

基于主曲线的指纹细节特征提取方法^{*})

唐庆适 苗夺谦 张红云

(同济大学计算机科学与工程系 上海200092)

摘 要 要提高指纹识别的识别率,关键是指纹特征的提取。主曲线是主成份分析的非线性推广,它是通过数据分布“中间”并满足“自相合”的光滑曲线,较好地反映了数据分布的结构特征。本文尝试使用主曲线这种新的方法来提取指纹的细节特征。实验结果表明利用主曲线来提取指纹的结构特征是可行的,它为指纹特征提取的研究提供了一条新途径。

关键词 主曲线,指纹,细节特征,特征提取

A Method Based on Principle Curves for Minutiae Extraction in Fingerprint Images

TANG Qing-Shi MIAO Duo-Qian ZHANG Hong-Yun

(Dept. of Computer Science and Engineering, Tongji University, Shanghai 200092)

Abstract The extraction of fingerprint feature is critical to improve the rate of fingerprint recognition. Principal curves are nonlinear generalizations of principal components analysis. They are smooth self-consistent curves that pass through the “middle” of the distribution and they preferably reflect the structural features of the data. In this paper, principal curves are used to extract the minutiae of fingerprint. From the result of experiment, it is proved that using principal curves for minutiae extraction is feasible.

Keywords Principal curve, Fingerprint, Minutiae, Minutiae extraction

1 引言

自动指纹身份鉴别系统(AFIS)的研究是20世纪60年代在一些发达国家率先开始的,其真正成熟则得益于80年代电脑的发展和光学扫描技术的革新。在网络和信息时代高速发展的现代社会中,出于安全、保密和保护合法性权限等原因,人们必须面对各种各样的身份验证,以指纹作为身份验证越来越受到社会的关注^[4]。在AFIS中,指纹特征的提取是实现指纹识别的关键环节。

指纹的常用的特征分为基于结构的特征和基于统计的特征^[10]。指纹的细节特征点(Minutiae)就是基于结构的特征,它主要指的是纹线端点(Ridge Ending)和纹线分叉点(Bifurcation)(定义如图1所示)^[3,9]。纹线端点指的是纹线突然结束的位置,纹线分叉点指的是纹线一分为二的位置。大量的统计结果表明,使用这两类特征点就足以描述指纹的唯一性。

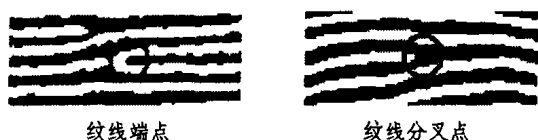


图1 指纹细节特征点定义

迄今为止,人们已经对指纹细节特征提取算法进行了很多的研究工作。目前大多数的指纹细节提取算法都是基于细化图的基础之上的。总的来说,细化是一种标准操作,目前已有许多细化算法,如Pavlidis算法,Deutsch算法和Hilsirh算法^[5~7],但这些采用数字形态学方法的常规模板构造方法(如

腐蚀和细化)缺乏几何直观性,也不符合人对指纹是按曲线理解而非像素点的特点。所以我们尝试用一种新的方法来提取指纹的细节特征点。

主曲线的概念是在1984年由Hastie和Stuetz提出^[1]。主曲线是通过数据分布“中间”并满足“自相合”的光滑曲线。其目的是根据给定的数据集求出一条曲线,使得这条曲线对给定数据集是某种意义下的对偶。形象地说,希望能寻找通过数据分布“中间”的曲线,使它能真实地反映数据的形态,即曲线是数据集合的“骨架”,数据集合是这个曲线的“云”。由此可见,主曲线对数据的分布结构信息保持性好。主曲线的理论基础是寻找嵌入高维空间的非欧氏低维流形,也是线性主成份的非线性推广。虽然在主曲线的原理中使用了较复杂的数学,但由于主曲线的这些性质和优点以及广泛的应用前景,在20世纪90年代后期已引起国外计算机科学家的关注。目前国际上关于主曲线在计算机方面的应用有:线性对撞机中电子束运行轨迹的控制、图像处理中辨识冰原轮廓、手写体数字的主曲线模板化等。

