



三维重建算法研究综述

摘要

三维重建技术能够应用在医学中的三维 CT 图像、机器人中的路径规划及考古工作中的遗迹的展示等诸多领域中。三维重建效果的完整性、精确性与稀疏点云重建 (Structure From Motion, SFM)、稠密点云重建 (Multi-View System, MVS) 有着直接关系,本文对这两个方法展开了具体的介绍。稀疏点云重建中主要介绍了特征点检测与匹配以及 SFM 重构方法,本文对近几年关于特征点检测与匹配的研究进行了总结,将 SFM 重构方法分为全局式、增量式、混合式,并进行了详细介绍。稠密点云重建中主要介绍了深度图估计,就传统的几何计算法、深度学习与几何计算相结合的方法、基于深度学习的方法进行了总结,同时介绍了三维重建的评价指标。最后对基于图像的三维重建进行了总结,并列举出了在未来可能面对的问题及发展趋势。

关键词

三维重建;图像;稀疏点云重建;稠密点云重建;评价指标

中图分类号 TP391

文献标志码 A

收稿日期 2019-11-20

资助项目 国家自然科学基金 (61773219, 61701244); 国家重点研发计划重点专项课题 (2018YFC1405703)

作者简介

张彦雯,女,硕士生,研究方向为计算机视觉三维重建.526531735@qq.com

胡凯(通信作者),男,博士,副教授、高级实验师,研究方向为机器人.nuistpanda@163.com

1 南京信息工程大学 自动化学院,南京,210044

2 南京信息工程大学 江苏省大气环境与装备技术协同创新中心,南京,210044

0 引言

三维重建指对三维物体建立适合计算机表示和处理的数学模型,经过十多年的发展,已经取得了巨大的成就。随着科学技术的发展,利用计算机构建出真实景物的三维模型,已经成为很多领域开始深层研究之前必不可少的一部分。如在医学治疗中,利用三维 CT 图像判断病情;在文物保护中,将文物进行三维复原,供游客参观。除此之外,在游戏开发、工业设计、航天航海等诸多领域,三维重建发挥着不可或缺的作用。目前,研究者们主要通过三种方式来获取三维模型:一是直接根据传统的几何建模技术;二是基于多幅不同角度拍摄的图像,运用计算机视觉理论进行三维建模,该方式为基于图像的三维重建;三是利用三维扫描设备对目标进行扫描,然后重建目标的三维模型,该方式为基于点云的三维重建。在上述三种方法中,传统的几何建模技术发展最为成熟,利用该方法可以得到非常精确的三维模型,因此该方法已经广泛应用于诸多领域。然而采用这种建模技术周期长、操作复杂,对于许多不规则的自然或人为物体,建模效果与真实场景仍存在差异。因此基于图像与基于点云的三维重建成为了三维重建领域中的热点。

基于图像的三维重建主要是将二维图像恢复成三维模型,其分为基于单幅图像的三维重建与基于多幅图像的三维重建。基于单幅图像的三维重建不能应用于大规模场景,因此本文主要介绍了基于多幅图像的三维重建。

基于点云的三维重建主要是将点云数据恢复成图形、图像,然后通过计算机处理形成三维模型。其主要流程有:数据配准^[1]、点云数据预处理^[2]、分割^[3]、三角网格化^[4-5]、网格渲染。由于点云获取设备获取点云角度单一,因此如何对多个点云数据进行配准拼接一直是研究的重点和难点。

20 世纪 90 年代 Debevec 等^[6]提出了用参数几何体来表示初始模型的方法。在 21 世纪初,学者们不断对传统的几何思想进行研究,至此通过参数几何法进行三维重建的研究已经较为成熟,同时在这 10 年间不断有应用于三维重建的软件问世。考虑到 3DMAX 和 Maya 等软件建模周期长、工作量大且不能对场景中隐藏细节进行还原,Chen 等^[7]基于 Windows 环境开发了一个多视图三维重建系统,该系统主要将基于多视图三维重建的算法进行集成,只需将各个视角的

图像作为输入,便可以得到三维模型.相对于传统的软件,该系统操作简单且三维模型的精确度高、细节处理效果更好.随着科技的发展和社会的进步,各行各业对于三维重建的需求日益增多,三维重建的发展又达到了新的阶段.基于图像的三维重建主要包括稀疏点云重建和稠密点云重建.稀疏点云重建(Structure From Motion, SFM)包括特征点检测与匹配以及 SFM 重构两个部分.初始的检测算子容易受到光照、尺度、旋转等因素的影响, Lowe^[8-9] 提出了 SIFT (Scale Invariant Feature Transform) 算法解决了以上因素的影响,至此稀疏点云重建的精度得到了大幅度提高.随着学者们不断创新与发展, SURF (Speeded Up Robust Features)、FAST (Features from Accelerated Segment Test)、ORB (Oriented FAST and Rotated BRIEF) 等优秀算法被相继提出.利用特征点检测与匹配,计算特征点个数达到要求的图像对,为 SFM 重构中计算初始图像对及 R 、 T 求解做铺垫. SFM 重构大致分为三类:增量式重构、全局式重构、混合式重构.其中增量式重构、全局式重构为传统方法.2017年, Cui 等^[10] 提出了混合式 SFM; 2018年, Cui 等^[11] 提出了一种外环图像匹配、内环增量式校准的方法; 2018年, Zhu 等^[12] 提出了一种嵌套分解的分布式框架用于全局式重构.在稠密点云重建 (Multi-View System, MVS) 中,深度图的计算尤为重要,其发展之初计算方法主要是计算几何参数以求得图像深度信息,例如 Gallup 等^[13] 提出的 Plane-Sweeping、Bleyer 等^[14] 提出的 PatchMatch 等.2010年前后,随着深度网络的发展,国内外一批精英学者将深度网络模型与几何参数法相结合以计算出图像深度,如 Huang 等^[15] 提出的 Deep MVS.2018年,香港科技大学训练了一个深度网络模型 MVSNet^[16],该网络模型能够脱离传统的几何计算,通过输入多视角图像得到深度图.2019年, Chen 等^[17] 等提出了 Point MVSNet 神经网络框架.基于深度网络的三维重建仍在不断发展创新.

本文主要介绍基于图像的三维重建的算法,并简要地介绍三维重建效果的评价指标.基于图像的三维重建的流程主要包括:稀疏点云重建、稠密点云重建、点云网格建模、三维语义建模、三维矢量建模.其中稀疏点云重建是三维重建中搭建场景框架的过程,稠密点云重建是三维重建中增强场景稠密性的过程,三维重建效果的完整性与精确性与这两个步骤有直接关系.本文针对稀疏点云重建与稠密点云

重建的发展进行系统综述.

