

**TONGJI UNIVERSITY**

专业型硕士研究生学位论文选题报告及工作计划

**课 题 名 称 基于多源融合三维重建的**

**语义结构化地图构建研究**

**学 号 1732940**

**研 究 生 何士波**

**专 业 控制工程**

**所在院、系 电子与信息工程学院控制系**

**导 师 岳继光**

**副 导 师 董延超**

**选 题 时间 2018年10月16日**

**同济大学研究生院**

**2018年10月24日**

1. 研究问题

|  |
| --- |
| 1. 在预研究的基础上提出应用研究（设计）中的科学问题   本课题仅针对大尺度场景进行三维重建工作，例如地下车库，工厂车间等，建立该类环境下的三维结构地图。基于视频流输入的三维重建系统的研究，包括通过视觉和IMU多源信息融合处理，经过运动恢复结构流程，建立高精度的稀疏点云，利用多视角几何技术生成稠密点云，抽象语义结构元素，融入先验知识进行全局优化建图。  本课题中的关键研究问题如下：   1. 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，降低了操作的复杂度，但是构建出的模型效果往往精度不高，噪音较大，并且构图的成功率较低，因此如何提高构建模型的精度是本课题需要着重考虑和研究的第一个关键问题。 2. 以单纯视觉的方式进行构图无法确定初始化尺度问题，在构图追踪的过程中也难以提供一个较好的初始位姿，并且在运动模糊或者大角度转向时无法精确构图，因此考虑加入IMU信息进行多源融合来提高构图的准确性，以解决上述问题，这是本课题研究的第二个关键问题。 3. 在得到高精度的稀疏点云后，能够通过多视角几何技术生成稠密点云。在得到稠密点云后，由于缺少语义和结构化信息很难直接被人类理解，也很难将环境中的语义结构先验知识融入到优化流程中。因此如何从稠密点云中提取语义元素，并根据语义结构先验知识约束得到语义化结构地图，是本课题研究的第三个关键问题。 4. 课题来源、选题依据和背景情况   （1）课题来源  项目来源： 国家自然科学基金  项目名称： 复杂环境多鱼眼相机协同结构化vSLAM及评估体系研究  （2）选题依据和背景情况  人类通过双眼来探索与发现世界，在接收外部信息的方式中，有不到三成来自于听觉、触觉、嗅觉等感受器官，而超过七成、最丰富、最复杂的信息则通过视觉进行感知的。计算机视觉便是一种探索给计算机装备眼睛（摄像头）与大脑（算法）的技术，以使计算机能够自主独立的控制行为、解决问题，同时感知、理解、分析外部环境。20世纪60年代，计算机视觉得到了最初的发展。该阶段的研究重心主要体现在如何从二维图像中恢复出如立方体、圆柱体等立体化的三维形状，解释各个物体的空间位置关系。1982年David Marr从信息处理的角度对数学、神经生理学、计算机图形学等学科的研究成果进行了归纳总结，并在此基础上提出了一系列计算机视觉理论。得益于这个完整明确的理论体系，计算机视觉得到了蓬勃的发展。它的核心思想是从二维图像恢复三维结构。图1.1展示的是经典Marr视觉信息处理过程。    图1.1 视觉信息处理过程  近年来，图像的三维重建在计算机视觉中发挥了很大的作用，并且在质量和性能上有了较大的提升。其主要应用是自动地对于难以建模的对象建模，加快了图像运用的建模过程。这种技术需要处理大量的数据，可以使用于室内和室外的场景, 而不受控制的环境通常影响室外场景，如密集建筑群，或者复杂的原始森林等。对于这些场景，虚拟现实和计算机模拟可以被用来分析工作环境和工作难度等方面，三维图像重建技术本身被视为一个生成3D模型的技术。快速有效完整重建类似于雕塑三维物体目标的三维模型成为目前的研究方向。由连续图像的三维重建主要是指从二维图像序列中的获取物体的信息并进行三维重建。然而,这个领域并没有引起人们足够的重视，因此本文将对三维重建的具体原理以及改善展开讨论。   1. 课题的研究目标以及理论意义和实际应用价值   本课题的研究目标是能够实现视觉和IMU信息的多源融合，解决三维重建过程中精确度不高，场景无法闭合等问题，并以重建后的稀疏点云为基础，构建稠密点云，提取语义元素，再根据语义结构先验知识约束最终生成一个高精度，强鲁棒的语义地图。  本课题所提出的基于多源融合三维重建的语义结构地图构建方法对空间态势感知和理论研究具有重要意义，针对目前构图精度低误差大，且构图成功率低等问题，提出了一种新的基于多源信息融合的策略来改善现状。此外，关于由稠密点云到语义结构地图的构建，提出了通过提取语义元素，以及先验知识约束的方式来优化地图生成过程。  从二维图像重建三维立体具有重要的研究价值和潜在经济社会价值，其核心技术是通过运动来恢复结构，从最早的70年代摄影测量领域开始研究图像匹配，至今已有近半个世纪的时间。该问题的研究基础奠定于2000年之前摄影测量人的工作，而真正快速发展是主要得益于最近十几年计算机视觉领域的刺激和贡献。加之近年来数字（智慧）城市的发展要求，这类应用对于精度、效率、场景的语义化理解等都有要求，实景三维场景又比较复杂，因此还需要投入更多的人力物力与研究促进其不断发展。  时至今日，世界一流的三维重建系统已经可以全自动地恢复实景三维场景，也可以实现物体的高精度三维建模。三维重建系统在不同的应用领域有着不同的预设条件和技术要求，主要包括医学领域的重建系统，机器人导航相关实时重建系统，工业领域包括3D打印在内的室内高精度重建系统，以及摄影测量领域实景三维重建系统，此外在设备检测与监视、医学图像处理、文物保护、机器人视觉、自动导航、工业产品外观设计与生产等领域也越来越多的被引入。 |

