

基于图像的三维重建典型技术综述

孙海迅, 罗健欣, 张艳艳, 郑义桀, 潘志松

(中国人民解放军陆军工程大学 指挥控制工程学院, 江苏 南京 210001)

摘要: 三维重建是计算视觉的重要方面, 近年来快速发展并在自动驾驶、虚拟现实等方面得以广泛应用。其中, 基于图像的三维重建方法发展较为成熟, 算法流程清晰, 备受研究者关注。主要技术包括: 特征点检测与匹配、稀疏点云重建、稠密点云重建、三维表面重建、纹理图像创建等。梳理图像三维重建完整流程, 选取重点环节介绍典型技术, 并进行分析比较, 最后对基于图像的三维重建技术进行总结和展望。

关键词: 图像三维重建; 稀疏点云重建; 稠密点云重建; 典型技术

DOI: 10.11907/rjdk.221275

开放科学(资源服务)标识码(OSID):



中图分类号: TP391.9

文献标识码: A

文章编号: 1672-7800(2023)009-0009-16

A Review of Typical Techniques of Image-based 3D Reconstruction

SUN Haixun, LUO Jianxin, ZHANG Yanyan, ZHENG Yijie, PAN Zhisong

(Command and Control Engineering College, Army Engineering University of PLA, Nanjing 210001, China)

Abstract: 3D reconstruction is an important aspect of computer vision. It has developed rapidly in recent years, and has been widely used in automatic driving, virtual reality and other fields. Among them, the image-based 3D reconstruction method is relatively mature, the algorithm process is clear. In this review, we do not only show the complete process from feature point matching, sparse point cloud reconstruction, dense point cloud reconstruction, 3D surface reconstruction and texture image creation, but also introduce typical technologies in depth, and make analysis and comparison. Finally, the the 3d reconstruction technology is summarized and prospected.

Key Words: image-based 3D reconstruction; sparse point cloud reconstruction; dense point cloud reconstruction; typical technology

0 引言

三维重建是指在计算机中建立现实物体或场景三维模型的关键技术。随着计算机视觉的发展, 三维重建在医学影像判读、文物修复、自动驾驶、虚拟现实等领域有着广泛应用。按照重建方法, 可将其分为传统方法和深度学习方法两大类。传统方法主要是对于在不同视角获取的物体照片, 利用相关算法, 恢复其三维结构。深度学习方法是通过训练神经网络预测物体的三维模型^[1]。深度学习方法对算力要求较高, 而传统基于图像的三维重建价格低廉且算法相对成熟, 因而应用更加广泛。

基于图像的三维重建主要包括以下流程: 图像特征点检测与匹配、稀疏点云重建、稠密点云重建、三维表面重建、表面纹理贴图。本文针对以上流程中的关键环节, 选取几种代表性方法进行综述。

1 特征点检测与匹配

特征点是图像中亮度变化剧烈的点或图像边缘曲线上曲率的极大值点, 它同周围的邻近点有着明显差异^[2], 目前已有多种特征点检测方法^[3]。在三维重建过程中, 可以通过对不同视角下图像特征点的检测与匹配, 求解相机姿态, 进而获取三维点云。本文重点介绍两种经典方法

收稿日期: 2022-03-15

基金项目: 国家自然科学基金项目(62076251)

作者简介: 孙海迅(1990-), 男, 中国人民解放军陆军工程大学指挥控制工程学院硕士研究生, 研究方向为计算机视觉、三维重建; 罗健欣(1984-), 男, 博士, CCF会员, 中国人民解放军陆军工程大学指挥控制工程学院讲师, 研究方向为计算机视觉、计算机图形学、网络与多媒体通信; 张艳艳(1986-), 女, 硕士, 中国人民解放军陆军工程大学指挥控制工程学院讲师, 研究方向为机器学习、三维重建; 郑义桀(1991-), 男, 中国人民解放军陆军工程大学指挥控制工程学院硕士研究生, 研究方向为计算机视觉、SLAM、三维重建; 潘志松(1973-), 男, 中国人民解放军陆军工程大学指挥控制工程学院教授、博士生导师, 研究方向为模式识别、机器学习、神经网络。本文通讯作者: 潘志松。

——Harris角点检测^[4]和SIFT特征点检测^[5]。

1.1 Harris角点检测

角点指图像上两个方向上边缘线的接合点。在一幅图像上取一小块特征区域,如果区域的灰度沿各方向都有较大变化,则认为窗口内存在角点。

设 (x, y) 为图像上某处像素坐标, $I(x, y)$ 代表该点处对应的光度, (u, v) 代表位移, $I(x+u, y+v)$ 代表位移后的光度, $w(x, y)$ 代表权重函数,本文选择高斯函数,则可用 $E(u, v)$ 度量特征区域位移后与位移前的光度差异,以此判断区域内是否包含角点。

$$E(u, v) = \sum_{x, y \in \Omega} w(x, y) [I(x+u, y+v) - I(x, y)]^2 \quad (1)$$

其中,对 $I(x+u, y+v)$ 进行一阶泰勒展开,即:

$$I(x+u, y+v) = I(x, y) + uI_x + vI_y \quad (2)$$

其中, I_x, I_y 表示图像在 (x, y) 坐标点处的亮度分别对 x, y 的偏导数,因而可将上式重写为:

$$E(u, v) \approx [u, v] M \begin{bmatrix} u \\ v \end{bmatrix}, M = \sum_{x, y \in \Omega} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (3)$$

通过对 M 矩阵的特征值进行分析即可判断 Ω 区域是否包含角点,当 M 的两个特征值 λ_1, λ_2 都比较大时,则认为 Ω 区域包含角点。同时,文献[4]给出了如下判别方法,当 R 大于某阈值,认为 Ω 区域存在角点。

$$R = \det M - k(\text{trace}M)^2 \quad (4)$$

1.2 SIFT特征点检测

文献[5]提出了一种对旋转、缩放、明暗保持不变的特征点检测方法,通过高斯差分算子(Difference of Gaussian, 简称DOG算子)寻找特征点。高斯空间与高斯差分空间如图1所示。

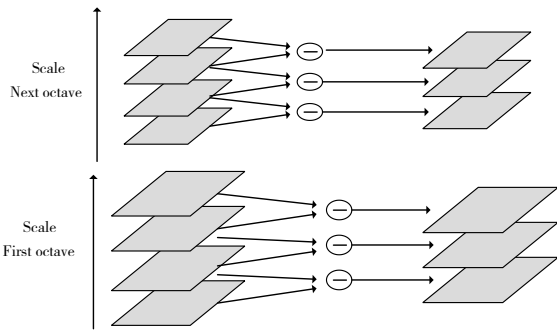


Fig. 1 Gaussian space and Gaussian difference space

图1 高斯空间与高斯差分空间

首先对原始图像进行降采样形成不同的组(octave),在每组内对图像用不同的高斯核函数(标准差分别为 $\sigma, k\sigma, k^2\sigma, \dots$)做卷积形成多张图像。计算方法为:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (5)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2 + y^2)/2\sigma^2} \quad (6)$$

其中, $I(x, y)$ 表示原图像, $G(x, y, \sigma)$ 表示标准差为 σ 的核函数, $L(x, y, \sigma)$ 表示卷积之后生成的一幅高斯图像,以此构建高斯图像空间(Gaussian Space)。

同一组内相邻两幅图像相减得到高斯差分图像空间(DoG Space),计算方法为:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \quad (7)$$

$D(x, y, \sigma)$ 表示生成的一幅高斯差分图像,并且每一组内有效差分 S 与高斯图像空间的层数 N 满足如下关系:

$$N = S + 3 \quad (8)$$

在高斯差分图像空间求极值得到图像特征点,即对每个点与它所在图像与上下相邻两层图像周围26个像素点的亮度进行比较,如果大于或者小于这26个点,则作为极值点。

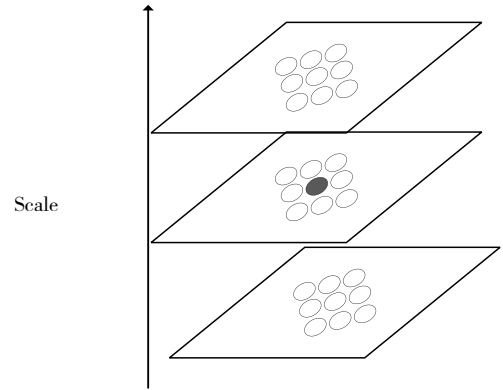


Fig. 2 Extremum point of the image

图2 图像极值点

进一步地,通过在亚像素位置上寻找极值点的更精确位置能够从本质上改善特征点的匹配稳定性^[5],图像极值点如图2所示。于是,在 $x_0 = (x_0, y_0, \sigma_0)^T$ 处对函数 $D(x, y, \sigma)$ 进行泰勒展开得:

$$D(x) = D(x_0) + \nabla D(x_0)^T (x - x_0) + \frac{1}{2} (x - x_0)^T \nabla^2 D(x_0) (x - x_0) \quad (9)$$

令 $\delta x = x - x_0$,将 $D(x)$ 对其求偏导使之之为0,得:

$$\delta x = -\nabla^2 D(x_0)^{-1} \nabla D^T(x_0) \quad (10)$$

于是精确的极值位置为 $x = x_0 + \delta x$,对应的函数值:

$$D(x) = D(x_0) + \frac{1}{2} \nabla D(x_0)^T (x - x_0) \quad (11)$$

其中,偏导数矩阵 $\nabla D^T(x_0), \nabla^2 D(x_0)$ 是用像素点附近点的差分形式近似表示^[5]。

此外,类似于Harris角点检测的方法,对图像平面上 x 处的二阶Hessian矩阵的特征值施加一定约束,能剔除某些只在一个方向上与其他点有明显差异的点,使得检测到的特征点更加稳定。约束条件为:

$$\frac{\text{trace}(H)^2}{\text{Det}(H)} < \frac{(r+1)^2}{r}, r = 10 \quad (12)$$

图像的主方向通过统计梯度直方图的方法确定,将图像的 x 轴旋转到主方向上,算法才具有旋转不变性。

SIFT特征描述子(128维)同样是统计局部区域梯度信息而生成,通过最近邻搜索方式进行匹配,并将最近邻与次近邻的距离之比限定在较小范围内,以此得到较为精确

的匹配对。

SIFT 特征点检测比 Harris 角点检测更鲁棒, 对旋转、尺度缩放、亮度变化能保持不变, 但实时性不高、运算速度较慢, 因而出现了诸如 SURF^[6-7] 等方法的改进。

2 稀疏点云重建

稀疏点云重建是根据图像上特征点的匹配关系, 通过运动恢复结构(SFM)生成稀疏点云的方法。运动恢复结构是指通过图像间的对应关系, 重构相机外部参数(位置和姿态)和内部参数(焦距和径向畸变)。在实现 SFM 的过程中, 通常包含三角测量, 以此求解三维点坐标, 得到稀疏点云。

2.1 基本概念

2.1.1 针孔相机模型

针孔相机模型的坐标系包括: 世界坐标系、相机坐标系、归一化像平面坐标系和物理像平面坐标系。世界坐标系是物体所在现实世界的坐标系, 实际计算时常用相机的初始位置建立世界坐标系; 相机坐标系是在以相机中心为坐标原点建立的坐标系, 其 Z 轴通常指向要拍摄的物体。世界坐标系到相机坐标系的坐标变换如图 3 所示, 相机坐标系下的坐标变换如图 4 所示。