本文第一节介绍稀疏点云重建,主要包括特征点检测与匹配、SFM 重构;第二节主要介绍稠密点云重建,尤其对深度图的计算进行详细介绍;第三节主要介绍三维重建效果的评价指标;第四节对三维重建进行总结,并列举出在未来可能面对的问题及发展趋势.

1 稀疏点云重建

稀疏点云重建是通过相机的移动来确定目标的空间和几何关系的.首先输入一个场景的多视角图像,检测每张图像中的特征点,并对特征点进行匹配,然后运用 SFM 重构算法,恢复出相机的参数、相机之间的相对位置关系及场景的稀疏点云.最后捆绑调整优化相机位姿、相机参数、场景结构.

稀疏点云重建的流程为:特征点检测,特征点匹配,相机位姿(旋转矩阵 R 、平移矩阵 T) 初始值的求解,捆绑调整优化相机位姿、相机参数、场景结构.

1.1 特征点检测与匹配

特征点检测与匹配的作用是计算特征点个数达到要求的图像对,使得图像与图像间一直进行传递,从而形成一条轨迹,为 SFM 重构中的初始图像对及 R 、 T 求解做铺垫.如图 1 给出了利用 ORB 算法进行特征点检测与匹配的效果.

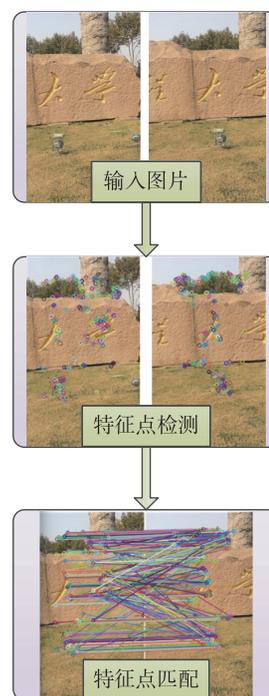


图 1 特征点检测与匹配

Fig. 1 Feature point detection and matching

1.1.1 特征点检测

局部特征点,主要分为斑点和角点两类特征点,斑点是指图像中像素或灰度值大的区域,角点是指图像中物体的拐角或线条交叉的部分.斑点检测主要包括二阶的拉普拉斯高斯边缘提取算法(Laplacian Of Gaussian,LOG)、尺度不变特征变换算法(Scale Invariant Feature Transform,SIFT)^[8-9]、加速稳健特征算法(Speeded Up Robust Features,SURF)^[18]等.角点检测主要有Harris角点特征提取算子、加速分割测试特征提取(Features from Accelerated Segment Test,FAST)^[19]、特征点检测算子(Oriented FAST and Rotated BRIEF,ORB)^[20]等.

1999年SIFT算子被Lowe^[8]提出,并于2004年

得到完善^[9],该算法主要是寻找多尺度空间下的关键点,计算关键点的位置、方向、尺度信息,以生成特征点描述子.SIFT算子提取的关键点精度高、不易受光照噪声的影响,但是该算子也存在计算量大、实时性能不高的缺点.Bay^[18]提出了SURF算子,与SIFT算法不同的是,SURF特征点主方向的分配不是通过梯度计算,而是利用了Haar小波,SURF的计算速度是SIFT的3至7倍.在Susan角点^[21]检测的基础上,Rosten等提出了FAST^[19]算子,随后很多学者在该算子的基础上进行改进.随着对特征点描述子的深入研究,近几年提出了很多新的算子,例如Susan^[21]、ORB^[20]等.表1对经典特征点检测算法进行了对比,并且提供了代码资源.

表1 特征点检测算法优缺点

Table 1 Advantages and disadvantages of feature point detection algorithms

特征点检测算法	优点	缺点	代码链接
SIFT ^[8-9]	在处理发生尺度、旋转变化大的场景时具有极大优势,面对光照、噪声具有很好的鲁棒性	计算维度大,运算速度很慢	https://github.com/pitZER/SiftGPU
SURF ^[18]	较SIFT计算量小,耗时短.具有旋转不变性、尺度不变性,鲁棒性好	运算速度比较慢	https://github.com/kinathru/SURFDetector
FAST ^[19]	运算速度快	容易受噪声因素干扰,鲁棒性差.不具备旋转不变性	https://github.com/Mofafa/FAST-Corner-Detection
ORB ^[20]	运算速度快,具有稳定的旋转不变性	不具备尺度不变性	https://github.com/raulmur/ORB_SLAM2

1.1.2 特征点匹配

特征点匹配即通过寻找两幅图像中相似的特征点,来确定两幅图像的位置关系.SIFT和SURF算子在对特征点匹配时都是通过两特征点间的欧式距离计算出匹配度的.SURF算子中加入了Hessian矩阵,能够判断出两个特征点的方向,相比于SIFT算子,SURF算子匹配的精确度更高.SURF算子需要对A视角图像的某个特征点与B视角图像上的每个特征点进行匹配,因此耗时较长.为解决问题,崔振兴等^[22]将特征点邻域内像素之间的差值形成的特征向量与SURF特征描述子相结合形成68维特征向量;李红波等^[23]提出了一种基于距离约束的改进SURF算法,该算法针对不同尺寸的图像,构建高斯金字塔图像层提取合适数目的特征点,采取特征点优化处理的方法使得提取的特征点分布均匀,并引入KD-tree索引使得特征点匹配更加准确快速;Xu等^[24]提出了一个加速算法,该算法采用了一种新的索引模式来寻找最相似的图像,然后通过一种新的接近最近邻搜索算法来匹配特征向量,大大加快了匹配速度;Sergeeva等^[25]提出了一种异构

背景的前景检测技术,在特征点匹配后运用该技术去除图像背景,该算法对复杂背景的目标重建效果很好.

1.2 SFM 重构

SFM重构的思想是利用相机运动轨迹来估算相机参数.相机在不同视点摄取多幅图像,利用这些图像计算出相机的位置信息以及运动轨迹,从而在空间坐标系下生成三维点云,恢复出物体的空间结构.由于初步匹配过程形成的特征点较为粗糙,因此在估算相机参数之前,常用几何约束删除匹配错误的特征点对.SFM重构的方法主要有增量式和全局式两种.

1.2.1 增量式重构

增量式重构是从两张图像中重建出部分场景以及两相机的 R (旋转矩阵)、 T (平移矩阵),然后添加图像,利用PnP计算出第三幅图像的 R 、 T ,通过三角化重建出更多的场景.由于在求位姿与三维点时会有大量噪声干扰,后续计算结果会发生漂移,因此引入BA(Bundle Adjustment)算法进行优化.在每次添加图像完成三角化后都需要进行优化.增量式重

构的步骤为:初始化、PnP 解算、三角化、BA 算法优化.其中较为重要的步骤是利用 PnP 算法求解相机位姿. PnP 有很多种类,如 P3P^[26]、EPnP^[27-28]、DLT^[29]、UPnP^[30]和 MRE^[31-32].

