

利用全局仿射模型进行卫星图像快速三维重建

陈豹^{1,2,3}, 王品贺⁴, 董秋雷^{1,2,3}

1. 中国科学院自动化研究所 多模态人工智能系统全国重点实验室, 北京 100190;

2. 中国科学院大学 人工智能学院, 北京 100049;

3. 中国科学院脑科学与智能技术卓越创新中心, 北京 100190;

4. 东北大学 计算机科学与工程学院, 沈阳 110169

摘要: 基于多视角卫星遥感图像的三维场景重建是遥感领域中一项极具挑战性的任务。针对现有方法重建速度慢的问题, 本文提出了一种利用全局式仿射模型进行卫星图像快速重建方法。该方法首先将多视角卫星图像裁剪成一组局部图像块并计算相应局部场景的三维仿射点云, 然后引入一种基于局部点云的全局式仿射运动矩阵估计算法计算出每个视角对应的相机仿射运动矩阵, 在此基础上利用少量地面控制点恢复出场景的三维欧氏结构。在两个公共数据集上的实验结果表明, 本文方法的重建速度、精度和完整性在大多数情况下均优于3种主流文献方法。

关键词: 三维重建, 卫星遥感图像, 仿射成像模型, 欧氏结构恢复, 全局仿射矩阵

中图分类号: TP701/P2

引用格式: 陈豹, 王品贺, 董秋雷. 2024. 利用全局仿射模型进行卫星图像快速三维重建. 遥感学报, 28(6): 1576–1587

Chen B, Wang P H and Dong Q L. 2024. Fast 3D reconstruction of satellite images via the Global Affine Model. National Remote Sensing Bulletin, 28(6): 1576–1587 [DOI: 10.11834/jrs.20222039]

1 引言

近年来, 随着卫星成像技术的飞速发展 (张过等, 2019), 基于高分光学遥感图像的三维重建技术, 在城市规划、地理测绘、定位导航等诸多领域展现出巨大的应用潜力 (李畅等, 2014), 并受到了研究学者的广泛关注。

大多数文献中的卫星光学影像重建方法采用有理多项式相机RPC (Rational Polynomial Camera) 模型 (Hartley 和 Saxena, 1997) 进行场景重建。作为通用传感器模型, RPC一般包含78个多项式系数及10个归一化常数。目前, 基于RPC模型的重建算法通常先完成两视角重建, 然后基于地面高程图 (DSM) 融合算法完成多幅视图的重建工作。其重建流程主要包括: (1) 采用视差 (D'Angelo 和 Reinartz, 2012; Kuschik, 2013; Rupnik 等, 2017) 或光流 (Bosch 等, 2016) 获得图像之间的稠密对应关系, (2) 通过捆绑调整优化光学卫星图像

RPC模型的估计残差, (3) 基于RPC模型进行三角化恢复图像对应的三维场景结构。Bosch 等 (2017) 利用捆绑调整优化成对光学遥感图像附带偏差补偿的RPC模型, 以满足模型之间的最大一致性, 并使用RPC模型三角化恢复对应地面场景的三维结构。S2P (De Franchis 等, 2014) 利用两视图之间的指向误差校正替代了复杂的非线性捆绑调整, 且基于RPC模型高程迭代实现了成对光学遥感图像的三维重建过程。在此基础上, Facciolo 等 (2017) 分析了影响两视图重建精度的主要因素, 并依据这些因素对数据集中的图像对进行启发式排序, 最后选择性地融合了S2P生成的独立的两视图重建结果。需要指出的是, 尽管上述基于RPC模型的卫星图像三维重建方法已经展现出较高的重建精度, 然而这些重建方法都难以避免求解关于RPC模型的复杂三阶多项式, 且估计RPC模型需要至少39组地面控制点及相关图像的对应关系, 因此重建过程需要耗费大量的计算

收稿日期: 2022-01-30; 预印本: 2022-07-08

基金项目: 国家自然科学基金 (编号: 61991423, U1805264); 中国科学院战略先导专项 (编号: XDB32050100); 北京市科技计划 (编号: Z211100011021004)

第一作者简介: 陈豹, 研究方向为三维计算机视觉。E-mail: chenbao2019@ia.ac.cn

通信作者简介: 董秋雷, 研究方向为模式分类、视觉定位与三维建模。E-mail: qldong@nlpr.ia.ac.cn

时间。

值得注意的是, 计算机视觉领域中提出了大量的基于光学图像的三维场景重建方法 (Govindu, 2004; Snavely 等, 2008; Wilson 和 Snavely, 2014; Schonberger 和 Frahm, 2016; 于英 等, 2019; 薛俊诗 等, 2020)。然而, 这些方法往往只适用于处理面阵相机拍摄的图像, 无法处理卫星上搭载的推扫式线阵相机所获取的图像, 也就是说, 这些计算机视觉领域中的三维重建方法并不能直接处理卫星图像三维重建任务。针对这一情况, 研究学者不断探索如何将视觉领域的重建算法应用于遥感图像。Fraser 等 (2014) 利用地面控制点和图像点之间的真实对应关系直接拟合仿射相机模型, 进而通过拟合生成的仿射相机模型实现光学遥感图像的仿射重建, 但该方法针对整幅光学遥感图像进行操作, 从而重建精度较低。Zhang 等 (2019) 则利用已知的 RPC 模型拟合弱透视投影模型, 并基于 COLMAP 引入了深度重参数化方法实现了基于计算视觉重建管道的光学遥感影像三维重建。Wang 等 (2022) 提出了一种基于仿射成像模型的增量式光学卫星图像重建方法, 该方法不再依赖于 RPC 模型, 只需要预先获得至少 4 个地面控制点即可实现场景重建。

综上所述, 尽管近年来文献中提出了许多基于光学卫星图像的三维重建方法, 使得关于光学卫星图像的三维重建技术取得了快速的发展, 然而这些方法或者需要反复地执行捆绑调整优化以减少近似模型误差, 或者需要 RPC 模型中的大量参数参与重建计算, 往往需要耗费大量的时间和计算资源。因此, 如何在保持场景重建精度的前提下降低三维重建时间成为一个亟待解决的问题。

针对上述问题, 本文提出了一种利用全局式仿射模型进行卫星图像快速重建方法。本文方法假设大尺寸卫星遥感图像中的局部图像块符合仿射成像模型。首先, 将输入的多视角卫星图像裁剪成一组含有重叠区域的小尺寸图像块, 并利用这些图像块分别计算出相应的局部场景仿射点云。然后, 设计一种基于局部点云的全局式仿射运动矩阵估计算法, 该算法利用局部点云之间的相对仿射变换一次性地计算出每个拍摄视角对应的相机仿射运动矩阵。最后, 利用获得的每个拍摄视角下的相机仿射运动矩阵以及少量地面控制点 (不少于 4 个) 重建出稠密的场景三维点云。