2．文献综述（文献综述不得少于2000字）

|  |
| --- |
| 1. 国内外在该研究方向的研究现状及发展动态   本课题主要涉及到的理论方法有及时定位与地图构建（Simultaneous localization and mapping，简称SLAM）运动，运动恢复结构（Structure from Motion，简称SfM），线条语义提取，3D场景重建，多源信息融合等，其中3D场景重建和结构化语义地图是本课题的重点研究内容。下面将从以上几个方面对国内外研究现状和发展动态进行描述。  （1） SLAM的国内外研究现状  即时定位与地图构建是让机器人在未知环境中持续地构建环境地图，并同时在地图中给自己定位。最早的SLAM技术还不是使用视觉的方法，而是使用声波传感器或者激光以及惯性测量单元实现环境建模和自身定位。直到21世纪，Stephen Se等人首次使用图像的特征点实现视觉SLAM[1],之后由Davison使用EKF框架实现了最早的单目实时SLAM系统[2]，奠定了单目系统的基础；Davison在2007年成功实现基于单相机的纯视觉SLAM系统，算法的关键是在线建立2D点到3D点的映射关系，并且使用实时运动模型估计相机的位置[3]；Mur-Artal使用ORB特征点作为地图构建特征点，大幅度降低了点云的数量，并且使用回环检测的方法使定位与建图的精度都大幅提升[4]；随着硬件计算能力和数据储存的提升，提取目标深度信息的技术得到了很大的发展，戚传江等人使用2D slam的解决方案，采用多传感器数据融合的方法，完成多自由度位姿检测[5]，拓展了SLAM的应用场景；Whelan的实验通过使用体积融合的方法实现了实时大范围的稠密RGB-D的SLAM系统[6]，通过这个研究，使用SLAM系统作三维环境构建以及实时相机定位成为可能。  应用到目标跟踪领域，单纯点云集还是无法满足要求，因此需要将点云数据语义化，Reiger使用关系树的方法实现物体的语义识别[7]，这项技术对于目标跟踪是很重要的；之后Sarkar在Reiger的研究基础上结合FastSLAM的方法，使得识别速度更快，鲁棒性更强[8]；Zhang, G等人使用基于线条的SLAM算法[9]提高物体识别的准确率，该方法能够对物体的边沿与轮廓进行稳定的识别。  （2）三维重建的国内外研究现状  照相机/摄像机是将一个三维场景或物体投影到二维平面上，但是在降维的过程不可避免地会损失存在信息，而利用三维重建技术，就是从获取到的二维图像中复原原始三维场景或物体，三维重建的结果如图2.1所示。    图2.1 三维重建结果示意图  CMU大学的Tomasi和Kanade[10]等人首先开发出了基于图片的三维重建系统，并利用仿射分解法对摄像机进行了标定，得到摄像机运动参数，然后重构出物体的空间模型。随后，INRIAB Bougnoux等人[11] [12]利用未标定的SfM和摄像机自标定等算法，开发了一个提升型三维重建模型。Berkeley大学的Debevoc[13]等人完成了著名的建筑物重构系统Façade，该系统要求首先得到建筑物的粗略几何模型和摄像机运动参数。Shum等人开发的人机交互式重构系统，利用物体的一组全景图，即从各个角度得到物体的n张图片，然后对这些图像进行处理，重构出其三维实体，或者将场景表示成一系列按深度划分的分层的集合。Faugeras等采用摄像机的自标定方法，利用分层重构等经典的方法，从图像序列中重构出建筑物的形貌。