增量式重构的具体流程是:首先输入第 1、第 2 张图像计算出相机位姿并重建出部分场景,然后输入第 3 张图像进一步调整相机位姿并重建场景,直到输入所有不同视角的图像,得到所有相机的初始位姿和初始三维点.如图 2 所示,通过输入第 1、第 2 张图像,重建出第 1 个场景,由于 2 张图重建出的场景效果不是很好,图 2 中没有将场景 1 展示出来;再添加第 n 张图像与之前重建的场景 $N-1$ 融合获得下一步的新场景,直到所有图像添加结束,重建出全部场景.

增量式 SFM 在对大规模场景进行三维重建时,由于误差累积会造成场景漂移以及反复地捆绑调整会耗费大量的时间,为了解决这些问题,段建伟^[33]将图优化理论与增量式 SFM 相结合,在构造图优化模型时,将重投影误差平方和设置为代价函数从而进行优化,以达到减少误差的效果.而刘彬等^[34]提出了线性的增量式 SFM,与传统增量式 SFM 相比改进了三维重建初始像对的筛选策略,通过将局部的光束法优化、局部外点过滤以及局部三角化三者组合的方式减少误差.

不同于上述方法,2018 年 Cui 等^[11]提出了一种外环图像匹配、内环增量式校准的方法.如图 3 所示下方箭头为外环,上方箭头为内环.外环是一个正交匹配环,通过迭代选择图像相似图的正交 MSTs (Maximum Spanning Trees) 进行图像匹配,内环是一个增量式相机校准循环.在 MSTs 进行图像匹配时,对于每个匹配的图像对,首先估计对应的极线几何,包括第 j 张图像相对第 i 张图像的相对旋转矩阵 \mathbf{R}_{ij} 、相对平移矩阵 \mathbf{t}_{ij} ,其中相机绝对位姿 $\{\mathbf{R}_i, \mathbf{C}_i\}$ 与相对位姿 $\{\mathbf{R}_{ij}, \mathbf{t}_{ij}\}$ 的关系如式(1)所示,其中 λ_{ij} 为尺度因子, \mathbf{R}_i 为相机绝对旋转矩阵, \mathbf{C}_i 为相机中心矩阵.

$$\begin{cases} \mathbf{R}_{ij} = \mathbf{R}_j \mathbf{R}_i^T, \\ \lambda_{ij} \mathbf{t}_{ij} = \mathbf{R}_j (\mathbf{C}_i - \mathbf{C}_j). \end{cases} \quad (1)$$

对于每一对匹配的图像,如果两个相机都经过校准,参考相对位姿 $\{\mathbf{R}_{ij}^{\text{ref}}, \mathbf{t}_{ij}^{\text{ref}}\}$ 可以通过式(1)计算得到.那么相对旋转误差为

$$\varepsilon_1 = \text{acos} \frac{\text{tr}(\mathbf{R}_{ij}^T \cdot \mathbf{R}_{ij}^{\text{ref}}) - 1}{2},$$

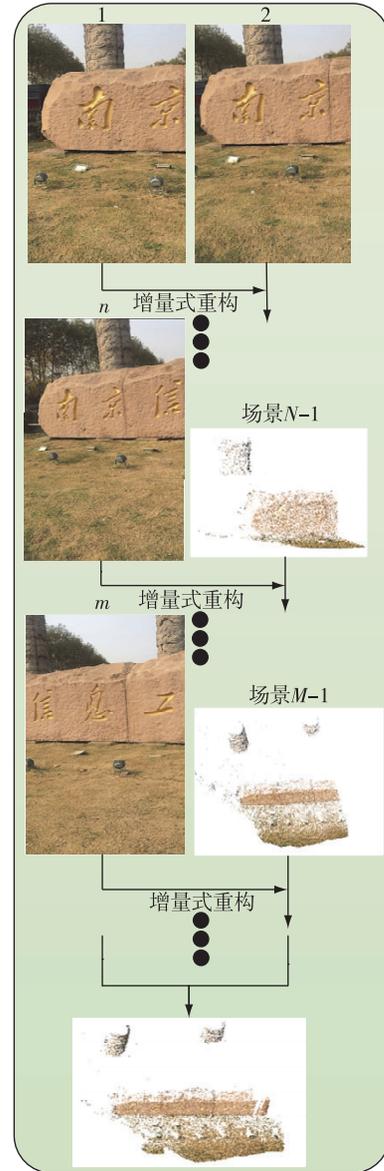


图 2 增量式重构结构

Fig. 2 Incremental SFM structure diagram

相对平移误差定义为

$$\varepsilon_2 = \text{acos} (\|\mathbf{t}_{ij} - \mathbf{t}_{ij}^{\text{ref}}\|).$$

设定一个阈值 β_2 , 如果 $\varepsilon_1, \varepsilon_2$ 大于 β_2 , 那么该匹配对被认为是错误的, 在当前迭代中被忽略. 实验表明即使存在多个错误匹配, 但是通过上述筛选, 在前三个正交 MTSs 中, 大多数匹配对都具有精确的相对位姿估计, 更适合摄像机标定^[11]. 通过不断地图像匹配和相机标定, 两个循环的收敛速度都很快并且排除了大量的冗余对, 该方法在处理各种图像数据集时效率高、鲁棒性强.

1.2.2 全局式重构

全局式 SFM 如图 4 所示, 它将所有图像添加到

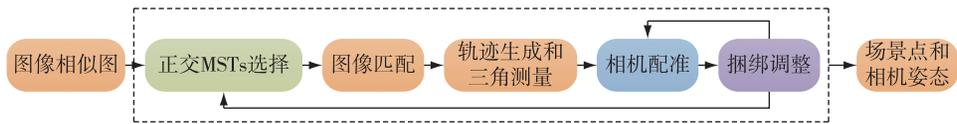


图3 内外环增量式重构^[11]

Fig. 3 Incremental SFM of inner and outer rings^[11]

场景中,计算所有摄像机的旋转矩阵、平移矩阵和所有摄像机的位置,然后进行三角化.全局式 SFM 采用的是将相机相对位姿均匀分布的方法,该方法在处理大规模重建时,相机之间的相关性不同,会导致重建效果差.于是,2018 年 Zhu 等^[12] 基于全局式 SFM 提出了一种嵌套分解的分布式框架.首先将原始的摄像机分布图分割成多个子图,对每个子图迭代至收敛;然后连接多个子图之间的变量,对连接多个子图的变量进行迭代优化.第一步优化为局部运动平均,第二步优化为全局运动平均.该方法中局部相机姿态在全局运动平均中是固定的,降低了内存占用.实验结果表明在全局式重构中加入该分布式框架具有较高的鲁棒性.

