

基于局部和全局的特征提取算法及在人脸识别中的应用

张国印 楼宋江 程慧杰 王庆军

(哈尔滨工程大学计算机科学与技术学院 哈尔滨 150001)

摘要 提出了一种基于局部和全局特征的特征提取算法。该算法不仅能保持数据集的局部性,同时也考虑了数据集的全局性,使得降维后的数据既能保持邻近关系,又能从整体上较好地重构和展现。PCA()能较好地展现原数据集,LPP能保持局部邻近关系,算法结合了这两个算法的思想,但由于LPP没有考虑类别信息,故先对LPP进行改进,给出了一种有监督的局部保持投影算法,使得提出的算法能更加有利于分类问题。通过人脸识别实验,验证了算法的正确性和有效性。

关键词 特征提取,局部性,全局性,LPP,主成分分析算法,人脸识别

中国法分类号 TP391 **文献标识码** A

Feature Extraction Algorithm Based on Locality and Globality and its Application in Face Recognition

ZHANG Guo-yin LOU Song-jiang CHENG Hui-jie WANG Qing-jun

(Faculty of Computer Science and Application, Harbin Engineering University, Harbin 150001, China)

Abstract A feature extraction algorithm based on locality and globality was proposed, which on one hand takes the locality of the data into consideration, on the other hand takes the globality of data into account. Consequently, the data after dimension reduction not only preserves the locality relationship, but also reconstructs and represents the original data perfectly. PCA (Principal Component Analysis) can represent the data nicely and LPP (Locality Preserving Projection) can preserve the locality relationship, so the algorithm just hybrid the two of them. But LPP does not utilize the classification information, so first LPP-based algorithm is improved, a supervised version is given, which results in that the algorithm is more suitable for classification task. Experiments in face recognition validate the correctness and effectiveness of the proposed algorithm.

Keywords Feature extraction, Locality, Globality, LPP, PCA, Face recognition

1 引言

作为生物特征识别技术的一个主要方向,人脸识别涉及模式识别、图像处理、计算机视觉等多个研究领域。人脸识别由于具有成本较低,不侵犯使用者隐私权、应用前景广阔等特点,因此具有十分重要的研究意义和实际应用价值,是目前的一个研究热点。

特征提取是人脸识别中一个非常重要的环节,如何提取有效的判别特征是解决问题的关键。在过去几年的时间里,出现了很多人脸识别算法,其中,基于子空间的算法起到了主导作用,最著名的是主成分分析^[1](Principal Component Analysis)和线性判别分析^[2](Linear Discriminant Analysis)。主成分分析算法的主要思想是通过估计数据的二阶统计性质来发现数据的本质线性维数,它是在全局最小重构误差的条件下把高维观察数据投影到低维空间上,通过求解数据点协方差矩阵的最大几个特征值所对应的特征向量所张成的子空间而得到。线性判别分析算法就是求出一个线性子空间,使得所有

样本在这个子空间中,类内样本散度最小,类间样本的散度最大,用它降维后的低维嵌入坐标非常有利于样本的分类。

上述两种方法是线性算法,但人脸空间更可能存在于非线性子空间中。最近相关人员提出了几种流形学习算法,主要用于非线性降维,比如等距映射(ISOMAP)^[3]、局部线性嵌入算法(LLE, Locally Linear Embedding)^[4]、拉普拉斯特征映射(LE, Laplacian Embedding)^[5]等。但是它们都不能解决“out of sample”问题,也就是不能直接映射新数据,并且都是非监督学习算法,因此不能直接用于人脸识别。基于以上问题,He等人提出了一线性算法LPP^[6],该算法是拉普拉斯特征映射的线性版本,能保持数据集的局部邻近关系。

PCA虽然能保持和展现人脸的主要特征,但是它没有考虑人脸的局部特征,有的时候局部特征对分类问题能起到关键作用。LPP算法虽然考虑了局部特性,但是不能把握整体,且没有考虑类别信息,因此不适合分类问题。本文基于以上分析,提出了一种兼顾局部和全局的特征提取算法,并应用于人脸识别,取得了较好的实验结果。

到稿日期:2008-09-24 返修日期:2009-02-27

张国印(1962-),男,博士,教授,主要研究方向为嵌入式、信息安全、模式识别;楼宋江(1982-),男,博士生,主要研究方向为模式识别、机器学习, E-mail: lousongjiangac@163.com.

