第20周汇报

——张溢炉

1. 《Single View Point Omnidirectional Camera Calibration from Planar Grids》总结
2. 内容摘要

本文介绍利用棋盘格标定全向相机的方法

1. 总结及理解
2. sectionII中摄像机成像模型可用于EuRoc中的鱼眼相机，在sectionIII中，自己根据vins源码推导出了求逆过程。
3. 启发
4. 《Unified Temporal and Spatial Calibration for Multi-Sensor Systems》总结
5. 内容摘要

本文介绍了一种offline利用棋盘格联合标定多传感器系统时间偏置和传感器空间相对位姿的最大似然估计的方法，可以将本文VI方法扩展到多传感器。

1. 总结及理解
2. 之前很多方法是先标定时间偏置再标定相对位姿
3. 获取多传感器时间戳的方法，这里产生时间戳时刻为接收到传感器数据那一刻。面对的主要问题：A、产生时间戳时刻和真实时刻存在延迟且不同传感器不一样；B、不同传感器时钟的起始时间可能不同；C、不同时钟的周期可能存在细微差别（可能一个是0.9999s一个是1.00002s），这样会造成累积误差。我们的目的让记录的时间戳尽可能接近采样时刻（这样可以保证同一时间戳对应所有传感器的数据是在同一地点同一时刻获取的），但延迟是不可避免的，最好每个传感器延迟是固定不变的（频率高（时间分得更细）通过时间戳+-td便能和频率低时间对齐，当然不能完全对齐，因为数据是离散的），但往往存在随机延迟。以下方法都使用了同一个时钟，故不存在B、C的问题，
4. 硬件同步：使用同一个硬件时钟同时触发各个传感器，如果记录的时间戳是触发后得到图像的时间，从获取到传感器参数（真实时刻，采样时刻）到得到图像并记录时间戳还有一段时间，这段时间称为传感器内部延迟（internal sensor delays-logic delay（逻辑处理）、filters delay（获取原始数据过程中可能需要滤波）
5. 软件同步（适合自己两轮机器人，都已联网，需要先搞清楚摄像头和激光雷达功能包使用什么时钟系统）：NTP[2]、TICSync[3]。往往不同设备可能不支持相同软件。
6. （本文面向对象）使用某一传感器时钟作为中心时钟，通过触发或轮询的方式得到各传感器读数，或者接收传感器以固定周期发布的数据。由此产生的延迟如下，a）、b）不会带来延迟1。
7. communication delays：其他传感器传到中心时钟所在传感器之间存在延迟，额外增加。一般为常数。
8. internal sensor delays：filters或logic，a)中有介绍。一般为常数。
9. stochastic（随机） delays：可能由终端、多线程或者随机函数等带来。环节越多带来随机延迟的可能性越大，比如c）相对于a）、b）。[4]中包含去除随机延迟的方法。
10. 不相关的量一起估计，可能存在有害的相互干扰[7]，但本文证明放在一起好于分开估计
11. 本文创新点（与之前方法的不同点）：
12. 联合估计，连续时间（将状态通过B-spline逼近为连续状态）
13. VI联合估计具体方法
14. 使用仿真和真实数据集做了实验，证明可以在几个高频传感器周期得到估计
15. 利用了IMU加速度数据，相对5、6
16. 重要文献

* 去除随机延迟：4
* 连续时间最大似然估计：7
* 与本文方法相似：13
* B-spline：用于将状态变成时间连续的，近似结果与阶数和节点数有关。14

1. 启发
2. 《Online Temporal Calibration for Monocular Visual-Inertial Systems》总结
3. 内容摘要

本文是vins中online与其他变量一起进行迭代非线性捆集优化的时间偏差标定的方法。

1. 总结及理解
2. 本文方法产生时间戳时刻在采样时刻之后，它们的差称为时间offset（延迟）td（几ms-几百ms），见图1。我们使用的是时间戳，首先假设其是准确的，延迟一般是未知的常数，也是本文假设，td产生的原因如下：
3. 触发（triggering）延迟：如相机的曝光时间，td和曝光时间成比例关系
4. 传输（transmission）延迟：数据传输过程
5. 时钟不同步：指时钟起始时间不一样
6. 另外，如果传感器使用不同时钟采集信息且周期存在累积偏移，将很难融合，能否看成随时间变化的线性函数
7. 创新点，重点与上篇论文对比：
8. Online
9. 通过大量实验展示了online时间偏差标定的作用（未标定时，vins在EuRoc数据集上延迟超过6ms会对结果造成较大影响）
10. 将该方法加入开源项目验证
11. 算法：重要假设：td是未知常数，像素在归一化平面几帧时间的运动速度不变。
12. Temporal offset（td）：timu=tcam+td
13. Feature Velocity on Image Plane：像素点在两帧之间归一化平面移动速度
14. 表示3D点的两种方法
15. 直接用世界坐标系的3D坐标表示
16. 使用某些帧或一帧的深度或深度倒数来表示，可以通过该帧位姿恢复3D坐标
17. 重要文献

* 标定方法：19，20，21（上一篇），22（online，非迭代）
* 可作ground truth方法：optitrack（p3668）
* 评价方法：28、29

1. 启发
2. 《An Efficient Solution to the Five-Point Relative Pose Problem》总结
3. 内容摘要

本文介绍了利用5点法得到两帧或三帧及以上图像up-to-scale位姿的方法

1. 总结及理解
2. 可以综合利用5点法和8点法，5点法在侧移时表现好，8点法在非侧移时表现好。
3. 对于点共面的情况，2帧时会增加模糊自由度，3帧可以减少模糊自由度。
4. 5、6点法需要知道内参，而7、8点法不用
5. 5点法步骤：具体见论文
6. Extraction of the nullspace of a 5\*9 matrix
7. Expansion of the cubic constraints (5) and (6)
8. Gauss-Jordan elimination with partial pivoting on

the 10\*20 matrix A.

1. Expansion of the determinant polynomial of the

3\*3 polynomial matrix B to obtain the tenth degree polynomial (14).

1. Extraction of roots from the tenth degree polynomial.
2. Recovery of R and t corresponding to each real root and point triangulation for disambiguation.
3. 重要文献

* 6、7、8点法：p762左下角
* 计算多项式近似文献：32

1. 启发
2. 《Robust Initialization of Monocular Visual-Inertial Estimation on Aerial Robots》总结
3. 内容摘要

更详细的介绍了vins初始化的过程。

1. 总结及理解
2. 需要先对VIO中一些值进行较好的初始化，否则在VIO优化时会陷入局部最优甚至崩溃。初始化先通过优化的方法得到较好的角速度偏差、速度、重力向量和尺度（不包括相机和IMU相对位姿，会造成什么影响呢？），然后使用VIO优化位姿、速度、角速度和加速度偏差，以及相机和IMU相对位姿。
3. 本文为了更快初始化，忽略了加速度偏差的初始化，具体看IV-E，有图表展示
4. 很多方法都是在静止或沿固定轨迹运动时进行初始化，但本文方法是在动态、随意运动中初始化
5. 初始化时，视觉需要保证足够的视差，IMU需要保证足够的刺激（加减速和旋转）
6. 10%的scale误差是可以接受的，实际上VIO可以容忍30%
7. 创新点

* A robust visual-inertial initialization procedure, which

can provide initial states for on-the-fly aerial robots.

* At the system level, we applied the proposed approach into sliding-window based monocular visualinertial system. Onboard closed-loop autonomous flight

experiments are performed.

* Open source code for the community

1. 重要文献

* 预积分：25、8、26
* 作者文献：7、8、12
* 用于稠密地图、导航、规划：33

1. 启发

第19周汇报

——张溢炉

1. 《Laser–visual–inertial odometry and mapping with high robustness and low drift》总结
2. 内容摘要

本文概述了一种使用相机、3D激光雷达和IMU的多传感器融合的slam算法。不同于之前的KF或因子图优化，本文使用了序列的、多层的pipeline 的形式，由粗到细进行处理。利用IMU估计初值，使用VIO进行优化，最后使用3D激光雷达进一步细化，同时建立稠密、准确的3D地图。第一篇能在跳跃等剧烈运动下进行高频位姿估计并建立3D地图的论文。

1. 总结及理解
2. 本文解决问题

* 机体猛地运动，造成图像模糊等退化的现象，如奔跑、跳跃
* 动态物体存在造成特征点稀疏、误匹配等
* 在某个传感器关掉后，系统仍能正常工作

1. 创新点

* 使用了相机、激光雷达和IMU建立了pipeline处理程序实现SLAM，并取得较好效果
* 动态的重配置，可以绕过不工作的模块，如VIO或激光雷达
* 该pipeline采用两级体素表示和多线程处理实现来加速扫描匹配，见P9-6.3
* 该pipeline可以在已有地图定位，即可以重利用地图。相机特征图和laser点云特征图都保存好了方便匹配使用，作场景识别。
* 该方法可以在动态环境、剧烈运动和大范围环境中使用

1. 本文方法：如下图所示，先利用利用IMU估计初值，使用VIO进行优化，最后使用3D激光雷达进一步细化,并不断通过反馈修正频率高、精度较低的模块。由最后的laser模块建立3D地图。VIO和laser都可以修正IMU的漂移和偏差。先使用相机、laser、IMU数据计算出位姿，再通过位姿将雷达测得的点投影到相机坐标系建立3D地图
2. 重要假设：漂移积累的频率低于传感器自身测量频率，即认为传感器在几个周期内是不会漂移的。利用高频传感器处理高频、快速的运动，利用低频传感器减少或消除漂移。如正常工作时，VIO模块4个IMU周期反馈准确值给IMU，减小IMU的漂移。
3. IMU prediction subsystem

* 将IMU加速度变换到相机坐标系，联系旋转带来的离心力，见式（4），式下有参考文献。所以将IMU和相机相距比较近可以减小离心力项影响。
* 偏置计算：将VIO或laser得到结果作为准确值返回给IMU模块用于估计偏置，认为返回值是准确的，再通过式（3）（4）计算偏置。将上一次偏置用于当前VIO？

1. VIO subsystem：从图4和本节内容可看出VIO不仅用了相机和IMU数据，还用了laser数据，相机和IMU数据用于motion估计时作为约束，laser在depth association时与相机特征点关联获得深度。与相机特征点关联后的laser点深度才能用于视觉跟踪。

* 利用相机三角化和laser共同填补点云，以laser点优先。怎么知道laser和相机指向同一个点，pose估计精度够吗？
* 基于关键帧的方法，选择原则是两帧之间相同特征或重合区域低于某个阈值
* Motion estimation：先不管深度，只计算旋转和位移，将视觉和IMU融合的方法待研究？
* Depth association
* 将相机特征点与laser点关联：将所有3Dlaser点投影到相机坐标系的单位圆（相当于归一化平面？怎么投影呢？方便和归一化的相机坐标系比较？），提取单位圆上与相机特征点最靠近的三个laser点，然后将这三个laser点再投射到笛卡尔（空间）坐标系，通过它们的距离判断关联的有效性。因为建立的是稠密地图，3D点很多，使用KD树（有文献）来加快搜索
* 通过三角化计算深度：对于通过laser点关联还得不到深度的点，使用三角化等方法恢复深度，文中有文献

