



西北工业大学

本科毕业设计论文

题目 面向室内环境的三维重建及相机位姿估计研究

专业名称 计算机科学与技术

学生姓名 范洁

指导教师 杨涛

毕业时间 2013年6月

摘要

真实环境的三维重建一直以来都是计算机视觉领域的一个重要课题,有着广泛的应用前景。本文针对室内环境三维场景重建和相机位姿估计方法进行研究。主要包括:基于并行跟踪与地图构建系统的相机位姿估计研究、基于 Kinect 的单帧三维点云稠密重建以及基于 RGB-D 数据的全局环境三维重建。论文的主要工作及创新点如下:

- (1) 针对适用于小型工作空间的并行跟踪与地图构建算法进行了详细介绍,并利用该算法系统输出相机内、外部参数,对移动平台进行位姿估计,提出了用比较的方法来估计其位姿估计的精度;
- (2) 针对 Kinect 的单帧三维点云稠密重建问题,提出通过 Kinect 在得到彩色图像的同时得到深度信息,并使用深度信息计算得到场景的三维点云图,再根据每个像素点对应的彩色信息对其进行着色,进而得到真实环境的三维重建;
- (3) 针对 RGB-D 数据的三维全局场景重建问题,根据最流行的 RGB-D SLAM 算法在完成构建完整三维地图时遇到的问题,提出了三个方面的改进。一是将 SIFT 算法替换为 SURF 算法进行特征提取和匹配,提高了此阶段的效率;二是使多层改进子地图划分标准图抽象可以更好的反映环境的拓扑结构,并具有更高的效率;三是在地图输出阶段添加冗余点去除模块,使得最后得到的地图数据量减少,同时为后续操作提供更准确的数据。结果证明该改进在建立三维地图时是完全有效的。

关键词: 三维重建, 并行跟踪与地图构建, 点云, 位姿估计, 深度信息

ABSTRACT

Three-dimensional reconstruction of real environment is an important topic in the field of computer vision, which has wide application prospects. In this dissertation, the research focused on 3D scene reconstruction of indoor environment and pose estimation of the camera, including: camera pose estimation based on parallel tracking and mapping system (PTAM), dense reconstruction based on the Kinect 3D point cloud and three-dimensional reconstruction of indoor environment based on RGB-D sensor. The main contributions are summarized as follows:

- (1) Described the entire process of PTAM algorithm for small workspace in detail, and using the PTAM algorithm to get the parameters and the pose of the camera, and estimation accuracy using comparative methods ;
- (2) To reconstruct the environment using 3D point cloud of Kinect, the Kinect's color image and the depth of information can be obtained at the same time, we can use the depth information to compute 3D point cloud scene and finish its dense reconstruction, and then, according to the color information of each correspond pixel, the reconstruction of the scene has been colored, getting the real three-dimensional reconstruction of the environment.
- (3) For the 3D environment reconstruction with RGB-D sensor, use the famous RGB-D SLAM algorithm to build a complete 3D map, and then propose three improvements. First, use SURF instead of SIFT to finish feature extraction and matching, which can improve the speed in this stage; Second, advance the submap dividing method, which makes the multi-graph abstraction can better reflect the topology of the environment, and become more efficient; Third, add a redundant point removing module in the map output step, which can reduce the resulting map data and provide more accurate data for the follow-up operations to. What's more, the results demonstrate the establishment of three improvements for in indoor 3D map building is completely valid.

Keywords: three-dimensional reconstruction, PTAM, point cloud, camera pose estimation, depth information

目 录

摘要.....	I
ABSTRACT.....	II
第一章 绪论.....	1
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	2
1.3 三维重建的研究关键点及难点.....	5
1.4 章节安排.....	5
第二章 基于 PTAM 的相机位姿估计算法.....	7
2.1 引言.....	7
2.2 PTAM 的相关预备知识.....	7
2.2.1 摄像机模型.....	8
2.2.2 坐标变换.....	9
2.2.3 摄像机内外参标定.....	10
2.2.4 图像金字塔.....	11
2.3 并行跟踪与地图重建算法.....	12
2.3.1 地图的表示.....	13
2.3.2 跟踪系统.....	14
2.3.3 地图的构建系统.....	16
2.3.4 理论结果.....	19
第三章 基于 PTAM 的实时移动平台的位置姿态估计.....	22
3.1 引言.....	22
3.2 移动平台的搭建.....	22
3.3 数据采集与预处理.....	23
3.4 基于移动平台的位姿估计及精度评估.....	25
3.5 本章小结.....	30
第四章 基于 RGB-D 传感器融合的移动平台室内三维重建.....	31
4.1 引言.....	31
4.2 Kinect 基本工作原理.....	31
4.3 Kinect 数据采集与预处理.....	34
4.4 基于 PTAM 位姿估计及 Kinect 的室内局部三维重建.....	36

西北工业大学 本科毕业设计论文

4.5	基于 RGB-D 多传感器融合的三维重建系统.....	39
4.5.1	硬件系统.....	39
4.5.2	软件系统.....	39
4.6	三维场景重建结果.....	43
4.7	实验结果分析.....	44
4.8	本章小结.....	44
第五章	结束语.....	46
5.1	工作总结.....	46
5.2	展望.....	47
参考文献	48
致谢	52
毕业设计小结	53
学术论文、专利、科研项目与获奖	54

第一章 绪论

1.1 研究背景与意义

近年来，有关移动机器人领域的研究非常活跃，并且发展迅速，这主要是由于两个方面的原因：第一，智能移动机器人应用范围广泛，包括制造业、服务业以及国防军事工业等诸多领域；第二，深空探测和海洋开发两大高尖端科技领域的快速发展也是促使移动机器人发展的客观因素和潜在动力。

在当前的实际的应用中，移动机器人的工作空间信息一般都是未知的，当派遣移动机器人进入工作环境时，就需要对其工作环境进行准确而有效地探测，并构建出完整的工作环境空间地图。移动机器人只有在准确的了解了所处的环境基础上才能进行切实可行的导航、避障、路径规划。所以，未知环境中移动机器人环境地图创建正成为移动机器人研究领域中的热点问题，受到众多的研究人员的关注。

环境地图能够提供移动机器人的路径规划和运动控制。随着移动机器人地图创建技术的不断发展，出现了许多在实际应用中安全可靠的环境地图创建方法。但是，至今为止大多数的研究成果都是面向二维的环境探测和地图创建方法，关于三维地图创建的研究工作和论文数量还很少。图 1-1 为在 2005 年美国国防部举办的 DARPA 挑战赛中获胜的斯坦福大学无人驾驶车 Stanley 的三维环境感知系统和创建的三维环境地图。

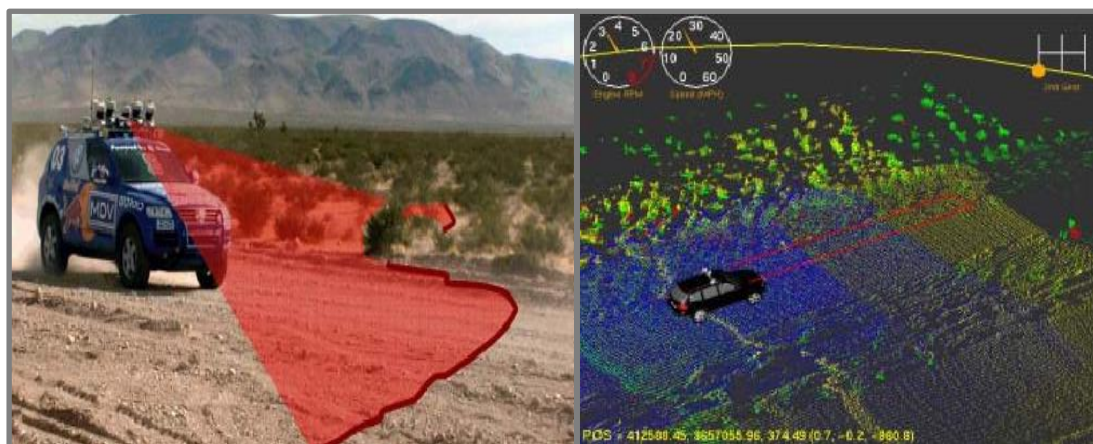


图 1-1 美国斯坦福大学无人驾驶车 Stanley 的环境信息获取示意图和重建三维环境地图

计算机视觉的研究目标是使计算机具有通过二维图像认知三维环境信息的能力。如何根据二维图像获得三维信息是计算机视觉研究的重要目的。室内环境是人类生活必不可少的重要活动场所，人类日常大部分活动都是在室内进行的。因此计算机的室内环境三维重建^{[1][2]}具有重要研究意义和应用价值，可以广泛应用于机器人室内导航、室内装潢设计、视觉监控以及计算机三维动画等多个领域。

目前，国内外在三维重建方面的研究主要集中在两个方面：1、基于声光测距装置的方法，直接测量出物体表面的三维坐标；2、基于计算机视觉的方法，利用摄像机成像模型计算出物体的三维结构。后者由于具有低成本、使用方便等特点，已得到越来越广泛的研究。

1.2 国内外研究现状

视觉是最接近于人类的感知方式，能够提供丰富的环境信息和特征，为应用现代数字技术解决机器人定位等问题提供了良好的信息来源。美国国家航空航天局(National Aeronautics and Space Administration, NASA)自主研发的“机遇号”和“勇气号”火星车，都采用了视觉里程计。

60年代，Roberts^[3]通过计算机程序从数字图像中提取出诸如立方体、棱柱体等多面体的三维结构，并对物体形状及物体的空间关系进行描述。Robert的研究工作开创了以理解三维场景为目的的三维机器视觉的研究。

70年代中期，麻省理工学院(MIT)人工智能(AI)实验室吸引了国际上许多知名学者参与机器视觉的理论、算法、系统设计的研究。其中的 David Marr 教授，在 1977 年提出了不同于“积木世界”分析方法的计算视觉理论^[4]，该理论在 80 年代成为计算机视觉研究领域中的一个十分重要的理论框架。Marr 认为视觉过程是一个信息处理过程，相应的研究应分为三个层次：计算理论层次，表示与算法层次，硬件实现层次^{[5][6]}。Marr 理论强调了当时并不受人重视的计算理论层次，他把立体视觉在这一层次所应完成的主要任务规定为定量地恢复出图像所反映的场景中的三维物体的形状和空间位置，并将这一恢复过程进一步分成如下前、中、后三个视觉处理阶段，如图 1-2 所示。

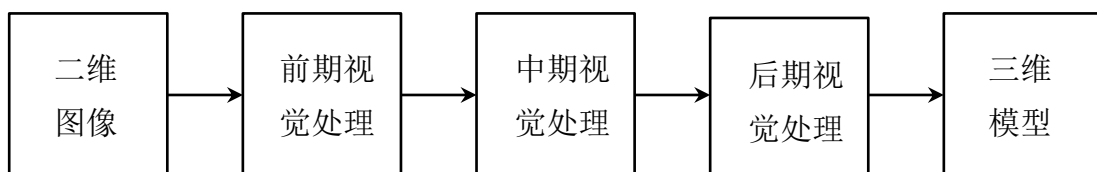


图 1-2 Marr 视觉理论框架

三维重建是计算机视觉研究领域中的一个重要分支,也是用计算机模拟人眼视觉功能所需要完成的至关重要的一步,即恢复物体的三维信息。国内外学者在这方面已做了大量的工作。从应用目的看,这方面的工作分为两大类:一种是获得视野范围内的立体信息,如同人的视觉,这在工业机器人视觉、军事侦察、地质等领域有着广泛应用;另一种是以所摄物体为中心,获得物体表面全方位立体信息,这主要是用于机械制造、影视、广告业中。本文内容主要是针对第一种应用而提出的,关于第二种应用不在本文研究之列,这里不进行过多的描述。

目前,基于图像重建模型的方法主要有以下几种:单幅图像重建方法,立体视觉方法^[7],未定标图像法,侧影轮廓法,深度图像重建的方法。

1) 基于单幅图像重建模型

由单幅图像恢复物体的三维模型主要是依据图像中包含的光照信息、纹理及焦距等。目前,针对单幅图像,可以通过编码结构光照明主动视觉技术,在欧式空间中进行重建;也可以根据图像中的灰度信息求解图像中各像素点的表面法矢信息,进而求得各点的深度信息;也可以通过人机交互的方式对相关约束条件进行指定,或根据人的已知经验获取信息,可以有效提高像素点的坐标精度,有利于重建高精度模型。

单幅图像重建方法对于目标物体有着严格的要求,包括物体的形状、大小、曝光程度及对光照的反射属性等,因此,只适用于一些特殊的场景。为了解决重建过程中的一些固有缺陷,理想的方法是引入知识库或加入人为操作,由用户交互完成一些指定工作,即容易获得目标物体的逼近模型。

2) 立体视觉法

立体视觉方法主要是在两幅图像中寻找对应点,再由对应点关系求解物体表面该点的实际坐标。立体视觉应用了三角测量原理:首先获取两幅已定标图像,且这两幅图像中应存在重叠区域,即真实物体表面的某一部分在两幅图像中均产生投影。然后在两幅图像中寻找一对对应点。连接两幅图像的投影点及该对应点,产生的两条直线将在空间中某处相交,则该交点为物体表面对应点的位置。若能获得物体表面所有点的坐标信息,即可生成物体的整体坐标,便于进行最后的模型建立。

对应点的寻找时立体视觉方法中的重要部分,常用的匹配特征有点、线、面,匹配算法有区域匹配、特征匹配和相位匹配方法。根据匹配影像数可分为:双视

匹配^[8]、三视匹配和多视匹配^[9]。目前，主要通过引入极限约束提高匹配效率和精度。

3) 基于未定标图像重建方法

一般的重建方法均需对相机进行标定，以确定相机的内外参数。近年来，未定标图像重建方法越来越受到人们的青睐，希望通过未定标图像序列恢复场景模型及相机的相关参数。预定标工作的减少或取消使得三维重建更加方便：首先，在获取图像的过程中可以根据具体情况移动相机、调整焦距，而且这些改动并不会影响重建结果。其次，对于一些无法标定的图像或视频，也可以用来重建模型。目前，对于未定标图像的重建主要按照几何空间层次(射影空间、仿射空间、度量空间)逐级重建，且每个层次的重建工作独立进行、互不影响。

4) 基于侧影轮廓线重建方法

物体在图像上的侧影轮廓线是理解物体几何形状的一条重要线索。从不同的投影点对同一物体进行观察的时候，在每个视角对应该图像上都会得到该物体的一条轮廓线。这条轮廓线与该视角下的投影点共同构成了一个锥体，多个视角下的锥体相交就构成了该物体的空间包络，这个空间成为物体的可见外壳。当观察视角足够多的时候，形成的锥体也足够多，这些锥体相交产生的可见外壳可认为是该物体的一个逼近。由于该算法需要多个视角下获得图像，即相机需要多次移动，图像的定标工作就变得非常复杂、非常重要，现有的工作一般使用机械转台等装置辅助定标过程。

5) 基于深度图像重建方法

深度图像中存储了矩阵形式的像素点，每个像素点包含了物体表面该对应点的深度信息。通常，可以通过结构光、激光计量器、声纳及计算机视觉的方法获取深度图像。在获取深度图像过程中，物体表面容易产生遮挡，无法通过单幅图像获得物体的完整信息，所以需要在多个视角获取深度图像，因此，一般需要将多幅深度图像合并到一起，这样才能获得物体的完整信息，并进一步将物体完整的表示出来。利用深度图像对物体进行重建的关键步骤是图像配准及曲面表示。由于深度图像包含了物体的三维几何信息，在重建时可以提高模型重建的精度，因此对一些精度和复杂度要求较高的物体进行重建时有更好的效果。

1.3 三维重建的研究关键点及难点

基于视觉的三维重建问题研究难点主要有以下几点^[10-13]:

1) 图像特征的提取

图像特征的提取主要包括：特征点的检测和特征点的描述。

特征点的检测：从某个尺度或角度上看，被监测到的特征点具有丰富且显著的结构信息，这些信息稳定、易于匹配，但可能没有实际直观视觉上的意义。通常情况下，取图像灰度变化的局部极值点作为特征点，如边界、角点、轮廓等。

特征点的描述：即用特征描述符来表示特征点，是一种特征向量。不同的算法对应不同的描述符，一个特征可以用多个描述符表示。从而特征描述符组成特征空间，特征空间决定了图像中哪些点参与匹配，哪些点被忽略。

常用特征算子有：Harris 算子、Moravec 算子、Forstner 算子、SIFT 算子等。

2) 图像特征的匹配

在特征提取的基础上进行特征匹配，基于特征的匹配所处理的图像一般包含的特征有颜色特征、纹理特征、形状特征、空间位置特征等，而这些特征由特征向量表示，那么，图像特征的匹配就可以转化为特征向量间的匹配。包括以下步骤：

特征匹配：通过比较特征向量的相似度获得多幅图像之间对应的匹配点，从而建立各个图像之间特征的匹配对应关系。常用的相似性度量函数有：欧式距离、Mahalanobis 距离、Manhattan 距离、“City Block” 距离、Minkowsky 距离等。

消除错误匹配：不管选用哪种相似性度量函数及特征描述符，都会存在错误的匹配，就需要对基于距离约束对匹配点进行校正。常用的候选消除算法有：RANSAC 随机抽样一致性算法、极限约束关系等。

