**《计算机视觉》大作业**

基于单目视觉的三维重建算法

|  |  |  |
| --- | --- | --- |
| **实验小组成员**  **（学号+班级+姓名）** | **分工及主要完成任务** | **成绩** |
|  |  |  |
|  |  |  |
|  |  |  |

山东大学

2021年3月

**三维重建sfm算法**

**摘 要：** 三维重建技术通过深度数据获取、预处理、点云配准与融合、生成表面等过程，把真实场景刻画成符合计算机逻辑表达的数学模型。这种模型可以对如文物保护、游戏开发、建筑设计、临床医学等研究起到辅助的作用。

**关键词：** 三维重建；稀疏重建；sfm算法

1.三维重建简介

1.1相关概念

彩色图像与深度图像：彩色图像也叫作RGB图像，R、G、B三个分量对应于红、绿、蓝三个通道的颜色，它们的叠加组成了图像像素的不同灰度级。RGB颜色空间是构成多彩现实世界的基础。深度图像又被称为距离图像，与灰度图像中像素点存储亮度值不同，其像素点存储的是该点到相机的距离，即深度值。深度值指的目标物体与测量器材之间的距离。由于深度值的大小只与距离有关，而与环境、光线、方向等因素无关，所以深度图像能够真实准确的体现景物的几何深度信息。通过建立物体的空间模型，能够为深层次的计算机视觉应用提供更坚实的基础。

PCL：PCL（Point Cloud Library，点云库）是由斯坦福大学的Dr.Radu等学者基于ROS(Robot Operating System，机器人操作系统)下开发与维护的开源项目，最初被用来辅助机器人传感、认知和驱动等领域的开发。2011年PCL正式向公众开放。随着对三维点云算法的加入与扩充，PCL逐步发展为免费、开源、大规模、跨平台的C++编程库。

PCL框架包括很多先进的算法和典型的数据结构，如滤波、分割、配准、识别、追踪、可视化、模型拟合、表面重建等诸多功能。能够在各种操作系统和大部分嵌入式系统上运行，具有较强的软件可移植性。鉴于PCL的应用范围非常广，专家学者们对点云库的更新维护也非常及时。PCL的发展时至今日，已经来到了1.7.0版本。相较于早期的版本，加入了更多新鲜、实用、有趣的功能，为点云数据的利用提供了模块化、标准化的解决方案。再通过诸如图形处理器、共享存储并行编程、统一计算设备架构等领先的高性能技术，提升PCL相关进程的速率，实现实时性的应用开发。

在算法方面，PCL是一套包括数据滤波、点云配准、表面生成、图像分割和定位搜索等一系列处理点云数据的算法。基于不同类型区分每一套算法，以此把整合所有三维重建流水线功能，保证每套算法的紧凑性、可重用性与可执行性。此类常用的算法模块均具有回归测试功能，以确保使用过程中没有引进错误。测试一般由专门的机构负责编写用例库。检测到回归错误时，会立即将消息反馈给相应的作者。因此能提升PCL和整个系统的安全稳定性

点云数据：点云数据通常出现在逆向工程中，是由测距设备获取的物体表面的信息集合。其扫描资料以点的形式进行记录，这些点既可以是三维坐标，也可以是颜色或者光照强度等信息。通常所使用的点云数据一般包括点坐标精度、空间分辨率和表面法向量等内容。点云一般以PCD格式进行保存，这种格式的点云数据可操作性较强，同时能够提高点云配准融合的速度。本文研究的点云数据为非结构化的散乱点云，属于三维重建特有的点云特点。

坐标系：

图像坐标系分为像素和物理两个坐标系种类。数字图像的信息以矩阵形式存储，即一副像素的图像数据存储在维矩阵中。图像像素坐标系以为原点、以像素为基本单位，U、V分别为水平、垂直方向轴。图像物理坐标系以摄像机光轴与图像平面的交点作为原点、以米或毫米为基本单位，其X、Y轴分别与U、V轴平行。

摄像机坐标系由摄像机的光心及三条、、轴所构成。它的、轴对应平行于图像物理坐标系中的、轴，轴为摄像机的光轴，并与由原点、、轴所组成的平面垂直。

世界坐标系是考虑到摄像机位置具有不确定性，因此有必要采用世界坐标系来统一摄像机和物体的坐标关系。世界坐标系由原点及x、y、z三条轴组成。

1.2三维重建流程

三维计算机视觉在计算机视觉是偏基础的方向，随着2010年阿凡达在全球热映以来，三维计算机视觉的应用从传统工业领域逐渐走向生活、娱乐、服务等，比如AR/VR，SLAM，自动驾驶等都离不开三维视觉的技术。