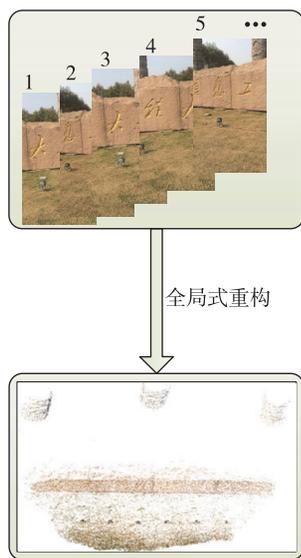


图4 全局式重构结构图

Fig. 4 Global SFM structure diagram

1.2.3 混合式重构

2017 年, Cui 等^[10] 提出了混合式 SFM, 主要思想是将前两者进行融合, 利用全局式估计摄像机旋转矩阵, 增量式估计摄像机位置, 然后进行局部捆绑调整.

2 稠密点云重建

由于 SFM 用来做重建的点是由特征匹配提供

的, 导致匹配点稀疏, 重建效果并不是很好, 所以需要 MVS 进行完善. 稠密点云重建 (MVS) 的目的是通过最佳方法对不同图像上的同一个点进行匹配, 增强重建场景的稠密性.

2.1 一致性度量

如图 5 所示, 假设图 b 中存在一点 P' 与图 a 中点 P 在不同视角的同一点, 由于几何约束, 点 P 与相机 a 的返回射线和点 P' 与相机 b 的返回射线一定相交于一点, 点 P 所对应的点 P' 一定位于极线上. MVS 主要任务是通过一种最佳搜索的方法, 匹配不同图像的同一点. 接下来工作就是在图片上的一条线上进行探测, 寻找两张图片上的同一点, 主要方法为逐像素判断, 由此提出了一致性度量函数.

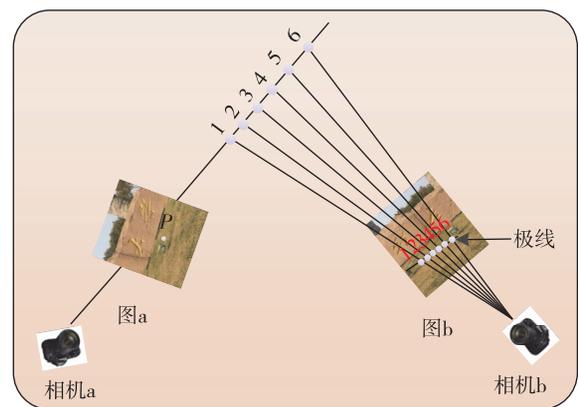


图5 匹配点的几何约束示意图

Fig. 5 Geometric constraint of matching points

两视图图像一致性的度量方法有: 平均绝对差算法 (Mean Absolute Differences, MAD)、差值平方和 (Sum of Squared Differences, SSD)、绝对差值和 (Sum of Absolute Differences, SAD)、归一化互相关性 (Normalized Cross-Correlation, NCC) 等. 这三个算法大同小异, 采用的方法都是遍历整个搜索图, 寻找到与模块图最为相似的部分, 不同的只是计算方法不一样.

为了减少搜索范围以及增加匹配率, 刘小群^[35] 提出了基于视差梯度的区域匹配算法, 该算法引入 NSAD 函数, 同时对不同区域采用不同大小的窗口进行匹配. 张一飞等^[36] 提出了融合边缘特征的 Edge-

Gray 算法,该算法在 SAD 算法匹配之前先计算出边缘特征图,在匹配过程中融合了边缘特征图和灰度图.

2.2 深度图的计算

深度图是指将相机到场景中各点距离作为像素值的图像.在 MVS 中对于深度图的计算是一大重点.

2.2.1 几何法

深度图的计算方法有很多,如 Yang 等^[37]提出了 Plane-Sweeping 算法,而 Gallup 等^[13]将该算法用在三维重建中.文献^[13]主要思想是将深度范围内分割一个一个平面,通过观察投影到摄像机上的点的颜色是否一致来判断这些点是否位于可见物体的表面上.但是该算法难以应对遮挡问题,另外,当纹理弱、光照变化大时还会重建出很多杂点.为了解决这些问题,Shu 等^[38]采用平面扫描立体的方法提出了一种深度区间估计方法:首先确定重建目标投影出的二维范围,然后将二维范围反投影到三维空间,得到多个三维平面;接着将来自其他视图的平面投影到视图 j 的坐标系中,视图 j 的深度区间等于所有平面相交的 z 轴的范围;最后计算每个视图的深度图.王跃嵩^[39]对 Plane-Sweeping 算法进行了改进,使用稀疏点云来确定需要进行局部平面扫描的平面,同时提出了一种迭代的方法,该方法能够消除杂点,避免将被遮挡点的分数拉低.

Bleyer 等^[14]将 Barnes 等^[40]提出的 PatchMatch 算法应用到了深度图计算中.PatchMatch 算法通过随机初始化,按照从左到右、从上到下的方法进行传播及随机搜索,不断地进行迭代,寻找出最合适的 patch,达到图像修复的目的.而 Bleyer 等^[14]正是运用 PatchMatch 算法根据一个平面找到一个近似的最近邻.首先在每个像素点中存储深度值和深度值的法向量,并将深度值和深度值的法向量随机初始化,然后把相邻像素的深度值和深度值法向量转换到当前像素上得到一个空间平面,将空间平面投影到图像上,比较两者的一致性差别,若后者的一致性更优,则用相邻像素的存储值替换当前像素的存储值,如此进行传播及随机搜索.该方法得到了较好的亚像素,但是在面对缺失纹理的区域时,效果不佳.为了解决这些问题,学者们对 PatchMatch 算法进行了改进.Besse 等^[41]提出了一种方法能够解决连续性标签处理问题,同时该算法比 PatchMatch 更精确.该方法主要是将 PatchMatch 与粒子置信度传播算法整合到 BP 框架下.Li 等^[42]利用自底向上的方法设计了

两层匹配,用预先训练好的卷积神经网络得到像素块的相似度,将该相似度作为底层代价,将底层代价与利用超像素块单元构建的顶层匹配代价相组合从而构建匹配能量.在视差优化过程中,与超像素间不同的是,该算法提出了利用 GC(Graph Cut) 优化超像素内的每一像素点的标签,可以降低在纹理性弱的区域误匹配的概率.

2.2.2 几何与深度学习相结合的方法

Wang 等^[43]提出了一种双通道法生成深度图像:首先利用暗通道生成初始深度图,然后用一个互补的亮通道来识别初始深度图中可能的错误估计,最后利用语义信息训练一个深度模型对初始图像中错误的深度估计进行修正.该方法能够解释场景元素之间的高阶语义关系,具有定量与定性的优势.Huang 等^[15]提出了 DeepMVS 的方法用来计算视差图,前期需要通过 SFM 对相机进行校准以及得到每张图像的姿态,通过图像预处理,将参考图像和各级视差图输入到网络中,整个网络模型包括三部分: PatchMatch 匹配网络、卷内特征聚合网络、卷间特征聚合网络,最后输出视差图,后续可以通过视差图求得深度图.该方法在面对纹理信息较少的区域和薄结构时相对于其他算法的效果更好,但是不能预判包含树木或草地的植被区域的差异且处理速度慢.