2 基于局部和全局的特征提取 FELG

2.1 PCA 和全局特征提取

PCA 的基本思想是将数量众多的原始相关变量转换为数量较少的相关变量。通常的做法是将原始变量进行线性组合为若干综合变量,这些综合变量之间相互独立,并且尽可能表示原始变量所包含的信息,选取最大的几个主成分进行分析,这样就可以在尽可能少地损失原有信息的基础上,降低数据的维数,提高运算效率。

数学形式表示为:找到一映射 $Y=W^T X$,把图像 X 投影到低维空间中的特征向量 Y 。其中 $X \in R^D, Y \in R^d, d \ll D, W$ 表示投影矩阵。考虑数据空间中的样本 $\{X_i\}_{i=1}^n, \mu$ 代表它们的均值,这样样本的协方差矩阵可表示为:

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)(X_i - \mu)^T \quad (1)$$

其目标函数为:

$$W_{PCA} = \arg \max_W \sum_{i=1}^n (y_i - \bar{y})^2 = \arg \max_W \|W^T \Sigma W\| \quad (2)$$

其中 \bar{y} 为映射后数据的平均值。

主成分分析投影矩阵的列向量是由 Σ 的前 k 个最大特征值所对应的特征向量组成。

PCA 的主要目标是投影后的图像达到最小重构误差或最小均方意义下的最优表达。因此它是一种全局特征提取方法,并且广泛应用于人脸识别。

2.2 改进 LPP 和局部特征提取

LPP 的目标是保持数据之间的相似性,即原始空间上相邻的数据点在投影空间上也保持相应的邻近关系。假定数据

$$S_{ij} = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{\beta})(1 - \exp(-\frac{\|x_i - x_j\|^2}{\beta})) & x_j \text{ 是 } x_i \text{ 的临近数据点, 并且属于同一类} \\ \exp(-\frac{\|x_i - x_j\|^2}{\beta})(1 + \exp(-\frac{\|x_i - x_j\|^2}{\beta})) & x_j \text{ 是 } x_i \text{ 的临近数据点, 但不属于同一类} \\ 0 & x_j \text{ 不是 } x_i \text{ 的临近数据点} \end{cases} \quad (5)$$

2.3 兼顾局部和全局的特征提取 FELG

同 PCA 和 LPP 不一样,这里研究了数据的整体特性,又研究了样本数据间的局部关系,并考虑了数据的类别信息,使得不同模式的类别尽可能分开。在低维子空间中,不同类彼此远离,同一类彼此靠近,且保持了数据的邻近关系。因此算法从两个方面考虑了目标函数:

$$\max_W \|W^T \Sigma W\| \quad (6)$$

$$\min_W \sum_{ij} (y_i - y_j)^2 S_{ij} \quad (7)$$

$$\text{Constraint: } W^T XDX^T W = 1$$

由以上两节得到的目标函数为:

$$\max_W \|W^T \Sigma W\| \quad (8)$$

$$\min_W W^T XDX^T W \quad (9)$$

$$\text{Constraint: } W^T XDX^T W = 1$$

这里的 S_{ij} 由式(5)给出。

所以可以化为单一目标函数:

$$\max(W^T \Sigma W - W^T XDX^T W) \quad (10)$$

$$\text{s. t: } W^T XDX^T W = 1$$

那么以上最大化问题可以利用 Lagrange 乘数法来解决。

$$L(W, \lambda) = (W^T \Sigma W - W^T XDX^T W) - \lambda(W^T XDX^T W - 1)$$

$$\text{令 } \frac{\partial L(W, \lambda)}{\partial W} = 0, \text{ 得到}$$

集 $X = \{x_1, x_2, \dots, x_n\}$ 含有 n 个样本点,每个样本点 x_i 属于 D 维欧式空间,即 $x_i \in R^D$,且样本数据来自一本征维数为 d ($d \ll D$) 的非线性流形。对于原始空间中的一数据点 x_i ,寻找一投影 W ,使得 $y = W^T x$,通过求解以下最小化问题得到最佳的投影:

$$W_{opt} = \arg \min_W \sum_{ij} (y_i - y_j)^2 S_{ij} = \arg \min_W \sum_{ij} (W^T x_i - W^T x_j)^2 S_{ij} = \arg \min_W W^T XDX^T W \quad (3)$$

