# 三维重建的一般方法以及改进

# 3.1引言

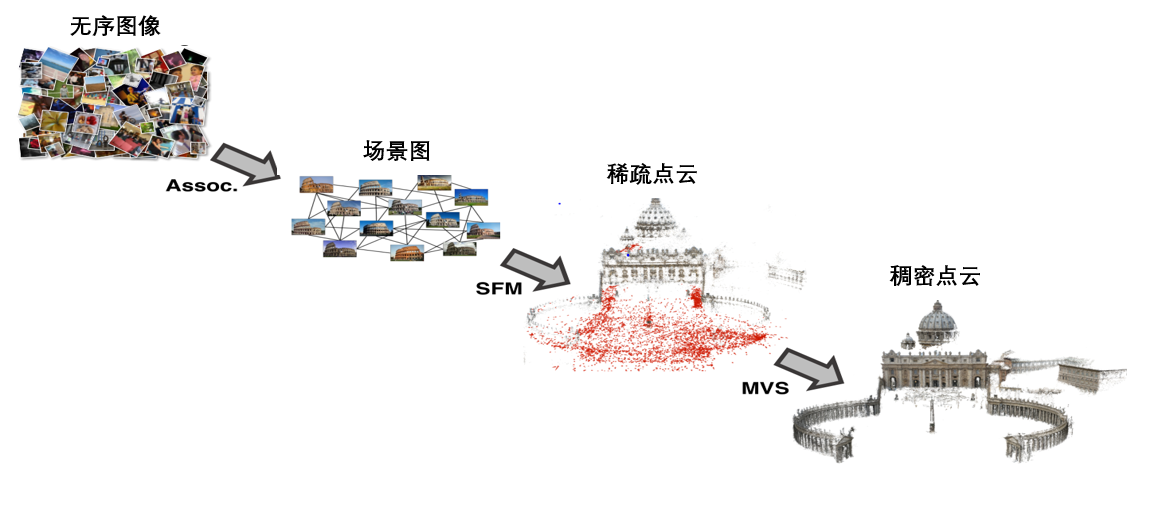
根据图3生成的地图，通过对地图的解析，可以获取每一个二维码的三维位置坐标，如表1所示。

在第二章，本文提出了一种结合二维码的SLAM视觉系统，利用该系统可以

# 3.2三维重建的一般方法

## 3.2.1一般方法概述

三维重建是指从三维图像中复原三维场景或者物体的过程，整个流程的输入为无序的图片即可，输出可以得到三维重建后的稀疏点云和稠密点，大致流程如图所示。



对于以上流程进一步进行细化，整个三维重建的过程可以划分为以下几个主要的步骤：

·1）2D图像采集：多角度拍摄或者从视频中提取到一组图像序列，将图像序列作为整个系统的输入。

2）特征点提取和匹配：根据拍摄到的图像，提取每张图像之间的特征点，并进行特征点的匹配。

3）稀疏点云：根据匹配结果估计特征点的深度，提取出稀疏点云，并估计相机的位姿和参数。

4）稠密点云：根据优化后的相机参数和匹配结果，获得稠密点云

5）纹理映射：根据以上点重建物体表面，进行纹理映射。

三维重建的一般步骤可以简化为如图~\ref{fig:3Dconstr\_pipiline}所示。

三维重建主要分为多个步骤，目前很多开源的系统都可以完成其中的部分环节，对于完整的三维重建流程还需要多个系统相互连接实现，表是当前三维重建系统的简要对比。本文在考虑到各个系统流程的完整性和合理性，以及在实际测试过系统之间的效果差异后，选择Colmap作为作为三维重建的工具。

## 3.2.2三维重建详细过程解析

3.2.2.1 获取图像

目前常规的三维重建仅需要输入无序图片，在构图的过程中，可以极大地降低操作的复杂度，并且对于相机的内参和外参也无需提前提供给整个三维重建的系统，在特征点匹配的过程中，这些参数都可以通过计算得到。对于输入的图像，可以通过随着时间流单帧拍摄的方式获取或者通过截取视频流的方式得到。在获取图像的过程中，需要注意保证每连续两帧图像之间尽可能保证30%的重叠区域，相邻两帧之间的旋转角度为30度到45度之间，且物体中的每一个点至少能被三帧图像观测到，对于场景，合理的图像获取方式如图所示。

3.2.2.2 特征点检测和匹配

目前，在三维重建领域中图像的特征匹配就是以特征点为基础而进行的，所以，如何定义和找出一幅图像中的特征点就非常重要。常见的特定点检测和匹配主要包括SUFT，SIFT，ORB，FAST，Harris角点等，

SUFT：解决特征检测中的尺度不变性问题，具备较高的计算效率。

SIFT：以SUFT为基础，基于浮点内核计算特征点，有着更加精确的空间位置和尺度。

ORB：满足实时性的速度，但是不具备旋转，尺度不变性且噪声敏感

Harris角点：能够较好的检测角点，进行精确的定位。

对于本文中的三维重建，选择SIFT作为特征点的检测和匹配方法，SIFT和其他方法相比较，有以下优点：

1）SIFT特征是图像的局部特征，其对旋转、尺度缩放、亮度变化保持不变性，对视角变化、仿射变换、噪声也保持一定程度的稳定性；

2. 独特性好，信息量丰富，适用于在海量特征数据库中进行快速、准确的匹配；

3. 多量性，即使少数的几个物体也可以产生大量的SIFT特征向量；

针对其计算耗时的问题，本文选择CUDA进行硬件加速，另外还考虑到三维重建本身是一个离线处理的过程，对算法的实时性没有过高要求。

在获取到每张图像上的特征点后，需要对图像之间建立匹配关系，常用的方式可以采用计算欧式距离的办法：1）暴力匹配，对所有的特征点都进行穷举，计算其对应距离。2）邻近搜索,建立KD树,缩小搜索范围,能提高效率,但也有可能不是最优,所以邻域取值是关键,越大越准确,越大计算量越大