Katholieke大学的Pollefeys等提出物体表面自动生成系统，该系统是在内参数可以改变的情况下，对摄像机采取了自标定的技术，该系统只要求用摄像机绕物体周围一周，拍到一系列的图像，就可以自动实现自标定和分层重构。  （3）运动恢复结构的国内外研究现状  运动恢复结构（SfM）能够利用两个场景或者多个场景自动恢复相机运动和场景结构，是一种能够自动完成相机追踪与运动匹配的自校准技术。在过去的一段时间内，由无序的图像进行SfM发生了巨大的进化。早期的自我校准度量重建系统是[14]SfM系统的基础，以此为启发，越来越多的大规模重建系统被开发出来，处理了数万计的互联网图片。随后，又提出了各种SfM策略，包括增量式 [15] [16]SfM,层次式[17]SfM和全局式[18] [19] [20]SfM，其中增量式SfM是最流行的重建无序照片集合策略。尽管SfM目前使用广泛，且现有系统具备较高的鲁棒性，但我们仍然没有设计出一个真正通用的SfM系统，虽然现有的系统已经相当先进，但鲁棒性、准确性、完整性等问题仍然是增量SfM的关键问题[22]，阻止了其作为通用方法的进一步使用。  （4） 线条语义提取的国内外研究现状  由点云到线条的提取是计算机图形学，模式识别等领域非常基础的算法，因此一直是非常重要的研究内容之一。目前已有很多基于多边形网格模型(Mesh)的轮廓线的提取算法[23,24,25,26,27,28,29,30,31]。三维网格模型上轮廓线提取主要是基于物体空间的算法，提取的轮廓比较精确，且应用更为广泛。针对点云模型，文献[32]中提出了一种基于图像空间点云模型的轮廓交互绘制算法，通过绘制点云模型上轮廓处的点画处物体轮廓线，得不到轮廓的空间结构信息。文献[33]中提出了点云模型上基于图像和物体空间的混合算法，首先逐点判断模型上哪些点是轮廓点，然后将轮廓点投影到平面，利用颜色缓存标记不同轮廓点，然后通过连接平面上的点得到三维模型上的轮廓。算法针对小规模的模型可以达到实时绘制，但针对大规模模型，逐点判断轮廓点的算法效率达不到实时要求，另外，由于是物体空间和图像空间的混合方法，算法的扩展性不好．比如针对一些提示性轮廓，由于投影到二维平面后轮廓线交叉比较多，提取轮廓线时带来很大局限。  （5）多源信息融合的国内外研究现状  多源信息融合（简称为信息融合）是指组合和合并多个来源的信息或数据以便形成一个统一结果的技术，是对人或动物利用各种感官来获取信息并通  过大脑综合分析来认识客观世界的一种功能模拟。信息融合要处理的数据／信息可能具有多种不同的特性，其中最大的挑战就是待融合的信息具有不完善性，主要表现为信息的不确定、不精确、不完整、模糊、多义、冲突等[34,36,37]。至今人们已提出了多种信息融合模型，如JDL模型、Omnibus 模型、模型、STDF模型等[35,38,39]其中应用最广泛的是模型及其演化版本。   1. 研究问题在本研究领域应用上的地位与价值   随着科学技术的日新月异，计算机视觉的应用日益受到各行业的关注和重视，如设备检测与监视、医学图像处理、文物保护、机器人视觉、自动导航、工业产品外观设计与生产等领域。计算机视觉技术为人们带来了机遇，也带来了挑战。三维重建作为计算机视觉技术中最为热门的研究方向之一，涉及到包括图像处理、立体视觉、模式识别等多个学科体系。利用计算机建立表达现实客观景物的三维模型，并以此来满足生产和生活的需要。随着工业化进程的不断发展，多种技术的实现均有赖于目标物体三维信息的获取。三维重建现已被广发的应用于生活和科研工作中，特别是在医学治疗、文物保护、游戏开发、工业设计、航天航海等方面，展现出了极强的生命力和影响力。 |