3) 维数灾问题

在三维重建过程中，维数灾问题是不可忽视的一大难题。维数灾问题是由整个重建系统状态导致的，从单个图像中提取许多特征点，每个位置进行位姿估计要包括 6 个自由度，且实际环境空间中存在的标志性特征成千上万，此时就会引发维数灾的问题。为解决此问题，必须对特征进行筛选来降低维数。

1.4 章节安排

本文针对面向室内环境三维重建及相机位姿估计研究，在阅读大量文献和参考并行跟踪与地图构建（PTAM）算法^{[14][15]}及 vSLAM^[16-26]的基础上，各章节内容安排如下：

第一章：绪论部分主要介绍了本课题的研究背景与意义，基于计算机视觉的三维重建的国内外研究现状，及研究热点与难点，同时介绍了本文的主要研究内容和组织结构。

第二章：针对相机位姿的研究，介绍并行跟踪与地图重建（PTAM）算法原理，以及介绍相关联的知识细节，从理论上解释采用 PTAM 算法可以进行位姿估计并描绘出移动平台的运动轨迹，同时可以进行跟踪与地图重建。

第三章：在第二章理论知识介绍的基础上，依据 PTAM 算法原理，从实际出发，构建一个完整的地面方位与性能评估平台，采集数据，整理数据库，对 PTAM 的精度和鲁棒性进行定量评估。

第四章：在 PTAM 位姿估计的基础上，独具创新与 Kinect-Xbox360 相结合，利用 Kinect 具有易获取深度信息传感器的特点，构建一个完整的地面实验与性能评估平台，采集数据，整理数据库，对室内场景进行三维重建；与此同时，提出三种改进现有算法的方法，运用改进的 RGB-D-SLAM 算法^[27-31]，真正的实现了室内三维场景的重建。

第五章：结束语。对本文的主要工作和研究成果进行总结，分析存在的不足与局限性，对今后的工作以及相关领域的发展前景进行展望。

第二章 基于 PTAM 的相机位姿估计算法

2.1 引言

关于以移动机器人为代表的移动设备平台，自主导航是一大主流难题，要完成许多任务，需要移动平台具备“人的感官”，允许机器人在未知环境中，依靠自身所带的传感器（如相机）递增式地创建环境地图，通过解算相机位姿，给出机器人所在位置，因此，应运而生的 SLAM 算法随即成为自主导航领域的一种关键技术。从最早由 Smith、Self 和 Cheeseman 于 1988 年提出以来，其应用价值被很多学者认为是实现真正全自主移动机器人导航的关键^{[32][33]}。根据机器人要完成实时地图重建与跟踪才能确保其在未知环境中自如运动的问题，越来越多的研究者在 SLAM 的基础上，研究出了 EKF-SLAM^[34]、Mono-SLAM、粒子滤波 SLAM^[35]等。在算法上对时间花销、定位精确程度、获取环境信息等方面做了改进。

近几年出现了一种新的算法——并行跟踪与地图重建（Parallel Tracking and Mapping, PTAM）。通过相机视觉模型，在考虑同等时间代价的情况下，该算法比 MonoSLAM 算法更加精确；与 EKF-SLAM 算法比较而言，比起其采用状态空间描述法的递推最优估计理论，该算法能容纳更多特殊的变化；在关键帧数目的选取上也有一定优势。

本论文将采用此算法，并对其进行了相应研究，在本章中，具体介绍与其相关的理论知识，同时在理论上给出预想结果。

2.2 PTAM 的相关预备知识

在计算机视觉中，通过图像获得外界信息与周围环境，图像是空间物体通过成像系统在像平面上的投影。图像上每一个像素点的灰度反映出空间物体表面某点的反射光的强度，该像素点在图像上的位置则与空间物体表面对应点的几何位置有关。计算机三维重建是指利用摄像机获取到的图像中的而为信息，求解出物体的空间位置、形状等三维信息，从而能够识别和重建物体或真是场景的技术。涉及到摄像机的模型、摄像机的标定、摄像机的校正以及图像的尺度化——图像金字塔等基本知识，在本章节中，我们从它们的内容和原理等方面进行详细介绍，并且在后面章节会有所涉及这些内容要点。

2.2.1 摄像机模型

摄像机模型是摄像机标定的基础与核心，只有确定了成像模型，才能进一步确定出摄像机的内外参数。三维空间中的物体到像平面的投影关系即为成像模型。理想的投影成像模型是光学中心的投影，也称针孔模型或线性摄像机模型。针孔模型假设物体表面的反射光都经过一个小孔从而投影到像平面上，即符合光的直线传播原理。针孔模型主要由光心（投影中心）、成像平面和光轴组成。由于针孔太小，小孔成像的透光量不足，因此需要长时间曝光，并且很难得到清晰的图像。实际摄像系统通常由透镜或者透镜组组成。这两种模型具有相同的成像关系，即像点是物点和光心的连线与图像平面的交点。因此，可以用针孔模型作为摄像机成像的模型。

在此，仅介绍对于一块薄凸透镜的成像系统。

对于仅有一块理想的薄凸透镜的成像系统来说，要想得到一个成缩小的实相，物距 u ，相距 v ，焦距 f 之间就必须满足^[24]：

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f} \quad (2-1)$$

当 u 远大于 f 时，取 v 与 f 近似相等，如果取透镜中心为三维空间坐标系原点，那么，三维物体成像于透镜焦点所在的像平面上，如图 2-1 所示。

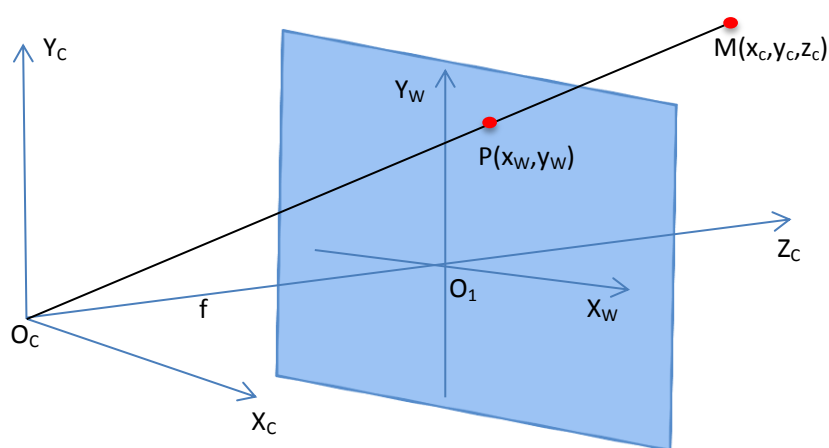


图 2-1 摄像机模型

2.2.2 坐标变换

计算机视觉中一般需要用到三个坐标系，分别是：世界坐标系 $WCS(x_w, y_w, z_w)$ 、摄像机坐标系 $CCS(x_c, y_c, z_c)$ 和图像坐标系 $WCS(r, c)$ （如图 2-2 所示）。由设备采集到的图像数据是直接由图像坐标系表示的，也就是，图像上的像素的行和列的位置 (r, c) 。而被观测物体的客观位置、大小、形态等必须通过世界坐标系 (x_w, y_w, z_w) 才能给人以真实感受。然而这两个坐标系之间的转换又要通过摄像机坐标系来得到。

1) 世界坐标系（World Coordinate System）

在真实环境中，选择一个参考坐标系来描述摄像机和所关注物体的位置该坐标系就是世界坐标系。摄像机坐标系和世界坐标系之间可以用旋转矩阵 R 和平移向量 T 相互转换，由此，空间中任一点 P 在世界坐标系和摄像机坐标系下的齐次坐标分别是 $(x_w, y_w, z_w, 1)^T$ 和 $(x_c, y_c, z_c, 1)^T$ ，且存在如下关系：

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^r & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \quad (2-2)$$

其中， R 是 3×3 的正交矩阵， t 是三维平移向量， $0 = (0, 0, 0)^T$ ， M_1 是两个坐标系之间的关系矩阵。

3) 图像坐标系（Pixel Coordinate System）

摄像机采集的数字图像在计算机内可以以数组形式存在，数组中的每一个元素（称为像素，Pixel）的值即是该图像点的亮度（或称灰度，若为彩色图像，

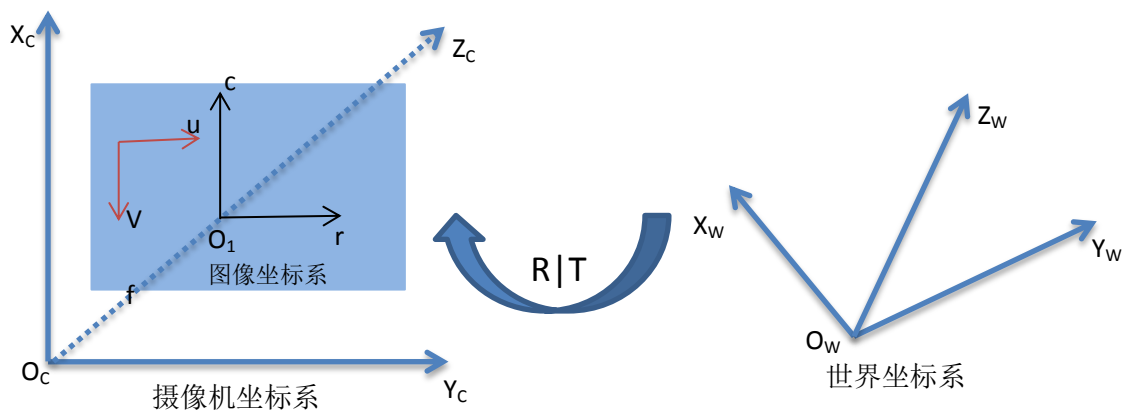


图 2-2 三个坐标系及其间转换

则图像像素的亮度由红、绿、蓝三种颜色的亮度组成)。如图 2-2 所示, 图像坐标系 $r-c$, 每一个像素的坐标可以表示为 (r, c) , r 和 c 分别表示该像素在数组中所在的行数和列数。因此, (r, c) 是以像素为单位的图像坐标系的坐标。

2) 摄像机坐标系(Camera Coordinate System)

摄像机坐标系是和观察者密切相关的坐标系。摄像机坐标系和图像坐标系相似, 差别在于摄像机坐标系处于 3D 空间中而图像坐标系位于 2D 平面里。摄像机坐标系中, 摄像机位于原点, x 轴向右, z 轴向前(朝向屏幕内或摄像机方向), y 轴向上(不是世界的上方而是摄像机本身的上方)。摄像机成像几何关系如图 2-2 所示, 其中 OC 表示摄像机的光心, 轴 XC 和轴 YC 与成像平面坐标系的 u 轴和 v 轴平行, ZC 轴表示摄像机的光轴, 其与图像平面垂直。光轴与图像坐标系平面相交于点 $O1$, 点 OC 、轴 XC 和轴 YC 共同组成的直角坐标系叫做摄像机坐标系。 $OCO1$ 为摄像机的焦距。

2.2.3 摄像机内外参标定

1) 外部参数

摄像机的外部参数用来描述摄像机坐标系与世界坐标系的关系, 表明摄像机位于世界坐标系下的具体位置的姿态。由式 2-2 知, 其可用旋转矩阵 R 和平移向量 T 来表示, R 是单位正交矩阵, 该矩阵满足 6 个正交约束关系, 包含有 9 个元素, 故实际旋转矩阵只有 3 个独立的参数, 再加上平移向量 T 的 3 个参数, 总共有 6 个独立外部参数。

2) 内部参数

内部参数有摄像机的内部结构决定, 而与摄像机所处的外部环境及位置无关, 由摄像机模型可知, 世界坐标系中一点 P 在相机坐标系下的坐标 P_c 和图像中对应点的像素坐标 $(u, v)^T$ 有如下关系:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \begin{pmatrix} x_c/z_c \\ y_c/z_c \\ 1 \end{pmatrix} = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_c/z_c \\ y_c/z_c \\ 1 \end{pmatrix} \quad (2-3)$$

其中, 图像的主点坐标为 (x_0, y_0) , 通常位于图像中心, f_x 、 f_y 是以像素为单位的焦距。可以看出欲确定相机坐标系和图像坐标系的相对关系, 就需要确定主点坐标以及焦距, 共 4 个参数, 称此 4 个参数为相机内参数。

2.2.4 图像金字塔

图像金字塔 (Image Pyramid) 是以多分辨率来解释图像的一种有效但概念简单的结构。图像金字塔最初用于机器视觉和图像压缩, 一幅图像的金字塔是一系列以金字塔形状排列的分辨率逐步降低的图像集合, 金字塔的底部是待处理图像的高分辨率表示, 而顶部是低分辨率的近似。当向金字塔的上层移动时, 尺寸和分辨率就降低。如图 2-3 所示, 是经过卷积迭代等处理的金字塔图像序列。

金字塔是结合降采样操作和平滑操作的一种图像表示方式。它的一个很大的好处是, 自下而上每一层的像素数都不断减少, 这会大大减少计算量; 而缺点是自下而上金字塔的量化变得越来越粗糙, 而且速度很快。

对于一幅尺寸为 $2^J \times 2^J$ 的图像, 则其图像金字塔中间级的尺寸是 $2^j \times 2^j$, 其中 j 的理论取值范围为 $0 \leq j \leq J$ 。但是当 j 很小时, 图像将丢失大部分信息, 所以一般限制只用 $P+1$ 级来减少原来图像近似值的尺寸, 也就是 j 的实际取值是 $(J-P) \leq j \leq J$ 。也就是说通常限制它们只使用 P 级来减少原始图像近似值的尺寸。而一般处理这种分辨率变化, 简单的方法就是隔行取值来生成低分辨率的图像, 复杂的方法就涉及到图像的内插和过滤等步骤, 当然复杂的方法使得生成的低分辨率下的图像更加平滑和真实。对某幅原始图像生成了一系列的低分辨率图像之后, 这些图像和原始图像一起就形成了图像金字塔。以 2.3.2 节跟踪系统涉及到的图像金字塔为例, 可生成类似如下的金字塔, 如图 2-4 所示。

金字塔是结合降采样操作和平滑操作的一种图像表示方式。它的一个很大的好处是, 自下而上每一层的像素数都不断减少, 这会大大减少计算量; 而缺点是自下而上金字塔的量化变得越来越粗糙, 而且速度很快。



图 2-3 图像序列金字塔

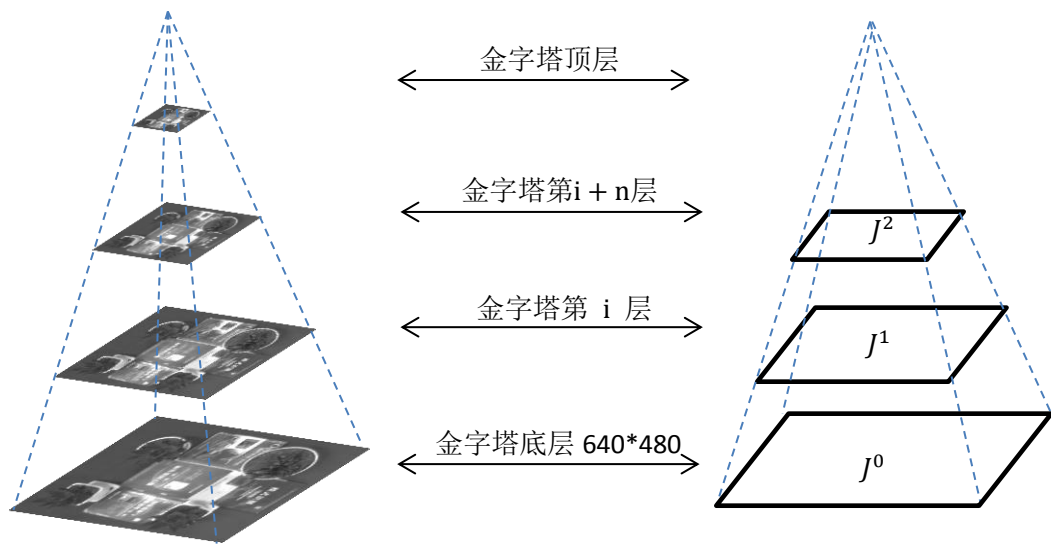


图 2-4 PTAM 跟踪系统图像金字塔

对于一幅尺寸为 $2^J \times 2^J$ 的图像，则其图像金字塔中间级的尺寸是 $2^j \times 2^j$ ，其中 j 的理论取值范围为 $0 \leq j \leq J$ 。但是当 j 很小时，图像将丢失大部分信息，所以一般限制只用 $P+1$ 级来减少原来图像近似值的尺寸，也就是 j 的实际取值是 $(J-P) \leq j \leq J$ 。也就是说通常限制它们只使用 P 级来减少原始图像近似值的尺寸。而一般处理这种分辨率变化，简单的方法就是隔行取值来生成低分辨率的图像，复杂的方法就涉及到图像的内插和过滤等步骤，当然复杂的方法使得生成的低分辨率下的图像更加平滑和真实。对某幅原始图像生成了一系列的低分辨率图像之后，这些图像和原始图像一起就形成了图像金字塔。以 2.3.2 节跟踪系统涉及到的图像金字塔为例，可生成类似如下的金字塔，如图 2-4 所示。

当然，有时候也会对原图像进行扩展，生成分辨率更高的图像，然后把这个图像放到金字塔的更底层。这时候就需要更多处理，使得生成的图像保持真实性，避免色块和噪声等的产生。