本文的主要贡献在于: (1) 设计了一种基于局部点云的全局式仿射运动矩阵估计算法, 该算法能够一次性地计算出每个拍摄视角对应的相机仿射运动矩阵, 有效地降低了相机仿射运动矩阵的估算时间。(2) 基于上述全局式仿射运动矩阵估计算法, 本文提出一种利用全局式仿射模型进行卫星图像快速重建方法。不同于增量式视觉重建方法 (Zhang 等, 2019; Wang 等, 2022) 与基于 RPC 模型的重建方法 (Bosch 等, 2016; Facciolo 等, 2017), 该方法不依赖于 RPC 模型且无需反复执行捆绑调整。

2 研究方法

2.1 重建流程

本文工作采用 Wang 等 (2022) 引入的假设: 大尺寸卫星遥感图像中的局部小尺寸图像块近似符合仿射成像模型。基于该假设, 本文提出了一种利用全局式仿射模型进行卫星图像快速重建方法, 包括以下主要步骤 (图 1 为重建流程图):

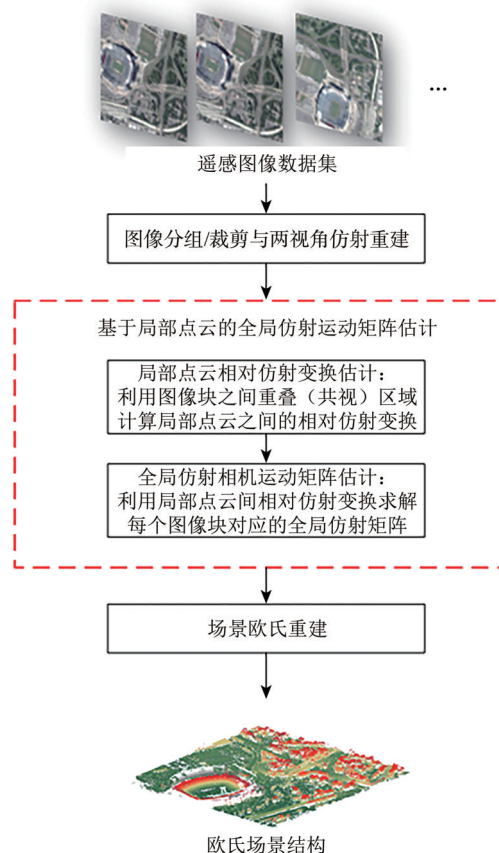


图1 所提方法的重建流程

Fig. 1 Structure of the proposed method

(1) 图像分组/裁剪与两视角仿射重建。对于光学遥感图像数据集, 该模块根据图像的拍摄日期进行分组, 并将每组图像子集裁剪成具有重叠区域的小尺寸图像块。对于任意两视角下具有公共区域的图像块, 应用因式分解算法同时获取局部仿射坐标系下的仿射相机矩阵和局部点云。

(2) 基于局部点云的全局式仿射运动矩阵估计。计算不同局部点云之间存在的相对仿射变换, 利用这些相对仿射变换一次性地计算出每个图像块对应的全局仿射运动矩阵。

(3) 场景欧氏重建。利用每个图像块对应的相机全局仿射运动矩阵以及至少4个地面控制点, 重建出场景的稠密欧氏结构。在下文中, 本文将对上述步骤进行详细介绍。

2.2 图像分组/裁剪与两视角重建

本节采用 Wang 等 (2022) 的处理方式, 对数据集中的图像进行图像分组/裁剪与两视角重建。具体如下:

(1) 图像分组/裁剪: 考虑到基于成对光学遥感的三维重建精度容易受到季节、天气、光照、阴影等因素影响, 利用图像拍摄日期 (相邻7天为一组) 对输入的多视角图像进行分组。将每组原始的卫星图像裁剪为尺寸为 1000×1000 的图像块, 且使得相邻的图像块具有尺寸为 500×1000 的重叠区域, 裁剪方式如图2所示。后续将对每组图像分别进行重建。

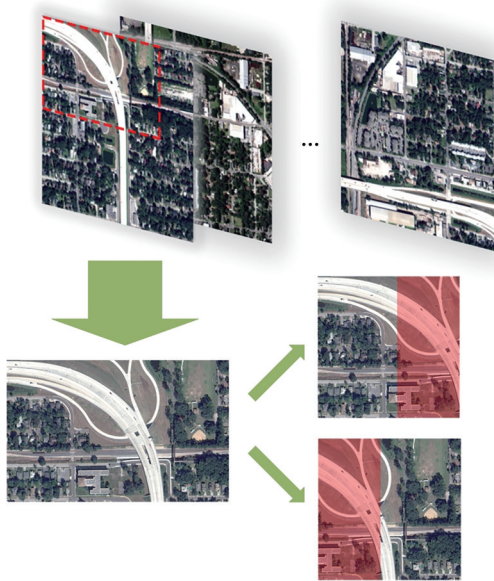


图2 图像裁剪示意图

Fig. 2 Illustration on image cropping

(2) 两视图仿射重建: 对于每个图像块, 使用 SIFT (Lowe, 2004) 算法提取特征点并计算不同图像块之间的特征点对应关系。对于每对具有不少于4个对应点的图像块, 通过因式分解算法 (Hartley 和 Zisserman, 2004) 计算每个图像块在局部仿射坐标系下的仿射运动矩阵及相应的仿射点云。具体地, 给定任意两视图下具有公共区域的两个图像块 (不失一般性, 记为第1、2图像块)。假设这两个图像块有 n (大于等于4) 个对应点, 并令 $X_i \in \mathbf{R}^3$ 表示第 i 个三维点的非齐次坐标, $x_i^k \in \mathbf{R}^2$ 表示在三维点 X_i 在第 k ($k=1, 2$) 个图像块对应的特征点的非齐次坐标。

仿射相机模型定义为

$$x_i^k = M^k X_i + t^k \quad (1)$$

式中, 仿射运动矩阵 $A^k \in \mathbf{R}^{2 \times 4}$ 由线性变换矩阵 $M^k \in \mathbf{R}^{2 \times 3}$ 和位移向量 $t^k \in \mathbf{R}^2$ 构成。

通过分别对这两个图像块的坐标进行中心化 (即对于每个图像块, 将每个特征点坐标减去图像块中所有特征点坐标的均值), 式 (1) 可以化简为

$$\mathbf{W} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_n^1 \\ x_1^2 & x_2^2 & \cdots & x_n^2 \end{bmatrix} = \begin{bmatrix} M^1 \\ M^2 \end{bmatrix} [X_1 \ X_2 \ \cdots \ X_n] \quad (2)$$

考虑到外点的影响, 等式 (2) 可以转化为如下关于 M 的最小化问题:

$$\min_{M, X} \|\hat{\mathbf{W}} - M\mathbf{X}\|_F^2 \quad (3)$$

式中, $\hat{\mathbf{W}} = \begin{bmatrix} \hat{x}_1^1 & \hat{x}_2^1 & \cdots & \hat{x}_n^1 \\ \hat{x}_1^2 & \hat{x}_2^2 & \cdots & \hat{x}_n^2 \end{bmatrix}$, $\hat{x}_i^k$ 表示图像特征点的量测值, $M = [M^1, M^2]^T$, $X = [X_1 \ X_2 \ \cdots \ X_n]$, 最小化问题可以通过奇异值分解直接求出闭式解。