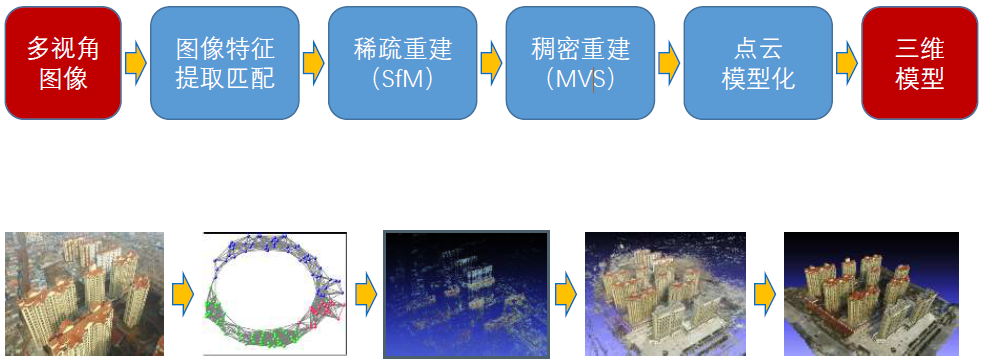
三维重建包含三个方面，基于SFM的运动恢复结构，基于Deep learning的深度估计和结构重建，以及基于RGB-D深度摄像头的三维重建。

SfM（Structure From Motion），主要基于多视觉几何原理，用于从运动中实现3D重建，也就是从无时间序列的2D图像中推算三维信息，是计算机视觉学科的重要分支。广泛应用于AR/VR，自动驾驶等领域。虽然SFM主要基于多视觉几何原理，随着CNN的在二维图像的积累，很多基于CNN的2D深度估计取得一定效果，用CNN探索三维重建也是不断深入的课题。

深度学习方法呈现上升趋势，但是传统基于多视几何方法热情不减，实际应用以多视几何为主，深度学习的方法离实用还有一定的距离。

本文主要介绍基于单目monocular的三维重建方法，SfM三维重建。

三维重建基本流程：

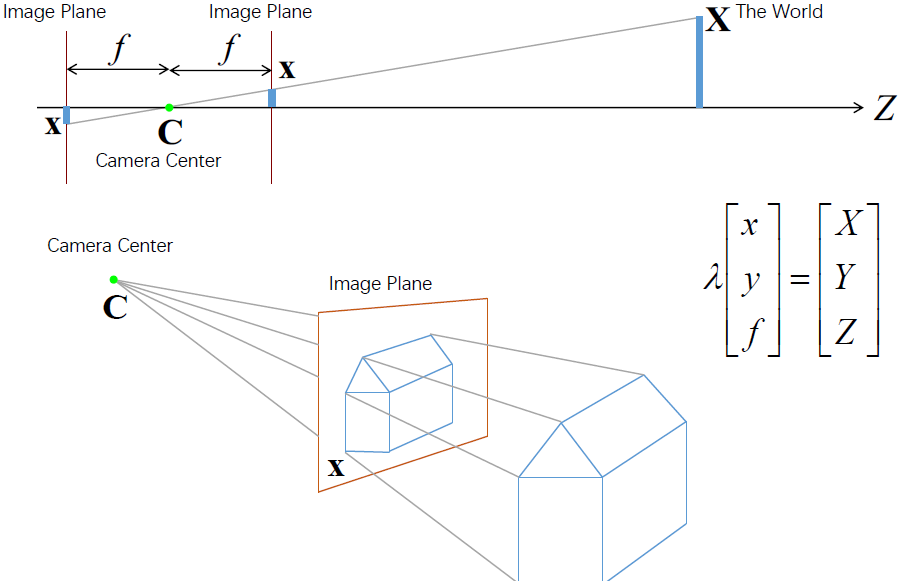


2.sfm与三维重建

2.1小孔相机模型

在计算机视觉中，最常用的相机模型就是小孔模型（小孔成像模型），它将相机的透镜组简化为一个小孔，光线透过小孔在小孔后方的像面上成像。小孔模型成的是倒像，为了表述与研究的方便，我们常常将像面至于小孔之前，且到小孔的距离仍然是焦距f，这样的模型与原来的小孔模型是等价的，只不过成的是正像，符合人的直观感受。在这种情况下，往往将小孔称作光心（Optical Center）。

小孔模型是一种理想相机模型，没有考虑实际相机中存在的场曲、畸变等问题。在实际使用时，这些问题可以通过在标定的过程中引入畸变参数解决，所以小孔模型仍然是目前最广泛使用的相机模型。