1. Scan matching subsystem：通过scan matching（最小化scan特征与map的距离）进一步优化motion，文中有相关文献，优化过程中还考虑了VIO和IMU的约束

* 将VIO作为位姿基点，用IMU积分进行插值，完美利用了前者漂移少，后者频率高的优点，使1帧laser中有了200hz/5hz=40个位姿点。那么一帧中按40等分将不同等分测到的laser点通过位姿变换变换到1个等分上，这就是去畸变过程？
* 避免了重复选点，见式12和图8
* 截断（不考虑）laser一部分区域，相同大小物体在近的地方阶段的
* 通过特征值和特征向量判断点集的类型？原理是什么？
* 边和面的点的距离公式如式（14）（15）
* Motion estimation：最小化scan两帧之间3D点的距离的到位姿，同时考虑了IMU约束，采用了牛顿梯度下降的方法进行优化，P9有论文
* Map in voxel：two-level voxel，如图10，将所有3Dvoxel用于建立稠密地图，用边和面特征点用来实现scan matching，进一步优化位姿。
* 平行处理进行加速，p10，laser部分计算量较大，并行可以加快运算速度，给IMU和VIO足够的时间

1. On robustness：IMU被认为在所有时刻都能稳定工作，相机对突然光照变化敏感，影响的是照片质量，另外在光线过暗环境不能工作，在低纹理、剧烈运动（造成图像模糊）容易跟踪丢失。Laser在无结构（空旷）环境（指laser测量距离内）不工作。用特征值代替某个方向（特征向量）解的好坏，为什么？怎么判断其好坏？（看参考文献）IMU给定VIO初值，VIO给定laser模块初值。进行非线性优化时只优化好的方向，另外方向保持初值不变。
2. 实验
3. Test with single-axis scanners：将传感器

* Accuracy test
* Robustness test
* Aggressive motion test

1. Test with custom-built contour
2. Test of localization map merging
3. 重要文献：

* 双目论文：P3
* 单目论文：P3
* RGB-D论文：P3
* VIO（包括松耦合和紧耦合）：P3
* Laser恢复motion：：P3
* 2Dlaser和IMU融合：P3
* IMU补偿卷帘门：P3
* 使用其他传感器辅助laser（作者参考文献）：P3
* 作者之前相关论文：P3
* 针孔模型和标定：p4
* 相机和laser相对位姿、laser和IMU相对位姿标定：P4
* 参考MAP作估计：p4
* IMU偏置校正：P5
* VIO（作者之前论文）：P5
* VIO位姿估计：P6-5.2
* Scan matching subsystem（用到作者以前论文，还有常见scan matching论文）：P7
* KD树和并行计算：P9
* Robustness（作者之前论文）：p10

1. 启发
2. 本文程序是并行还是串行呢？三个模块串行，laser模块采用了并行加速？
3. 通过传感融合，解决不适合相机运动的低纹理、快速运动、光照突变和光线过暗的情况，不适合laser的无结构环境（空旷地），补偿IMU的漂移和偏差。
4. 三个模块并行运行吗？三个模块怎么进行反馈更新？新旧值如何替换？
5. 点云地图只需知道空间某点有没有物体，而不需知道有没有颜色，所以可以通过相机三角测量和laser得到的深度进行互补的方式建立稠密地图
6. 可否利用IMU进行插值对scan一帧扫描过程中的机体位姿进行修正？文中P7是不是这个意思？
7. 《VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator》总结
8. 内容摘要

本文概述香港科技大学VINS方法。针对仅视觉方法无法恢复尺度，利用IMU的整合测量可以显著改善运动跟踪, 提出基于优化的紧耦合的方法用于获得高度准确的位置信息，通过融合预积分IMU测量的视觉--惯性测距进行特征观察，取得很好效果。包含了初始化、闭环检测和地图重利用模块。

1. 总结及理解
2. VI的优点：

* 纯单目视觉不能观察尺度，但单目相机和IMU的结合可以观察尺度，横滚角和俯仰角。
* IMU测量值的积分可以弥补相机在光线变化、低纹理和运动模糊场景丢失的跟踪表现

1. 之前VI缺点：

* 严格的初始化，由于没有直接的距离测量，不能直接将视觉和IMU信息融合。

思考：深度相机可以直接和imu融合吗？

* 漂移（使用闭环检测、重定位he 全局优化解决）

1. VINS的创新点

* 可以开始于未知位置的鲁棒性好的初始化,可以获得尺度信息
* 紧耦合，VI外部参数校正和IMU偏差校正
* 在线重定位和4DOF全局图优化
* 姿态图可以重利用，可保存，加载和融合多个子地图

1. 本文方法的整体框架如图2，主要分为4个部分，measurement preprocessing、initialization、VIO、pose graph optimization。其中VIO和pose graph optimization分为2个线程。
2. measurement preprocessing
3. V预处理：提取和匹配两帧图像特征，使用极线约束RANSAC方法排除一些误匹配。对IMU数据进行预积分。

选取关键帧原则：

* 足够视差，为了避免纯旋转带来视差但不能三角化的情况，使用IMU进行补偿，补偿只影响帧选择，不加入旋转的计算
* 跟踪质量：跟踪到的特征点（2帧之间）

1. IMU预积分：具体公式见论文，因为两帧之间偏差在第二帧时刻才能算出，一开始用第一帧时刻偏差估算，但实际第二帧时刻计算出才是对的，若2个偏差变化很小，可用一阶泰勒近似，否则算完后需重新积分。
2. initialization：松耦合，视觉和IMU分别计算姿态，视觉计算出的位移是没尺度信息的，最后需要两者对齐恢复尺度。
3. vision-only：使用SfM（offline-先收集一些帧（滑窗）再一起计算）来恢复相机位姿和3D坐标点。
4. 滑窗内最新帧分别与其它所有帧稳定跟踪（跟踪特征点超过30个）和有足够视差（特征点在2帧距离超过20个像素）
5. 先通过文献[34]五点法2D-2D进行2帧三维重建（无尺度信息）
6. 再通过文献[35]pnp3D-2D重建初始化的其他帧
7. 最后使用[36]BA优化。
8. 以C0帧作为参考坐标系，按论文中公式进行转换
9. VI对齐
10. 校正IMU角速度偏差，以相机得到的旋转量作为参考。可以像霄卓一样先测个初值，给个好的初始偏差再进行优化。
11. 初始化偏差后，计算速度，重力向量和尺度
12. 进一步优化重力向量
13. 根据重力向量得到C0坐标系和真实世界坐标系（重力方向与z轴平行）的旋转量，将所有以C0坐标为参考的变量变换到世界坐标，即世界坐标系z轴与重力方向平行，原点与C0重合
14. VIO（包含了relocalization）：基于非线性优化的紧耦合方式，其使用滑动窗口。使用[38]ceres solver进行非线性优化，见公式14。Relocalization包括闭环检测和丢失后重定位
15. 总公式
16. IMU的残差：
17. 视觉测量的残差：使用通用观测模型
18. 边缘化：减少计算复杂度，schur[39]。边缘化的2种可能，次新帧为关键帧边缘化窗口中最老帧的测量，次新帧不是关键帧时，边缘化次新帧测量，建立与其上一帧IMU连接，注意最新帧就是当前帧，插入新帧，边缘化一个旧帧。边缘化主要原则是在窗口中保存空间分隔大的关键帧，这能保证三角化时有足够的视差，最大化加速度计在大激励下准确性。边缘化可能带来次优解，但可以接受。
19. Motion-only VIO：只优化固定数量IMU的位姿和速度，将相机和IMU相对位姿，特征点深度，偏差和老的IMU状态看成固定值。可以加快估计速度，适用于移动机器人，计算时间为原来1/10（50ms变成5ms）左右。
20. IMU帧率姿态估计：产生结果可用于无人机，IMU不断以VIO的结果作为起点进行估计。

Relocalization：利用了图像信息的丰富性来识别，再

1. Loop detection：文献29(DBOW2),40(brief)。在VIO基础上再增加500多个使用BRIEF描述的特征用于重定位和闭环检测，之后使用时间和几何一致性验证得到闭环候选帧。
2. Feature retrival:在当前滑窗和闭环帧寻找匹配点。利用2D-2D,3D-2D的变换使用RANSAC方法去除异常点。图10能直观的看出该方法的用处。
3. Tightly coupled relocalization：在式14的基础上增加闭环帧到当前滑窗的视觉约束，并将闭环帧的位姿作为常数。闭环帧数越多，结果越准确和光滑。
4. Global pose graph optimization：重定位优化完以后，优化了当前滑窗的变量，为了优化滑窗之前的位姿，需要进行全局优化。
5. four accumulated drift direction：在式9中可以计算当前帧相对世界坐标系的重力向量，那么可以结合当前计算的机体pose和重力向量相对[0,0,g]的变化求出roll和pitch，yaw取垂直于重力向量的平面，即欧拉角先绕世界坐标z轴旋转。
6. adding keyframe into the pose graph：每个关键帧作为一个顶点，各顶点包含两种边-序列边，直接从VIO得到；闭环边。
7. 4-DOF pose graph optimization：闭环边需要加huber核去除异常边，但序列边不用，因为VIO基本去除了所有异常边。VIO、重定位和闭环检测各线程分开跑，可以同时处理。
8. Pose graph merging：地图保存格式是一样的，当前地图与之前地图出现相同地方时，也能像闭环检测一样将两个地图关联起来。
9. Pose graph saving：保存边和顶点，还有关键帧的描述子，不保存原始图像。
10. Pose graph loading：
11. 直接法需要好的初始化，两帧之间重合区域要大一些，非直接法需要额外的提取和匹配特征，非直接法在工程中应用更广，其更成熟鲁棒性更好，直接法更容易建立稠密地图，其本身是像素级的。
12. 用离散的方法求积分（适合计算计算方式）：zero-order hold (Euler), first-order hold

(midpoint), and higher order (RK4)。本文使用Euler方法，见文献19、24.

1. 重要文献（VI方面）

* 作者之前的相关工作：6-8
* 松耦合：9-10
* 基于EKF紧耦合：11-13
* 基于图优化紧耦合：14-19
* 直接法VI：22
* IMU预积分参考文献：23，16（作者实验室，主要），19（biaes），24（biaes）。
* 初始化：17，25，26，27，28，18（基于ORBSALM），34（2D-2D），35(3D-2D),
* 跟踪方法涉及文献：31、32、33（RANSAC）
* 评价方式：42-44，RMSE
* 相机模型：鱼眼模型45
* 标定工具：toolkit46

1. 启发
2. 《A review of visual inertial odometry from filtering and optimisation perspectives》总结
3. 内容摘要

本文概述VIO的发展，文章发表于2015年，本文将介绍基于滤波和基于优化的方法。

1. 总结及理解
2. simultaneous localisation and mapping (SLAM) techniques and structure from motion (SFM)分别是robotic和computer vision research communities的研究主题。
3. filter-based approaches

IMU（内部传感器）提供角速度和加速度，可以提供内部的运动模型作为预测步得到状态的先验分布，相机（外部传感器）通过特征提取和匹配提供外部的观测模型（似然分布）作为更新步得到状态的后验分布。

1. IMU data driven dynamic model：运动模型，预测步用，因为IMU噪声和偏差的存在，随着时间的增长，累积误差会越来越大，因而需要观测模型限制误差的边界
2. Error state representation and updating：用四元数表示状态偏差
3. Visual measurement model and updating：测量模型。更新步用。建立测量模性的方法有松耦合和紧耦合方法[13]。
4. Optimization-based

为了获得实时性：使用多线程、使用关键帧和滑动窗口

1. Feature alignment：将各帧相同特征点进行对齐（最小化重投影误差）
2. Dense alignment：使用稠密方法或半稠密方法提取特征点，22-23使用了RGB-D相机，除了使用图像信息外，还可使用深度信息进行优化[24]。
3. Inertial measurement term：加入IMU项。分为松耦合[11]和紧耦合[31，19]两种方法，松耦合方法直接融合相机和IMU单独估计的结果，紧耦合将IMU加入优化过程。
4. Links

