

分带 K-均值聚类的平面标靶定位

陈西江^{1,2} 花向红^{1,2} 杨荣华³ 张青华⁴

(1 武汉大学测绘学院,武汉市珞喻路 129 号,430079)

(2 武汉大学灾害监测与防治研究中心,武汉市珞喻路 129 号,430079)

(3 重庆大学土木工程学院,重庆市沙正街 174 号,400030)

(4 成都市勘察测绘研究院,成都市一环路北三段 70 号,610081)

摘要:提出了一种分带 K-均值聚类的平面标靶定位方法。根据标靶与测站距离的限制条件,推导了较大噪声点的剔除公式,在整体最小二乘拟合平面的基础上增加了噪声点二次剔除的方法,对经过噪声点剔除的点云数据进行分带、聚类处理。同时,对每一带的聚类中心进行均值化处理,得到每一带的中心点,通过求取不同带中心的均值来确定标靶中心点。实验结果表明,分带 K-均值聚类的平面标靶定位模型较适合于平面标靶同名点确定。

关键词:点云数据分带;K-均值聚类;噪声点去除;粗差剔除;平面标靶定位

中图分类号:P207.1

标靶定位是点云数据处理中必不可少的环节,对于球体的定位主要是采用最小二乘法^[1-2]及文献[3]提出的稳健的标靶球定位方法,文献[4]将球体标靶作为点云配准的约束条件;鲁铁定等提出了基于总体最小二乘(TLS)的标靶球定位。目前,对于平面标靶定位算法的研究较少。Lichit 等根据标准平面标靶的反射特性,提出了三种标靶定位方法^[5],该方法在强返回信号的点云含有噪声的情况下无法获得精度较高的标靶中心点。基于此,本文对标靶点云数据进行噪声剔除,提出了一种分带 K-均值聚类算法。

1 分带 K-均值聚类算法

1.1 点云数据的分带

点云数据的分带即对圆形的平面标靶围绕初始中心点进行分带,根据公平聚类的原则,每一带的点云数相差不能太大,其分带过程如下:

1) 根据式(1)计算点云数据的初始中心:

$$[\bar{x}, \bar{y}, \bar{z}] = \frac{1}{n} \sum_{i=1}^n [x_i, y_i, z_i] \quad (1)$$

2) 以初始中心 $(\bar{x}, \bar{y}, \bar{z})$ 为原点,对圆形标靶

进行分带,为了使各带点云数据相当,需要各带包围面积相等,即需满足 $S_1 = S_2 = \dots = S_m$,假设各带的包围半径分别为 R_1, R_2, \dots, R_m ,则有式(2):

$$\pi R_1^2 = \pi(R_m + R_{m-1} + \dots + R_1)^2 - \pi(R_{m-1} + \dots + R_1)^2 \quad (2)$$

在第一带包围半径 R_1 已知的情况下,解得各带的包围半径:

$$\left. \begin{aligned} R_2 &= (\sqrt{2} - 1)R_1 \\ &\vdots \\ R_m &= (\sqrt{m} - \sqrt{m-1})R_1 \end{aligned} \right\} \quad (3)$$

式中, m 为分带数。

3) 根据式(3)将所有带的包围半径相加,如式(4),假设圆形标靶半径为 r ,根据式(4)可解得第一带包围半径 R_1 的限制条件,如式(5)所示:

$$R_1 + R_2 + \dots + R_m = \sqrt{m}R_1 \quad (4)$$

$$R_1 = r / \sqrt{m} \quad (5)$$

4) 利用前三步对图 1 点云数据进行分带,分带结果如图 2 所示。

1.2 K-均值聚类算法

K-均值聚类算法^[6]是用来处理数据聚类问题的算法,它是将含有 n 个数据点的集合 $X =$

收稿日期:2012-12-15。

项目来源:国家自然科学基金资助项目(41174010,41074025);精密工程与工业测量国家测绘地理信息局重点实验室开放基金资助项目(PF2011-17)。

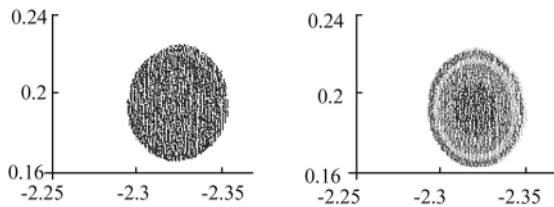


图1 分布图

Fig. 1 Distribution

图2 分带图

Fig. 2 Zoning

$\{x_1, x_2, \dots, x_n\}$ 划分为 k 个聚类中心 $C_j (j=1, 2, \dots, k)$, 根据样本数目或随机选取 k 个数据点作为 k 个聚类的中心, 计算每个数据点到初始聚类中心的距离, 即评价数据集中的其他对象与这些聚类中心的相似度来进行点集聚类的分配, 根据分配后的点集重新计算聚类中心。

2 噪声点的去除

2.1 基于距离限制的噪声点去除

标靶反射强度中心的数据容易受噪声污染。因此, 需对较大噪声点进行剔除, 由于平面标靶反射片的半径只有 3 cm, 旋转角很小, 由此造成的角度偏差可忽略不计, 根据测距精度可大致确定噪声点的位置, 剔除噪声, 其具体步骤如下。

1) 根据文献[7]提出的反射率式(6), 提取反射片中的点云数据:

$$P_R(\rho) = P_T \frac{\cos a}{\rho^2} \pi r \eta_{\text{sys}} \eta_{\text{atm}} \quad (6)$$

式中, a 为入射角; r 为表面反射率; η_{sys} 为系统损失; η_{atm} 为大气损失。

2) 计算反射片中点云数据集离坐标系原点的距离 l_i :

$$l_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \quad (7)$$

3) 对 d_i 进行缩小化处理。

4) 根据扫描仪的测距精度 $\pm \sigma$ mm/50 m, 确定由测距精度造成的最大测距误差 2δ , 如式(8), 并将 l_i 与 2δ 进行对比分析:

$$2\delta = 2l_i\sigma/50 \quad (8)$$

如果 $l_i > 2\delta$, 剔除 l_i 所对应的点集, 否则, 保留 l_i 所对应的点集。

2.2 基于总体最小二乘平面拟合粗差的剔除

1) 将文献[8]中二维正交 TLS 模型 $a(x - \bar{x}) + b(y - \bar{y}) = 0$ 延伸到三维正交 TLS 模型, 如式(9), 并利用特征值分解计算待定参数 a, b, c :

$$a(x - \bar{x}) + b(y - \bar{y}) + c(z - \bar{z}) = 0 \quad (9)$$

2) 根据式(10)计算点云数据集到拟合平面的距离:

$$d_i = \frac{|a(x_i - \bar{x}) + b(y_i - \bar{y}) + c(z_i - \bar{z})|}{\sqrt{a^2 + b^2 + c^2}} \quad (10)$$

3) 根据测距精度 $\pm \sigma$ mm/50 m, 计算点集到拟合平面距离 d_i 的限差 δ :

$$\delta \approx \sigma \sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}/50 \quad (11)$$

4) 判断 d_i 与 δ 的关系, 如果 $d_i > \delta$, 则 d_i 所对应的坐标点删除, 反之则保留。

5) 利用保留下来的点集重新进行平面拟合, 重复前 4 步, 直至所有点集满足 $d_i \leq \delta$ 。

3 基于分带 K-均值聚类的平面标靶定位算法及精度评定

算法具体实施过程如下。

1) 根据式(6)计算反射片区域的反射率, 提取反射片区域的点云数据。

2) 根据 § 2.1 方法剔除较大的噪声点。

3) 利用 § 2.2 方法剔除余下的噪声点。

4) 对剔除粗差后的点云数据集进行分带, 根据总的点云数据数确定最佳带数。

5) 对每一带的点云数据利用 K-均值进行聚类分析, 根据 § 1.2 确定最佳的类数 k 及每一类的聚类中心 C_j^k 。

6) 根据式(12)得到每一带的中心点

$$\bar{C}_j = \frac{1}{k} \sum_{i=1}^k C_j^i \quad (12)$$

7) 根据式(13)计算得到标靶中心点

$$\bar{C} = \frac{1}{m} \sum_{i=1}^m \bar{C}_j^i \quad (13)$$

为了评价标靶定位精度的高低, 需要对确定的标靶中心点进行配准, 其配准结果的精度采用点位中误差累积差及点位差两种评价体系, 具体如式(14)所示:

$$\begin{cases} m_1 = n \sqrt{\sum_{i=1}^n (\Delta x^2 + \Delta y^2 + \Delta z^2)/n} \\ m_2 = \sum_{i=1}^n (|\Delta x| + |\Delta y| + |\Delta z|) \end{cases} \quad (14)$$

式中, n 为坐标点数; $\Delta x, \Delta y, \Delta z$ 为配准坐标误差。

4 实验分析

为了验证本文方法的效果, 分别将 ILRIS-3D 扫描仪布设在 Z1、Z2 站, 并在其周围布设 4 个反射片, 根据式(6)进行反射率的计算, 并提取反射

片中的点云数据,如图 3 所示。

由图 3 可知,标靶中心出现了较大噪声点,采用基于距离限制方法对较大噪声点进行剔除,剔除之后的点云数据如图 4 所示。图 3 中共有 6 251 个坐标点,图 4 中共有 3 782 个坐标点。

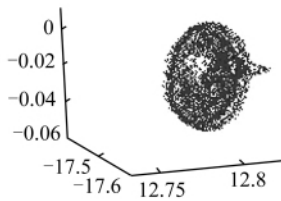


图 3 噪声点云
Fig. 3 Point Cloud with Noise

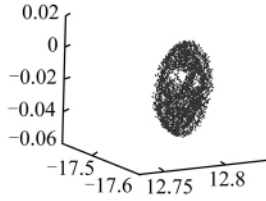


图 4 剔除噪声的点云
Fig. 4 Point Cloud Without Noise

根据图 4 中的点云数据,利用 § 2.2 方法进行粗差剔除的迭代,至第六次时待定参数及点集数不再改变,即最终用于分带聚类分析的点集数为 3 597 个,如图 5 所示。将剩余的点集分成 6 带,并对每一带进行 K-均值聚类分析,得出所有带的聚类中心,如图 6 所示。依据最佳聚类数及不同聚类数所对应的聚类中心,计算每一带聚类中心的均值,得到各带的中心点,结果见表 1。

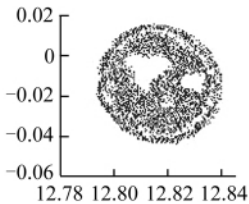


图 5 剔除噪声点
Fig. 5 Points Removing Noise

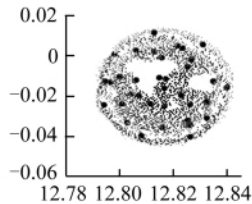


图 6 聚类中心
Fig. 6 Cluster Center

从图 6 可以看出,在数据缺失区域及其近邻区域包含有聚类中心值。因此,对各带不同聚类数的聚类中心进行均值化解各带中心点,可在一定程度上弥补数据缺失的影响,同时,为了确定标靶中心点,需要根据式(13)进一步对各带中心

表 1 各带中心点

Tab. 1 Center of Different Zoning

	<i>x</i>	<i>y</i>	<i>z</i>
第一带	12.808	-17.510 3	-0.024 87
第二带	12.809	-17.509 6	-0.017 52
⋮			⋮
第六带	12.809	-17.51	-0.015 68

点进行均值化处理,其所确定的标靶中心点如图 7 所示。由图 7 可知,由于受数据缺失的影响,初始中心点偏重于数据量比较多的中下区域,而本文方法所确定的中心点在初始中心点的上方,即所受数据缺失的影响较小。