从主曲线定义可知,它与中轴线(与轮廓线等距离的穿过图形的光滑曲线)定义有明显的相似性。鉴于此,我们结合主曲线的思想提出了一种新的思路来提取指纹的特征点,即不通过细化的指纹图提取指纹的细节点,而是首先使用主曲线获取每根指纹线的骨架,然后提取指纹的细节特征点。

2 主曲线定义及相关算法

在这一部分,我们给出了主曲线的定义^[1,2]以及Balazs Kegl提出的推广的多边形主曲线算法^[2]。

^{*})本文得到国家自然科学基金项目(No. 60175016)和上海市教委“曙光计划”项目的资助。唐庆适 硕士研究生,研究方向:计算机软件与理论,主曲线理论,指纹识别。

2.1 主曲线定义

定义1 设随机向量 $Y=(Y_1, Y_2, \dots, Y_p)$ 的概率密度为 $g_y(y)$, 如果通过 Y 中间的一条曲线 $f(s)$ 满足:

$$f(s) = E(Y | s_f(y) = s)$$

则称 $f(s)$ 是的一条主曲线。其中 $s_f(y)$ 是数据点 y 投影到曲线 $f(s)$ 上 s 点的值, 即

$$s_f(y) = \sup\{s: \|y - f(s)\| = \inf_{\tau} \|y - f(\tau)\|\}.$$

定义2 光滑曲线 $f(t)$ 是主曲线当且仅当:

- (i) $f(t)$ 不自相交
- (ii) $f(t)$ 在任一有界的 R^d 子集中长度有限
- (iii) $f(t)$ 是自相合的, 即 $f(t) = E[X | t_f(X) = t]$.

由主曲线定义可知: 主曲线上每个点是所有投影至该点的数据点的条件均值, 它满足自相合性。主曲线的理论基础是寻找嵌入高维空间的非欧氏低维流形, 也是线性主成分的非线性推广, 它能真实地反映数据的形态。图2是一个简单的例子, 从该图中可发现主曲线有两个明显的优点: 一方面对数据的信息保持性好, 另一方面它较好地勾画出了原始信息的轮廓。

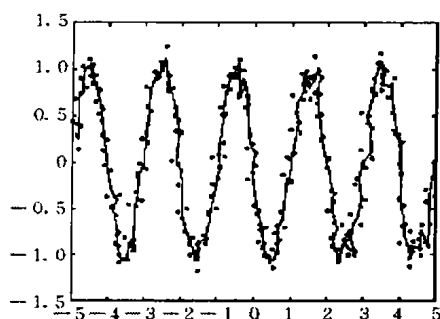


图2 主曲线

2.2 推广的多边形主曲线算法

推广的多边形(PL)主曲线算法是由 Balazs Kegl 提出的。该算法主要由以下几步组成:

① 初始化步: 首先采用传统的 Suzuki-Abe 细化算法获取原始图像模板的近似骨架图 G_{in} 。由两个集合 V, S 构成, 其中 $V = \{v_1, v_2, \dots, v_n\} \subset R^d$ 是顶点的集合, $S = \{(v_{i1}, v_{j1}), \dots, (v_{ik}, v_{jk})\} = \{s_{i1j1}, \dots, s_{ikjk}\}$ 是边的集合。

② 拟合—光滑步: 这一步的目的是调整骨架图 G_{in} 的光滑性, 使之更好地拟合图形。对给定数据集 $X_n = \{x_1, x_2, \dots, x_n\}$, 用距离惩罚函数 $E(G) = \Delta(G) + \lambda P(G)$ 取最小来优化骨架图, 其中 $\Delta(G) = \frac{1}{n} \sum_{i=1}^n \Delta(x_i, f)$ 是表示数据集中点到图形 G 中曲线的距离平方的平均值, $P(G) = \frac{1}{k+1} \sum_{i=1}^{k+1} P_v(v_i)$ 是关于 G 中曲线的平均曲率惩罚函数。 $\Delta(G)$ 值较小表示骨架图可以较好地拟合数据, $P(G)$ 值较小能保证骨架图的光滑性较好。在这一步中首先执行投影步, 在投影步中把数据集 X_n 划分到属于骨架图顶点和边的最近邻区; 其次做顶点优化步, 即调整骨架图 G 中顶点和边的位置, 使得距离惩罚函数 $E(G)$ 取得局部最小值。