基于优化（核心gauss-Newton迭代方法思想）的方法可以看作通过不断迭代优化得到最优状态的maximum likelihood（ML）方法，基于滤波（核心KF思想）的可以看作maximum a posterior（MAP）方法，当先验未知时两者等同（数值分析）。基于优化的方法加上来自内部传感器或其它源的regularisation term（正则化项） 或prior term（先验项）与滤波方法相似。

1. Iterated EKF update：on-line方法，包含预测和更新步，在更新步进行了迭代更新。
2. Smoothing-based approaches：off-line方法[33]。可以使用BA方法也可以使用KF方法（可前向也可后向）。
3. Marginalisation to keyframes：为了减少时间和空间复杂性，要滤除一些冗余信息（帧），选择有用的帧[19,24]
4. Moving horizon estimation：将MAP方法分为2部分
5. State observability and parameter identifiability

分析状态的可观测性和参数的可识别性，有利于分析和减少状态估计的误差。State和具有random walk processes的时变参数（如IMU偏差）需进行可观测性分析（文献见原文），而如相机内参等时不变参数需进行可识别性分析。

VIO重要参数：

* 相机内参：焦距，中心点，镜头畸变，一般看作时不变
* IMU参数：角速度和加速度偏差，一般看作时变的
* 空间（spatial）参数：IMU和相机之间相对位姿，一般看作时不变
* 时间（temporal）参数：相机和IMU之间的延迟

1. 重要文献：

* 基于滤波文献：11（EKF完整描述）、19
* 四元数error描述，VINS也用了：12
* ICP：25、26、27（GPU）
* 相机内参在线标定：40
* IMU和camera之间位姿估计：41-43
* IMU和camera时间对齐：44、45（重要）

1. 启发
2. 《A Review of Visual-Inertial Simultaneous Localization and Mapping from Filtering-Based and Optimization-Based Perspectives》总结
3. 内容摘要

本文发表于2018年，分别介绍了基于滤波和基于优化（根据后端处理方式）的VIO方法。通过实验得出结论：优化方法方法有更好的精度和占用更少的内存，滤波方法占用更少的CPU。优化方法目前研究较多，紧耦合更多。

1. 总结及理解
2. P3的4种特征点
3. P4中VIORB提出的四个步骤
4. VIORB参考源码（非原文作者）：<https://github.com/jingpang/LearnVIORB>
5. 多线程存在不确定性，可以跑多次取中位数。
6. 重要文献

* 表1列出了近十年的VIO方法
* review：19-23
* 线面特征：84-88
* 光流法三种不同计算方法p4：89-91
* VIO数据集：102-106，附录A
* P12提到的DL、多传感器、主动SLAM、动态环境相关文献

1. 启发
2. 《Visual SLAM and Structure from Motion in Dynamic Environments: A Survey》总结
3. 内容摘要

本文发表于2018年，总结了动态环境一些方法，文中3个表和第6部分总结了各方法的优缺点。

1. 总结及理解
2. 下图很好总结了各种方法，定位和建图过程中，最理想的就是把确定自身的位姿，同时建立环境中静态点、动态点3D位置。下图总结了2种实现方式：A+B和C。而标准VSLAM假设环境是静态的，只根据静态点估计自身的位姿和环境中静态点的3D位置。注意objects是指环境中的目标，由多个特征点feature组成。
3. A. Robust Visual SLAM:图2  
   1. Motion Segmentation：将动态特征和静态特征分开。标准VSLAM中分割的方法是RANSAC，如果静态点占大部分，该方法是能很好工作的。Sampson距离也会用于排除异常点。  
   1. Background/Foreground Initialization：利用先验知识识别出了静态特征（background）和动态特征（foreground）。该方法分为了Background和Foreground Initialization2种方法，具体见原文。能否使用光流或深度学习的方法先把动态物体识别出来。  
   2. Geometric Constraints（注意是否在知道自身位姿的前提下？）：极线约束；flow vector bound(FVB)，可以处理退化的情况，如动态物体沿极线运动；多个3D点与匹配点像素坐标的连线交于一点；动态特征会使跟踪效果明显下降，使用两种方法计算，一种添加新特征一种不添加，从结果判断新特征是不是动态的；重投影误差；  
   3. Optical Flow：运动物体在图中成块出现，若运动物体块大于静态物体，会出现估计自身位姿出错。难计算退化的情况，如物体和相机以相同速度作平面运动。文献4用到了mahalanobis距离。  
   4. Ego-Motion Constraints：通过里程计、IMU得预先计算出自身位姿。  
   5. Deep Learning  
   2. Localization and 3D Reconstruction：这里只使用静态点。  
   1. Feature Based：短基线可以使用optical flow-based techniques(e.g., Kanade-Lucas-Tomashi (KLT) tracker [100])；长基线可以用SIFT等描述子。  
   2. Deep Learning：预训练不适合几何估计
4. B. Dynamic Object Segmentation and 3D Tracking  
   1. Dynamic Object Segmentation：将A.1得到的动态特征，划分为不同集合，该集合可能属于同一object。  
   1. Statistical Model Selection：用不同运动模型匹配，类似F、H初始化。  
   2. Subspace Clustering：将高维空间降到低维后再进行分类，类似PCA降维。  
   3. Geometry：极线约束等。  
   4. Deep Learning  
   2. 3D Tracking of Dynamic Objects：跟踪动态目标。  
   1. Trajectory Triangulation：前提是物体运动轨迹已知或满足某种形式  
   2. Particle Filter
5. C. Joint Motion Segmentation and Reconstruction  
   1. Factorization：不适用透视相机。  
   1. Multibody Structure from Motion (MBSfM)  
   2. Nonrigid Structure from Motion (NRSfM
6. 重要文献：表1、2、3和图2、4、7有总结

* 1point方法：137
* 鲁棒估计方法(可以处理异常点和误匹配)：RANSAC [37], PROSAC [22], MLESAC [158]

1. 启发

第18周汇报

——张溢炉

1. 《Improving Grid-based SLAM with Rao-Blackwellized Particle Filters by Adaptive Proposals and Selective Resampling》总结
2. 内容摘要

本文是激光slam算法gmapping相关的论文，其是一种改进的基于栅格地图的slam算法，使用了具有自适应自适应建议分布和选择性重采样的RBPF（Rao-Blackwellized Particle Filters)。

1. 总结及理解

* RBPF中每个粒子携带一个单独的环境地图。存在2个缺点：
* 复杂性，这是根据构建精确地图所需的粒子数量来衡量的。减少粒子数是主要的挑战。
* 重采样过程可能会消除正确的粒子。
* 文中提出解决上述问题的方法：
* 提出更准确的建议分布，将运动和测量都考虑了，提高了粒子滤波预测步骤的准确性，通过评估由扫描注册程序获得的与粒子相关的最可能位姿的可能性来计算的。在生成新粒子时，最后的读数将被考虑在内，从而可以根据一个更可靠(因此也更准确)的模型来估计系统的演化，这个模型比仅使用[5]和[8]中最后的里程计读数得到的模型更准确
* 自适应重采样策略，只允许在需要时执行重采样步骤，从而保持合理的粒子多样性。监测Neff控制的自适应重采样。
* 1

1. 启发
2. 《Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters》总结
3. 内容摘要

本文是激光slam算法gmapping相关的论文，其是一种改进的基于栅格地图的slam算法，也是对fastslam的改进，它的建议分布与fastslam 2.0相似。使用了具有自适应自适应建议分布和选择性重采样的RBPF（Rao-Blackwellized Particle Filters)。该方法可以减少获取栅格地图的粒子数量，是对上一篇论文的改进。相对之前的工作，本文方法粒子数降低了1个数量级。

1. 总结及理解

* 常见RBPF步骤

1. 采样：根据建议分布从上一时刻粒子集采样生成新的粒子集；
2. 计算重要性权重：计算每个粒子的重要性权重，其值为目标发布与建议分布的商；
3. 重采样：根据重要性权重对粒子集进行采样，权重大的粒子更可能保留下来。采样完后，使所有粒子的权重相等；
4. 地图估计：根据上面求得的轨迹和所有观测计算地图；

* RBPF中每个粒子携带一个单独的环境地图。存在2个缺点：
* 复杂性，这是根据构建精确地图所需的粒子数量来衡量的。减少粒子数是主要的挑战。
* 重采样过程可能会消除正确的粒子。
* 解决上述两个问题常见方法是：改进建议分布，使建议分布与目标分布更相似，如文献13-15，本文方法来源于文献13；有选择的重采样，保持粒子集的多样性。
* 文中提出解决上述问题的方法：
* 提出更准确的建议分布，将运动和测量都考虑了，提高了粒子滤波预测步骤的准确性，通过评估由扫描注册程序获得的与粒子相关的最可能位姿的可能性来计算的。在生成新粒子时，最后的读数将被考虑在内，从而可以根据一个更可靠(因此也更准确)的模型来估计系统的演化，这个模型比仅使用[5]和[8]中最后的里程计读数得到的模型更准确
* 自适应重采样策略，只允许在需要时执行重采样步骤，从而保持合理的粒子多样性。监测Neff控制的自适应重采样，可以减少粒子消失的风险。Neff越小，权重差别越大。
* 本文方法步骤：

1. 根据文献19的方法使用里程计数据做出位姿的初始估计；
2. 根据上一步位姿估计和上一帧扫描地图进行扫描匹配，得到最佳位姿估计。若扫描匹配失败，则使用运动模型估计位姿，直接跳到第5）步；
3. 在上一步最佳估计附近采样K个粒子的位姿，每个粒子的均值和方差由测量和运动联合的模型计算，其中归一化因子根据文献17求出
4. 根据上一步计算的每个粒子的均值和方差的高斯分布采样K个粒子的位姿，作为新的粒子集
5. 使用递归的方法更新权重，当Neff小于阈值时进行重采样。
6. 根据测量和位姿估计地图。

* 定位需要全局一致的地图，建图需要精确的定位。这两者是相互依赖的。
* 建议分布使用考虑了运动模型和测量模型更准确分布，有以下两个优点：提高了地图精度；减少了估计误差。
* 当测量模型精度明显高于运动模型，如激光雷达，测量模型的方差会比运动模型小很多，当以运动模型作为建议分布时，生成在测量模型均值附近的粒子会偏少，从而造成后验误差较大
* 当处于凌乱的场景或者里程计被强噪声影响时，滤波器可能出现过于自信的问题，这类问题可以通过多模式的扫描匹配完成，各个节点分别进行重复采样
* 扫描匹配失败的原因：测量信息少或者相邻两帧重叠部分较少
* 本文使用方法参考文献（IV）
* RBPF：文献2、8
* fastslam 1.0、fastslam 2.0：文献5、6
* 扫描匹配“vasco”：文献20（网站：<http://carmen.sourceforge.net>）、21
* 测量模型“beam-endpoint model”：文献22（《概率机器人》第6章）
* 运动模型：文献23（《概率机器人》第6章里程计模型），作者尝试过更精确的模型也未取得明显提升
* 栅格地图：文献12
* 文献19的方法使用里程计数据做出位姿的初始估计
* 文献[13]：改进的建议分布、获得递归的计算权值的公式
* 疑问
* 为什么可以用运动模型、运动和测量模型作为建议分布，而不能直接用测量模型作为建议分布

自己的想法：测量模型需要状态xt作为条件，运动模型能计算出xt的近似值

1. 启发
2. 《Bayesian Map Learning in Dynamic Environments》总结
3. 内容摘要

本文介绍了建立栅格地图的方法。比较了EM算法和贝叶斯推理的方法，EM算法容易陷入局部最优，从而导致机器人定位和建图失败，而贝叶斯方法保留了多个假设，鲁棒性更好。最后介绍了RBPF。