附：参考文献

|  |  |
| --- | --- |
| 序号 | 文献目录（作者、题目、刊物名、出版时间、页次） |
| **1** | Se S, Lowe D G, Little J J. Mobile Robot Localization and Mapping with Uncertainty using Scale-Invariant Visual Landmarks[J]. International Journal of Robotics Research, 2002, 21(8):735-760. |
| **2** | Davison A J. Real-Time Simultaneous Localisation and Mapping with a Single Camera[C]// IEEE International Conference on Computer Vision, 2003. Proceedings. IEEE, 2008:1403. |
| **3** | Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29(6):1052. |
| **4** | Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. IEEE Transactions on Robotics, 2015, 31(5):1147-1163. |
| **5** | 戚传江, 温程璐, 潘思宇. 基于多传感器数据融合的移动机器人六自由度姿态估计[J]. 2017. |
| **6** | Whelan T, Kaess M, Johannsson H, et al. Real-time large-scale dense RGB-D SLAM with volumetric fusion[J]. International Journal of Robotics Research, 2015, 34(4-5):598-626. |
| **7** | Reiger R, Sarkar A. SLAM using Relational Trees and Semantics Zhengzhou / China Dec. 8-10th, 2011[C]// Icmmm. 2011. |
| **8** | Sarkar A, Reiger R, Roy S, et al. SLAM Using Relational Trees and Semantics[J]. Advanced Materials Research, 2012, 452-453(452-453):648-653. |
| **9** | Zhang G, Lee J H, Lim J, et al. Building a 3-D Line-Based Map Using a Stereo SLAM[J]. IEEE Transactions on Robotics, 2015:1-14. |
| **10** | Tomasi C, Kanade T. Shape and motion from image streams under orthography: a factorization method[J]. International Journal of Computer Vision, 1992, 9(2):137-154. |
| **11** | Bougnoux S, Robert L. Totalcalib: a fast and reliable system for off-line calibration of image sequences[C]// 1997. |
| **12** | Debevec P E, Taylor C J, Malik J. Modeling and rendering architecture from photographs:a hybrid geometry- and image-based approach[C]// University of California at Berkeley, 1996:11-20. |
| **13** | Beardsley P A, Torr P H S, Zisserman A. 3D Model Acquisition from Extended Image Sequences[C]// European Conference on Computer Vision. Springer-Verlag, 1996:683-695. |
| **14** | Dellaert F, Seitz S M, Thorpe C E, et al. Structure from motion without correspondence[C]// Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. IEEE, 2000:557-564 vol.2. |
| **15** | Fitzgibbon A W, Zisserman A. Automatic camera recovery for closed or open image sequences[C]// European Conference on Computer Vision. Springer-Verlag, 1998:311-326. |
| **16** | Mohr R, Quan L, Veillon F. Relative 3D reconstruction using multiple uncalibrated images[M]. Sage Publications, Inc. 1995. |
| **17** | Pollefeys M, Van Gool L, Vergauwen M, et al. Visual Modeling with a Hand-Held Camera.[J]. International Journal of Computer Vision, 2004, 59(3):207-232. |
| **18** | Gherardi R, Farenzena M, Fusiello A. Improving the efficiency of hierarchical structure-and-motion[C]// Computer Vision and Pattern Recognition. IEEE, 2010:1594-1600. |
| **19** | Crandall D J, Owens A, Snavely N, et al. SfM with MRFs: Discrete-Continuous Optimization for Large-Scale Structure from Motion.[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(12):2841-2853. |
