

基于高斯混合云变换的绝缘子图像聚类分析

杨耀宗¹ 王江贞² 王栋² 王志伟²

1 中国铁路兰州局集团有限公司机务部, 甘肃兰州, 730030;

2 中国铁路兰州局集团有限公司兰州西机务段, 甘肃兰州, 730070;

摘要: 兰新线途径三大沙漠和四大风区, 部分区段处于八级强风以上的天气, 被带动的沙尘持续冲击绝缘子伞裙表面造成冲蚀磨损, 使憎水性图像的直方图峰值过多而严重影响分割迭代效率与准确性, 为解决该问题, 本文设计了一种改进的高斯混合模型云变换算法。将K-means++应用到图像信息初始粒度的选择和细化过程中, 为高斯混合模型提供优化后的先验初值。选用两组损伤程度不同的绝缘子为实验样本, 分别采用K-means、高斯混合模型云变换(A-GCT)、及改进的A-GCT进行仿真对比, 结果表明, 在分割该类恶劣环境中划痕较多的绝缘子图像时, 本文算法的迭代时间比经典A-GCT减少一倍左右, 且误分率和水珠误分个数均有所降低。

关键词: 高斯云变换; 图像分割; 复合绝缘子; K-means++

DOI:10.69979/3041-0673.25.04.047

引言

从兰州到新疆乌鲁木齐全长约1776km的兰新线, 途经各种如高温高寒、强风沙尘等恶劣环境。兰新线穿过的区域风期逾200多天, 且新疆区段最大风速达37.6m/s, 环境中沙尘浓度很大^[1]; 列车顶部复合绝缘子常年处于此类环境中, 有研究者发现在新疆区段发生的绝缘子闪络事故中, 强风沙尘导致的占50%^[2]。车顶绝缘子闪络会造成列车的停运事故, 对铁路运营带来安全隐患和经济损失^[3, 4], 因此定期检测、评价绝缘子憎水性等级意义重大。

基于图像的等级判定方法逐步取代了传统的喷水分级、表面张力等方法, 不同的图像处理算法及环境会得到不一样的结论。董政呈等^[5]设计了一个憎水性检测装置, 为模拟现场环境将距离光源等条件“标准化”, 用基于边缘的图像分割方法提取水珠进行等级判定。闫康将绝缘子样本分别放置于小雨、中雨、大雨中得出憎水性图像, 与喷壶喷洒得到的图像进行仿真对比, 结果表明不同降雨条件下憎水性随时间变化的规律^[6]。上述相关文献大都模拟了一个“标准化”的实验环境采用不同的算法去判定憎水性等级, 绝缘子损伤程度远不及所处在西北地区风沙环境中的绝缘子, 算法差异性对迭代效率的变化影响不大。敬兴东用砂纸磨蚀伞裙模拟真实环境, 用FLUENT仿真伞裙表面气流特性, 通过测量接触角测定憎水性等级从而分析沙尘对憎水性的影响^[7]。

以上文献研究内容都集中在兰新线强风沙环境对绝缘子闪络特性及憎水性影响的因素上面, 测定等级的方法依旧是传统的喷水法结合数字图像处理, 但图像信息本身存在着复杂性和相关性, 传统分割方法只考虑了模糊性或随机性, 很少将二者结合去详细阐述图像分割

时的不确定性问题。云模型的提出对处理不确定性信息提供了很好的工具^[8]。姚红等人以云理论为基础分析了彩色图像的分割, 采用非均匀量化的方法对彩色信息进行量化, 结果表明该方法优于传统的K-means等方法^[9]; 以上文献基本阐述了云模型于其他新颖算法的结合, 说明云模型在图像分割领域的应用广泛, 同样忽略了图像初始信息粒度粗略问题。

K-means++是在K-means上发展而来的一种算法。在效率方面比K-means提高近20%, 算法迭代难度也减小了几个数量级。

综上所述, 针对兰新线强风沙环境下绝缘子图像信息的复杂度高、峰值过多、迭代耗时长的问题, 本文设计了一种改进的高斯混合云变换算法, 利用K-means细化图像峰值的初始粒度, 为高斯混合模型提供优化后的输入值, 从而提高了高斯混合模型算法的迭代效率, 改善了聚类效果。