③ 重构步: 这一步是对拟合—光滑步的补充完善。在重构步中利用骨架图的几何性质对顶点和边的结构进行修改, 消除或校正初始图形的瑕疵。例如删除短分支、删除小圈等。

推广的多边形(PL)主曲线程序流程图见图3所示。

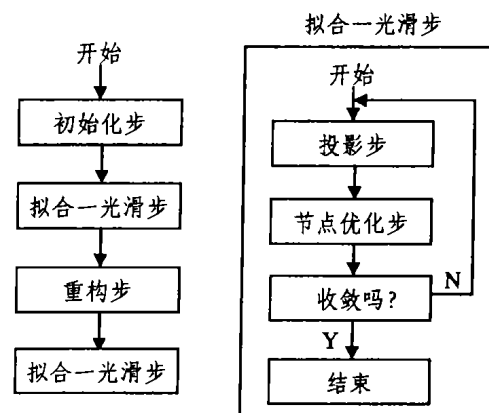


图3 推广的 PL 主曲线算法流程图

3 利用主曲线提取指纹细节特征点

在本节, 我们提出利用了推广的多边形主曲线算法来提取指纹细节特征点的方法。具体的思路如下:

首先, 首先我们利用一张已有的指纹图片(见图4)从中将每条指纹线抽取出来, 本文使用了两条指纹线作为实例进行演示(见图5)。



图4

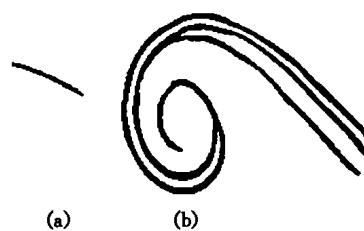


图5

然后利用在2.2节中所述的推广的 PL 主曲线算法得到指纹线的主曲线。我们首先对图5(a)的指纹线进行分析。指纹线图5(a)的主曲线如图6所示。



图6

从图6可以看出, 主曲线很好地拟合了指纹线。在图6中, 该指纹线是由一条主曲线组成的, 在这里我们用 $A \sim B$ (A, B 是主曲线的两个端点) 表示这条主曲线。本文在表1给出主曲线 $A \sim B$ 的数据。

表1

	数据点——用(x,y)坐标表示
A [^] B	(51.7608, 76.7328)
	(54.6401, 76.1603)
	(59.2785, 75.0582)
	
	(79.3904, 65.3305)
	(82.2814, 63.6042)
	(84.2326, 62.4981)

从表1可以看出,主曲线 A[^]B 是一个数据点集,而数据点集中的第一个点和最后一个点就是主曲线的两个端点。而本文正是通过对主曲线端点进行分析从而提取出指纹的细节特征点(算法思想在本节最后阐述)。接下来我们分析指纹线图5(b)的主曲线(见图7)。

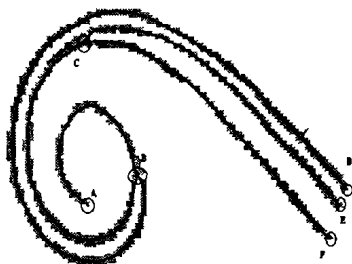


图7

通过图7可知,图5(b)中的指纹线的骨架一共由5条主曲线(A[^]B, B[^]C, B[^]D, C[^]E, C[^]F)组成,这五条主曲线分别是五个数据点集,每个数据点集的第一个和最后一个点就使曲线的两个端点。

基于以上的实验结构,我们在这一节的最后一部分提出提取细节特征点算法:首先遍历主曲线的数据点集的两个端点;如果端点只在一个数据集中出现,那么这个点是纹线端点;如果端点在三个数据集中出现,那么这个点是纹线分叉点。

通过上述算法,可以得到图6中的点 A 和 B 以及图7中的点 A 是纹线端点;图7中的 B、C 两点是纹线分叉点;而图7中的点 D、E、F 由于是指纹的边界,所以过滤它们,不作为细节

特征点提取。

结论 在本文,我们提出了利用主曲线来进行指纹细节特征点提取的方法,这是一种新的提取指纹特征点的手段。从实验的结果看,该方法对大部分的指纹线都取得了比较好的效果,当然,对一些比较复杂的指纹线,例如曲率比较大的指纹线以及指纹中心部分,还有一些不尽如人意的地方,这些地方需要我们在下一步的工作中进行改进。通过对所做实验进行分析,我们认为利用主曲线对指纹的细节特征点的提取在理论和实践上都是可行的。我们计划在今后的研究中完善算法以便取得更好的效果,同时我们还要将这种方法与传统的特征提取方法进行比较,分析它与传统方法在效率和性能上的优劣。除此之外,我们还计划把主曲线扩展到对指纹的宏观特征分析上,以及实现指纹的识别。

