using two different datasets. The experimental results show that the face recognition method based on PPCA is superior to the method based on the traditional PCA.

Keywords: Principle component analysis; Probabilistic principle component analysis; Face recognition

收稿日期: 2003-04-08; 修订日期: 2003-08-14

作者简介: 刘直芳(1974-), 女, 四川达州人, 博士生, 主要从事图像处理、计算机视觉、模式识别、人工智能和智能交通方面的研究。

0 引言

人脸自动识别是模式识别和图像处理等学科的一大研究热点,它在身份鉴别、信用卡识别、护照核对以及监控系统等方面有着广泛的应用。人脸图像由于受光照、表情以及姿态等因素的影响^[1],同一个人的脸像矩阵差异较大。因此,在进行人脸识别时,所选取的特征必须对上述因素具备一定的稳定性和不变性。当前大多数人脸识别算法是基于无监督统计方法^[2],其中 PCA 是当前用得最多的方法^[3]。Sirov 和 Kirty^[4]首先用 PCA 算法来表示人脸, Turk 和 Pentand^[5]提出了著名的基于 PCA 变换的人脸识别算法——特征脸识别方法。

主成分分析是从可观测的显式变量中提取信息,组成不可直接观测的隐含变量。它采用的主要原则是使方差最大,以尽可能多地保留原变量所包含的信息,同时又能用尽可能少的主成分替代原有变量,从而使问题变得简单。概率 PCA 是传统 PCA 的延伸,它由 Tipping 和 Bishop^[6]等人首先提出,其目的是为 PCA 定义一个恰当的概率模型。在传统 PCA 中,子空间外的信息只是简单的丢弃,然而在 PPCA 中,这些信息将作为高斯噪声进行估计。对于概率 PCA 模型,可以通过最大可能函数或 EM 算法来估计模型参数而得到最佳概率模型。PPCA 与传统 PCA 相比,它定义了一个恰当的概率模型,这个模型能很容易地延伸为混合模型^[7],同时该模型的参数能用 EM 算法训练获得。

1 隐含变量模型和 PCA

1.1 隐含变量模型

一个隐含变量模型是从 d 维观测矢量数据 t 中寻找 q 维隐含变量矢量 x :

$$t = y(x; w) + \varepsilon \quad (1)$$

式中 $y(x; w)$ 是隐含变量 x 关于参数 w 的函数; ε 是与 x 独立的噪声。隐含变量中最典型的是因子分析^[6]。在公式(1)中的 $y(x; w)$ 函数是变量 x 的线性函数:

$$t = Wx + \mu + \varepsilon \quad (2)$$

式中 t 为观测变量; x 为隐含变量,且 $x \sim N(0, I)$ 分

布;参数 W 为 $d \times q$ 的因子矩阵(因子荷载);参数 μ 为均值; ε 为高斯噪声,且 $\varepsilon \sim N(0, \Psi)$, Ψ 为对角线矩阵。根据公式(2),则观测数据矢量服从下列分布:

$$t \sim N(\mu, C) \quad (3)$$

式中 $C = \Psi + WW^T$

1.2 PCA 算法

PCA 算法是一种线性投影技术,对于一般的 PCA 算法:

$$U = W^T(t - \mu) \quad (4)$$

式中 U 是 q 维投影矢量; t 为 d 维原始观测矢量数据; W 是协方差矩阵 S 的前 q 个正交的特征矢量。

$$S = \frac{1}{N} \sum_{n=1}^N (t_n - \mu)(t_n - \mu)^T \quad (5)$$

则 W 满足:

$$Sw_j = \lambda_j w_j, \quad W = (w_1, w_2, \dots, w_q)$$

其最佳的线性重建为:

$$\hat{t}_n = WU + \mu \quad (6)$$

PCA 是基于投影在重建方差最小的投影方法,其最小投影方差为:

$$\zeta = \sum \|t_n - \hat{t}_n\|^2 \quad (7)$$

在 PCA 中,其丢失的方差为:

$$\sigma_{\text{PCA}}^2 = \frac{1}{(d-q)} \sum_{j=q+1}^d \lambda_j \quad (8)$$

2 概率 PCA 和基于 PPCA 的人脸识别

2.1 概率模型

根据参考文献[6]和[7],给出概率 PCA 的概率模型。假设噪声服从 $\varepsilon \sim N(0, \sigma^2 I)$ 和隐含变量 x 的先验分布为:

$$p(x) = (2\pi)^{-\frac{d}{2}} \exp\left\{-\frac{1}{2} x^T x\right\} \quad (9)$$

则根据隐含变量公式(2)可知,在隐含变量 x 的条件下观测数据 t 的概率分布为:

$$p(t | x) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\left\{-\frac{1}{2\sigma^2} \|t - Wx - \mu\|^2\right\} \quad (10)$$