2.3 基于局部点云的全局仿射运动矩阵估计

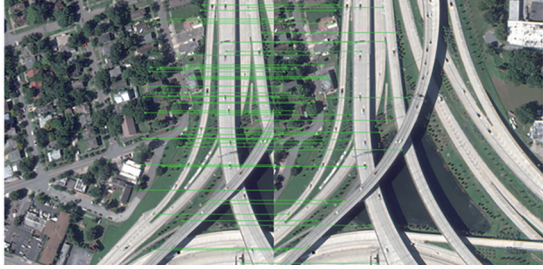
通过因式分解获得所有小图像块的局部仿射重建结果后, 本文根据小图像块之间的共视区域 (或重叠区域) 可以将各小图像块的局部仿射运动矩阵及仿射重建点云注册到同一个全局仿射坐标系。

基于此, 本文提出了一种基于局部点云的全局仿射运动估计算法。该算法分为如下两个主要步骤:

(1) 局部点云间的相对仿射变换估计。如图3所示, 无论是不同视角下小图像块之间的 SIFT 匹配, 还是同一视角相邻小图像块重叠区域中重复出现的 SIFT 特征点, 都可以为求解局部仿射重建点云之间的相对仿射变换提供足够的约束条件。具体地, 给定任意一对具有公共区域的局部点云 A

和 B , 令仿射矩阵 $H_{mn} \in \mathbf{R}^{4 \times 4}$ 表示两者之间满足的仿射变换关系。利用 A 和 B 对应的 SIFT 特征点索引, 得到这两组局部点云中的对应点, 记为 X_m 和 X_n 。理想情况下, 这些对应点满足如下等式关系:

$$H_{mn} X_m = X_n \quad (4)$$



(a) 相邻图像块之间的 SIFT 匹配

(a) SIFT matching between adjacent image tiles



(b) 不同视角下图像块之间的 SIFT 匹配

(b) SIFT matching between image tiles under different viewing angles

图 3 不同图像块之间的 SIFT 匹配

Fig. 3 SIFT matching between different image tiles

(2) 全局仿射相机运动矩阵估计。根据步骤 (1), 可以得到一组局部点云间的相对仿射变换矩阵, 本文将利用这些相对仿射变换矩阵计算所有相机 (对应于每个图像块) 在同一个仿射坐标系下的相机运动矩阵。

将所有参与两视角仿射重建的图像对构建无向图 $G = (V, E)$, 如图 4 所示, 其中 V 表示节点的集合, E 表示边的集合, 无向图 G 中的每一个节点 $v_m \in V$ 表示一组图像对。若节点 m 和 n 对应的图像对之间有重叠区域或共视区域, 则认为节点 m 和 n 之间存在一条边 $e_{mn} \in E$ 。假设节点 m 和 n 对应的点云的绝对仿射变换矩阵分别表示为 H_m 和 H_n , 节点 m 和 n 之间的相对仿射变换表示为 H_{mn} , 则它们之间满足

$$H_{mn} H_m = H_n \quad (5)$$

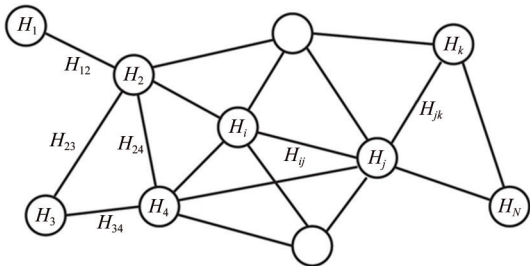


图 4 点云间相对仿射变换构建的无向图 G

Fig. 4 Undirected graph G constructed by relative affine transformation between point clouds

若图 G 存在 N 个节点, M ($M > N$) 条边 (节点之间的相对仿射变换), 则式 (5) 用矩阵形式可

利用任意 4 组对应点, 即可求得 H_{mn} 。考虑外点的影响, 本文还使用 RANSAC (Fischler 和 Bolles, 1981) 算法计算满足约束方程式 (4) 的最大一致集, 然后利用最小二乘方法求得优化的相对仿射变换 H_{mn} 。

表示为

$$\begin{pmatrix} H_{12} & \hat{0} & \hat{0} & \cdots & \hat{0} & \hat{0} \\ H_{13} & \hat{0} & \hat{0} & \cdots & \hat{0} & \hat{0} \\ \hat{0} & H_{23} & \hat{0} & \cdots & \hat{0} & \hat{0} \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ \hat{0} & \hat{0} & \hat{0} & \cdots & H_{N-1,1} & \hat{0} \\ \hat{0} & \hat{0} & \hat{0} & \cdots & \hat{0} & H_{NN-1} \end{pmatrix}_{4M \times 4N} \begin{pmatrix} H_1 \\ H_2 \\ H_3 \\ \vdots \\ H_{N-1} \\ H_N \end{pmatrix}_{4N \times 4} = \begin{pmatrix} \hat{0} & I & \hat{0} & \cdots & \hat{0} & \hat{0} \\ \hat{0} & \hat{0} & I & \cdots & \hat{0} & \hat{0} \\ \hat{0} & \hat{0} & \hat{0} & \cdots & \hat{0} & \hat{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ I & \hat{0} & \hat{0} & \cdots & \hat{0} & \hat{0} \\ \hat{0} & \hat{0} & \hat{0} & \cdots & I & \hat{0} \end{pmatrix}_{4M \times 4N} \begin{pmatrix} H_1 \\ H_2 \\ H_3 \\ \vdots \\ H_{N-1} \\ H_N \end{pmatrix}_{4N \times 4} \quad (6)$$

式中, I 表示 4×4 的单位矩阵, $\hat{0}$ 表示 4×4 的零矩阵。式 (6) 可表示为

$$PY = QY \quad (7)$$

式中, $P \in \mathbf{R}^{4M \times 4N}$ 是由 H_{mn} 和 $\hat{0}$ 构成的稀疏矩阵, $Q \in \mathbf{R}^{4M \times 4N}$ 是由 I 和 $\hat{0}$ 构成的稀疏矩阵。 Y 表示 $[H_1, H_2, H_3, \dots, H_{N-1}, H_N]^T_{4N \times 4}$ 。但由于存在外点, 式 (7) 在一般情况下不成立, 所以原问题转化为

$$\begin{aligned} \min_Y & \|SY\|_F^2 \\ \text{s.t.} & \|Y\|_F^2 = 1 \end{aligned} \quad (8)$$

式中, $S = P - Q$, 对式 (8) 进行 Frobenius 范数优化即可将此最小化问题转化为有约束的最小二乘进行求解, 得到图 G 中每个节点对应的绝对仿

射变换矩阵。

假设无向图 G 中节点 i 对应的局部仿射点云表示为 X_i ，对应的绝对仿射变换矩阵表示为 H_i ，可以通过 $\hat{X}_i = H_i X_i$ 将该组点云变换到某一个仿射坐标系下。依次对无向图 G 中所有节点对应的局部点云进行仿射变换，从而将所有局部点云变换到一个统一的全局仿射坐标系，进而求解得到整个场景的稀疏仿射结构。