1. 总结及理解

* EM算法：

1. E步：给定地图，计算定位。
2. M步:给定地图，计算位姿。

* 动静态环境可以用概率pc表示，pc=0表示静态环境
* 航迹推算会因为执行器的噪声而存在滑动和漂移的问题
* 重要文献
* RB：文献4、10
* 栅格地图：文献12

1. 启发
2. 《Rao-Blackwellised Particle Filtering for Dynamic Bayesian Networks》总结
3. 内容摘要

本文证明了RBPF比标准的粒子滤波（PF）更精确，并在具有径向基函数网络的非平稳在线回归和机器人定位与地图构建这两个问题中得到了验证。

1. 总结及理解

* 本文符号与其他文献有所不同：y表示测量，z表示待估计变量，包括r和x
* RBPF步骤

1. SIR（Sequential importance sampling）：使用递归方法计算权重
2. 重采样：根据权重对粒子集进行采样，可以避免SIR的退化

* 重要文献：引用文献中画2个圈的
* MCMC步：使用马尔可夫转换核（Markov transition kernel）产生变种，保持粒子的多样性，参考画星星文献

1. 启发
2. 《A Survey of Simultaneous Localization and Mapping》总结
3. 内容摘要

本文介绍了激光SLAM、VSLAM、多传感器融合SLAM以及深度学习等多种类型SLAM算法，可以用于查找相关文献。

1. 总结及理解
2. II-激光slam

* 雷达：按结构可分为机械式雷达、混合固态激光雷达和固态激光雷达。文中介绍了雷达品牌。

1. III-VSLAM、VIO：相机品牌，有些相机是带IMU的
2. IV-激光视觉的融合
3. 激光SLAM可靠性高、鲁棒性好
4. 重要文献

* 综述：55（V）、56（V，tookit）、57（mono-V）、127（VIO）、218（语义）
* 激光slam：7（2Dlidar review）、
* IMU+lidar：46（解决走廊这种低纹理）、
* Dynamic：47（lidar）、
* VSLAM：71-72（VIORB）、115(RGB-D SLAM)
* VIO：126（同步VI数据）、143（避免相机和IMU之间标定）
* 语义slam：220、221、222
* VI标定(P7)：224（kalibr）、113（vins-fusion）、130（MSCKF）、225-226、227-226（end to end）
* Camera-depth(P7)：229
* Lidar-IMU(P7)：230-233（看文中介绍）
* Camera-lidar(P7)：234-240(看文中介绍)
* V-激光（p8）：249-266（看文中介绍），255（Vloam），257（与自己情况相似）

1. 启发

第17周汇报

——张溢炉

1. 《Motion and structure from motion in a piecewise planar environment》总结
2. 内容摘要

本文讲述了单应矩阵的求解方法

1. 总结及理解

* 当特征点位于同一个平面时，可以通过4对点求出相机位姿。
* 点共面满足如下条件

nTp1=d

其中n为平面的法向量（在第一帧相机坐标下），d为相机中心到平面的距离，p1为点在第一帧相机坐标系下的3D归一化坐标

* 由p2=Rp1+t1得p2=（R+tnT/d）p1，那么x2=K-1(R+tnT/d)Kx1，其中x为像素坐标，

单应矩阵为：H=K-1(R+tnT /d)K，因为尺度不确定性，其具有一个自由度

* 由于尺度不确定性，可令：p2=（dR+t1nT）p1，令A= dR+tnT
* 通过8个点组成8个非齐次方程，将其转为超定方程，通过SVD求解超定方程求出A
* 通过SVD方法分解，当A的3个奇异值不同时得到8组n、R、t解，当3个中有2个相同时得到4组解。3个相同时得到一组未定义的n的解和一组有定义的解。

1. 《Accurate Non-Iterative O(n) Solution to the PnP Problem》总结
2. 内容摘要

本文讲述了一种由4组3D-2D点的不需要迭代的求解方法；EPnP，其时间复杂度为O（n）。

1. 总结及理解

* 3D（世界坐标P）-2D求解方法：首先，利用3D-2D点使用PnP方法求出2D点在当前帧坐标系的3D坐标p2（有尺度信息）；然后使用3D-3D的方法求出位姿，此时P=Twc2p2
* 很多PnP方法重在求解点在相机坐标系下的深度。而本文方法使用4个不在同一平面的control point（不是特征点的坐标）的加权和来求解。
* EPnP步骤
* 给出n个3D路标点，取四个控制点c，将n个点分别用控制点表示，由以下线性方程求出每个点的alpha，其中cj取个坐标的质心，其余三个控制点使用PCA求出，那么可得含4个未知数的4个方程组，可求出alpha
* 方程两边乘齐次变换矩阵T转到相机坐标系，方程仍然满足，则有
* 那么由相机的投影规则有，其中每个点的alpha值都可由第一步方程求出
* 可得如下方程，其中M为2n\*12，x为12维向量

Mx=0

* 将M进行SVD分解，由最小的N个0奇异值（其不会严格为0，取最小的N个奇异值）对应右特征向量矩阵的向量求得x，理论上得6个点M刚满秩，由于尺度不确定性，会超定一个方程。实际中，常用15个点。但代码中使用ransac方法只用了4个点
* 当取N个向量时，求解方法不尽相同,都是利用相同点在不同坐标系的距离相等，具体看论文。N<=3时求解较简单，N>=3时需使用relinerization technique[16]。N越大，结果的精度会越高，但实际上，常将所有N（较小的N）的不同组合求出来，然后选取重投影误差最小的解。代码中先估值betas，在使用高斯-牛顿迭代求解
* 当所有点位于同一个平面时，只取3个control point即可，其他方法类推。

1. 《Least-Squares Fitting of Two 3-D Point Sets》总结
2. 内容摘要

本文讲述在已知3D-3D点的情况下求解R、t的最小二乘解方法，只需计算一次，用于orb-slam2闭环检测中

1. 总结及理解

* 步骤如下，具体证明看原文
* 计算各点的去重心坐标
* 计算H
* 使用SVD分解H，得到U、V
* 取R=X=VUt，计算det（R）=1,若为-1，第三行取反
* 再求t
* 1

1. 启发
2. 《Real-Time Loop Closure in 2D LIDAR SLAM》总结
3. 内容摘要

讲述谷歌cartographer算法

1. 总结及理解

* 匹配算法：
* scan-to-scan matching：会累计误差，最终误差会很大
* scan-to-map matching：减少了误差的累计（本篇论文使用方法）
* pixel-accurate scan matching：最后有说这个算法在后台用于将scan点集和最近的submap进行匹配，生成loop closing的约束条件。
* 优化算法
* 粒子滤波方法（particle filter）
* 基于图优化（graph-based）
* 1

1. 启发
2. 《A new robot navigation algorithm based on a double-layer ant algorithm and trajectory optimization》总结
3. 内容摘要

本文介绍了double-layer ant colony algorithm（DL-ACO）算法，它使用了2层ACO，2层ACO单独、依次运行。首先，parallel elite ant colony optimization (PEACO)方法产生初始的无碰撞的路线，然后使用turning point optimization algorithm (TPOA)方法优化路径的长度、光滑性和安全性。最后，使用piecewise B-spline方法提高光滑性，更适合履带车通过。本文最后，通过实验证明方法的有效性和一致性。

1. 总结及理解

* 全局路径规划要解决的问题是：找到起点和终点之间无碰撞（动态和静态障碍物）和满足其他约束（距离、运动学等）的路径
* 评估路径的指标：长度、光滑性、安全性
* 本文主要贡献
* 增强函数（式1）和双层结构扩大了搜索范围，避免了过早收敛
* 路径优化方法优化了长度和光滑性
* piecewise B-spline在不影响安全性的情况下，优化路径的转角
* 本文方法相对其他方法能产生更短、更光滑的路径，此外，双层结构也增强了算法的稳定性
* 本文算法的缺点：本文算法需要精度高的地图，从而限制了它的应用场景。此外，本文算法只适用于静态场景。
* 传统的方法artificial potential field [4]和cell decomposition [5]被广泛应用，易于执行，但是计算量大，容易在复杂问题中掉入陷阱
* 与传统方法相比，启发式算法在路径规划中更有效。包括neural network [7], fuzzy logic

technique [8]，genetic algorithm (GA) [9], particle swarm optimization (PSO) [10]

ant colony algorithm (ACO) [11]

* PSO和GA方法由于初始值是随机选择的，相同情况下生成的结果具有不确定性。这会影响路径规划器的鲁棒性
* ACO被认为最有效、鲁棒性好的方法，在不同的场景中被频繁使用[12]-[15]。下面是ACO方法的拓展方法
* 文献[16]提出了三维的方法：unmanned combat aerial vehicle (UCAV)
* 为了在动态环境中使用，文献[17]在路径评估中增加了fuzzy cost function
* 文献[18]提出了基于动态路径重计算和改进的侦察（scout）蚁算法，可以完成机器人在未知环境的导航
* 但ACO存在收敛速度低、过早收敛的问题，此外，由于单步移动最大距离的受限造成的路线曲折，从而影响路线的长度和光滑性。下面是几种为了克服上述缺陷而设计的算法
* 文献[19]提出了hybridized ACO,它将局部搜索与已存在的ACO相结合使计算量减小
* 文献[20]提出了adaptive polymorphic ant colony algorithm（APACA），它采用了自适应的状态转移策略和自适应的信息更新策略
* 文献[21]提出了两层ACO，其将启发式算法分为了预处理层和路径规划层，这样可以避免局部最优。之后这种结构被广泛采用[22]-[25]，其可以扩大搜索范围、加快收敛速度
* 路径不平滑会加长运行时间，减少电机的寿命。平滑的路径能减少能力的消耗[26]，广泛应用的平滑方法有B-spline，如文献[31]、[32]。由于B-spline方法在某些场合会影响路径安全性，故本文提出piecewise B-spline的方法，它仅在拐角处作平滑处理。
* 算法步骤概述，详细公式见原文
* PEACO：产生初始路径

1. 计算增强因子：从式1可看出，迭代次数t越小，q越小
2. PEACO处理：
3. 初始化：将蚁群分为两个亚种群，初始它们参数，两个亚种群独立运行
4. 选择：没有到达终点的蚂蚁根据转移概率选择下一步
5. 信息素更新：当所有蚂蚁到达终点后，更新地图的信息素，以便下一次迭代使用
6. 达到最大迭代次数后，选择两个亚种群中最好的路径提供给下一层使用

* TPOA：通过移除不必要的转角等算法优化路径的长度、光滑性和安全性
* piece B-spline：使转角处变光滑
* 文献[35][36]介绍了运动控制规则
* 本文算法的运行时间和路径质量都与地图的分辨率正相关，在实际使用时要均衡两者