参考文献

- 1 Hastie T. Principal Curves and surfaces. Laboratory for Computational Statistics. Stanford University, Department of Statistics: [Technical Report 11]. 1984
- 2 Kegl B. Principal Curves: Learning, Design, and Applications: [Dissertation for Ph. D]. 1999
- 3 Lin Hong. Automatic personal identification using fingerprints [D]: [Dissertation for Ph. D.]. Michigan State University, 1998. 5~46
- 4 Hong L, Jain A. Integrating faces and fingerprints for personal identification. IEEE-PAMI, 1998, 20(12): 1295~1307
- 5 Hrechak A K, McHugh J A. Automated fingerprint recognition using structural matching. Pattern Recognition, 1990, 23(7): 893~904
- 6 Verma M R, Majumder A K, Chatterjee B. Edge Detection in fingerprint. Pattern Recognition, 1987, 20(5): 513~523
- 7 Malleswara Rao T Ch. Feature extraction for Fingerprint classification. Pattern Recognition, 1976, 8: 181~191
- 8 张军平,王珏. 主曲线综述[J]. 计算机学报, 2003, 26(2): 129~146
- 9 尹义龙,宁新宝,张晓梅. 改进的指纹细节特征提取算法. 中国图象图形学报, 2002, 7(12): 1302~1306
- 10 张雄,贺贵明. 基于宏观曲率的指纹特征提取和分类. 计算机研究与发展, 2003, 40(3)

(上接第174页)

评估值,由于其搜索空间大,因此收敛到最优解的概率较低。BHCM 算法通过尝试所有可能的取值来发现最优的聚类数目,由于直接使用 K-means 方法,因此没能发现最优评估值 0.44。aiNet 方法存在收敛到最优评估值的概率较低的问题(即偏离理想聚类中心集),尽管在 Iris 数据集上收敛到正确聚类数目的概率较高,但对于本文中的两个人工数据集,收敛到正确聚类数目的概率约为 0.5。这表明新算法不仅能够获得正确的聚类数目,且在同样的计算量(K-means 的计算次数)下收敛到最优值的概率最高,即收敛速度更快。

结论 文中给出了一个全新的动态聚类算法,它通过免疫网络算法快速地求出一个候选聚类中心集,然后使用 Tabu 搜索在候选聚类中心集上实现动态聚类。通过仿真试验证明了新算法具有较高的收敛概率,并极大地提高了动态聚类过程的收敛速度。

在大量实际的聚类应用中,如网络入侵检测,数据挖掘,由于事先很难获得理想的聚类簇数量,因此本算法将具有广

阔的应用前景。

参考文献

- 1 Karkkainen I, et al. Dynamic local search for clustering with unknown number of clusters[A]. IEEE 16th Intl. Conf. on Pattern Recognition[C], Quebec CANADA, 2002, 2: 240~243
- 2 Hall L O, et al. Clustering with a genetically optimized approach [J]. IEEE Trans. on Evolutionary Computation, 1999, 3(2): 103~112
- 3 Krowi R. Genetic algorithms for clustering: a preliminary investigation. IEEE In: Proc. of the Twenty-Fifth Hawaii Intl. Conf. on System Sciences, 1992, 4: 540~544
- 4 Xu Hong-Bing, et al. Fuzzy tabu search method for the clustering problem. In: IEEE In: Proc. of the first Intl. Conf. on Machine Learning and Cybernetics, 2002. 876~880
- 5 行小帅,等. 基于免疫规划的 k-means 聚类算法. 计算机学报, 2003, 26(5): 605~610
- 6 Timmis. Artificial immune system: an novel data analysis technique inspired by immune network theory. Wales university, 2001
- 7 de Castro L N. An Evolutionary Immune Network for Data Clustering[A]. In: Proc. of the IEEE SBRN, 2000. 84~89
- 8 Bezdek J C, Pal N R. Some new indexes of cluster validity. IEEE Transactions on Systems, Man and Cybernetics, Part B, 1998, 28(3): 301~315