利用稀疏点云和图像块之间的投影关系可计算相机的初始全局仿射运动矩阵，表示为

$$x_i^k = A^k X_i \quad (9)$$

式中，图像 I^k 对应的第 i 个仿射三维点的齐次坐标表示为 X_i ，对应的图像特征点的非齐次坐标表示为 x_i^k ，则仿射运动矩阵表示为 $A^k \in \mathbb{R}^{2 \times 4}$ ，仅需4组2D与3D之间的对应关系即可线性求解。

为了改善稀疏场景结构以及精确求解相机矩阵，本文对稀疏三维点云和初始的相机全局仿射运动矩阵进行如下捆绑调整BA (Bundle Adjustment) (Triggs等, 2000):

$$\min_{A^k, X_i} \sum_{i=1}^N \sum_{k=1}^M \delta_{ik} \|x_i^k - \lambda(A^k, X_i)\|_{\text{huber}} \quad (10)$$

式中， x_i^k 表示第 k 幅图像中与三维点 X_i 对应的图像点坐标， λ 表示三维点与图像域之间的投影关系。若第 k 个相机能观测到三维点 X_i ，那么 δ_{ik} 为1，否则 δ_{ik} 为0。 $\|\cdot\|_{\text{huber}}$ 表示损失函数为 Huber 范数 (Huber, 1981)， A^k 和 X_i 均为需要优化的变量。本文使用 Ceres solver (Agarwal 和 Mierle, 2012) 库中的 Levenberg-Marquardt 算法求解式 (10) 的非线性方程。

2.4 场景欧氏重建

对于2.2节分组后的每组图像块，利用仿射立体校正算法 (Ozcanli等, 2015) 和 MGM 立体匹配算法 (Facciolo等, 2015) 获取图像块之间的稠密对应关系。然后根据图像块之间的稠密对应关系和每个图像块对应的全局仿射运动矩阵，通过仿射三角测量直接计算得到地面场景的稠密仿射点云。进而考虑到场景仿射结构与欧氏结构之间仅存在一个空间仿射变换，因此相似于文献 (Wang等, 2022)，利用大于等于4个地面控制点拟合该空间仿射变换矩阵，从而获得每组图像块对应的场景欧氏结构。最后，利用 Facciolo 等 (2017) 的融合地面高程图的方法，对每组图像块对应的欧

氏结构生成高程图并进行配准，合并为完整的场景三维结构。

3 数据结果处理与分析

在本节中，本文所提出的方法与3种主流重建方法：S2P (Facciolo等, 2017)、COLMAP (Zhang等, 2019) 及 JHUAPL (Bosch等, 2016) 进行对比评测。

3.1 数据集与评价方法

3.1.1 数据集

为了评估重建方法的有效性，本文使用国际公用的 MVS3DM 数据集和 DFC2019 数据集进行对比评测。

(1) MVS3DM 数据集 (Bosch等, 2016): 该数据集包含阿根廷圣费尔南多附近 100 km² 区域的 50 张遥感影像。数据集中的图像均是在 2014 年—2016 年由 WorldView-3 卫星采集得到。在此基础上，霍普金斯大学提供了 8 个观测点的预处理遥感图像数据集以及对应的激光雷达真值图像。

(2) DFC2019 数据集 (Bosch等, 2019): 该数据集由美国佛罗里达州的杰克威尔逊和内布拉斯加州的奥马哈两个城市 106 个观测站点的 2783 张图像构成。数据集中的图像均由 WorldView-3 卫星在 2014 年—2016 年采集得到，每张 RGB 图像的尺寸大小为 2048 × 2048 像素。

3.1.2 评测方法

与 Bosch 等 (2017) 的评测方法类似，首先，将数据库中提供的真实点云与重建方法计算出的点云映射到相同的地理网格中，分别生成地面高程图；然后，将这两个高程图进行配准对齐；最后，统计对齐后的两个高程图之间的像素值差异。评测过程主要采用重建精度、完整度、时间三者作为评价指标。其中精度表示中值高度误差 ME (Median Error)，完整度 CP (Completeness) 表示误差小于 1 m 的重建点云与整体点云的比值，时间 (Time) 表示重建运行的时间。

3.2 实验结果对比

由于两个评测数据集图像采集时间的跨度均在一年以上，因此数据集中的图像均可划分为多组图像，并同时对其进行处理。为了更加公平地评估算法的重建效率和重建性能，本文分别对每

个站点对应的一组图像以及所有图像进行了重建结果对比。

在实验1中，本文对不同站点对应的数据集的某一组图像进行重建结果对比，图像分组方法采用2.2节的分组方法，即每组图像包含3—8张遥感图像；而实验2则完成不同站点对应的数据集中所

有图像的重建工作，即每个数据集包含47—50张遥感图像。

（1）实验1。本文在14个站点完成了地面场景的重建工作，重建精度和完整度的对比结果如表1所示。

表1 不同站点对应的一组图像的重建结果对比
Table1 Comparison of experimental results for the same group of satellite images

观测站点	重建方法	完整度/%	精度/m	观测站点	重建方法	完整度/%	精度/m
Explorer	JHUAPL	65.8	0.387	MasterPark	JHUAPL	49.6	0.853
	COLMAP	61.7	0.276		COLMAP	20.0	0.452
	S2P	72.4	0.319		S2P	51.0	0.322
	Ours	72.6	0.285		Ours	51.3	0.320
MasterPro.1	JHUAPL	60.3	0.602	MasterSeq.1	JHUAPL	64.8	0.460
	COLMAP	63.1	0.302		COLMAP	62.4	0.312
	S2P	67.1	0.340		S2P	67.2	0.369
	Ours	70.4	0.324		Ours	70.9	0.361
MasterPro.2	JHUAPL	55.2	0.807	MasterSeq.2	JHUAPL	56.7	0.806
	COLMAP	31.9	0.364		COLMAP	58.9	0.323
	S2P	63.6	0.392		S2P	64.6	0.381
	Ours	65.7	0.356		Ours	66.8	0.367
MasterPro.3	JHUAPL	42.1	1.566	MasterSeq.3	JHUAPL	49.1	1.235
	COLMAP	10.9	0.632		COLMAP	33.9	0.373
	S2P	40.6	0.380		S2P	48.5	0.317
	Ours	40.9	0.376		Ours	50.1	0.310
JAX_033	JHUAPL	2.0	14.070	OMA_247	JHUAPL	72.2	0.344
	COLMAP	51.3	0.443		COLMAP	33.4	0.224
	S2P	76.8	0.267		S2P	77.8	0.162
	Ours	77.4	0.261		Ours	78.0	0.156
JAX_068	JHUAPL	64.9	0.603	OMA_287	JHUAPL	72.7	0.279
	COLMAP	45.1	0.344		COLMAP	52.5	0.245
	S2P	68.2	0.462		S2P	80.4	0.183
	Ours	69.3	0.407		Ours	80.3	0.180
JAX_175	JHUAPL	55.9	0.814	OMA_381	JHUAPL	93.0	0.129
	COLMAP	41.4	0.451		COLMAP	73.9	0.112
	S2P	65.7	0.223		S2P	96.4	0.076
	Ours	66.0	0.209		Ours	96.5	0.073