第16周汇报

——张溢炉

1. 《Fast Relocalisation and Loop Closing in Keyframe-Based SLAM》总结
2. 内容摘要

本文介绍了ORB-SLAM算法中重定位和闭环检测的方法

1. 总结及理解

* 提取500个角点时使用non-maximum suppression会改善角点分布，但是会降低匹配表现（特征点少了），故提取1000个角点
* 自动初始化时是2D-2D问题，重定位时3D-2D问题，闭环检测是3D-3D问题
* 在做一致性检测时，之所以选择3个方向，是因为在不同深度旋转的角度不一样
* 闭环检测：认为loop帧数据是比较准确的
* 选取候选帧
* 计算当前帧和闭环帧之间的相似变换Sk,l，算的过程中利用RANSAC作了几何检验。首先计算匹配点；利用无三维坐标匹配点同一张图周围最近3个有3D维坐标（两帧图都是用loop帧的路标点）的点计算该匹配点3D坐标（相机坐标系下），丢弃距离超过10个像素的点；对两帧进行上述操作后得到3D-3D点，然后利用文献[26]方法计算相似变换矩阵；然后验证，匹配点重投影小于6个像素表示是内点，如果在RANSAC前70次迭代中找到支持40%匹配点为内点的相似变换，则表示成功；最后利用所有内点在计算相似变换Sk,l
* 优化pose，因为BA有huber函数，漂移太大很难收敛。具体方法见文献[25]。detaSij是通过相似变换算出，闭环内两两帧之间变换，SjwSiw是由pose计算得出，r为residual error，理想状态等于1，这时两种方法恢复的值相同。目的是不断迭代优化pose，然后重复计算相似变换，使目标函数接近于0。协方差矩阵/\是指定的，不为单位矩阵时，表示不同自由度之间相互影响，但其常是单位矩阵。优化方法：g2o+LM+CHOLMOD



优化完pose后，利用如下公式校准每个路标点xj，使用Ti作为参考，接近于loop帧的位姿

* 路标点融合，如在loop看到的点，在当前帧又看到了，就会出现同一个物理点，在两帧中计算出了两个不同的3D坐标。将同一物理点（匹配点）在loop帧上世界坐标和当前帧该点的世界坐标一起投影到当前帧，若相距低于20像素，擦除当前帧的世界坐标。然后按同一的方法，融合loop周围关键帧与当前帧共视路标点，直到没有点可以擦除停止。在循环上诉操作，融合loop帧及其相邻帧与当前帧之前帧，直到没有点可以擦除。在点被擦除后就需要更新graph的边，包括帧之间以及帧与路标点的边
* 校正未参与闭环检测过程的新帧，在闭环检测当前帧之后前端又处理了一些新的帧。只需校正最新帧与闭环检测当前帧即可，其他帧会很快被局部优化。校正公式如下，这里只需利用尺寸校正的尺寸sk，l，注意校正前后标志



* 重定位：
* 为了更快的选出候选帧，不像闭环检测那样设置最小阈值alpha和一致性检测。在后面的几何检测以及在重定位之后跟踪超过20帧才会插入关键帧的设置，可以增强其鲁棒性。
* 使用闭环检测相似的方法恢复一些3D点，然后利用3D-2D+RANSAC计算位姿，如果在RANSAC前178次迭代中找到支持40%匹配点(至少20个点)为内点，则将该帧作为参考帧，若接下来超过20帧跟踪成功，重定位才算真正成功

1. 《Scale Drift-Aware Large Scale Monocular SLAM》总结
2. 内容摘要

本文主要讲述使用7dof优化的方法，没有说到两帧之间相似变换的求解方法，具体方法见下一篇。

1. 总结及理解

* 单目相机有7个自由度会漂移，因为它是通过累加的方法计算尺寸，而双目、深度相机只有6个自由度，与单目相似，当前变换是不断利用上一次位姿获得的。这也是靠前的位姿和尺寸会更准确一些的原因
* sim（3）的李代数表示形式如下

1. 《Closed-form solution of absolute orientation using unit quaternions》总结
2. 内容摘要

本文讲述一种计算已知一些点在自己坐标系下三维坐标的两帧图像之间的相似变换方法，该方法不用迭代，只需计算一次，使用于3个点及以上的情况。其中，将旋转用四元数表示。

1. 总结及理解

* 求解两坐标系之间相似变换方法分类：
* 单目具有7个自由度，那么取三个点就能找到9个约束，去掉其中两个约束便能求出解
* 迭代的方法：最小二乘方法
* close-form solution：如本文方法，给出匹配点在两个坐标系的坐标，其不需要迭代，只需一次计算，另外，其并不需要很好的初始值。
* 计算平移量：首先得到很多左右两个坐标系的匹配点。如下公式理解：假如r为向量，想象相机的带着一个坐标系在运动，先旋转，将两个坐标系方向变成一样，再变换尺寸，因为远近不一样，物理不变的长度，在近的坐标系变长，当然因为视角不同，物体形状还会畸变，这里先不考虑，故需要放缩一下；最后平移，r0是在右坐标系下的平移量。
* 具体方法
* 有选择的丢弃一些约束，使用已知在左右两个坐标系坐标的不在一条直线三个匹配点计算旋转。以其中2点作为x轴，2个点中一个作为原点，以3个点平面作为xy平面，将各轴化为单位向量。得到如下图坐标系，左右坐标系进行相同操作，这样就得到三个点在左右坐标系下的坐标系。因为原点相同，所以丢弃了平移，将各轴化为单位向量解决了尺寸问题，这样两坐标系只存在旋转，得到如下旋转公式，先转换为世界坐标系，再转换到右坐标系



* 本文方法：给出n个匹配点在左右两个坐标系的3D坐标（已知），最小化重投影误差，如下公式：



* 通过中心化可化为如下公式，第二项为0，第一项与s、R相关，第三项与r0相关
* 令第三项为0，求得平移
* 求尺度s，第一项可化为不对称和对称两种情况：
* 不对称情况：两个坐标系测量精度差别很大时，将第一项进行恒等变换有：



令第一项为0有，求得尺度：



* 对称情况：两个坐标系测量精度差别不大时，将第一项进行恒等变换有



令第一项为0，这里s与R无关，求得尺度

* 对于以上两种情况，D越大越好，无论D多大约束的存在最后结果肯定大于0，D越大值越小，D公式如下
* 然后使用单位四元数可将其化为



* A3证明欲求q使qTNq最大，只需求得N的最大特征值对应的特征向量作为q即可，从而求得旋转q（四元数）
* 然后使用上面s公式求出s
* 最后使用上面r0公式求出r0
* 当匹配点共面时存在两种特殊的旋转：旋转前后两坐标系交于一条直线和旋转轴为面的法向量。具体计算方法见论文

1. 《g2o: A General Framework for Graph Optimization》总结
2. 内容摘要

本文讲述g2o的原理

1. 总结及理解

* 整个框架如下图，g2o采用迭代方法操作，图中灰色部分error function和田操作（累加方式）是需外部指定的。另外，为了更高效，可以指定J；为了设计合适的求解方法，可以自己添加linear solver。在求解前可以先使用plain、schur方法消元。



* 本文实验使用了CHOLMOD（cholesky）、PCG来求解。前者计算速度快，但不能利用参数的块结构。后者迭代次数多，计算速度慢，但能利用参数块结构。
* Node是状态变量（姿态、3D坐标）,边是约束，一般约束求最小值，最小值一般为0，如最小重投影误差，J矩阵是约束对状态变量求导。
* 重要文献：
* LM：18

1. 《Double Window Optimisation for Constant Time Visual SLAM》总结
2. 内容摘要

本文采用inner和outer窗口用于用于局部地图优化，利用了pose-point（Ti）和pose-pose（Ti,j）两种约束来优化。该方法能自动建立合适的graph，动态选择并优化inner和和outer窗口

1. 总结及理解

* Full BA太花时间，使用局部地图来优化能保持较好的实时性，包括以下方法
* Active windows：选取全部关键帧的一部分，通常窗口边界的关键帧是固定的，那些与它有共视点的帧会被包含。这种方法比较适合大范围、少闭环的场景，但不适合小范围、闭环比较多的情况，可能会因为包括的关键帧太多而很难收敛
* Relative representation：比较经典的是relative bundle adjustment（RBA），它优化的是两帧之间的相对位置。此方法适合少闭环情况，在闭环比较多的情况精度会下降。
* Pose-pose reduction：相对BA，其只优化pose，不优化路标点，路标点不变。
* Double window：以关键帧Vref作为参考帧，与其及其邻居相连权重最大的前M1帧被取为inner窗口W1，接下来M2帧被取为outer窗口W2，一般M1<<M2。W1帧使用BA优化，W2中一些帧也使用BA优化，W2的所有帧使用pose optimization。两者不需迭代进行，共同优化以下函数即可，优化完后以Vref为参考恢复计算了相对位姿的帧
* 闭环检测：用另一坐标系求出的路标点（两者的匹配点）投影到自己坐标系求出相似变换或齐次坐标，方便后面求残差



* 小闭环检测，首先文中定义了当前帧Vi的两级邻居N1、N2，小闭环候选帧是在N2中找到的，将Vi求出的的路标点投影到Vref求出位姿Tloop，然后删除一些重复的点。最后最小化下面目标函数优化



* 大闭环优化：和之前方法相似

第15周汇报

——张溢炉

1. 《ORB: an effcient alternative to SIFT or SURF》总结
2. 内容摘要

本文提出了ORB (Oriented FAST and Rotated BRIEF)特征，有效结合了FAST特征点和BRIEF描述子，并给FSAT加上了方向，从而使BRIEF具有旋转不变性

1. 总结及理解

* 对FAST的改进版：oFAST（Orientation FAST）,其主要有以下几点改进
* 为了解决FAST错误检测到线的问题。本文采取的方法是：先调低阈值获得大于N个角点，再使用Harris corer measure [11]中的方法对角点进行排序取前N个角点
* 为了使FAST获得尺度不变性，本文采取的方法是：在图像金字塔的每层提出角点
* 为了给定FAST角点的方向，本文使用intensit centrid（IC） [22]提出的方法，并将其与BIN和MAX的梯度方法在噪声干扰的的情况下作了比较（结果如下图），证明IC方法具有更好的效果
* 对BRIEF的改进版：rBRIEF，其主要有两个目的，具体步骤在论文有注释

1. 使每位描述子具有高方差，从而使不同关键点描述子之间具有更好的区分性
2. 使描述子各位之间相关性低，从而使各位之间更加独立，下图是通过相关性选择前后（左右）的点对，比较直观的感觉就是左右有较多接近平行且靠得比较近的点

* 从实验结果来看，ORB特征具有较好的旋转不变性。但是使用图像金字塔的方法并没有使其具有较好的尺度不变性，还需改善

1. 《Bags of Binary Words for Fast Place Recognition in Image Sequences》总结
2. 内容摘要

本文提出了一种从一串经过的图片中识别场景（Place Recognition）的新方法，其以FAST+BRIEF为特征，处理速度可达到22ms/帧，比之前的方法快了一个数量级，而且它能保证检测到的都是正确的

1. 总结及理解

* Place Recognition对回环检测和重定位非常重要
* 存储图像描述子的树如下图，其中words由所有描述子的特征向量组成，direct index用于由图片查找其对应特征，inverse index用于由特征查找出现过的图片。先使用训练集图片获得足够大的树，然后再在运行更新direct index和inverse index。假如有M个特征（words），先使用K-means++将其聚成k类，然后再根据设定树的深度Lw，继续聚类。这样在查询特征时可以将时间复杂度变为log级。



* 在建树过程中使用TF-IDF（term frequency-inverse document frequency）来表示特征的重要性: TF表示某特征在某张图像中频率，越大越好，IDF表示某特征在全部图像特征中频率，越小越好，所以用倒数
* 闭环检测的步骤如下，结合文后实验能更好理解各步的作用
* dataset query：通过查询KD树得到当前图片的M维表示向量，通过如下公式计算归一化的相似度分数，设定最小阈值得到候选的图片