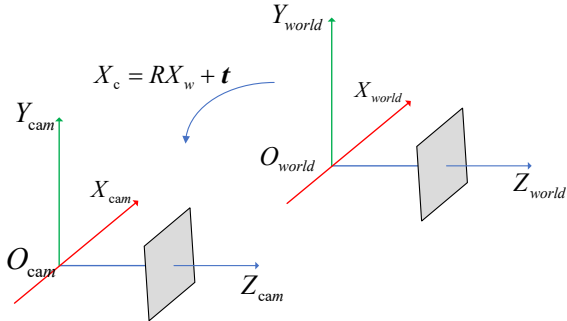


Fig. 3 Coordinate transformation from world coordinate system to camera coordinate system

图 3 世界坐标系到相机坐标系的坐标变换

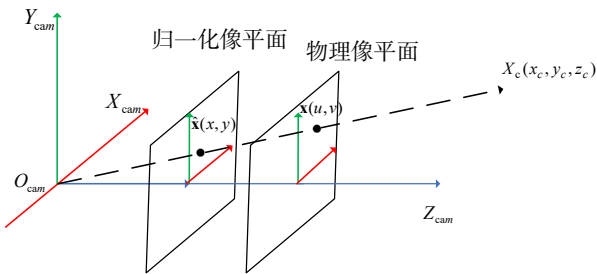


Fig. 4 Coordinate transformation in camera coordinate system

图 4 相机坐标系下的坐标变换

设世界坐标系下某一点 $X_w(x_w, y_w, z_w)$, 在相机坐标系下坐标为 $X_c(x_c, y_c, z_c)$, 二者可通过旋转平移变换得到, 于是满足以下关系:

$$X_c = RX_w + t \quad (13)$$

写成齐次形式:

$$\begin{bmatrix} X_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ 1 \end{bmatrix} \quad (14)$$

归一化像平面是人为设置的虚拟平面, 将其 z 坐标(距离相机光心距离)设为 1。设点 X_c 的归一化坐标为 (x, y) , 则:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x_c \\ z_c \\ y_c \\ z_c \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad (15)$$

物理像平面平行于归一化平面, 指相机 CCD 阵列所在的平面。物理像平面坐标也即图像的像素坐标, 归一化坐标乘上焦距 f 和分辨率 α (可以合并写为 f_α), 再进行坐标平移后(因为物理像坐标通常以图像左上方为原点)即可得到物理像平面坐标。

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_\alpha x + u_0 \\ f_\beta y + v_0 \\ 1 \end{bmatrix} \Rightarrow \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f_\alpha & 0 & u_0 \\ 0 & f_\beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad (16)$$

因此, 从空间点的世界坐标 X_w 到图像坐标的变换关系为:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f_\alpha & 0 & u_0 \\ 0 & f_\beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f_\alpha & 0 & u_0 \\ 0 & f_\beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad t] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = \frac{1}{z_c} K [R \quad t] \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (17)$$

其中, K 为内参矩阵, $[R \quad t]$ 为外参矩阵。

2.1.2 径向畸变矫正

相机的针孔模型与真实的相机成像不尽相同, 真实的相机会由于透镜形状引起图像畸变——径向畸变。因此, 要从二维图像中准确提取度量信息, 必须进行摄像机标定。Zhang^[8] 提出一种利用棋盘格标定相机的方法, 标定过程仅需打印棋盘格, 并从不同方向拍摄图片, 不仅灵活方便, 而且鲁棒性好, 如图 5 所示。

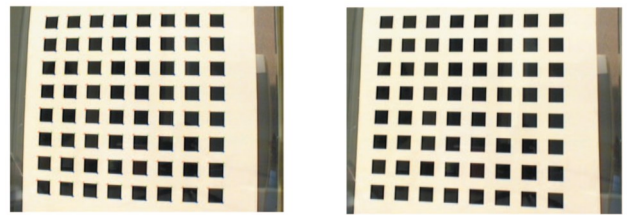


Fig. 5 Comparison before and after correction of radial distortion

图 5 径向畸变矫正前后对比

该方法假设畸变后的归一化坐标与畸变之前坐标存在如下关系:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4) \begin{bmatrix} x \\ y \end{bmatrix}, r^2 = x^2 + y^2 \quad (18)$$

其中, k_1, k_2 为径向畸变系数, r 为坐标点到归一化平面中心点的半径。畸变后的图像坐标为:

$$\begin{bmatrix} \tilde{u} \\ \tilde{v} \end{bmatrix} = f \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \quad (19)$$

在理想情况下:

$$\begin{bmatrix} u \\ v \end{bmatrix} = f \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} \quad (20)$$

联立以上等式, 可得:

$$\begin{bmatrix} (u - u_0)r^2 & (u - u_0)r^4 \\ (v - v_0)r^2 & (v - v_0)r^4 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} \tilde{u} - u \\ \tilde{v} - v \end{bmatrix} \quad (21)$$

其中, \tilde{u}, \tilde{v} 是通过观察棋盘格上已知的某点在未矫正图像上的位置而得到, u, v 是理想情况下利用前面的投影关系计算得到, r 是该点到棋盘中心的距离。在棋盘格上选定多点, 即可利用最小二乘法解方程得到 k_1, k_2 。之后, 在矫正后的图像上, 即可利用如下关系得到点 (u, v) 对应的 \tilde{u}, \tilde{v} , 并将矫正前坐标 (\tilde{u}, \tilde{v}) 处的颜色值作为矫正后图像上 (u, v) 坐标的颜色值, 生成矫正后的图像。

2.2 相机姿态恢复

在对极几何^[9]中, 常通过图像坐标点的对应关系, 求解基本矩阵或单应性矩阵进而求解相机姿态。

2.2.1 通过基本矩阵与本质矩阵求解相机姿态

如图6所示, X 为一空间三维点, O_1, O_2 为两个相机中心, 则 XO_1O_2 所在的平面称为极平面, π_1, π_2 为两个相机的物理像平面, x_1, x_2 分别为 X 在两个物理像平面的像素坐标。设 x_1', x_2' 分别为 x_1, x_2 对应的归一化平面坐标, $[R, t]$ 为 O_2 坐标系相对于 O_1 坐标系的坐标变换矩阵, 两个相机的内参矩阵分别为 K_1, K_2 , 则有如下约束关系:

$$x_2^T F x_1 = 0, x_2^T E x_1 = 0 \quad (22)$$

其中, $E = [t]_{\times} R$ 被称为本质矩阵, $F = K_2^{-T} E K_1^{-1}$ 被称为基本矩阵。

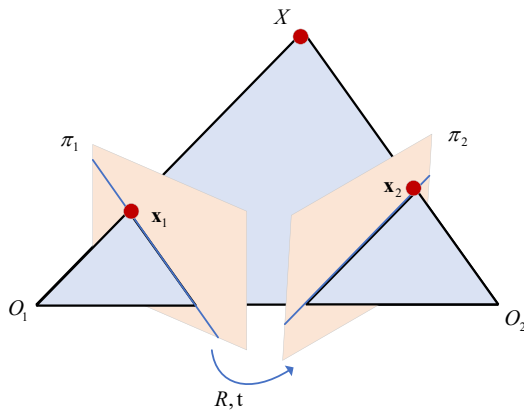


Fig. 6 Diagram of epipolar geometry

图6 对极几何示意图

在图像特征点检测与匹配的基础上, 利用以上约束关系, 即可求得基本矩阵和本质矩阵, 进而得到相机的外参数(相机姿态)。文献[10]给出了利用8对在图像上匹配的特征点计算基本矩阵 F 的方法, 称为8点法。具体为:

设 $x_1 = (u_1, v_1, 1)^T, x_2 = (u_2, v_2, 1)^T$ 为一对匹配的特征点, 根据约束 $x_2^T F x_1 = 0$, 即:

$$(u_2, v_2, 1) \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} = 0 \quad (23)$$

若令 $f = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^T$, 则有:

$$[u_1 u_2, u_2 v_1, u_2, u_1 v_2, v_1 v_2, v_2, u_1, v_1, 1] f = 0 \quad (24)$$

也即每一对匹配点能对 F 矩阵提供一个约束, 而若令 F 的任一元素归一化(所有元素同时除以某一元素), 并不影响 $x_2^T F x_1 = 0$ 的成立, 即 F 有8个自由度, 因此至少需要8对匹配的特征点即可求得 F , 当匹配的特征点大于8时, 也可得利用最小二乘法求解。

为增加8点法的鲁棒性, 文献[11]采用了将图像坐标归一化的方法, 也称为归一化8点法。文献[12]提出要对 F 矩阵增加奇异值约束。文献[13]提出了一种根据奇异值约束对矩阵 F 进行重构的方法, 即在利用8点法得到 F 的基础上, 将 F 进行SVD分解:

$$F = USV^T, S = \text{diag}(\sigma_1, \sigma_2, \sigma_3) \quad (25)$$

利用奇异值约束将基本矩阵重构为: $F \cdot = \text{Udiag}(\sigma_1, \sigma_2, \sigma_3)V^T$, 这样可以使得 $F - F \cdot$ 的Frobenius范数最小。

在实际应用中, 特征点可能因噪声产生误匹配而使得基本矩阵的估计不准确。为解决这一问题, 文献[14]采用随机抽样一致性(RANSAC^[15])方法估计基本矩阵, 具体为:

(1) 计算采样次数。

$$M = \frac{\log(1 - z)}{\log(1 - p^k)} \quad (26)$$

其中, z 代表期望采样达到的成功率, M 代表为达到这一成功率至少需要的采样次数, k 代表每次采样求解基本矩阵所需要的点数, p 代表所有样本点中内点(匹配准确的特征点)的概率。

(2) 对匹配的特征点集中随机采样8对, 利用8点法估计初始矩阵 F 。

(3) 用Sampson距离^[16]度量点集中所有点, 将点划分为内点和外点(误匹配的特征点), 并记录当前计算所得 F 矩阵对应的内点个数。一对特征点的Sampson距离可表示为:

$$d(x_1, x_2) = \frac{(x_2^T F x_1)^2}{(F x_1)_x^2 + (F x_1)_y^2 + (x_2^T F)_x^2 + (x_2^T F)_y^2} \quad (27)$$

其中, x_1, x_2 表示不同图像上的一对特征点, F 为步骤(2)计算得到的矩阵, 下标 x, y 分别表示某一向量的 x 分量和 y 分量。

(4) 重复步骤(2)(3) M 次, 选择内点个数最多的 F 矩阵和相应的内点。

(5) 对所有内点重新执行步骤(2)得到基本矩阵 F' 。

相机姿态的恢复需要求解本质矩阵。本质矩阵 E 是

一个秩为 2 的矩阵, 具有 5 个自由度, 有两个相等的奇异值和一个 0 奇异值^[17]。与求解基本矩阵 F 所用的 8 点法类似, 本质矩阵的求解可以采用 5 点法^[18], 也可在求得基本矩阵 F 的基础上, 利用以下关系求本质矩阵初始解:

$$\hat{E} = K_2^T F K_1 \quad (28)$$

同样地, 要对 E 施加奇异值约束并重构^[9]:

$$\hat{E} = U \text{diag}(\sigma_1, \sigma_2, 0) V^T \quad (29)$$

$$E = U \text{diag}\left(\frac{\sigma_1 + \sigma_2}{2}, \frac{\sigma_1 + \sigma_2}{2}, 0\right) V^T \quad (30)$$

在文献[9]中证明, SVD 分解不唯一, 如果可以分解为式(30), 则也一定可以分解为: $E = U \text{diag}(1, 1, 0) V^T$ 。如果可以将本质矩阵进行 SVD 分解, 则 $E = [t]_x R$ 有如下 4 种可能情况:

$$[R|t] = [UWV^T|u_3]; [UWV^T - u_3]; [UW^T V^T|u_3]; [UW^T V^T - u_3] \quad (31)$$

其中, u_3 代表 U 矩阵的第 3 列, $W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

从图 7 可以看出, 4 种情况中只有一种情况是空间点 X 同时位于两个相机的镜头前方, 因而只要用一个点做测试, 看它是否位于两个相机的前方, 即可选择出正确的相机姿态。