2.3 并行跟踪与地图重建算法

PTAM(Parallel Tracking and Mapping, 并行跟踪与地图重建)^[35]是由 Georg Klein 等人于 2007 年提出的增强现实的一种算法。其理论基础在 Georg Klein 等人发表的文章中有陈述。如下图所示，PTAM 的最大特点是将跟踪和地图重建分开来看，将其分别在两个并行的线程中进行，此种想法与市场发展相适应，目

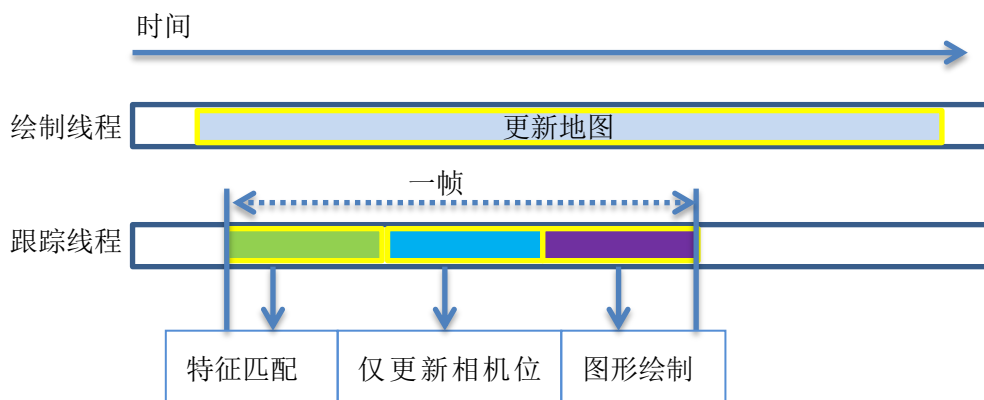


图 2-5 PTAM 系统流程及框架

前来看，越来越多的电脑具有双核或以上多核的配置。同时，PTAM 能自动锁定采集到的图像中的特征点(如场景中物体的边缘和角点)，识别出上千个特征点，并不使用标记，对现有空间进行分析识别。随着摄像头的移动可以动态添加或者减少特征点。它不像 ARToolKit 那样需要有特殊的参照物，因此具有更高的鲁棒性。其定位平面的方法是根据同一个特征点的移动描绘出弧线，再根据所有特征点的弧线确定一个特定的平面。

该算法使用双核 CPU，其中一个核用来处理特征点的侦查和锁定，另外一个核则用来定位出空间中特定平面（如图 2-5 所示）。使用高性能的 CPU，其处理速度能达到采集帧数级。

PTAM 作为新近提出的技术，具有自动识别特征点的功能，不需要特定的标志物，这在大多数情况下确实提供了很大的方便。PTAM 因为没有特定的参照物，所以不能保证两次识别出的平面的坐标系完全相同。ARToolKit 因为使用了特定的标志物，所以能确保识别出平面的坐标系是相同的。

另一方面，PTAM 因为要对大量的特征点进行自动识别和追踪，所以既要求有高分辨率的摄像头来采集清晰的特征点，又要求有高性能的 CPU 来进行处理。而 ARToolKit 对机能要求相对较低，甚至能在手机、PDA 等设备上运行，所以有更好的适应性。

2.3.1 地图的表示

地图由一系列位于世界坐标系下的 M 点特征组成，这些点均。每个点代表了世界坐标系下的局部平面的纹理补丁。地图中的第 j 个点的世界坐标可表示为

$p_{jw} = (x_{jw} \ y_{jw} \ z_{jw} \ 1)^T$ ，同时，每个点含有一个单元补丁常态 n_j 和一个源像素作为参考。

该地图还包含 N 个关键帧，每个关键帧均位于以该视角下摄像头为坐标原点的坐标系，用 k_i 表示第 i 个关键帧。那么，此坐标系与世界坐标系之间的转换用 E_{k_iw} 来表示。与此同时，每个关键帧也存储了一个四层灰度图像金字塔，该图像的每个像素用 8 位表示；零级存储 640×480 像素的摄像头快照，并像这样向下采样，第三级为 80×60 像素。

组成每个补丁特征的像素不单独保存，而每个点特征含有一个源关键帧，通常，以第一个关键帧为观察视点。因此，每个地图点对每个单一源关键帧存储一个参考帧，在此关键帧的单一源金字塔等级，像素也位于此金字塔等级。在源金字塔等级，补丁对应于 8×8 像素正方形，在世界坐标系中，一个补丁的大小和形状取决于金字塔等级，从源关键帧相机中心的距离以及补丁常态的方向。

例子表明之后的地图可能包含 M 为 2000 到 6000 个点和 N 为 40 到 120 个关键帧。

2.3.2 跟踪系统

PTAM 将跟踪与地图构建分开来看，既然地图构建不依赖于跟踪，那么，就没有必要使用映射的每一帧图像。在跟踪过程中，视频流中的许多帧都包含了冗余信息，特别是当摄像机静止时，由此可想，我们可以把精力主要集中在处理一些较小且更有用的关键帧上。

PTAM 跟踪系统采用 Unibrain Fire-I 视频摄像头来采集数据。该款摄像头配备了 2.1mm 的广角镜头，获得图像的像素为 640×480 ，频率为 30Hz，采用 YUV 色彩模型，最终用于跟踪的为 8bpp 的图像。

跟踪过程主要分为以下几个阶段：

1) 校准摄像机模型，采用针孔摄像机投影模型，根据 2.2.1 和 2.2.2 小节知识校准摄像机，主要用到 `CamProj()`函数；对摄像头进行位姿估计，并利用坐标系之间的转化关系将世界坐标系下的点转化到相机坐标系下，形成地图点 P ，完成投影工作；

2) 从摄像机采集一帧图像，由先前的位姿估计产生一个运动估计模型，根据此估计模型，将点投影到图像上；在图像中，进行大尺度的特征搜索，大概 50 个左右，即补丁搜索。补丁搜索的过程如下：在当前帧中，为了找到一个单一的地图点 P ，我们需在该点的预测图像位置的周围执行固定范围的图像搜索。要

执行此搜索，必须先考虑第一次观察到的和当前摄像机位置之间视点的改变，即补丁的扭曲变化。我们用一个变形矩阵 A 来执行仿射扭曲特征化，即

$$A = \begin{bmatrix} \frac{\partial u_C}{\partial u_S} & \frac{\partial u_C}{\partial v_S} \\ \frac{\partial v_C}{\partial u_S} & \frac{\partial v_C}{\partial v_S} \end{bmatrix} \quad (2-4)$$

其中， $\{u_S, v_S\}$ 表示对应于补丁源金字塔等级和垂直像素的位移，和 $\{u_C, v_C\}$ 表示对应于当前相机帧的零级（最大像素位移）金字塔的像素位移，即位于金字塔底层。这个矩阵的反投影单元像素位移位于补丁平面上的源关键帧的金字塔级，然后投影到当前（目标）帧。执行这些投影确保变形矩阵的补偿（第一顺序）不仅改变了角度和尺度，而且改变了关于图像的镜头畸变大小。

上式矩阵 A 的行列式是用来决定应该搜索位于金字塔哪一层级的当前帧的补丁。行列式 A 对应的区域，方像素以及单一源像素将占据整个分辨率图像； $\det(A)/4$ 表示金字塔第一级的相应区域，依此类推。选择目标金字塔层级为 l 层，这样 $\det(A)/4^l$ 就最接近单元目标顶层，也就是说，我们试图找到在金字塔等级中与其尺度最相似的匹配的补丁。例如，一个 8×8 像素补丁搜索模板是从金字塔底层利用 $A/2^l$ 和双线性插值共同生成的。从各个像素值中减去平均像素强度，以提高一些光线变化的适应能力。接下来，在一个以预测位置为圆心，以固定值为半径的圆的围绕范围内，最终发现最佳匹配模板位于目标金字塔等级。这是通过对零均值 SSD 的得分评估得到的，这些得分来自圆搜索区域内的所有 FAST 角点的位置以及选择最小差值的位置。如果这此预设的阈值之下，那么该补丁被认为已找到。

在某些情况下，特别是处于金字塔的高等级，整数像素位置就不能够准确地产生平滑的跟踪结果。补丁位置可以通过执行一个迭代的最小误差来进行细化。我们用反向合成方法^[1]，最大限度地减少平移和平均补丁强度差异。但是，执行跟踪每个补丁的计算量是昂贵的。

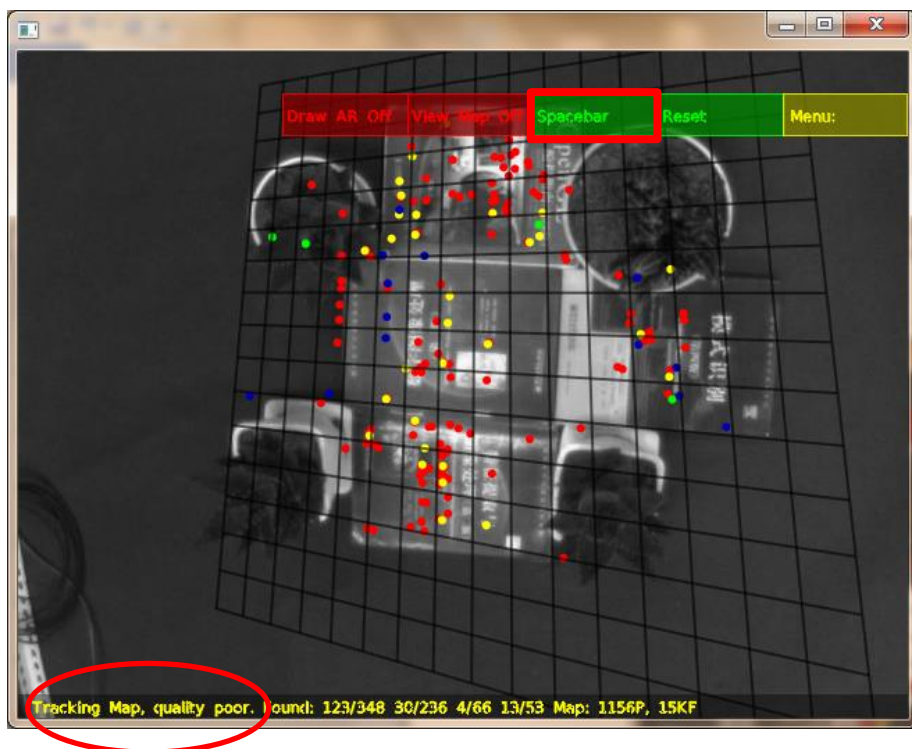
3) 依据上一步中的补丁搜索以及粗匹配来更新相机位姿。由一个给定的集合 S 表示已成功观测到的补丁，进而一个摄像头的姿态更新就可以计算出来。每个观测产生一个已被找到的补丁位置 $(\hat{u} \hat{v})^T$ （称为零级像素单元），并假定为有 $\sigma^2 = 2^{2l}$ 乘 2×2 的单位矩阵（还是在零级像素单元）的测量噪声。姿态更新的迭代计算是对误差的目标函数进行最小化鲁棒重投影得到的。

4) 为了提高跟踪系统摄像机快速移动时加速度的恢复能力, 补丁搜索和姿态更新要进行两次。(2) 中提到最初的粗略搜索仅搜索 50 个构图点, 它们出现在当前帧图像金字塔的最高层级, 该搜索(亚像素细化)在一个大的搜索径向范围内执行。从这些测量值, 计算出一个新的姿势。继而, 在图像中, 当剩余潜在的可见图像补丁增至 1000 时, 这些补丁会被重投影回图像, 而且, 补丁搜索会在这个更严格的搜索区域内再次进行。而亚像素细化仅执行在一个高层级补丁的子集中。最后一帧的位姿由图像测量的粗测和细化的计算结果共同决定。

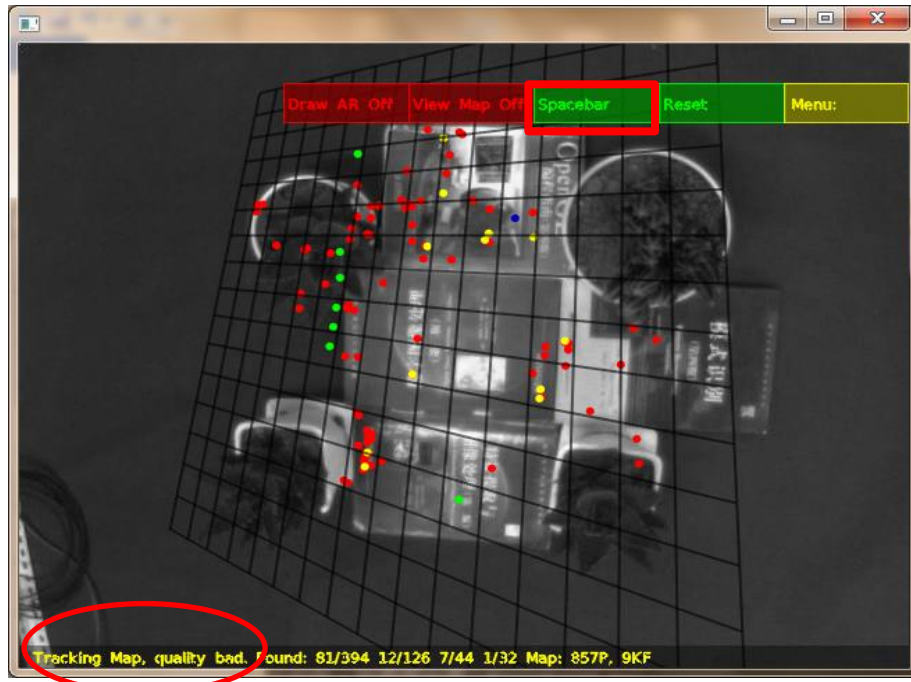
5) 跟踪系统尽可能的要实现鲁棒性, 但是实际中不可避免地会出现跟踪失败的现象。处于这方面的考虑, 跟踪系统对捕获的每一帧都进行了跟踪质量的估计。以特征观测的成功比作为衡量标准。如果此比例低于一定的阈值, 则被认为跟踪质量差, 忽略此帧不用于地图构建, 继续执行跟踪过程; 如果此比例低于一个更低的阈值并超过几帧, 则认为跟踪丢失, 即跟踪失败。此时, 可以按下 reset 键, 重新进行跟踪。如图 2-6 所示。出现跟踪失败或者跟踪质量差的情况, 可能是由于遮挡, 模糊等情况的出现。

2.3.3 地图的构建系统

地图构建系统主要是建立 3D 点云图的过程。地图的构建发生在两个不同的阶段: 第一阶段, 采用立体技术[]构建初始地图; 第二阶段, 通过映射和匹配过程, 不断的对初始地图进行扩增和更新。其中, 关键帧是由跟踪系统来增加的。



(a) 跟踪质量差不用于地图构建



(b) 跟踪质量极差跟踪失败

图 2-6 跟踪质量与绘制关系

地图构建的流程图如图 2-7 所示，详细的描述了整个地图构建的过程。

1) 地图初始化

当系统首次启动时，我们采用五点立体算法来进行地图初始化，其方式与 [10,18,20] 相似。首先用户将相机置于要进行跟踪的工作区，并按下 Spacebar 键。同时，系统的第一个关键帧将被存储到地图上。接着，1000 个二维补丁跟踪初始在金字塔的最低水平级，在显著的图像位置（最大 FAST 角）。然后，用户可以顺利地平移（并可能旋转）相机使其略有偏移位置并再次按下此键。通过平滑的运动跟踪二维补丁，从而，第二个关键按钮提供了第二个关键帧和特征对应，其来自从五点算法和 RANSAC 估计基本矩阵和三角化基本地图。通过捆绑调整细化而成最终的地图。

初始地图可以任意缩放，并起初用一台摄像机。为了在有意义的地方和规模得以增强，地图首先要归一化单位。这就要假设相机在立体点对间平移 10 厘米左右。接着，旋转和平移地图，主平面位于 $z = 0$ 的世界坐标系下。这一步采用基于仿射不变量的 RANSAC 算法：许多点集合是随机被选来假设一个平面，而剩余的点则用于一致性测试。获得的假设是细化而成的，其需要通过评估一致性集合的空间均值和方差，以及形成平面法向量的协方差矩阵的最小特征向量。