* Match Grouping：由于时间上靠近的图片非常相似，可以将相似的图片分成不同组，得到不同组的相似分数H，公式如下，取最高分的一组用于下一步：



* Temporal Consistency（连续性检测）：计算之前query image是否与之前各组<vt−Δ t, VT1 >, . . . , <vt−kΔ t, VTk >，如果有k组通过测试，就取当前图片与T’组得分<vt,vt’>最高的作为候选帧进入下一步
* Efficient Geometrical Consistency：通过两帧之间的匹配点计算出基础矩阵F用于最后验证，保证至少有 12对匹配点
* 实验过程中发现FAST+BRIEF匹配的特征都是中远距离的，中远距离在相机移动时尺度变化较小，间接说明了其尺度不变性较差
* 以下是实验的设定参数，在调参时可以作为参考



1. 《ORB-SLAM: A Versatile and Accurate Monocular SLAM System》总结
2. 内容摘要

本文介绍使用ORB特征的单目SLAM算法：ORB-SLAM。包含了自动初始化、追踪、建图、重定位、回环检测等多个环节，可以在一个环境中长时间运行，取得了前所未有的表现。

1. 总结及理解

* parallel tracking and mapping (PTAM)的缺点：
* 缺少闭环检测，不能减小累积漂移误差和增强全局一致性
* 不能充分处理闭塞的情况
* 重定位时对视角变化的不变性低
* 地图初始化时需要人为干预
* ORB-SLAM有以下几个优点
* 整个系统使用同一种特征：ORB特征，使其更加高效、简单和可靠
* 可以在大的环境中实时运行
* 在essential graph上实现实时闭环检测，优化位姿
* 实时重定位，具有很好的视角和光照不变性
* 健壮的自动初始化
* 剔除冗余的帧和地图点，这也是使其能在环境中长期运行的原因
* 为了实现实时性，ORB-SLAM将整个程序分为tracking、local mapping、loop closing3个线程，整体框架如下图。文中第4-第7部分详细介绍了自动初始化、tracking、local mapping、loop closing的过程和细节，文献11中更详细介绍了重定位和闭环检测，需对照代码更进一步理解。
* 各线程的作用如下
* Tracking：根据摄像头传来的帧进行实时定位和决定何时插入关键帧。关键帧是摄像头所有帧中比较重要的一些帧，选择条件见V-E
* Local mapping：处理新的关键帧及优化局部地图来减小累积误差
* Loop closing：检测回环，进行full BA减小累积误差，建立全局一致地图
* 文中三幅图（graph）的关系：spanning treeessential graphcovisibility graph
* 如果获取词袋使用的图足够多，那么可以将其用到其他场景的应用

1. 《ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo,and RGB-D Cameras》总结
2. 内容摘要

ORB-SLAM2是在ORB-SLAM的基础上进行改进，新增了将双目摄像头和RGB-D摄像头作为传感器时的预处理程序，同时实现了地图的重利用。

1. 总结及理解

* 下图为ORB-SLAM2的整体框架图，与ORB-SLAM相比，tracking和loop closing两部分不同。Tracking一开始需对三种不同相机做不同预处理，另外开启时会进行重定位重用建好的地图，识别自己的位置。Loop closing将sim3换成了se3，当进行full BA时，临时开启第四个线程
* 如下是双目和深度摄像头预处理过程，预处理后其他代码共用，使代码整个框架更清晰，便于移植，可读性高



* 双目和深度摄像头解决了单目尺度不确定性，可以从单帧图像获得深度，初始化更简单。也解决了单目旋转的问题
* 近特征点（小于基线的40倍）能很好的估计相机平移、旋转及特征点深度，远特征点只能很好的估计旋转，估计平移和特征点深度不准。这一点被用在了tracking插入关键帧的条件，有利于在视野开阔的室外场景下运行
* 程序中共进行了三次BA优化：tracking线程motion-only BA只优化位姿；local mapping线程local BA优化相机位姿和3D点；loop closing线程full BA优化闭环内所有位姿和3D点
* 下周阅读文献：
* Bundle adjustment a modern synthesis
* closed-form solution of absolute orientation using unit quaternions
* Double window optimisation for constant time visual SLAM
* Fast Relocalisation and Loop Closing in Keyframe-Based SLAM
* g2o: A General Framework for Graph Optimization

第14周汇报

——张溢炉

1. 《Distinctive Image Features from Scale-Invariant Keypoints》总结
2. 内容摘要

本文介绍了经典Scale Invariant Feature Transform (SIFT)特征。它具有尺度不变性（scale invariant）和旋转（rotation invariant）不变性，在仿射畸变（affine distortion）、 3D视点变化（change in 3D viewpoint）, 噪声增加（addition of noise）、 光照变化（change in illumination）的情况下匹配都具有较好的鲁棒性。另外，本文主要将SIFT特征用于物体识别，自己主要关注SIFT特征部分

1. 总结及理解

* SIFT能从一张500\*500的图像中提取2000个左右特征，这对物体识别比较有帮助，在混乱背景下识别一个小物体至少需要3个特征才能比较可靠的识别
* Harris角点是经典的关键点，但是它对尺寸变化敏感
* SIFT特征提取主要分为四个步骤

1. 尺度空间的极值点检测（Scale-space extrema detection）：得到的极值点为,包含位置和尺度信息。其中（x，y）是极值点的坐标，为极值点尺度，对应高斯核而得方差。计算步骤如下

* 利用高斯核、图像金字塔和difference-of-Gaussian（DOG）得到如下图所示高斯金字塔，左边为不同分辨率下通过不同方差高斯核得到的图像，右边为经DOG操作后的图像。具体计算公式见论文：如下图中，s=2，k=
* 取每组octave右边的中间两层来获得极值点，获取方法如下图，将带X的点与周围26个点比较，若该点大于或小于所有点，将其作为极值点

1. 关键点定位（Keypoint localization）：通过子像元插值和删除边缘效应对第一步得到的极值点进行筛选，最后获得关键点。子像元插值和删除边缘效应如下

* 子像元插值：利用泰勒公式得到连续情况下的D(xˆ)，当其低于某个值是将该特征点丢弃，文中在像素值归一化的情况下取0.03
* 删除边缘效应：为了得到稳定的特征点，只是删除DoG响应值低的点是不够的。由于DoG对图像中的边缘有比较强的响应值，而一旦特征点落在图像的边缘上，这些点是不稳定的。一方面图像边缘上的点是很难定位的，具有定位歧义性；另一方面这样的点很容易受到噪声的干扰而变得不稳定。一个平坦的DoG响应峰值往往在横跨边缘的地方有较大的主曲率，而在垂直边缘的方向有较小的主曲率。文中计算出hessian矩阵H后，通过如下公式筛选，取

1. 分配方向（Orientation assignment）：计算关键点局部梯度方向，将其作为特征点的方向
2. 计算描述子（Keypoint descriptor）：通过关键点周围区域梯度计算关键点描述子。如下图，将特征点周围8\*8的区域的梯度换算成每个方块包含8个方向矢量的2\*2描述子，其中包括2\*2\*8个描述子。文中取的是4\*4\*8个描述子。将描述子的计算与关键点方向结合起来便能达到旋转不变的目的



* DOG计算结果和使用拉普拉斯算子计算结果相似，即求出的是二阶导数
* SIFT特征的一个缺点是计算量太大，不适合用于对实时SLAM
* 待解决问题
* 将关键点与周围26点比较得到的极值点是角点和边缘点吗？其中数学含义是怎么样的？
* 描述子具体的计算过程是怎么的？

1. 《SURF: Speeded Up Robust Features》总结
2. 内容摘要

本文概述了具有尺度和旋转不变性的SURF (Speeded Up Robust Features)特征，相比SIFT具有更好的repeatability（可重复性，在不同视角的图像中被提取出来）, distinctiveness（区分度，区分不同特征点）,robustness（鲁棒性，在光照变化、视角变化、噪声等干扰下的可靠性高）。此外，它 计算量更小，实时性更高，这得益于其积分图像。

在一些不需要考虑旋转不变性的场合（如平面机器人导航，其只沿垂直轴旋转）可以使用upright SURF(U-SURF)，U-SURF不具备旋转不变性，不仅计算量更小，而且区分性更好

1. 总结及理解

* SIFT是经典的特征提取方法，但是计算量大，SURF是其改进型，计算量小一些，ORB在两者的基础上的改进型，计算更有效
* 特征提取和匹配的三个主要步骤
* 提取关键点，如角点、斑点和T形点。其需具备较好的repeatability
* 计算描述子，根据关键点周围的像素点计算其描述子，需具备较好的distinctiveness、robustness
* 匹配，根据描述子之间相似程度进行匹配，如计算两者表示向量的马氏距离、欧氏距离或汉明距离
* 本文第二部分对关键点和描述子的做了很好的回顾，有需要可以阅读其中相关论文加深了解
* SURF使用方形滤波器计算近似hessian矩阵来检测关键点，根据关键点周围Haar-wavelet响应的分布来计算描述子，两者都得益于积分图像来加速计算。描述子只使用了64D，这也加速了描述子计算和匹配时间。另外，作者还展现了一种新的索引方法，不仅加快了匹配速度，也增强了描述子的鲁棒性。特征提取也可分为以下四步
* 特征点检测：SURF使用Hessian矩阵来检测特征点，该矩阵是x,y方向的二阶导数矩阵，可测量一个函数的局部曲率，其行列式值代表像素点周围的变化量，特征点需取行列式值的极值点。用方型滤波器取代SIFT中的高斯滤波器，利用积分图（计算位于滤波器方型的四个角落值）大幅提高运算速度。
* 特征点定位：与SIFT类似，通过特征点邻近信息插补来定位特征点。
* 分配方向：通过计算特征点周围像素点x,y方向的哈尔小波变换，并将x,y方向的变换值在xy平面某一角度区间内相加组成一个向量，在所有的向量当中最长的(即x、y分量最大的)即为此特征点的方向。
* 特征描述子：选定了特征点的方向后，其周围像素点需要以此方向为基准来建立描述子。此时以5\*5个像素点为一个子区域，取特征点周围20\*20个像素点的范围共16个子区域，计算子区域内的x、y方向(此时以平行特征点方向为x、垂直特征点方向为y)的哈尔小波转换总和Σdx、ΣdyΣdx、Σdy与其向量长度总和Σ|dx|、Σ|dy|Σ|dx|、Σ|dy|共四个量值，共可产生一个64维的描述子。
* 待解决问题
* 理解hessian矩阵和特征点之间联系
* 理解哈尔变换

1. 《Machine learning for high-speed corner detection》总结
2. 内容摘要

本文介绍了一种使用机器学习提取FAST（Features from Accelerated Segment Test）角点的方法

1. 总结及理解

* 本文第一部分系统介绍角点检测的发展，有需要可以阅读其中相关论文加深了解
* 快速提取角点的方法FAST：首先取如下图所示中心点p周围的16个点，若有连续n个点像素值大于或小于点p，且都超过阈值t，则认为p为角点。当取n=12时，称其为FAST-12。相对于n<12的情况，FAST-12有一种快速判断不是角点的方法，先判断1、9点中是否有一个大于或小于p且超过阈值t，若没有则p点不是角点；若有，则继续计算5、13，判断四个点（1，5，9，12）中是否有三个点满足阈值要求，若不满足，则不是角点；若满足。则继续判断是否有连续12个点满足条件。
* FAST-12存在下面四个问题

1. 当n<12时获得的候选点比较多
2. 检测出来的角点不是最优的，因为它的效果取决于要解决的问题和角点的分布情况
3. 对于角点分析的结果被丢弃了
4. 检测到的很多角点都是连在一起的

