广东工业大学硕士学位论文

（工学硕士）

基于视觉 SLAM 的移动机器人闭环检测研究

杨孟军

二○一八年六月

分类号：　　学校代号：11845

UDC：　　密级：　　学　　号：2111515027

广东工业大学硕士学位论文

（工学硕士）

基于视觉 SLAM 的移动机器人闭环检测研究

杨孟军

指导教师姓名、职称：　苏成悦　教授

学科(专业)或领域名称：电子科学与技术

学 生 所 属 学 院：　物理与光电工程学院

论 文 答 辩 日 期：　2018 年 6 月 2 日

A Dissertation Submitted to Guangdong University of Technology

for the Degree of Master

(Master of Engineering Science)

Research on loop closure detection of Mobile Robot based on Visual SLAM

M.E.Candidate:Mengjun Yang

Supervisor: Prof. Chengyue Su

June 2018

School of Physics & Opto-electronic Engineering

Guangdong University of Technology

Guangzhou, Guangdong, P. R. China, 510006

摘 要

随着移动机器人领域的发展，移动机器人的智能自主性得以提高，然而对未知环境进行探测建图并实时定位导航是移动机器人技术的研究难点。移动机器人在未知环境下根据自身位置估计和传感器数据，自主智能地实现自身定位和建立周围环境地图，这一过程称为同时定位与地图构建（Simultaneously Localization And Mapping，SLAM）。视觉传感器具有信息量丰富、轻量级、便宜等优点，将 SLAM 与视觉传感器相结合已成为机器人自主导航的研究热点。SLAM 的前端——视觉里程计（Visual Odometry,VO）可以实现移动机器人的定位，但其仅考虑相邻时间帧上的关联，移动机器人之前运动产生的误差将不可避免地累积到下一个时刻，使得整个 SLAM 会出现累积误差，长期估计的结果将不可靠，无法构建全局一致的轨迹和地图。如果能够成功检测出闭环，则可以显著地减小累积误差，并以此作为地图是否需要更新校正的依据，对于提高大规模 SLAM 的鲁棒性有重大意义。词袋模型是当前主流的闭环检测方法，然而此方法利用的是人为设计的特征，其中人类专业知识和见解在开发过程中占主导地位，这有着有很大的应用局限性，而且这些特征地提取耗费大量的时间。在光照变化明显的的场景中时，词袋模型的方法会忽略了环境中很多有用的信息，造成闭环检测的准确度不高。深度学习技术的目的是从可用于分类的原始数据中学习表示数据的方法，闭环检测本质上来说很像一个分类问题，因此深度学习技术为典型的闭环检测问题带来了新的方法。

本文提出将在图像分类与检索表现非常好的vgg16-places365卷积神经网络模型应用于视觉SLAM闭环检测中，将该模型作为图像特征提取器，在数据集New College上与别的闭环检测方法进行了对比验证，在闭环检测的速度和准确性上取得了很好的检测效果，为视觉SLAM闭环检测提出了一种新的方法，在移动机器人定位导航工程应用领域内，具有一定的创新性。本研究主要工作和成果如下：

1. 本文阐述了视觉SLAM框架中各个部分算法原理，包括传感器数据、前端视觉里程计、后端优化、闭环检测、建图。详细介绍了闭环检测的相关原理。

2. 本文描述了整个 SLAM 问题转化为数学模型的理论推导过程，SLAM 问题由状态估计转换为最小二乘法，然后通过迭代的方法求解最小二乘问题，这样整个 SLAM 问题得到求解。

3. 本文对特征点检测算法进行了研究，分析对比SIFT、SURF和ORB特征描述子方法。详细介绍了词袋模型的整个实现过程，包括字典的形成，相似度的计算等，并对此进行了实验验证。为满足SLAM实时性强的系统特性，实验采用ORB特征描述子来提取特征，这样增强了图像特征匹配的旋转不变性与实时性。

4. 本文提出将 vgg16-places365 卷积神经网络模型应用在移动机器人的视觉 SLAM 闭环检测上，可视化了每层提取到的特征，使之更加形象化。详细介绍了 vgg16-places365 卷积神经网络模型的框架，介绍了模型的训练参数的设置，给出了实验用的闭环检测方法，用余弦相似度来计算模型提取到的特征向量之间的相似性。此外，本文将基于 vgg16-places365 卷积神经网络模型的闭环检测方法与几种传统基于人工设计特征（BoVW、GIST 等）的闭环检测方法以及其他几种基于深度学习模型的闭环检测方法在数据集 New College 上进行了对比实验，给出了实验结果及分析，实验结果表明基于 vgg16-places365 卷积神经网络模型在闭环检测上的 PR 性能和特征提取时间性能上具有比较好的优势。

关键词：SLAM；特征提取；闭环检测；词袋模型；卷积神经网络；相似度

Abstract

With the development of the field of mobile robots, the intelligent autonomy of mobile robots can be improved. It is a difficult research point for the mobile robot technology to detect the unknown environment and create real-time navigation. The mobile robot autonomously realizes its own positioning and establishes an environmental map based on its own position estimation and sensor data in an unknown environment. This process is called Simultaneously Localization And Mapping (SLAM). Vision sensors have the advantages of rich information, lightweight, and low cost. Combining SLAM with vision sensors has become a research hotspot for robot autonomous navigation. Visual Odometry can achieve the positioning of mobile robots, but it only considers the correlations on adjacent time frames, and the error caused by the motion of the mobile robot will inevitably accumulate to the next moment., so that the cumulative error would occur in the entire SLAM. The results of long-term estimates would be unreliable, or we cannot build globally consistent trajectories and maps. If the loop closure detection is successful, the cumulative error can be significantly reduced and used as a basis for whether or not the map needs to be updated. It is of great significance to improve the robustness of the large-scale SLAM. The Bag-of-Word model is the current mainstream loop closure detection method. However, this method uses the characteristics of human design, in which human expertise and insights dominate the development process. This has great application limitations, and it takes a lot of time to extract characteristics. In the scene where the light changes significantly, the Bag-of-Word model method ignores many useful information in the environment, resulting in a low accuracy of loop closure detection. The purpose of deep learning technology is to learn representational data from raw data that can be used for classification. Loop closure detection is essentially a classification problem, which brings a new approach to the typical loop closure detection problem.

This paper applies the latest vgg16-places365 convolutional neural network model applied to image classification and retrieval for the first time to visual slam loop closure detection. The paper uses this model as an image feature extractor, compares vgg16-places365 convolutional neural network model with other loop closure detection methods in the New College dataset. The result shows that the vgg model has achieved a good effect on the speed and accuracy of loop closure detection, so a new method has been

explored for visual slam loop closure detection. It has certain innovation in the engineering field of mobile robot localization and navigation.The main work and results of this study are as follows:

1. This paper describes the principle of each part of the visual SLAM framework, including sensor data, front-end visual odometry, back-end optimization, loop closure detection, and construction. The paper focused on the related principles of loop closure detection in detail.

2. The mathematical model of the whole SLAM module is deduced in this paper and the state estimation is converted to the least squares method, and then the iterative method is used to solve the least squares problem, so that the whole SLAM problem is solved.

3. In this paper, the feature point detection algorithm is studied, and the SIFT, SURF and ORB feature descriptor methods are analyzed and compared. The whole process of Bag-of-Word model is introduced in detail including the formation of the dictionary, calculation of similarity. At last we do the experiment to verify. In order to satisfy the strong real-time system characteristics of SLAM, the experiment uses ORB feature descriptors to extract features, which enhances the Rotation invariance and real-time performance of image feature matching;

4. This paper first applies the vgg16-places365 convolutional neural network model to visual slam loop closure detection. This paper also visualizes the features extracted from each layer to make it more visualized. The framework of the network model of vgg16-places365 convolutional neural network is introduced in detail, then the training parameters of the model are introduced and then the loop closure detection method for experiment is given using the cosine similarity to calculate the two eigenvectors' similarity. Finally the paper gives the results and analysis of the experiment. This experiment is tested on the New College data set. Besides comparing with a number of traditional methods based on artificial design features such as BoVW, GIST, etc., the paper also compares vgg16-places365 convolutional neural network model with the other several methods of deep learning model, and it was found that PR performance and feature extraction time performance are superior.

Keywords: SLAM; Feature extraction; Loop closure detection; BoW; Convolutional neural network; Similarity

目 录

CONTENTS

第一章 绪 论

1.1 选题背景及意义

移动机器人是灵活性较高和自主能力较强的自动化综合体，可具有感知、规划、协同等与人相似的能力，涉及了传感技术、机械电子、自动化控制、计算机科学、人工智能等各个领域学科的研究。随着机器人领域的新一代技术的飞速发展，移动机器人能够在复杂危险的确定性环境下，代替或帮助人类完成一些难以进行的工作任务，如灾难救援、处理爆炸物、医疗护理等。此外，移动机器人也应用于服务业、娱乐业等便民生活方面。因此，移动机器人的各项技术已经得到全世界的研究人员的广泛关注。对移动机器人的探索研究开始于20世纪60年代末期，1966年至1972 年Nils Nilssen和Charles Rosen 等人研发出了自主移动机器人[1]。同一时期，美国和前苏联研制研发了移动机器人，将其投射到月球进行探测。美国“探测者”在地面站的控制下完成了在月球上挖掘等一些简单操作。苏联的“登月者”在无人驾驶的状态下降落在月球表面收集土壤和岩石样品。1997年美国NASA研制了第一台用于在火星上从事科学考察工作的自主式移动机器人，该移动机器人使用了视觉传感器和激光传感器。21世纪初美国iRobot机器人公司推出的扫地机器人Roomba[2]，具有自主避障和自主回充等功能，自此移动机器人开始逐渐步入人们的日常生活，智能机器人技术在世界范围内得到了大力发展。近二十几年来，我国加强智能机器人技术的研究力度，成功研发出一系列工业机器人和特种机器人[3]。一些高校科研基地也取得了丰硕的研究成果，清华大学自主研发了THMR-V智能车，该移动机器人能在路面上自主行驶、自主避让和巡航[4]；上海交通大学的FRONTER-1自主移动机器人装有多种视觉传感器，操作性简单，反应速度快[5]；哈尔滨工业大学研制的服务机器人装备了机械臂和激光等多种传感器，可受远程控制，是一款真正物联网意义上的机器人[6]。移动机器人也逐步走向产业化，涌现一批小米、科沃斯等品牌的全自动扫地机器人，还有用于餐饮、商场等场所的智能服务机器人。移动机器人的自主导航技术是机器人领域的研究热点之一，Leonard和Durrant-Whyte首次提出要实现移动机器人在陌生环境下自主导航[7,8]，需要解决三个因素：“机器人在哪儿？”、“机器人的周围环境是怎样的？”以及“机器人怎样到达目标位置？”。前两个因素表示移动机器人在环境中实时定位和创建地图，后一个因素表示移

动机器人的避障和路径规划的功能。当知道机器人周围环境的地图时，有许多有效的办法能对机器人进行比较准确的定位，同理当知道机器人的位置也有很多办法能对机器人周围环境进行建图。如机器人处于室外环境时，可以用GPS等来进行定位解决机器人自主导航需要解决的前两个因素。但是当机器人处于室内或矿井下等没有GPS的环境时，用GPS的方法就无法解决移动机器人的定位和建图问题，在这些环境中就要用到本文所讲的SLAM。移动机器人通过自身一系列不同类型的传感器感知未知环境的信息，实现对环境进行构建地图并能够同时自我定位与导航，这一过程叫做同时定位与建图(Simultaneous Localization And Mapping)，简称SLAM。

SLAM整个问题非常复杂，其包含的定位与建图互相之间相辅相成是一个“鸡生蛋、蛋生鸡”的问题[9]，自八十年代末，SLAM被提出后，经过许多学者几十年的研究，解决SLAM问题的方法得到巨大进展，各种SLAM相关的算法以及解决方案被提出，对于SLAM技术的研究主要经历基于滤波器方法和基于图优化方法这两个阶段。滤波器的方法是指若机器人的观测数据是已知的，用以前时刻的状态来估计机器人的当前姿态。基于滤波器方法有卡尔曼滤波、扩展卡尔曼滤波（EKF）、粒子滤波PF等方法[10-12]，尽管许多研究学者们对基于滤波器方法进行了很多的优化和改进，但由于基于滤波器的SLAM方法的本身的线性化问题，在长时间运动和大规模构建地图时存在更新效率低，易造成累积误差。20世纪末Lu和Milios提出了基于图优化的SLAM的概念[13]。机器人在环境中运动时，其位姿误差会不断增加，因此建图的质量会下降，为了克服这种不断增长的累积误差，机器人必须感知以前观测过的特征，识别这些特征来校正更新地图。该过程涉及机器人在一个循环中运动以及将新的观测与已建地图关联的问题，因此这个过程在SLAM中称为闭环(Closing The Loop)。但是，在机器人位姿不确定性较大的情况下，往往比较难以将当前的路标与之前观测过的路标关联起来，通过传感器识别以前观测过的地方或者路标，并进行正确关联，即形成闭环，非常具有挑战性。而如果检测到回环，将其信息提供给后端进行处理，能显著的减少机器人移动过程中的累积误差从而提高后端优化处理的准确性，因此闭环检测对于整个SLAM系统精度与鲁棒性的提升具有重要意义[14]。鲁棒的SLAM，不仅具有重要的理论意义，也将加快移动机器人在人类日常生活的应用，具有广泛的实用价值和应用前景。激光传感器是SLAM中应用最广的传感器，大规模环境下基于激光的SLAM通常采用算法本身产生的位姿估计，存在累积误差的问题，从而可能导致闭环检测失败，而视觉特征所包含的信息丰富，在多视点配准上更适于闭环检测，因此本文主要研究视觉SLAM中的闭环

检测。

1.2 视觉 SLAM 概述及研究现状

由于视觉传感器具有成本低、质量轻、图像信息量大等优点，视觉 SLAM 是目前机器人导航方面的研究热点之一。移动机器人通过视觉传感器观测周围环境目标位置，从未知起点开始采用创建环境地图，同时利用生成的环境地图进行实时定位。根据视觉传感器类型不同，视 SLAM 分为基于单目视觉方法、基于双目视觉方法和基于 RGBD 深度视觉方法。一个完整稳定有效的视觉 SLAM 过程包括视觉传感器获取环境信息，前端视觉里程计，后端数据优化，闭环检测以及建图这些环节。解决 SLAM 问题实际上就是对状态估计进行解决处理，按照解法过程的区别有滤波器的方法与图优化的方法。