其中 S_{ij} 刻画两数据 x_i, x_j 之间的相似性:

$$S_{ij} = \begin{cases} \exp(-\|x_i - x_j\|^2) / t & x_i, x_j \text{ 为邻近点} \\ 0, & \text{其它} \end{cases}$$

附加约束条件 $W^T XDX^T W = 1$,拉普拉斯算子 $L = D - S, D$ 为对角矩阵,其元素为对称矩阵 S 的列向量(或行向量)元素之和,即 $D_{ij} = \sum_j S_{ij}$ 。最后可以通过以下广义特征方程求解:

$$XDX^T W = \lambda XDX^T W \quad (4)$$

假定 w_0, w_1, \dots, w_{d-1} 为特征方程(4)的 d 个解,它们所对应的特征值分别为 $\lambda_0, \lambda_1, \dots, \lambda_{d-1}$,那么可以得到如下映射: $x_i \rightarrow y_i = W^T x_i$,其中 y_i 为 d 维向量, W 为变换矩阵。

LPP 自提出之后,广泛应用于各类识别任务中,但是由于它是一种非监督学习,没有考虑数据的类别信息(Label),所以结果并不理想。为了解决这一问题,本文先对 LPP 算法进行改进,提出了一有监督的局部保持投影。

这里主要对刻画数据 x_i, x_j 之间的相似性度量 S_{ij} 进行改进,使得算法考虑数据集的类别信息,在同等距离条件下,同一类内数据点的相似性比不同类内的相似性大得多,该权值的计算引用了文献[7]:

$$(\Sigma - XDX^T)W = \lambda XDX^T W \quad (11)$$

假定 w_0, w_1, \dots, w_{d-1} 为特征方程(11)的 d 个解,它们所对应的特征值分别为 $\lambda_0, \lambda_1, \dots, \lambda_{d-1}$,那么就可以得到如下映射: $x_i \rightarrow y_i = W^T x_i$ 。其中 y_i 为 d 维向量, W 为变换矩阵。这样图像 X 就能通过 $Y = W^T X$ 嵌入到低维子空间。

3 实验结果及其分析

根据以上提出的算法就可以提取出图像的特征,然后在 ORL 和 Yale 标准人脸库上进行实验。ORL 人脸库有 40 个人,每人 10 张图片,一共 400 张,图像是在不同时期、不同光线强度下拍摄的,其中包括姿态、光照和表情的变化。图 1 是取自 ORL 人脸库第一个人的样本。Yale 人脸库有 15 人,每人 11 张图片,一共 165 张,也受表情、光照等因素的变化。图 2 是来自 Yale 人脸库第一个人的样本。

为了有效地计算,所有图像都缩放成 32×32 像素,即每一图像都可以看成是 $D = 32 \times 32 = 1024$ 维向量。每人分别随机选择 3, 4, 5, 6, 7 幅图像用作训练样本,其它图像用于测试样本,重复 20 次,最后取平均值作为识别结果,采用 NN(最近邻近)分类器进行分类。FELG 与算法 PCA 和 LPP 进行了比较,实验结果如表 1 和表 2 所列。

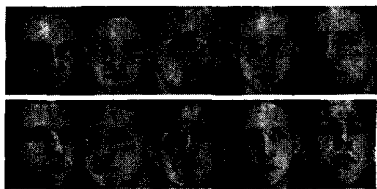


图1 ORL 人脸图像的样本



图2 YALE 人脸图像的样本

表1 算法在 ORL 人脸库中的识别结果(%)

训练样本数	PCA	LPP	FELG
3	78.2	82.1	85.2
4	83.7	87.6	91.3
5	86.8	89.8	93.9
6	89.1	93.3	96.7
7	92.4	94.1	97.9

表2 算法在 Yale 人脸库中的识别结果(%)

训练样本数	PCA	LPP	FELG
3	70.9	73.1	79.6
4	73.4	75.6	82.8
5	73.9	76.2	84.5
6	75.3	78.5	86.9
7	76.8	79.7	87.6