* 本文提出利用机器学习的方法解决FAST-12的前三个问题，但机器学习有一个问题，训练集不可能包含所有样本，故预测时需借助模型的泛化能力，那么结果有一定的置信度
* 首先根据以下公式将每个中心点p的周围16个点分成3类，组成一个关于中心点p的16维的特征向量，并将每个中心点用布尔型Kp标注为是否是角点
* 然后使用这些数据训练ID3决策树
* 最后使用训练好的决策树来预测角点
* 另外，本文提出使用极大值抑制（Non-Maximal Suppression，NMS）的方法解决FAST-12的第4个问题。方法如下

1. 对所有检测到的角点构建一个打分函数（如下式）。就是像素点与周围16个像素点差值的绝对值之和
2. 考虑两个相邻的角点，并比较它们的分值。
3. 分值较低的角点将会被删除。

* 导数值易受噪声干扰
* 简单介绍了评价指标consistency of corner numbers(CCN)、ACU
* 使用机器学习方法提取FAST角点的方法具有以下优缺点
* 优点

1. 提出速度快
2. 对于变化大的图片和不同种类特征具有较高的可重复性

* 缺点

1. 抗干扰能力不太好
2. 可能对一个像素宽的直线作出响应
3. 依赖于阈值的选择

* 待解决问题
* 阅读文中相关论文理解角点与H矩阵及其特征值之间的联系？
* ID3的训练方法？

1. 《Faster and Better: A Machine Learning Approach to Corner Detection》总结
2. 内容摘要

本文是上一篇文章方法的改进版，它更快、更好

1. 总结及理解

* 较上一篇本文对提取角点的方法做了更好更全的分类，可以选择阅读
* 本文提出了FAST-ER（Enhanced Repeatability），它在原来算法里提高特征点检测的重复度，重复意味着第一张图片内的检测的点，也可以在第二张图片上检测出来。由于一些形变较大因素造成的形变，很难通过简单且固定的模板将所有的角点检测出来，而原来的FAST算法其决策树并不能最优的实现区分角点（实现最优的重复率）。FAST-ER就是针对这样的问题而提出的，其主要是通过模拟退火（也有通过最速下降法的）优化原先决策树的结构，从而提高重复率。
* FAST-ER选择了如下图所示中心点周围48个像素点来判断是否是角点



* FAST-ER的作用和总结：因为每个迭代过程中，都需要对重新应用新的决策树进行检测，而且16个变换中每一个都需要对应一个候选树，所以这样的检测算法并不十分有效，因此，从效率上考虑，上述的算法一般用于产生训练数据，之后获得较为精确的角点检测结果，我们就可以通过原来的FAST算法来产生单个树
* 待解决问题
* FAST-ER使用模拟退火方法训练的具体过程

1. 《BRIEF: Binary Robust Independent Elementary Features》总结
2. 内容摘要

本文提出了BRIEF（Binary Robust Independent Elementary Features）描述子，它是一种使用二进制串表示的描述子，使用汉明距离做匹配，占用存储更小，计算速度更快

1. 总结及理解

* BRIEF使用对关键点周围一块S\*S区域的（x，y）像素点对做比较来获得它的描述子，如下图所示，实线表示一对比较的点，其使用(X, Y) ∼ i.i.d. Gaussian(0, S2/25)分布获得。提取BRIEF描述子包括以下几步
* 使用高斯核对图像进行平滑处理
* 使用某种分布选择关键点周围的点对，常用如上图所示高斯分布。然后比较点对之间的大小获得关键点的描述子，点对和比较顺序选定之后不再改变
* BRIEF-k表示有k\*8对点作比较

1. 启发

要保持特征点的尺度不变性，常使用图像金字塔，不同距离拍摄的照片形成图像金字塔一样的效应。旋转不变性一般通过计算关键点周围像素特点（如梯度）确定关键点的主方向，这里旋转一般指摄像头绕与照片平面垂直轴旋转，这样关键点的主方向能反映拍摄时旋转的角度，然后将这个角度作为描述子的考虑因素提高匹配的鲁棒性

第13周汇报

——张溢炉

1. 《Mapping Technology in Visual SLAM: A Review》总结
2. 内容摘要

本文主要介绍一些传统的建图方法

1. 总结及理解

* 对于移动机器人来说，SLAM作为底层技术，可以用于上层定位、导航、避障、重建和可视化等，稀疏地图可以用于机器人定位，但导航、避障和重建需要用到稠密地图。
* 视觉SLAM可以用于建立稀疏、半稠密和稠密地图。
* 基于特征的方法主要用于建立稀疏地图，Mur-Artal [9]提出了一种建立半稠密地图的方法。直接法可以用于建立半稠密和稠密地图，对环境特征和图像质量不太敏感。稠密地图可以用建立网格地图和八叉树地图，用于导航和避障
* ORB-SLAM的各个优点，它可以用于室内和室外环境：

1）新三维点的匹配更高效

2)闭环检测可以解决长期漂移积累误差

3)目标函数相位可以非常有效地求解

* 建立稀疏地图计算量较小，可以使用板上的处理器。但建立半稠密和稠密地图计算量较大，通常需要用GPU
* 未来的发展方向和应用有3D实时建图、建立网格地图和拓扑地图、增强现实AR、医疗应用

1. 《CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction》总结
2. 内容摘要

本文介绍了将CNN与半稠密SLAM方法结合的单目相机稠密建图方法，另外还在图中加入了语义信息

1. 总结及理解

* 动机：克服单目相机的一些局限，如绝对尺寸不确定性、旋转和低纹理区域遇到的问题
* 模型如下图所示，模型中有三个通道，中间的是CNN通道，可以通过预训练好的CNN模型生成全局的初始化深度，CNN模型建立的深度在物体边缘比较模糊；这正好与半稠密方法互补，故再利用最上层的半稠密方法来精细化物体的边缘；最下层通道与上两层相互独立，增加语义信息有助于智能体对周围环境的理解和相互作用。注意，因为实验使用的摄像头与采集数据集使用的摄像头不同，其内参也会不一样，所以使用CNN预测的深度需要经过如下公式调整



* 下图展现的是调整CNN预测参数（with adjustment）和精细化（Refinement）带来的效果提升
* 文中使用的两个数据集是：the ICL-NUIM dataset [8] and TUM RGB-D SLAMdataset [26]
* 语义标记中常用的四个标记是：floor（如地板）,vertical structure（如墙）, large structure（如家具）, small structure（如杯子）
* 实验时，CNN和语义分割模型是在GPU运行的，其他运算在CPU上运行。目前来看深度学习的计算量太大，需要使用到CPU和GPU，不太适合计算能力弱、实时性要求高的小型移动机器人，但5G等通讯技术在速度上的提升，可以考虑将数据传输到云端处理。此外，像汽车这种大型的移动机器人，可以搭载GPU使用深度学习来做SLAM，有利于提高自动驾驶的安全性
* 未解决问题：具体公式还需在理解一些基础公式（如g2o、Huber function等）后来理解

1. 《Past, Present, and Future of Simultaneous Localization and Mapping: Toward the

Robust-Perception Age》总结

1. 内容摘要

本文较上周两篇对SLAM过去30年的发展做了比较完整的回顾，不仅仅是视觉SLAM。可将SLAM的发展分为2个时代：古典时代（1986-2004，概率方法）、算法分析时代（2004-2015）

1. 总结及理解

* 建图主要有两个作用：第一，给上层算法使用，如路径规划、避障和人机交互等；第二，可以减小视觉里程计估计的误差
* 以下分别是超声波、摄像头、雷达、激光的优缺点图，单个传感器有一些局限，融合多种传感器可以取长补短，适应不同环境



* visual-inertial navigation (VIN)方法融合了相机和IMU，不用做闭环检测，取长补短。相机适合长时间移动测位姿，但相机不能判断是物体动还是相机动。IMU擅长快速移动情况，可以用于判断相机的短时间的移动
* 状态估计的两种方法
* 历史上很长一段时间，研究者们使用滤波器，尤其是扩展卡尔曼滤波器（EKF），只关心当前状态
* 近年来普遍使用非线性优化方法，使用所有时刻采集的数据进行状态估计，被认为优于滤波器，而成为主流方法
* SLAM的常见公式有：maximum a posteriori estimation, factor graph optimization, graph-SLAM, full smoothing, smoothing and mapping(SAM)
* SLAM的分类如下图，从算法、环境、传感器、地图类型、传感器的输入进行了分类
* Do robots need SLAM?

当机器人可以通过已知路标可靠定位时不需要SLAM，目前，一些缺少定位机构（如GPS）的地方适合用SLAM，SLAM主要用于室内

* 没有闭环的VIO（视觉-惯性里程计）引发的多传感器融合技术
* SLAM=里程计+后端（非线性优化+闭环检测），闭环检测可以减小漂移，同时完整的SLAM能够放置错误的数据关联和感知混淆，鲁棒性更好
* 全局一致的地图在很多应用都需要
* Is SLAM solved?这个问题需要从以下三方面考虑
* 机器人：运动类型（动态-机器人是否是运动的、最大速度）、传感器（分辨率、采样率）、计算资源
* 环境：2D或3D、是否有现成路标、动态目标数量、对称性、感知混淆的可能
* 性能要求：精度、地图类型（稀疏、稠密）、成功率要求（自动驾驶要求百分之百）、估计延迟、运行时间、环境大小。
* 机器人步入第三个时期-robust-perception age，主要面临以下几个需求
* Robust Performance：在长时间和大范围条件下失败率低，具有失败恢复的能力，能够根据内外部条件自动调节参数（如排除异常点的阈值）
* High-Level Understanding:超越几何重建，如加入语义
* Resource Awareness:能够在普通的传感器和处理器上运行
* Task-Driven Perception:根据任务采集重要数据，丢弃冗余数据，及自适应地选择合适的地图表达方式
* Maximum a posterior（MAP）在线性高斯情况下与KF的估计结果一样。实践证明MAP方法优于贝叶斯方法。
* Factor graph的优点
* 对问题而可视化
* 可以建模包含各种变量和因子及任意连接的复杂的推理问题
* BA起源与MAP，与高斯假设下的MAP相似，不同之处在于（4）
* (4)中的因素不像BA中那样局限于对射影几何的建模，而且包括了多种传感器模型
* (4)通常需要增量地解决:在机器人移动的每一步都提供了新的度量。
* 视觉里程计是Short-term data association，回环检测是long-term data association
* Slam失败的两个方面
* 算法：数据关联失败（感知混淆或去除false negative点造成关联太少而失败）；未建模的动态环境。
* 硬件：传感器精度下降
* 鲁棒性的挑战
* Failsafe SLAM and Recovery:后端优化影响鲁棒性的原因：其为非凸优化，依赖于初始值；异常点会造成整个优化失败，故需加入huber缓解。
* Robustness to HW（hardware） Failure:
* Metric Relocalization: 当前的特征描述符缺乏足够的不变性来在这种情况下可靠地工作。
* Time Varying and Deformable Maps:通常假设环境是刚体和静态的，但实际环境会刚刚相反
* Automatic Parameter Tuning:
* 解决长时间大环境的方法，室外有GPS不就好了吗？建图好像没太大必要
* Node and Edge Sparsification:去除信息少的
* Out-of-Core (Parallel) SLAM:多线程
* Distributed Multirobot SLAM:多机器人，涉及地图的融合
* 长时间大环境面临的挑战
* Map Representation:地图会随着时间和范围变大越来越大
* Learning, Forgetting, and Remembering:
* Robust Distributed Mapping:多机器人建图时的鲁棒性
* Resource-Constrained Platforms:机器人的计算能力和内存都是有限的
* metric地图（有尺度信息的）的表示，2D激光slam常用栅格地图和特征地图
* Landmark-Based Sparse Representations：特征地图
* Low-Level Raw Dense Representations：点云
* Boundary and Spatial-Partitioning Dense Representations：如给出多边形的边，即将空间分成不同的多边形，包括曲边。如TSDF
* High-Level Object-Based Representations：语义
* 地图表示面临的挑战
* High-Level Expressive Representations in SLAM:
* Optimal Representations:选择合适的表示
* Automatic Adaptive Representations:根据任务环境等自适应的选择合适的地图
* 新传感器
* Range cameras:深度相机
* Light-field cameras:光场相机，拍摄光强和射线角度
* Event-based cameras:事件相机
* 重要文献，不太重要的在论文标记
* 其他概述：8、70、65（算法分析方法）、TABLE 1
* 多传感器融合：83（VI）、163(VIO)、175(VIO)、95（VIO）、106（表现好）、139（表现好）、260（V激光）
* 稀疏矩阵的优化：125、126、146、204
* 动态环境的方法：检测、丢弃动态目标（180）、对动态物体进行建模（12、253）。时间独立的多地图建模动态环境（61）、时变的单地图建模环境（140）。
* 语义：第六部分，仔细看文章
* 主动SLAM：文章第八部分
* 其他：论文上画三角地方