2.2.3 基于深度学习的方法

考虑到传统的 MVS 难以处理弱纹理区域、瘦小的结构以及反射和投射的表面,基于深度学习的 MVS 算法应运而生.前者主要通过提取深度信息计算深度图,后者主要通过机器学习算法训练模型计算深度图.2018 年 Yao 等^[16]提出了非结构化多视点三维重建网络(MVSNet).MVSnet 是一个端到端的神经网络,如图 6 所示.第一部分将不同视角的图像输入到网络中获取输入图像的特征图.第二部分为文献^[16]创新部分:通过一个基于单应变换的 2D 转 3D 的变换方法,将特征图变换到一个 256 个深度平面上,将转换后的 3D 特征图称为 Feature Volume.式(2)表示将 N 张多视角图像变换成 N 个特征量,其中 Deep Features 为将图像卷积后提取到的深度特征.式(3)表示将第 i 张特征图变换到深度为 d 的特征平面上的单应矩阵 $H_i(d)$,将特征图变换到特征量不同的深度平面上,至少有一个深度是正确的.其中 I 为单位矩阵、 K 为相机内部参数、 R 为相机旋转矩阵、 t 为相机位移向量、 d 为深度变量、 n 为参考图像相机的主轴.

$$\text{Deep Features}\{F_i\}_{i=1}^N \xrightarrow{\text{warp}} \text{Feature Volumes}\{V_i\}_{i=1}^N, \quad (2)$$

$$H_i(d) = K_i R_i \left(I - \frac{(t_1 - t_i) n_1^T}{d} \right) \cdot R_1^T K_1^T. \quad (3)$$

第三部分将 N 个特征量通过逐个像素的方差计算合并成一个花费量.第四部分花费量经过一个四级的U-Net 结构来生成一个可能量,从而得到每个深度下每个像素的可能性大小.第五部分结合上述每个像素在不同深度上的概率分布通过计算深度的数学期望得到初始深度图.第六部分将初始深度图与原图像串成一个四通道输入得到深度残差,进而与初始深度图相结合得到最终深度图.该网络框架可以适

用于任意张视图输入,截止到 2018 年 4 月 18 日之前在公开 Tanks and Temples 数据集上的测试排名位居第一,其运行速度更快,不需要任何的微调,具有良好的泛化性.

在上述 MVSNet 中对花费量正则化占用了大量内存,使网络难以对高分辨率场景进行重建.为改进这个缺点,Yao 等^[44]于 2019 年对原来的 MVSNet 进行优化提出了 R-MVSNet,其主要是针对 MVSNet 的第四部分进行了改进,如图 7 所示.R-MVSNet 用 GRU 单元代替原来的 3D CNN,GRU 单元可以沿深度方向收集上下文信息,同时可以控制上一个深度平面的更新及重置信息的程度,这样占用的内存会

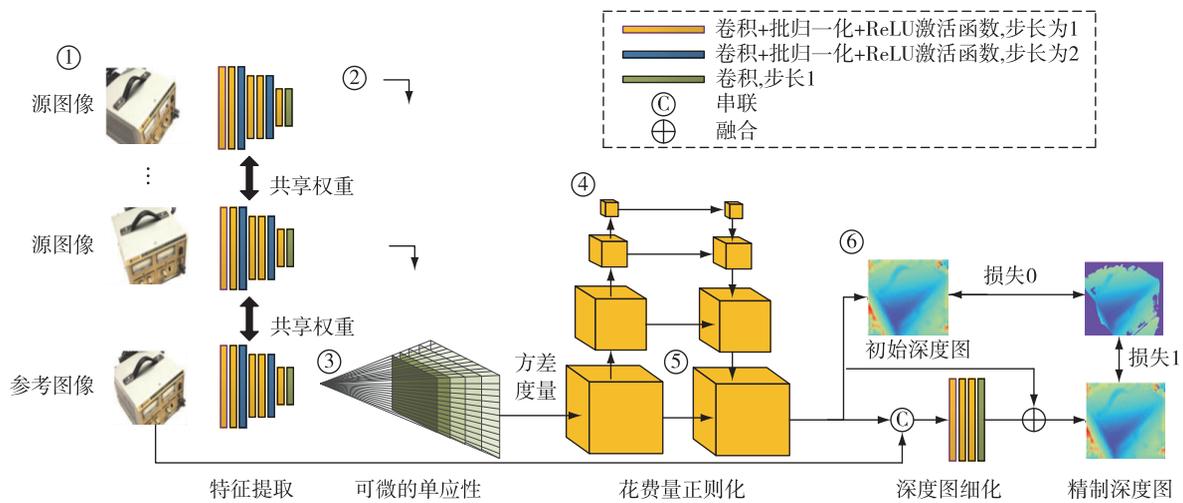


图 6 MVSNet^[16] 框架

Fig. 6 The network of MVSNet^[16]

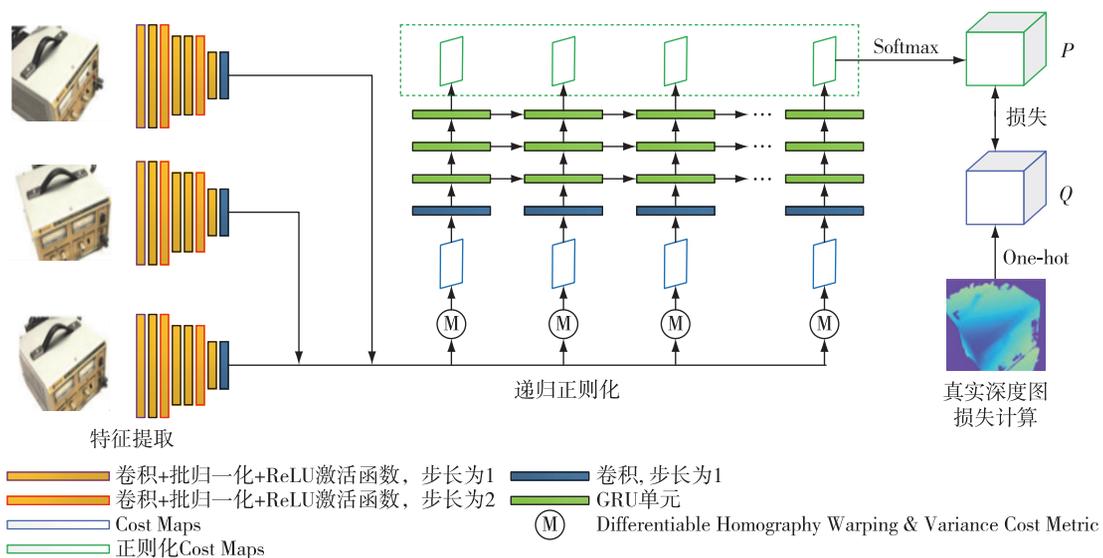


图 7 R-MVSNet^[44] 部分结构

Fig. 7 Part of the R-MVSNet architecture^[44]

大幅度减少同时效果也没有下降.