| **20** | Wilson K, Snavely N. Robust Global Translations with 1DSfM[C]// European Conference on Computer Vision. Springer, Cham, 2014:61-75. |
| **21** | Sweeney C, Sattler T, Hollerer T, et al. Optimizing the Viewing Graph for Structure-from-Motion[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2015:801-809. |
| **22** | Schönberger J L, Frahm J M. Structure-from-Motion Revisited[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016. |
| **23** | Benichou F, Elber G. Output Sensitive Extraction of Silhouettes from Polygonal Geometry[C]// Pacific Conference on Computer Graphics and Applications. IEEE Computer Society, 1999:60. |
| **24** | Dyken C, Reimers M. Real-time linear silhouette enhancement[J]. Mathematical Methods for Curves & Surfaces Troms, 2005, 152(2):135-143. |
| **25** | Gooch, Bruce, Sloan, et al. Interactive technical illustration[J]. 1999. |
| **26** | Hertzmann A, Zorin D. Illustrating smooth surfaces[C]// Conference on Computer Graphics and Interactive Techniques. ACM Press/Addison-Wesley Publishing Co. 2000:517-526. |
| **27** | Kirsanov D, Sander P V, Gortler S J. Simple silhouettes for complex surfaces[C]// Eurographics/acm SIGGRAPH Symposium on Geometry Processing. Eurographics Association, 2003:102-106. |
| **28** | Olson M, Zhang H. Silhouette Extraction in Hough Space[J]. Computer Graphics Forum, 2010, 25(3):273-282. |
| **29** | Saito T, Takahashi T. Comprehensible rendering of 3-D shapes[C]// Conference on Computer Graphics and Interactive Techniques. ACM, 1990:197-206. |
| **30** | Wang L, Tu C, Wang W, et al. Silhouette smoothing for real-time rendering of mesh surfaces[J]. IEEE Trans Vis Comput Graph, 2008, 14(3):640-652. |
| **31** | Wang L, Tu C, Wang W, et al. Silhouette smoothing for real-time rendering of mesh surfaces[J]. IEEE Trans Vis Comput Graph, 2008, 14(3):640-652. |
| **32** | Winkenbach G, Salesin D H. Computer-generated pen-and-ink illustration[C]// Conference on Computer Graphics and Interactive Techniques. ACM, 1994:91-100. |
| **33** | Xu H, Nguyen M X, Yuan X, et al. Interactive Silhouette Rendering for Point-Based Models[J]. Vis.pku.edu.cn, 2004. |
| **34** | 韩崇昭, 朱洪艳, 段战胜等. 多源信息融合[M]. 清华大学出版社, 2010. |
| **35** | 彭冬亮, 文成林, 薛安克. 多传感器多源信息融合理论及应用[M]. 西安电子科技大学出版, 2010. |
| **36** | Khaleghi, Bahador, Khamis, et al. Multisensor data fusion: A review of the state-of-the-art[J]. Information Fusion, 2013, 14(1):28-44. |
| **37** | 徐晓滨, 文成林, 刘荣利. 基于随机集理论的多源信息统一表示与建模方法[J]. 电子学报, 2008, 36(6):1174-1181. |
| **38** | Blasch E, Llinas J, Lambert D, et al. High Level Information Fusion developments, issues, and grand challenges: Fusion 2010 panel discussion[C]// Information Fusion. IEEE, 2010:1-8. |
| **39** | Blasch E P, Breton R, Valin P, et al. User information fusion decision making analysis with the C-OODA model[C]// Proceedings of the, International Conference on Information Fusion. IEEE, 2011:1-8. |