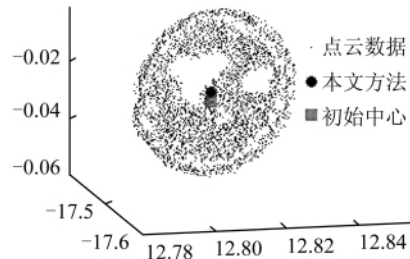


图 7 标靶中心点
Fig. 7 Target Center

同理,得到其余三个标靶的中心点,并进行配准,根据式(14)得到配准精度如表 2 所示。

表 2 配准精度

Tab. 2 Registration Accuracy

	本文方法	软件	模糊 C-均值聚类	K-均值聚类
m_1	0.116 4	0.124 1	0.126	0.123 6
m_2	0.069 3	0.085 35	0.083 6	0.080 4

由表 2 可知,在标靶含有噪声的情况下,本文方法的标靶定位精度较优于软件及传统方法。根据配准参数将 Z1 站点点云数据(图 8(a)所示)配准到 Z2 站点(图 8(b)所示),并与软件的配准结果对比分析,结果如图 8(c)、图 8(d)所示。由图可知,通过本文标靶定位方法所获得的配准参数对点云的配准效果与软件的配准效果精度相当。



(a) Z1 站点云



(b) Z2 站点云



(c) 本文配准结果



(d) 软件配准结果

图 8 配准结果

Fig. 8 Registration Results

参 考 文 献

- [1] Shakarji C M. Least-squares Fitting Algorithms of the NIST Algorithm Testing System[J]. Journal of Research of the National Institute of Standards and Technology, 1998, 103:633-641
- [2] Yuriy R, Milan H, Lars E S. Determination of the Optimal Diameter for Spherical Targets Used in 3D Laser Scanning[J]. Survey Review, 2005, 38(297): 243-253
- [3] 官云兰, 詹新武, 程效军, 等. 一种稳健的地面激光扫描标靶球定位方法[J]. 工程勘察, 2008, 10: 42-45
- [4] 赵煦, 周克勤, 闫利, 等. 基于激光点云的大型文物景观三维重建方法[J]. 武汉大学学报·信息科学版, 2008, 33(7): 684-687
- [5] Lichit D D, Stewart M P, Tsakiri M, et al. Benchmark Tests on a Three-dimensional Laser Scanning System [J]. Geometrics' Research Australasia, 2000, 72:1-23
- [6] Macqueen. Some Methods for Classification and Analysis of Multivariate Observations[C]. The 5th Berkeley Symp Math Statist, USA, 1967
- [7] Soudarissanane S, Lindenbergh R, Menenti M, et al. Scanning Geometry: Influencing Factor on the Quality of Terrestrial Laser Scanning Points[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2011, 66: 389-399
- [8] 张贤达. 矩阵分析与应用[M]. 北京: 清华大学出版社, 2004

第一作者简介: 陈西江, 博士生, 研究方向为三维激光的点云数据处理。

E-mail: cxj_0421@163.com

Planar Target Location Based on the Zoning K-means Clustering

CHEN Xijiang^{1,2} HUA Xianghong^{1,2} YANG Ronghua³ ZHANG Qinghua⁴

(1 School of Geodesy and Geomatics, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Hazard Monitoring and Prevention Research Center, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(3 School of Civil Engineering, Chongqing University, 174 Shazheng Street, Chongqing 400030, China)

(4 Chengdu Institute of Survey and Investigation, 70 the North Three Section of First Ring Road, Chengdu 610081, China)

Abstract: This paper presents planar target location methods based on zoning K-means clustering aiming at the impact of the noise points and missing data to planar target. The formula of elimination to larger noise points is derived based on the distance constraints between the target and the station. The second elimination of noise points is added to the paper on the basis of the Total Least-squares fit planar. The zoning and cluster of point cloud data is conducted after removing noise points, and the center of each zoning is determined by the mean processing of cluster center of each zoning. The target center is determined by calculating the mean of different zoning center. The experimental results demonstrate that the planar target location model based on the zoning K-means clustering is suited for the determined of the planar target conjugate.

Key words: point cloud data zoning; K-means clustering; removal of noise points; elimination of gross errors; planar target location

About the first author: CHEN Xijiang, Ph.D candidate, majors in point cloud data processing of 3D laser.

E-mail: cxj_0421@163.com