故观测数据 t 的分布为:

$$p(t) = \int p(t|x)p(x)dx = (2\pi)^{-\frac{d}{2}} |C|^{-\frac{1}{2}} \times \exp\left\{-\frac{1}{2}(t-\mu)^T C^{-1}(t-\mu)\right\} \quad (11)$$

式中 $C = \sigma^2 I + WW^T$ 为观测模型方差。

通过贝叶斯概率公式 $p(x|t) = \frac{p(x)p(t|x)}{p(t)}$, 可以得到隐含变量 x 的后验分布:

$$p(x|t) = (2\pi)^{-\frac{q}{2}} |\sigma^{-2}M|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\{x - M^{-1} \times W^T(t-\mu)\}^T (\sigma^{-2}M) \{x - M^{-1}W^T(t-\mu)\}\right] \quad (12)$$

式中 $M = \sigma^2 I + W^T W$ 。注意: M 的大小为 $q \times q$, 而 C 的大小为 $d \times d$ 。在该模型下, 观测数据的 log 函数为:

$$L = \sum_{n=1}^N \ln\{p(t_n)\} = -\frac{N}{2} \{d \ln(2\pi) + \ln |C| + \text{tr}(C^{-1}S)\} \quad (13)$$

2.2 PPCA 的 EM 算法

为了得到概率模型的参数 W 和 σ^2 , 在 PPCA 的 EM 算法中, 将隐含变量 x 看作“不完全”的数据, “完全”数据包含观测矢量和这些隐含变量^[6,7]。给定一个观测矢量 t_n , 并不知道相应的隐含变量 x_n , 但是知道观测数据和隐含变量的联合分布 $p(t_n, x_n)$, 根据公式(13)能得到“完全”数据的 log 函数。首先在 EM 算法中的 E-步, 计算给定观测 t_n 下的隐含变量 x_n 的后验概率 $p(x_n|t_n)$, 在 M-步, 通过最大化“完全”数据的 log 函数的期望, 获得新参数 \tilde{W} 和 $\tilde{\sigma}^2$ 。

因此, “完全”数据的 log 函数为:

$$L_c = \sum_{n=1}^N \ln\{p(t_n, x_n)\} \quad (14)$$

利用公式(9)和(10)得到 $p(t_n, x_n)$:

$$p(t_n, x_n) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\left\{-\frac{\|t_n - Wx_n - \mu\|^2}{2\sigma^2}\right\} \times (2\pi)^{-\frac{q}{2}} \exp\left\{-\frac{1}{2}x_n^T x_n\right\} \quad (15)$$

在 E-步中, 得到关于 $p(x_n|t_n, W, \sigma^2)$ 分布的期望:

$$\langle L_c \rangle = -\sum_{n=1}^N \left\{ \frac{d}{2} \ln \sigma^2 + \frac{1}{2} \text{tr}(\langle x_n x_n^T \rangle) + \frac{1}{2\sigma^2} \|t_n - \mu\|^2 - \frac{1}{\sigma^2} \langle x_n \rangle^T W^T (t_n - \mu) - \right.$$

$$\left. \frac{1}{2\sigma^2} \text{tr}(W^T W \langle x_n x_n^T \rangle) \right\} \quad (16)$$

式中

$$\begin{aligned} \langle x_n \rangle &= M^{-1} W^T (t_n - \mu), \\ \langle x_n x_n^T \rangle &= \sigma^2 M^{-1} + \langle x_n \rangle \langle x_n \rangle^T \\ M &= \sigma^2 I + W^T W \end{aligned} \quad (17)$$

在 M-步中, 最大化 $\langle L_c \rangle$ 关于参数 W 和 σ^2 的偏导数, 并使其为 0, 得到新参数 \tilde{W} 和 $\tilde{\sigma}^2$:

$$W = \left[\sum_n (t_n - \mu) \langle x_n^T \rangle \right] \left[\sum_n \langle x_n x_n^T \rangle \right]^{-1} \quad (18)$$

$$\tilde{\sigma}^2 = \frac{1}{Nd} \sum_{n=1}^N \left\{ \|t_n - \mu\|^2 - 2 \langle x_n^T \rangle \times \tilde{W}^T (t_n - \mu) + \text{tr}(\langle x_n x_n^T \rangle \tilde{W}^T \tilde{W}) \right\} \quad (19)$$

