

# 基于Delaunay三角剖分和K-means聚类的指纹检索模型

王干 王思坤 王彦力  
成都理工大学

[摘要] 本文从指纹检索的切入点出发, 结合统计分析方法提取指纹细节点的相关指标, 根据现有细节点数据建立指纹检索模型, 最后结合数据分析及检索结果计算检索穿透率, 验证基于Delaunay三角剖分和K-means聚类的指纹检索模型的准确性和有效性。

[关键词] 指纹识别; 检索穿透率; Delaunay三角剖分; K-means聚类

【DOI】 10.12252/j.issn.2096-627X.2021.12.1171

## 一、引言

近些年, 生物识别技术发展相当迅速, 该技术的特点是利用人类生物特征进行身份认证。其优势十分明显, 由于人类的生物特征通常具有唯一性、可测量性、终身不变性等特点, 可以使得认证准确率很高。常见的生物特征种类繁多, 如指纹、虹膜、人脸、静脉、掌纹、耳廓、DNA、笔迹和声音等<sup>[1]</sup>。因此目前常见的生物识别技术主要用于识别以上生物特征。在生物特征识别领域, 指纹作为最具独特性与持久性的生物特征之一, 被广泛应用于身份识别。

## 二、方法介绍

### (一) Delaunay三角剖分

假设是有限点集, 为由点集中的点作为端点构成的封闭线段, 为的集合。点集的三角剖分就是一张图<sup>[2]</sup>, 该图满足条件:

1. 没有相交边

2. 图中所有的面都是三角面, 且所有三角面的合集是点集的凸包。所谓凸包, 就是包含点集的最小凸多边形, 满足点集的所有点均在的边界上或的内部, 用表示点集的凸包。

3. 假设中的一条边, 若满足存在一个圆经过的两个端点, 圆内不含点集中任何其他的点, 这一特性又称为空圆特性。如果点集的三角剖分的边集的任意元素都是Delaunay边, 则称该三角剖分为Delaunay三角剖分<sup>[3]</sup>。

### (二) K-means聚类算法

均值聚类是基于样本集合划分的聚类算法。均值聚类将样本集合划分为个子集, 构成个类, 将个样本分到个类中, 每个样本到其所属类的中心的距离最小, 每个样本只能属于一个类, 所以均值聚类是硬聚类<sup>[4]</sup>。

## 三、模型建立

首先, 文章中将采用Delaunay三角剖分算法选取三角形。其次, 提取每个三角形的面积特征、平均角度特征, 并结合K-means聚类方法, 对指纹数据进行分类。最后, 对每个类进

行如下操作, 依次是计算各样本到聚类中心的距离; 再根据距离最大原则对指纹数据集进行过滤, 最终保留检索结果。

## 四、指标提取

对于相同手指指纹数据, 构建Delaunay三角剖分, 剖分的每个三角形的三个顶点到重心点的距离之和作为三角形的颜色编码, 结果图1所示:

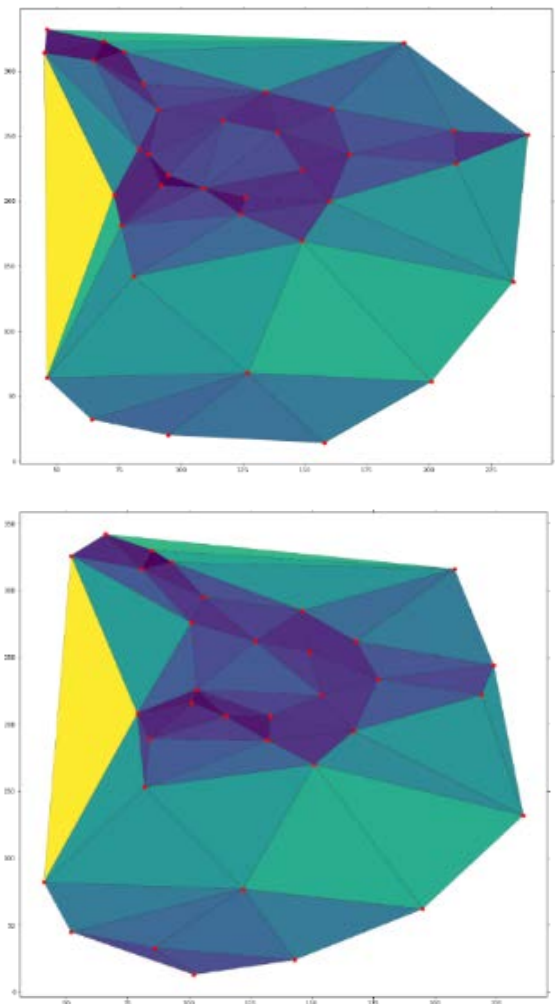


图 2 10024\_0和10024\_1

结果显示: 点越密集的地方, 图片上的颜色越深。同时, 基于上述Delaunay三角剖分能够看到, 对于相同手指之间, 其Delaunay三角剖分在直观上和整体上具有相似性, 对于局部的三角形, 二者也具有一定的位置统一性, 且颜色的深浅度基本