2019年Chen等^[17]提出了PointMVSNet神经网络框架,主要分为两大部分.第一部分首先将不同视角的图像输入到网络中,生成粗糙深度预测图,然后转换成粗糙点云.第二部分利用图像金字塔网络提取三维点云中每个点在各视角图像中的纹理特征,构建特征增强点云,然后利用点流模型迭代出精确而稠密的点云.该方法直接将场景作为点云处理,计

算量大幅度减少,保留空间连续性,精度、速度、灵活性提高显著,重建结果更为高效、精确(表2).

利用基于图像的方法进行三维重建,图8为拍摄的RGB图像(通过摄像机摄取多幅不同视点的RGB图像),图9为稀疏点云重建效果,图10为稠密点云重建正面效果,图11为稠密点云重建侧面效果.

表2 深度图计算方法比较

Table 2 Comparison of depth map calculation methods

深度图的计算方法	文献	优点	源码
改进的 Plane-Sweeping	文献[38]	能够较好地处理复杂场景,准确率高、速度快	https://github.com/jhwei/Multi-view_Stereo (原始算法代码)
	文献[39]	能够应对遮挡问题,以及当纹理弱、光照变化大的情形,操作简单	
改进的 PatchMatch	文献[41]	能够解决连续性标签处理问题,该算法比 PatchMatch 更精确	https://github.com/ivanbergonzani/patch-match-stereo (原始算法代码)
	文献[42]	能够降低在纹理性弱的区域错误匹配的概率	
双通道法	文献[43]	能够解释场景元素之间的高阶语义关系,具有定量与定性的优势	
DeepMVS	文献[15]	在面对纹理性弱的区域和薄结构时更有优势	https://github.com/phuang17/DeepMVS
MVSNet	文献[16]	可以适用于任意张视图输入,其运行速度更快,精度高,不需要任何的微调,具有良好的泛化性	https://github.com/YoYo000/MVSNet
PointMVSNet	文献[17]	直接将场景作为点云处理,计算量大幅度减少,保留空间连续性,精度、速度、灵活性提高显著,重建结果更为高效、精确	https://github.com/callmeray/PointMVSNet



图8 RGB图像
Fig. 8 Original RGB image



图9 稀疏点云重建效果
Fig. 9 Rendering reconstructed by SFM



图10 稠密点云正面重建效果
Fig. 10 Front view reconstructed by MVS

3 评价三维重建的指标

在完成三维点云重建后,需要评价三维重建结果的优劣性.三维重建的评价方法大致分为有真实模型和无真实模型两类.无真实模型的评价是通过对比输入数据与重建之后的数据一致性程度来实施的,其方法有重建误差法、积分误差法等.重建误差法

通过计算输入点到重建表面点的平均欧式距离来评



图 11 稠密点云侧面重建效果

Fig. 11 Side view reconstructed by MVS

价,通常距离越小,误差越小,重建效果越好.重建误差法受噪声的干扰较大,当出现几个偏移较大的重建点时会影响到最后的整体结果,所以最后得到的结果较为片面.积分误差法是由 Zhou 等^[45]提出来的,通过计算输入点到重建表面距离最近点距离的平均值进行评价,通常误差越小,重建效果越好.积分误差法虽然采用取平均的方法来避免噪声点的干扰,但是却没有考虑到其他因素.总体来说无真实模型的评价手段单一,没有考虑局部特征.Song 等^[46]提出了 3DGIC 算法,该算法无需提供真实模型,并且该算法包括了全局一致性和局部准确性,该算法的主要思想是计算局部表面描述符联合概率的累积分布,其中累积分布表示全局统计量,而局部表面描述符用于评估局部精度,将全局和局部的关联进行合并,从而得到一个更紧凑的评估.

有真实模型的评价是通过重建之后的模型与真实模型进行对比来实施的,主要方法有形状误差法、分解法、相位矩不变量法、多指标评价法等.Seitz 等^[47]提出了一种以准确性和完整性为指标的三维重建结果的量化评价方法,该方法需要借助设备获得物体的高精度三维数据,不利于大规模场景重建的评估,成本高、效率低.为了解决上述问题,公明等^[48]提出了一种基于图像相位信息不变性的评价方法,该方法首先要获得物体在某一视点下的二维投影图及该物体在相近视点下的图像,然后对两张图像进行边缘提取,计算两张边缘图像的相位矩不变量,从而构造特征向量,最后通过比较特征向量的距离进行三维重建的评估.但是该方法计算量大,不稳定.张雯等^[49]提出了一种评价三维模型几何质量的方法,将三维模型分解到基元点、基元线和基元面,以局部精度作为评价指标,通过计算点的距离偏移,线的长度插值、距离偏移、角度偏移,面的边数差、面积差值及距离偏移、角度偏移等综合数据,来评价三维重建的指标,但是该方法容易出现点的数

据缺失情况.

上述的三维重建评价方法中主要有三个缺陷:

- 1) 部分算法只能分析真实模型的场景;
- 2) 部分算法只能测量重建的精度,但是不能测量完整性,导致评价结果存在偏差;
- 3) 大多算法只考虑到几何精度,并没有将纹理模型的视觉精度添加到评价指标中,在某些需要视觉精度的场景中应用意义不大.

Waechter 等^[50]提出了一种以视觉质量为评定指标的方法,该方法将获取的图像分为训练集和评价集,使用基于图像建模的流程重构训练集,利用评价集的摄像机参数渲染视图,最后通过图像的差异性度量将这些效果图与评价集进行比较.该方法是一种灵活的评价方式,不需要重建物体的真实数据,能够评价重建物体表面的纹理质量,并且能够显示和定位重建不精确的地方.

对于三维重建结果评价的研究仍在发展的过程中,还没有十分完善.由上述文献可以看出,对三维重建效果的研究不能只针对一个因素进行评价,应该综合多个因素从局部到全局进行评价.

4 总结与展望

4.1 总结

在游戏开发、工业设计、航天航海等诸多领域,三维重建发挥着不可或缺的作用.本文基于图像的三维重建方法展开了综述,重点介绍了特征点检测与匹配、SFM 重构、MVS 的研究现状,还介绍了三维重建的评价指标.