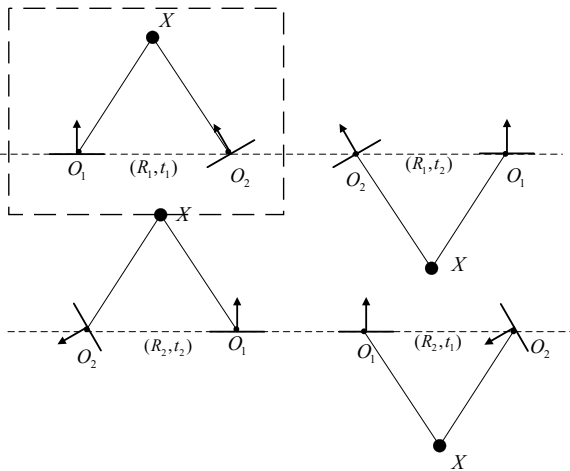


Fig. 7 Choice of camera pose

图 7 相机姿态选择

2.2.2 用单应性矩阵求相机姿态

文献[19]指出, 当图像上匹配的特征点均来自于空间中同一个平面时, 无论有多少这样的匹配点, 它们对基本矩阵 F 的约束不会多于 4 个, 这将导致求解基本矩阵 F 的结果不稳定。此外, 当相机姿态只有旋转没有平移时(从 O_1 变换到 O_2 是纯旋转, $t = 0$), $E = [t]_x R = 0$, 难以通过对 E 的分解求解 R 。

如图 8 所示, 若空间中一平面 π 方程为: $n^T X + d = 0$, 则空间点 X 对应的两幅图像上的投影 x_1, x_2 有如下关系:

$$x_2 = H x_1, \quad H = K_2 \left(R - \frac{t n^T}{d} \right) K_1^{-1} \quad (32)$$

其中 K_1, K_2 分别为两个相机的内参矩阵。若相机为

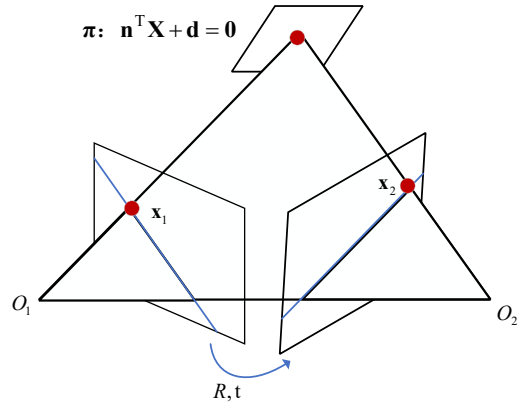


Fig. 8 Diagram of homography transformation

图 8 单应性变换示意图

纯旋转变换, 即 $t=0$, 则 $H = K_2 R K_1^{-1}$ 。对于 H 的求解方法, 同样可以采用类似于求解基本矩阵 F 的方法, 利用两幅视图上匹配点的约束关系 $x_2 = H x_1$, H 矩阵有 8 个自由度, 而一对匹配点能提供两个约束, 因此至少要 4 对匹配点即可求解 H ^[20]。此外, 可以采用 RANSAC 的方法估计 H , 然后便可对 H 进行分解以求解 R 和 t ^[20]。

2.3 三角测量

在已知相机参数和匹配点的前提下, 即可利用三角测量恢复空间点的三维坐标, 具体方法如下:

2.3.1 直接线性变换法

文献[9]提出了直接线性变换法。设第 i 个相机的投影矩阵为: $P_i = K_i [R_i, t_i]$, 空间点的三维坐标为 $X = [x, y, z, 1]^T$, 对应在第 i 个视角的图像坐标为 $x_i = [u_i, v_i, 1]^T$, d_i 表示 X 点在第 i 个视角下的深度, 即相机坐标系下的坐标 z_c 。则:

$$d_i x_i = P_i X \quad (33)$$

两边同时叉乘 x_i , 得:

$$x_i \times (P_i X) = 0 \quad (34)$$

由 $P_i X = \begin{bmatrix} P_{i1} X \\ P_{i2} X \\ P_{i3} X \end{bmatrix}$, 得: $v_i (P_{i3} X) - P_{i1} X = 0$
 $u_i (P_{i3} X) - P_{i2} X = 0$
 $u_i (P_{i2} X) - v_i (P_{i1} X) = 0$

其中, 第 3 个方程与前两个线性相关, 则一个观测视角能对 X 提供两个约束, 而 X 有 3 个自由度, 因此要求解 X , 至少需要图像上一对匹配点, 也即至少要从两个视角对其进行观测。当点数 N 大于 2 时, 还可以构建方程组利用最小二乘法求解:

$$AX = 0, A = \begin{bmatrix} u_1 P_{13} - P_{11} \\ v_1 P_{13} - P_{12} \\ \dots \\ u_i P_{i3} - P_{i1} \\ v_i P_{i3} - P_{i2} \\ \dots \\ u_N P_{N3} - P_{N1} \\ v_N P_{N3} - P_{N2} \end{bmatrix} \quad (35)$$

2.3.2 RANSAC方法

在存在外点的情况下,可以用RANSAC方法求解^[15]。主要步骤如下:①RANSAC采样次数,设置内点阈值(重投影误差);②随机采样一对视角,计算空间点三维坐标;③计算每个视角中的重投影误差,统计内点个数;④重复步骤②、③直到满足采样次数,选择内点个数最多的视角;⑤利用所有内点重新计算三维点坐标。

此外,在利用相机匹配对进行三角测量计算三维点时,可以从图像中提取EXIF初始值以获取焦距^[21-22]。

2.4 PnP问题

PnP问题指根据已知的三维点坐标和对应的二维点,求解相机内外参数的问题,这在增量式SFM中经常被用到^[21]。

常用的方法为直接线性变换法,设相机的投影矩阵为: $P = K [R, t]$,空间点的三维坐标为 $X = [x, y, z, 1]^T$,对应的图像坐标为 $x = [u, v, 1]^T$, d 表示 X 点的深度,即相机坐标系下的坐标 z_c 。则:

$$d \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K [R, t] X = K [R, t] X = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} X =$$

$$\begin{bmatrix} p_1^T \\ p_2^T \\ p_3^T \end{bmatrix} X \quad (36)$$

$$u = \frac{p_1^T X}{d} = \frac{p_1^T X}{p_3^T X} = \frac{X^T p_1}{X^T p_3}, v = \frac{p_2^T X}{d} = \frac{p_2^T X}{p_3^T X} = \frac{X^T p_2}{X^T p_3} \quad (37)$$

$$\begin{aligned} X^T p_1 - u X^T p_3 &= 0 \\ X^T p_2 - v X^T p_3 &= 0 \end{aligned} \quad (38)$$

P 矩阵共有12个参数,而一对3D-2D对应点能提供2个约束,因此至少需要6对匹配点即可求解 P 。因为 $P[:, 1:3] = KR$,且 K 为上三角阵,而 R 为正交矩阵,可以对 $P[:, 1:3]$ 进行QR分解,即可求得 R^{-1} 和 K^{-1} ,进而求得 K 和 R 。又 $P[:, 4] = Kt$,则 $t = K^{-1}P[:, 4]$,得平移向量 t 。

文献[23]提出一种计算PnP问题的Kneip算法,利用三维点构建一个平面作为过渡坐标系,进而求解相机的内外参数。此外,同样可以用RANSAC方法实现Kneip算法。求解PnP问题还有其他几种方法,如用4对不共面的点进行求解的P3P方法^[24],利用4对以上不共面或者3对共面点进行求解的ePnP法^[25],以及利用深度学习与传统方法相结合的EPro-PnP方法^[26],实现端到端的三维位姿预测。

2.5 捆绑调整

由于噪声的影响,以上初步求得相机内外参数和三维点坐标还不够准确,捆绑调整(Bundle Adjustment)^[27]可以同时三维点坐标和相机参数进行非线性优化。捆绑调整如图9所示。

设 $\chi_{ij} = 1$ 表示第 i 个点在第 j 个相机中可见, $\chi_{ij} = 0$ 表示第 i 个点在第 j 个相机中不可见; $X_i = (x_w, y_w, z_w, 1)^T$ 表示第 i 个三维点的世界坐标; P_j 表示第 j 个相机的内参以及世界坐标系与第 j 个相机的外参数; \hat{u}_{ij} 表示第 i 个点在第 j 个

相机中的观测点; u_{ij} 表示第 i 个点在第 j 个相机中的投影点。

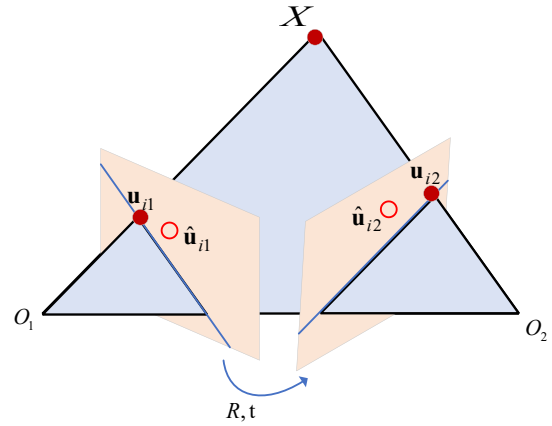


Fig. 9 Diagram of bundle adjustment
图9 捆绑调整示意图

则在有 n 个三维点和 m 个视角时,可构建损失函数:

$$g(\theta) = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} \|\hat{u}_{ij} - u_{ij}(P_j, X_i)\|^2 = \frac{1}{2} \sum_i^n \sum_j^m \chi_{ij} e_{ij}^2 \quad (39)$$

其中, $\theta = (P_1, \dots, P_m, X_1, \dots, X_n)$ 为待优化的量。对无约束非线性最小优化问题对于最优化问题 $g(\theta) = \frac{1}{2} \|x - f(\theta)\|^2$ 通常可用以下几种方法。

2.5.1 最速下降法

假设 $g(\theta)$ 在 θ_i 处可微,则它在 θ_i 处有泰勒展开(展开到一阶)。

$$g(\theta) = g(\theta_i) + (\nabla g(\theta_i))^T \delta\theta \quad (40)$$

其中, $\theta = \theta_i + \delta\theta$,当 $(\nabla g(\theta_i))^T \delta\theta < 0$ 时可保证 $g(\theta)$ 的值在下降,当 $\delta\theta$ 取梯度的反方向时,可达到最快下降速度。算法流程如下:

Step1: 给定初始点 θ_0 ,迭代终止阈值 ε 和步长 λ ,令 $t=0$ 。

Step2: 计算 $\nabla g(\theta_t)$,若 $\|\nabla g(\theta_t)\| \leq \varepsilon$ 停止迭代,输出 θ_t ,否则转下一步。

Step3: 取 $\theta_{t+1} = \theta_t - \lambda \nabla g(\theta_t)$, $t = t + 1$,转Step2。

2.5.2 牛顿法

最速下降法是收敛的,但最终的收敛是线性的,而且通常很慢,并且在某些情况下无法找到二阶多项式的最小值,此时可用牛顿法^[28]。

假设 $g(\theta)$ 在 θ_i 处二阶可微,且假定二阶导数 $\nabla^2 g(\theta)$ 总是正定,则它在 θ_i 处二阶近似函数 $Q(\theta)$ 极小值作为下一次迭代点 θ_{i+1} 。同样,将 $Q(\theta)$ 泰勒展开为:

$$Q(\theta) = g(\theta_i) + (\nabla g(\theta_i))^T \delta\theta + \frac{1}{2} \delta^T \theta \nabla^2 g(\theta_i) \delta\theta \quad (41)$$