保持一致。

基于此,通过Delaunay三角剖分来提取每一张手指数据的图像信息和特征。主要包括如下三个方面:基于位置坐标的特征提取、基于三角形面积的特征提取和基于角度的特征提取。

首先是基于坐标的特征提取。通过计算重心坐标进行特征提取,部分结果如表1所示。

表1 重心坐标值

序号	x_bar	y_bar	z_bar
10024_0	120.97	186.93	178.47
10024_1	156.45	245.00	157.91

然后是基于三角形面积的特征提取。通过计算所有三角形的面积,基于所有三角形的面积来构造特征统计量,提取的特征有数量、最小值、最大值、极差、平均值、方差、标准差、第一四分位数、中位数、第三四分位数、四分差、变异系数、偏度、峰度、总面积。

最后是基于角度的特征提取。将手指的每一张细节点角度进行区间的整合。将0到359依次八个等度数的区间,依次为[0,45),[45,90),[90,135),[135,180),[180,225),[225,270),[270,315),[315,360),计算每一个区间的平均度数,此时得到8个关于度数的统计量,如果某一个区间类不存在,则填充为0。

综合上述结果,即对每个指纹数据总共提取了26个特征。

## 五、模型分析

基于11000个样本特征进行聚类分析。通过采用不同的聚类方法和确定不同的聚类个数,得到最终的聚类模型。根据聚类的结果计算每一类中的所有样本到聚类中心的距离进行升序排序,根据最终的过滤比例进行数据过滤,剩余的指纹样本即保存下来的样本。

最终将样本聚为10类,统计每一类的样本个数,结果如表2所示:

表2 聚类之后的样本量分布情况统计

聚类结果	1	2	3	4	5	6	7	8	9	10	总和
样本量	994	1278	1220	519	1494	703	982	330	1463	2017	11000

通过计算样本到其聚类中心的距离然后进行排序能够有效降低时间复杂度和空间复杂度,加快筛选速度,一定程度上提高精确率。对应第1类的994个样本计算到聚类中心的距离,然

后对距离结果进行降序排序,根据给定的比例剔除距离最大的这一部分样本,最终保留下来的样本作为下一步精确匹配的结果,在此基础上,结合指纹检索模型并针对不同的过滤比例,得到的穿透率如表3所示:

表3不同模型穿透率

过滤比例	80%	90%	95%	97%
穿透率	0.5320	0.4040	0.3500	0.3320

从上表可知,不同的过滤水平下,目标指纹高度相似的指纹之间存在重合,一定程度上验证了基于三角剖分算法(Delaunay)和K-means聚类算法的指纹检索模型的准确性。但是也能够看到,对于不同的待匹配指纹,重合的情况不尽相同,说明模型的进度有待提高。

## 六、结语

基于三角剖分算法(Delaunay)和K-means聚类算法的指纹检索模型的精度和穿透度整体来说并不高,还有很大的提升空间。一方面,导致穿透度较低的主要原因可能和特征提取的过程有直接关系,可能需要使用其他方法进行特征提取。另外一方面,对于指纹数据的降噪和指纹数据位置的修正需要进一步修正,这也会直接导致模型的穿透度较低,匹配效果价差。

## 参考文献:

- [1] 汤正刚.自动指纹识别系统中的关键技术研究[D].杭州电子科技大学,2019.
- [2] 李婷.基于辅助数据的指纹对齐及指纹检索中的三角形提取方法[D].北京邮电大学,2009.
- [3] 杨敏.指纹的匹配方法研究[D].大连理工大学,2004.
- [4] 徐莎莎.基于集成学习初始化的k均值聚类算法研究[D].华南农业大学,2019.
- [5] 朱韵.隐空间K均值聚类[D].上海交通大学,2018.

**作者简介:**王千(1998—),男,安徽蚌埠人,硕士在读,职称:无,研究方向:大数据技术及应用。

王思坤(1996—),男,四川广元人,硕士在读,职称:无,研究方向:大数据技术及应用 成都理工大学。

王彦力(1995—),男,四川巴中人,硕士在读,职称:无,研究方向:大数据技术及应用。