1 云理论

1.1 云模型

云模型于1995年被李德毅院士提出, 兼顾了概念的随机性和模糊性, 为不确定性信息的处理提供了一个很好的工具。

云变换将任何一个频率分布 $f(x)$ 转换成多个不同幅值的高斯分布的叠加, 叠加的云越多, 误差越小, 表达式为:

$$f(x) \rightarrow \sum_{i=1}^n a_i \cdot C(Ex_i, En_i, He_i) \quad (1)$$

式中 $f(x)$ 是频率分布; a_i 是幅值系数; n 是云变换后生成概念的个数。

1.2 高斯混合模型云变换算法

高斯混合模型 (GMM) 将问题域中的概念转化成高斯分布或多个高斯分布的叠加。样本 x 由 GMM 描述的概率密度函数为:

$$p(x|\Theta) = \sum_{m=1}^M \omega_m N(x|\mu_m, \Sigma_m) \quad (2)$$

式中, $N(x|\mu_m, \Sigma_m)$ 是密度函数, ω_m 是权值, 且满足 $\sum_{m=1}^M \omega_m = 1$, μ_m 是均值, Σ_m 是协方差矩阵, 根据阈值和迭代次数, 用最大期望 (EM) 算法得出 GMM 的参数 Θ , 再通过概念含混度 “CD” 将 GMM 和云变换结合形成高斯混合模型云变换算法。

含混度 $CD=3He/En$, 概念划分的清晰程度可根据相邻分布之间的 CD 来表征; $CD=0$ 说明弱外围元素不交叠, 概念划分清晰; 否则含混度越大, 概念重叠程度越高。

通过高斯混合模型和云变换的结合, 解决了云变换只能处理一维图像数据的问题, 将图像数据分布转化为多层次粒度的概念^[10]。

高斯混合模型云变换包括启发式和自适应两种。H-GCT 根据经验值给定概念数量; A-GCT 无需人为指定概念数量, 根据数据对概念的贡献程度来计算概念数量, 输出结果为按一定规则生成的多个高斯分布。规则利用含混度 CD 与概念的交叠程度来度量, 如表 1 所示:

表 1 含混度和交叠程度的联系

含混度 CD	交叠程度
1	重合
(0.6354, 1)	骨干区交叠
(0.5004, 0.6354)	基本区交叠
(0.2, 0.5004)	外围区交叠
(0, 0.2)	弱外围区交叠
0	弱外围不交叠

由于 H-GCT 不能解决概念混乱、多粒度优化的问题, 因此本文直接讨论 A-GCT。利用含混度交叠程度指定迭

代策略, 根据表 1 设置对应的含混度阈值, 超过阈值则剔除, 按此策略来调用 H-GCT 来完成该算法。该算法有效解决了概念信息粒度的变化问题, 更符合现实情况下的信息聚类。

2 基于改进 A-GCT 的绝缘子图像分割

K-means++ 聚类算法先对数据的初粒度进行细化, 将细化后的值作为高斯云变换的初值进行迭代计算, 得出最优解^[11]。K-means 算法将数据根据相似度聚类到任意的数据中心, 通过 n 次迭代实现聚类的方法^[12]; 而 K-means++ 是对 K-means 随机选择初值中心的一种改进的聚类方法。

步骤 1: 输入样本 $\{x_i | i=1, 2, 3, \dots, N\}$, 含混度阈值 β 。

步骤 2: 统计样本集频率分布的波峰数量, 将其作为概念数量初值。

步骤 3: 利用 K-means++ 聚类方法对步骤 2 中的波峰数

值进行聚类分析, 得到 M 个细粒。

步骤 4: 利用高斯混合模型将以上粒度聚类成 M 个高斯云模型 $C(Ex_k, En_k, He_k)$, $k=1, 2, \dots, M$ 粒

步骤 5: 根据上式 (2)-(4) 来计算缩放比例 α_k 及高斯云数字特征, 并按照 CD 值将其排序

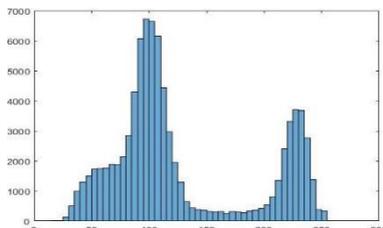
步骤 6: 按顺序对 M 个高斯云的含混度 CD 进行判断, 如果 $He_k/En_k > \beta$, 那么概念数量 $M = M - 1$