注：最优重建结果在表中以粗体显示。

由表1可见，在完整度方面，本文所提方法在12个站点的完整度表现优于3种对比算法，仅在2个站点略低于S2P和JHUAPL方法，因此可以说明本文所提算法在单组图像中重建完整性较好。而在重建精度方面，本文所提出的方法在14个站点中的9个站点下的表现均优于3种对比算法，从

而展现出了良好的重建性能。本文所提方法在3个观测站点的重建可视化结果如图5所示，地面高程图和误差图对应的彩色条状图如图6所示，通过将本文方法生成的地面高程图与地面真值图像对比可以发现，本文方法重建的结构较为完整。本文在站点 MasterPro.2 与其他重建方法重建结果对比

的可视化结果如图7所示。由图7可见, 本文方法重建的建筑结构细节更加明显, 地面场景更加完整; 其次, 相比于3种对比算法, 本文所提方法在站点 MasterPro.2下表现出了更小的误差以及更高的完整度。为进一步验证本文算法的重建精度, 本文从 MVS3DM 以及 DFC2019 两个公开数据集中

挑选了部分复杂场景区域, 图8给出了若干复杂场景卫星图像示例, 其中包括建筑密集、存在大量阴影以及地面对象复杂等复杂场景图像示例, 表2则展示了本文方法与3种对比方法在上述数据上的重建结果。从表2可以看出, 本文方法在大多数情况下取得了最高的重建精度与完整度。