对 θ 求梯度令其为0,可得:

$$\nabla Q(\theta) = \nabla g(\theta_i) + \nabla^2 g(\theta_i) \delta\theta = 0 \quad (42)$$

$$\delta\theta = -(\nabla^2 g(\theta_i))^{-1} \nabla g(\theta_i) \quad (43)$$

算法流程如下:

Step1: 给定初始点 θ_0 , 迭代终止阈值 ε 和步长 λ , 令 $t=0$ 。

Step2: 计算 $\nabla g(\theta_t)$, 若 $\|\nabla g(\theta_t)\| \leq \varepsilon$, 停止迭代, 输出 θ_t , 否则转下一步。

Step3: 取 $\theta_{t+1} = \theta_t - \lambda(\nabla^2 g(\theta_t))^{-1} \nabla g(\theta_t)$, $t = t + 1$, 转 Step2。

$$\begin{aligned} g(\theta) &= \frac{1}{2} p^T(\theta) p(\theta) \\ &= \frac{1}{2} p^T(\theta_t) p(\theta_t) + \delta^T \theta (\nabla p(\theta_t)) p(\theta_t) + \frac{1}{2} \delta^T \theta (\nabla p(\theta_t)) \nabla p(\theta_t)^T \delta \theta \end{aligned} \quad (45)$$

令 $J^T(\theta_t) = \nabla f(\theta_t) = \nabla p(\theta_t)$, 则:

$$g(\theta) = \frac{1}{2} g(\theta_t) + \delta^T \theta J^T(\theta_t) p(\theta_t) + \frac{1}{2} \delta^T \theta J^T(\theta_t) J(\theta_t) \delta \theta \quad (46)$$

对 θ 求梯度令其为0, 可得:

$$J^T(\theta_t) p(\theta_t) + J^T(\theta_t) J(\theta_t) \delta \theta = 0 \quad (47)$$

而 $\nabla g(\theta_t) = J^T(\theta_t)(f(\theta_t) - x) = J^T(\theta_t) p(\theta_t)$, 故:

$$\delta \theta = -(J^T(\theta_t) J(\theta_t))^{-1} \nabla g(\theta_t) \quad (48)$$

与牛顿法相比, 用 $J^T(\theta_t) J(\theta_t)$ 代替 $\nabla^2 g(\theta_t)$, 降低了运算复杂度。

2.5.4 Levenberg-Marquardt 法

最速下降法是用平面在局部拟合损失函数, 牛顿法是用二次曲面拟合, 当 $\nabla^2 g(\theta)$ 在某区域负定时, 牛顿法可能收敛到局部极大值, 高斯-牛顿法在 $J(\theta)$ 不满秩时可能失效^[28]。Levenberg-Marquardt 法^[30-31]是将最速下降法和高斯-牛顿法有机结合, 使用“信赖阈”控制下降速度的方法。当收敛速度较快时, 信赖阈增大, 算法趋向于高斯-牛顿法; 当收敛速度慢时, 信赖阈减小, 算法趋向于最速下降法, 优点是只用到一阶雅各比矩阵, 计算速度快, 对初始值有一定的鲁棒性。

对于最优化问题 $g(\theta) = \frac{1}{2} \|x - f(\theta)\|^2$, 可得 $\nabla g(\theta_t) = -J^T(\theta_t)(x - f(\theta_t))$, 其中, $J^T(\theta_t) = \nabla f(\theta_t)$ 。另根据以上推导, $\nabla^2 g(\theta_t) \delta \theta = -\nabla g(\theta_t)$, 可以将 $\nabla^2 g(\theta_t)$ 近似替换为 $J^T(\theta_t) J(\theta_t) + \frac{1}{\lambda} I$, λ 为信赖阈半径。于是构建增量正规方程:

$$(J^T(\theta_t) J(\theta_t) + \frac{1}{\lambda} I) \delta \theta = J^T(\theta_t)(x - f(\theta)) \quad (49)$$

通过求解增量正规方程得到 $\delta \theta$ 。可见, 当 λ 趋向无穷大时, $(J^T(\theta_t) J(\theta_t)) \delta \theta = -\nabla g(\theta_t)$, 近似于高斯-牛顿法; 当 λ 趋近于0时, $\delta \theta = -\nabla g(\theta_t)$, 近似于最速下降法。

Levenberg-Marquardt 法的主要流程为:

Step1: 选取初始值 θ_0 , 终止控制阈值 ε , 令 $e_0 = \|x - f(\theta_0)\|^2$, $\lambda_0 = 10^{-3}$ 。

Step2: 计算 $J^T(\theta_t)$ 。

Step3: 构造增量正规方程 $(J^T(\theta_t) J(\theta_t) + \frac{1}{\lambda} I) \delta \theta = J^T(\theta_t)(x - f(\theta))$ 。

Step4: 通过求解增量正规方程, 得到 $\delta \theta$ 。

2.5.3 高斯-牛顿法

牛顿法虽然速度快但计算量大, 需要保存二阶 Hessian 矩阵的逆矩阵, 高斯-牛顿法与牛顿法类似, 但区别是它基于 $f(\theta)$ 在 θ 附近的线性近似^[29], 即设 $p(\theta) = f(\theta) - x$, 对其泰勒展开为:

$$p(\theta) = p(\theta_t) + (\nabla p(\theta_t))^T \delta \theta \quad (44)$$

如果 $\|x - f(\theta_0 + \delta \theta)\|^2 < e_k$, 令 $\theta_{t+1} = \theta_t + \delta \theta$; 如果 $\|\delta \theta\| < \varepsilon$, 终止迭代; 否则令 $\lambda_{t+1} = 10\lambda_t$, 执行 Step3。如果 $\|x - f(\theta_0 + \delta \theta)\|^2 \geq e_k$, 令 $\lambda_{t+1} = 0.1\lambda_t$, 执行 Step3。

2.6 运动恢复结构分类

典型的运动恢复结构方法有增量式 (Incremental SFM)^[32-35]、全局式 (Global SFM)^[36-37]、分层式^[38-39]。

2.6.1 增量式运动恢复结构

主要步骤包括: 图像特征点检测与匹配、构建图像连接图、初始化、增量式重建新的视角、全局捆绑调整。

(1) 图像特征点检测与匹配。采用上文所述方法, SIFT 方法检测的特征点比较稳定, 其应用较广泛^[33-34]。

(2) 图像连接图构建。每一张图像形成图上的一个结点, 具有特征匹配点的两幅图像之间有边相连, 结点的边越多, 表示与其具有匹配的特征点的其他图像越多。从图 10 可以看出, 白天和夜晚拍摄的图像因为光度差异会在连接图上形成两个簇^[33]。

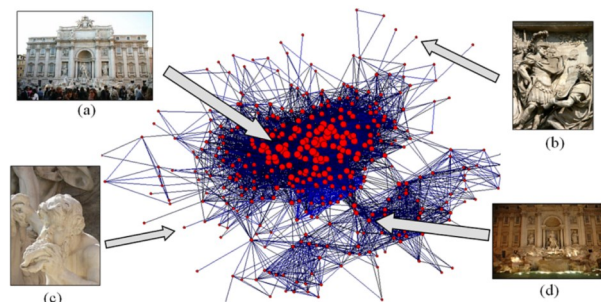


Fig. 10 Construction of image connecting graph

图 10 图像连接图构建

(3) 初始化。选择一对初始相机进行 Tracks 重建 (三角测量) 恢复三维点坐标, 然后进行 Tracks 滤波和捆绑调整以优化估计的三维点坐标和相机内外参数。

其中, 初始相机对的选取要遵循一定的原则: 匹配点足够多、基线足够长、满足单应性的匹配点尽量少^[33-34]。文献[34]指出, 由于冗余数据的存在, 从图像连接图中的密集位置进行初始化通常会使得重建更加鲁棒和准确。

Tracks 滤波是指滤除 Tracks 重建中生成的错误点 (包括无穷远处的点和重投影误差过大的点)^[21, 33], 以降低后续运算复杂度。

(4) 新的视角增量式重建。选择能看到已被估计的三

维点最多的视角,利用PnP方法^[21]或者RANSAC方法^[33]求解新视角的相机姿态,然后对单个相机姿态进行捆绑调整。下一步对新视角能看到的点进行Tracks重建和滤波,以此优化三维点的坐标。

(5)全局捆绑调整。对所有已加入的相机内外参数和三维点坐标进行一次全局的非线性优化,由于这一方法运行时间较长,可以在加入多个视角之后运行一次^[21]。

增量式运动恢复结构如图11所示。其主要优点是:对误匹配的特征点鲁棒、场景结构可在调整中不断优化。缺点是:对初始视角的选取和视角添加顺序敏感、大场景下的累计误差会导致场景漂移、捆绑调整重复进行导致效率低。文献[35]提出一种与RANSAC相结合提升精度的增量式运动恢复结构方法。

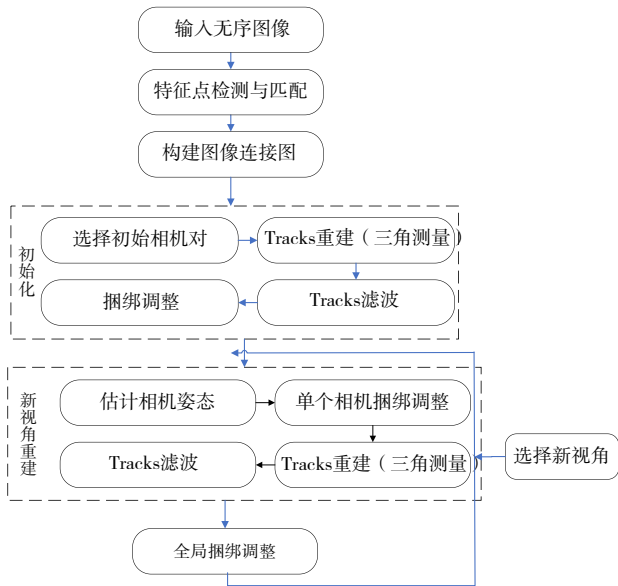


Fig. 11 Incremental structure from motion
图11 增量式运动恢复结构

2.6.2 全局式运动恢复结构

由于捆绑调整比较耗时,全局式运动恢复结构同时估计所有相机的旋转矩阵和平移向量,生成三维点后,只运用一次捆绑调整进行优化。文献[36]提出一种估计全局相机姿态的方法。

首先,在生成图像连接图之后,利用任意两幅图像的特征点匹配关系,计算两个相机的相对旋转矩阵 R_{ij} 和平移向量 t_{ij} ,通过最小化如下关系求解所有相机相对于世界坐标系的旋转矩阵 R_i 与 R_j : $\min_{\mathbb{R}} \|R_{ij} - R_j R_i^T\|$ 。

其次,进行全局旋转矩阵滤波,根据求得的 R_i 与 R_j 优化图像连接图,即当 $\|R_{ij} - R_j R_i^T\| > \epsilon$ 时,认为相机 i 与 j 之间存在错误的连接边,要删除。

再次,估计全局平移向量 C_i 与 C_j ,用约束关系

$$\min_c \left\| l_{ij} - \frac{C_i - C_j}{\|C_i - C_j\|} \right\| \text{ 求解 } C_i \text{ 与 } C_j, \text{ 其中, } l_{ij} = R_j^T t_{ij}.$$

$$\text{再进行全局平移向量滤波: } \left\| l_{ij} - \frac{C_i - C_j}{\|C_i - C_j\|} \right\| > \epsilon \text{ 时, 认为}$$

为相机 i 与 j 之间存在错误的连接边,要删除。

最后,利用求得的相机姿态求解三维点并进行一次全局的捆绑调整。

文献[37]提出一种多种信息融合的方法,在求解过程加入了GPS等辅助信息,对结果进行优化。

全局式运动恢复结构如图12所示。其主要优点是:没有误差积累,误差均匀分布在连接图上;对初始相机选取和相机添加顺序不敏感;捆绑调整仅执行一次,重建效率高。不足之处主要体现在:相机位置求解时对匹配外点敏感,鲁棒性不足;连接图边界被过滤,容易造成部分图像丢失,导致重建不完整。