包括用户交互，地图的初始化大约耗时 3 秒钟。

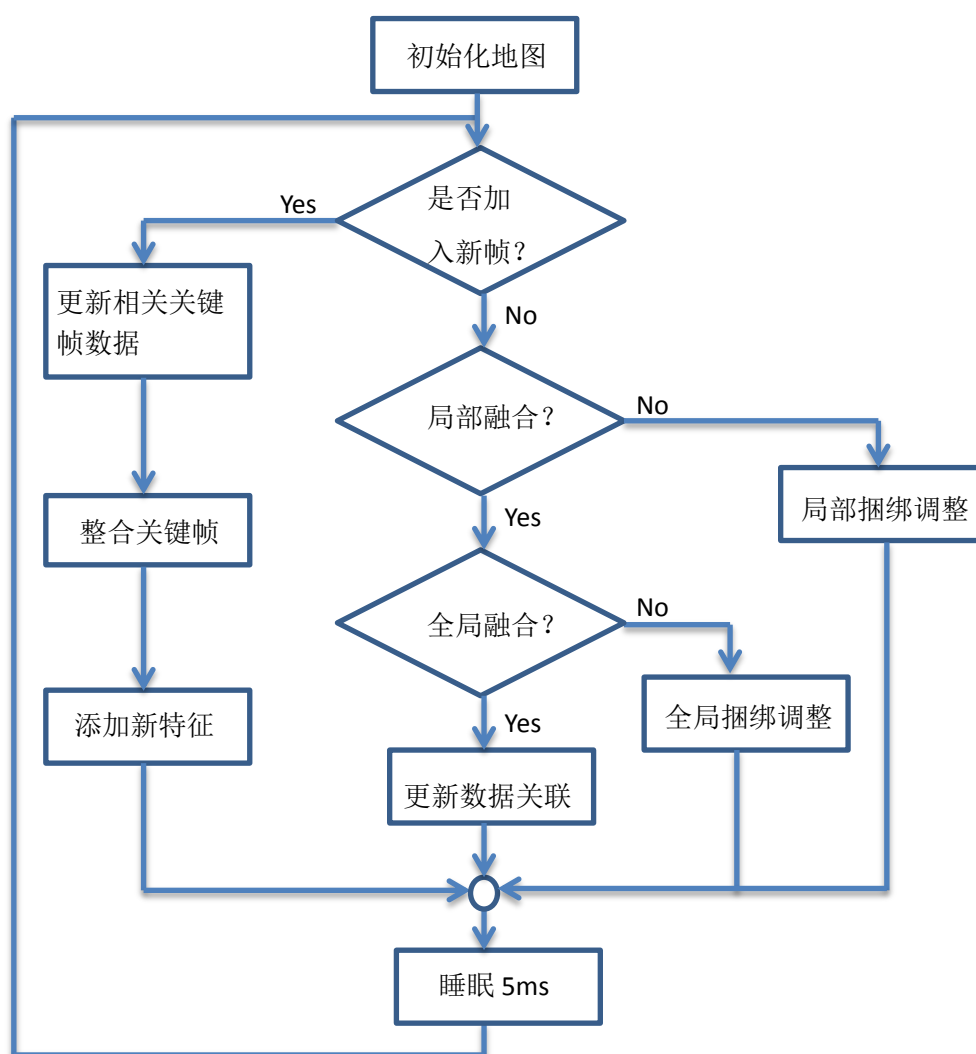


图 2-7 异步地图重建过程流程图

2) 添加关键帧

最初的地图仅包含两个关键帧，并且描述的是一个较小空间。由于摄像机的移动而远离其初始位置，新的关键帧和地图特征被添加到系统中，进而扩增地图。可以被添加的关键帧需要满足以下条件：跟踪质量必须好，与被添加的上一个关键帧时间间隔必须超过 20 帧；相机必须以最小距离远离已加入地图的最近关键点。最小距离要求避免常见单目 SLAM 问题，并要保证立体基线的新特征三角化。采用的最小距离取决于特征的平均深度，以致当相机接近一个表面时，关键帧间隔很近，当观测到远处的墙壁时，需进一步划分。

对于关键帧的每一级金字塔，跟踪系统已经计算出了一个 FAST 交点的集合。在 Shi-Tomasi^[36]的基础上，无极大抑制和阈值被用来缩小这个集合并为每层金字

塔找出最突出点。其次，离可成功观测的已有特征近的突出点将被丢弃。那么，剩下的突出显著点就成为构建新地图的候选点。

新地图的点需要深度信息。这对单个关键帧来说是不可实现的，而要三角化其他视点。就摄像头位置而言，我们选择地图上已经存在的最近关键帧作为第二视图。这两种视图之间使用极线搜索来建立的对应关系：在第二视图中，采用零均值 SSD 法，对极线旁角点周围的像素补丁和候选地图点做比较，并仅在相等金字塔层级进行搜索。此外，我们不搜索无限极线，但对新的候选点（取决于在新关键帧中现有点的深度贡献）的可能深度做以先前假设。如果已经发现一个匹配，那么三角化新地图点，并插入到地图中。

3) 捆绑调整和数据关联细化优化地图

通过捆绑调整迭代来降低误差，以提高鲁棒性。这种方法不仅可以对相机位姿进行优化同时可以对场景中的特征点进行优化全局捆绑调整是对所有的关键帧进行的，而局部调整仅是对于某个新关键帧，其周围相邻的四个关键帧用以优化调整。如果关键帧被添加到地图时，捆绑调整正在进行，那么，立即中断捆绑调整，并将新关键帧添加到底图中之后继续进行捆绑调整操作。

当捆绑调整已经收敛，并且没有新的关键帧再被添加时，线程有空闲时间，此时可以用于地图的优化。像捆绑调整一样，只要一个新的关键帧到达，数据关联就要被中断。这主要是通过新的测量在老的关键帧上执行；在老的关键帧中测量新创建的地图特征，或重新测量异常值。由跟踪系统做出的测量可能是不正确的。这经常发生在世界包含的区域重复模式。这种测量给予低的权重捆绑调整中使用的 M-估计。如果他们位于杜克估计的 0-权重区域，他们就会被标记为离群值。每个离群测量值在删除前都会有“第二次机会”：在关键帧中重新测量它，使用该特征的预测位置并在比之前更严格的搜索区域内进行跟踪。如果发现新的测量，重新插入到地图中。经过这样的测量后，仍然被认为是一个离群值，则将永久从地图中删除。

2.3.4 理论结果

1) 耗时方面

对于大多数的图像序列，尽管地图的尺寸一直在增加，跟踪过程执行大约都为 20ms。其中，有两种例外情况：若图像序列较多，1) 跟踪会约在 1320 帧时跟丢，并且系统将尝试恢复，每帧需花费 90ms。2) 若正常操作情况下，约在 1530 帧时，跟踪时间每帧需花费 30ms。以上是当摄像机移动远离办公桌时。在这里，所花费的时间决定哪个特征被测量并变得显著。如表 2-1 所示，对一个典

型帧跟踪所要花费的时间。其中，关键帧的准备包括帧的捕获，YUV 转换为灰度图像金字塔以及 FAST 角点的检测。特征投影主要是决定哪些特征是可见的，哪些是待测量的。补丁搜索和姿态更新的耗时主要取决于计算扭曲矩阵 A、搜索各幅图对应补丁以及每级金字塔角点密度。

表 2-1 M=4000 时的跟踪过程耗时

各个阶段	准备关键帧	特征投影	补丁搜索	姿态更新	总计
耗时 (ms)	2.2	3.5	9.8	3.7	19.2

该算法所适用的场景范围只是针对“小工作空间”。但是，跟踪系统具有很好的地图扩展性，通常系统一般保持大约 6000 个点和 150 个关键帧可用。对于捆绑调整，单独映射的耗时是很难获取的，他们之间存在着很大差异。而我们采用异步的方式，可以获得不同大小地图完成捆绑调整时的时间。如表 2-2 所示。

表 2-2 不同大小地图的捆绑调整耗时

关键帧数	2-49	50-99	100-149
本地捆绑调整 (ms)	170	270	440
全局捆绑调整 (ms)	380	1700	6900

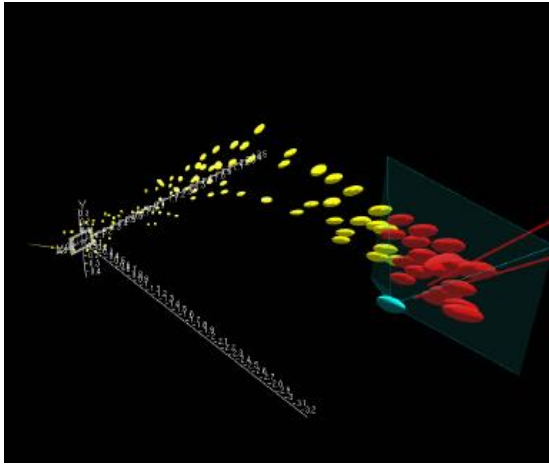
上述均为耗时平均量。当地图增长超过 100 个关键帧时，全局捆绑调整跟不上，总是几乎被迫中止，对大于 150 个关键帧的全局地图收敛可能需要耗时几十秒。

2) 与 EKF-SLAM^[36] 的比较

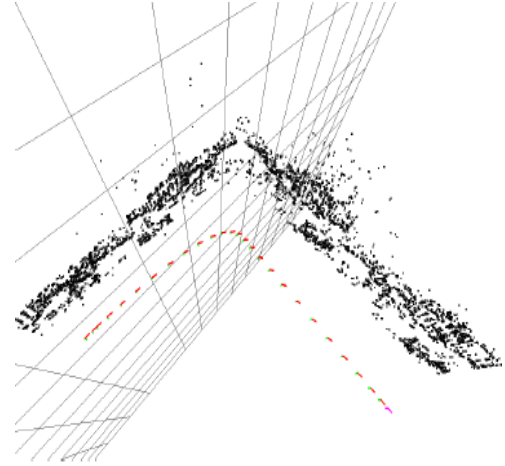
以一个墙角为例，当摄像机沿直角墙运动时，从一堵墙运动到另一堵墙时，该过程中 EKF-SLAM 会明显缩小墙角的角度，直角轨迹变成了近似曲线轨迹，并生成稀疏的特征地图，而 PTAM 算法会生成一个相对稠密的特征地图，轨迹近似规则直角。比较两者的轨迹，进行分析可以看出，误差主要存在于 z 轴方向上。如图 2-8 (a) 所示，黄色轨迹表示 EKF-SLAM 算法的轨迹，如图 2-8 (b) 所示，红色为 PTAM 算法的轨迹。

针对“SLAM 摆动”现象，EKF-SLAM 系统通常要保证持续平滑的摄像机运动，在正确的深度下，有效的初始化新特征。若摄像机固定时，跟踪抖动就会

导致深度初始化的错误。相比之下，PTAM 就更具有鲁棒性，当摄像机聚焦于兴趣点，用户可以暂停一下，然后到下一个兴趣点，或者是相同点的不同视角。此



(a) EKF-SLAM 轨迹



(b) PTAM 轨迹

图 2-8 EKF-SLAM 轨迹与 PTAM 轨迹

外，PTAM 采用图像金字塔，极大地提高了系统快速运动和与其关联的运动模糊的耐受性。同时，当一些特征被遮挡或损坏时，大量特征的使用会减少可见跟踪抖动，进而提高整个系统的鲁棒性。

具体的系统评估情况将在下一章节具体介绍。

第三章 基于 PTAM 的实时移动平台的位置姿态估计

3.1 引言

PTAM 将跟踪和地图重建看作两个独立的任务，在双核计算机中利用并行线程处理：一个线程处理鲁棒跟踪不稳定手持式运动的任务，而另一个，处理从先前观察到的视频帧中提取点特征并产生 3D 地图的任务。这允许使用计算量大的批量优化技术，通常不与实时操作有关：其结果是一个能由成千上万的地标生成详细地图的系统，并以帧速率进行跟踪，其精确度和鲁棒性可以与当前基于模型系统的最高技术水平相媲美。

3.2 移动平台的搭建

为了减少移动摄像机大的晃动，便于之后姿态估计，在此简化 PTAM 的实验平台，如图 3-1 所示，将一个摄像头固定于可水平运动以及垂直运动的自制组建滑杆，以确保摄像头可以按照预想的轨迹（直线）运动。根据第二章中 PTAM 运行的要求，采集图像的摄像头频率需为 30Hz，获得的图像像素需为 640x480，鉴于此，实验中选用 1394 摄像头，分辨率为 1280x1024，帧率为 45fps。



图 3-1 评估数据采集平台

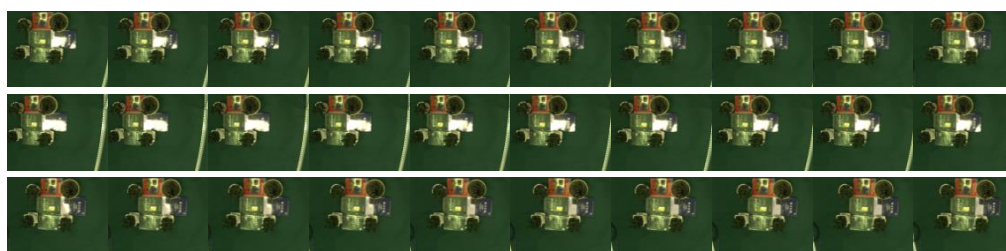
3.3 数据采集与预处理

数据采集阶段，我们最后使用的数据一共有两组：第一组，我们以视频作为输入，手持摄像头随意移动所得到的数据，共 67 帧。为了说明，在图 3-2 (a) 中，我们将视频转化为图像序列做以显示。第二组，我们直接以图像序列作为输入。为了后期评估时利于实验结果的对照，此组数据利用已搭建好的数据采集平台采集到 30×30 的图像序列“矩阵”，共 900 帧，如图 3-2 (b) 所示，只列出了前 30 帧（第一行）的图像序列。第三组，将 1394 摄像头固定于三脚架顶端作为采集设备获取数据，运动大致是以所拍物体为中心，以 0.8m 为半径的圆，共 189 帧，如图 3-2 (c) 所示。

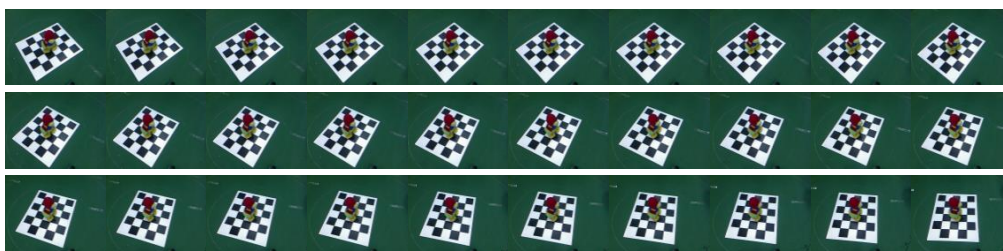
采集到的数据在使用之前需要进行预处理，首先应该考虑透镜畸变带来的影响。畸变的原因主要是摄像头中透镜的引入，理论上是有可能定义一种透镜而不引入为畸变的，但现实世界中没有完美的透镜，只要存在制造加工，就会出现或多或少的畸变。因为制作一个“球形”透镜比制作一个理论上理想的透镜容易得多，而且从机械加工方面考虑，透镜和成像仪要始终保持平行是很难做到的。透镜畸变主要有两种，一种是径向畸变，来自于透镜形状；另一种是切向畸变，来自整个摄像机的组装过程。此处主要以径向畸变为例做以说明。



(a) 第一组采集评估数据



(b) 第二组采集评估数据



(c) 第三组采集评估数据

图 3-2 采集到的数据

实际摄像机的镜头总是在成像仪的边缘产生显著的畸变。这个现象来源于“筒形”或“鱼眼”影响。对于某些透镜，光线受折射与反射的影响，在远离透镜中心的地方比靠近中心的地方显得更加弯曲。对常用的普通透镜来说，这种现象更加严重。筒形畸变在便宜的网络摄像机中非常厉害，但在高端摄像机中不明显，因为这些透镜系统做了很多消除径向畸变的工作。实验中采用的摄像头所产生的畸变不明显，因此，数据处理阶段省去了这一步骤。

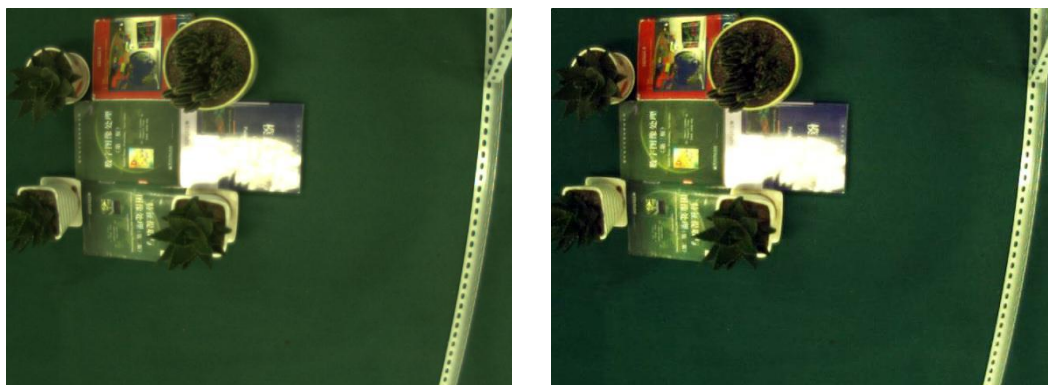
再次，由于光照等因素的影响，所采集到的图像除了发生畸变，还需要进行色亮度的矫正。色亮度矫正是利用一块标准的色彩卡(Color Checker)来进行的。主要原理是，色彩卡上有 24 个不同颜色的方格，每个方格颜色都有标准的 RGB 值，相机拍摄到的色彩卡的每个格子又会有各自的 RGB 值，那么软件就可以根据这 24 组结果分别对 RGB 三条通道拟合出三条对应的曲线，这三条曲线分别表示了 RGB 三通道拍摄值和标准值的对应关系，那么，对摄像机拍摄的图像都可以根据这三个对应关系做相应的变换来达到色亮度矫正的目的。如图 3-3 所示，显示了色亮度矫正前后的图像。