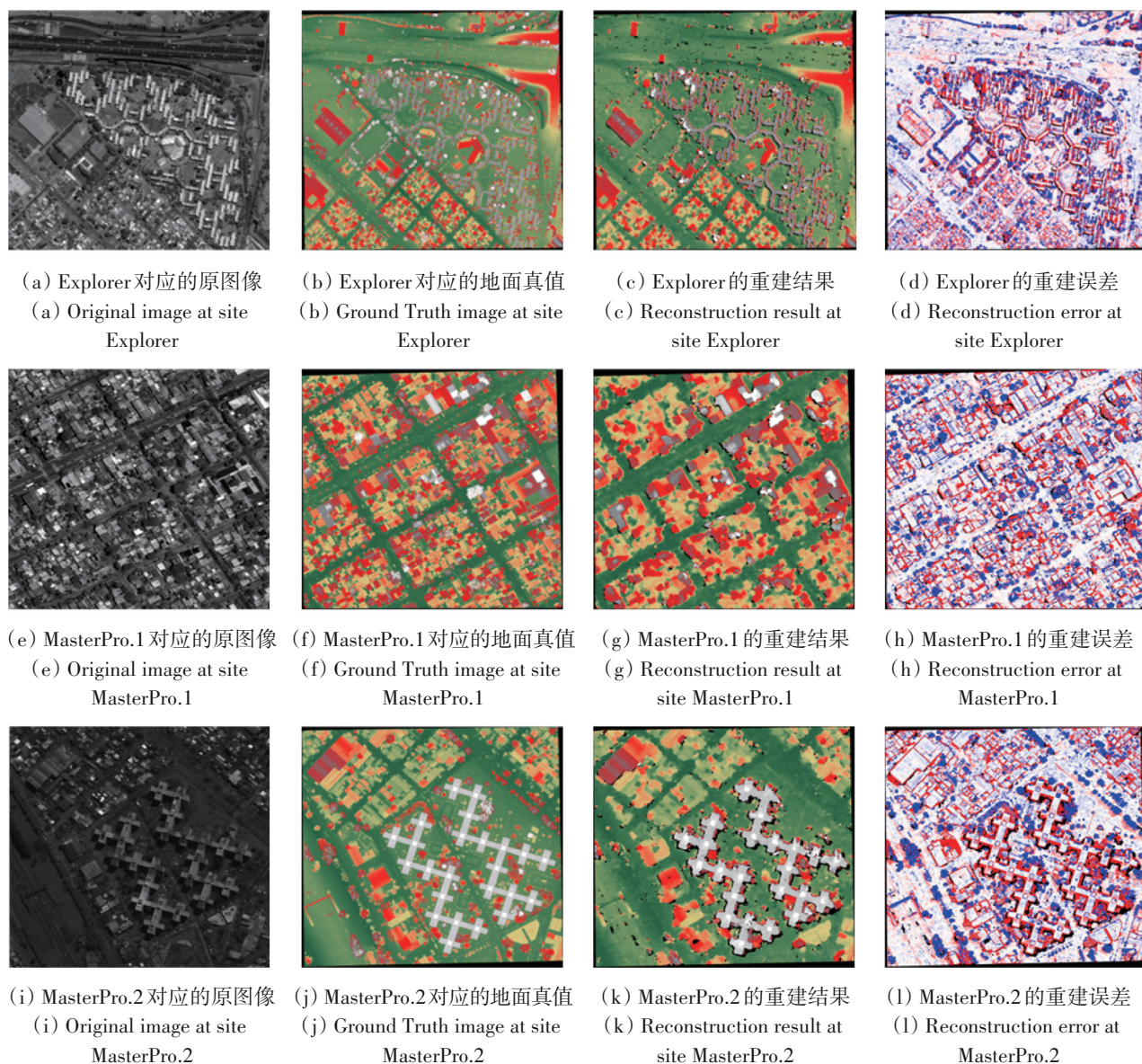


图5 本文方法在3个站点的重建可视化结果

Fig. 5 Visualization results of our method at three sites

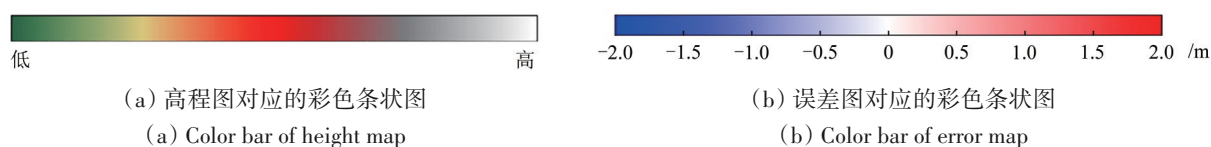


图6 地面高程和重建误差对应的彩色条状图

Fig. 6 Color bars of height map and error map

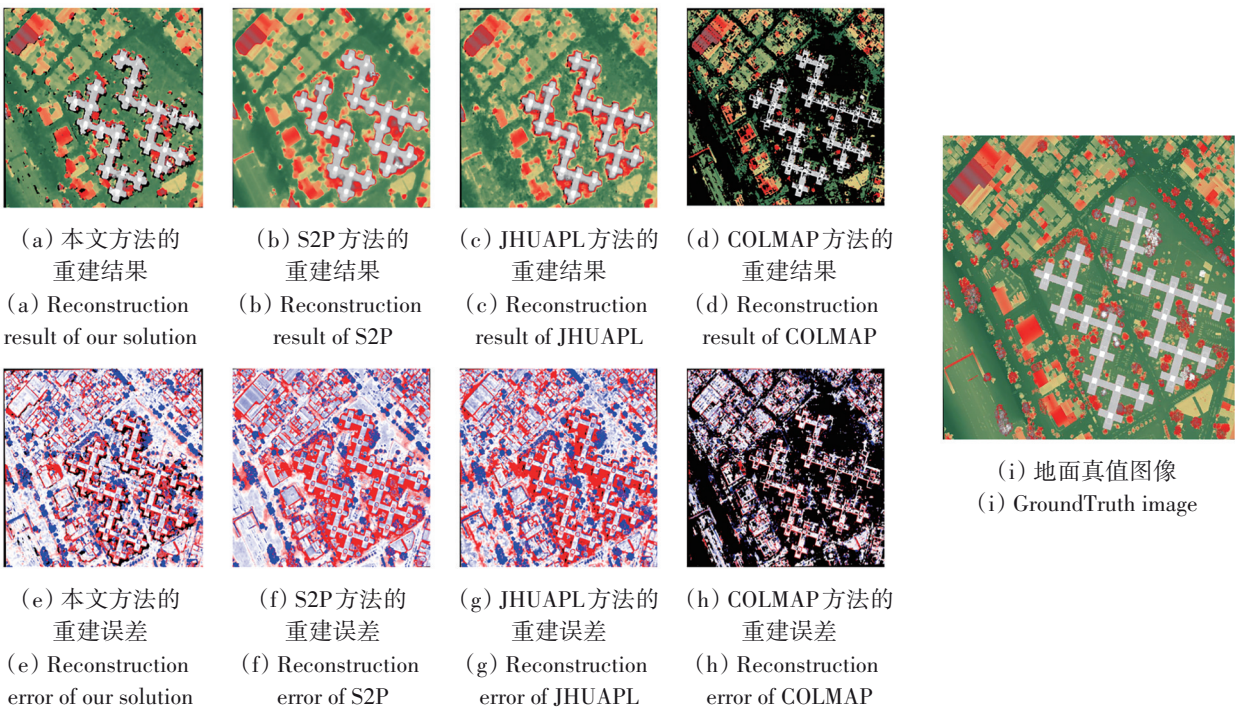


图7 不同方法在站点 MasterPro.2的重建结果及误差对比
Fig. 7 Comparison of reconstruction results and errors at site MasterPro.2

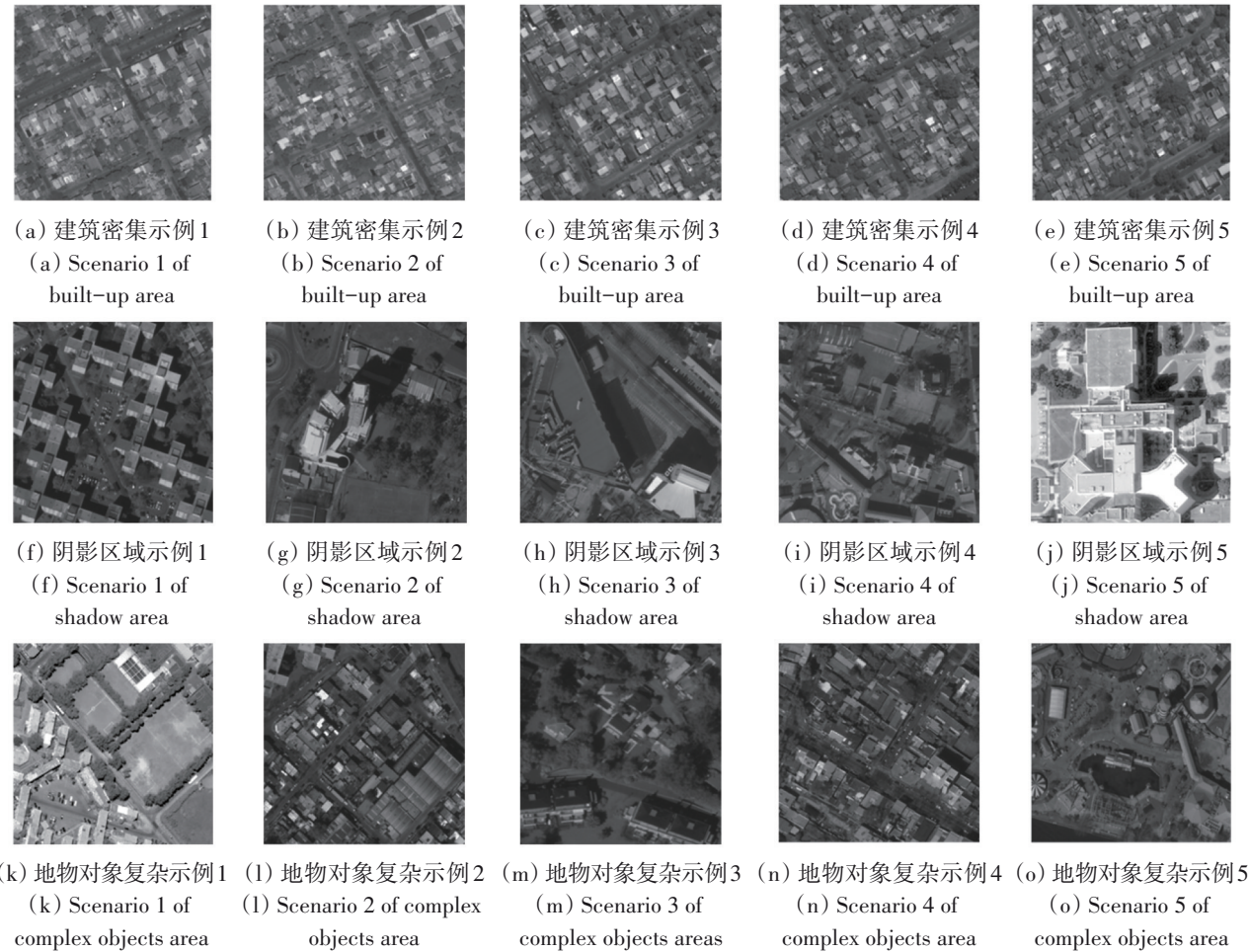


图8 复杂场景卫星图像示例
Fig.8 Satellite images under complex scenarios

表2 复杂场景卫星图像的重建结果对比

Table 2 Comparison of experimental results of satellite images under complex scenes

场景类别	测试场景	JHUAPL		COLMAP		S2P		本文	
		完整度/%	精度/m	完整度/%	精度/m	完整度/%	精度/m	完整度/%	精度/m
建筑物密集	a	62.3	0.609	66.5	0.369	69.5	0.348	70.0	0.345
	b	58.0	0.685	60.8	0.410	67.3	0.379	68.0	0.360
	c	52.3	0.934	71.0	0.393	68.9	0.348	70.2	0.378
	d	49.7	1.010	68.6	0.429	65.8	0.396	68.9	0.384
	e	58.9	0.734	65.9	0.423	65.9	0.474	66.3	0.421
存在大量阴影	f	53.0	0.543	28.7	0.368	54.3	0.621	56.3	0.503
	g	46.1	1.232	15.3	0.607	49.2	0.422	49.7	0.403
	h	47.5	1.013	24.1	0.575	55.6	0.638	60.2	0.572
	i	39.8	0.872	19.9	0.685	57.4	0.693	60.2	0.590
	j	33.2	1.652	55.9	0.388	62.4	0.312	57.6	0.328
地物对象复杂	k	67.5	0.417	58.7	0.326	68.4	0.373	65.8	0.283
	l	64.8	0.367	44.2	0.360	65.4	0.473	68.4	0.354
	m	29.2	2.295	11.2	1.347	34.5	1.451	41.6	1.177
	n	37.7	1.784	34.5	0.835	47.3	0.727	51.7	0.696
	o	53.4	0.791	33.9	0.353	53.3	0.603	65.3	0.483