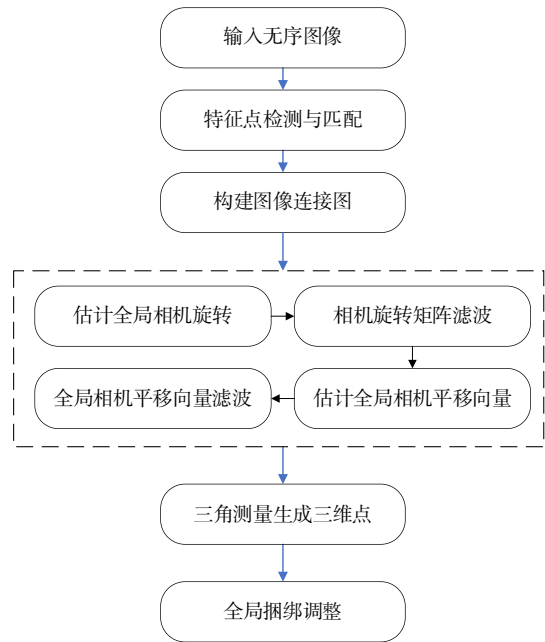


Fig. 12 Global structure from motion
图12 全局式运动恢复结构

2.6.3 分层式运动恢复结构

文献[38]提出基于聚类方式的分层式运动恢复结构,如图13所示。其主要步骤为:利用已知视角创建一个局部模型进行捆绑调整,加入一个新的视角,两个模型融合捆绑调整。这种方法计算效率高,且算法稳定,不依赖初始相机的选取,避免了大场景中的误差积累和漂移,但规则设计较为复杂。

3 稠密点云重建

基于图像的三维重建中,稠密点云的获取通常基于多视图立体技术(Multi-view Stereo, MVS)。该技术的优点是成本低、精度较高、图像来源广;缺点是计算速度慢。

MVS中常用的一个基本假设是光度一致性假设(Photo-Consistency),是指同一空间的点在不同视角的投影,光

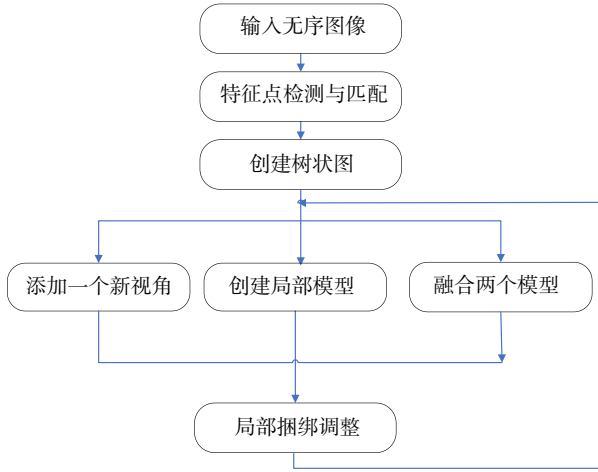


Fig. 13 Hierarchical structure from motion

图 13 分层式运动恢复结构

度应该相同,重建的关键是恢复光度一致性的点。光度一致性度量方式有:误差平方和(SSD, Sum of Squared Difference)、归一化交叉协方差(NCC, Normalized Cross Correlation)。具体为:

$$\rho_{SSD} = \|f - g\|^2 \quad (50)$$

$$\rho_{NCC} = \frac{(f - \bar{f})(g - \bar{g})}{\delta_f \delta_g} \quad (51)$$

其中, f 和 g 分别为特征描述子的特征向量, \bar{f} 、 \bar{g} 为描述子的均值向量, δ_f 、 δ_g 为描述子向量标准差。

基于多视图立体技术(MVS)的稠密点云重建主要有基于体素的方法、基于深度图融合的方法和基于空间 patch 扩散的方法。

3.1 基于体素的方法

基于体素的方法是指将空间划分为很多体素,通过图像序列的约束计算体素内三维点的方法。空间划分包括两种类型:一种是规则体素的划分^[40],即将空间划分为小立方体;另一种是不规则体素的划分^[41-42],即将空间划分为小四面体,如图 14 所示。

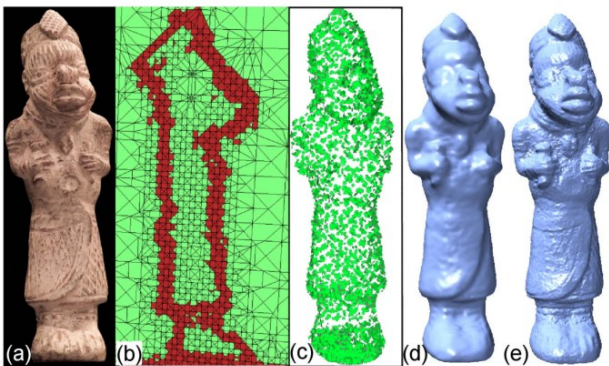


Fig. 14 Dense reconstruction based on voxels

图 14 基于体素的稠密重建

这种方法的优点是:生成规则的点云、便于提取物体的平面。其缺点是:精度受到空间划分分辨率的影响,仅适用于小场景,单个物体,遮挡较少的场景,难以处理高精

度、大规模的场景。

3.2 基于深度图融合的方法

基于深度图融合的方法^[43]是利用SFM产生的稀疏点云构建种子栅格(patch),通过区域生长对 patch 进行优化后生成不同视角下的深度图,再将不同视角深度进行融合的稠密点云重建方法。该方法主要包括以下 3 个步骤:数据预处理、区域生长、深度图融合,如图 15 所示。

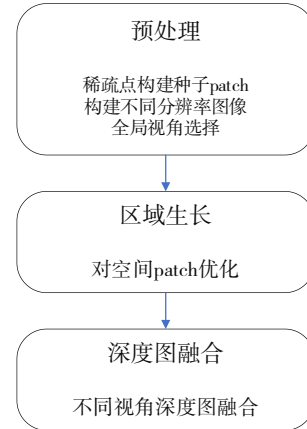


Fig. 15 Dense reconstruction based on depth map fusion

图 15 基于深度图融合的稠密重建

3.2.1 数据预处理

patch 是以空间中的 3D 点中心建立一个很小的栅格,栅格的分辨率与图像分辨率以及 3D 点的位置相关。patch 的中心点即为 3D 点的位置,patch 的朝向即为 3D 点的法向量,patch 上的每个点可以投影到不同视角中计算 NCC 用于度量光度一致性。

其预处理包括:通过 SFM 重建的稀疏点构建种子 patch、构建不同分辨率的图像以处理图像分辨率不同情况、全局视角选择。全局视角选择是指选择满足特定条件的视角用于区域生长。一般要满足以下条件:①图像具有相同的内容,外观——通过共享 sift 特征点的个数;②图像具有足够大的视差(宽基线)——通过视线的夹角;③图像具有相似的分辨率——通过计算投影球半径。

为使得计算加速,在生成某一视角的深度图时,文献^[43]从以上全局视角中选择 4 个视角,称为局部视角。局部视角的选择要求:NCC 值大于一定的阈值,并且和已经选择的视角的极平面要足够分散(不共面),如图 16 所示。

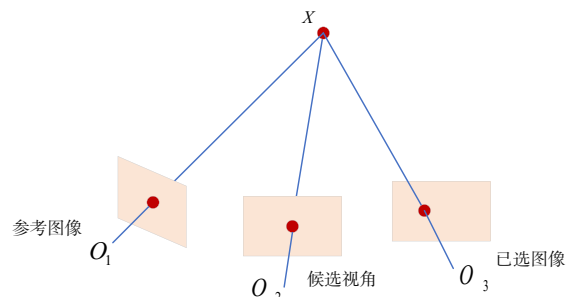


Fig. 16 Diagram of view selection

图 16 视角选择示意图

3.2.2 区域生长

在建立初始种子 patch 后,可以利用种子 patch 逐步扩张的方法在其邻域生成新的 patch。其主要步骤包括:①对稀疏点云中的种子 patch 以置信度(光度一致性)建立优先级队列;②估计初始稀疏种子点的深度;③对每个种子点进行非线性深度优化;④每次优化完后,如果图像上的对应像素没有深度值或当前像素的置信度值高于邻域像素一定范围,则将其像素添加到队列中。

其中,非线性优化的数学模型为:

$$E = \sum_{ijk} (I_R(i, j) - c_k \cdot I_k(i, j))^2 \quad (52)$$

其中, i, j 是 patch 中的采样点, k 为视角个数。 $I_R(i, j)$ 表示点在参考图像上的光度, c_k 是颜色尺度, 由于 $I_k(i, j)$ 表达式中含有深度信息, 因而可以利用梯度下降等优化方法对 E 进行优化, 进而求解像素点深度。优化求解结果如图 17 所示^[43]。

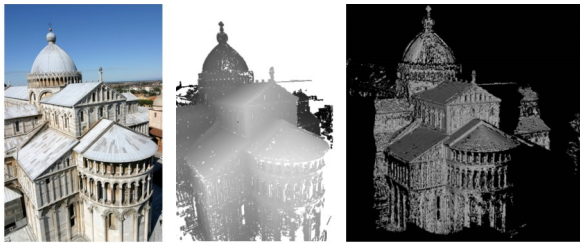


Fig.17 Reconstruction effect based on depth map fusion

图 17 基于深度图融合的重建效果

3.2.3 深度图融合

深度融合是指在三维空间合并位置相近的点。融合过程同样要对点施加一定的约束, 通常包括一致性约束、可视性约束。一致性约束是指相同的点在不同视角下的深度应当具有一致性; 可视性约束是指在图像上可视的三维点在融合后的深度图上不能被遮挡。如图 18 所示, A 和 C 将被认为是错误的点, 予以剔除, B 将被选择为正确的点。

深度图融合方法的特点为: 邻域视角选择提升了深度

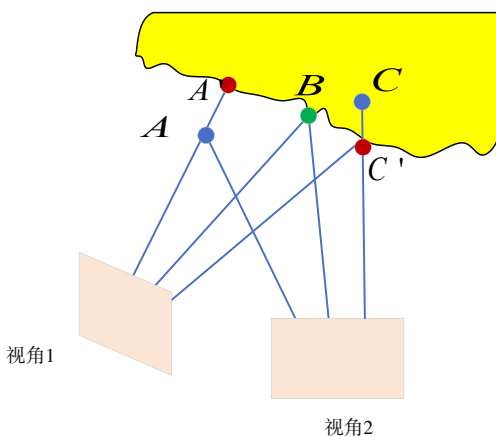


Fig. 18 Diagram of consistent point selection

图 18 一致点选择示意图

估计的准确度, 对一些有遮挡、障碍物较多的场景能较好地处理, 适用场景广泛, 只用到光度一致性约束和可视性约束, 算法实现简单。

3.3 基于空间 patch 扩散的方法

与深度图中在二维图像上进行 patch 扩张不同的是, 基于空间 patch 扩散的方法^[44]采用的是在空间中对三维 patch (大小为 5x5 共 25 个三维点) 进行扩散, 最终使 patch 覆盖物体表面。