步骤 7: 如此循环步骤 4-6, 最终形成阈值内的 m 个高斯云模型, 按照 CD 对 m 个高斯云进行排序。

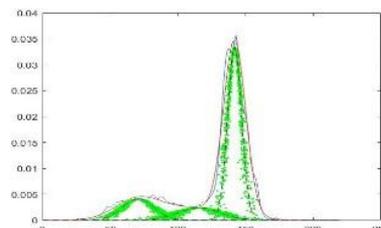
本文以兰局机务段两种不同损伤程度的实况绝缘子为例, 两组图像的直方图以及经典高斯混合模型云变换过程如图 1 所示



(a1) 憎水性较好绝缘子



(b1) 灰度直方图



(c1) A-GCT 分割结果

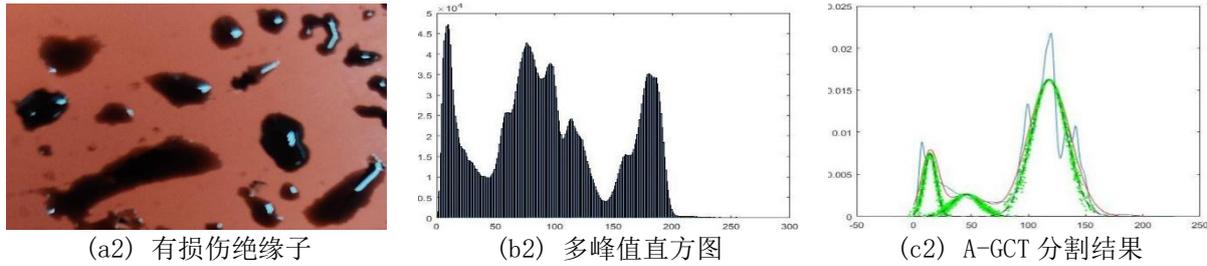


图 1 迭代过程比较

图 1 为两组图像迭代过程比较；b2 为损伤较严重绝缘子的直方图，从图中可明显看出峰值较多，图像复杂度与直方图直接相关，可通过灰度级出现情况和各灰度级上像素的个数定量描述，即峰值的数量。对上述两组图像用不同的方法进行仿真对比，如图 2 所示：

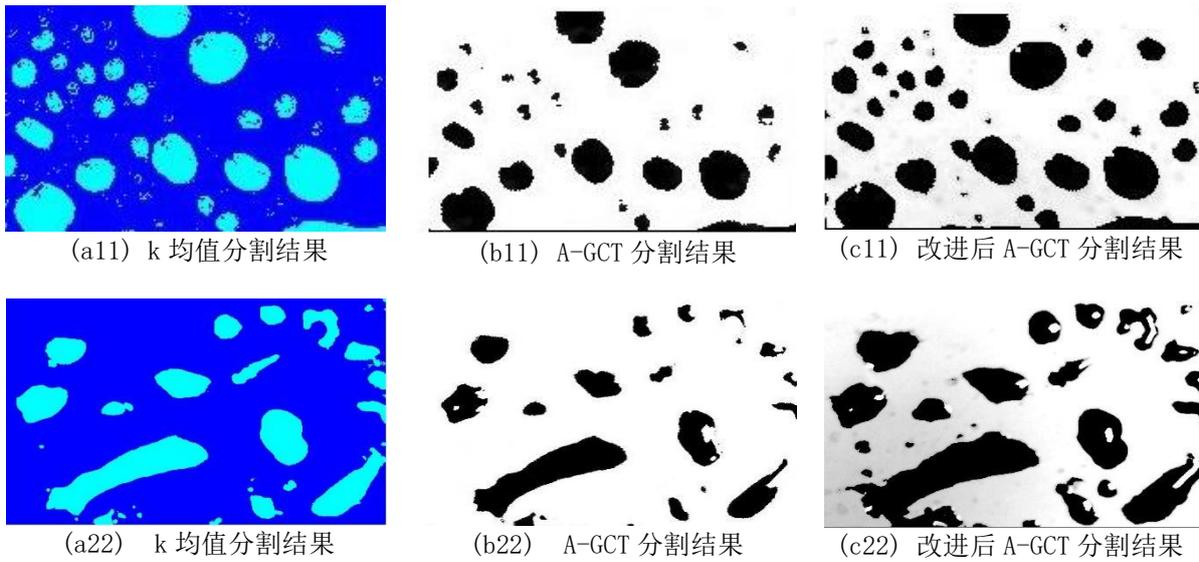


图 2 分割结果比较

针对两幅损伤程度不同绝缘子图像，分别利用三种算法对其进行仿真比较；如图 2 所示：K-means 分割后存在大量的水滴微粒，且边界的连续性弱，这是由于 K-means 在选择迭代中心时的随意性所致；通过引入 K-means++ 来优化像素初值，选择最优迭代中心，迭代效率更高且结果更加准确；即图 c11 相比于图 c22，迭代效率更高、边缘分割更加连续、结果更接近真实值。