**3.**研究内容

|  |
| --- |
| 1. 明确研究对象、研究内容及工作范围   现阶段三维重建问题大多都是基于无序图像的增量式重建，这类方法使用简单，且方法成熟，对于一些特征丰富的简易场景能够呈现出很好的效果。但是对于一些较为特殊现实的场景，（例如停车场等室内场景各个视角的相似度较高，场景表面特征容易变化）时，在三维重建时就会出现场景匹配紊乱，场景无法形成闭环的问题。本课题的主要研究对象是基于多源融合三维重建的语义结构化建图系统，主要的研究内容为针对研究对象，建立出一个鲁棒性较高，误差率低的三维场景，包括：根据相机的运动恢复场景结构获得初步三维场景；再随后融合相机角速率和加速度信息（IMU信息）提升三维重建场景的准确性；对于完成的三维重建场景，根据其点云信息提取出线条语义信息，构建语义结构化地图。   1. 主要研究内容   （1）基于视频流的三维重建系统研究与开发  常规的针对三维场景的重建，一般采用无序图片作为输入，通过衡量各幅图片之间的相似度来判断输入的无序图片之间的相互关系，再通过特征点的匹配得到Scene Graph，随后根据SfM流程建立稀疏点云模型，如图3.1所示。但是现实停车场等室内场景中的非连续帧图片会存在相似度较高的问题：相距较远的部分视图较为相似而导致，以及部分连续帧图片之间特征匹配度较低的问题：场景在不同的视角下会得到不同的图像（例如建筑物表面反光，场景中存在动态物体的情况等），这些情况都会给图片之间的特征匹配造成极大的干扰，从而导致稀疏点云的建模达不到预期结果。    图3.1 运动恢复结构示意图  分析无序图片输入对三维重建整体过程的负面影响，揭示室内场景利用视频流进行三维重建易导致失败的原因，研究如何利用视觉运动追踪技术对视频流数据进行有序关联采样，并解决远距离图像误匹配问题，探索利用有序图像数据进行三维重建的全局优化方法，开发基于视频流的面向停车场等室内环境的三维重建系统是本课题的基本研究内容。  （2）视觉IMU多源融合高精度三维重建研究  一般情况下，三维重建系统仅利用被构建场景的图像信息或者视频流信息作为输入进行三维构图，但多数场景下会因为图像噪音及分解力问题导致最终的三维重建结果误差较明显。对于目前的单目三维重建系统，还存在初始化的尺度问题难以确定，以及在相机急转向的情况下出现的运动模糊和追踪失效问题。对于双目三维重建系统而言，也存在精度不高和鲁棒性较低的问题。这些问题的存在会对最终的三维重建效果带来很大影响。  揭示基于纯视觉数据进行三维重建效果低于预期的原因，在三维重建系统中研究基于图模型的视觉、惯导多源信息的融合机制，解决纯视觉三维重建系统的初始化不稳定、大角度转向追踪鲁棒性差、尺寸歧义等问题，提高三维重建系统的精度和鲁棒性是本课题的重点研究内容。  （3）语义结构化地图构建方法研究  点云地图分为稀疏、半稠密及稠密地图，分别由稀疏特征点、半稠密边缘轮廓特征及稠密像素点作为特征进行地图构建。点云地图多用于计算机视觉的定位应用，但是由于缺少语义和结构化信息很难直接被人类理解，也很难将环境中的语义结构先验知识融入到优化流程中。为提高三维重建地图的友好程度，便于定位及路径规划等应用设计，本研究内容将进行语义结构化地图构建方法研究。  本课题基于稀疏点云地图利用多视立体成像技术生成稠密/半稠密点云，融合图像和稠密/半稠密点云数据检测并抽象出语义结构元素，研究融入语义结构先验知识的全局优化建图方法，建立语义结构化构图三维重建系统是本项目的重点研究内容。 |

4．拟解决的关键技术

|  |
| --- |
| 明确工作中的关键技术难点，提出解决的方法   1. 有序视频流代替无序图片作为输入问题   原本三维重建的输入为无序图片，现在将场景的视频流输入到SLAM中，得到关键帧和各关键帧的相关关系，将这些数据替代原本无序图片作为三维重建的输入数据，是课题拟提出的关键技术之一。   1. 视觉信息与惯性测试单元信息多源融合问题   利用单目相机在运动的过程中，不断收集图像信息的同时加入每一帧图像对应的相机的惯性测量单元信息（包含物体三轴姿态角(或角速率)以及加速度信息），提出一种能够解决三维重建过程中的多源信息融合问题，是课题拟提出的关键技术之一。   1. 从稠密点云中提取语义元素构成结构化语义地图的问题   在得到场景的高精度稀疏点云后，利用多视角几何技术将生成稠密点云，由于稠密点云中的点为图像的像素点构成，本身不具备语义属性，因此需要通过点的结构属性提取出语义元素，再根据语义结构先验知识约束得到语义化结构地图，是课题拟提出的关键技术之一。 |