将公式(17)代入公式(18)和(19), 可以得到:

$$\tilde{W} = SW(\sigma^2 I + M^{-1}W^T SW)^{-1} \quad (20)$$

$$\tilde{\sigma}^2 = \frac{1}{d} \text{tr}(S - SWM^{-1}\tilde{W}^T) \quad (21)$$

通过公式(20)和(21)逐步迭代到算法收敛。从公式(20)和(21)可以看出 PPCA 的 EM 算法只与观测数据的协方差矩阵 S 有关。

2.3 基于 PPCA 的人脸识别

前面, 已介绍概率 PCA 算法的原理, 在此直接将该方法用于人脸识别, 只对识别的步骤进行大致叙述。将一幅 $l \times h$ 大小的人脸图像按列构成一个 $d = l \times h$ 维的列向量, N 为训练集的样本数, 则构成训练人脸样本的图像矩阵为 t 。利用上面给出的 PPCA 算法进行人脸识别的步骤如下:

- (1) 对原始训练人脸样本数据 t 进行标准中心化处理得到 t' 。
- (2) 初始化参数 W_0 和 σ_0^2 , 迭代停止条件为 $esp = 0.0001$, 同时其迭代次数大于 5。
- (3) 通过 EM 算法获得概率 PCA 的模型参数 W 和 σ^2 , 即得到人脸图像隐含变量所在的子空间。
- (4) 将训练人脸样本库中每一个人脸图像投影到模型所获得的子空间 $\rho^{(i)} = W^T t'$ 构成特征搜索空间 Ω 。
- (5) 对任一待识别的人脸图像 t_{test} , 将其中心化处理 $t'_{\text{test}} = t_{\text{test}} - \bar{t}$ 。
- (6) 将中心化后的人脸图像 t'_{test} 在特征子空间上进行投影 $\rho_{\text{test}} = W^T t'_{\text{test}}$ 。

(7) 将 ρ_{test} 在特征搜索空间 Ω 中进行匹配, 匹配准则采用余弦距离 $d = \min(d(\rho^{(i)}, \rho_{\text{test}}))$ 。

3 实验结果

采用两种影像集对两种算法分别进行测试, 第一个训练样本集是基于姿势和表情改变的影像库, 称为 P 集, 该影像库里包含 ORL 人脸数据库, 该训练集共有 225 张图像, 有 75 个人, 每个人有 3 种不同姿势或表情图像作为训练样本, 其中有 40 人为 ORL 数据库中的标准人脸, 其他 35 个人为自己影像库里的人, 其相应的测试集为 421 张图像, 与训练集图像不同, 其中 49 人, 每人有 7 种不同的图像, 另外 26 人有 3 种不同表情或者姿势的图像作为测试集。第二个训练样本集是基于光照改变的影像库, 称为 L 集, 该影像库是利用 P 集中的部分人脸图像进行光照处理得到的, 该训练集共有 90 张图像, 有 30 人, 每个人有 3 种不同方向光照下的图像, 其相应的测试集为 150 张图像, 每个人有 5 种与训练集不同光照方向的图像, 文中所用数据集如表 1 所示。

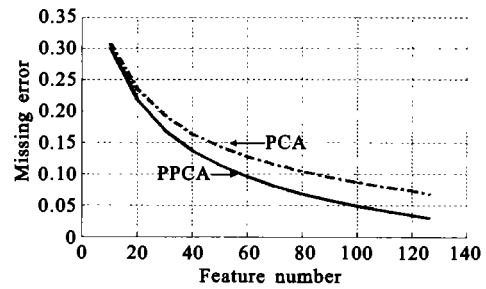
表 1 人脸训练和测试数据集

Table 1 The data sets of face train and test

Data sets	The number	The number	The number
	of person	of train set	of test set
Pose and express (P set)	75	225	421
Lighting (L set)	30	90	150

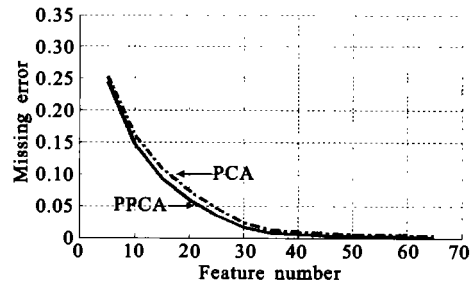
根据公式(8)和(21)分别求出 PCA 丢失的方差和 PPCA 的噪声方差(丢失的方差), 如图 1 所示, 从图 1 可以看出, PPCA 的方差收敛速度快于 PCA。同时图 2 给出了 PCA 和 PPCA 两种方法对不同数据集的特征脸结果。对于 PPCA, 因其模型考虑了隐含变量和噪声信息, 因此根据公式(2)求出的原始影像将优于 PCA, PPCA 根据隐含变量和噪声信息还原的图像如图 3 所示。根据不同的特征数, PCA 和 PPCA 的识别率如图 4 所示, 从图 4 可以看出, 对于 P 集, 利用 PPCA 算法在 75 个特征数下其识别率达到 88.5%, 而 PCA 算法在相同特征数下其识别率只有 85.1%, 尽管 PPCA 的识别率大于 PCA 的识别率, 但它们之间相差不是很大; 对于 L 集, 利用 PPCA 算法在 30 个特征数下其识别率为 91.3%, 而利用 PCA 算法则只有 71.3%, PPCA 算法识别率明显高