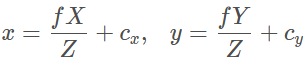


2.2内外参矩阵

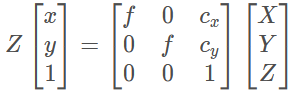
设空间中有一点P，若世界坐标系与相机坐标系重合，则该点在空间中的坐标为(X, Y, Z)，其中Z为该点到相机光心的垂直距离。设该点在像面上的像为点p，像素坐标为(x, y)，那么(X, Y, Z)和(x, y)是一个简单的相似三角形关系，会得到：



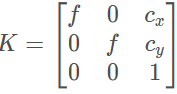
但是，图像的像素坐标系原点在左上角，而上面公式假定原点在图像中心，为了处理这一偏移，设光心在图像上对应的像素坐标为 ( c x , c y )，则



有



其中矩阵



称为内参矩阵，因为它只和相机自身的内部参数有关（焦距，光心位置）。

外参矩阵：一般情况下，世界坐标系和相机坐标系不重合，这时，世界坐标系中的某一点P要投影到像面上时，先要将该点的坐标转换到相机坐标系下。设P在世界坐标系中的坐标为X，P到光心的垂直距离为s（即上文中的Z），在像面上的坐标为x，世界坐标系与相机坐标系之间的相对旋转为矩阵R（R是一个三行三列的旋转矩阵），相对位移为向量T（三行一列），则



其中[R T] 是一个三行四列的矩阵，称为外参矩阵，它和相机的参数无关，只与相机在世界坐标系中的位置有关。

相机的标定，即为通过某个已知的目标，求取相机内参矩阵的过程。最常用的标定目标就是棋盘格。用相机对棋盘格从不同角度拍摄多张照片，然后将这些照片导入标定程序或算法，即可自动求出相机的内参。

2.3特征点检测与匹配

特征的检测和匹配在许多计算机视觉应用中是一个重要的组成部分，例如无缝拼接，三维重建等。其中兴趣点特征是很重要的一类特征，而目前应用最广泛的兴趣点特征检测方法就是SIFT检测算法，该检测算法所得到的特征点不仅在位置上能够稳定识别，而且具有尺度不变性和旋转不变性。

SIFT特征不只具有尺度不变性，即使改变旋转角度，图像亮度或拍摄视角，仍然能够得到好的检测效果。由于SIFT算法有很多操作，会令人不解，算法可以分为以下几个部分：

1. 构建尺度空间

这是一个初始化操作，通过生成尺度空间来创建原始图像的多层表示以保证尺度不变性。

2. LoG近似

使用Laplacian of Gaussian能够很好地找到找到图像中的兴趣点，但是需要大量的计算量，所以使用创建更容易的表示来近似它。

3. 找到关键点

利用很快的近似，我们可以找到特征点，它们是Difference of Gaussian图像的极大极小值。

4. 除去不好的特征点

边界和低亮度区域是不好的特征点，除去它们以使得算法有效和鲁棒，在这里使用近似Harris Corner检测器。

5. 给特征点赋值一个方向

为每个特征点计算一个方向，依照这个方向做进一步的计算，这个操作有效地取消了方向的影响，使得算法具有旋转不变性。

6. 生成SIFT特征

最后，利用位置上的尺度和旋转不变性，能够生成一个表示，它能帮助唯一地识别特征。通过这个表示，我们可以很容易识别寻找的特征。

2.3捆绑调整BA

Bundle Adjustment译为光束法平差或者束调整、捆绑调整。BA问题主要是对三维点位置和相机参数进行非线性优化，而我们可以把BA问题看成最小化重投影误差问题，同时这也是一个非线性最小二乘问题。

由前面可知，我们可以通过特征点检测与匹配，得到多对特征匹配点，使用对极几何可以求得相机的位姿。我们可以通过空间点P的坐标求得空间点P投影在图像中的像素坐标，但是这个像素点坐标是先通过位姿估计，再通过公式计算出来的，相当于这个像素坐标是估算出来的。这个过程称为重投影。如果完全没有误差，那么两者的坐标是一样的，但这是不可能的，不管是实际测量出来的像素坐标还是通过位姿估算+重投影计算出来的像素坐标，都是有误差的，将实际值和估算值一减，就得到了重投影误差。

由于两张图片中有很多特征匹配点，将所有特征点重投影误差求和，取平方，再乘以1/2，这就构建了一个非线性最小二乘问题了，将里面的位姿、空间点作为优化的对象，最小化重投影误差，即得到一个优化后的结果，这就是所谓的BA。