5. 工作方法

|  |
| --- |
| 1. 选择科学的工作方法，制订完整的技术路线   以视频流作为在线SLAM的输入数据，进行有序化采样，得到一个包含关键帧图片与关键帧间相互关系的Keyframe Database。除了视觉信息之外，再加入IMU信息进行多源信息融合，作为三维重建的输入，利用全局性SfM解决场景无法闭合的问题，最终得到高精度的稀疏云。随后，使用MVS多视角几何技术得到半稠密/稠密点云，从图像及3D点云数据中提取语义元素，再加上语义结构先验知识作为约束条件优化生成结构化语义地图。图5.1为本课题三维重建以及语义地图构建的技术路线图。    图 5.1 技术路线示意图   1. 工作方案的可行性分析，预设工作中可能遇到的难点，提出解决的方法 2. 研究方案   1.基于视频流的三维重建系统研究方案  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。  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。  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。    图 5.2 离线3D重建技术流程图  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。  2.基于图模型的视觉、惯导多源信息融合方案  在三维重建系统中单纯依靠视觉进行三维重建会出现初始化不稳定、大角度转向追踪鲁棒性差、尺寸歧义等问题。本方案提出引入IMU惯导信息进行多源信息融合以解决上述问题。图5.3是基于语义地标及特征点的factor graph图模型示意图。    图5.3 基于语义地标及特征点SLAM/3D重建的factor graph图模型  该图模型涵盖了语义地标及其矢量模型、传统3D地图特征点、单目相机系统以及车载惯导测量等信息。图中各个节点的下标表示按时间出现的顺序，随着相机的运动会有越来越多的变量节点及其观测约束节点被加入到图模型中。对于一个给定的图模型及相应的观测数据，其Maximum a Posteriori解是通过滚动优化X使后验概率最大。  3.检测并抽象语义结构元素，融入先验知识全局优化建图  通过上述过程可以得到高精度的稠密点云，现在需要提取语义结构线条，例如地下车库场景中的路标，墙体立柱等。现在采用Line 3D方法来提取场景中的线条信息，效果如图5.3所示。    图5.4 关于城市场景的稀疏点云模型、稠密点云模型以及3D线条模型  为了生成3D线条模型，首先需要在不同图像的二维线段之间建立对应关系，随后计算这些线段之间的对应关系。由于在匹配的过程中会建立大量的对应关系，需要提出一些相应的方法来减少误匹配，保留正确的匹配对。随后，给2D线段分配3D位置关系，如图5.5（a）所示。由于在同一位置存在大量的冗余线段，需要通过聚类方法消除冗余线段进行重构，最终效果如图5.5（b）所示。    图5.5 （a）3D线条模型（b）聚类重构后的模型  在抽象语义的过程中，融入先验知识提供约束条件，优化抽象结果，最终建立结构化地图。   1. 研究方案可行性分析   研究方案1）以视频流为输入的三维重建研究：  目前的以视频流作为输入进行三维重建工作已经相当成熟，其中存在的主要问题为：采集数据不适当导致无法进行三维重建以及在运动恢复结果过程中，输入的数据信息是无序的导致最终构图效果欠缺。本方案提出了关于如何采集数据来满足后续的重建工作，以及通过以SLAM的结果来代替原本的无序图片，其中SLAM结果包含关键帧信息以及关键帧之间的时序关系。目前已解决数据收集方式的问题，并且SLAM的结果输出与无序图像的输入相比较，仅扩充了图相间的时序关系，因此所提方案失败风险较小。  研究方案2）视觉IMU多源融合三维重建研究：  目前多源信息融合技术以及相当成熟，在数据融合的动态过程中，能够使信息量递增，并且使无序信息变得有序。在本方案的实际研究过程中，视觉信息和IMU信息本身并不具备较大的冗余性，并且IMU信息主要用于修正视觉信息带来的误差或者错误影响，因此所提方案失败风险较小。  研究方案3）结构化语义地图构成研究：  在收集到稠密点云或者半稠密点云后，先由点云提取结构化语义线条，再由语义线条构成语义地图。在这个过程中，关键在于得到高精度的点云结果，并去除噪音点和错误点，然后对优化后的点云结果提取线条，因此所提方案失败风险较小。 |

6.预期结果和结论

|  |
| --- |
| 对研究问题的解答进行科学预设，提出预期的工作成果  1）建立一个基于视频流和IMU多源融合的高精度三维重建系统，能够针对各种室内外的大尺度建筑物场景实现三维重建；  2）建立一个完善的语义结构化地图构建系统，对于停车库等大尺度场景能将误差控制在20cm以内。 |

7. 工作基础

|  |
| --- |
| 1. 科学评估研究条件和实验条件以及自己的专业基础和导师的专业把控能力   目前对三维重建已有大量的理论和实际研究，而且对运动恢复结构过程以及由稀疏点云转化为稠密点云过程已有大量实验验证，无论在理论还是在实际操作上都具有可行性，如图7.1所示，对室外某一假山场景构建的点云情况。    图7.1 假山场景构建的点云模型  本人现已掌握相机成像模型理论知识，以及三维空间位姿转换关系，对信息融合的方法也有一定的工程经验，利用SLAM对某一建筑物场景运行后的结果，如图7.2所示。    图7.2 SLAM运行建筑物场景的结果图  曾在导师的指导下，使用Normal Iterative Closest Point（NICP）方法构建点云模型，效果如图7.1所示， 对由视觉信息构建点云模型有深入的了解。同时指导老师在图像和SLAM领域有着丰富的教学和实践经验，曾负责多项相关研究项目，能给予我专业的指导。    图7.3 NICP生成点云模型示意图   1. 所需经费，经费来源，开支预算（工程设备、材料须填写名称、规格、数量）  * 经费来源   国家自然科学基金《复杂环境多鱼眼相机协同结构化vSLAM及评估体系研究》经费   * 开支预算（元）  1. 设备费：18000   相机设备：3000  存储设备：3000  电脑主机、显示器以及外部设备：12000   1. 论文发表版面费：3000 2. 资料费：1000 3. 其他费用：1000   总计：23000 |