1.2.1 基于滤波器的视觉 SLAM 研究现状

基于滤波的 SLAM 是 SLAM 研究历史上最早解决 SLAM 问题的方法，20 世纪 90 年代初 Smith 等人最早提出基于卡尔曼滤波求解 SLAM 问题[15]。21 世纪初 Chiuso 等人[16]利用单目视觉信息，运用扩展卡尔曼滤波[17]以及增量式地图的方法解决 SLAM 问题。卡梅隆大学研究者使用粒子滤波器实现了一种 Fast SLAM[18,19]，避免了卡尔曼滤波中的计算复杂度和数据关联错误敏感度。Grisettiyz 等人[20]采用一种自适应的粒子预测分布方法，减少粒子数目，进而降低维护地图时的存储空间。Sibley 等人[21]提出一种滑动窗滤波器 SLAM，即在滑动窗口中不仅保留当前位姿，还包括之前一段连续的位姿，以提高滤波器算法的精确度。国内有学者使用平方根容积卡尔曼滤波计算 SLAM 后验概率密度，以减小线性化误差，提高了机器人 SLAM 定位精度[22]。

1.2.2 基于图优化的视觉 SLAM 研究现状

用滤波器的 SLAM 方法是通过不断地迭代机器人曾经看到信息来实现机器人位姿状态估计以及对周围环境地图的建立，由于其线性化点不一致，无法多重线性化，导致信息矩阵秩增加，会严重地造成累积误差。而用图优化的 SLAM 方法则是通过利用机器人过去看到过的信息来估计机器人的位姿状态，建立好机器人周围环境地图。滤波器的方法计算复杂度比用图优化的方法的计算复杂度要高的多，如当观测到的路标数目为 N 时，滤波器方法的运算复杂度为 $O(N^3)$，图优化的方法运算复杂度为 $O(N)$ [23]。

最早 1997 年 Lu 和 Milios 提出通过图优化来解决 SLAM 问题[24]，但当时研究者们认为此方案的计算非常复杂因而没被广泛应用。Thrun 等人基于前人的研究基础上提出一种 Graph SLAM 方法[25]，利用稀疏化降低计算量，通过位姿间的约束绘制地图，进而简化为一个最小二乘估计问题。牛津大学 Klein 等人[26]首次提出一种基于关键帧光束平差法 BA 的单目视觉 SLAM 系统，名为 Parallel Tracking and Mapping（PTAM）。Forster 等人于 2014 年提出的一种基于稀疏直接法的视觉里程计——Semi-Direct Monocular Visual Odometry（SVO）[27]。Mur-Artal 等人[28]基于 PTAM 的算法框架提出一种 ORB-SLAM 系统，该系统利用 ORB 描述子进行特征匹配和重定位，通过方位图优化闭环回路，ORB-SLAM 是现代 SLAM 系统中做的非常完善，非常易用的系统之一。LSD-SLAM 是 J.Engle 等人于 2014 年提出的 SLAM 工作[29,30]，它将直接法应用到了半稠密的单目 SLAM 中。

1.3 闭环检测概述及研究现状

机器人运动过程中，在获取图像数据、特征提取与匹配和运动估计等几个方面都会存在误差，即使用 Bundle Adjustment 进行局部的或者全局的优化，仍然会存在累积误差。机器人运动的时间越长，这种误差影响越明显。判断机器人当前位置是否是其曾到达过的地方，即成功地检测闭环可以减少前端视觉里程计产生的累积误差，通过闭环检测对所有姿态结果进行优化，这是消除累积误差最有效的方法，闭环检测是一个比 Bundle adjustment 更加强烈、更加精准的约束。

1.3.1 基于人工设计特征的闭环检测研究现状

检测闭环的关键点在于决定观察到的图像与图像之间的相似性，许多方法都是通过将两个时刻下采集到的图像进行配准，因此闭环检测问题从另外一种角度来看是指图像的配准问题。对于大多数人为设计特征进行图像配准的算法，闭环检测方法采用的是视觉词袋模型(BoVW)[31]，词袋模型对图像中视觉特征描述子进行聚类，建立词典，然后对于给定的图像在词袋中找到对应的单词。常见的视觉特征 SIFT、SURF、ORB 等[32-34]被广泛应用，并取得了显著的效果，如在 FAB-MAP[35]中引入了 BoVW，由于 SIFT 和 SURF 等本地图像特征在构建 BoVW 描述子时具有不变性，FAB-MAP 实现了优异的性能，成为闭环检测研究的标准基线算法之一。Cummins 等人[36]使用 FAST 算子提取图像的局部特征，然后应用到增量式的闭环检测中，提高了闭环检测的效率。Liu Y[37]

等人使用图像的 GIST 描述子提取图像的全局特征，在最近的视觉 SLAM 研究中受到欢迎。这些技术都是使用人为设计的特征，为了实现所需的特征，其中人类专业知识和见解在开发过程中占主导地位。这有着有很大的应用局限性，这些特征的提取耗费大量的时间，同时在光照变化明显的的场景中时，这些方法忽略了环境中有用的信息，造成闭环检测的准确度不高。

1.3.2 基于深度学习的闭环检测研究现状

深度学习技术的目的是从可用于分类的原始数据中学习表示数据的方法，闭环检测本质上来说很像一个分类问题，这为典型的闭环检测问题带来了新的方法。最近几年，研究者们尝试将深度模型应用到闭环检测中，使用深度学习来解决这个问题。Xiang Gao 等人[38,39]使用 Autoencoder 网络模型来提取图像特征，并使用相似度矩阵方法检测提取到特征的相似度来检验闭环，在公开的数据集上取得了比较好的效果。卷积神经网络（Convolutional neural network，CNN）的最新进展[40]激励研究 CNN 作为对现有图像描述子弱点的潜在解决方案。在许多研究中，CNN 从视觉数据特征提取抽象层次的能力已经超过基于人为设计特征解决方案的性能[41,42]。特别是 CNN 在图像分类和图像检索任务[43]方面的卓越成就是非常令人鼓舞的。考虑到视觉闭环检测类似于图像分类和图像检索，因此基于 CNN 的特征的能力用于设计视觉闭环检测问题的解决方案是合理的。HE Yuanlie 等人[44]使用 FLCNN（Fast and lightweight convolutional neural network）提取图像特征，并计算相似度矩阵，进一步提高闭环检测的实时性和准确性。YifanXia 等人[45]提出利用级联深度学习模型 PCANet（principal Component Analysis Net）来提取图像特征然后应用到闭环检测中，实验结果比人工设计的特征闭环检测的效果要好。Yi Hou 等人[46]利用 PlaceCNN 提取图像特征进行闭环检测，在不同光照，以及相同光照下在数据集上的闭环检测效果要远优于其余几种人工设计特征的方法。

1.4 本文主要内容

本文主要开展了对基于视觉 SLAM 的闭环检测算法的研究，首先用现在较为广泛的闭环检测方法——基于人工设计特征的词袋模型来检测闭环的实验，然后在此基础上再进一步探索用现在发展较为迅速的深度学习方法来做闭环检测，提出了将现在场景分类表现优秀的 vgg16-places365 卷积神经网络模型应用闭环检测，取得了很好的实验效果。本论文的章节如下所示：

第一章为绪论，介绍了本文研究的背景与意义，对视觉 SLAM 及闭环检测研究现状进行阐述，介绍了本文研究工作的主要内容和论文结构。

第二章为视觉 SLAM 以及闭环检测原理介绍，详细介绍了视觉 SLAM 的框架，描述了视觉 SLAM 各个模块的原理以及实现的功能，其中重点介绍了闭环检测的相关原理，作为后续章节工作的基础。

第三章为基于词袋模型的闭环检测，首先介绍了 SIFT、SURF 和 ORB 等人工特征的原理，然后介绍词袋模型的原理以及实现流程，最后采用基于 ORB 特征的词袋模型来检测闭环的实验。

第四章为基于卷积神经网络的闭环检测，提出了将场景分类表现优秀的 vgg16-places365 卷积神经网络模型应用于闭环检测，叙述了该模型的各层次在深度学习框架 caffe 下的架构，以及模型的训练参数，然后用从上述模型提取的特征来计算相似度方法来判断是否产生闭环，通过在专门用来做闭环检测的标准数据集 New College 上进行实验测试，实验对比了基于 vgg16-places365 卷积神经网络模型和其余模型在闭环检测上的性能以及特征提取时间，并对之进行了分析。

最后为总结与展望。总结了本文所完成的一些研究工作，之后还探讨了本文所没研究到到的一些问题，并提出了该实验在今后一些改进办法。

第二章 视觉 SLAM 以及闭环检测原理介绍

2.1 视觉 SLAM 原理介绍

移动机器人通过自身一系列不同类型的传感器感知未知环境的信息，实现对环境进行构建地图并能够同时自我定位与导航，这一过程叫做同时定位与建图（SLAM），其框架如图 2-1 所示。

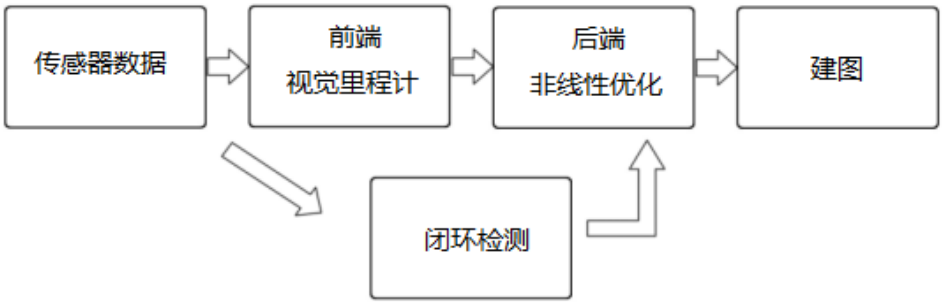

图 2-1 整体视觉 SLAM 流程图

Fig.2-1 Overall vision SLAM flow chart

2.1.1 传感器

本文探究的是视觉，其传感器数据指的是通过相机来获取的，视觉传感器目前主要有三大类：单目相机、双目相机、RGB-D 深度相机。从字面意思来看，单目即是一个摄像头，双目则是有两个摄像头，而 RGB-D 相机包换一个普通 RGB 摄像头能获得彩色图片，此外其还包含能测量深度的摄像头。本文会在后面详细介绍它们的工作原理

图 2-2 单目相机

Fig.2-2 Monocular camera

图 2-3 双目相机

Fig.2-3 Binocular camera

图 2-4 深度相机

Fig.2-4 Depth camera

基于单目相机的 SLAM，即只用一支摄像头就能够完成 SLAM。单目相机如图 2-2 所示。这种传感器很简单、便宜，日常生活中也经常可见，像手机等智能设备上都配备，所以目前对单目 SLAM 研究非常热。单目相机利用的是针孔相机模型。相比别的双目传感器和 RGB-D 传感器，单目有个最大的问题，就是不能确切地知道深度。由于不能知道绝对深度，在仅仅一幅图像中，是不能够确定物体的真实大小的。如图 2-5 所示。仅仅这一幅图像里，不容易判断出这些手掌上的小人是离得较远的真实人的人还是假的模型。要想估计图像中的物体距离相机的距离以及其真实的大小，需要依靠相机的运动。通过三角测量法，利用相机的运动来求解相机运动并估计像素的空间位置。

图 2-5 手掌上的是真人还是模型？

Fig.2-5 Is it human or model on the palm of your hand?

针孔相机模型描述了单目相机的成像模型，而双目相机通过多个相机之间的基线，估计空间点的位置。双目相机如图 2-3 所示。然而，仅根据一个像素，是无法确定这个空间点的具体位置的。这是因为，从相机光心到归一化平面连线上的所有点，都可以投影至该像素上。只有当空间点的深度确定时（比如通过双目或 RGBD 相机），才能确切地知道它的空间位置。双目相机的距离估计是比较左右眼的图像获得的，通过计算同时采集到的图像之间的视差（左右图的横坐标之差），来估计每一个像素的深度，并不依赖其他传感设备，所以它既可以应用在室内，亦可应用于室外。当想计算每个像素的深度时，其计算量与精度都将成为问题，而且只有在图像纹理变化丰富的地方才能计算视差。但双目相机标定很复杂，视差的运算量非常大，既很耗计算机内存资源又耗时间。一般需要 FPGA 设备对其进行提速才能提高速度满足要求。

相比于双目相机通过视差计算深度的方式，RGB-D 相机的做法更为“主动”一些，它能够测量每个像素的深度。目前市场上主流的 RGB-D 传感器相机有微软的 Kinect 和华硕的 Xtion PRO。在媒体娱乐方面，RGB-D 传感器使用 3D 人体运动捕获算法实现

动作捕捉和手势识别等功能，可以让人们无须手持控制器，直接通过手势操纵游戏。在三维重建方面，RGB-D 传感器可以同时获取室内环境的彩色图像信息和深度图像信息，相比立体相机和 TOF 相机[47]，具有轻便、便宜、信息完整等优点。RGB-D 传感器实物图如图 2-4 所示，传感器主要由一颗 PS1800 芯片、一个 RGB 摄像头、一个 IR 摄像头和一个 IR 投射器组成。PS1800 芯片是一个多感应系统级的芯片，具有极高的并行运算能力，主动投射红外光源从而对图像进行编码，能够同步获取深度图像和彩色图像。RGB 摄像头用于捕捉彩色图像信息，IR 摄像头和 IR 投射器组成了三维结构光深度传感器，用于捕捉深度图像信息。不过现在大多深度相机测量范围较小、易受太阳光影响等这方面的原因，在 SLAM 方面，主要用于室内环境，而对于室外环境则较难应用。

2.1.2 视觉里程计

视觉里程计又称为视觉前端，视觉 SLAM 前端是对获取的 RGB 图像进行提取特征点以及计算描述算子，利用特征描述算子对相邻两帧图像进行特征匹配，进而通过获取匹配特征点在对应的深度图像中的深度信息，得到相邻两帧匹配特征点对应的三维点坐标，根据匹配特征点与三维点的对应关系估计相机 6D 运动变换，并对局部相机姿态进行优化，得到粗略的相机运动估计，为后端提供比较好的初始值。视觉里程计按是否需要提取特征分为基于特征法的视觉里程计和基于直接法的视觉里程计。基于特征视觉 SLAM 前端可分为特征提取与描述子计算、特征匹配、运动估计以及局部姿态优化这几个步骤。这是现在视觉里程计最主要的做法。视觉 SLAM 前端流程图如图 2-6 所示。

