

基于多级索引的指纹分类算法

王文涛¹, 尹义龙¹, 戴鸿君¹, 王文会²

(1. 山东大学计算机科学与技术学院, 济南 250101; 2. 中共河北省委党校, 石家庄 050061)

摘要: 为提高大规模网络化指纹识别系统的检索速度和准确率, 提出一种基于多级索引的指纹分类算法, 将质量特征分为质量高和质量差 2 类, 对于质量高的指纹利用指纹图像的 3 个特征——指纹类别、宏观曲率和平均周期建立三级索引, 以逐级缩小检索空间。实验结果表明, 该算法准确率高、检索速度快, 具有良好的实时性。

关键词: 指纹检索; 质量特征; 宏观曲率; 平均周期

Fingerprint Classification Algorithm Based on Multi-level Index

WANG Wen-tao¹, YIN Yi-long¹, DAI Hong-jun¹, WANG Wen-hui²

(1. College of Computer Science and Technology, Shandong University, Jinan 250101;

2. The CPC Hebei Provincial Committee Party School, Shijiazhuang 050061)

【Abstract】 In order to improve the index speed and the accuracy of the network-based fingerprint recognition system for the large population, a fingerprint classification algorithm based on multi-level index is proposed. Quality feature is used to classify fingerprints into two types: low quality and high quality. For the low quality fingerprints, three features, including ridge shape, macroscopic curve and average period are used to establish three-step index, which decreases the index space gradually. Experimental results show this algorithm has more accuracy, and improves the index speed effectively. It has better real-time performance.

【Key words】 fingerprint index; quality feature; macroscopic curvature; average period

1 概述

随着网络技术的快速发展, 指纹识别技术的网络化应用越来越广泛, 用户的规模也越来越大, 出现了面向上万人甚至更大规模的网络化指纹考勤系统等应用。这些系统中都面临着巨量指纹数据库的检索问题, 这对指纹分类算法提出更高要求。

国内外学者已经对指纹分类算法进行了大量的研究。传统的指纹分类算法沿用指纹学上的分类法, 把指纹分为 5 类: 拱形(arch), 尖拱形(tent arch), 左旋形(left loop), 右旋形(right loop)和漩涡形(whorl)^[1]。但是当数据量增大时, 传统的分类方法并不能有效地缩小待匹配指纹的数量, 而且对低质量指纹的处理效果不佳, 容易产生分类错误。

为解决上述问题, 文献[2]提出一种改进的基于指纹奇异点的指纹分类模型, 结合指纹方向场的对称性质, 设计一种寻找指纹中心区对称线的方法, 在此基础上, 根据奇异点和指纹对称线的位置关系, 提出一种允许类交叉的指纹分类算法, 把指纹分为 12 类。文献[3]利用 core 点和 delta 点的位置、数目和方向判断指纹的类别, 并结合局部特征建立三级索引, 将指纹分为 8 类, 这种分类策略仍然采用传统的分类思想, 可划分的类别数有限, 即使将指纹数据库分类, 每个子类的指纹数量仍然较大。多级分类取决于选取特征的稳定性和可区分性以及分类的级数, 是分类检索技术的发展趋势, 但是这些分类算法都没有考虑低质量指纹的分类问题, 而且仍然面临分类指纹类别数少的问题。

为解决大规模指纹数据库的索引问题, 本文提出基于多级索引的指纹分类算法。根据指纹图像的质量特征, 将指纹

分成质量差和质量高 2 类, 对于质量高的指纹, 利用指纹图像的 3 个特征——纹线距离、纹形类别和宏观曲率建立三级索引。实验结果证明, 该算法分类准确率高、速度快, 适用于大规模指纹数据库的分类检索。

2 方法总体描述

对于大型指纹数据库, 通过一对一对来检索指纹非常耗时。为提高检索速度和检索效率, 利用指纹图像的 4 个特征——质量指标、纹形类别、宏观曲率和平均周期, 建立四级索引。如果输入指纹被判定为质量较差, 则与数据库中所有质量差的指纹进行匹配, 直到匹配得分最高的被认定为相匹配指纹。如果输入指纹被判定为质量好, 则按照图 1 进行检索。每一级检索都通过指纹图像的一种特征对待匹配指纹进行过滤, 并最终获得最小的待匹配指纹集合。详细的检索过程如下:

(1) 使搜索空间为整个指纹数据库, 记为 F。

(2) 根据输入图像的质量指标, 如果质量差, 直接与包含所有标记为质量差的指纹集合 L 进行细节匹配; 否则, 读取数据库中所有质量高的指纹为一个集合, 记为 G。