基于图像的三维重建不需要精密的测量仪器,成本低,同时该方法计算出了相机的内外参数,使得重建效果更为精确,具有快速、简便、逼真的优点,能较好地实现现实事物的虚拟化,但是其仍存在一些缺点:

- 1) 在对物体的细节特征重建上不是很理想;
- 2) 不能重构出实时变化的场景,实时性不强,不能应用于实时地图重建的操作中;
- 3) 基于图像的三维重建将场景中的物体都变成了二维对象,用户很难获取到需要的信息;
- 4) 基于图像的三维重建对照相机与摄影设备也有一定要求,大量的图像文件需要足够的存储空间.

基于点云的三维重建也是一大研究热点.基于点云的三维重建需要精密的仪器,成本高,却很难达到对大规模场景进行重建的要求,但是其计算量小、

精度高.基于图像与基于点云的三维重建都有各自的优点以及局限性,在今后的三维重建过程中,需要根据重建场景的规模以及自身拥有的设备等,选择合适的手段进行重建,重建得到的图像会具备更好的性能.

4.2 展望

1) 基于图像的三维重建在特征检测与匹配过程中,耗时长、精度低、无法识别物体表面的纹理信息,因此需要对特征点检测与匹配该部分的内容进行更深层次的研究,从而促使基于图像的三维重建发展更加成熟.

2) 对重建工作进行评价的算法并不是很多,发展较为缓慢,整体研究不够成熟,未来对于评价算法的研究应该更加注重整体与局部的结合,以及视觉精度与几何精度的结合,以更全面地评价三维重建的效果.

3) 在工业应用中,由于各种外界因素的干扰,导致原始的三维重建方法并不适用于所有的环境中,因此研究出能够面对各种不可控因素的三维重建方法成为一件迫在眉睫的事情.

4) 由于基于图像的三维重建计算量大,很难完成实时三维重建,希望在日后的研究中能够缩减计算量,将基于图像的三维重建运用到实时重建上.

参考文献

References

- [1] Mohd Isa S N, Abdul Shukur S A, Rahim N A, et al. A review of data structure and filtering in handling 3D big point cloud data for building preservation [C] // 2018 IEEE Conference on Systems, Process and Control (IC-SPC), 2018: 141-146
- [2] He H, Wang H L, Sun L. Research on 3D point-cloud registration technology based on Kinect V2 sensor [C] // 2018 Chinese Control and Decision Conference (CCDC), 2018: 1264-1268
- [3] 孙金虎. 点云模型分割与融合关键技术研究 [D]. 南京: 南京航空航天大学, 2013
SUN Jinhu. Research on key technologies of point cloud segmentation and fusion [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2013
- [4] 王向军, 邢峰, 刘峰. Delaunay 三角剖分和仿射约束的特征相同多物体同名点立体匹配 [J]. 光学学报, 2016, 36(11): 1115004
WANG Xiangjun, XING Feng, LIU Feng. Stereo matching of objects with same features based on delaunay triangulation and affine constraint [J]. Acta Optica Sinica, 2016, 36(11): 1115004
- [5] 孙庆科, 何云涛, 陈瑞强, 等. 基于结构光的三维点云重建方法研究 [J]. 计算机应用研究, 2016, 33(9): 2873-2877
SUN Qingke, HE Yuntao, CHEN Ruiqiang, et al. Research on 3d point cloud reconstruction based on structured light [J]. Computer Application Research, 2016, 33(9): 2873-2877
- [6] Debevec P E, Taylor C J, Malik J. Modeling and rendering architecture from photographs: a hybrid geometry-and image-based approach [C] // Proceedings of the 23rd annual Conference on Computer Graphics and Interactive Techniques, 1996: 11-20
- [7] Chen X, Wu Q F, Wang S Z. Research on 3D reconstruction based on multiple views [C] // 2018 13th International Conference on Computer Science & Education (ICCSE), 2018: 1-5
- [8] Lowe D G. Object recognition from local scale-invariant features [C] // Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999. DOI: 10.1109/ICCV.1999.790410
- [9] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110
- [10] Cui H N, Gao X, Shen S H, et al. HSfM: hybrid structure-from-motion [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1212-1221
- [11] Cui H N, Shen S H, Gao W, et al. Progressive large-scale structure-from-motion with orthogonal MSTs [C] // 2018 International Conference on 3D Vision (3DV), 2018, DOI: 10.1109/3DV.2018.00020
- [12] Zhu S Y, Zhang R Z, Zhou L, et al. Very large-scale global SfM by distributed motion averaging [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4568-4577
- [13] Gallup D, Frahm J M, Mordohai P, et al. Real-time plane-sweeping stereo with multiple sweeping directions [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2007. DOI: 10.1109/CVPR.2007.383245
- [14] Bleyer M, Rhemann C, Rother C. PatchMatch stereo-stereo matching with slanted support windows [C] // British Machine Vision Conference, 2011
- [15] Huang P H, Matzen K, Kopf J, et al. DeepMVS: learning multi-view stereopsis [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2821-2830
- [16] Yao Y, Luo Z X, Li S W, et al. MVSNet: depth inference for unstructured multi-view stereo [C] // European Conference on Computer Vision, 2018: 767-783
- [17] Chen R, Han S F, Xu J, et al. Point-based multi-view stereo network [C] // IEEE International Conference on Computer Vision, 2019: 1538-1547
- [18] Bay H. SURF: Speeded Up Robust Features [C] // European Conference on Computer Vision, 2006, 110(3): 404-417
- [19] Rosten E, Drummond T. Machine learning for high-speed corner detection [C] // European Conference on Computer Vision, 2006: 430-443
- [20] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C] // International Conference on Computer Vision, 2011, DOI: 10.1109/IC-