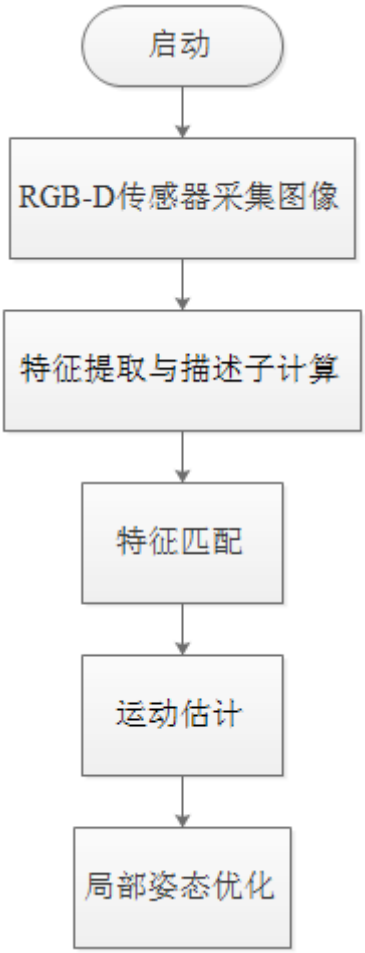

图 2-6 视觉 SLAM 基于特征方法前端流程图

Fig.2-6 The flow diagram of visual SLAM front-end based on feature

目前视觉 SLAM 中特征提取算法应用最为广泛有 SIFT、SURF 和 ORB。SIFT 算法提取的特征点数量多，误差小，但运算速度慢。SURF 算法提取的特征点数量较少，但是运算速度比 SIFT 算法快，而 ORB 是目前运算速度最快的一种特征提取算法。具体更详细介绍将在第三章中展示。

特征匹配是视觉 SLAM 中极为重要的一步，它解决了 SLAM 中的数据关联问题，即确定当前看到的路标与之前看到的路标之间的对应关系。在实际应用中，一般通过描述子与描述子之间的距离来判断其是否相似。目前惯用的度量描述子与描述子距离的方法有四种：欧氏距离、马氏距离、汉明距离、海宁格距离。对于图像配准来说最为简单的配准方式是暴力匹配。暴力匹配指的是对从图像中提取到的每一个关键点及其描述子与其余所有特征的描述子之间的上述四种之一的距离进行测量，然后按测量的值的大小进行排列，取其中最近的距离对应的特征点当做配准点。暴力匹配法虽然简单，但是当要匹配的特征点很多时，它的计算量非常大，非常耗时间与资源，这不

符合 SLAM 中实时性要求，这种情况下可用快速近似最近邻（FLANN）算法来配准。如图 2-7 所示为基于 ORB 的 FLANN 算法特征匹配图像。

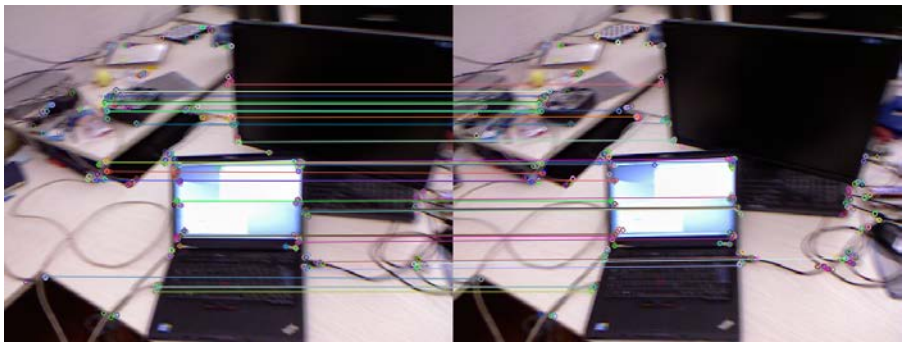

图 2-7 ORB 特征提取与匹配

Fig.2-7 ORB Feature Extraction and Matching

通过关键点的提取以及其对应的描述子的计算就可以得到其特征点，然后用不同图像的特征进行配准就能确定图像与图像之间相同特征之间的单一映射关系。之后根据这些配准好的的特征点就可以求出相机的运动，即是运动估计。对于单目相机来说，有图像知道的是 2D 点像素坐标，根据两组 2D 点可以用对极几何来解决。当相机为双目或者深度相机时，可以通过相机知道了空间距离信息，用两组 3D 点可以根据迭代最近点（Iterative Closest Point，ICP）方法来估计运动[48]。当知道 3D 点的坐标和它们在相机平面上的投影位置时，用 PnP（Perspective-n-Point）方法估计相机的运动[49]。因为这些运动估计方法不是本文讨论的重点，这里不过度展开。

对于直接法的视觉里程计主要是根据图像的像素信息来计算相机的运动。基于特征点的方法是通过最小化重投影误差来优化相机运动，而基于直接法中，最小化的不是重投影误差而是最小化光度误差。直接法基于灰度不变假设，直接根据像素亮度信息来估计相机运动，其可以完全不用特征点地提取，节省了特征的计算时间，也避免了特征缺失的情况。基于直接法的视觉里程计可以构建半稠密甚至稠密的地图，而基于特征点的方法不能做到。虽然基于直接法的视觉里程计有这些优势，但是其也有着很明显的缺点，最小化光度误差函数是非凸的，使得优化算法容易进入极小，只在运动很小时直接法才表现比较好。再有就是，灰度值不变是一个很强的假设，如果相机拍摄时自动曝光，会使图像整体变亮或变暗，这会破坏灰度不变假设，使得算法失败。

2.1.3 SLAM 后端

前端视觉里程计能给出一个短时间内的轨迹和地图，但由于不可避免的误差累积，

这个地图在长时间内是不准确的。所以，在视觉里程计的基础上，需要构建一个尺度、规模更大的优化问题，以考虑长时间内的最优轨迹和地图，这就是本节所讲的视觉 SLAM 后端优化。视觉 SLAM 可以归成两大类：基于滤波器的视觉 SLAM 和基于图优化的视觉 SLAM。基于滤波器的视觉 SLAM 无法重线性化和秩增加，而且要求系统计算能够快速完成线性化和更高的更新效率，因而其很难被用于大型环境的地图建立。基于图优化的方法是考虑当前时刻状态与之前所有状态的关系，此时将得到非线性优化为主题的优化框架。图优化是一种非线性化与图论相结合的理论。而图优化的 SLAM 利用位姿图来构建相机的运动，利用从图像中提取到的特征转变为与相机位姿的约束关系，通过解 BA 非线性优化来估计相机的所有位姿。做 BA 优化时会考虑全局的信息，而不是像滤波 SLAM 只考虑当前位姿与上一个时刻的位姿。图是由节点和边构成，在基于图优化的 SLAM 中，优化图中的节点(node)或顶点(vertex)表示相机在不同时刻的位姿，边(edge)可以表示不同时刻位姿之间的约束，基于图优化的 SLAM 后端的目标是调整优化图中机器人位姿节点所处的位置，使其尽量满足边所表示的约束关系，获得一张最优位姿图。在以图优化框架的视觉 SLAM 算法里，Bundle Adjustment (BA)[50] 起到了核心作用。在早期，包含大量特征点和相机位姿的 BA 计算量过大，不适合实时计算。近年来由于有研究者发现 SLAM 的稀疏性质，利用图优化方法求解视觉 SLAM 问题逐渐成熟，图优化也已经成为求解 SLAM 问题所公认最好的方法。

2.1.4 SLAM 框架之闭环检测

闭环检测（Loop Closure Detection），是指机器人能够识别出它以前到达过的地方。本文主要针对视觉 SLAM 的闭环检测模块做研究，本节将更加详细介绍闭环检测的相关原理为后续章节作铺垫。前端提供特征点的提取和轨迹、地图的初值，这些值传给后端后，由后端对其进行优化处理。因为视觉里程计仅考虑相邻时刻帧间的关系，那么前面产生的误差将会不断的转移到后一个时刻，这样整个算法将会出现累积误差，长期估计的结果将不准确，不能更准确地建立全局一致的轨迹和地图。如果检测成功，可以显著地减小累积误差，从而校正出现偏差的轨迹与地图，如图 2-8 所示，这有利于实现大规模健壮的 SLAM。

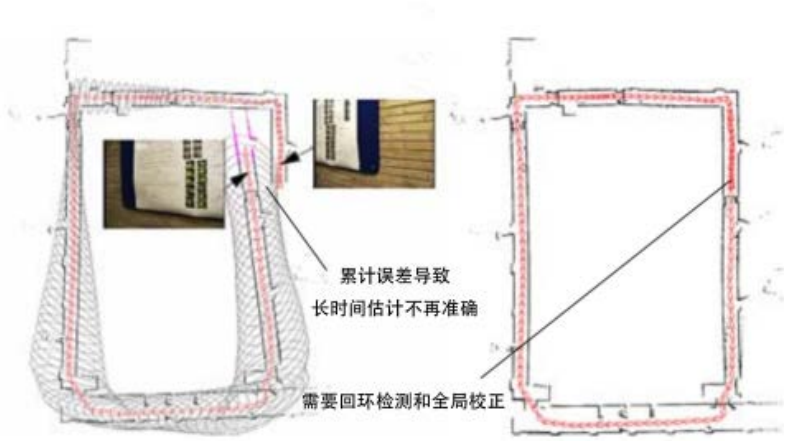

图 2-8 闭环检测的校正效果示意图

Fig.2-8 Schematic diagram of the correction effect of loop closure detection

机器人运动过程中，在获取图像数据、特征提取与匹配和运动估计等几个方面都会存在误差。即使使用 Bundle adjustment 进行局部的或者全局的优化，仍然会存在累积误差，机器人运动的时间越长，这种误差影响越明显。通过闭环检测对所有姿态结果进行优化是消除累积误差最有效的方法，闭环检测是一个比 Bundle adjustment 更加强烈、更加精准的约束。对于闭环检测方法有很多，主要分为以下四种：

1. 暴力检测的方法，将新获取的关键帧与之前所有的关键帧一一比较，整体运算效率相当低，严重影响了视觉 SLAM 的实时性。

2. 基于随机检测的方法，从之前所有关键帧中随机选取 N 帧，选出其中与当前帧相似的帧，但当构建场景地图面积较大，导致闭环检测运算效率很低。

3. 基于 KD-Tree 的方法，先是将特征构建成 KD-Tree，这样可以用来与当前帧进行检索对比，此方法是目前主流的检测方法之一。

4. 基于 BoVW 的方法，BoVW（视觉词袋）模型最早出现在神经语言程序学(NLP)和信息检索(IR)领域，近年来被广泛应用于计算机视觉中。基于 BoVW 模型的闭环检测的运算效率高，且适用于大规模场景的地图创建。

现在对视觉 SLAM 闭环检测大部分做法是采用视觉词袋模型（Bag-of-view-Words，BoVW），视觉词袋模型闭环检测的做法流程如下：首先从大量图像中提取到许多视觉特征，然后对这些视觉特征进行聚类，形成一个“词典”，对于要检测的图像用词典为其描述其含有哪些单词，这样就形成了一个描述向量，通过检测代表图像的描述向量之间的相似度来判断是否产生了闭环。因此实际上这是比较图像之间的相似性。相似度高的则认为是机器人在同一地方所看到的场景，从而认为其产生了闭环。鉴于此，

有研究者因此将闭环检测看做是一个图像分类问题，使用模式识别的方法来做闭环检测，尤其是近几年，有一些学者将深度学习应用在闭环检测上，为闭环检测探索出了许多新的方法。

从人类的角度看，能够以很高的精确度感觉到“两张图像是否相似”或“这两张照片是从同一个地方拍摄的”这件事实。因此闭环检测算法被希望也能够得出如此的结果。当事实上两张图像是从同一个地方拍摄，那么闭环检测算法也应该给出“这是闭环”的结果。反之，事实上两张图像是从不同地方拍摄的，那么程序也应该给出“这不是闭环”的判断。然而程序检测的结果可能不如所期待的那样，可能出现表 2-1 中的四种情况。

表 2-1 闭环检测的结果分类

Table 2-1 The result classification of loop closure detection

算法/事实	是闭环	不是闭环
是闭环	真阳性（True Positive）	假阳性（False Positive）
不是闭环	假阴性（False Negative）	真阴性（True Negative）

其中真阳性与真阴性很好理解，指的是检测出的结果与事实上是符合的。假阳性是指算法检测出来是闭环，而事实上所检测的两幅图片不是同一个地方拍摄的。如图 2-9 所示。假阴性是指事实上所检测的两幅图片是同一个地方拍摄的，而算法检测结果却不是闭环。如图 2-10 所示。

图 2-9 假阳性场景

Fig.2-9 False Positive scene

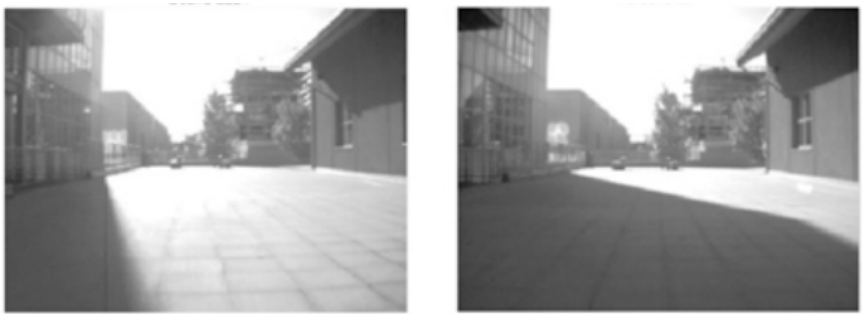

图 2-10 假阴性场景

Fig.2-10 False Negative scene

优秀的闭环检测算法要最大可能地能够识别出假阳、假阴这两种情况。学者们常一般用 precision-recall 曲线来鉴定一个闭环检测算法的优劣，测试它的 P 和 R 值，然后做出一条 precision-recall 曲线，其中 P 值为准确率（precision），R 值为召回率（recall），其计算公式如下：

$$precision = \frac{TP}{TP + FP} \tag{2.1}$$

$$recall = \frac{TP}{TP + FN} \tag{2.2}$$

其中 TP 代表正确检测出的闭环数目，FP 称假阳性代表真实上不是闭环但是检测的结果是闭环的数目，FN 又称假阴性代表没有检测到的真实的闭环数目；准确率描述的是算法所提取的所有闭环中，确实是真实闭环的概率。召回率则描述的是在所有真实的闭环中，被正确检测出来的概率。