BA问题被归结为一个非常大的非线性最小二乘问题，通常用Levenberg-Marquardt(LM)算法求解。

2.5稀疏重建

Structure from Motion（SfM）是一个估计相机参数及三维点位置的问题。一个基本的SfM pipeline可以描述为:对每张2维图片检测特征点（feature point），对每对图片中的特征点进行匹配，只保留满足几何约束的匹配，最后执行一个迭代式的、鲁棒的SfM方法来恢复摄像机的内参（intrinsic parameter）和外参(extrinsic parameter)。并由三角化得到三维点坐标，然后使用Bundle Adjustment进行优化。

根据SfM过程中图像添加顺序的拓扑结构，SfM方法可以分为:

·增量式（incremental/sequential SfM）

·全局式（global SfM）

·混合式（hybrid SfM）

增量式SfM首先使用SIFT特征检测器提取特征点并计算特征点对应的描述子（descriptor），然后使用ANN（approximate nearest neighbor）方法进行匹配，低于某个匹配数阈值的匹配对将会被移除。对于保留下来的匹配对，使用RANSAC（RANdom Sample Consensus）和八点法来估计基本矩阵（fundamental matrix），在估计基本矩阵时被判定为外点（outlier）的匹配被看作是错误的匹配而被移除。对于满足以上几何约束的匹配对，将被合并为tracks。然后通过incremental方式的SfM方法来恢复场景结构。首先需要选择一对好的初始匹配对，一对好的初始匹配对应该满足：

（1）足够多的匹配点；

（2）宽基线。之后增量式地增加摄像机，估计摄像机的内外参并由三角化得到三维点坐标，然后使用Bundle Adjustment进行优化。

增量式SfM从无序图像集合计算三维重建的常用方法。增量式SfM可分为几个阶段：图像特征提取、特征匹配、几何约束、重建初始化、图像注册、三角化、outlier过滤、Bundle adjustment等步骤。

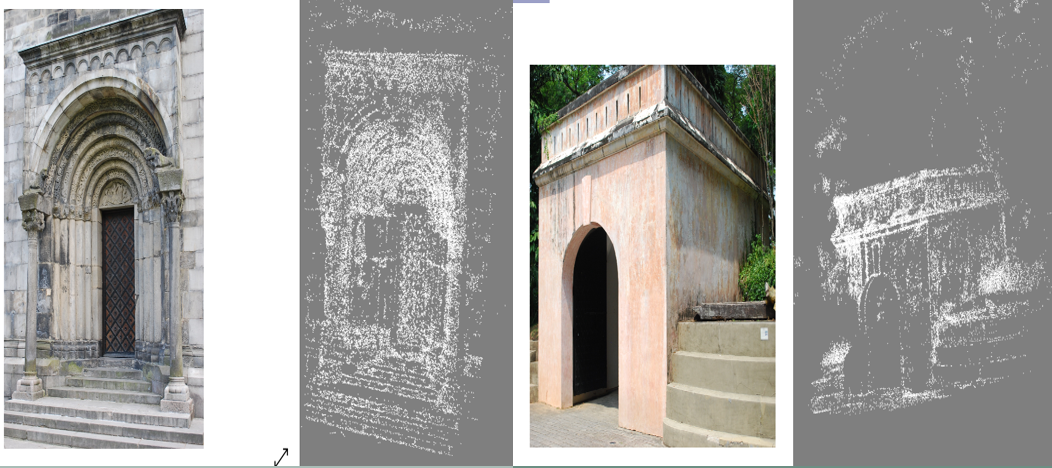
全局式sfm估计所有摄像机的旋转矩阵和位置并三角化初始场景点。

优势：将误差均匀分布在外极几何图上，没有累计误差。不需要考虑初始图像和图像添加顺序的问题。仅执行一次捆绑调整，重建效率高。

缺点：鲁棒性不足，旋转矩阵求解时L1范数对外点相对鲁棒，而摄像机位置求解时相对平移关系对匹配外点比较敏感。场景完整性，过滤外极几何边，可能丢失部分图像。

混合式SfM在一定程度上综合了incremental SfM和global SfM各自的优点。HSfM的整个pipeline可以概括为全局估计摄像机旋转矩阵，增量估计摄像机位置，三角化初始场景点。用全局的方式提出一种基于社区的旋转误差平均法，该方法既考虑了对极几何的精度又考虑了成对几何的精度。基于已经估计的相机的绝对旋转姿态，用一种增量的方式估计相机光心位置。对每个添加的相机，其旋转和内参保持不变，同时使用改进的BA细化光心和场景结构。

3.sfm实现效果



山大图书馆稀疏重建：