(3) 根据输入图像的纹形类别, 在指纹集合 G 中检索出同

基金项目: 山东省高新技术自主创新工程专项基金资助项目(2007 ZCB01030); 山东省自然科学基金资助重点项目(Z2008G05)

作者简介: 王文涛(1980—), 男, 硕士研究生, 主研方向: 模式识别, 图像处理; 尹义龙, 教授、博士; 戴鸿君, 讲师、博士后; 王文会, 硕士

收稿日期: 2009-10-10 **E-mail:** wangwt@126.com

类别的指纹，得到记录集 H。

(4) 计算指纹图像的宏观曲率，根据宏观曲率在 H 中检索出符合条件的指纹记录集 C。

(5) 获取输入图像的平均周期，根据平均周期在记录 C 中检索指纹记录，得到记录集 P。

(6) 进行细节匹配。

(7) 平均周期的计算存在一定误差，对于质量较高的指纹，如果在当前类别没有找到匹配指纹适当增大搜索半径，则继续检索数据库，进行细节匹配，直到找到匹配指纹或所有指纹匹配结束。

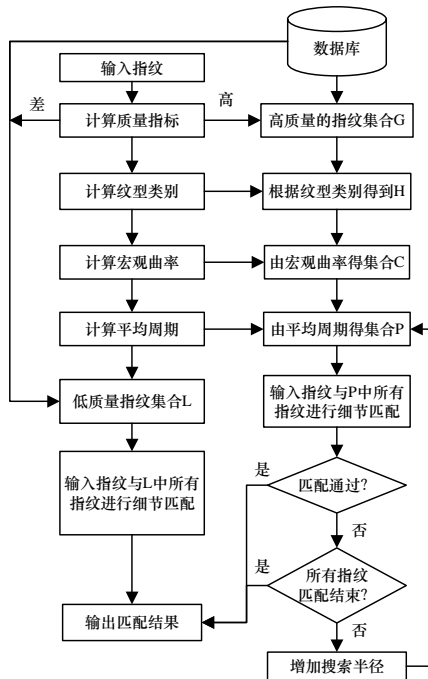


图 1 检索流程

随着检索级数的增加，计算复杂度也随之增加，但待匹配集合越来越小，对于大规模指纹数据库可以明显地提高检索速度。但是对于数量不多的小型指纹数据库，因为待匹配样本较少而计算复杂度太高，所以多级检索策略反而不能充分地发挥算法自身的优势。

3 特征的定义与提取

为适应本文提出的多级索引算法，本文分别选取指纹的质量特征、指纹类别、宏观曲率和平均周期作为用于每一级数据库检索的特征，并对每种特征进行面向本系统的适用性研究。

这几个特征都是指纹图像的固有特征，具有极高的稳定性，而且特征之间相互独立，可区分性很强，因此，能够适用于本文提出的基于多级索引的指纹分类算法。

3.1 质量特征

质量指标被用来评价指纹图像的质量，不同的指标反映了指纹图像在不同方面的质量好坏情况。本文采用空域梯度的方法判断指纹的质量^[4]。

实验针对不同质量的指纹图像使用该指标进行分类，质量得分分布如图 2 所示。可以看出，在梯度指标值 1.5~2.5 之间，不同质量的指纹图像有一定的重合。由于质量差的指纹识别较困难，是造成指纹识别错误的主要原因，因此本文首先把梯度标值小于 1 的指纹图像标记为质量差，对于这些指纹，不再进行纹形、宏观曲率和平均周期的计算和分类。