2.1.4 SLAM 框架之建图

SLAM 包含定位与建图两部分。在经典 SLAM 中的地图指的是路标点的集合，当知道了所有的路标点在哪里时，就可以完成了建图。前面所介绍的视觉里程计和后端 BA 优化事实上都是建模了路标点的位置，并对它们进行了优化。地图可以分为稀疏地图与稠密的地图。稀疏地图指主要由路标组成的地图。对于移动机器人的定位，稀疏的地图就够用了。而对于稠密地图来说，其是要构建所有看到东西。对于移动机器人的导航，就要用到稠密的地图，比如二维栅格地图，如图 2-11 所示。当然出于视觉效果提出的要求而产生的 3D 点云模型也是地图的一种，如图 2-12 所示。

图 2-11　2D 栅格地图

Fig. 2-11　2D grid map

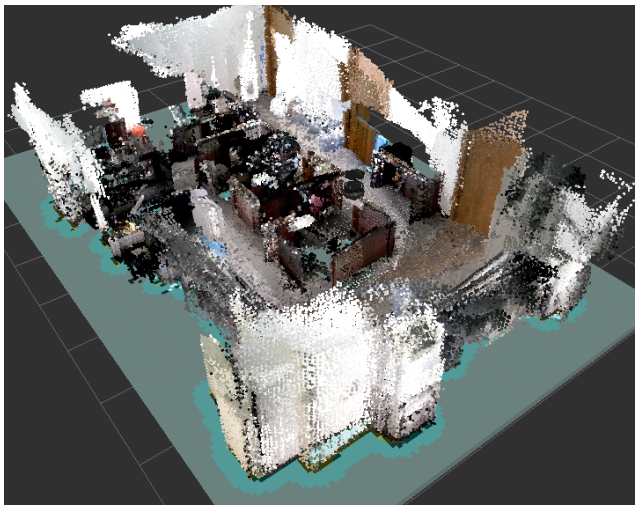

图 2-12　3D 点云地图

Fig. 2-12　3D point cloud map

2.2　SLAM 问题的数学表述

由于相机通常是在某些时刻采集数据的，所以只关心这些时刻的位置和地图。这就把一段连续时间的运动变成了离散时刻 $t=1,...,k$ 当中发生的事情。在这些时刻，用 x 表示机器人自身的位置。于是各时刻的位置就记为 $x_1,...,x_k$，它们构成了机器人的轨迹。而对于地图，设地图是由许多个路标（Landmark）组成的，而每个时刻，传感器会测量到一部分路标点，得到它们的观测数据。不妨设路标点一共有 N 个，用 $y_1,...,y_N$ 表示它们。现在考虑从 $k-1$ 时刻到 k 时刻，机器人的位置 x 是如何变化的，机器人会安装传

感器测量有关自身运动的参数，测量出来的但不一定直接是位置信息，还可能是加速度、角速度等信息。然而，无论是什么信息，都能使用一个通用的、抽象的数学模型如（2.3）式所示：

$$x_k = f\left(x_{k-1}, u_k, w_k\right) \tag{2.3}$$

这里u_k是运动传感器的读数（有时也叫输入），w_k为噪声，f为一般函数，整个函数可以指代任意的运动传感器，成为一个通用的方程，此方程称为运动方程。机器人在x_k位置上看到某个路标点y_j，产生了一个观测数据$z_{k,j}$。同样用一个抽象的函数h来描述这个关系，如（2.4）式所示：

$$z_{k,j} = h\left(y_j, x_k, v_{k,j}\right) \tag{2.4}$$

这里$v_{k,j}$是这次观测里的噪声。

因此 SLAM 过程可总结为两个基本方程如（2.5）式所示：

$$\begin{cases} x_k = f\left(x_{k-1}, u_k, w_k\right) \\ z_{k,j} = h\left(y_j, x_k, v_{kj}\right) \end{cases} \tag{2.5}$$

由于运动方程在视觉 SLAM 中没有特殊性，因此主要讨论观测方程。假设在x_k处对路标y_j进行了一次观测，对应到图像上的像素位置$z_{k,j}$，观测方程可以表示成（2.6）式所示：

$$sz_{k,j} = K\exp\left(\xi^{\wedge}\right)y_j \tag{2.6}$$

这里K为相机内参,s为像素点的距离。同时这里的$z_{k,j}$和y_j都必须以齐次坐标来描述，且中间有一次齐次到非齐次的转换。对机器人位姿状态的估计，等同于在已知传感器数据u和观察到的数据z的条件下，计算状态x的条件概率分布$P\left(x \mid z,u\right)$。现在仅仅考虑只有图片这样的观测数据z，没有u，等同于估计$P\left(x \mid z\right)$的条件概率分布。由贝叶斯法则可得（2.7）式所示方程：

$$P\left(x \mid z\right) = \frac{P(z \mid x)P(x)}{P(z)} \propto P(z \mid x)P(x) \tag{2.7}$$

$P\left(x \mid z\right)$是后验概率，$P(z \mid x)P(x)$是似然部分，$P(x)$是先验概率。由于直接求后验分布有难度，转而求一个后验概率最大化（Maximize a Posterior，MAP）时的状态最优估计，得（2.8）式所示：

$$x^*_{MAP} = \arg\max P(x|z) = \arg\max P(z|x)P(x) \tag{2.8}$$

当机器人位姿未知时，此时先验就无从知道。这样就可以求解 x 的最大似然估计，如（2.9）式所示：

$$x^*_{MLE} = \arg\max P(x|z) \tag{2.9}$$

2.2.1 最小二乘的引出

假如在高斯分布下，那么最大似然有比较简化的形式。又有对于观测模型来说，对于某一次观测有如（2.10）式所示：

$$z_{k,j} = h(y_j, x_k) + v_{k,j} \tag{2.10}$$

假设了噪声项 $v_k \sim N(0, Q_{k,j})$，所以观测数据的条件概率为（2.11）式所示：

$$P(z_{j,k} | x_k, y_j) = N(h(y_j, x_k), Q_{k,j}) \tag{2.11}$$

上式依然是一个高斯分布，使用最小化负对数的方式，来求一个高斯分布的最大似然，得如下（2.12）式所示：

$$-\ln(P(x)) = \frac{1}{2}\ln\left((2\pi)^N \det(\Sigma)\right) + \frac{1}{2}(x-u)^T \Sigma^{-1}(x-u) \tag{2.12}$$

在最小化上式的 x 时，第一项与 x 无关，可以略去。于是，只要最小化右侧的二次型项，就得到了对状态的最大似然估计。代入 SLAM 的观测模型，可以得如下（2.13）式：

$$x^* = \arg\min\left((z_{k,j} - h(x_k, y_j))^T Q_{k,j}^{-1}(z_{k,j} - h(x_k, y_j))\right) \tag{2.13}$$

即是最小化误差的平方。因此，对于机器人所有时刻的运动和任何时刻的观测，其所得数据与估计值之间的误差为（2.14）和（2.15）所示：

$$e_{v,k} = x_k - f(x_{k-1}, u_k) \tag{2.14}$$

$$e_{y,j,k} = z_{k,j} - h(x_k, y_j) \tag{2.15}$$

并求该误差的平方之和（2.16）所示：

$$J(x) = \sum_k e_{v,k}^T R_k^{-1} e_{v,k} + \sum_k \sum_j e_{y,k,j}^T Q_{k,j}^{-1} e_{y,k,j} \tag{2.16}$$

这就得到了一个总体意义下的最小二乘问题（Least Square Problem）。

2.2.2 最小二乘的解法

最小二乘问题是一个非线性的问题，对于这类问题，通常用迭代的方式来求解，从一个初始值出发，不断地更新当前的优化变量，使目标函数下降。具体步骤可列写

如下：

1. 给定某个初始值 x_0。
2. 对于第 k 次迭代，寻找一个增量 Δx_k，使得 $\left\| f\left(x_k + \Delta x_k\right) \right\|_2^2$ 达到极小值。
3. 若 Δx_k 足够小，则停止。
4. 否则，令 $x_{k+1} = x_k + \Delta x_k$，返回 2.

让求解导函数为零的问题，变成了一个不断寻找梯度并下降的过程。一阶和二阶梯度法和 Gauss-Newton 法和 Levenberg-Marquadt 法是视觉 SLAM 的优化问题上也被广泛采用的方法，大多数优化库都可以使用它们确定增量 Δx_k。

2.3 本章小结

本章对视觉 SLAM 框架中各个部分算法的原理作了大概介绍，其中传感器数据、视觉里程计、后端优化、建图只是粗略介绍，因这些不是本文的重点，并未展开。本文主要是针对闭环检测部分进行研究，详细地介绍了闭环检测原理，为后文的内容做好铺垫。此外，本章提供了整个 SLAM 问题的数学表述的推导过程，SLAM 问题由一个状态估计问题转化为一个最小化二乘问题，这为整个 SLAM 提供了数学理论上的支持。

第三章 基于人工设计特征的闭环检测

通过全局场景识别、判断机器人当前位置是否处于已访问过的场景这就是所谓的闭环检测，对于实现闭环检测总体上来说分为两中情况：基于几何关系的闭环检测与基于外观(Appearance based)的闭环检测[51]。基于几何关系是指当发现当前相机运动到了之前的某个位置附近时，检测其是不是闭环，但是由于视觉里程计运动估计时不断累积的误差，往往没法正确的发现“运动到了之前的某个位置附近”这件事实，所以闭环检测也就变得不可能了[52]。另一种方法是基于外观的，它和前端视觉里程计和后端的估计都无关，仅仅根据两幅图像的相似性确定闭环检测关系。这种做法摆脱了累积误差，使得闭环检测模块成为 SLAM 系统中一个相对独立的模块。自 21 世纪初被提出以来，基于外观的闭环检测方式能够有效的在不同场景下工作，成为了视觉 SLAM 中主流的做法，并被应用到实际的系统中去[53-55]。在基于外观的闭环检测算法中，核心问题是如何计算图像间的相似性。例如对于图像 A 和图像 B，要设计一种方法，计算它们之间的相似性评分：s(A,B)。这个评分会在某个区间内取值，当它大于一定的阈值就认为出现了一个闭环。从直观上看图像能够表示成矩阵，最直接做法就是让两个图像相减，然后取某种范数，但是这样做不够好，不能很好的反应图像间的相似性，因为像素灰度是一种不稳定的测量值，它严重受环境光照和相机曝光的影响。另外，当相机视角发生少量变化时，即使每个物体的光度不变，它们的像素也会在图像中发生位移，造成一个很大的差异值。因此需要一种更加可靠的方式，学者们提出了像视觉里程计那样用人为设计的特征点来做闭环检测，对两个图像的特征点进行匹配，只要匹配数量大于一定值，就认为出现了闭环。

3.1 人为设计的特征提取方法

目前视觉 SLAM 中特征提取算法应用最为广泛有 SIFT、SURF 和 ORB。SIFT 算法提取的特征点数量多，误差小，但运算速度慢。SURF 算法提取的特征点数量较少，但是运算速度比 SIFT 算法快。而 ORB 是目前运算速度最快的一种特征提取算法。本节将介绍这三种特征提取算法，因本章后面实验将用到 ORB 算法，因此将对 ORB 算法进行详细介绍，其余两种只是粗略介绍。

（1）SIFT 算法

SIFT 全称为 Scale Invariant Feature Transform，尺度不变特征变换。SIFT 算法是由 David Lowe[56]于 20 世纪末提出的一种特征提取算法。SIFT 算法检测与描述图像中的局部特征，能在空间尺度中找到极值点，进而提取出位置不变、尺度不变、旋转不变的特征点。SIFT 可以获取丰富信息量，但运算复杂，计算量大，比较耗时，实时性比较差。

（2）SURF 算法

SURF 算法（加速鲁棒特征算法）是由 Herbert Bay 等人[57]于 2008 年提出基于 SIFT 算法改进的一种算法，其运行效率比 SIFT 算法更快。

（3）ORB 算法

ORB(Oriented FAST and Rotated BRIEF)算法是由 Ethan 等人[58]于 2011 年基于结合 BRIEF(Binary Robust Independent Elementary Features)特征描述算法[59]和 FAST(Features from Accelerated Segment Test)角点提取算法[60]的优点提出了一种性能优越的二进制特征提取算法。BRIEF 算法的优点在于提取特征点速度极快，但该算法不具备旋转不变性，对噪声敏感以及不具备尺度不变性，因此 ORB 算法引入 FAST 算法思想来弥补 BRIEF 算法的不足。

1. 特征点提取。FAST 算法通过检测局部图像中是否存在角点，无需计算二阶导数，也就省略了去噪声的步骤，从而提高了运算效率。在图 3-1 中，要检测 p 是否特征点就要先用围绕 p 的 16 个点的灰度值和 p 的值相减，用相减得到值与事先设置好的阈值进行对比，当大于阈值的差值个数足量时则可以同意点 p 是特征点，如式(3.1)所示。

$$N=\sum_{1\le i\le 16}|I(x_i)-I(p)|>t \tag{3.1}$$

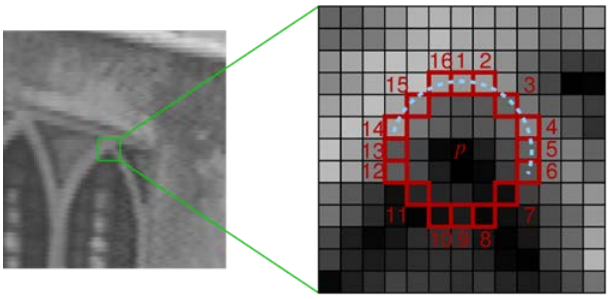

图 3-1 角点检测

Fig.3-1 Corner detection

2. 特征点描述子。ORB 算法选择了 BRIEF 特征描述，其优点在于特征描述算子结构简单，能够更快地完成匹配，但其缺点是不具备旋转不变性。因此 ORB 算法使用了