于 PCA 算法识别率。可以看出, 无论是 P 集还是 L



(a) P 集的“丢失”方差

(a) The lost variance of P set



(b) L 集的“丢失”方差

(b) The lost variance of L set

图 1 “丢失”数据的方差

Fig. 1 The lost variance in different sets



(a) P 集的 PCA 特征脸

(a) The eigenface of P set by PCA



(b) P 集的 PPCA 特征脸

(b) The eigenface of P set by PPCA



(c) L 集的 PCA 特征脸

(c) The eigenface of L set by PCA



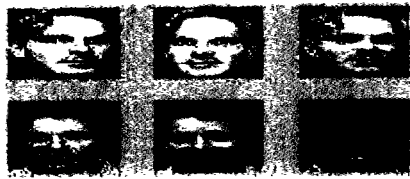
(d) L 集的 PPCA 特征脸

(d) The eigenface of L set by PPCA

图 2 PCA 和 PPCA 特征脸

Fig. 2 The eigenface of PCA and PPCA

集,利用 PPCA 算法进行人脸识别其识别率高于 PCA 算法。图 5 给出部分识别结果。



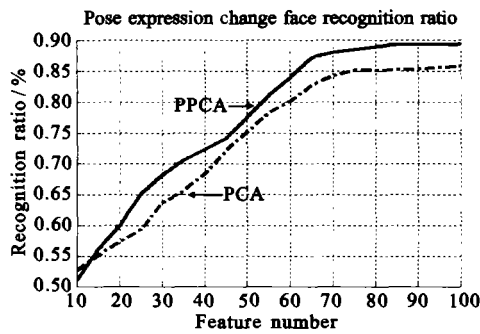
(a) P 集根据 PPCA 概率模型得到的原始图像
(a) The reconstruction face by PPCA in P set



(b) L 集根据 PPCA 概率模型得到的原始图像
(b) The reconstruction face by PPCA in L set

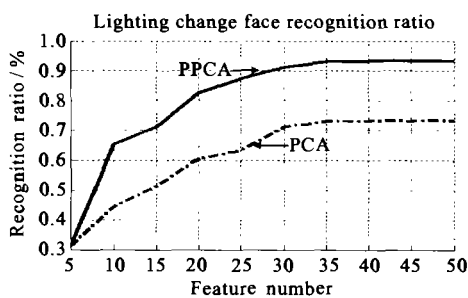
图 3 PPCA 根据公式(2)得到的图像

Fig. 3 The original face by PPCA in different sets using formula (2)



(a) P 集识别率

(a) The recognition ratio of P set



(b) L 集识别率

(b) The recognition ratio of L set

图 4 不同特征数下的识别率

Fig. 4 The recognition ratio in different feature numbers

4 结 论

提出一种基于 PPCA 的人脸识别算法,该算法



(a)



(b)

图 5 P 集(a)和 L 集(b)的部分识别结果

Fig. 5 The part recognition results of (a) P set and (b) L set

通过建立概率模型以及 EM 算法对模型参数的估计,获得隐含变量以及噪声方差信息,使得通过 PPCA 模型恢复得到的原始观测数据效果优于传统的 PCA 算法。利用不同的人脸数据进行测试实验,结果表明,PPCA 人脸识别算法的识别效果优于 PCA 算法的识别效果。

参 考 文 献:

- [1] 杜平. 基于特征脸空间重建的彩色人脸检测方法[J]. 红外与激光工程, 2002, 31(1): 9-13.
- [2] Bartlett M S. Face image analysis by unsupervised learning and redundancy reduction [D]. San Diego: University of California, 1998.
- [3] 凌旭峰. 基于支撑向量机的人脸识别技术[J]. 红外与激光工程, 2001, 30(5): 318-322.
- [4] Sirovich L, Kirby M. Low-dimensional procedure for the characterization of human faces [J]. J Opt Soc Amer A, 1987, 4(3): 519-524.
- [5] Turk M A, Pentland A P. Eigfaces for recognition [J]. J Cognitive Neurosci, 1994, 3(1): 71-86.
- [6] Tipping M E. Probabilistic principle component analysis [J]. Journal of the Royal Statistical Society, 1999, 61(3): 611-622.
- [7] Tipping M E. Mixture of probabilistic principle component analyzers [J]. Neural Computation, 1999, 11(2): 443-482.