注：最优重建结果在表中以粗体显示。

在重建出的10组局部场景点云中，本文手动框选建筑物等主要地物，并统计本文方法与3种对比方法对于每种地物的平均重建精度与完整度，并在表3中展示了相应的结果。从表3可以看出，本文方法对于建筑物等主要地物的重建精度和完整度均优于3种对比方法。

表3 不同方法针对主要地物的重建结果对比

Table3 Comparison of experimental results for main objects

主要地物	重建方法	完整度/%	精度/m
建筑物	JHUAPL	53.2	0.876
	COLMAP	51.6	0.537
	S2P	59.3	0.465
	本文	60.7	0.435

注：最优重建结果在表中以粗体显示。

(2) 实验2。本文在4个站点完成了数据集所有图像的重建工作，重建精度和完整度的对比结果如表4所示。

由表4可见，本文所提方法在4个站点的重建效率均优于3种对比算法。与JHUAPL方法相比，本文所提方法无需对复杂的RPC模型参数进行捆绑调整优化，因此能够提升重建效率；与COLMAP方法相比，本文方法只需一次捆绑调整优化即可求解每个相机的全局仿射运动矩阵；而

与S2P方法相比，本文方法直接线性求解得到稠密三维点云，无需进行复杂的高程迭代，因此大幅减少了重建时间。此外，在融合步骤中，S2P方法需要对前50对两视图重建结果进行融合，相较之下本文方法只需对每组图像块生成的高程图进行融合，从而进一步提高了重建效率。综上所述，本文所提出的方法在重建效率方面优于3种对比算法。

值得注意的是，本文统计的基于RPC方法的重建时间是建立在卫星供应商提供RPC模型参数的基础上。如果卫星图像没有提供对应的RPC模型参数，那么3种对比算法的重建效率将大幅度下降。相比之下，本文方法由于采用仿射成像模型而无需RPC模型，因此在重建效率的表现上没有受到影响，有着更广泛的适用性。

在完整度方面，本文方法在4个站点的重建完整度均优于3种对比算法；而在重建精度方面，本文所提出的方法在大部分站点下的实验结果展现出了最优的效果。

综上所述，在大多数情况下，本文方法在重建速度、精度与完整性方面优于COLMAP、S2P与JHUAPL 3种对比算法。

表4 不同站点对应的所有图像的重建结果对比
Table4 Comparison of experimental results for all remote sensing images

观测站点	重建方法	完整度/%	精度/m	时间/min
Explorer	JHUAPL	72.4	0.323	210.6
	COLMAP	72.5	0.315	86.7
	S2P	80.1	0.317	70.1
	本文	80.4	0.310	56.3
Site1	JHUAPL	70.3	0.390	200.7
	COLMAP	66.8	0.450	83.8
	S2P	74.0	0.486	71.4
	本文	74.9	0.455	57.4
Site2	JHUAPL	59.7	0.593	128.9
	COLMAP	63.4	0.393	50.7
	S2P	73.1	0.524	43.8
	本文	73.3	0.522	37.6
Site3	JHUAPL	43.9	1.345	121.4
	COLMAP	50.0	1.632	48.9
	S2P	58.6	1.497	42.1
	本文	59.1	1.311	35.2

注：最优重建结果在表中以粗体显示。

4 结 论

本文提出了一种基于全局式仿射模型估计的卫星图像快速重建方法，该方法假设大尺寸卫星遥感图像中的局部图像块符合仿射成像模型，并引入了一种基于局部点云的全局式仿射运动矩阵估计算法，得益于此，本文只需通过一次捆绑调整优化，便能计算出每个局部图像块对应相机的全局仿射运动矩阵，大大减少了重建运行时间。经实验验证，该方法在保证完整度和精度的前提下能有效地提高遥感影像的重建效率。

然而，受图像中重复纹理、弱纹理、遮挡等诸多因素的影响，本文方法重建出的场景三维点云中往往会存在一定数量的噪声点、空洞和外点。如果将场景地物间的先验信息和规则作为约束条件引入到场景三维重建过程中，则可降低噪声、减少空洞和外点数量，进而改善重建性能。在未来的工作中，本文拟深入研究如何将场景地物间的先验信息与规则有效地引入到本文方法中以进一步改善重建性能。

参考文献(References)

Agarwal S and Mierle K. 2012. Ceres solver: Tutorial and reference.

Google Inc, 2(72): 8.

Bosch M, Kurtz Z, Hagstrom S and Brown M. 2016. A multiple view stereo benchmark for satellite imagery//Proceedings of the 2016 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). Washington, USA: IEEE: 1-9 [DOI: 10.1109/AIPR.2016.8010543]

Bosch M, Leichtman A, Chilcott D, Goldberg H and Brown M. 2017. Metric evaluation pipeline for 3D modeling of urban scenes. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-1/W1: 239-246 [DOI: 10.5194/isprs-archives-XLII-1-W1-239-2017]

Bosch M, Foster K, Christie G, Wang S A, Hager G D and Brown M. 2019. Semantic stereo for incidental satellite images//Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, USA: IEEE: 1524-1532 [DOI: 10.1109/WACV.2019.00167]

D'Angelo P and Reinartz P. 2012. DSM based orientation of large stereo satellite image blocks. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 39(B1): 209-214 [DOI: 10.5194/isprsarchives-XXXIX-B1-209-2012]

De Franchis C, Meinhardt-Llopis E, Michel J, Morel J M and Facciolo G. 2014. An automatic and modular stereo pipeline for pushbroom images. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, II-3: 49-56 [DOI: 10.5194/isprsannals-II-3-49-2014]

Facciolo G, De Franchis C and Meinhardt E. 2015. MGM: A significantly more global matching for stereovision//Proceedings of the British Machine Vision Conference. Swansea: BMVA Press [DOI: 10.5244/C.29.90]

Facciolo G, De Franchis C and Meinhardt-Llopis E. 2017. Automatic 3D reconstruction from multi-date satellite images//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE: 1542-1551 [DOI: 10.1109/CVPRW.2017.198]

Fischler M A and Bolles R C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6): 381-395 [DOI: 10.1145/358669.358692]

Fraser C S, Dare P M and Yamakawa T. 2004. Digital surface modeling from spot 5 HRS imagery using the affine projective model//Proceedings of the XXth International Society for Photogrammetry and Remote Sensing Congress. 35: 385-388

Govindu V M. 2004. Lie-algebraic averaging for globally consistent motion estimation//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. Washington: IEEE: I [DOI: 10.1109/CVPR.2004.1315098]

Hartley R I and Saxena T. 1997. The cubic rational polynomial camera model//Proceeding of the Image Understanding Workshop. New Orleans, USA: Citeceer: 649-653 [DOI: 10.1.1.2.5489]