主方向来引导 BRIEF，其矩阵 S：

$$S=\begin{pmatrix} x_1 & x_2 & \cdots & x_{2n} \\ y_1 & y_2 & \cdots & y_{2n} \end{pmatrix} \tag{3.2}$$

ORB 算法在特征点提取时已经确定了特征点的方向角 θ。利用 θ 和其对应的旋转矩阵 R_θ 来矫正 S，为 S_θ。

$$S_\theta = R_\theta S \tag{3.3}$$

$$\text{其中，}\ R_\theta=\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{3.4}$$

即可得旋转不变的特征描述子：

$$g_n(I,\theta) := f_{nd}(I)\,|\,(x_i,y_i)\in S_\theta \tag{3.5}$$

综合上述关于 SIFT 算法、SURF 算法以及 ORB 算法的描述，ORB 算法提取特征点以及计算描述子的运算速度最快。因此，为了实现实时的视觉 SLAM，本文在特征点提取和特征描述子计算方面采用 ORB 算法。

3.2 词袋模型

Bag-of-words model 最早出现在自然语言处理和信息检索领域。这个模型是许多个独立出现词汇的集合，它不考虑文本的语法和语序这些元素。BoW 是使用一组无序的单词(words)来表达一段文字或一个文档。最近几年以来，BoW 模型被大规模地应用到机器视觉中。

词袋的目的是用“图像上有哪几种特征”来描述一个图像，如图 3-2 所示。根据这样的描述，可以度量这两个图像的相似性。在目前流行的 SLAM 过程中，词袋模型是闭环检测的主流做法，包括号称鲁棒与效果兼备的 ORB_SLAM。使用词袋模型分为几个步骤：确定单词的概念，许多单词放在一起组成了字典，确定一幅图像中出现了哪些在字典中定义的概念，用单词出现的情况描述整幅图像。这就把一个图像转换成了一个向量的描述，该向量描述的是“图像是否含有某类特征”的信息，比单纯的灰度值更加稳定。又因为描述向量代表的是“是否出现”，而不管它们在哪儿出现，所以与物体的空间位置和排列顺序无关，因此在相机发生少量运动时，只要物体仍在视野中出现，就认为描述向量不发生变化。最后比较上一步中的向量描述的相似程度。

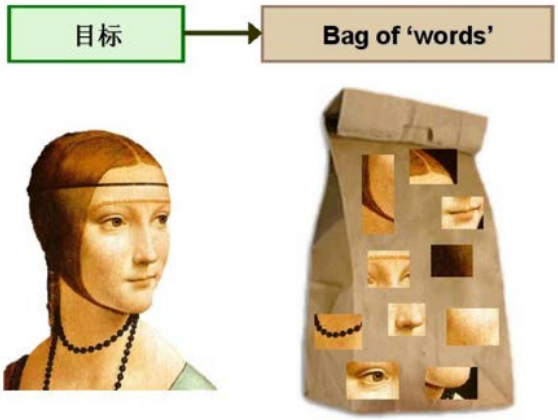

图 3-2 词袋模型

Fig.3-2 Bag-of-words model

3.2.1 字典

字典由很多单词组成，而每一个单词代表了一个概念。一个单词与一个单独的特征点不同，它不是从单个图像上提取出来的，而是某一类特征的组合。所以，字典生成问题类似于一个聚类（Clustering）问题。聚类问题是无监督机器学习中一个特别常见的问题，用于让机器自行寻找数据中的规律的问题。词袋模型的字典生成问题亦属于其中之一。对大量的图像提取了特征点，用经典的 K-means（K 均值）算法[61]找一个有 k 个单词的字典，每个单词可以看作局部相邻特征点的集合。

3.2.2 K-means 算法

K-means 算法是一种硬聚类算法，是一个简易明了的方法，在机器学习的无监督学习中被广泛使用。下面对它的原理做简要介绍，假设当有 N 个数据，要将其归为 k 个类别，K-means 的流程如下：

1. 随机选取 k 个中心点： $c_1,...,c_k$ ；
2. 对每一个样本，计算与每个中心点之间的距离，取最小的作为它的归类；
3. 重新计算每个类的中心点。
4. 如果每个中心点都变化很小，则算法收敛，退出；否则返回 1。

本章后续实验采用一种 k 叉树来表达字典如图 3-3 所示。它的思路很简单，类似于层次聚类，是对 k-means 的深化。假定有 N 个特征点，希望构建一个深度为 d ，每次分叉为 k 的树，训练字典时，逐层使用 K-means 聚类。根据已知特征查找单词时，亦可逐层比对，找到对应的单词。其步骤如下：

1. 在根节点，用 k-means 把所有样本聚成 k 类 。这样得到了第一层。
2. 对第一层的每个节点，把属于该节点的样本再用 k-means 再聚成 k 类，得到下一层。
3. 依此往下推，最后得到叶子层。叶子层即为所谓的 Words。

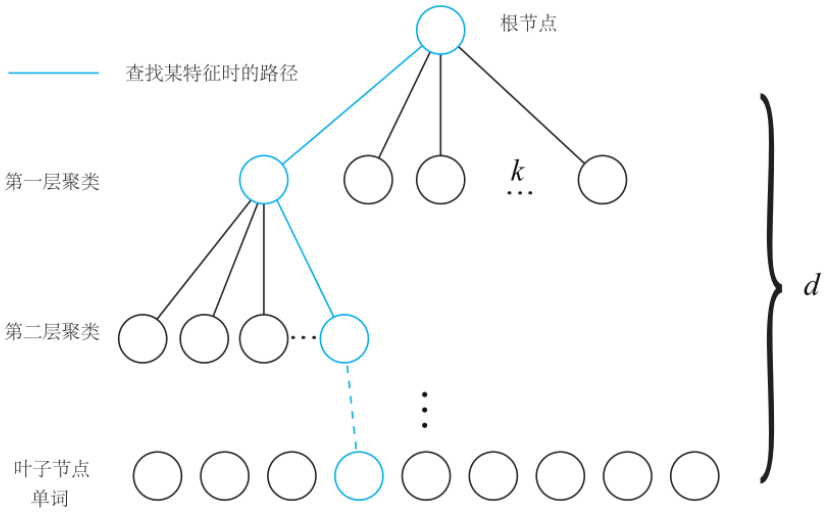

图 3-3 K 叉树字典示意图

Fig.3-3 K-fork tree dictionary schematic

实际上，最终在叶子层构造了 word，在快速查找时使用，用到了树的中间节点。这样一个 k 分支，深度为 d 的树，可以容纳 k^d 个单词。另一方面，在查找某个给定特征对应的单词时，用此方法只需将它与每层聚类中心比较，总共只需要比较 d 次，就能够快速找到最后的 word，这样查找效率就能大大提升。

K-means 方法可以把已经提取的大量图像特征聚类成一个包括 k 个单词的词典，根据图像中特征点，查找其字典中相对应的单词。用单词出现的情况描述整张图像，将一个图像转换成了一个向量的描述，因此图像间的相似性转化为描述向量的相似性了。计算描述向量的方式有很多种，L_1 范数形式是常用的一种，后续的实验会详细介绍。

3.3 基于词袋模型的闭环检测实验

3.3.1 实验环境

测试的硬件环境主要包括 AMD A8-4500M@1.9GHz 8 核处理器，8G 内存。测试的系统环境为 Ubuntu14.04，内核版本为 Linux 3.16.0-43-generic。装有 opencv3.4 开源库用来对图像进行处理。还装有开源库 DBoW3，DBoW3 是一个开源的 C++库，它能实

现图像特征排序。本实验用其将图像转变成视觉词袋，然后建立词典。

3.3.2 标准测试数据集

为了测试验证的方便性和准确性，本次实验数据来自于 TUM 数据集，该数据集是由慕尼黑工业大学(TUM)计算机视觉小组提供的 RGB-D 数据集 ，RGB-D 数据集包括了由 RGB-D 相机获取的 RGB 图像与深度图像，以及由 8 个高速追踪相机所构成的高精度运动捕捉系统所记录 RGB-D 相机的真实位姿信息，即 Ground Truth 数据。本文选取十张 RGB 图像来测试基于人工设计特征 ORB 的词袋模型算法，它们来自一次实际的相机运动轨迹，实验图像如图 3-3 所示，可以看出第一张图像与最后一张图像明显采自同一个地方，本次测试要做的就是程序能否检测到这件事情。

图 3-3 演示实验中使用的十个图像，采集自不同时刻的轨迹

Figure 3-3 The ten images used in the experiment, collected from different moments

3.3.3 相似度计算

本节讲述通过词袋模型来计算任意图像之间的相似度。有了字典之后，给定任意特征 f_i，只要在字典树中逐层查找，最后都能找到与之对应的单词 w_j。假设从一张图像中提取了 N 个特征，找到这 N 个特征对应的单词之后，相当于拥有了该图像在单词列表中的分布。考虑到，不同的单词在区分性上的重要性并不相同，因此希望对单词的区分性或重要性加以评估，给它们不同的权值以起来更好的效果。在文本检索中，常用的一种做法称为 TF-IDF（Term Frequency–Inverse Document Frequency）[62,63]，或译频率-逆文档频率。TF 指的是图像中出现频率越高的部分区分度越高，IDF 指的是在词典中出现的频率越低区分度越高。当建立词典时考虑 IDF，该 word 的 IDF 为：

$$\mathrm{IDF}_i = \log\frac{n}{n_i} \tag{3.6}$$

上式中 n 为所有特征数量，n_i 为叶子节点中特征点的数量。TF 是指某个特征在单个图像中出现的频率。那么 TF 为：

$$\mathrm{TF}_i = \frac{n_i}{n} \tag{3.7}$$

上式中 n 为单个图像中单词出现的总次数，n_i 为单词 w_i 出现的次数。所以它的权重等于 TF 乘 IDF 之积：

$$\eta_i = \mathrm{TF}_i \times \mathrm{IDF}_i \tag{3.8}$$

考虑权重以后，对于图像 A，它的特征点可对应到许多个单词，组成它的 Bag-of-Words：

$$\mathrm{A} = \{(w_1,\eta_1),(w_2,\eta_2),...,(w_N,\eta_N)\} = v_\mathrm{A} \tag{3.9}$$

通过词袋，可以用单个向量 v_A 描述了一个图像 A。这个向量 v_A 是一个稀疏的向量，它的非零部分指示出图像 A 中含有哪些单词，而这些部分的值为 TF-IDF 的值。给定两幅图片 A 和 B，可以得到描述向量 v_A 和 v_B，然后通过 L_1 范数计算这两幅图像之间的相似性，如公式（3.10）所示；

$$s(v_\mathrm{A} - v_\mathrm{B}) = 2\sum_{i=1}^{N}|v_{\mathrm{A}i}| + |v_{\mathrm{B}i}| - |v_{\mathrm{A}i} - v_{\mathrm{B}i}| \tag{3.10}$$

3.3.4 实验过程以及结果

首先对十张目标图像提取 ORB 特征并存放至 vector 容器中，提取的 ORB 特征如图 3-4 所示，然后调用 DBoW3 的字典生成接口生成字典。在 DBoW3::Vocabulary 对象的构造函数中能够指定树的分叉数量以及深度，这里使用了默认构造函数，也就是 k = 10，d = 5。这是一个小规模的字典，最大能容纳 10000 个单词。对于图像特征，本文亦使用默认参数，即每张图像 500 个特征点，最后把字典存储为一个压缩文件。

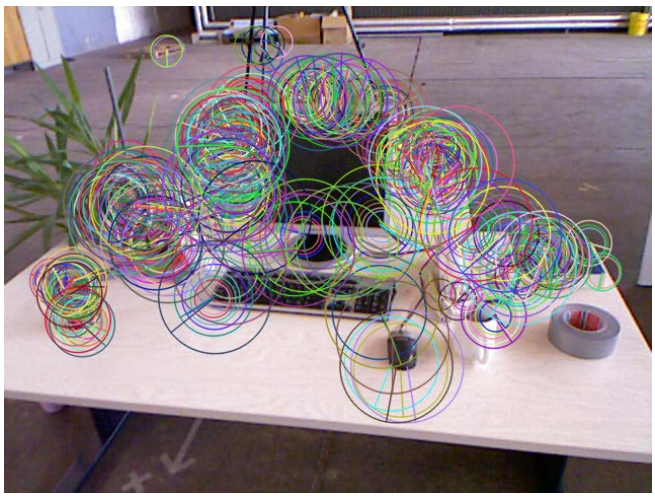

图 3-4 提取的 ORB 特征点

Figure.3-4 Extracted ORB feature points

上面已生产了字典，然后使用生成的字典来生成 Bag-of-Words，可以求得每个图像的 Bag-of-Words 描述向量，BoW 描述向量中含有每个单词的 id 和权重，它们构成了整个稀疏的向量，最后比较这两个描述向量，运用 DBoW3 计算到一个分数，计算的方式由上小节中的相似度计算中所定义。程序运行结果可得相似度值如图 3-5 所示。由图 3-5 中图片与图片中的相似度的值可以看出明显相似的是第一张图和第十张图(在 c++ 中下标为 0 和 9），相似度评分约 0.0525，而其他图像约在 0.02 左右。

```
image 0 vs image 0 : 1
image 0 vs image 1 : 0.0234552
image 0 vs image 2 : 0.0225237
image 0 vs image 3 : 0.0254611
image 0 vs image 4 : 0.0253451
image 0 vs image 5 : 0.0272257
image 0 vs image 6 : 0.0217745
image 0 vs image 7 : 0.0231948
image 0 vs image 8 : 0.0311284
image 0 vs image 9 : 0.0525447
```

图 3-5 程序运行结果图

Figure.3-5 program running results

3.4 本章小结

本章简要介绍视觉闭环检测的实现思路与过程，之后就基于人工设计特征的闭环检测展开详细介绍，包括一些人工设计特征算法如 SIFT 算法、SURF 算法和 ORB 算法，详细介绍了 ORB 算法的原理。介绍了现如今主流的闭环检测方法——词袋模型，

包括字典的建立、K-means 算法、相似度理论等。最后对基于词袋模型的闭环检测进行了实验验证，本次实验采用 ORB 算法来提高在特征点提取和特征描述子计算速度，极大提高了算法的实时性。