从实验结果中可以看出,各类算法在 ORL 人脸库中的结果比较理想,这是因为其人脸的表情、光照等因素变化较少,适合提取关键特征。从实验数据可以看出,LPP 算法的识别率高于 PCA,这是因为 PCA 只是较好地低维空间中展现

(上接第 269 页)

成份特征根、方差贡献率以及方差累计贡献率(如表 3 所列)和主成份因子得分系数表(如表 4 所列),最终获得所需要的属性约简。

表3 主成份特征根及方差贡献率

成份	特征根 λ	方差贡献率的比重(%)	累积方差贡献率(%)
1	2.829	70.728	70.728
2	0.613	15.331	86.059
3	0.500	12.500	98.559
4	0.058	1.441	100.000

表4 主成份因子得分系数矩阵

主成份	1	2	3	4
P	0.303	0.616	0.655	-2.258
M	0.341	0.012	-0.378	3.226
S	0.274	0.341	1.195	-0.217
X	-0.266	1.065	0.000	1.339

当选择特征值大于 1 的主成份时,累计方差贡献率只有 70%,因此以 85%作为阈值选出前两个主成份 F_1, F_2 , 4 个主成份各自的方差贡献率为权重,根据式(1)计算出各属性的得分分别为 0.309, 0.243, 0.246, 0.025。在非核属性中,属性 P 的得分最高,因此选择属性 P 加入到核属性集合中,可以得到决策表的一个约简 $\{S, X, P\}$,可以求出 $M_A = M_{\{S, X, P\}}$ 。

结束语 当数据量很大时,应用粗糙集算法进行数据的分析处理时间复杂度较高,因此如何能在有限的时间内求出

了原数据,而 LPP 考虑了人脸的局部信息,人脸空间更可能存在于非线性子空间中,LPP 在一定程度上能解决这类问题。GELG 不仅从全局和局部角度出发,而且也对数据的类别信息进行了综合,因此识别效果较理想。

结束语 本文提出了一种基于局部特性和全局特性的新的特征提取算法,既考虑了人脸的全局信息又兼顾了局部信息;构造出了更加准确、鲁棒的识别系统,并且考虑了类别信息,因而具有良好的识别性能。实验结果表明,算法可以有效提取出特征,并使不同类尽可能分离,因而可以提高识别率。

参考文献

- [1] Turk M, Pentland A. Eigenfaces for recognition[J]. Cognitive Neurosci, 1991, 3(1): 71-86
- [2] Belhumeur P N, Kriegman D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19: 711-720
- [3] Tenenbaum J B, de Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290: 2319-2323
- [4] Rowies S, Saul L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290: 2323-2326
- [5] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396
- [6] He X, Niyogi P. Locality preserving projections[M]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2004
- [7] Zhao Haitao. Combining labeled and unlabeled data with graph embedding[J]. Neurocomputing, 2006, 69: 2385-2389

最佳约简一直是科研工作者的一个研究重点。启发式方法是近年来研究的热点,它将粗糙集方法和现有的一些比较成熟的、效率较高的统计分析方法相结合,从而获得属性约简和相应的规则。

在本文提出的基于相容关系和主成份分析方法相结合的粗糙集分析方法中,如何进一步降低求解相容矩阵的时间复杂度是下一步要继续研究的问题。

参考文献

- [1] Wong S K, Ziarko W. On optimal decision rules in decision tables[J]. Bulletin of Polish Academy of Science, 1985, 33: 693-696
- [2] 黄兵, 周献中, 张蓉蓉. 基于信息量的不完备信息系统属性约简[J]. 系统工程理论与实践, 2005(4): 55-60
- [3] Kryszkiewicz M. Rough set approach to incomplete information system[J]. Information Sciences, 1998(112): 39-49
- [4] 王国胤. Rough 集理论在不完备信息系统中的扩充[J]. 计算机研究与发展, 2002, 39(10): 1238-1243
- [5] 张腾飞, 王锡准, 肖建梅. 不完备信息系统的一种属性相对约简算法[J]. 计算机工程, 2007(5): 184-186
- [6] 闫德勤. 概率差别矩阵与不完备信息系统属性约简[J]. 计算机科学, 2005(8): 164-166
- [7] 付昂, 王国胤, 胡军. 基于信息熵的不完备信息系统属性约简算法[J]. 重庆邮电大学学报: 自然科学版, 2008, 32(5): 586-592