(a1) 原始拍摄图像

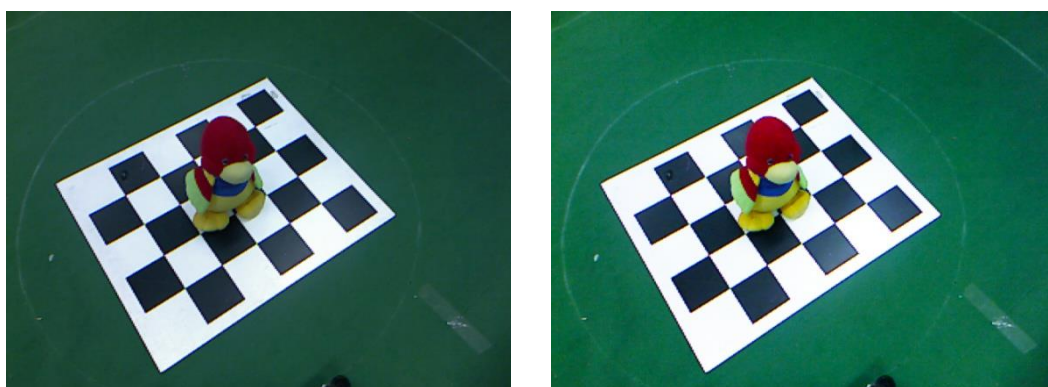


(a2) 色亮度矫正后图像



(b1) 原始拍摄图像

(b2) 色亮度矫正后图像



(c1) 原始拍摄图像

(c2) 色亮度矫正后图像

图 3-3 原始图像与矫正后图像

3.4 基于移动平台的位姿估计及精度评估

由于设备限制，该阶段我们创建了这样的实验环境：采用 AMD APU A6-3400M 四核处理器，主频为 1.40GHz，内存为 4GB，分辨率为 768，显卡芯片为 AMD Radeon HD 6520G 核心显卡+AMD Radeon HD 6650M 独立显卡的笔记本电脑。为了更加准确的进行评估 PTAM 精度同时保持系统的稳定性，我们将原有的摄像头采集数据直接作为输入，改为了将其转化为图片序列，通过读取图片序列来进行跟踪。

对于位姿估计，我们使用 1394 摄像头作为采集设备采集并做处理，将处理后的图像数据作为 PTAM 算法的输入，经计算得到 R 和 T 的参数，再以此作为输入，并用 Matlab 画出移动平台的运动轨迹。

为了更直观的进行精度评估，我们主要采用以下两个方面：一方面，与上一章理论介绍中提到的按照墙角运动，将 PTAM 的输出轨迹与 EKF-SLAM 的轨迹作以比较，如图 3-4 所示。可以看出，同一坐标系下，PTAM 只是在实际轨迹的

附近有小幅度的偏移，而 EKF-SLAM 的轨迹则有较大的偏移，因此，说明 PTAM 的精度比 EKF-SLAM 高。

另一方面，从前期为验证 PTAM 估计精确度建立的数据库中提取三组数据作为 PTAM 的输入，经计算画出轨迹，并将画出的轨迹与 Acts 跟踪软件画出的轨迹或理论轨迹作以比较，依次如图 3-5、3-6、3-7 所示，可以主观的看出，PTAM 的精度比 Acts 跟踪软件的精度高。在确定了 PTAM 精度的基础上，我们将其作为三维重建中的一部分，在下一章做以详述。

这里需要说明的是，该精度验证部分均是事先人为设定的轨迹，根据预想轨迹进行摄像头运动的。这样做有利于进行对比，观察是否与预期估计相符，进而得出有效结论。

下面开始进行各组数据的验证：

- 1) 根据上一章末尾的叙述，摄像机沿 90 度墙角进行折线运动，实际轨迹理论上应该是 90 度的直角，如下图，用黑色细线表示，同一坐标系下的 PTAM 算法和 EKF-SLAM 算法的轨迹对比如下图。红色轨迹表示 PTAM 的轨迹，蓝色轨迹表示 EKF-SLAM 的轨迹。由下图在理论轨迹振荡的幅度可以看出，PTAM 的精度明显高于 EKF-SLAM 的精度。

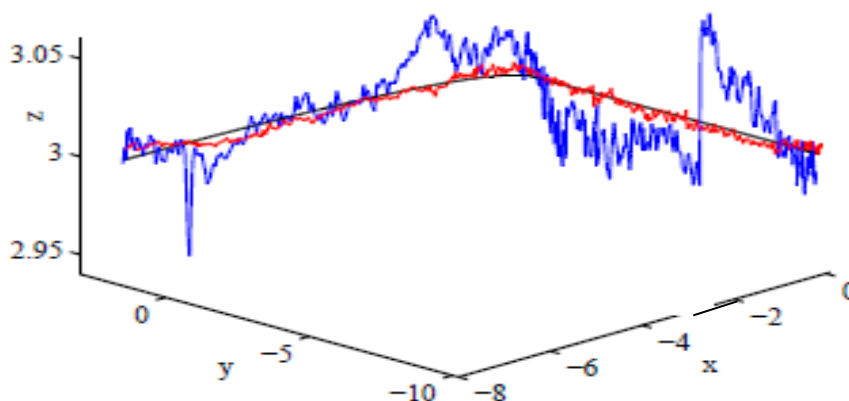
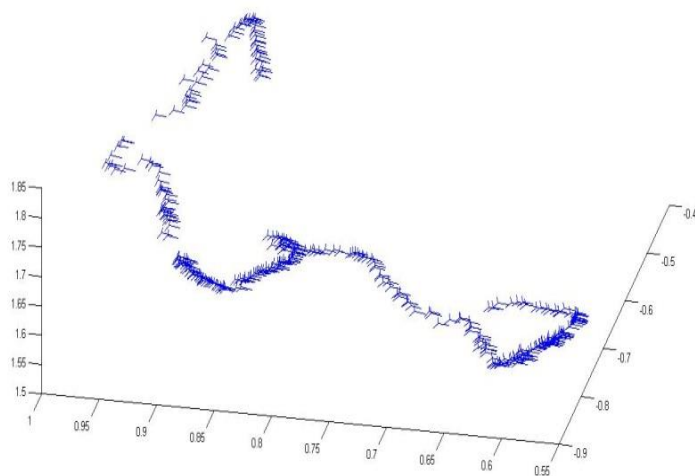
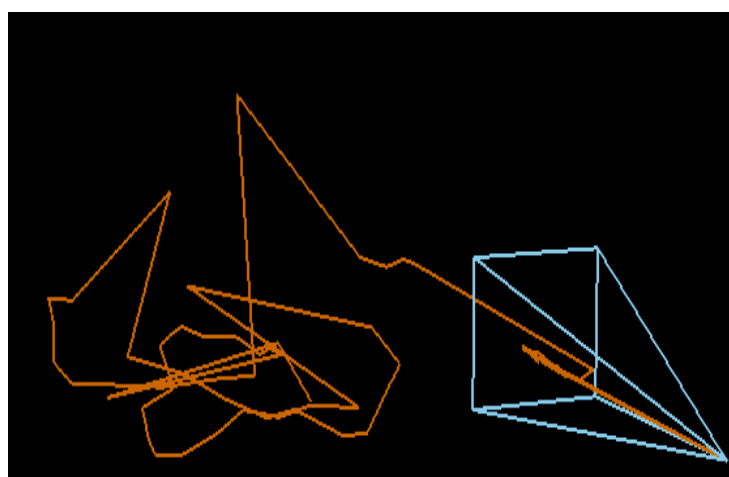


图 3-4 PTAM 与 EKF-SLAM 轨迹的对比

- 2) 使用第一组数据作为 PTAM 的输入并画出其运动轨迹，该组数据采集时为手持 1394 摄像头进行随意运动，但为了在小范围内运动较长时间，故刻意运动略微如“蛇形”，如下图所示。再用 Acts 跟踪软件输出的轨迹，相比之下，Acts 的输出与实际刻意的大致轨迹相差甚远，而 PTAM 的则较为相近。



(a) PTAM---第一组数据运动轨迹

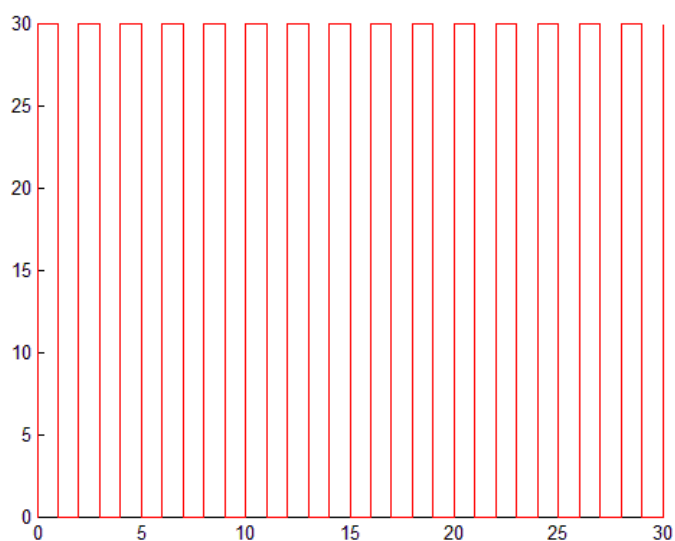


(b) Acts 跟踪软件---第一组数据运动轨迹

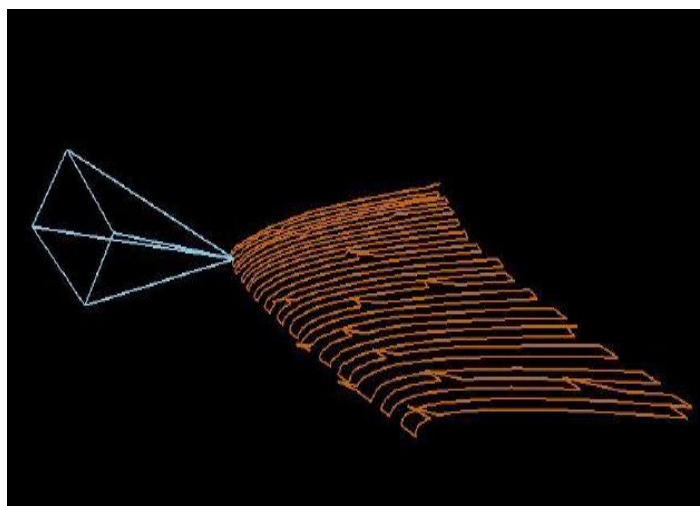
图 3-5 第一组数据结果的对比

3) 使用第二组 900 张数据作为 PTAM 的输入并画出其运动轨迹，第二组数据的拍摄方式为定点拍摄，采集平台如图 3-1 所示，摄像机的镜头垂直于地面向下，按照水平和竖直移动每 1cm 来捕获一帧图像的方法采集数据，若把每一帧图像看作一个点，则此组数据可以形成一个 30×30 的图像矩阵，将所有帧图像进行编号，方便之后的选取。

首先，利用 Matlab 画出理论运动轨迹，如图 3-6 (a) 所示，然后利用 Acts 跟踪软件绘制出全部数据库的数据。可以看出，并非每一条直线都是端直的，有 7 个地方存在误差，与前一列图像有点的重合，除此之外，其他地方绘制较好。

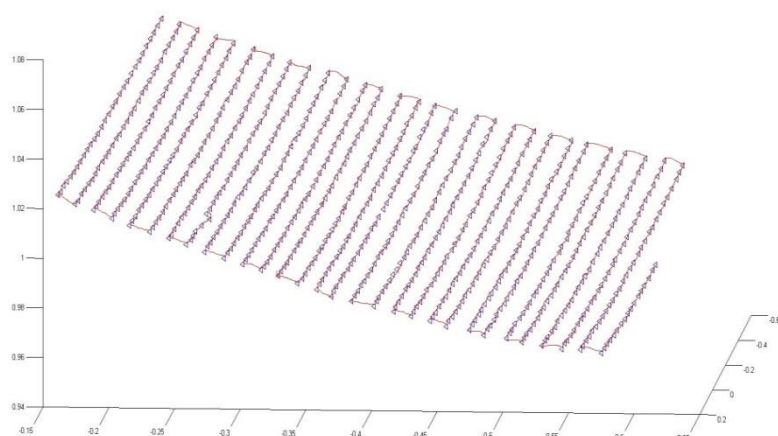


(a) 理论---第二组数据运动轨迹



(b) Acts 跟踪软件---第二组数据运动轨迹

接下来，是用 PTAM 运行的结果，在此需要说明的是，该组数据总共为 900 帧图像，下图中可以看出初始的轨迹没有被绘制出，这是因为 PTAM 系统需要手动建立初始地图，在改写时并没有将其加入绘制中，因此，前 21 帧被用于地图的初始化，经计算得到的 R、T 中没有将初始地图的数据包括，仅包括了后面初始化后用于地图绘制的 879 帧的图像数据信息，后面不再赘述。观察下图，可以看出，除了有一个地方存在明显的偏差，其他绘制与理论绘制结果很为相近，相比之下，PTAM 的轨迹和理论轨迹相符。



(c) PTAM---第二组数据运动轨迹

图 3-6 第二组数据结果的对比

- 4) 使用第三组数据作为 PTAM 的输入并画出其运动轨迹，第三组数据的采集，是将 1394 摄像头固定于三脚架顶端，并按照事先画出的圆形轨迹移动，因此理论上的轨迹应该是一个圆形，但关于设备及场地的局限和手动的晃动的影响是不可避免的，如下图所示。

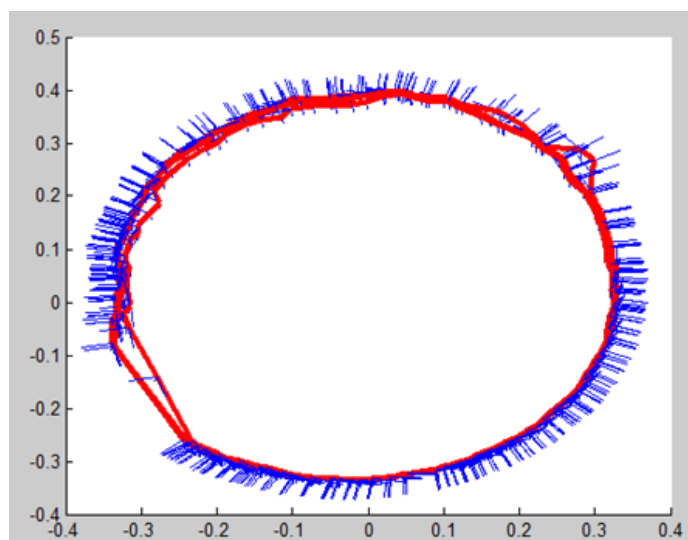


图 3-7 第三组数据的部分数据的运动轨迹

经过 S 矩阵型(包括直线、直角)、圆形以及随意移动采集设备画出的轨迹，综合可以看出，PTAM 算法的精度相当高，可以用于实际位姿的估计，并能取得可观的效果。

3.5 本章小结

本章在上一章介绍的理论的基础上，对 PTAM 的性能做了定性的评估。以对照的方式给出了 PTAM 的跟踪位姿的相对准确度。结果表明，PTAM 的性能较好，其跟踪的鲁棒性和准确性已经达到了相当高的水平。

但是还存在一些局限性：由于该系统规定地图初始化需要手动操作，以避免捕获到质量差的不可用帧，此过程需要花费 3 秒。并规定该系统仅用于小型工作空间，当前得到验证的仅是工作室内的某个感兴趣区域。此外，PTAM 系统还不能达到，在任何未知场景中，任何未经训练的用户，可以运用自如的地步。

同时在验证数据库的建立中，还是会受客观存在设备的局限、非机械严格等间隔移动以及无意识可能产生的晃动等一些影响，这些都是不可避免的需要被考虑进去。

第四章 基于 RGB-D 传感器融合的移动平台室内三维重建

4.1 引言

Kinect^[36]是美国微软公司于2010年11月4日推出的Xbox360游戏机体感周边外设的正式名称，起初名为Natal，意味初生。它实际上是一种3D体感摄影机，利用即时动态捕捉、影像辨识、麦克风输入、语音辨识、社群互动等功能让玩家摆脱传统游戏手柄的束缚，通过自己的肢体控制游戏，并且实现与互联网玩家互动，分享图片、影音信息。Kinect彻底颠覆了原始游戏单一的手动操作，是人际互动的更加完全的展现出来。一经上市，有关其的应用开发日新月异，稍有怠慢就会永远被甩在身后。不过，Kinect目前带给我们的仍只是一个充满无限可能的远景以及体感游戏娱乐之余的科研悸动，由于一些基于Kinect的应用研究仍为复杂。

2011年的SIGGRAPH(Special Interest Group for Computer GRAPHICS, 计算机图形图像特别兴趣小组)会议上，KinectFusion首次展示了实时、廉价、轻便的室内场景的三维重建，使我们向“无处不在的数字化”迈进了一大步。

从某方面来看，Kinect也是性能强大、价格低廉的3D摄像机。Kinect不但可以玩游戏，它还成为机器人的眼睛，帮助机器人认路，源于它集成了多个传感器，可同时获得彩色图像、3D深度图像以及声音信号，结合PrimeSense所开发的Light Coding技术，真正拍摄出3D影像。这是它能够帮助机器人认路的关键。

本章利用Kinect设备可以获得深度信息等的一些优点，将其与PTAM相结合，共同协作完成对室内场景的三维重建过程。由于Kinect设备的限制，本文不能将达到高精度的三维重建，但是可以用于娱乐、展示等目的，并将大众轻松的与虚拟世界以及显示世界相连接，得到真实感受。