第四章 基于卷积神经网络的闭环检测

深度学习技术的目的是从可用于分类的原始数据中学习表示数据的方法，闭环检测本质上来说很像一个分类问题，这为典型的闭环检测问题带来了新的方法。CNN 从视觉数据特征提取抽象层次的能力已经超过基于人为设计特征解决方案的性能。特别是 CNN 在图像分类和图像检索任务方面的卓越成就是非常令人鼓舞的。考虑到视觉闭环检测类似于图像分类和图像检索，因此基于 CNN 的特征的能力用于设计视觉闭环检测问题的解决方案是合理的。本人在阅读了许多文献查阅相关资料发现近年来 vgg16 模型在图像分类以及检索方面的杰出表现，尤其是在特征提取方面 vgg16 是非常优秀的。尽管现在 resnet 和 inception 网络等等具有很高的精度和更加简便的网络结构，但是在特征提取上，vgg 比 resnet 和 inception 网络表现更加优秀，而到目前为止，本人在查阅了大量文献后发现研究者们还没有将 vgg16 模型应用到视觉 SLAM 的闭环检测中，为了进一步提高闭环检测在时间上和 PR 曲线上的性能，在这一章中，本文首次提出了将最新的图像分类与检索效果表现很好的深度模型 vgg16-places365 卷积神经网络模型应用于视觉 SLAM 的闭环检测。本文首先对该模型进行了部分参数调整的重新训练，然后将该模型在数据集 New College 上进行验证，让该模型作为图像特征提取器对图像提取特征，用提取到的特征进行相似度的比较，当相似度大于某个事先设定的阈值时就认为检测到了闭环，作出模型在该数据集上的准确率-召唤率（precision-recall，PR）曲线，该曲线被经常用来检验闭环检测算法的优劣。为了验证基于 vgg16-places365 卷积神经网络模型算法的优劣，本文将其与传统的人工设计特征的闭环检测算法以及最近几年学者们提出的基于其余深度模型的闭环检测算法在数据集 New College 上的 precision-recall 性能和特征提取速度的性能进行了对比，实验结果表明基于该模型的闭环检测在速度和准确性上取得了很好的检测效果相比于其他算法在闭环检测的表现出一定的优势,为视觉 SLAM 闭环检测提供了一种新的方法。

4.1 实验模型

4.1.1 vgg16-places365 卷积神经网络结构框架

在这篇研究中使用开源的深度学习框架 Caffe[64]来提取基于 vgg16-places365CNN

的特征。在表 4-1 中简要概述了在 Caffe 中该模型的体系结构及其每层维度。该 CNN 模型是一个多层神经网络，主要由三层类型组成：五个卷积层(conv1，conv2，conv3，conv4，conv5)，五个最大池化层(pool1，pool2，pool3，pool4，pool5)和三个完全连接的层(fc6，fc7，fc8)。输入层是一幅 224×224×3 的三通道图像，卷积层和全连接层的激活函数均采用修正线性单元[65](Rectified Linear Unit,ReLU)。ReLU 是较为常用深受研究者喜爱的激活函数，究其原因是其收敛速度快，且效果突出。常用的 ReLU 函数为：

$$f(x)=\max(x,0) \tag{4.1}$$

4.1 式中，当 $x>0$ 时，输出 x；当 $x\le 0$ 时，输出为 0。

表 4-1　vgg16-places365CNN 模型在 caffe 中的架构以及每层的特征维度

Tab.4-1　The architecture and the feature dimension of each layer of the vgg16-places365 CNN model in Caffe

层次	卷积层												
	conv1	pool1	conv2	pool2	conv3	pool3	conv4	pool4	conv5	pool5	fc6	fc7	fc8
维度	3211264	802816	1605632	401408	802816	200704	401408	100352	100352	25088	4096	4096	365

对于卷积层，convx（x=1,2）由 convx_1 和 convx_2 组成，convx（x=3,4,5）由 convx_1 和 convx_2 和 convx_3 组成，该模型使用的卷积核都很小为 3x3 ,这种卷积核是表示左右、上下、中心这样的模式的最小单元，可以捕捉到横、竖以及斜对角像素的变化。还有很特殊的 1x1 的卷积核这是做空间的线性映射。poolx（x=1,2,3,4,5） 均采用最大池化（Max Pooling）分别接在 conv1_2、conv2_2、conv3_3、conv4_3、conv5_3 后面减少特征图的大小。最大池化层为相关特征提供平移不变性并同时减小其尺寸。事实上，它也是通过合并底层本地信息来构建抽象表示的过程。而对于全连接的层，前一层中的所有神经元都完全连接到当前层的每个单个神经元。全连接层主要应用在神经网络的高层网络中，这是因为其有利于图像分类和图像检索应用。借助深层架构，CNN 能够在各种抽象层次上学习高级语义特征。当输入一幅 RGB 图像到该模型后，可以提取到每层特征可视化图如图 4-1 所示。此外，与浅层次的卷积和池化层相比，由后面的实验结果也可以看出 pool5 这样的更深层次的池层对于视觉环路闭合检测效果特别突出，因为它仍然保留输入图像的大部分空间信息并且导出输入图像的更丰富的语义表示。fc6 和 fc7 对前面卷积层和池化层提取到的特征进行融合和分类，维度均

为 4096。输出层含有 365 个神经元，表示数据集的类别数目。

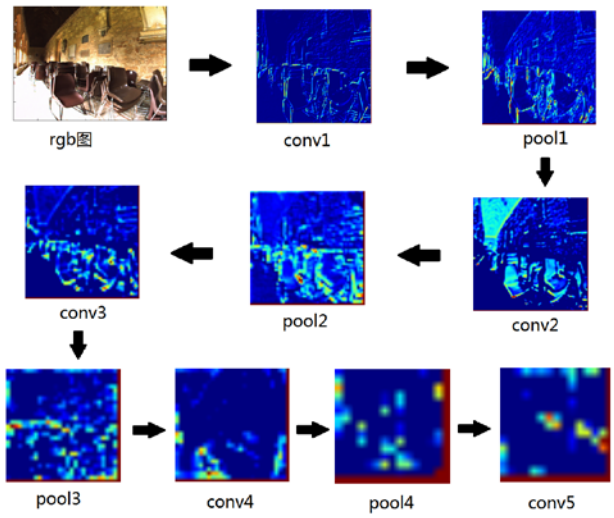

图 4-1 每层特征的可视化图

Fig.4-1 Visualization of each feature

4.1.2 vgg16-places365 卷积神经网络的训练

本文使用 vgg16-places365[66]卷积神经网络模型来作为整个图像的特征提取器，实验用 Places365-Standard 数据集对该模型进行重新训练，该数据集是由麻省理工学院收集的用于完成场景识别和场景理解任务的数据集，整个数据集包含 180 多万张场景图片，分为 365 个场景类别。当输入图像通过该网络时，每层的输出被认为是一个特征向量，随着网络层次的不断加深，获得的表达场景语义信息也越丰富，用这些特征向量来实现后续的闭环检测。因为场景识别是一个图像多分类问题，所以该网络最后采用了 Softmax 分类器对输入图像进行分类。Softmax 是经常用到的多分类器，它计算的是每个类别的几率。Softmax 用在多分类的情况下，其计算公式如下：

$$p_j = \frac{e_j^o}{\sum_k e^{o_k}} \tag{4.2}$$

其中 o 为网络模型的参数矩阵； k 为分类数；

本文对输入图像进行减均值的预处理，将处理后的图像输入到 vgg16-places365 中最终会得到一个 k 维的概率向量，进而通过公式（4.2）可预测输入图像的类别 $\hat{k}$：

$$\hat{k} = \arg\max_{1 \le i \le k} \{p_i\} \tag{4.3}$$

Softmax 分类器对应的损失函数为交叉熵损失函数（loss function），因此该网络将其作为网络的损失函数，损失函数为：

$$L(\theta)=-\frac{1}{m}\left[\sum_{i=1}^{m}\log\frac{e^{\theta_{y^{(i)}}^{T}x^{(i)}}}{\sum_{l=1}^{k}e^{\theta_{l}^{T}x^{(i)}}}\right] \tag{4.4}$$

其中 m 为每个训练批次的样本数量； θ 为网络模型的参数矩阵；$x^{(i)}$ 为第 i 个图片样本；$y^{(i)}$ 为第 i 个样本真实标签；k 为分类数。由上公式可知 loss function 是非凸的，不能通过解析的方法求解，要用迭代的方法来解决非凸问题。solver 的主要作用就是交替调用前向（forward)算法和后向（backward)算法来迭代更新参数，从而最小化损失函数的值。本文使用随机梯度迭代算法（Stochastic Gradient Descent），来求解最优参数。该网络中的权重和偏置等参数由以下公式更新：

$$V_t=\eta\cdot V_{t-1}-\lambda\cdot\nabla L(\theta_{t-1}) \tag{4.5}$$

$$\theta_t=\theta_{t-1}+V_t \tag{4.6}$$

其中 λ 是负梯度的学习率，η 是冲量系数为上一次梯度值的权重，V_t 为第 t 次迭代的参数更新值，θ_t 是第 t 次迭代的参数值。本文重新对模型进行了训练，将该模型中基准学习率 λ 设为 0.01，在迭代次数达到每 30 万次时将学习率乘以 0.1，在冲量系数上做了调整，将冲量系数 η 由 0.9 设为 0.95，这样可以使用 SGD 的深度学习方法更加稳定以及快速。

4.2 闭环检测实验流程

用前面训练好的 vgg16-places365 卷积神经网络模型来提取要验证图片的特征。将要验证的图片输入该模型后，就得到了隐藏单元每层中数据为该图片的抽象特征， 由于相似的输入会导致类似的特征，两个任意场景 (m,n) 的差异可以用它们的隐藏层的响应来表示。在基于外观的闭环检测算法中，核心问题是如何计算图像之间的相似性，本文采用余弦相似度来计算场景 (m,n) 的相似度，余弦相似度衡量的是 2 个向量间的夹角大小，通过夹角的余弦值表示结果，因此 2 个向量的余弦相似度为：

$$\cos\theta=\frac{m\cdot n}{\|m\|*\|n\|} \tag{4.7}$$

分子为向量 m 与向量 n 的点乘，分母为二者各自的 L_2 相乘，即将所有维度值的平方相加后开方。余弦相似度的取值为[-1,1]，值越大表示越相似。

闭环检测处理的图像在相邻帧间图像的特征很相似，容易被错误的认为是回环。所以在在寻找闭环时，需要设置比较图像的帧间范围，两幅比较的图像不能过于近，假设与第 k 帧图像附近相邻的图像的数量大小是 l，则与第 k 帧图像相似度比较的图像应该为从第 1 帧到第 $k-l$ 帧。本文提出的基于 vgg16-places365 卷积神经网络闭环检测方法如图 4-2 所示。对于采集的第 k 帧图像，首先对其进行减均值等预处理，然后将图像输入到 vgg16-places365 卷积神经网络模型中，提取效果表现最好的 fc6 层的输出作为图像的特征向量($Vector_k$)，计算其与先前图像之间的特征向量($Vector_{1...K-L}$)之间的相似度，若相似度大于事先设定的阈值，则认为检测到了闭环。

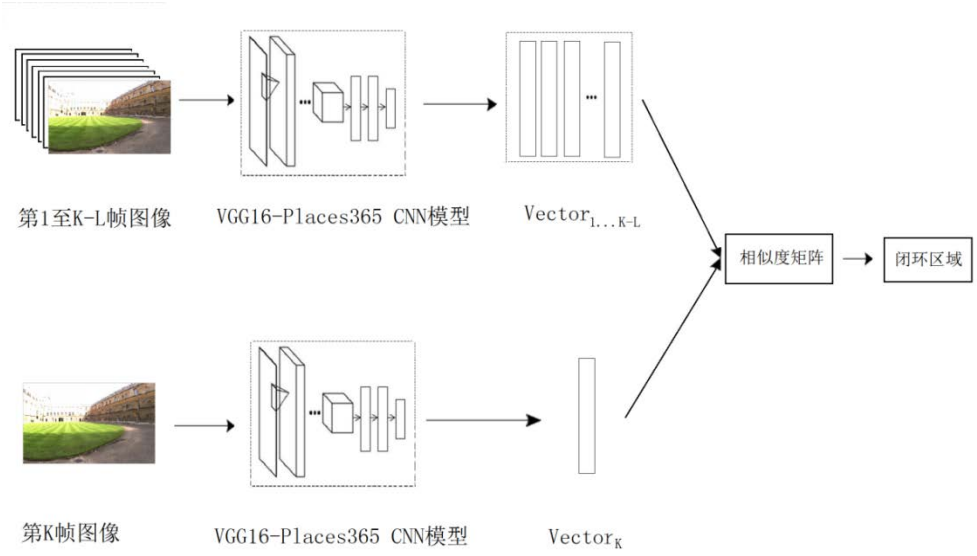

图 4-2 闭环检测方法示意图

Fig.4-2 Sketch map of loop closure detection

4.3 实验结果与分析

4.3.1 实验环境

为了验证基于 vgg16-places365 卷积神经网络闭环检测方法的性能，本文将其与基于传统人工设计特征的 BoVW 以及其余几种基于深度模型的闭环检测方法进行了比较，分别比较了它们之间准确率、召唤率以及提取所花的时间等这些性能。实验数据集是闭环检测数据集 New College。本次实验所用电脑的硬件配置为：CPU-3.5GHz，RAM-32GB，显卡 GTX-1070。软件平台系统为 ubuntu，在其上使用 caffe 框架进行网络模型的搭建；除此之外还安装了 CUDA8.0，使用 CUDA 并行计算结构，这充分利用了 CPU 和 GPU 各自的优点；安装了 cuDNN 加速库用来给 GPU 加速，最后使用 caffe

的 python 接口用 python 程序来加载模型对图像进行特征提取与相似度计算。

4.3.2 实验数据集

New College 数据集是由牛津大学移动机器人团队收集的专门用于视觉 SLAM 的闭环检测算法的评估验证数据集。这个标准数据集包含了 1073 对图像，它们是由在移动平台的左边和右边各放置一个摄像头，移动平台每行进 1.5 m 采集一次图像来进行收集。图 4-3 给出了数据集的示例图像。数据集中还给出了形成闭环区域的真实标注，标注以矩阵的形式给出，若图像 i 与图像 j 形成闭环区域，则 (i, j) 对应的数值为 1，否则为 0。闭环检测是 L 值的设置方式与文献[15]相同，将 New College 数据集的 L 值设为 100。