重建流程如图 19 所示。

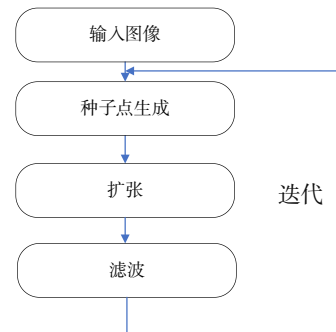


Fig. 19 Reconstruction process based on patch diffusion

图 19 基于空间 patch 扩散重建流程

初始种子点生成采用 SIFT 等特征点; 扩张过程对已重建三维点的邻域进行匹配, 滤波过程采用两种约束去除噪声点: 光度一致性约束和可视性约束。

初始 3D patch 的生成步骤: ①在图像上均匀计算 SIFT/Harris 特征; ②沿极线进行搜索找到匹配特征点; ③对匹配对, 通过三角化建立 patch, 法向量指向参考图像, 通过光度一致性约束对可视图像进行筛选; ④对 patch 位置和法向量进行优化。

空间 patch 扩散步骤为: ①将三维 patch 投影到图像上; ②如果相邻 cell 没有 patch 且深度连续, 则进行 patch 扩散, 新建 patch 的法向量初始值等于邻域 patch, 新建 patch 的位置设置为当前 cell 的视线和邻域 patch 所在平面的交点; ③计算初始 patch 的可视图像, 并进行优化。

扩张过程中并没有考虑到可视性的情况, 滤波过程的主要目的是利用各种约束对这些不合理的点进行筛选。可用的约束包括可视性约束, 可视图像个数小于一定阈值, 图像邻域中的 cell 同时也是空间邻域, 其比例小于一定阈值 (0.25)。对于每一个 patch, 统计它所在及相邻 image cell 中的所有 patch。如果 3D 空间中与相邻的 patch 比例小于一定值 (0.25), 则认为是孤立点, 将被移除。

PMVS 和基于深度图的重建方法类似, 基本原理也是基于光度一致性, 并且采用区域生长的方式进行稠密重建, 同样适用于杂乱有遮挡的大规模场景。算法优点是: 适用性强, 适用于各种形状的物体, 能够直接重建出完整的稠密点云, 不需要进行单独的深度图融合; 重建过程中同时考虑了可视性约束, 能够得到更加准确的重建结果。其缺点是不适用于非朗伯面 (Non-Lambertian), 并容易产

生空洞, 计算量非常大。

近年来, 也有针对以上算法不足的改进方法, 如: 文献 [35] 利用并行计算提升点云重建速度; 文献 [45] 针对在小范围场景三维重建过程中存在离群点的现象, 将 PMVS 算法与统计分析法相融合提出一种改进的点云滤波算法, 提高了重建精度。

当前, 稠密点云重建面临的主要问题与挑战是: ①弱纹理和朗伯面——不能满足光度一致性约束; 通常需要先验形状约束进行重建; ②细长结构重建——很难得到像素级别精度的深度值; 通常的解决方法需要进行语义级别的重建; ③动态场景重建——图像序列的重建方式可以捕捉动态场景, 但是基本只适合于实验室环境; 动态物体的重建需要结合形状先验、高级语义信息等。

4 三维表面重建

在得到稠密点云后, 利用各种算法可以从点云重建物体表面。三维物体的表面表达方式包括: 边界表示法、空间划分法、构造体素法。边界表示法用四边形、三角形、参数曲面或者非参数的曲面表示物体形状, 具有稳定性强、灵活性强、有助于表示表面细节。空间划分法将物体内部划分成细小连续实体以描述物体形状。构造体素法通过集合运算(并、交、差)将简单形状组合成复杂形状, 只能用来表示较为简单的实体, 无法用于形状复杂或表面精细的物体。

计算机视觉中最常用的三维模型表示方式为三角网格(Triangular Mesh)。其基本结构包括: 顶点(Vertex)、面片(Facet)、边(Edge)。点、线和面上的各种属性包括: 颜色、法向量、纹理坐标(Texture Coordinate)。

常用的三维表面重建方法有: 基于二元分割的表面重建方法、基于符号距离场的表面重建方法。

4.1 基于二元分割的表面重建方法

其主要思想是将空间划分成一组小单元(体素/四面体), 采用二元分割思想将这些小单元划分成内部和外部两类, 介于内部和外部之间的面即为物体表面, 如图 20 所示^[43]。

在稠密点云的基础上进行表面重建主要经历的步骤

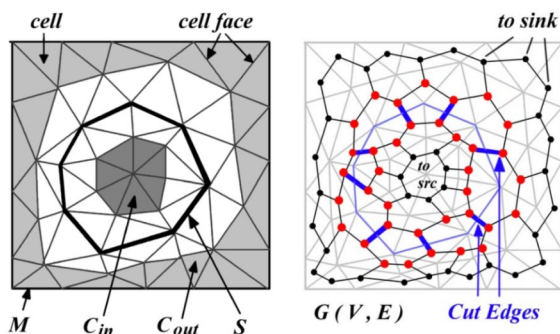


Fig. 20 Surface reconstruction based on binary segmentation

图 20 基于二元分割的表面重建

如下:

4.1.1 德劳内三角剖分

德劳内三角剖分(Delaunay Triangulation)是指将点云中的相邻点用一种特定的方式连线, 使得任意相邻的 4 个点构成空间中的一个四面体。德劳内三角剖分具有以下特性:

(1)空球特性。任意 4 个点的外接球内都不含有其他点。

(2)最大化最小角。德劳内三角剖分最大化四面体上三角形的最小角。

(3)对偶性。德劳内三角剖分和维诺图(Voronoi Diagram)是互为对偶的关系, 如图 21 所示^[43]。

文献 [46] 给出了一种增量式进行德劳内三角剖分的方法。

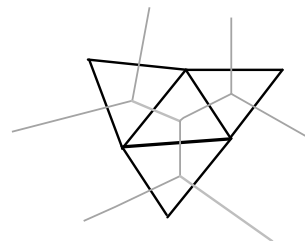


Fig. 21 Duality of Delaunay triangulation and Voronoi diagram

图 21 德劳内三角剖分与维诺图的对偶关系

三维表面的生成问题将转化为四面体的二元标记问题, 即将一剖分四面体标记为“外部”, 另一剖分四面体标记为“内部”, 内外交界处的空间曲面即为物体表面。

4.1.2 用图分割方法提取物体表面

图分割(Graph Cut)是指对一特殊的有向加权图的某些边进行切分, 将图分割为没有交集的两个部分, 使得边的权重损失最小^[46-48]。这一图的特殊之处在于包含两个特殊的节点: 源节点和汇节点, 其中源节点只有出边, 而汇节点只有入边, 如图 22 所示, 边的粗细表示权重大小, 切分目标是选取权重最小的边切分。

在对稠密点云进行德劳内三角剖分后, 根据其其与维诺图的对偶关系生成维诺图: 即使四面体的中心作为维诺图

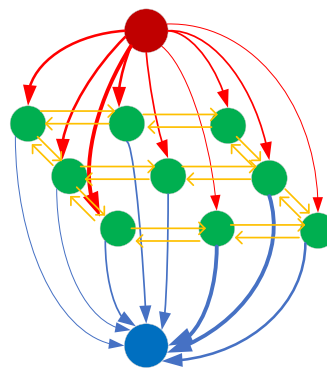


Fig. 22 Diagram of graph cut

图 22 图分割示意图

各小区域的中心,四面体的三角形表面作为维诺图的边。在一定约束条件下,对维诺图的边赋予一定权重并将其看作有向加权图,于是便可进行图分割提取物体表面,如图23所示^[43]。

对维诺图的边赋予权重的方法包括以下几种:可见性约束、光度一致性约束、平滑性约束^[46]、剪影约束^[43]、表面质量约束^[47]。

重建结果如图23所示^[43]。

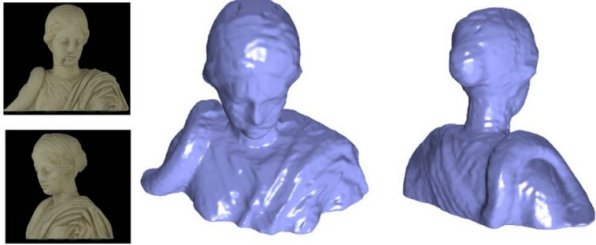


Fig.23 Surface reconstruction based on graph cut
图23 基于图分割的表面重建

4.1.3 网格细节优化

用二元分割方法提取的物体表面,由于点云存在噪声或者空洞,会导致原始网格存在噪声或者空洞,无法捕捉场景细节,因而需要对网格细节进行优化(Mesh Refinement)。

文献[44]提出一种基于光度一致性的网格细节优化方法,即利用多视角图像再次计算相同点在不同视角下的重投影误差,通过同时或者逐个移动网格顶点的位置,使重投影误差最小,也使光度一致性最好。度量光度一致性的能量函数为:

$$E_{\text{error}}(S) = \sum_{i,j} \int_{\Omega_{ij}^S} h(I_i, I_j^S)(x) dx_i \quad (53)$$

其中, $h(I_i, I_j^S)(x)$ 表示图像 i 和 j 之间在像素 x 处与光度一致性呈单调递减的某一函数; I_j^S 表示将图像 i 通过曲面 S 重投影到图像 i 上; Ω_{ij}^S 表示重投影的有效区域; $E_{\text{error}}(S)$ 的含义即为所有两两视角对下相同区域的重投影误差之和; $E_{\text{error}}(S)$ 越小,说明生成的网格越准确。

要优化网格顶点位置,即要求以上能量函数关于网格顶点的梯度,应利用梯度下降法进行优化。但由于网格顶点的位置变化实际上会影响其周围三角面片上点的位置,于是先计算关于某顶点周围三角面片点的梯度,再采用插值方法计算这一顶点的梯度,插值方法采用重心坐标法。 $E_{\text{error}}(S)$ 对 $[X_i]$ 的梯度为其对网格上 X_i 一周所有点梯度的加权和具体如图24所示。

$$\frac{dE_{\text{error}}(S)}{dX_i} = \int_S \varphi_i(x) \nabla E dx \quad (54)$$

其中, $\mathbf{v}(x)$ 表示顶点 X_i 一周三角网格上的任意一点, $\varphi_i(x)$ 表示 $\mathbf{v}(x)$ 相对于 X_i 的重点坐标, ∇E 是 $E_{\text{error}}(S)$ 相对于 $\mathbf{v}(x)$ 的梯度。

基于二元分割的表面重建方法具有以下特点:①算法

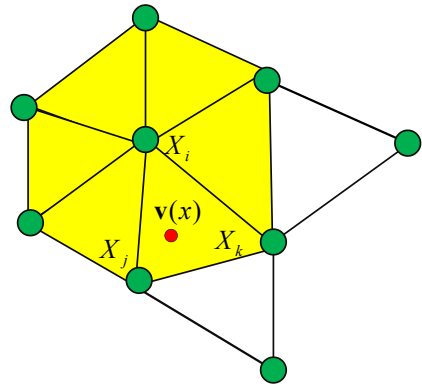


Fig. 24 Grid vertex position optimization
图24 网格顶点位置优化

流程简单,容易实现;②适用于大规模场景以及小物体重建,能够处理复杂表面;③重建效果依赖德劳内三角剖分的质量,容易受到噪声和外点影响;④通常需结合网格细节优化获取更加精细的表面细节。

4.2 基于符号距离场的表面重建方法

基于符号距离场的表面重建流程如图25所示。

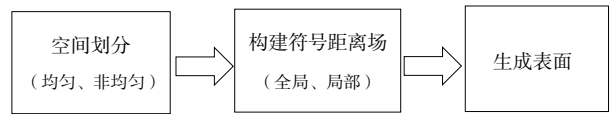


Fig. 25 Surface reconstruction based on symbol distance field
图25 基于符号距离场的表面重建流程

4.2.1 空间划分

空间划分主要是用于对空间中的隐函数进行离散化,以便于对划分区域的每个顶点计算一个符号距离值。空间划分方式可以采用均匀和非均匀两种。均匀划分是将空间划分为大小相等的体素网格(Grid),非均匀方式是用八叉树(Octree)结构将空间进行划分,即在点密集处空间划分较细,树的顶点较小;点稀疏处空间分辨率较低,树的顶点较大,点的分辨率可以根据三维点到相邻点的平均距离确定。