4.2 Kinect 基本工作原理

为了提高Xbox游戏的可玩性，让用户实时全身参与其中，实现自然的用户交互，微软推出了一款名为Kinect的周边设备。此款设备刚一问世就凭借平均每天133,333台的销量打破了吉尼斯世界纪录，被评为有史以来销售最快的电子消费产品，这个数字超过了苹果公司iPhone和iPad的同期销售量。在娱乐方面，

Kinect 对游戏中的人物进行操纵，提高了游戏的参与性与趣味性，可以说它让体感操控表现的淋漓尽致。在研究开发方面，世界各地的研究以及开发人员开始并利用这个廉价的深度摄像头，大胆开发出富有创意的应用。

Kinect 的外观（如图 4-1 所示），由图可看到它一共有两个摄像头和一个基于激光红外发射器（从左往右，分别是激光红外发射器 I，RGB 摄像头 II，IR 摄像头 III），中间一个用于获取彩色图像，两边的用于获取深度图像。具体来说，为了使 Kinect 在可见光的环境下工作，则选用红外发射器；在红外发射器的旁边有一个红外 CMOS 传感器，用于感知红外线，通过此传感器 Kinect 把接收到的数据通过内部硬件系统分析得到深度信息，从而构成 3D 深度感应器。

由微软的官方说明中得知，Kinect 的可用距离为 1.2 到 3.5m，然而通过实验证明，Kinect 的实际可用距离为 0.4 到 7.0m。然而，当物体位于 4 米以外时，深度信息会变得非常不敏感，建议在 0.8~3.5 米范围内使用。在此深度范围内，可以分割为 2048 个等级，能够获取 6 平方米范围内所有物体的深度信息，横向角度范围为 57 度，纵向为 43 度。Kinect 每秒可以处理 30 帧的深度信息，每一帧由 640x480 个像素组成，每个像素用 11 位表示。同时，值得一提的是，其在识别人体动作的时候，精度可以达到 4cm。

Kinect 以主动方式获取深度信息，简要过程为：通过红外发射器（摄像头 I）发射随机结构光点阵，同时由红外线传感器（摄像头 III）捕捉这些点阵信息，通过对捕捉信息的处理，就可以获得每个像素的深度值。具体过程为：Kinect 采用人眼看不见的 class1 激光，红外发射器透过镜头前光栅向外界发射红外线模版，即将激光均匀的分布在被测量空间中，再透过红外线摄像机记录下空间中的每个散斑，保存原始数据为模版，再由居中的摄像头同步捕捉深度图像，并将捕捉的数据传入主机由感测晶片 PS1080 和体感侦测装置处理，与内部模版进行匹配，最终得到每个像素的 3D 深度值。

由于 Kinect 比市面上最便宜的深度摄像机的价格要降低十倍左右，普通用户完全可以接受，同时，这对于软件的普及也有一定的推动作用。因此，本文使用 Kinect 作为深度摄像头来获取环境的深度信息。与此同时 Kinect 是首个将深度摄像头和彩色摄像头相结合的设备，对于三维场景的重建起着至关重要的作用。

Kinect 不是源自传统的 3DV 的 ToF（time of flight）摄像头，因此，它不是基于 ToF 原理，也与结构光测量技术不同，而是采用了 PrimeSense 的一种三维测量技术，即光编码（Light Coding）技术。光编码技术和 ToF 技术的区别如表 4-1 所示。

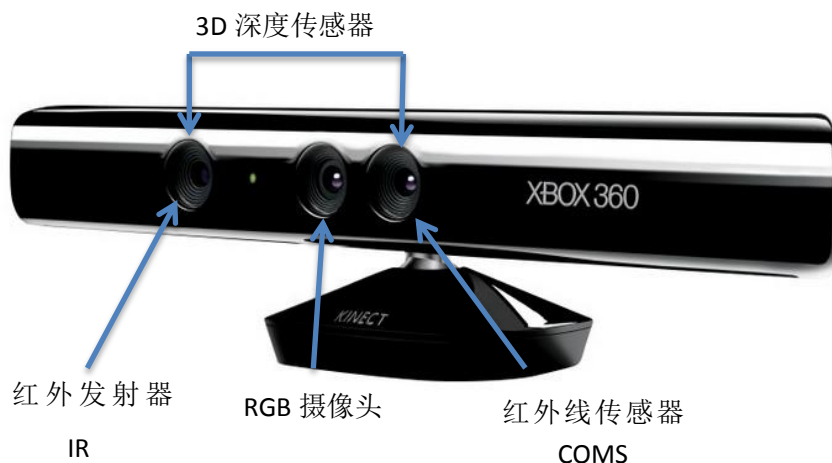


图 4-1 Kinect XBOX 360 的外观

表 4-1 光编码技术和 ToF 技术的区别

类别 \ 属性	光源	编码	装置	原理
ToF 技术	脉冲光	二维编码	加速度传感器+红外线	反射时间差测距与定位
光编码技术	连续光（近红外线）	“体编码”	感测晶片 PS1080+体感侦测装置+红外线	根据激光散斑图案的变化测距与定位

光编码，顾名思义就是利用光源照明对需要测量的空间编码，说到底还是属于结构光技术。但与传统的结构光技术不同的是：从光源发射出去的光并不是一幅周期性变化的二维的图像编码，而是一个具有三维景深的“体编码”。这种光源叫做激光散斑（Laser Speckle）。当激光照射到粗糙物体或穿透毛玻璃后会形成随机衍射斑点。这些散斑具有高度的随机性，而且会随着距离的不同变换图案。也就是说，空间中任意两处的散斑图案都是不同的。只要在空间中打上这样的结构光，整个空间就都被做了标记，把一个物体放进这个空间，只要寻找物体上面的散斑图案，就可以知道这个物体的位置。当然，在这之前要把整个空间的散斑图案都记录下来，所以要先做一次光源的标定。

在 PrimeSense 的专利上，标定的方法是这样的：每隔一段距离，取一个参考平面，把参考平面上的散斑图案记录下来。假设用户活动空间是距离电视机 1 米到 4 米的范围，每隔 10cm 取一个参考平面，那么标定下来我们就可以得到保

存的 30 幅散斑图像。需要进行测量的时候，拍摄一副待测场景的散斑图像，将这幅图像和我们保存下来的 30 幅参考图像依次做互相关运算，这样我们会得到 30 幅相关度图像，而空间中有物体存在的位置，在相关度图像上就会显示出峰值。把这些峰值一层层叠在一起，再经过一些插值，就可以得到整个场景的三维形状。

由于微软还没有向外公布深度信息的具体计算步骤，但可以通过反向工程得出其计算过程如公式 4-1 所示。

$$z(d) = a_1 \times \exp\left(-\left(\frac{d-b_1}{c_1}\right)^2\right) + a_2 \times \exp\left(-\left(\frac{d-b_2}{c_2}\right)^2\right) \quad (4-1)$$

其中， $a_1 = 3.169 \times 10^4$ ， $b_1 = 1338.0$ ， $c_1 = 140.4$ ， $a_2 = 6.334 \times 10^{18}$ ， $b_2 = 2.035 \times 10^4$ ， $c_2 = 3154.0$ 。

由上述介绍可知，Kinect 使用廉价的深度数据采集设备，其工作原理是使用红外光感设备采集深度信息，若要在光照条件过于复杂和有明显热源的环境中是，采集到的数据会有很强的干扰，因此，对用户所使用的环境有一定的要求。

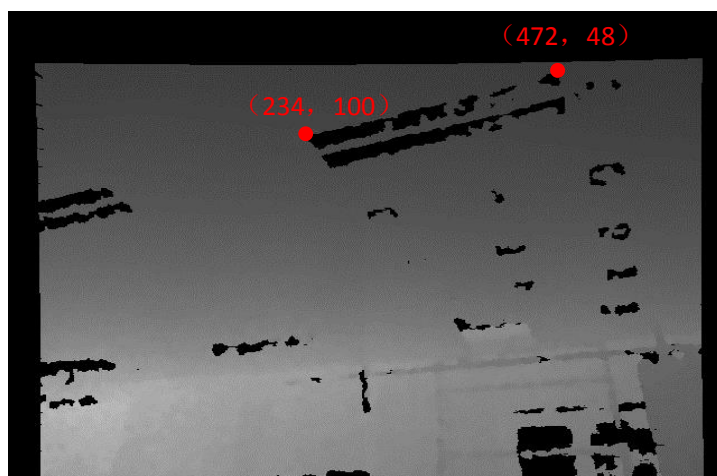
4.3 Kinect 数据采集与预处理

由于 Kinect 自身的条件限制，即 Kinect 需要电源，不能处于阳光直射的环境，阴影处的与室内环境相差不大，但不能用于阴影处与阳光直射区交界会出现相当大的误差，并且 Kinect for Xbox 不支持近景模式，有效视野范围为 0.8~3.5m。若想把深度图和彩色图合并显示，就会发现两者图像直接合并在一起会有误差，即彩色图像上的同一像素点在其对应的深度图像上的坐标不一样。

经实验可知，相对近处 y 轴方向相差 20 至 50 个单位像素，远处相差甚大。究其原因，Kinect 的深度摄像头和彩色摄像头大致在同一水平线上，但两者相距一定距离（实验选用的相距约 2.5cm），而且镜头本身的参数也完全不同，所以两个摄像头所采集到的图像画面会有些许的差异。如图 4-2 所示，该画面为用 OpenGL 显示的彩色图像和深度图像，所选场景为所在实验室，主要观察日光灯对应的深度图和彩色图的区域，图中已标出坐标以便比较。



(a) 室内场景的彩色图像



(b) 室内场景的深度图像

由于三维重建工作是以深度信息为基础，若存在视差就会给三维重建带来不准确偏差，因此，我们需要先对 Kinect 的彩色摄像头和深度信息获取摄像头进行校正，便于后续三维重建中对点云着色工作的进行。下图表示校正前后的对比图，为了便于观察，我们将彩色图像和深度图像合并在同一坐标系下，可以清楚的看到两者的变化。



(c) 校正前室内场景的深度与彩色合并后图像

图 4-2 室内场景的彩色图像与深度图像

根据 Heresy'space 的文章^[37]的理论，改写了可执行文件 Kinect.dll。经过校正，效果如图 4-3 所示。



图 4-3 校正后的室内场景的深度与彩色合并后图像

由上图可以看出，经过校准所采集到的深度图像和彩色图像几乎重合，达到了预想的效果，这样有利于生成 3D 点云。

4.4 基于 PTAM 位姿估计及 Kinect 的室内局部三维重建

第三章已经证明，PTAM 估计的相机姿态和轨迹准确度比较高，可以将其估计的结果用于三维重建。上一节介绍了 Kinect 数据采集及预处理，则可以获得

彩色图片序列，同时获得对应的深度信息。将彩色图片序列作为输入，并利用第二章所讲内容，可以求解出相机的姿态并画出轨迹。

但是，存在一个问题，由 Kinect 得到的点云坐标的坐标系和由 PTAM 得到的相机坐标的坐标系并不是同一个坐标系，要想进行三维重建就必须利用第二章的知识进行坐标系的变换。经实验证明，微软开发的 Kinect SDK-v1.7 只能实现场景中一帧的重建结果，也就是说，并没有将相机不同视角下的场景放在同一个坐标系下显示。要想实现这一点，就需要计算出每一帧中相机的外部参数，即旋转矩阵 R 和平移向量 T 。一旦有了相机的外部参数，就能够将相机不同视角下的场景转换到同一个坐标系下显示。

本论文获取的点云坐标均是定义在 OpenNI 坐标系下^[38]的，OpenNI 的坐标系一般这样规定：将相机的光心作为坐标轴原点，相机的前方作为 Z 轴的正方向，相机的上方作为 Y 轴正方向，相机的右方作为 X 轴的正方向，符合左手定则。此处需作说明，PTAM 做定义的坐标系满足右手定则，因此在计算过程中存在一个负号的差异。如图 4-4 所示。

那么，将 OpenNI 定义的坐标系的相机坐标系和 PTAM 定义的坐标系的相机坐标系分别带入式 2-2，进行求解可以得到以下关系：

$$\begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = R^{-1} \left(\begin{bmatrix} X_{OpenNI} \\ Y_{OpenNI} \\ Z_{OpenNI} \end{bmatrix} - T \right) \quad (4-2)$$

由此可以看出，我们就可以把得到的所有帧的数据先保存，再经上式转换到同一坐标系下显示出来，就可以得到多帧场景的三维重建结果。以下是实验验证重建的场景以及实验的重建结果。如图 4-5、4-6 所示。

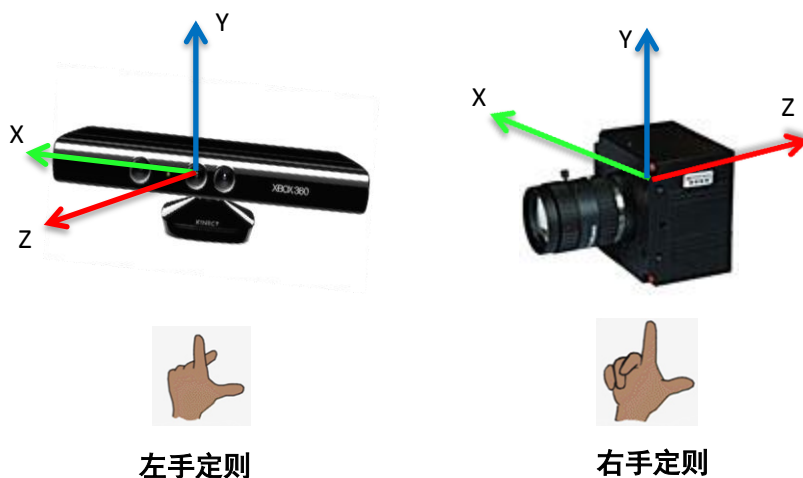


图 4-4 OpenNI 及 PTAM 定义的坐标系

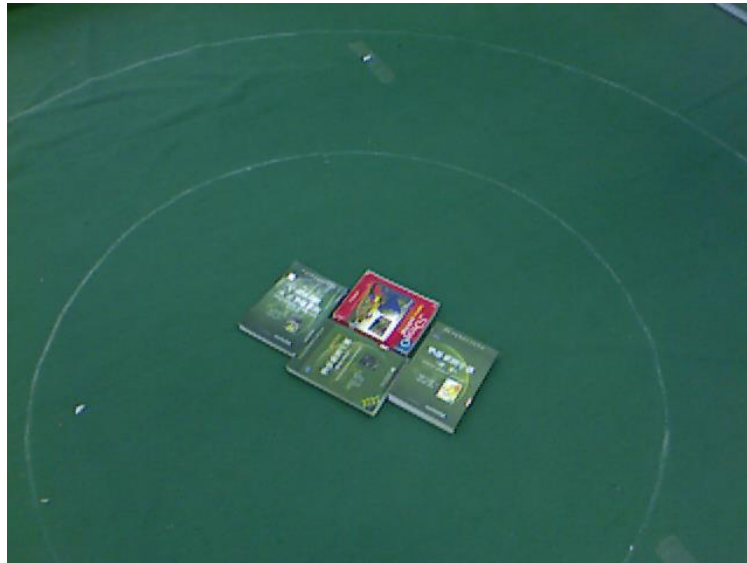


图 4-5 需要重建的场景



图 4-6 场景重建结果

上述验证结果数据来源于第三章对 PTAM 系统精度估计时所建数据库中的第三组数据，Kinect 的移动轨迹为沿所画白色圆形内圈移动。可以看出，重建结果令人满意。但是，虽然 PTAM 系统能计算出精度很高的相机参数，但是对输入源的要求很高，比如，相邻帧之间如果没有匹配的特征点，那么 PTAM 系统整个跟踪过程就会失效，导致计算出的相机参数出错。并且使用 PTAM 来估算相机参数的这种方法不太适用于大场景。

4.5 基于 RGB-D 多传感器融合的三维重建系统

4.5.1 硬件系统

以微软生产的 Kinect 为代表的配备 RGB-D 摄像头^[39]的装置系统是一种新奇的传感器系统,它可以在捕获 RGB 彩色图像的同时捕获到每个像素的深度信息。我们以移动机器人为载体,将 Xbox360 版的 Kinect 固定于其顶部,构建成简单的地面平台系统,如图 4-7 所示。



图 4-7 三维重建移动平台

让上述移动机器人具有上下高度伸缩的功能,故可以在三维重建时调节高度,由于 Kinect 可以在 0.8~3.5m 内重建效果较为理想,再加上高度的调节以及 Kinect 自身角度的自由调节,足以完成在室内三位重建的任务。

4.5.2 软件系统

设计的软件系统主要是在前人研究^[40-45]的基础上,对其进行算法改进,并大胆尝试新思路,得到了预期的效果。主要过程步骤如下:

1 特征提取与匹配

采用 SURF 算法进行特征提取与匹配。特征的提取利用计算近似的 Hessian 矩阵的行列式的极值来确定特征点的位置。输入一幅图像 I , 对图上每一个像素点 $x=(x, y)$, 利用公式(4-3) 计算出其在 x 处尺度为 σ 的 Hessian 矩阵 $H(x, \sigma)$ 。