图 4-3 New College 数据集的同一位置左、右摄像头采集的示例图像

Fig.4-3 Sample images of the left and right cameras collected at the same position in New College

4.3.3 PR 性能比较

为了找出最适用于视觉环路闭合检测的基于 vgg16-places365 卷积神经网络的特征，本文评估了来自所有层（conv1、conv2、conv3 层除外）的基于该网络特征的性能。图 4-4 显示了在 New College 数据集上基于 vgg16-places365 卷积神经网络中的某些层比较的实验结果(precision-recall 曲线)。由图 4-4 可以看出随着召回率的不断增加，在最开始时各种算法的准确率是都为 100%的，前文中提到过召回率代表的是在所有的真实的闭环中，被正确无误地检测出来的闭环的概率。召回率低代表算法严格，它检测出来的闭环也将更加的少，因此准确率很高。随着召回率的增加，对于闭环检测的算法将变得慢慢不是那么严格，准确率也将慢慢下降。由图 4-4 还能清晰的知道对于卷积层和最大池化层来说,可以看出相同召回率的情况下，准确率性能在逐层递增；而最

终完全连接的层性能却是在不断下降，造成这个原因是对于卷积层来说深度越深其对特征的抽象程度越高，所能体现的语义信息也越丰富，所以层次越深其越能更好的代表图像特征，因此性能是在逐渐增加的。对于全连接层来说，因为全连接成会丢失空间信息，所以用高层次的全连接层提取的特征视觉环路闭合检测的效果不是很好，表明深度学习的深度和编码图像时空间信息的深度的重要性。由图可以看出 fc6、pool5、conv5 三层的特征在闭环检测上的性能优于其它层。其中以 fc6 层性能最佳，当召唤率低于 40%时，其准确率为 100%；当召唤率达到 40%时，其准确率也达到 90%左右。

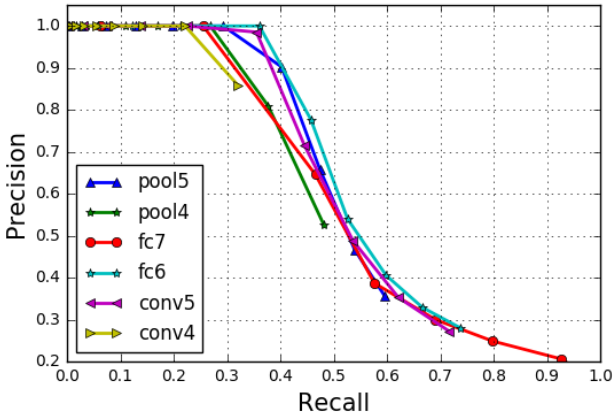

图 4-4 基于 vgg16-places365 CNN 模型的各个层次特征的闭环检测 Precision-Recall 曲线比较

Fig.4-4 Comparison of Precision-Recall curve for loop closure detection of various layer features of vgg16-places365 CNN Model

为了比较不同算法对视觉闭环检测的有效性，本文还做了与上述类似的实验。实验同样在 New College 数据集上对基于 vgg16-places365 卷积神经网络的描述子的两个最优表现层特征以及基于人工设计的特征(BoVW，GIST)的性能进行了比较，其实验结果如图 4-5 所示。可以看到，根据两个评估标准，基于 vgg16-places365 卷积神经网络 fc6 层的描述子的表现胜过人工设计的描述子的表现，在相同的召回率下其准确率要高约 3%~4%。这是因为人工设计的特征主要依靠设计者的先验知识，很难利用大数据的优势。由于依赖手工调参数，特征的设计中只允许出现少量的参数，而利用深度模型学习得到的数据特征对大数据的丰富内在信息更有代表性，深度学习可以从大数据中自动学习特征的表示，其中可以包含成千上万的参数。除此之外，利用人工设计的特征得到的是局部特征，而利用深度学习得到的是全局特征。

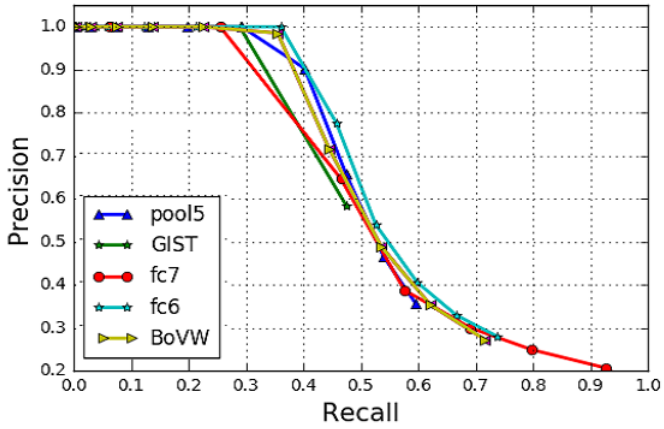

图 4-5　基于 vgg16-places365 CNN 模型与传统人工设计特征的 PR 曲线对比

Fig.4-5 Comparison of PR Curve of vgg16-places365 Model and Traditional Artificial design features

本文除了对比传统的人工设计特征视觉闭环检测的有效性外，还对比了其它最新深度学习在闭环检测性能上的对比，实验同样在 New College 数据集上对基于 vgg16-places365 卷积神经网络的描述子的两个最优表现层特征以及最近几年基于深度学习的 PlaceCNN、Autoencoder、FLCNN 等的性能进行了比较，其实验结果如图 4-6 所示。由图知，根据两个评估标准，基于 vgg16-places365 卷积神经网络 fc6 层的描述子的表现略微优于其余基于卷积神经网络最优层的描述子（FLCNN、PlaceCNN）的表现，当召回率相同时有着更高的准确率。这是因为基于 vgg16-places365 卷积神经网络结构更复杂，每层特征维度更多，因此其更能比较好的体现特征。除此之外，由图还可知基于卷积神经网络的描述子要优于基于 Autoencoder 在闭环检测上的性能，这是因为基于卷积神经网络的模型对于场景分类问题表现更加好。

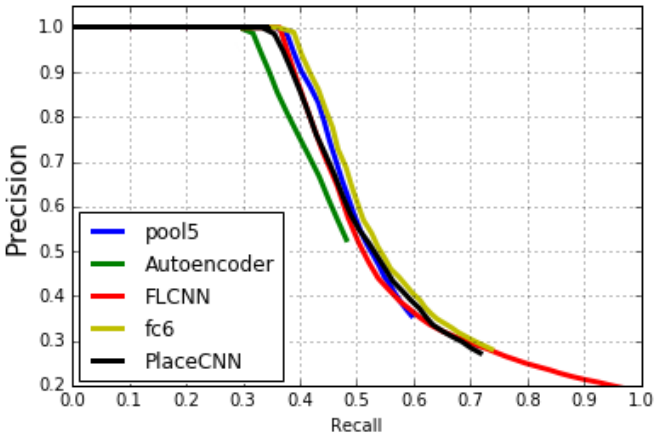

图 4-6　基于 vgg16-places365 CNN 模型与基于其余深度学习模型的 PR 曲线对比

Fig.4-6 Comparison of PR Curve of vgg16-places365 Model and Other Deep learning models

4.3.4 时间性能比较

评估图像描述子的另一个重要考虑因素是其计算效率。这个效率是通过提取描述子所需的时间和描述子的长度来衡量的。本文比较了基于 vgg16-places365 卷积神经网络的闭环检测与其它两种人工设计特征的闭环检测以及其它三种基于深度模型闭环检测算法所需要的时间，得到结果如表 4-2 所示。这里的时间是 1000 张图像的平均值，其中深度模型只包括特征提取时间，不包括加载输入图像或卷积神经网络模型的时间。可以看出在 CPU 上时，基于 vgg16-places365 描述子比人工设计特征的描述子更有效，每幅图像平均需时 0.139 s，比其他两种人工设计特征方法快约 5~10 倍。如果在 GPU 上提取特征，则速度还要快一个数量级达到 0.014 s。由表 4-2 还可知道基于 vgg16-places365 卷积神经网络模型特征提取速度要快于基于 Autoencoder 和 PlaceCNN 两模型。因 FLCNN 是一种专门设计的快速、精简的卷积网络模型，有着很好的实时性，vgg16-places365 卷积神经网络模型在提取特征速度上要稍微慢于 FLCNN 模型，虽速度上要略慢，但是差距很小，而且由图 4-6 可知，vgg16-places365 卷积神经网络模型在闭环检测上的 PR 性能要优于 FLCNN 模型。

表 4-2 用不同特征描述子来计算所得的平均每幅图像所花费时间

Tab.4-2 The cost time spent on the average of each image with different feature descriptors

特征	人工设计的特征		vgg16-places365		Autoencoder	PlaceCNN	FLCNN
	BoVW	GIST	CPU	GPU	GPU	GPU	GPU
时间（s）	0.567	1.787	0.139	0.014	0.020	0.052	0.013

4.4 本章小结

本章是这次课题研究所做工作最重要的部分。给出了此次课题研究的想法来源，提到了本文的创新之处在于将一个在图像分类和检索表现很好的深度模型应用到了视觉 SLAM 的闭环检测上，在移动机器人的定位与导航工程领域内具有一定的应用创新性。介绍了 vgg16-places365 卷积神经网络模型的框架，用大型的场景识别标准数据集重新对该模型进行了训练，在训练过程中修改了部分训练参数以作出适应性调整。给出了实验用的闭环检测方法流程，在 NewCollege 数据集上进行实验，用 vgg16-places365 卷积神经网络模型作为场景图像的特征提取器对图像提取特征，利用余弦相似度来计算提取到的两特征向量之间的相似性，从而得出相似度矩阵，根据相似度矩阵判断闭

环区域。为了进行对比验证，本实验对比了在 New College 数据集上与传统的人工设计特征的闭环检测算法，以及最近几年学者们提出的基于其他深度模型的闭环检测算法的性能表现，在实验结果及分析部分给出了对比结果与原因分析。实验结果表明对于 vgg16-places365 卷积神经网络模型的各层特征表现来说，fc6 层在检测精度和表示的紧凑性方面表现最佳。此外，实验结果表明基于 vgg16-places365 模型闭环检测的性能要优于人工设计的特征以及其余三种深度模型在闭环检测上的性能，在同样的召唤率下其准确率更高。本文还在时间性能上进行了对比，基于 vgg16-places365 卷积神经网络模型的图像特征提取比基于人工设计特征提取在 CPU 上的速度要快一个数量级，而在 GPU 上提取速度更是比传统人工设计特征快约两个数量级，在 GPU 上也比基于 Autoencoder 和 PlaceCNN 要快 2 到 3 倍，虽速度上略慢于基于 FLCNN 模型的方法，但差距很小，也有着良好的实时性。

总结与展望

SLAM 技术是实现移动机器人自主导航的关键技术之一，由于视觉传感器具有便宜、丰富信息量、轻巧等诸多优点，视觉 SLAM 的研究是移动机器人视觉导航中的核心问题之一。本文介绍了移动机器人视觉 SLAM 的研究历程，以及移动机器人 SLAM 问题的研究背景，然后针对其中闭环检测模块展开研究分析，因为闭环检测是 SLAM 中的一个关键问题，是机器人判断自己当前位置是否位于已访问过的环境区域，成功检测出闭环，可以显著地减小前端视觉里程计的累积误差，并以此作为地图是否需要更新校正的依据，对于提高大规模 SLAM 的鲁棒性有重大意义。本文完成的主要工作总结如下：

1．查阅大量国内外文献资料，分析了移动机器人视觉导航中的视觉 SLAM 问题，概述了视觉 SLAM 研究现状，并且阐述了闭环检测的研究现状，

2．阐述了视觉 SLAM 框架中各个部分算法原理，重点详细介绍了闭环检测的相关原理，之后对整个 SLAM 模块的数学模型进行了推导，并给出了求解的方法。

3．如今视觉 SLAM 闭环检测的主流做法为传统的基于人工设计特征的词袋模型方法。比较了 SIFT、SURF 和 ORB 这三种特征点算法的优缺点，本文最终使用了 ORB 算法实现特征提取。然后用视觉词袋模型（BoVW）方法来判断是否产生闭环，实验过程如下：利用 K-means 算法对提取到的特征进行聚类生成字典，用生成的字典得到每幅图像的描述向量，然后通过计算图像的描述向量之间的相似度判断是否检测到了闭环。实验结果表明第一张图片与第十张图片相似度最高，从外观上很明显可以看出这两张图片是一个地方，因此最终成功地检测出了闭环。

4．利用深度学习的方法对闭环检测进行研究，首次将在图像分类与检索表现优秀的 vgg16-places365 卷积神经网络模型应用在视觉 SLAM 闭环检测上，本文还可视化了每层提取到的特征，使之更加形象化。详细介绍了 vgg16-places365 卷积神经网络模型的框架，然后介绍了模型的训练参数的设置，然后给出了实验用的闭环检测方法，用余弦相似度来计算两特征向量之间的相似性，最后在标准的闭环检测数据集 New College 上进行测试，分别和几种传统基于人工设计特征的方法（BoVW、GIST 等）以及其他几种深度学习模型的方法进行了对比验证，实验结果表明本文采用的 vgg16-places365 卷积神经网络模型在闭环检测的 PR 性能和特征提取时间性能上具有

比较好的优势，为视觉 SLAM 闭环检测提供了一种新方法。

在本论文中，只对一些方面做出了研究，仍有许多问题有待于进一步研究与完善，具体包括以下几个方面：

1．对于主流的传统人工特征方法在特征提取与描述子计算方面要尝试使用更新更快的算法进一步提高闭环检测算法的效率。

2．要设计更加精简、快速而又对于闭环检测效果比较好的深度模型，本文用的模型虽然效果比较好，但是训练时间长，提取特征文件所占资源也较大，因此更高配置的计算机资源也是必要。

3．本文采用对比验证试验还略显不够，没有考虑光照等条件变化对各种算法能否成功检测出闭环的影响的对比，后续尝试增加对该方面试验进行验证。

4．本文只在一个数据集上进行了验证，如果能在多个数据集上进行验证，将会更具有说服力。

参考文献

[1] Nilsson N J. A Mobius Automation:an Application of Artificial Intelligence Techniques[C]//Proceedings of the 1st international joint conference on Artificial intelligence. San Francisco:Morgan Kaufinann Publishers Inc, 1969:509-520.