除上述定性分析外，分别采用迭代时间、误分个数及误分率三个指标定量分析。误分率 ME 是指在图像分割过程中目标物体与背景颜色的误分概率，分析结果如表 2 所示：

表 2 定量分析

方法/组别	指标			
	误分个数	误分率	迭代时间	
组 1	K-means	4298	0.075	0.042
	A-GCT	874	0.027	0.351

组 2	改进的 A-GCT	615	0.019	0.287
	K-means	3025	0.088	0.058
	A-GCT	692	0.132	0.504
	改进的 A-GCT	526	0.064	0.194

分别采用三个指标对三种不同的方法进行定量分析：

组内：改进的 A-GCT 比 K-means 迭代时间多 0.24 秒左右，比 A-GCT 迭代时间缩短 0.7 秒左右；误分个数明显少于 k 均值，比经典 A-GCT 低 260 左右；误分率比 K-means 低近 0.05 且略高于经典 A-GCT；组 2 有相同的趋势，由于图像较复杂，误分率和误分个数都高于组 1。

组间：由于组 2 绝缘子损伤较严重，憎水性变差，水珠数量逐渐向水迹转化，因此误分个数少于组 1，误分率略高，且改进的 A-GCT 迭代时间比经典变换低 0.3 左右。

3 结论

列车顶部绝缘子受强风沙影响造成图像数据频率分布复杂、算法迭代效率低下。针对该问题,本文设计了一种改进的高斯混合云变换算法,将K-means++聚类应用到图像信息初始粒度的选择和细化中,通过优化GMM的输入提高算法的效率和效果,从而实现绝缘子图像的聚类分割。

取两种损伤程度不同的绝缘子,利用三种方法进行仿真比较:图2可明显看出K-means聚类方法分割后存在大量的小水滴,边界的连续性弱,改进的A-GCT优化了迭代中心的选择,迭代效率更高且结果更加准确;为定量描述分割效果,引入误分率ME进行定量分析,由表2可知当绝缘子损伤较大时,本文所设计算法迭代效率更高,且误分率比经典A-GCT低一倍左右。

参考文献

[1] 韩春刚,宋永顺.兰新高速铁路环境对动车组性能的影响分析及对策[J].铁路计算机应用,2014,(2).
[2] 司马文霞,程浩,杨庆.沙尘环境下绝缘子交流闪络特性及机理[J].高电压技术,2011,37(004):834-840.
[3] 孟健,梁德胜,张瑞刚.电网污闪事故的原因和预防措施[J].化工管理,2018,000(12):121.
[4] 吴广宁,石超群,张血琴.高速列车车顶绝缘子的研究进展[J].南方电网技术,2016,10(003):71-77.

[5] 董政呈,方彦军,王先培.基于嵌入式的复合绝缘子憎水性检测方法研究[J].绝缘材料,2015,48(07):67-72.
[6] 闫康.基于图像处理技术的复合绝缘子憎水性识别方法的研究[D].华北电力大学;华北电力大学(保定),2014.
[7] 许凯,秦昆,刘修国.高斯混合模型云变换算法及其在图像分割中的应用[J].武汉大学学报(信息科学版),2013,38(10):1163-1166+1183.
[8] 李德毅,杜鹃.不确定性人工智能.第2版[M].国防工业出版社,2014.
[9] 姚红,丁莎.基于云模型的彩色图像分割方法[J].电脑知识与技术,2014,10(22):5297-5300.
[10] 刘玉超.一种自适应的多粒度概念提取方法——高斯云变换[J].计算机工程与应用,2015,51(09):1-8.
[11] 刘旋,王国胤,罗小波.基于高斯云变换的遥感图像多粒度聚类[J].计算机科学,2017,44(09):23-27+52.
[12] 钟熙,孙祥娥.基于Kmeans++聚类的朴素贝叶斯集成方法研究[J].计算机科学,2019,46(S1):439-441+451.

作者简介:杨耀宗(1977-),男,汉族,甘肃兰州人,本科,高级工程师,研究方向:机电与动力工程。