由图所示，两帧之间大多数特征点都可以正确匹配，但依然存在存在部分匹配是错误的，在本文中，选择了RANSAC（随机抽样一致性）的方式来剔除错误的匹配对，以更加准确的估计相机位姿，RANSAC是指可以从一组包括局外点（错误匹配）的观测数据中，通过迭代的方式估计数学模式中的参数。

RANSAC的基本假设是：

（1）数据由“局内点”组成，例如：数据的分布可以用一些模型参数来解释；

（2）“局外点”是不能适应该模型的数据；

（3）除此之外的数据属于噪声。

局外点产生的原因有：噪声的极值；错误的测量方法；对数据的错误假设。

RANSAC也做了以下假设：给定一组（通常很小的）局内点，存在一个可以估计模型参数的过程；而该模型能够解释或者适用于局内点。

3.2.2.3 SFM（运动恢复结构）

在上一小结中可以获得初始的匹配关系，但是这种匹配关系不完全可靠，需要添加几何约束进行检测，该集合约束完全依赖于场景中的客观事实。可以通过基本矩阵F将匹配好的两帧图像之间的像素坐标进行关联，假设一个符合条件的匹配对像素坐标需要满足以下公式：

找到相机基线最大的像对，根据该像对，通过RANSC八点法计算本征矩阵，再通过对本征矩阵SVD分解得到第二个图像的**R、T**，在这一步需要进行畸变校正，然后根据R、T和矫正后的像点坐标三角计算出三维点。

当所有的两两匹配图像对被确定以后，可以开始计算相机的位姿（3\*3的旋转矩阵R，1\*3的平移向量t），摄像机的内参（焦距f，畸变参数k1，k2）。几何场景提供轨迹中的每个3D点Xj，通过投影方程，将3D点投影到摄像机的2D成像平面上，投影误差的定义为投影点和图像上真实点之间的欧式距离。如下图：

对于n个视角和m个轨迹，投影误差的目标优化方程可以写为:

。SFM算法的目标就是找到合适的相机和场景参数去优化这个目标函数，g是采用一个非线性最小二乘的优化方法求解，常用BA（光束平差法）来优化上述过程。

最后，不断添加新的摄像机和3D点进行BA。这个过程直到剩下的摄像机观察到的点不超过20为止，说明剩下的摄像机没有足够的点可以添加，BA结束。得到相机估计参数和场景几何信息，即稀疏的3D点云。

## 3.2.3当前三维重建存在的问题

根据3.2节的描述，可以发现通过现有的三维重建技术流程可以实现有多张连续图像到三维点云的转化，整个过程操作起来十分简便，且能够得到一个较好的结果。但是从整体实验速度，三维点云结果的精确性来看，依然还存在很多的问题。

1. 整个三维重建的过程十分耗时，由于三维重建中的图像时间没有时序信息，关键帧之前的匹配都通过暴力枚举的方式进行匹配，对于输入N张原始图像的系统，时间复杂度高达O（N2），由单帧图像到稀疏点云一般耗时为几分钟，由单帧图像到稠密点云耗时耗时更是会高度几个小时。
2. 三维点云的精度不高，噪音点较大，一方面是由于三维重建往往采取增量式的重建方式，导致最终场景重建的结果无法闭合；另外一方面也是由于对于某些重复度较高的场景，关联帧之间的匹配由于缺少时序信息，从而导致匹配精度较低，在解算相机位姿时，得到的结果也无法保证正确。

4 解决方案

1. 本方案的建图过程分为两个阶段：分别为在线SLAM阶段和离线3D重建阶段。
2. 在线SLAM系统实时获取前视单目相机，基于多通道信息融合SLAM技术快速建立scene graph及点云环境地图。考虑到在线SLAM为了满足实时性要求，只是对局部窗口进行增量优化，即使在闭环时也只进行轨迹和部分点云优化，因此所建立的scene graph及点云环境地图的精度较低。随着时间的推移不可避免的会出现误差累计，并且在闭环处出现场景无法闭合的情况。因此，现在仅通过SLAM获取KeyFrame DataBase，包含关键帧图像、关键帧之间的相互关系信息，来为后续的3D重建准备输入。
3. 离线3D重建以提高建图精度为目标，包括SfM和MVS两个主要环节，其流程如图5.2所示。
4. 
5. 图 5.2 离线3D重建技术流程图
6. 首先基于在线SLAM结果建立粗糙Scene Graph及点云环境地图，以全局大尺度地图及运动轨迹为优化目标，使用鲁棒性更高的特征描述算子，利用高性能计算机的快速处理能力进行Global Bundle Adjustment全局优化，得到更加精确的全局稀疏地图和运动轨迹。然后利用MVS（Multi View Stereo）技术对图像中的高梯度变化区域建立半稠密环境地图。