[2] Joseph L J. Robots at the tipping point:the road to irobot roomba[J]. IEEE Robotics & Automation Magazine, IEEE, 2006. 13(1):76-78.

[3] Jia Y H, Mei F X. Simple Path Planning for Mobile Robots in the Present of Obstacles[J], Journal of Beijing Institute of Technology, 2002.

[4] 王建农, 吴捷. 自主移动机器人的导航研究[J]. 机器人, 1997. 19(6): 461-473.

[5] 朴松昊, 洪炳熔. 一种动态环境下移动机器人的路径规划方法.机器人[J], 2003. 25(l):18-19.

[6] 罗荣华, 洪炳熔. 移动机器人同时定位与地图创建研究进展[J]. 机器人, 2004, 26(2):183-186.

[7] Leonard J, Durrant-Whyte H F, Cox I J. Dynamic Map Building for an Autonomous Mobile Robot[J]. International Journal of Robotics Research, 1990, 11(4):286-298.

[8] Leonard J，Durrant-Whyte H．Mobile Robot Localization by Tracking Geometric Beacons．IEEE Transactions on Robotics and Automation,1991，7(3)：376--382．

[9] Christian S ,Thomas K．Filter design for simultaneous localization and map building (SLAM)．In Proceedings of IEEE International Conference on Robotics and Automation, Washington,USA，2002．2737-2742．

[10] Leonard J, Durrant-Whyte H F. Mobile robot localization by tracking geometric beacons[J]. IEEE Transactions on Robotics and Automation, 7(4):376-382, 1991.

[11] Montemerlo M, Thrun S, Koller D, et al. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges[C]//International Joint Conference on Artificial Intelligence. Morgan Kaufmann, 2003, 133(1):1151-1156.

[12] 陈伟,吴涛,李政等. 基于粒子滤波的单目视觉 SLAM 算法[J]. 机器人, 2008, 30(3):242-247.

[13]Lu F, Milios E. Globally consistent range scan alignment for environment mapping[J]. Autonomous Robots, 1997, 4(4):333-349.

[14]H Wang, Hou L. Online mapping with a mobile robot in dynamic and unknown environments[J]. International Journal of Modelling, Identification and Control, 2008,4(4): 415-423.

[15]Smith R, Self M, Cheeseman P. Estimating uncertain spatial relationships in robotics[C]//Uai 86: Second Conference on Uncertainty in Artificial Intelligence. Elsevier, 1986, 5(5):435-461.

[16]Chiuso A , Favaro P, Jin H, et al. 3-D Motion and Structure from 2-D Motion Causally Integrated over Time: Implementation[C]// Computer Vision ECCV 2000. Springer Berlin Heidelberg, 2000, 24(4):523-535.

[17]Einicke G A, White L B. Robust extended Kalman filtering[J]. Signal Processing IEEE Transactions on, 1999, 47(9):2596-2599.

[18]Montemerlo M, Thrun S, Koller D, et al. FastSLAM:a factored solution to the simultaneous localization and mapping problem[J]. Archives of Environmental Contamination & Toxicology, 2003, 50(2):240-248.

[19]Montemerlo M, Thrun S, Roller D, et al. FastSLAM 2.0: an improved particle filtering algorithm for simultaneous localization and mapping that provably converges[C]//International Joint Conference on Artificial Intelligence. Morgan Kaufmann, 2003, 133(1):1151-1156.

[20]Grisettiyz G, Stachniss C, Burgard W. Improving Grid-based SLAM with Rao-Blackwellized Particle Filters by Adaptive Proposals and Selective Resampling[C]//IEEE International Conference on Robotics & Automation. IEEE, 2005, 312(20):2432-2437.

[21]Sibley G, Matthies L, Sukhatme G. Sliding window filter with application to planetary landing[J]. Journal of Field Robotics, 2010, 27(5):587-608.

[22]康轶非, 宋永端, 宋宇. 平方根容积卡尔曼滤波在移动机器人 SLAM 中的应用[J]. 机器人, 2013, 35(2):186-193.

[23]Strasdat H, Montiel J , Davison A J. Visual SLAM: Why filter[J]. Image & Vision Computing, 2012, 30(2):65-77.

[24]Lu F, Milios E. Globally Consistent Range Scan Alignment for Environment Mapping[J]. Autonomous Robots, 1997, 4(4):333-349.

[25]Thrun S, Montemerlo M. The graph SLAM algorithm with applications to large-scale mapping of urban structures[J]. The International Journal of Robotics Research. 2006, 25(5-6): 403-429.

[26]Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces[C]//IEEE & Acm International Symposium on Mixed & Augmented Reality. IEEE, 2007:1-10.

[27]C Forster, M Pizzoli, D Scaramuzza. “Svo: Fast semi-direct monocular visual odom-etry”[C]//Robotics and Automation (ICRA), 2014 IEEE International Conference on, p-p. 15–22, IEEE, 2014.(SVO)

[28]Mur-Artal R, Montiel J, Tardós J. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. IEEE Transactions on Robotics, 2015, 31(5):1147-1163.

[29]J. Engel, J. Sturm. “Semi-dense visual odometry for a monocular camera”[C]//in Proceedings of the IEEE international conference on computer vision, pp. 1449–1456,2013.

[30]J Engel, T Schops, and D Cremers. “Lsd-slam: Large-scale direct monocular slam”[C]// Computer Vision–ECCV 2014, pp. 834–849, Springer, 2014.(LSD-SLAM)

[31]D Filliat. A visual bag of words method for interactive qualitative localization and mapping[J]. International Conference onRobotics and Automation (ICRA), 2007:3921-3926.

[32]D Lowe. Distinctive image features from scale-invariant keypoints [J], International Journal of Computer Vision,2004, 60(2): 91-110.

[33]H Bay,Tuytelaars,Van Gool. Surf: Speeded up robust features [C]//Computer Vision-ECCV, 2006, 404-417.

[34]E Rublee, V Rabaud, K Konolige. Orb: an efficient alternative to sift or surf [C]// IEEE International Conference on Computer Vision (ICCV), 2011: 2564-2571.

[35]Mark Cummins and Paul Newman. Highly Scalable Appearance-Only SLAM - FAB-MAP 2.0. In Robotics [J]. Science and Systems (RSS), 2009:1-8.

[36]Cummins M, Newman P. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance [J]. International Journal of Robotics Research, 2008,

27(6):647-665.

[37]Liu Y, Zhang H. Visual loop closure detection with a compact image descriptor [C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, Vilamoura-Algarve, Portugal,2012:1051-1056.

[38]Gao X, Zhang T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system [J]. Autonomous Robots, 2017, 41(1): 1-18.

[39]Gao X, Zhang T. Loop closure detection for visual slam systems using deep neural networks [C]//TCCT of Chinese Association of Automation(CAA), 2015: 5851-5856.

[40]Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks [J]. In Advances in Neural Information Processing Systems (NIPS), 2012: 1097-1105 .

[41]K Chatfield, K Simonyan, A Vedaldi, A Zisserman. Return of the Devil in the Details: Delving Deep into Convolutional Nets [J]. British Machine Vision Conference, 2014: 1-11.

[42]Ji Wan, Dayong Wang, Deep Learning for Content-Based Image Retrieval: A Comprehensive Study [J]. ACM International Conference on Multimedia (MM), 2014: 157-166 .

[43]Artem Babenko, Anton Slesarev, Alexandr Chigorin. Neural Codes for Image Retrieval [J]. European Conference on Computer Vision , 2014,8689: 584-599 .

[44]何元烈,陈佳腾,曾碧.基于精简卷积神经网络的快速闭环检测方法[J].计算机工程,2017:1-6.

[45]Xia Y, Li J, Qi L, et al. Loop closure detection for visual SLAM using PCANet features [C] International Joint Conference on Neural Networks.Vancouver,Canada, IEEE,2016: 2274-2281.

[46]Hou Y, Zhang H, Zhou S. Convolutional Neural Network-Based Image Representation for Visual Loop Closure Detection [C]//International Conference on Information and Automation, Lijiang, China, 2015: 2238-2245.

[47]Gokturk SB, Yalcin H, Bamji C. A Time-Of-Flight Depth Sensor - System Description, Issues and Solutions[C]//Conference on Computer Vision & Pattern Recognition Workshop. IEEE, 2004:35-35.

[48] Rusinkiewicz S, Levoy M. Efficient variants of the ICP algorithm[C]//The Third International Conference on 3-D Digital Imaging and Modeling. IEEE, 2001: 145-152.

[49] V Lepetit, F Moreno-Noguer, P Fua. “Epnp: An accurate o(n) solution to the pnp problem,” International Journal of Computer Vision, vol. 81, no. 2, pp. 155–166, 2008.

[50] M Lourakis,Argyros. “Sba: A software package for generic sparse bundle adjustment,” ACM Transactions on Mathematical Software (TOMS), vol. 36, no. 1, p. 2,2009.

[51] D Hahnel, W Burgard, D Fox, and S Thrun. “An efficient fastslam algorithm for generating maps of large-scale cyclic environments from raw laser range measurements,” in Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on, vol. 1, pp. 206–211, IEEE, 2003.

[52] P Beeson, J Modayil, B Kuipers. “Factoring the mapping problem: Mobile robot map-building in the hybrid spatial semantic hierarchy,” International Journal of Robotics Research, vol. 29, no. 4, pp. 428–459, 2010. Times Cited: 16 Beeson, Patrick Modayil, Joseph Kuipers, Benjamin 16.

[53] R Mur-Artal, J Montiel,D Tardos. “Orb-slam: a versatile and accurate monocularslam system”[C]// IEEE Transactions on Robotics. 2015 :1147-1163.

[54] Y Latif, C Cadena, J Neira. “Robust loop closing over time for pose graph slam”[J]. The International Journal of Robotics Research, vol. 32, no. 14, pp. 1611–1626, 2013.

[55] Ulrich,Nourbakhsh. “Appearance-based place recognition for topological localization”[C]// in Robotics and Automation, 2000. Proceedings. ICRA’00. IEEE International Conference on, vol. 2, pp. 1023–1029, Ieee, 2000.

[56] Lowe D G. Distinctive Image Features from Scale-Invariant Keypoints[J]. International Journal of Computer Vision, 2004, 60(2):91-110.

[57] Bay H, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3):346-359.

[58] Calonder M, Lepetit V,Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//IEEE International Conference on Computer Vision. IEEE, 2011, 58(11):2564-2571.

[59] Strecha C, et al. BRIEF: Binary Robust Independent Elementary Features[C]//European Conference on Computer Vision. Springer, 2010, 6314:778-792.

[60] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]//European Conference on Computer Vision. Springer, 2006, 3951:430-443.

[61] S. Lloyd. “Least squares quantization in pcm”[C]// IEEE transactions on information theory, vol. 28, no. 2, pp. 129–137, 1982.

[62] J Sivic,A Zisserman. “Video google: A text retrieval approach to object matching in videos”[C]// in Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on,pp. 1470–1477, IEEE, 2003.

[63] S Robertson. “Understanding inverse document frequency: on theoretical arguments for idf”[J]. Journal of documentation, vol. 60, no. 5, pp. 503–520, 2004.

[64] Yangqing Jia, Evan Shelhamer, Jeff Donahue. Caffe: Convolutional Architecture for Fast Feature Embedding [J]. arXiv preprintarXiv:1408.5093, 2014:1-4.

[65] Shang W, Sohn K, Almeida D, et al. Understanding and Improving Convolutional Neural Networks via Concatenated Rectified Linear Units [C]\\ International Conference on Machine Learning(ICML).2016.

[66] Zhou B, Khosla A, Lapedriza A,et al. Places: An image database for deep scene understanding [J]. arXiv preprintarXiv:1610.02055, 2016:1-12.

攻读学位期间发表的论文

[1] 杨孟军、苏成悦、陈静、张洁鑫. 基于卷积神经网络的视觉闭环检测研究[J]. 广东工业大学学报.

学位论文独创性声明

本人郑重声明：所呈交的学位论文是我个人在导师的指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明，并表示了谢意。本人依法享有和承担由此论文所产生的权利和责任。

论文作者签名：杨孟华　日期：2018.05.28

学位论文版权使用授权声明

本学位论文作者完全了解学校有关保存、使用学位论文的规定："研究生在广东工业大学学习和工作期间参与广东工业大学研究项目或承担广东工业大学安排的任务所完成的发明创造及其他技术成果，除另有协议外，归广东工业大学享有或特有"。同意授权广东工业大学保留并向国家有关部门或机构送交该论文的印刷本和电子版本，允许该论文被查阅和借阅。同意授权广东工业大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印、扫描或数字化等其他复制手段保存和汇编本学位论文。保密论文在解密后遵守此规定。

论文作者签名：杨孟华　日期：2018.05.28

指导教师签名：[illegible]　日期：2018.05.28

致 谢

时间转瞬即逝，三年的研究生求学生涯即将结束，站在毕业的门槛上，回首往昔，奋斗和辛劳成为丝丝的记忆，甜美与欢笑也都尘埃落定。广东工业大学以其优良的学习风气、严谨的科研氛围教我求学，以其博大包容的情怀胸襟、浪漫充实的校园生活育我成人。

本论文是在导师苏成悦教授的悉心指导之下完成的。三年来，导师渊博的专业知识，严谨的治学态度，精益求精的工作作风，诲人不倦的高尚师德，朴实无华、平易近人的人格魅力对我影响深远。导师不仅授我以文，而且教我做人，虽历时三载，却赋予我终生受益无穷之道。本论文从选题到完成，几易其稿，每一步都是在导师的指导下完成的，倾注了导师大量的心血，在此我向我的导苏成悦教授表示深切的谢意与祝福！

另外感谢课题组成员林上飞、林君宇、王木华、张勇、张洁鑫、肖志聪、陈洪极、朱文杰、陈科诚、梁高鹏等对我在学习和科研上的帮助。有了他们的帮助，我的研究才能够顺利的完成，和他们相处的日子里也让我感受到团队的温馨。

还要感谢父母的养育之恩，以及在我求学生涯中给与我无微不至的关怀和照顾，一如既往地支持我、鼓励我，是他们近 20 多年来精神上和经济上的全力支持才使得我能够完成最后的学业，对此是万分敬意和感谢。

此外本文有部分图片来自网络，在这里特此感谢这些图片的提供者。

最后，值此毕业论文完成之际，我谨向所有关心、爱护、帮助我的人们表示最诚挚的感谢与最美好的祝愿。