

# 基于 LIDAR 数据的城市数字表面模型生成技术

汪承义, 赵忠明

(中国科学院遥感应用研究所, 北京 100101)

**摘要:** 激光雷达(LIDAR)数据是一种新型数据源, 它产生的是高密度点云数据。为了更加方便地应用这些数据, 首先要生成数字表面模型(DSM)。采用传统的方法生成城市 DSM, 对城市区域复杂性的考虑不足, 也没有对数据存在的缺值情况进行相应处理, 故无法生成高质量的城市 DSM。该文阐述了一种新颖的生成高质量城市 DSM 的方法, 兼顾城市的复杂性和 LIDAR 传感器本身的特点。试验证明, 该方法生成的 DSM 与传统方法相比, 具备更好的效果。

**关键词:** 激光雷达; 数字表面模型; 三角网; 插值算法; EM 算法

## Digital Surface Model Generation From LIDAR Data in Urban Areas

WANG Cheng-yi, ZHAO Zhong-ming

(Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101)

**【Abstract】** LIDAR data are new data source, and they generate a high spatial resolution “point cloud”. To use it, the first step generally is to translate it to Digital Surface Model (DSM). The traditional methods are limited in terms of generation of DSM in urban areas because the urban environment is extremely complex with many man-made objects of different features and heights, and those methods do not consider conditions when data are missing. DSM generated using these traditional interpolation methods cannot achieve the purpose needed for urban applications, such as precision. This paper presents a new method to generate DSM considering characteristics of cities and traits of LIDAR sensors. The DSM generated using this method has better quality than that using transitional ones.

**【Key words】** Light Detection And Ranging (LIDAR); Digital Surface Model (DSM); TIN; interpolation algorithm; EM algorithm

航空激光扫描是一项自动化生成数字地形模型和数字表面模型的新技术<sup>[1]</sup>。激光雷达(LIDAR)已经成为提取空间高精度三维信息的有效手段<sup>[2-3]</sup>。利用该技术, 可以更快地生成高精度的DSM。DSM反映了区域表面的高低起伏<sup>[4]</sup>; DSM定义如下<sup>[5]</sup>: DSM=DTM+非地形要素。对于城市区域的DSM, 建筑物在其中的影响更大, 是需要重点考虑的特征。由于LIDAR原始数据为三维点数据, 因此生成DSM, 插值算法必不可少。此外, 处理LIDAR原始数据是一个关键的环节, 它处理的结果直接关系到数据应用效果。尤其是城市区域, 由于存在各种复杂的人工建筑物, 生成高质量的DSM就变得更加繁琐。本文通过LIDAR数据插值生成DSM将是规则网格DSM。一般的插值算法会产生某些非预期的现象<sup>[6]</sup>, 如距离反比算法会产生波纹状现象, 而Spline算法则可能导致断裂面情况。

### 1 数据描述

本文使用的数据集是原始的LIDAR点数据, 采集数据的系统为 LH System ALS40 激光雷达系统。数据参数的水平采样间隔为 2.0 m~2.2 m, 25° 视场角和 30% 的重叠率。该传感器的工作波段为近红外波段。此波段激光束照射到纯水中时, 传感器无法获取回波而导致信息的缺失。然而, 当水中的泥沙或叶绿素含量增加时, 回波随之增加<sup>[7]</sup>。由于诸如上述复杂情况的存在, 因此要生成高质量的DSM, 需要考虑传感器的特性。