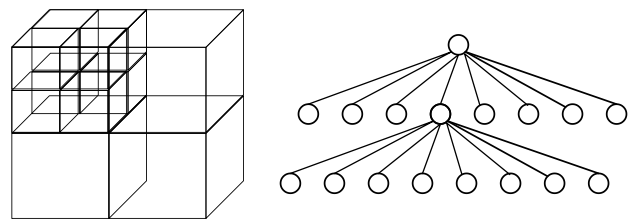


Fig. 26 Uniform (left) and non-uniform (right) spatial division
图26 均匀(左)与非均匀(右)的空间划分

4.2.2 符号距离场构建

文献[49]提出一种局部的符号距离场构建方法,用符号距离函数(Signed Distance Function, SDF)计算空间中各处的符号距离值,然后提取等值面作为物体表面。

符号距离函数是某度量空间中点 X 到某一边界的距离关于 X 位置的函数,当 X 在边界内时, SDF 为正;当 X 在

边界外时, SDF 为负。如图 27 所示, 0 所在位置即可认为是物体表面。

4	3	2	1	0	-1
3	2	1	0	-1	-2
3	1	0	-1	-2	-3
2	1	0	-1	-2	-3
1	0	-1	-2	-3	-4

Fig. 27 Diagram of symbol distance function

图 27 符号距离函数示意图

文献[49]所用的符号距离函数为:

$$F(\mathbf{x}) = \frac{\sum_i c_i w_i(\mathbf{x}_i) f_i(\mathbf{x}_i)}{\sum_i c_i w_i(\mathbf{x}_i)} \quad (55)$$

即在点 \mathbf{x} 附近邻域取 i 个点 \mathbf{p}_i , 在 \mathbf{p}_i 处建立局部坐标系 (x 轴与 \mathbf{p}_i 点的法向量 \mathbf{n}_i 方向相同, y 轴与 z 轴分别与 x 轴垂直), 再利用旋转矩阵 R_i 将 \mathbf{x} 旋转到局部坐标系下, 得到 $\mathbf{x}_i = R_i \cdot (\mathbf{x} - \mathbf{p}_i)$, 对以下基函数 $f(\mathbf{x}_i)$ 用 $c_i w_i(\mathbf{x}_i)$ 加权计算即可得到点 \mathbf{x} 处的符号距离值, 其中, $w_i(\mathbf{x}_i)$ 权重函数, c_i 为点 \mathbf{p}_i 的置信度 (NCC 值)。

$$f(\mathbf{x}_i) = f_x(x) f_y(y) f_z(z) = \frac{x}{\sigma^4 2\pi} \cdot e^{-\frac{1}{2\sigma^2}(x^2 + y^2 + z^2)} \quad (56)$$

局部符号距离场结果如图 28 所示^[49]。

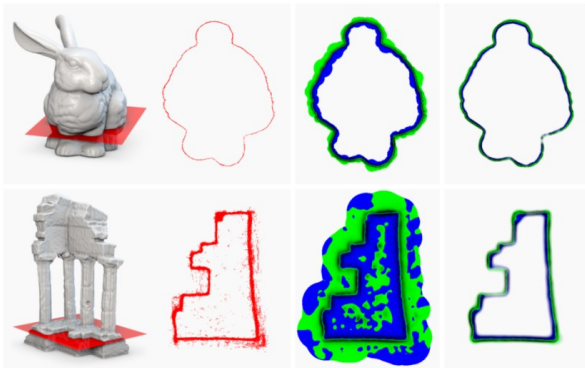


Fig. 28 Boundaries computed by symbolic distance function

图 28 符号距离函数计算边界

文献[50]提取了一种全局的符号距离场构建方法, 将向点看作是对隐函数梯度的采样, 如图 29 所示。全局的方法优点是鲁棒, 能够很好地处理噪声和空洞, 缺点是计算量大, 需要求解大规模的方程。

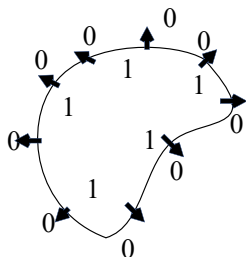


Fig. 29 Poisson surface reconstruction

图 29 泊松表面重建

4.2.3 移动立方体算法生成表面

移动立方体算法 (Marching Cube) 最早由文献[51]提出, 在得到每个体素 8 个顶点的符号距离值后, 通过插值方法生成三角形面片, 每个顶点有两种状态, 总共有 256 种, 但由于反转状态不变, 因此可以减少一半, 为 128 种, 再根据旋转不变形, 又可以减少到 14 种情况。可以认为这 14 种类类似于基, 经过旋转、反转可以得到 256 种状态所对应的结果 (其中 3 种情况), 如图 30 所示。

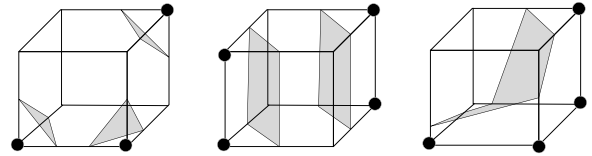


Fig. 30 Surfaces generated by marching cube method

图 30 移动立方体法生成表面

文献[52]提出一种专门针对八叉树的方法, 旨在解决相邻节点分辨率不一致时可能出现的裂缝问题。通过引入一组从八叉树的拓扑推导而来的二叉边树, 可以在不约束八叉树的拓扑结构或修改顶点值的情况下, 直接提取水密的多边形网格。

文献[53]提出一种将平面投影与区域生长相结合的表面重建方法, 降低了表面重建复杂度。文献[54]用统计滤波器对点云简化去噪, 并建立点云间拓扑结构, 缓解了重建表面粗糙、孔洞等问题。

5 纹理图像创建

为了生成更加逼真的场景和物体, 可以对生成的三维模型贴上纹理图像。文献[56]在文献[55]的基础上提出创建纹理图像管线, 主要分为如下步骤: 视角选择、全局颜色调整、泊松图像编辑。

5.1 视角选择

为每个三角面片选择唯一的视角, 用于获取纹理信息, 视角的选择应当考虑到以下因素: 图像尺度、图像细节丰富程度、图像可视性、邻域平滑性。具体方法为: 将视角选择建模成多标签分配问题, 为每一个三角形分配一个标签 (视角), 在网格上建立马尔可夫随机场, 每个面片代表一个顶点, 通过优化能量函数得到最优视角选择。能量函数分为数据项 E_{data} 和平滑项 E_{smooth} , 即:

$$E(1) = \sum_{F_i \in Faces} E_{data}(F_i, l_i) + \sum_{F_i, F_j \in Faces} E_{smooth}(F_i, F_j, l_i, l_j) \quad (57)$$

其中, F_i 表示第 i 个三角面片, l_i 表示分配给它的视角标签, 数据项 E_{data} 通过三角面片对应的二维图像上三角形内的平均像素梯度进行度量, 旨在选择图像分辨率较大的视角; 平滑项约束希望相邻三角形有相同的视角以保持邻域的平滑性, 从而减少缝隙, 为后续处理提供便利。选择好对应的视角后, 即可将网格投影到对应视角的可视图像

上,截取对应图像作为纹理图像,经过坐标归一化处理,即可作为三角面片的纹理坐标,后期通过纹理坐标可找到纹理图像上对应的贴片。

5.2 全局颜色调整

由于不同视角间存在相机曝光或者光照差异导致不同的纹理图像边界处存在明显缝隙,因而要为边界处的每个像素添加一个颜色调整量使得缝隙处的颜色差异尽量小,纹理图像内部相邻像素的调整量尽量相似。

具体方法:首先为三角面片的每个顶点对应像素添加一个调整量,再通过插值方式得到三角面片内部对应的每个像素的调整量,然后将缝隙处的顶点拆分成 n 个 (n 为标签个数),对其施加 n 个调整量,最后通过优化以下能量函数得到调整后的结果:

$$E_g = E_s(g) + \lambda E_i(g) \quad (57)$$

其中, $E_s(g)$ 的约束目标是使不同视角下同一顶点调整后的颜色尽量接近,而 $E_i(g)$ 的约束目标是使相同视角下不同顶点的调整量尽量接近, λ 为可调整系数。

5.3 泊松图像编辑

对于缝隙比较严重的区域,全局颜色调整并不能保证完全去除缝隙,还需要作进一步处理。泊松图像编辑算法原理为对前景图像和背景图像进行混合,使得融合后的图像满足:边界上的像素值与背景图像相同,前景区域内的梯度与引导梯度场相同^[57]。

这种方法最初用于图像融合。如图 31 所示^[57],要将源图像 g 插入到目标图像 S 的区域 Ω 中,该区域具有边界 $\partial\Omega$,目标图像的像素值与像素坐标的关系可表示为一个函数 f^* ,而合成后的图像在区域 Ω 中的像素值与像素坐标关系用 f 描述, v 代表一个引导梯度场,实际上是指源图像 g 的梯度。其中, $g, S, \Omega, \partial\Omega, f^*, v$ 都是已知量,而 f 是待求函数(融合后在区域 Ω 中的图像)。

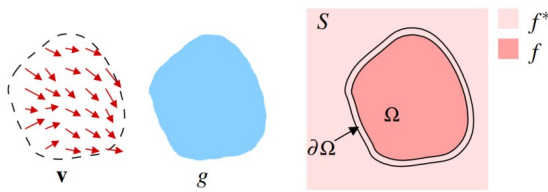


Fig. 31 Poisson image fusion

图 31 泊松图像融合示意图

优化的函数为:

$$\min_f \iint_{\Omega} |\nabla f - v|^2 \text{ wrt. } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (58)$$

其中, ∇f 表示图像融合后 f 的梯度,优化目标是使 f 边界上的像素值与背景图像相同, f 的梯度与源图像 g 的梯度 v 相同。求解以上优化函数可以转化为求解以下泊松方程:

$$\Delta f = \text{div } v \text{ over } \Omega, \text{ wrt. } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (59)$$

其中, Δf 表示 f 的拉普拉斯滤波结果,在某一像素处可以用相邻像素线性表示,而 $\text{div } v = \frac{\partial^2 g}{x^2} + \frac{\partial^2 g}{y^2} = \Delta g$,于是可将泊松方程转化成 $\Delta f = \Delta g$,用线性方程快速求解。泊松图像融合效果如图 32 所示^[57]。

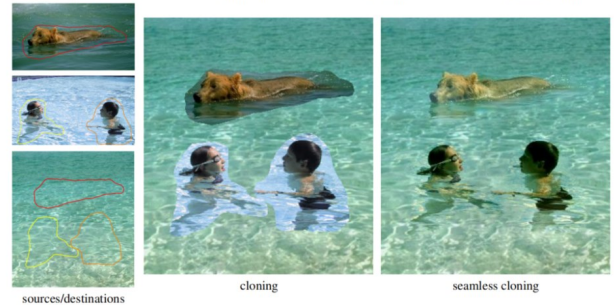


Fig. 32 Effect of Poisson image fusion

图 32 泊松图像融合效果

进行全局颜色调整后,只需在缝隙处进行泊松编辑即可得到边界平滑的纹理图像。此外,可以通过控制内部区域的范围(如由当前边界往前景区域扩充 3 个像素作为前景进行计算),从而减少计算复杂度。泊松图像编辑还应用于图像拼接,如文献[58]提出一种对无人机影像的拼接技术,取得了良好效果。

6 结语

与深度学习的三维重建方法相比,基于图像的三维重建算法流程清晰,具有明确的中间结果,是典型的“白盒”模型,不需要预训练神经网络,对各种场景和物体适应性强,可以全流程优化改善,增强实现效果。但也存在计算速度慢、过分依赖于图像特征点,且对弱纹理、朗伯面、纤细物体重建效果不佳等缺点。

鉴于以上不足,以下工作可能取得重要进展:①图像数据与视觉里程计、激光雷达等多传感器信息融合,不再依赖于纯图像,增加信息来源,既提升运算速度,又提高计算精度;②与深度学习方法相结合,实现在没有图像特征点或匹配对很少的情况下,得到相机的相对位置关系,并对弱纹理区域、朗伯面重建具有鲁棒性;③算法软件提升与基于 GPU 的硬件加速相结合,提升算法运行效率,逐步实现大场景下的实时三维重建。

参考文献:

[1] JIN Y W, JIANG D Q, CAI M. 3d reconstruction using deep learning: a survey [J]. Communications in Information and Systems, 2020, 20 (4): 389-413.