8．工作计划（含实验、实践、设计、写作）

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **阶段及内容** | **工作量估计**  **（时数**） | | **起讫日期** | **阶段成果形式** |
| **（1）** | **理论研究，查阅相关文献资料** | **200** | | **2018-10至2018-12** | **深入理解运动恢复结构，信息融合原理等，并了解课题相关的理论知识。** |
| **（2）** | **运行运动恢复结构初步demo** | **400** | | **2018-12至2019-1** | **选择部分大尺度场景进行三维重建建模，统一图像采集方法与输入方式。** |
| **（3）** | **将SLAM结果代替无序图片作为输入** | **600** | | **2019-1至2019-4** | **将包含关键帧信息和其相关时序关系的SLAM结果作为输入代替无序图片** |
| **（4）** | **视觉+IMU信息融合过程实现** | **400** | | **2019-4至2019-6** | **在视觉信息的基础上，融合IMU信息，优化相机建模过程** |
| **（5）** | **提取语义线条构建结构化语义地图** | **400** | | **2019-6至2019-8** | **建立根据稀疏点云得到的稠密点云，并提取语义线条，构建高精度语义地图** |
| **（6）** | **结果分析** | **400** | | **2019-8至2019-10** | **分析系统的性能，并得出结论** |
| **（7）** | **论文撰写** | **600** | | **2019-10至2020-1** | **完成论文撰写工作** |
| **合计** | **3000** |  |

9.同济大学硕士研究生学位论文选题报告评分表

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 评审项目 | 权重 | 评分标准 | | 得分（百分制） |
| A.研究问题 | 30% | 80~100分 | 研究问题明确，具有前沿性，反映出明确的应用研究中的科学内涵，具有较强的应用价值。 |  |
| 60~80分 | 研究问题明确，反映出应用研究中的科学内涵，具有应用价值。 |
| 60分以下 | 研究问题不明，缺乏理论意义和实用价值。 |
| B.理论基础和专门知识 | 20% | 80~100分 | 较好地掌握并能综合运用科学理论、方法和技术手段解决研究问题。 |  |
| 60~80分 | 基本掌握理论基础和系统专业知识。 |
| 60分以下 | 缺乏理论基础和系统知识。 |
| C.工作成果 | 20% | 80~100分 | 成果科学、完整、扎实，体现出明确的学术思想。 |  |
| 60~80分 | 研究成果完整，体现出一定的学术思想。 |
| 60分以下 | 研究成果不完整，未体现学术观点。 |
| D.工作方法 | 20% | 80~100分 | 工作方法科学合理，条理清楚，针对性强。 |  |
| 60~80分 | 制定了明确的工作方案，有针对性。 |
| 60分以下 | 研究方法不明确。 |
| E.报告报告 | 10% | 80~100分 | 报告严密、逻辑性强、文笔流畅，表达清楚。 |  |
| 60~80分 | 基本概念清晰、层次分明，表达较清楚。 |
| 60分以下 | 写作、表达较差。 |
| 总分 |  | 总分=0.3A+0.2B+0.2C+0.2D+0.1E | |  |

备注：评审专家只对五项指标每一项的最后一栏内打分（百分制），不必计算总分。

评审小组组成：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 组成 | 姓名 | 职称 | 单位 | 签字 |
| 导师 | 岳继光 | 教授 | 同济大学 |  |
| 成员 | 肖辉 | 教授 | 同济大学 |  |
| 何斌 | 教授 | 同济大学 |  |
| 苏永清 | 副教授 | 同济大学 |  |
| 沈润杰 | 副研究员 | 同济大学 |  |
| 李雪峰 | 副教授 | 同济大学 |  |
| 董延超 | 副教授 | 同济大学 |  |

注：此评分表作为硕士研究生课程成绩单必备的材料之一

年 月 日

10.评审意见

|  |
| --- |
| 导师（或导师组）对本课题的评价  导师签名  年 月 日 |
| 评审小组的审查结论  组长签名 组员签名  年 月 日 年 月 日 |
| 学科专业委员会意见  负责人签名  年 月 日 |