### 2 算法

原始的LIDAR数据是一系列离散点, 为了便于数据处理, 需要生成规则的网格数据, 因此重采样是必备的过程, 它包

含两个方面: 位置的重采样和数值的重采样。前者确定网格尺寸, 而后者关系到插值算法<sup>[8]</sup>。图1为该方法的处理流程。

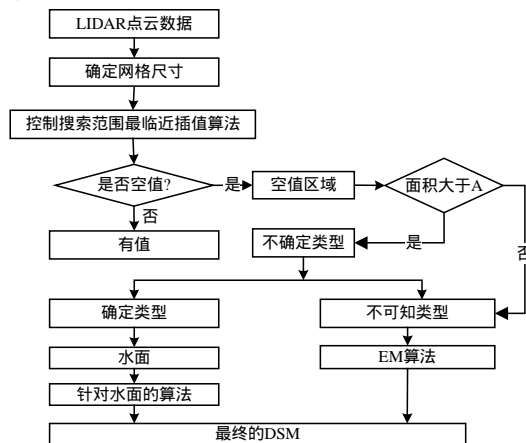


图1 算法流程

#### 2.1 网格尺寸确定

在进行重采样之前应首先确定网格尺寸。为了最大程度减少信息损失和避免信息冗余, 最好策略是每一网格包含一个并且只有一个点数据。网格中包含的数据多于一个会导致信息的损失, 反之, 网格过于密集会增加信息冗余。假设区域中存在  $n$  个数据点, 区域面积为  $A$ , 采用的网格尺寸<sup>[9]</sup>为

**基金项目:** 国家“863”计划基金资助项目(2003AA135010)

**作者简介:** 汪承义(1974-), 男, 博士研究生, 主研方向: 图像处理, 地理信息系统, 模式识别, 信号处理; 赵忠明, 研究员、博士生导师

**收稿日期:** 2007-02-20 **E-mail:** wangcycastle@163.com

$$1/\sqrt{n} \times A_0$$

## 2.2 插值算法

目前存在许多插值算法,对于某一种插值算法只能对一些特定的应用有效,事实上,没有一种插值算法适用于所有情况。因为插值算法是建立在对未知数值预测的基础之上的,而且该算法大多数是通过寻找临近点来建立相应模型进行预测未知点的数值,预测点离已知点越远,该点的贡献越小。

现今常用的插值算法有:距离反比算法,最临近插值算法,Spline 算法等。本文插值的目的是要反映城市区域剧烈的高程变化,而不是生成一个平滑的平面,所以选用最临近点作为基础的插值算法,因为此方法可以反映人工建筑物与周围环境的巨大差异。最终生成的 DSM 将不仅反映地形信息,而且包含地形上的建筑物和其他要素。

定义最临近点插值算法的搜索范围为网格尺寸的 2 倍,如果在搜索范围内存在数据点,则网格点数值为离网格最近的数据点,如在搜索范围不存在数据点,网格点数值为空。

图 2 显示该算法得到的结果。

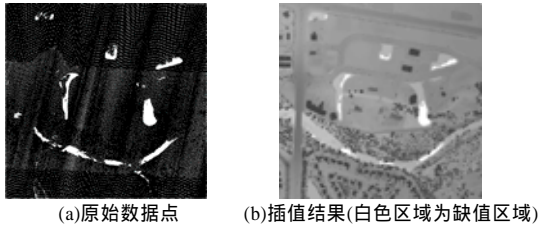


图 2 原始数据及初步处理结果

## 2.3 缺值区域处理

通过以上处理后,会得到一些缺值区域,要得到高质量的 DSM,对缺值区域的处理至关重要。缺值区域的存在都是由某些原因导致的,针对不同的缺值类型采用相应的处理方法才可以很好地解决缺值问题。本文缺值区域分为两种类型:已知缺值类型(当前只定义了水面类型)和未知缺值类型。针对已知类型采用与其相匹配的处理方法,而对于未知类型则使用 EM 算法。

### (1)水面缺值类型的处理

需要判断哪些缺值区域为水面类型,针对水面的特点,采用下述两个参数来定义。

$$P = \{A, \delta\} \quad (1)$$

其中,  $A$  为区域的面积;  $\delta$  的定义如下:

按顺时针方向连接边界点成一条曲线,并计算曲线上每一点的斜率,  $\delta$  是曲线上点斜率的标准差。

设点的高程分别为  $\{h_1, h_2, \dots, h_n\}$ , 每一点的斜率为  $g_i = (h_{i+1} - h_{i-1})/h_i$ ,  $g_0 = g_i/n$ ,  $\delta^2 = (g_i - g_0)^2/n_0$ 。

当  $A > 100$  并且  $\delta < 10$  时,则认为该缺值区域为水面区域。因为水面区域周边高程变化平缓,且只要面积达到一定大小的时候,在传感器上才有明显的反映,而小区域情况则可能由多种原因引起。

一旦确定缺值区域为水面类型,实现插值就相对容易,因为水面区域接近一个平面,可以通过缺值边界点生成近似平面实现插值。实现步骤如下:

- 1) 计算多边形中心点。
- 2) 对边界点的数值实现平滑。
- 3) 计算中心点的数值,取值为边界点的算术平均。
- 4) 根据边界点和中心点建立三角网,内部各点的数值采用三角网插值算法得到。

### (2)未知缺值类型的处理

对于非水面区域,本文都认为是未知类型的缺值区域,采用 EM(Expectation Maximization)算法实现插值。

EM 算法针对多变量多个记录但某些记录中存在某些变量的观察值缺失,它首先假定多变量数据记录符合一定的连续型随机变量分布的密度函数,该假定也解释数据产生的机理,是 EM 算法的出发点。设随机变量  $\zeta$  的分布是连续型的,密度函数  $f(x, \theta_1, \dots, \theta_k)$  的形状已知,但含  $k$  个未知参数  $\theta_1, \theta_2, \dots, \theta_k$ 。将  $\zeta_1, \zeta_2, \dots, \zeta_n$  分别代入  $x$ , 将所得  $n$  个相乘而得函数:  $L(\zeta_1, \dots, \zeta_n; \theta_1, \dots, \theta_k) = \prod_{i=1}^n f(\zeta_i, \theta_1, \dots, \theta_k)$  函数  $L$  称为似然函数。当子样  $\zeta_1, \zeta_2, \dots, \zeta_n$  固定时,  $L$  是  $\theta_1, \theta_2, \dots, \theta_k$  的函数。最大似然法的目标是:取使  $L$  达到最大值的  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$  作为  $\theta_1, \theta_2, \dots, \theta_k$  的估值。EM 算法是处理非完整数据时估计最大似然函数的通用的迭代方法。

EM 算法的基本步骤如下:

- 1) 用估计值代替缺失值。
- 2) 估计参数。
- 3) 假设新估计参数是满足一定要求的,重新估计缺值。
- 4) 再估计参数,重复以上 3), 4) 两步,直到收敛<sup>[10-12]</sup>。算法流程见图 3。

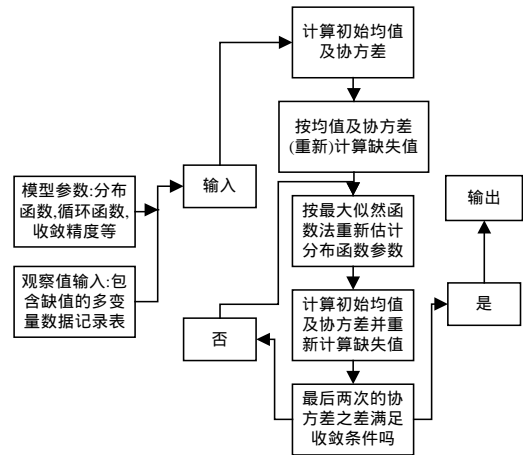


图 3 EM 算法流程

## 3 实验结果

图 4~图 7 是对于同一数据采用不同算法计算得到的结果。从图 4 的结果可以清楚地看到,本文采用的方法对于水面的描述更加准确,且对城市对象的高差变化和细微特征的刻画也很理想。



图 4 距离反比算法



图 5 三角网算法



图 6 最临近点算法



图 7 本文算法

(下转第 63 页)