[2] QIU J G, ZHANG J G, LI K. Image matching method based on Harris and Sift algorithm [J]. Academic Journal of Test Technology, 2009, 23 (3): 271-274.
邱建国,张建国,李凯. 基于Harris与Sift算法的图像匹配方法[J]. 测试技术学报,2009,23(3):271-274.

[3] YANG J R, ZHANG X Y, LUO P. A review of feature point detection

- methods based on image matching[J]. *Electronic Testing*, 2021(6):53–54,92.
- 杨济瑞,张晓燕,罗攀. 基于图像匹配的特征点检测方法综述[J]. *电子测试*, 2021(6):53–54,92.
- [4] HARRIS C, STEPHENS M. A combined corner and edge detector[C]//Manchester: Alvey Vision Conference, 1988.
- [5] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004,60(2):91–110.
- [6] BAY H, TUYTELAARS T, GOOL L V. SURF: Speeded up robust features[C]//Proceedings of European Conference on Computer Vision, 2006: 1–14.
- [7] HUANG H B, LI X L, XIONG W Y, et al. Overview of image registration based on SIFT algorithm[J]. *Journal of Software Guide*, 2019, 18(1):1–4. 黄海波,李晓玲,熊文怡,等. 基于SIFT算法的图像配准综述[J]. *软件导刊*, 2019, 18(1):1–4.
- [8] ZHANG Z Y. A flexible new technique for camera calibration[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(11): 1330–1334.
- [9] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision[M]. Cambridge: Cambridge University Press, 2003.
- [10] LONGUET-HIGGINS H C. A computer algorithm for reconstructing a scene from two projections[J]. *Nature*, 1981, 293: 133–135.
- [11] HARTLEY R I. In defense of the eight-point algorithm[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(6): 580–593.
- [12] CHESI G, GARULLI A, VICINO A, et al. Estimating the fundamental matrix via constrained least-squares: a convex approach[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(3): 397–401.
- [13] TSAI R Y, HUANG T S. The perspective view of three points[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, 6: 13–27.
- [14] SHAN X, WANG Y M, DONG J P. Matching method based on RANSAC algorithm[J]. *Journal of Shanghai Institute of Electrical Engineering*, 2006, 9(4): 66–69. 单欣,王耀明,董建萍. 基于RANSAC算法的基本矩阵估计的匹配方法[J]. *上海电机学院学报*, 2006, 9(4): 66–69.
- [15] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 381–395.
- [16] WENG J, HUANG T S, AHUJA N. Motion and structure from two perspective views: algorithms, error analysis and error estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(5): 451–476.
- [17] HUANG T S, FAUGERAS O D. Some properties of the E-matrix in two-view motion estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11: 1310–1312.
- [18] NISTER D. An efficient solution to the five-point relative pose problem[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(6): 756–770.
- [19] LUONG Q T, FAUGERAS O D. The fundamental matrix: theory, algorithms, and stability analysis[J]. *International Journal of Computer Vision*, 1996, 17(1): 43–75.
- [20] ROBERT L, FAUGERAS O D. Relative 3D positioning and 3D convex hull computation from a weakly calibrated stereo pair[J]. *Image and Vision Computing*, 1995, 13(3): 189–196.
- [21] FUHRMANN S, LANGGUTH F, GOESELE M. Mve—a multi-view reconstruction environment[C]//Eurographics Workshops on Graphics and Cultural Heritage, 2014: 11–18.
- [22] SNAVELY N. Bundler: Structure from motion (SFM) for unordered image collections[EB/OL]. <http://phototour.cs.washington.edu/bundler/>.
- [23] KNEIP L, SCARAMUZZA D, SIEGWART R. A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation[C]//Conference on Computer Vision and Pattern Recognition, 2011: 2669–2976.
- [24] GAO X S, HOU X R, TANG J, et al. Complete solution classification for the perspective-three-point problem[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(8): 930–940.
- [25] LEPETIT V, MORENO-NOGUER F, FUA P. EPnP: Efficient perspective-n-point camera pose estimation[J]. *International Journal of Computer Vision*, 2009, 81(2): 155–166.
- [26] CHEN H, WANG P, WANG F, et al. EPro-PnP: Generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation[C]//Proceedings of Conference on Computer Vision and Pattern Recognition, 2022: 2781–2790.
- [27] TRIGGS B, MCLAUCHLAN P F, HARTLEY R I. Bundle adjustment—a modern synthesis[C]//International Workshop on Vision Algorithms—Vision Algorithms: Theory and Practice, 1999: 298–372.
- [28] FRANSDEN P E, JONASSON K, NIELSEN H B. Unconstrained Optimization (3rd Edition)[EB/OL]. http://www.imm.dtu.dk/documents/ftp/publlec/lec2_99.pdf.
- [29] MADSEN K, NIELSEN H B, TINGLEFF O. Methods for non-linear least squares problems(2nd Ed.)[EB/OL]. http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3215/pdf/imm3215.pdf.
- [30] LEVENBERG K. A method for the solution of certain non-linear problems in least squares[J]. *Quarterly of Applied Mathematics*, 1944, 2(2): 164–168.
- [31] MARQUARDT D W. An algorithm for least-squares estimation of non-linear parameters[J]. *Journal of the Society for Industrial and Applied Mathematics*, 1963, 11(2): 431–441.
- [32] WU C. Towards linear-time incremental structure from motion[C]//3DV-Conference, 2013 International Conference on IEEE Computer Society, 2013.
- [33] SNAVELY N, SEITZ S M, SZELISKI R. Modeling the world from internet photo collections[J]. *International Journal of Computer Vision*, 2008, 80(2): 189–210.
- [34] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4104–4113.
- [35] YU Y, ZHANG Y S, XUE W, et al. An incremental motion restoration structure method with enhanced robustness and improved accuracy[J]. *Journal of Geodesy Surveying and Mapping*, 2019, 48(2): 207–215. 于英,张永生,薛武,等. 一种稳健性增强和精度提升的增量式运动恢复结构方法[J]. *测绘学报*, 2019, 48(2): 207–215.
- [36] CUI H, SHEN S, GAO W, et al. Efficient large-scale structure from motion by fusing auxiliary imaging information[J]. *IEEE Transactions on Image Processing*, 2015, 24(11): 3561–3573.
- [37] CHATTERJEE A, GOVINDU V M. Efficient and robust large-scale rotation averaging[C]//Sydney: 2013 IEEE International Conference on Computer Vision, 2013.
- [38] FARENZENA M, FUSIELLO A, GHERARDI R. Structure-and-motion

- pipeline on a hierarchical cluster tree [C]//Proceedings of International Conference on Computer Vision Workshop on 3-D Digital Imaging and Modeling, 2009: 1489-1496.
- [39] GHERARDI R, FARENZENA M, FUSIELLO A. Improving the efficiency of hierarchical structure-and-motion [C]//San Francisco: The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, 2010.
- [40] SEITZ S M, DYER C R. Photorealistic scene reconstruction by voxel coloring [J]. International Journal of Computer Vision, 1999, 35(2): 1-23.
- [41] SINHA S N, MORDOHAJ P, POLLEFEYS M. Multi-View Stereo via Graph Cuts on the Dual of an Adaptive Tetrahedral Mesh [C]//Rio de Janeiro: 2007 IEEE 11th International Conference on Computer Vision, 2007.
- [42] VOGIATZIS G, CARLOS H N E, TORR P H S, et al. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(12): 2241-2246.
- [43] GOESELE M, SNAVELY N, CURLESS B. Multi-view stereo for community photo collections [C]//Proceedings of the IEEE International Conference on Computer Vision, 2007: 1-8.
- [44] FURUKAWA Y, J. Accurate Ponce, dense, and robust multi-view stereo [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376.
- [45] KONG Q B, HE L, YUAN L, et al. Improved PMVS 3D reconstruction point cloud filtering algorithm [J]. Computer Applications and Software, 2021(4): 215-219.
孔庆博, 何丽, 袁亮, 等. 改进PMVS三维重建点云滤波算法 [J]. 计算机应用与软件, 2021(4): 215-219.
- [46] LABATUT P, PONS J P, KERIVEN R. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts [C]//Rio de Janeiro: 2007 IEEE 11th International Conference on Computer Vision, 2007.
- [47] VU H H, LABATUT P, J P PONS et al. High accuracy and visibility-consistent dense multiview stereo [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(5): 889-901.
- [48] KOLMOGOROV V, ZABIH R. Multi-camera scene reconstruction via graph cuts [C]//European Conference on Computer Vision (ECCV), 2002.
- [49] FUHRMANN S, GOESELE M. Floating scale surface reconstruction [J]. ACM Transactions on Graphics (ToG), 2014, 33(4): 1-11.
- [50] KAZHDAN M, BOLITHO M, HOPPE H. Poisson surface reconstruction [C]//Proceedings of the fourth Eurographics symposium on Geometry processing, 2006: 61-70.
- [51] LORENSEN W E, CLINE H E. Marching cubes: a high resolution 3D surface construction algorithm [J]. ACM siggraph computer graphics, 1987, 21(4): 163-169.
- [52] KAZHDAN M, KLEIN A, DALAL K, et al. Unconstrained isosurface extraction on arbitrary octrees [C]//Proceedings of the fifth Eurographics Symposium on Geometry processing, 2007: 125-133.
- [53] ZHANG X S, HUA S G. Point cloud surface reconstruction combined with plane projection and regional growth [J]. Mechanical and Electrical Engineering Technology, 2020, 49(6): 23-24.
张晓帅, 华顺刚. 结合平面投影与区域生长的点云表面重建 [J]. 机电工程技术, 2020, 49(6): 23-24.
- [54] PANG Z Y, ZHOU Z F, WANG L D, et al. Improved 3D reconstruction algorithm for point cloud data [J]. Advances in Laser and Optoelectronics, 2020, 57(2): 021102.
庞正雅, 周志峰, 王立端, 等. 改进的点云数据三维重建算法 [J]. 激光与光电子学进展, 2020, 57(2): 1-7.
- [55] LEMPITSKY V, IVANOV D. Seamless mosaicing of image-based texture maps [C]//Minneapolis: Conference on Computer Vision and Pattern Recognition, 2007.
- [56] WAECHTER M, MOEHRLE N, GOESELE M. Let there be color! large-scale texturing of 3d reconstructions [C]//Proceedings of European Conference on Computer Vision, 2014: 836-850.
- [57] P'EREZ P, GANGNET M, BLAKE A. Poisson image editing [J]. ACM Transactions on Graphics, 2003, 22(3): 313-318.
- [58] ZHOU J, XIE K, FU C, et al. Uav remote sensing image Mosaic technology combining SIFT feature points and Poisson fusion [J]. Bulletin of Surveying and Mapping, 2021(1): 94-98.
周杰, 解琨, 付超, 等. 结合SIFT特征点和泊松融合的无人机遥感影像拼接技术 [J]. 测绘通报, 2021(1): 94-98.

(责任编辑: 孙娟)