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (4-3)$$

其中, $L_{xx}(x, \sigma)$ 是 Gaussian 二阶偏导数 $\frac{\partial^2}{\partial x^2}$ 在 x 处对图像 I 的卷积。继而, 计算每个 H 的行列式:

$$\det(H_{approx}) = D_{xx} D_{yy} - (0.9 D_{xy})^2 \quad (4-4)$$

其中, D_{xx} 是盒子模版与图像的卷积, 用 D_{xx} 近似代替 L_{xx} 。接着, 判断行列式(4-4)的符号。若符号为负, 则不是局部极值点; 若符号为正, 则该点归类为极值点。

接着, 构建图像的尺度空间, 即图像在不同解析度下的表示。与 SIFT 建立金字塔的原理相反, SURF 通过不同尺寸盒子作为过滤模版和积分图像求取 Hessian 矩阵行列式的响应图像, 然后, 在响应图像上采用 3D 非极大值抑制, 求取各个不同尺度的斑点。使用 $3 \times 3 \times 3$ 的模版在三维空间内进行非极大值抑制, 根据预设的 Hessian 阈值 H 比较, 当 h 大于 H , 且比邻近的 26 个点的相应值都大时, 被选为特征点, 然后采用三维线性插值使其更精确。

为了保证特征矢量具有旋转不变性, 需要给每一个特征点分配一个主方向。以特征点为中心, 以 $6s$ 为半径的圆内, 对图像进行 Haar 小波响应运算, 设计一个以方向为中心, 张角为 $\pi/3d$ 的扇形滑动窗口, 以步长为 0.2 弧度, 转动滑动窗口, 并对窗口内所有点的 Haar 响应值进行累加, 主方向为最大的 Haar 响应累加值对应的方向。

进行完特征点提取后进行特征点的匹配, 有如下步骤:

1) 在计算 H 矩阵的行列式时, 同时计算出矩阵的迹。由于按亮度不同, 将特征点分为两种: ①特征点及其周围小邻域的亮度大于背景区域, H 矩阵的迹为

正；②特征点及其周围小邻域的亮度小于背景区域，H 矩阵的迹为负。依此特性，先对两个特征点的 H 矩阵的迹进行比较：同号，说明两个特征点具有相同对比度；异号，说明对比度不同，则放弃 2) 的相似性度量；

2) 在 1) 的基础上，对两个特征点描述子的相似性度量，采用欧式距离进行计算：

$$Dis_{ij} = \left[\sum_{k=0}^{k=n} (X_{ik} - X_{jk})^2 \right]^{1/2} \quad (4-5)$$

其中， X_{ik} 表示待配准图中第 i 个特征描述子的第 k 个元素， X_{jk} 表示参考图中第 j 个特征描述子的第 k 个元素， n 表示特征向量的维数。

对于待配准图上的特征点，计算其到参考图像上所有特征点的欧式距离，得到一个距离集合。通过对距离集合进行比较运算得到小欧式距离和次欧式距离。设定阈值，一般为 0.8，当最小欧式距离和次欧式距离的比值小于该阈值时，认为特征点与对应最小欧式距离的特征点是匹配的，否则，没有点与该特征点相匹配。阈值越小，匹配越稳定，但极值点越少。

2 位姿估计与优化

针对位姿估计和优化的三维重建中点云数据配准问题，为了防止 ICP 算法收敛到局部最小点，产生错误的配准结果，则先采用 RANSAC 配准算法进行粗匹配，并结合深度信息进行初始位姿估计，将计算出结果作为 ICP 配准算法的初始值，再次使用深度信息进行 ICP 算法对初始位姿的优化实现由粗到细的配准策略，并减少迭代次数，达到令人满意的收敛结果。

具体步骤如下：

1) 利用 RANSAC 算法划分出“局外点”和“局内点”，剔除匹配异常点，将得到的所有局内点随机建立点云集合，先进行粗匹配，设得到的两个点集分别是 S 和 T ，从源数据集 S 中随机抽取不在一条直线上的三个点 $\{S_a, S_b, S_c\}$ ，在目标点集 T 中，搜索对应的三个点 $\{T_a, T_b, T_c\}$ ，利用三对控制点进行欧式变换矩阵的估计，计算出 H_c 。然后计算在估计变换关系 H_c 误差阈值 δ 下，两个点集之间的一致程度，并利用对应点的深度信息得到初始位姿的估计；

2) 以 1) 计算出的结果作为 ICP 算法的输入，避免单调收敛到局部最小值，根据最小化目标函数估计配准参数 $H_k = \begin{bmatrix} R_k & T_k \\ 0 & 1 \end{bmatrix}$ ，并计算估计误差 d_k ；使用变换

矩阵 H_k 对数据集 P_k 更新, $P_{k+1} = H_k(P_0) = R_k P_0 + T_k$; 重复迭代, 当两次计算的估计误差变化小于阈值 τ 时, 即 $\|d_k - d_{k+1}\| < \tau$ 时, 停止。

3 Hog-man 地图优化

得到经优化的初始位姿后, 将其作为节点, 与其相邻的帧间的位姿关系作为边添加到图中, 进而采用 Hog-man 图优化算法进行地图的全局优化, 得到高精度的三维地图。

由于当想要建立一块区域完整的信息时, 会在这块区域周围采集较多的点属于一种包围式的建立。因此, 核心思想是根据相机观察中心的距离值(如式(4-7))作为标准将原始的图划分为多个子图, 以子图中的一个节点来表示该子图, 从而得到原始图一层的抽象, 对得到的图再依次进行抽象, 就得到一个多层图结构, 提取出原始图中的拓扑结构。

$$IP = P[0 \quad d \quad 0 \quad 1]^T \quad (4-6)$$

其中, 设 O 为摄像机中心, IP 为兴趣中心点, d 为 IP 到 O 的距离, 这里取 2m。假设当前相机的位姿用一个 4×4 的矩阵 P 表示, 则 IP 的坐标为使用观察中心代替相机中心作为子图分割标准, 更好地反应建立完整地图时的节点分布特点。

当有新节点加入到图中时, 首先是加入到原始图中, 然后查看是否改变子图划分, 如果有改变就需要更新高层图, 同时对最顶层进行优化, 只有当顶层的拓扑结构发生很大改变的时候才由顶层图向底层图反向传递, 更新底层图。这样就保证了实时的优化。

根据得到的优化结果进行简单叠加得到为了去除直接叠加所有帧带来的冗余点, 采用的基于空间体的点云融合方法。其主要思想是在对多帧点云数据进行融合时, 首先建立一个包含所有帧的空间体, 空间体以一定的分辨率划分为空间像素点, 每个点上将记录到环境中物体的距离, 然后依次对每一帧进行处理, 修改空间像素点的值。最后空间体中的零值像素即为最后的曲面上的点, 这样得到的曲面是具有最小二乘性质的曲面。同时为了反映采集到的每帧数据的一些特点, 在进行距离值叠加的时候, 采用权值函数, 最后的叠加方程如式(4-7) (4-8)。

$$D_{i+1}(x) = \frac{W_1(x)D_1(x) + W_{i-1}(x)D_{i-1}(x)}{W_1(x) + W_{i-1}(x)} \quad (4-7)$$

$$W_{i+1}(x) = W_i(x) + W_{i-1}(x) \quad (4-8)$$

$$W(x) = \frac{x-x^*}{Err(x^*)} \quad (4-9)$$

$$Err(dist) = dist^2 \times 0.0035 \quad (4-10)$$

实验中使用 Kinect 摄像头作为 RGB-D 摄像头，因此，根据 Kinect 传感器的误差方程(4-10)权值函数的选取如式(4-9)。其中 x^* 为单帧曲面上点的到相机中心的距离，经过计算最后得到 3D 点云图，即整个场景的三维重建效果图。

4.6 三维场景重建结果

进行三维重建的场景是室内一色彩较为丰富的感兴趣环境区域，以下是 Kinect 旋转 90 度和 180 度时重建的结果。如图 4-8、4-9 所示。

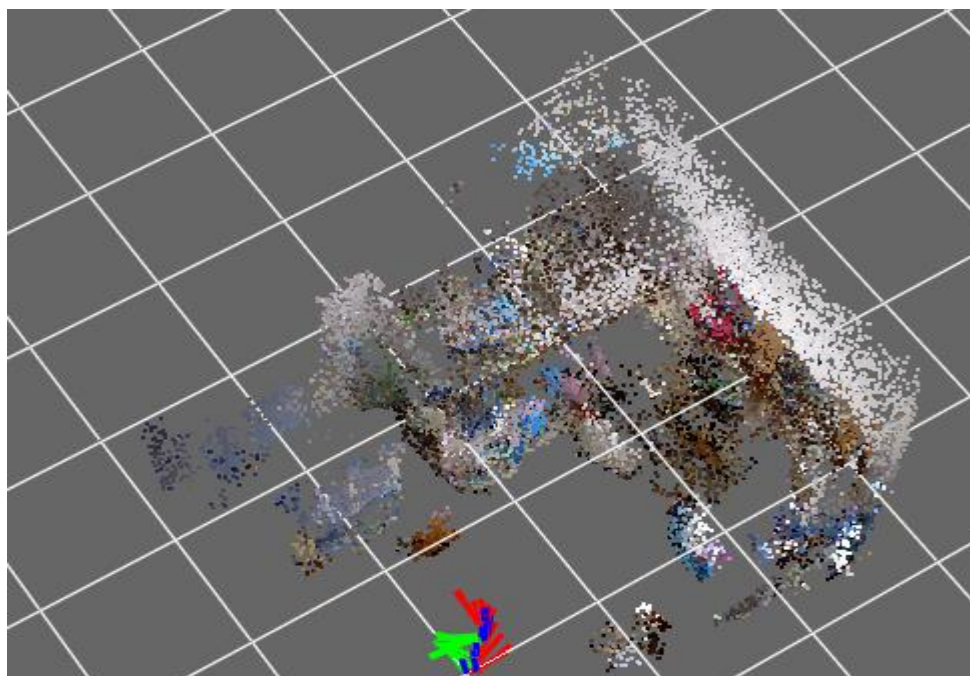


图 4-8 旋转90°的室内三维重建效果图

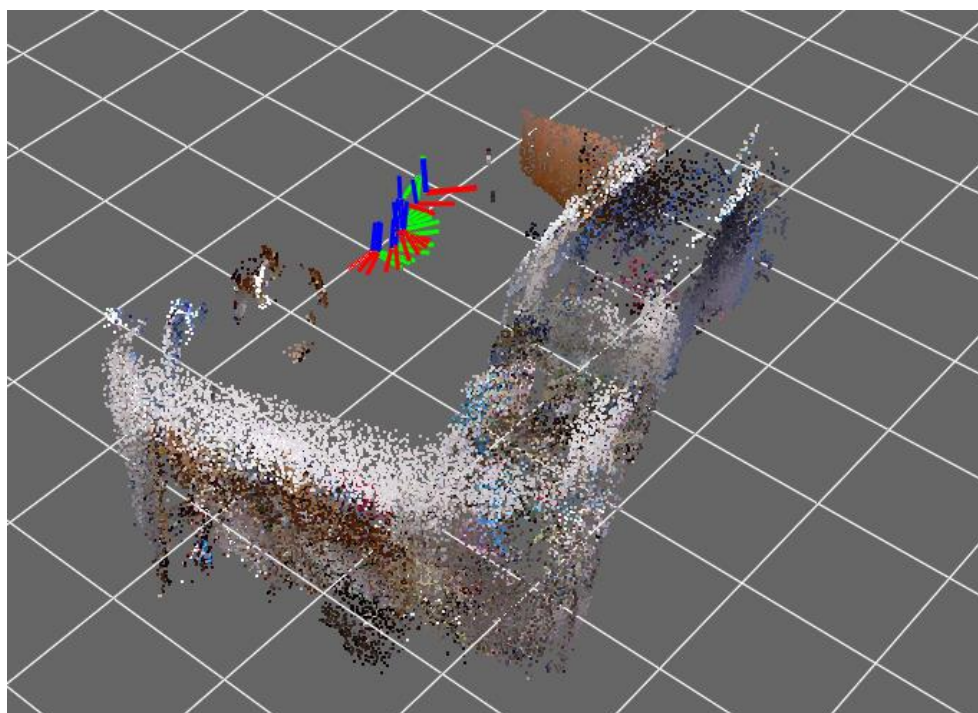


图 4-9 旋转180°的室内三维重建效果图

4.7 实验结果分析

从以上重建结果可以看出，稠密重建达到了不错的效果，为了说明其准确性，我们截取了一部分相机参数，其中平移向量 T 用其三个分量 T_1 、 T_2 、 T_3 ，旋转矩阵 R 绕 X 轴、 Y 轴、 Z 轴的角度分别用 φ 、 θ 、 ω 来表示。如下表 4-2 所示。

由所得到的数据分析可以得到，除了旋转矩阵绕 X 轴方向旋转的角度值每个间隔相差大约 10 左右，其他平移向量个分量及旋转矩阵绕 Y 轴和 Z 轴旋转的角度值没有大的变化，经过归一化分析，与 Kinect 匀速移动的角度差相比，两者对应。

4.8 本章小结

本章着重对基于 RGB-D 多传感器融合的移动平台室内三维重建进行了研究，主要工作包括以下几点：

(1) 本章详述了 Kinect 的工作原理，利用其自带的 RGB-D 多传感器可以在采集彩色图像的同时获取深度信息的优点，进行数据的采集。由于采集设备自身设计的不足，深度图像和彩色图像存在一定的误差，对其进行了矫正，并将采

表 4-2 平移向量各分量及旋转矩阵分别绕各坐标轴得到的参数

	平移向量 T			旋转矩阵 R		
	T_1	T_2	T_3	φ	ϑ	ω
1	0.361745	0.656592	0.222123	110.724039	1.328114	-4.557694
2	0.351568	-0.723133	0.172582	122.332265	-6.318515	2.163312
3	0.389416	-0.729430	0.177265	133.283609	-6.365624	1.605008
4	0.417756	0.672151	0.218636	-107.901505	-2.548138	-1.918024
5	0.413355	0.664369	0.223471	-114.308732	-1.286966	-2.57167
6	0.437177	0.676910	0.179113	-128.330505	-3.682801	-0.773887
7	0.593904	0.533957	0.302219	-141.272773	-2.616172	-0.299501
8	0.540621	0.581749	0.197769	-136.398136	-3.898397	-0.726061
9	0.540930	0.497355	0.195870	-142.758549	-4.274315	-1.193265
10	0.633448	0.280283	0.209914	-153.702067	-4.179787	-1.865838
11	0.658592	0.324099	0.112526	-141.494682	-5.707119	-1.721343
12	0.687362	0.164027	0.166947	-157.133148	-4.881384	-1.953846
13	0.703102	-0.256590	0.228318	-168.219178	-4.489408	-1.556988
14	0.710245	-0.350697	0.312153	173.030897	-3.414093	-1.401098
15	0.710334	-0.160617	0.200807	-183.113463	-4.800128	-1.689767

集到的数据按离线与实时来进行室内场景的三维重建。离线主要是对感兴趣区域分别进行单帧和多帧的重建，达到了不错的效果。

(2) 同时，提出了基于 RGB-D 多传感器融合的三维重建算法，并详述了其工作过程。主要包括下述步骤：特征提取匹配、位姿估计优化和图优化及图融合处理。(a) 使用 RGB-D 传感器采集图像，采用 SURF 算法：建立尺度空间，先利用 Hessian 矩阵确定候选点，然后进行非极大抑制，进而，对 RGB 颜色信息进行特征的提取，并对得到的稳定特征点进行相互匹配；(b) 经过点对点之间的配准后，采用基于仿射不变量的 RANSAC 算法，并结合深度信息进行位姿估计，经计算得到摄像机的初始位姿，然后再次使用深度信息进行 ICP 算法对初始位姿的优化。(c) 得到经优化的初始位姿后，将其作为节点，与其相邻的帧间的位姿关系作为边添加到图中，进而采用 Hog-man 图优化算法进行地图的全局优化，得到高精度的三维地图。最后，根据得到的优化结果进行简单叠加得到 3D 点云图，即整个场景的三维重建效果图。

(3) 本章提出的三维重建算法是对原有算法的改进。主要涉及以下方面：由于 SURF 算法在亮度变化匹配较好且计算速度较高，本章采用 SURF 算法进行特征匹配，因此在模糊处理和计算速度性能更优，从此阶段的速度来看，约为 SIFT 的 3 倍。由于改进了子地图的划分标准，使得能更好的生成大尺度的 3D 地图。同时，在地图生成部分，利用经典的基于空间体的多帧融合方法减少冗余点，不仅提高了实时性，更为后续操作处理提供了准确的数据。

第五章 结束语

5.1 工作总结

三维重建技术在计算机视觉领域中发挥着重要作用,不仅在军事上大有用武之地,在生活上也给人们带来了便利。近些年,关于三维重建技术的研究也取得了很大的进步。本文针对三维重建技术,开展了深入的研究,主要工作和创新点包括以下方面:

(1)了解了三维重建技术的研究背景和意义,分析了当前国内外主流三维重建技术的研究现状以及研究关键与难点,总结了其研究优缺点及适用场合,并总结出三维重建技术广阔的发展前景,发掘出了三维重建技术对于人类生活和军事科技发展等方面的重要性。