- CV.2011.6126544
- [21] Smith S, Brady J. SUSAN: a new approach to low level image processing [J]. International Journal of Computer Vision, 1997, 23(1): 45-78
- [22] 崔振兴, 曾威, 杨明强, 等. 一种改进的 SURF 快速匹配算法 [J]. 江苏师范大学学报(自然科学版), 2014, 32(3): 41-46
CUI Zhenxing, ZENG Wei, YANG Mingqiang, et al. An improved algorithm of image registration based on SURF [J]. Journal of Jiangsu Normal University (Natural Science Edition), 2014, 32(3): 41-46
- [23] 李红波, 赵永耀, 吴渝, 等. 一种基于距离约束的改进 SURF 算法 [J]. 系统仿真学报, 2014, 26(12): 2944-2949, 2956
LI Hongbo, ZHAO Yongyao, WU Yu, et al. Improved SURF algorithm based on distance constraint [J]. Journal of System Simulation, 2014, 26(12): 2944-2949, 2956
- [24] Xu H, Jin Y J, Wan W G. An efficient 3D reconstruction system for Chinese ancient architectures [C] // 2018 International Conference on Audio, Language and Image Processing (ICALIP), 2018: 221-225
- [25] Sergeeva A D, Sablina V A. Using structure from motion for monument 3D reconstruction from images with heterogeneous background [C] // 2018 7th Mediterranean Conference on Embedded Computing (MECO), 2018: 1-4
- [26] Gao X S, Hou X R, Tang J L, et al. Complete solution classification for the perspective-three-point problem [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8): 930-943
- [27] Lepetit V, Moreno-Noguer F, Fua P. EPnP: an accurate $O(n)$ solution to the PnP problem [J]. International Journal of Computer Vision, 2009, 81(2): 155-166
- [28] Chen P. An iterative camera pose estimation algorithm based on EPnP [M] // Lecture Notes in Electrical Engineering. Singapore: Springer Singapore, 2017: 415-422
- [29] Hartley R, Zisserman A. Multiple view geometry in computer vision [M]. Cambridge: Cambridge University Press, 2004
- [30] Penate-Sanchez A, Andrade-Cetto J, Moreno-Noguer F. Exhaustive linearization for robust camera pose and focal length estimation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(10): 2387-2400
- [31] Furukawa Y, Ponce J. Accurate camera calibration from multi-view stereo and bundle adjustment [J]. International Journal of Computer Vision, 2009, 84(3): 257-268
- [32] Yin Y K, Peng X, Li A M, et al. Calibration of fringe projection profilometry with bundle adjustment strategy [J]. Optics Letters, 2012, 37(4): 542
- [33] 段建伟. 基于图优化的增量式单目 SFM 三维重建方法 [J]. 江苏科技信息, 2019(6): 37-40
DUAN Jianwei. Incremental monocular SFM 3D reconstruction method based on graph optimization [J]. Jiangsu Science & Technology Information, 2019(6): 37-40
- [34] 刘彬, 刘学军, 张华, 等. 线性的增量式三维稀疏重建系统设计 [J]. 电光与控制, 2019, 26(7): 100-104, 109
LIU Bin, LIU Xuejun, ZHANG Hua, et al. Linear incremental 3D sparse reconstruction system design [J]. Electronics Optics & Control, 2019, 26(7): 100-104, 109
- [35] 刘小群. 基于双目立体视觉的立体匹配算法研究 [D]. 长沙: 中南大学, 2011
LIU Xiaoqun. Stereo matching algorithm based on binocular stereo vision [D]. Changsha: Central South University, 2011
- [36] 张一飞, 李新福, 田学东. 融合边缘特征的 SAD 立体匹配算法 [J]. 计算机工程, 2020, 46(4): 236-240, 246
ZHANG Yifei, LI Xinfu, TIAN Xuedong. SAD stereo matching algorithm combining edge features [J]. Computer Engineering, 2020, 46(4): 236-240, 246
- [37] Yang R G, Pollefeys M. Multi-resolution real-time stereo on commodity graphics hardware [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. DOI: 10.1109/CVPR.2003.1211356
- [38] Shu B, Qiu X J, Wang Z Q. Image based modeling via plane sweep based surface growing [C] // IEEE International Conference on Shape Modeling and Applications, 2009. DOI: 10.1109/SMI.2009.5170158
- [39] 王跃嵩. 多视图立体匹配三维重建方法 [D]. 武汉: 华中科技大学, 2015
WANG Yuesong. 3D reconstruction method of multi view stereo matching [D]. Wuhan: Huazhong University of Science and Technology, 2015
- [40] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing [C] // ACM Transactions on Graphics, 2009, 28(3): 24
- [41] Besse F, Rother C, Fitzgibbon A, et al. PMBP: PatchMatch belief propagation for correspondence field estimation [J]. International Journal of Computer Vision, 2014, 110(1): 2-13
- [42] Li L C, Zhang S L, Yu X, et al. PMSC: PatchMatch-based superpixel cut for accurate stereo matching [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(3): 679-692
- [43] Wang K, Dunn E, Tighe J, et al. Combining semantic scene priors and haze removal for single image depth estimation [C] // IEEE Winter Conference on Applications of Computer Vision, 2014: 800-807
- [44] Yao Y, Luo Z X, Li S W, et al. Recurrent MVSNet for high-resolution multi-view stereo depth inference [C] // The IEEE Conference on Computer Vision and Pattern Recognition, 2019: 5525-5534
- [45] Zhou H, Liu Y H. Accurate integration of multi-view range images using k-means clustering [J]. Pattern Recognition, 2008, 41(1): 152-175
- [46] Song R, Liu Y H, Zhao Y T, et al. An evaluation method for multiview surface reconstruction algorithms [C] // Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission, 2012: 387-394
- [47] Seitz S M, Curless B, Diebel J, et al. A comparison and evaluation of multi-view stereo reconstruction algorithms [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006. DOI: 10.1109/CVPR.2006.19
- [48] 公明, 葛娟, 曹伟国, 等. 相位矩不变量用于评价基于图像的三维重建结果 [J]. 计算机辅助设计与图形学

学报, 2013, 25(9): 1384-1391

GONG Ming, GE Juan, CAO Weiguo, et al. Evaluation of image-based 3D reconstruction using phase moment invariants[J]. Journal of Computer-Aided Design & Computer Graphics, 2013, 25(9): 1384-1391

[49] 张雯, 程亮, 张群, 等. 三维模型几何质量评价方法[J]. 测绘通报, 2014(7): 44-47

ZHANG Wen, CHENG Liang, ZHANG Qun, et al. An

evaluation method for the geometry quality of three-dimensional models [J]. Bulletin of Surveying and Mapping, 2014(7): 44-47

[50] Waechter M, Beljan M, Fuhrmann S, et al. Virtual rephotography: novel view prediction error for 3D reconstruction[J]. ACM Transactions on Graphics, 2017, 36(4): Article No.45

Review of 3D reconstruction algorithms

ZHANG Yanwen^{1,2} HU Kai^{1,2} WANG Pengsheng^{1,2}

1 School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044

2 Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAEET),
Nanjing University of Information Science & Technology, Nanjing 210044

Abstract Three-dimensional (3D) reconstruction technology can be applied in many fields such as 3D CT images in medicine, path planning in robots, and display of relics in archaeological work. The completeness and accuracy of 3D reconstruction are directly related to steps of Structure from Motion (SfM) and Multi-View System (MVS). In this paper, we introduce these two steps in detail. In the reconstruction of sparse point cloud, feature point detection and matching and SfM reconstruction methods are introduced and elaborated. We summarize the researches on feature point detection and matching in recent years, and classify the reconstruction methods of SfM into global and incremental ones, and introduce them in detail. In the dense point cloud reconstruction, the depth map estimation is introduced and elaborated, in which the traditional geometric calculation method, hybrid geometry-and deep learning-based approach, and deep learning based method are summarized. Meanwhile, the evaluation index of 3D reconstruction is introduced. Finally, the image based 3D reconstruction is summarized, and the problems facing the 3D reconstruction and possible future research directions are listed.

Key words 3D reconstruction; images; structure from motion; multi-view system; evaluation index