其中，“o”表示质量较好指纹图像的得分；“*”表示质量较差指纹图像的得分。

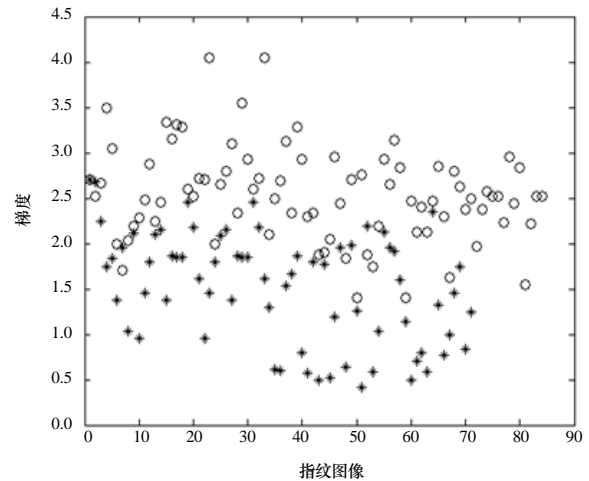


图 2 将梯度作为质量指标的得分分布

3.2 指纹类别

由于本文主要讨论实时指纹识别系统中的指纹检索问题，通过活体实时采集的指纹图像经常出现丢失 delta 点的现象，因此采用文献[5]基于指纹方向图的分类方法。其主要算法步骤如下：(1)对指纹图像进行分块，并对每一块指纹图像进行傅立叶变换。(2)获得指纹图像的方向图。(3)寻找 core 点并确定指纹图像的特征向量。(4)使用 k 近邻分类器和最小距离分类器来对指纹进行分类，将指纹分为上述 5 类。

3.3 宏观曲率

指纹极限的弯曲特性是描述指纹的重要信息，但是由于单条指纹曲线的布连续、分叉和不规则弯曲，因此提取单独一条指纹的曲率信息对于描述指纹的结构特征意义不大。通过观察发现，相邻几条指纹曲线组成的弯曲带的曲率变化对单段指纹曲线的曲率突变不敏感，同时，不同的指纹带一起反映了整个指纹的宏观弯曲性质^[6]。宏观曲率的提取算法描述如下：

(1) 为了减少处理对象的大小，提高算的速度，并将处理限制在一有效范围之内，保证算法的有效性，首先选择感兴趣区域。

(2) 脊跟踪，选出由几条相邻脊曲线组成的一个指纹带。

(3) 根据若干规则，对邻域中多个点进行合并。

(4) 用一条曲线拟合上面搜索到的多条脊线中的点，然后提取这条曲线的特征。

基于上面算法提取出来的特征，对指纹图像进行分类。根据这些特征的空间分布和指纹中心点的方向的关系，将指纹图像分为 3 类，如图 3 所示。

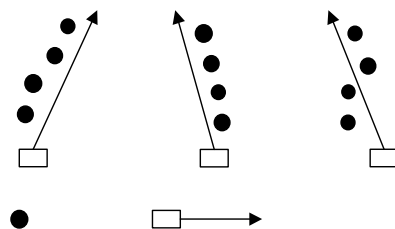


图 3 3类指纹图像

其中，第 1 类：特征点全部分布在中心点方向延长线的左边；第 2 类：特征点全部分布在中心点方向延长线的右边；

第3类:特征点部分分布在中心点方向延长线的左边,部分分布在中心点延长线的右边。

3.4 平均周期

指纹图像是由纹脊(或称为纹路)和纹谷组成的图像,平均周期是指纹的重要特征,通过计算指纹的平均周期并按照平均周期将数据库中的指纹进行排序,可有效地提高指纹的搜索速度^[7]。

平均周期的计算首先将 I 划分为不相交的大小为 $\omega \times \omega(50 \times 50)$ 的矩形区域,计算每一区域 W 的局部周期信息,然后根据局部周期信息计算平均周期信息。计算局部周期信息的基本方法是将各局部区域在一条穿过该区域中心并各点切线方向都和纹路走向垂直的曲线上投影,根据投影信号得到每个局部区域的周期。

假设数据库中指纹纹路的最小间距为 g_{\min} , 最大间距为 g_{\max} , 则将平均周期放大 δ 倍, 有 $\delta g_{\min} \leq T \leq \delta g_{\max}$ (T 为平均周期)。将指纹数据库按 T 分类, 每个 T 值作为一类, 则数据库中最多有 $\delta(g_{\max} - g_{\min}) + 1$ 类。给定常数 γ (正整数), 称之为搜索半径。 I 为输入指纹。在数据库中搜索与 I 相同的指纹的方法可描述如下:

- (1) 计算出 I 的平均周期 T ;
- (2) 在 T 类中查找指纹与 I 比对, 比对成功则转(7);
- (3) 将 i 赋值 1 到 γ , 分别做(4)和(5);
- (4) 在 $T+i$ 类中查找指纹与 I 比对, 比对成功则转(7);
- (5) 在 $T-i$ 类中查找指纹与 I 比对, 比对成功则转(7);
- (6) 拒绝 I , 结束;
- (7) 接受 I , 结束。

使用平均周期对指纹进行分类, 对于任何指纹都能计算出平均周期, 不会被拒绝分类, 而且类与类之间的距离可度量, 因此, 可以使用平均周期之差作为距离。而在数据库中搜索与输入指纹匹配的指纹时, 可以从与输入指纹距离最小的类开始进行搜索。因此, 系统中把平均周期放在多级索引的最后一级。