Hartley R and Zisserman A. 2004. Multiple View Geometry in Com-

- puter Vision. Cambridge: Cambridge University Press.
- Huber P J. 1981. Robust statistics//Lovric M, ed. International Encyclopedia of Statistical Science. Berlin: Springer [DOI: 10.1007/978-3-642-04898-2_594]
- Kuschik G. 2013. Large scale urban reconstruction from remote sensing imagery. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-5-W1: 139-146 [DOI: 10.5194/isprsarchives-XL-5-W1-139-2013]
- Li C, Xiong H, Tao S Y and Han Z K. 2014. Quantitative change detection of vegetation and three-dimensional terrain for geographical disaster areas from optical stereopair images. Journal of Remote Sensing (in Chinese), 18(6): 1258-1267 (李畅, 熊昊, 陶顺勇, 韩振坤. 2014. 地质灾害区光学立体影像植被与3维地形定量变化检测. 遥感学报, 18(6): 1258-1267) [DOI: 10.11834/jrs.20143304]
- Lowe D G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2): 91-110 [DOI: 10.1023/B:VISI.0000029664.99615.94]
- Ozcanli O C, Yi D, Mundy J L, Webb H, Hammoud R and Tom V. 2015. A comparison of stereo and multiview 3-D reconstruction using cross-sensor satellite imagery//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Boston, USA: IEEE [DOI: 10.1109/CVPRW.2015.7301292]
- Rupnik E, Daakir M and Deseilligny M P. 2017. MicMac – a free, open-source solution for photogrammetry. Open Geospatial Data, Software and Standards, 2(1): 14 [DOI: 10.1186/s40965-017-0027-2]
- Schonberger J L and Frahm J M. 2016. Structure-from-motion revisited//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 4104-4113 [DOI: 10.1109/CVPR.2016.445]
- Snaveley N, Seitz S M and Szeliski R. 2008. Modeling the world from internet photo collections. International Journal of Computer Vision, 80(2): 189-210 [DOI: 10.1007/s11263-007-0107-3]
- Triggs B, McLauchlan P F, Hartley R I and Fitzgibbon A. 2000. Bundle adjustment—a modern synthesis//Proceedings of the International Workshop on Vision Algorithms. Corfu: Springer: 298-372 [DOI: 10.1007/3-540-44480-7_21]
- Xue J S, Yi H, Wu Z H and Chen X N. 2020. A hybrid multi-view 3D reconstruction method based on scene graph partition. Acta Automatica Sinica, 46(4): 782-795 (薛俊诗, 易辉, 吴止钺, 陈向宁. 2020. 一种基于场景图分割的混合式多视图三维重建方法. 自动化学报, 46(4): 782-795) [DOI: 10.16383/j.aas.c180155]
- Wang P H, Shi L M, Chen B, Hu Z Y, Dong Q L and Qiao J Z. 2022. Pursuing 3-D scene structures with optical satellite images from affine reconstruction to Euclidean reconstruction. IEEE Transactions on Geoscience and Remote Sensing, 60: 5632214 [DOI: 10.1109/TGRS.2022.3213546]
- Wilson K and Snaveley N. 2014. Robust global translations with IDS-fM//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 61-75 [DOI: 10.1007/978-3-319-10578-9_5]
- Yu Y, Zhang Y S, Xue W and Wang T. 2019. A incremental structure from motion method of robustness enhancement and accuracy improvement. Acta Geodaetica et Cartographica Sinica, 48(2): 207-215 (于英, 张永生, 薛武, 王涛. 2019. 一种稳健性增强和精度提升的增量式运动恢复结构方法. 测绘学报, 48(2): 207-215) [DOI: 10.11947/j.AGCS.2019.20170665]
- Zhang G, Jiang Y H, Li L T, Deng M J and Zhao R S. 2019. Research progress of high-resolution optical/SAR satellite geometric radiometric calibration. Acta Geodaetica et Cartographica Sinica, 48(12): 1604-1623 (张过, 蒋永华, 李立涛, 邓明军, 赵瑞山. 2019. 高分辨率光学/SAR卫星几何辐射定标研究进展. 测绘学报, 48(12): 1604-1623) [DOI: 10.11947/j.AGCS.2019.20190469]
- Zhang K, Snaveley N and Sun J. 2019. Leveraging vision reconstruction pipelines for satellite imagery//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops. Seoul, Korea: IEEE: 2139-2148 [DOI: 10.1109/ICCVW.2019.00269]

Fast 3D reconstruction of satellite images via the Global Affine Model

CHEN Bao^{1,2,3}, WANG Pinhe⁴, DONG Qiulei^{1,2,3}

1. National Key Laboratory of Multi-modal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;
2. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China;
3. Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Beijing 100190, China;
4. School of Computer Science and Engineering, University of Northeastern, Shenyang 110169, China

Abstract: Three-dimensional (3D) scene reconstruction based on multiview satellite remote sensing images is a challenging task in the field of remote sensing. Most of the existing methods either have to perform bundle adjustment repeatedly or must calculate several parameters in the rational polynomial camera model, resulting in a relatively long reconstruction time. To solve the abovementioned problems, this study considers that the local small-sized patches in large-sized satellites could be approximately modeled by the affine

imaging model and proposes a fast 3D reconstruction method of satellite images based on global affine model estimation.

First, the input multiview satellite images are cropped into a set of small-sized patches with overlapping regions. For each pair of patches that have a sufficient number of point correspondences from two views, the corresponding 3D affine point cloud is calculated. Second, on the basis of the obtained local point clouds, a global affine camera motion estimation algorithm is presented for calculating the affine motion matrices of the cameras corresponding to all the patches in a unified coordinate system. Finally, the obtained affine camera motion matrices and a few ground control points are utilized to recover the Euclidean scene structure.

3D reconstruction is conducted for the same group of remote sensing images and all remote sensing images to verify the effectiveness of the method. The proposed solution is compared with three state-of-the-art methods (i.e., COLMAP, S2P, and JHUAPL). Experimental results on two public datasets (i.e., MVS3DM and DFC2019) show that the proposed method outperforms the three comparison algorithms in most cases with respect to speed, accuracy, and completeness. To verify further the reconstruction accuracy of the method, this study selects 15 complex scene areas from two public datasets, including complex scenes with built-up areas, shadow areas, and complex object areas. For 15 complex scenarios, the proposed method outperforms the three methods with respect to accuracy and completeness in most cases.

This study proposes a fast reconstruction method of satellite images based on the global affine model estimation algorithm. The method assumes that the local image tile in large-scale satellite remote sensing images conforms to the affine imaging model and introduces a global affine motion matrix estimation algorithm based on local point clouds. Consequently, the proposed solution can calculate the global affine motion matrix of each local image tile through only one bundle adjustment, considerably reducing the reconstruction running time. Experimental results show that the proposed method can quickly solve the global affine matrix corresponding to each image tile and realize fast 3D reconstruction of remote sensing images.

Key words: 3D reconstruction, satellite images, affine imaging model, Euclidean structure update, global affine matrix

Supported by National Natural Science Foundation of China (No. 61991423, U1805264); Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB32050100); Science and Technology Program of Beijing (No. Z211100011021004)