(2)详细了解了三维重建技术的有关基础知识和理论。介绍了以最简单的小孔成像为原理的摄像机针孔模型,并分别简介了图像坐标系、相机坐标系、世界坐标系以及图像金字塔。在以上理论的基础上,详细介绍了并行跟踪与地图绘制(PTAM)系统的原理与算法。

(3)在学习 PTAM 算法理论的基础上,根据实际需要,改写程序使其满足简便直观地对其进行相机位姿估计以及运动轨迹的精确度的评估。同时,为了满足数据要求,建立了一个图片数据库来将其作为输入,经计算求解出旋转矩阵 R 和平移向量 T ,分析各组数据的相机姿态和轨迹结果并得出评估结果,实验证明,PTAM 系统具有较高的精度。

(4)一方面,将 RGB-D 传感器与 PTAM 系统相结合,对场景进行稠密重建,通过 Kinect 获取 RGB 图像序列,并作为 PTAM 系统的输入求解出移动平台的位置与姿态并画出运动轨迹。与此同时,通过 Kinect 可以获取需要重建场景的深度信息,并由此得到场景的 3D 点云,然后使用 OpenGL 对点云进行了着色,并将得到的每一帧的 3D 点云经公式变换放于同一世界坐标系下。另一方面,对 RGB-D SLAM 算法进行改进,分别在于三个方面,一是将 SIFT 算法替换为 SURF 算法进行特征提取和匹配,提高了此阶段的效率;二是使多层改进子地图划分标准图抽象可以更好的反映环境的拓扑结构,并具有更高的效率;三是在地图输出阶段添加冗余点去除模块,使得最后得到的地图数据量减少,同时为后续操作提供更准确的数据。结果证明该改进在建立 3D 地图时是完全有效的。

5.2 展望

本文虽然在三维重建技术的研究上取得了一些进展,但是还存在一些问题值得深入研究的方面,主要包括以下方面:

(1) PTAM 系统与 RGB-D 传感器结合使用时的实时性问题

本论文没有实现 PTAM 系统与 RGB-D 传感器结合进行三维场景重建的实时性。究其原因:首先,本论文实验所使用的 Kinect Xbox 有一部分技术研究还被微软所垄断,研究者们个人的开发也是基于源技术的创新与开发,因此,还没有找到将 RGB-D 传感器获取的数据格式转换成 PTAM 能够读取的数据格式的方法或者两者通用的数据格式;其次,PTAM 系统在运行时,需要手动先对场景地图初始化,这必然导致人为的差异化,并且 PTAM 系统只能接受一个摄像机,当多个摄像机同时运行时,会提示出错。

(2) RGB-D 传感器自身的局限性

RGB-D 传感器用于三维重建是以深度信息(即距离)为基准进行重建,在使用时都会有明确的距离使用范围,以本文为例, Kinect Xbox 的有效视野范围为 0.8 ~3.5m,但是,经验证不同型号的 Kinect 会有差异,最远有效视野范围可达 5.5m。如果拍摄的距离要求小于 0.8m 或大于 3.5m 时,那么产生的结果就会不可靠,或者会发生该超出区域被不能被重建出的情况。

除此之外,不推荐在室外环境下使用 Kinect Xbox。Kinect 是利用结构光进行深度信息的捕获的,而当在室外阳光刺眼的情况下或明暗交界处, Kinect 的重建就会失效,并且本实验使用的 Kinect Xbox 必须要求插上电源,才能使用,这就无形中也给室外环境重建带来了不小的麻烦。在此推荐可以使用华硕的 WAVIXtion,完全克服了 Kinect Xbox 在硬件设备的缺陷。

虽然,三维重建技术目前在方法和设备上都有所限制,但是我们完全有理由相信,在研究者的努力下,三维重建技术会越来越完善,成为生活中和科研领域不可缺少的一部分。

参考文献

- [1] E. Bondarev, F. Heredia, Favier, R.J.J., Ma, L.. On photo-realistic 3d reconstruction of large-scale and arbitrary-shaped environments. *10th IEEE Consumer Communications and Networking Conference, Las Vegas, Nevada*, 2013.1, pp. 606-609.
- [2] 娄静涛. 面向室内三维重建的单幅全向图水平直线检测与定位. 国防科学技术大学硕士学位论文, 2009.
- [3] L.G. Roberts. Machine perception of three-dimensional solids. In *Optical and Electro-optical Information Processing*. MIT Press, Cambridge, 1965: 159-197.
- [4] D. Marr, Vision. W. H. *Freeman and Company*, 1982.
- [5] D. Marr. Vision. A computational investigation into the human representation and processing of visual information. *San Francisco Freeman and Company*, 1982.
- [6] Cochran, D. Steven, Medioni, Gerard, 3-D surface description from binocular stereo. *IEEE trails, PAMI*, 1992, 14(10):981-994.
- [7] D. Nist ér, O. Naroditsky, and J. R. Bergen. Visual odometry. *IEEE Intl. Conference on Computer Vision and Pattern Recognition(CVPR '04), Washington, D.C.2005. IEEE Computer Society*, pp.652–659.
- [8] Minoru Asada, Takamaro, T. Visual Tracking Of Unknown Moving Object By on Multisensor Fusion and Intelligent Systems. 1999.
- [9] Okutomi M,Kanade T.A mutiple-baseline stereo. *IEEE Transanctions on Pattern Analysis and Machine Intelligenc*.1993, 15(4):353.
- [10] 袁亮,姜光. 三维重建过程中的点云数据配准算法的研究.西安电子科技大学硕士学位论文, 2010, 01, pp.30-43.
- [11] 庄严, 王伟等. 移动机器人基于激光测距和单目视觉的室内同时定位和地图构建, 自动化学报, 2005, 31(6), pp. 925-933.
- [12] 武二永. 基于视觉的机器人同时定位与地图构建. 杭州, 浙江大学, 2007.
- [13] 林强,李宏东,顾伟康. 一种新颖实用的基于视觉导航的三维重建算法.传感技术学报, 2003年04期

- [14] K. Georg, M. David. Parallel tracking and mapping for small AR workspaces. *Active Vision Laboratory Department of Engineering Science University of Oxford*.
- [15] G. Klein and D. W. Murray. Parallel tracking and mapping on a camera phone. *In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR)*, 2009.
- [16] I. Mahon, S. B. Williams, O. Pizarro. Efficient view-based SLAM using visual loop closures. *Robotics, IEEE Transactions*, 2008, 24(5), pp. 1002-1014.
- [17] G. Silveira, E. Malis, P. Rives. An efficient direct method for improving visual SLAM. *Robotics and Automation, 2007 IEEE International Conference on. IEEE*, 2007, pp.4090-4095.
- [18] L. M. Paz, P. Piniés, J. D. Tardós, et al. Large-scale 6-DOF SLAM with stereo-in-hand. *Robotics, IEEE Transactions on*, 2008, 24(5), pp. 946-957.
- [19] J. Aulinas, Y. Petillot, J. Salvi, et al. The SLAM problem: a survey. *Proceeding of the 2008 conference on Artificial Intelligence Research and Development*. 2008, pp.363-371.
- [20] Y. Jeong, D. Nister, D. Steedly, R. Szeliski, and I.S. Kweon. Pushing the envelope of modern methods for bundle adjustment. *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, 2010, pp. 1474-1481.
- [21] A. S. Huang, A. Bachrach, P. Henry, et al. Visual odometry and mapping for autonomous flight using an RGB-D camera. *Proc. IEEE International Symposium of Robotics Research (ISRR)*. 2011.
- [22] S. Izadi, D. Kim, O. Hilliges, et al. KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera. *24th annual ACM symposium on User interface software and technology. ACM*, 2011, pp.559-568.
- [23] N. Engelhard, F. Endres, J. Hess, et al. Real-time 3D visual SLAM with a hand-held RGB-D camera. *RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden*. 2011.
- [24] N. Karlsson, J. Ostrowski. The vSLAM algorithm for robust localization and mapping. *Proceedings of the 2005 IEEE International Conference*, 2005, pp.24-49.
- [25] J. Valasek, K. Gumam. Vision based sensor and navigation system for autonomous air refueling. *Journal of Guidance, Control, and Dynamics*, 2005, 28(5), pp.979-989.

- [26] A. Nabil, V. Sazdovski. Low altitude airborne SLAM with INS aided vision system. *AIAA Guidance, Navigation and Control Conference and Exhibit, South Carolina, AIAA*. 2007.
- [27] E. Mouragnon, F. Dekeyser, P. Sayd, M. Lhuillier, and M. Dhome. Real time localization and 3D reconstruction. In *Proc. IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR '06), New York, 2006*, pp. 363–370.
- [28] J. Sturm, N. Engelhard, F. Endres, W. Burgard, D. Cremers .A benchmark for the evaluation of RGB-D SLAM systems , *Proc. of the International Conference on Intelligent Robot Systems (IROS)*, 2012.
- [29] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard, An evaluation of the RGB-D SLAM system, in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2012.
- [30] N. Engelhard, F. Endres, J. Hess, J. Sturm, W. Burgard. Real-time 3D visual SLAM with a hand-held camera, *Proc. of the RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum*, 2011.
- [31] J. Sturm, S. Magnenat, N. Engelhard, F. Pomerleau, F. Colas, W. Burgard, D. Cremers, R. Siegwart. Towards a benchmark for RGB-D SLAM evaluation , *Proc. of the RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf. (RSS)*, 2011.
- [32] 陈卫东, 张飞. 移动机器人的同步自定位与地图创建研究进展. *控制理论与应用*, 2005, 22(3), pp.455- 460.
- [33] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communcations of the ACM*, 1981, pp.381–395.
- [34] T. Bailey, J. Nieto, J. Guivant. Consistency of the EKF-SLAM algorithm intelligent robots and systems, 2006 IEEE/RSJ International Conference on. IEEE, 2006, pp. 3562-3568.
- [35] 郭利进, 王化祥, 孟庆浩等. 一种改进的粒子滤波 SLAM 算法. *计算机应用研究*, 2008, pp. 1698-1700.
- [36] J. Smisek, M. Jancosek, and T. Pajdla, 3D with Kinect, *ICCV Workshop on Consumer Depth Cameras for Computer Vision*, 2011.
- [37] Heresy'space 《透过 OpenNI 合并 Kinect 深度以及彩色影像资料》
http://kheresy.wordpress.com/2011/01/21/combine_depth_and_image_from_kinect/

- [38] PrimeSense, Willow Garage, SideKick and Asus, “Introducing OpenNI”
<http://www.openni.org>.
- [39] C. Kerl, J. Sturm, D. Cremers. Robust odometry estimation for RGB-D cameras.
IEEE Int. Conf. on Robotics and Automation (ICRA), 2013.
- [40] F. Steinbrücker, J. Sturm, and D. Cremers, Real-time visual odometry from dense RGB-D images, *ICCV Workshop on Live Dense Reconstruction with Moving Cameras*, 2011.
- [41] P. Henry , M. Krainin, E. Herbst, et al. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. *12th International Symposium on Experimental Robotics (ISER)*. 2010, pp. 22-25.
- [42] J. Sturm, S. Magnenat, N. Engelhard, et al. Towards a benchmark for RGB-D SLAM evaluation. *RGB-D Workshop on Advanced Reasoning with Depth Cameras at Robotics: Science and Systems Conf.(RSS)*, 2011, pp.1-3.
- [43] J. Shi and C. Tomasi. Good features to track. *IEEE Intl. Conference on Computer Vision and Pattern Recognition (CVPR'94)*, IEEE Computer Society, 1994, pp.593–600.
- [44] Xiaodong Li, A. Nemra, N. Aouf. 3D mapping based VSLAM for UAVs. *20th Mediterranean Conference on Control & Automation (MED)*, 2012, pp.348-352.
- [45] E. Hernandez, J. Neira. Visual SLAM with oriented landmarks and partial odometry. *21st International Conference of Electrical Communications and Computers (CONIELECOMP)*, 2011, pp.39-45.

致 谢

首先，谨以此文献给我的毕业设计指导老师杨涛副教授，衷心感谢他长久以来的悉心指导与教诲。从大三下学期开始，杨老师担任我们的数字图像处理课程的老师，其课堂上风趣幽默却不失治学的严谨，恰到好处的小小鼓励时常会激发大家勇于提出新想法新观点的勇气，同时也让我对图像这个领域充满了浓浓的求知欲和探索欲。也许，数字图像处理课每周的大作业“虐”怕了很多人，因此，毕业季大家都选择了相对轻松的毕设选题。但是，将这种“虐”变成一种习惯，何尝不是一件终身受益的选择呢。我选择了将“虐”进行到底，所以在此我要再次向杨老师道一声，“谢谢”。

特别要感谢实验室的师兄师姐们给予我的帮助！感谢非说自己不是博士的博士冉令燕师兄毕设初期的帮助，感谢张晓强师兄采集数据时每一步细心的讲解，感谢马文广师兄大大小小、理论非理论、后勤非后勤的各种各样的琐事的帮助与解答，感谢仝小敏师姐时不时的调节气氛，让大家以愉悦的心情工作，感谢陈挺、卓涛、肖彬师兄以及屈冰欣、宋征玺、张福俊师姐的帮助与关心。在大学的最后几个月里，可以融进这个温馨的大家庭里，实感荣幸、开心与满足。

特别要感谢和我奋战在统一战线的王斯丙战友，我相信他在研究生阶段会在这个大家庭里成长很多，更会收获很多。同时还有邓邓邓邓以及半个的室友苟丽、一个半室友南文倩，还有大大以及陪伴我大学四年的你们，每一个走进来又走出去的你们。

还有，Mon peti ami，谢谢你一直的陪伴，耐心的讲解，还有不懂的时候一次次地被骂哭，我会慢慢变得和你一样强大。还有，我亲爱的爸爸妈妈和姐姐，谢谢你们大家。

最后，还有感谢一根筋的自己坚持了这个课题，坚持了一段难忘的时光。在整个毕设期间，从刚开始的图像拼接但现在的三维重建，从简单的 $AX=B$ 到难懂的 RGB-D 算法，从一次次的小失败和小成功，从 word 排版到 ubuntu 系统下三维场景重建的研究尝试，我也要感谢自己可以坚持下来。

作者：范洁

2013 年 6 月 12 日写于西北工业大学

毕业设计小结

“一图胜千言”，这句话让我与图像处理结缘，进而我选择了数字视频研究组，在这个气氛活跃业绩显著的优秀团队中。在这里，我不但收获了专业知识，独立思考、分析解决问题的能力，而且培养的科学研究严谨精神，乐观迎接困难的生活态度。也许，我可以这样说，在做毕设短短的几个月时间内，我遇到的困难层出不穷，从硬件到软件，从安装到编译，我很开心自己一直在勇往直前，乐观的解决一个又一个的“横道石”。我深深体会到，挫折提升成功的道理。

现在毕业设计工作已经接近尾声，回望之前留下的每一个脚印，心里更多的是欣慰与欣喜，过去的每一天都记录着我的成长。在此，感谢团队里的老师悉心栽培以及师兄师姐的帮助，让我真正学会了科学研究，和之前自己称作的研究大相径庭，并更胜一筹。感谢杨老师的信任让我参与了国家自然科学基金等研究工作，以及在研究各方面给与我的支持。

有些路很远，走下去很累，可是，不走会后悔。这充实同时被困难与阻碍填满的成长路，我一直在路上。

学术论文、专利、科研项目与获奖

一、发表和撰写的学术论文

- [1] Fei Liu, Yinan Sang, Jie He, **Jie Fan**, Ruichao Li, Xiongyi Cui, Haoyu Li, Jie Chen, Northwestern Polytechnical University Team Entry for the 2012 AUVSI International Aerial Robotics Competition, International Aerial Robotics Competition (IARC), Aug. 2012, pp.1-12
- [2] Zhaoli Liu, Cunjian Zhang, Fei Liu, **Jie Fan**. Low-Carbon economic of undeveloped areas, Shandong Province Soft Science Research Projects, administration research, Nov.2012

二、已申请专利

- [1] 杨涛, 张艳宁, **范洁**, 王斯丙. 基于 RGB-D 多传感器融合的场景三维重建
- [2] 杨涛, 张艳宁, 王斯丙, 马文广, 仝小敏, **范洁**. 基于彩色深度融合的虚拟阵列合成孔径透视成像

三、参与科研项目

- [1] 国家自然科学基金面上项目“移动相机虚拟阵列合成孔径成像与视频目标跟踪研究”, 2013-2016, 参与移动平台相机位姿估计法评估研究;
- [2] 西北工业大学“翱翔之星”计划“基于多摄像机动态视觉感知网络的遮挡目标透视成像与鲁棒跟踪研究”, 2012-2014, 负责多传感器融合的场景三维重建研究;
- [3] 国家大学生创新项目(数模基地)“移动机器人路径规划算法研究及仿真系统平台设计与实现”, 2009-2012, 参与移动机器人路径规划算法研究;
- [4] 国家大学生创新项目(数模基地)“基因表达谱数据挖掘方法研究”, 2011-2012, 参与数据挖掘方法的研究;
- [5] 山东省软科学研究计划项目“欠发达地区低碳经济发展研究——以菏泽为例”, 项目编号 2011RKGB1080, 2012-2013, 参与数据库处理的工作。

四、所获奖项

- 2012年 分别获得第六届国际空中机器人大赛(IARC) *Innovative Design Award* 和 *Excellent Team Culture Award*
- 2011年 西北工业大学“工大正禾杯”数学建模竞赛二等奖