低质量的指纹严重影响了指纹识别的准确率和处理速度, 是指纹识别的难点, 也是自动指纹识别系统中难点。所以, 本文提出的基于多级索引的指纹分类索引方法, 首先通过质量指标把低质量的指纹分离出来, 作为一个类。在多级系统中, 错误是具有级联性质的, 对于质量高的指纹, 由于指纹类别和宏观曲率具有非常强的稳定性和一致性, 因此在第1级和第2级, 采用这2个特征进行逐级索引。平均周期的稳定性相对较差, 把平均周期的检索放在最后一级, 既保证了准确性, 又能保证分类检索的速度。

4 实验结果

实验中使用了指纹库中的 320 组同源指纹, 每组 10 枚, 共 3 200 枚指纹。如果同源指纹被分在同一类别, 则认为分类正确, 否则认为分类错误。测试结果表明, 其中有 7 组质量非常差, 使用质量评价算法全部判断正确。在一级纹形分类中, 由于采集不完全造成的无中心点的残缺的 32 枚指纹被分到其他类中, 如表 1 所示, 纹形分类错误率为 1.00%; 二级宏观曲率分类中错误分类的指纹共 30 枚, 如表 2 所示, 基本上都是受无中心点的影响, 所以, 分类错误率为 0.93%; 三级按照平均周期分类采用逐渐增大搜索半径的策略, 主要

用于进一步提高检索效率。由于第3级的输入指纹集合是第1级和第2级检索的结果, 如果前2级分类错误, 已经无法找到匹配指纹。实验结果表明, 这样的分类性能完全满足了大型指纹数据库分类算法的要求。

表1 测试样本的指纹类别分类结果

输出结果	待测样本类别				
	左旋	右旋	旋涡	拱形	尖拱
左旋	765	0	3	1	2
右旋	0	661	2	2	3
旋涡	1	2	580	0	0
拱形	3	4	2	882	0
尖拱	1	3	3	0	274

表2 测试样本的宏观曲率分类结果

输出结果	待测样本类别		
	第1类	第2类	第3类
第1类	949	4	2
第2类	8	1 229	6
第3类	3	7	992

输入不同类型的指纹时, 每类指纹在指纹库中所占比例不同, 因此, 具有不同的加速比。由于使用了多级索引, 而且没有类别交叉, 低质量的指纹是一类, 第1级按照指纹类别将指纹库分为5类, 第2级在第1级的基础上按照宏观曲率分为3类, 第3级在第1、第2级的基础上, 再按照平均周期分为10类~15类。假设输入查询指纹的类别概率分布与测试库指纹类别概率分布相同, 相对于不分类的指纹识别系统, 平均加速比介于151~226之间; 相对于传统的5类分法, 本文的平均分类加速比介于30~45之间。而一般的加速比都不超过10, 如文献[7]在533组2 663枚指纹库上进行测试, 其平均加速比为1.92, 所以, 本文提出的多级索引算法具有一定优势。

5 结束语

本文针对大型指纹数据库的检索效率问题, 提出基于多级索引指纹分类算法, 通过实验证明了其有效性和稳定性。准确计算所选指纹特征和提取新的特征并增加索引级数是下一步工作。

参考文献

- [1] Leong Chung Ern. Fingerprint Classification Approaches: An Overview[C]//Proc. of the 6th International Symposium on Signal Processing and Its Applications. [S. l.]: IEEE Press, 2001.
- [2] 马惠敏, 苏光大, 黄如克. 一种用于大库匹配的指纹分类算法[C]//第12届全国神经网络计算学术会议论文集. 青岛, 中国: [出版者不详], 2003.
- [3] 吴爱燕, 魏世泽. 一种基于网络的大型指纹数据检索方法[J]. 计算机科学, 2005, 32(8): 196-199.
- [4] Qi Jin, Shi Zhongchao, Zhao Xuying, et al. Measuring Fingerprint Image Quality Using Gradient[C]//Proc. of the SPIE'05. [S. l.]: IEEE Press, 2005.
- [5] 元秀燕, 尹义龙, 骆功庆, 等. 基于频谱能量的指纹分类[J]. 计算机工程与设计, 2008, 29(8): 2010-2013.
- [6] 张雄, 贺贵明. 基于宏观曲率的指纹特征提取和分类[J]. 计算机研究与发展, 2003, 40(3): 453-458.
- [7] 祝恩, 殷建平, 张国敏. 指纹纹路间距的计算[J]. 微电子与计算机, 2004, 21(10): 7-9.

编辑 陈文