1. 《Real-Time Semantic Mapping of Visual SLAM Based on DCNN》总结
2. 内容摘要

本文提出了一种使用DCNN模型建立语义地图，并具有较好的实时性

1. 总结及理解

* 动机：传统的SLAM只包含距离和颜色等低级信息，本文使用深度学习在地图中加入语义高级信息，并保持SLAM的实时性，促进机器人对陌生环境的理解，加强人机交互
* 本文的整体模型如下，首先使用ORB-SLAM选择关键帧，实时定位和建图，然后将关键帧通过语义分割模型进行处理，得到语义图，最后将ORB-SLAM建立的3D图与语义图合成。



* 语义分割模型如下，包括feature extraction、muti-scale、classification三个模块
* feature extraction：使用了一个1\*1的卷积层将64个通道减少到32个通道加速后面运算；使用dilated convolution卷积提高输出特征的分辨率
* muti-scale：先使用图像金字塔方法池化，再使用4个dilation为6, 12, 18，24及核为3\*3的dilated convolution并行处理，最后将结果串联得到最后特征
* classification：使用softmax进行分类



* 本文使用了NVIDIA Jetson TX2嵌入式模块来实现视觉SLAM的实时处理。它是一台模块化 AI 超级计算机，采用 NVIDIA Pascal™ 架构，具有 256 个CUDA核心。更棒的是，它性能强大，但外形小巧，节能高效，非常适合机器人、无人机、智能摄像机和便携医疗设备等智能边缘设备。它支持 Jetson TX1 模块的所有功能，同时可以铸就更大型、更复杂的深度神经网络。
* 实验结果如下图所示，虽然准确率相对ResNet101低0.091，但在各种分辨率下速度是ResNet101的3倍左右。从本文实验结果来看，将深度学习用于小型移动机器人具有一定可行性

1. 《Deep Residual Learning for Image Recognition》总结
2. 内容摘要

本文是ILSVRC 2015分类任务的冠军模型：ResNet

1. 总结及理解

* 动机：一般来说，随着模型深度增加，在不过拟合的情况下，训练误差和测试误差会更小，但是实验结果并不是这样，刚刚相反，出现了退化现象，实验结果如下图所示。ResNet旨在解决这个问题。假设：如果最优函数更接近于恒等映射而不是零映射，那么solvers应该更容易找到与恒等映射相关的扰动，而不是将函数作为一个新的函数来学习，这样学习更快。



* ResNet由堆叠的一个个小模块（如下图所示）组成，相对VGG的改变，ResNet的非线性层拟合的目标函数是F(x) := H(x) – x，H(x) := F(x) – x是任务待拟合的函数。其中X的映射有三种方式，从实验结果来看A\B\C对实验结果影响不大，但计算量逐个增大：

1. 当非线性层输出维度和x相同时，对应元素直接相加；不同的话，填充0
2. 当非线性层输出维度和x相同时，对应元素直接相加；不同的话，x进行线性变换，然后与F(x)对应元素相加
3. 任何情况下，x都进行线性变换，然后与F(x)对应元素相加



* 下图为plain网络和ResNet实验结果，细线为训练误差，粗线为测试误差。ResNet-34的训练和测试误差都小于ResNet-18，说明ResNet解决了退化的问题



* 待解决问题：batch normalization (BN) [16]

第12周汇报

——张溢炉

1. 《基于深度学习的视觉 SLAM 综述》总结
2. 内容摘要

本文介绍了同时定位与地图构建（SLAM），介绍了一些传统的方法和深度学习方法

1. 总结及理解

* 典型SLAM系统前后端如下图，传统的方法包括运动方程和观测方程，运动方程利用两帧图像计算相机位姿，观测方程用于后端优化。单目摄像头较难计算像素点的深度，需要通过两帧图像计算；双目可以通过单次拍摄计算深度；RGB-D摄像头可以直接获得深度。
* 视觉里程计：也称帧间估计，通过前后两帧图像计算摄像头位姿。
* 后端优化：后端接受不同时刻视觉里程计测量的相机位姿，以及回环检测信息，对它们进行优化，减小噪声的干扰和累积误差，得到全局一致性轨迹和地图。
* 回环检测：判断自身是否进入历史同一地点。回环检测发生时可触发 SLAM 后端全局一致性算法进行地图优化，消除累积轨迹误差和地图误差。闭环检测问题本质上是场景识别问题。
* 建图：根据估计的轨迹，建立与任务要求对应的地图。



* 人工设计的稀疏图像特征当前有很多局限性，一方面如何设计稀疏图像特征最优地表示图像信息依然是计算机视觉领域未解决的重要问题，另一方面稀疏图像特征在应对光照变化、动态目标运动、摄像机参数改变以及缺少纹理或纹理单一的环境等方面依然有较多挑战。
* 目前深度学习结合SLAM有3个方向
* 用深度学习方法替换传统SLAM一个或几个模块特征提取，如特征匹配，提高特征点稳定性，提取点线面等不同层级的特征点，深度估计，位姿估计，重定位；
* 在传统SLAM之上加入语义信息，如图像语义分割，语义地图构建；
* 端到端的SLAM，端到端是输入image输出action，没有定位和建图。 如机器人自主导航（深度强化学习）等。

本文介绍了基于深度学习的帧间估计、闭环检测和语义地图生成，目前深度学习多应用于 SLAM 局部的子模块，如定位模块或闭环检测模块。

* SLAM如果能和一些CV、nlp任务结合起来，会产生一些比较好玩的任务，比如：
* 建图时利用图像描述生成一段自然语言描述，并保存当前帧，直到下一帧与当前帧的相似度低于某个阈值时，再生成一段描述，当语句到达一定数量或者描述完一个场景，再利用nlp中生成结尾的任务生成一段总的描述
* 建图时识别图中物体，然后从外部知识获取物体的属性（如颜色、是否可移动等）、物体间的关系，建立某空间的语义图谱
* 与EQA任务结合，SLAM可以实时建图，可以更新某个空间被移动过的物体，但是实时建的图会有很多噪声，不像仿真环境中那么干净

以上任务在小范围比较合适，如家里、仓库等

* 传统方法与深度学习方法比较如下图

1. 《Ongoing Evolution of Visual SLAM from Geometry to Deep Learning:Challenges and Opportunities》总结
2. 内容摘要

本文介绍了视觉SLAM的一些传统的方法和与深度学习结合的方法，讲述了深度学习应用到视觉SLAM上面临的一些机遇和挑战

1. 总结及理解

* 传统的方法如下：
* 基于特征（Feature-Based）：首先通过特征提取和描述子匹配找出两帧图像相同的像素点，可以通过关键帧的位姿计算这些像素点的世界坐标，然后利用几何算法计算相机位姿。相机可以看成一个小孔成像的模型，对于单目相机具有尺度不确定性，需要通过两帧图像来计算像素点的深度。整个计算过程中是帧到帧累加计算，所以会有累积误差，这就需要利用后端和闭环检测来减小累积误差。后端采用图优化的方法，将一些路标点投影到相机成像画面，通过一些算法（如BA）使投影点和实际点尽可能接近。比较有代表性的特征有SIFT、SURF和ORB，目前ORB用的比较多，ORB-SLAM2兼容了单目、双目和深度相机，具有较好的实时性，目前应用比较广泛。得到位姿和深度后，建图还是比较容易，但单目使用这种方法只能建立稀疏的三维图，不能用于导航。
* 直接法：也称光流法，有稠密和半稠密的方法。假设相邻两帧图像的像素值不变，通过梯度下降的方法调整位姿，使两幅图像对应像素点的方差最小，这就要用到李代数，因为旋转公式有约束条件，不好做梯度下降。可以进行稠密和半稠密建图。目前比较好的方法是Direct Sparse Odometry (DSO)
* CNN、RNN、LSTM、encoder-decoder、无监督学习在视觉SLAM方面都有应用，引入了CV和NLP方面的大多数模型。文中主要提到深度学习在SLAM的深度估计、位姿估计、语义建图（如下图）三个方面结合比较多。
* 数据集
* 深度估计：KITTI [35], TUM [40], NYU [41]
* 重定位：7-scenes dataset [46]，Cambridge landmarks [33]
* 视觉里程计（VO）：KITTI [35], Robotcar [38],M’alaga [47], EuRoc MAV [39], NYU [41],TUM [40]
* 场景分割：PASCAL VOC [42], NYU [41], Synthia [43],Cityscapes [37], KITTI [35],ADK20 [45]
* 传统的方法在定位和建图已经具有比较好的准确性和实时性，但鲁棒性不太好，会面临特征稀少、光强变化、尺度不确定性、相机内参漂移等问题，将深度学习应用到SLAM可提高其鲁棒性、语义信息以及学习能力，可以使机器人更好地与环境交互。CNN可以提取大量的隐性图像特征，具有尺度不变性、旋转不变性。
* 监督学习需要数据集来驱动，但不同机器人面临不一样的未知环境，很难建立通用的数据集，建立数据集也需要花费大量的人力和时间，这是当前面临的瓶颈。无监督学习、传统方法与深度学习的结合、多传感器（雷达、超声波、激光、陀螺仪）的融合有利于解决目前面临的一些问题。

1. 《Visual SLAM for Automated Driving:Exploring the Applications of Deep Learning》总结
2. 内容摘要

本文介绍了视觉SLAM在自动驾驶上的应用，很多内容和之前两篇相似，在这就不再重复

1. 总结及理解

* 视觉SLAM比较适合自动驾驶的以下三个情景
* 停车：停车时要识别周围物体，选择车位，避开障碍物。在外面停车每次都得重新建图，对于生活中常去的停车的地方，如家里，可以将图存储下来，方便下次使用，虽然每次物体放置都不一样，但停车场整体结构一般不会变，有助于重定位
* 高速路：高速路需要比较高的实时性，每秒需要处理30张图片，传统方法实时性较好
* 城市驾驶：城市环境比较复杂，移动物体较多，DSO-SLAM表现较好
* 自动驾驶对安全性要求很高，因为一个失误可能就是人命，似乎很难找到一个置信度百分之百的深度学习模型，需要综合各种方法来防止意外的发生。因为世界各地环境都不相同，也很难找到一个通用的数据集
* 文后还提到了一些CNN Based Pipelines，如Tateno et al. [54]提出联合学习语义分割和建立深度地图

1. 《CVI-SLAM—Collaborative Visual-Inertial SLAM》总结
2. 内容摘要

本文介绍了多个智能体协作的Visual-Inertial SLAM，每个智能体上搭载了摄像头和Inertial Measurement Unit（IMU），与中央处理器协同定位和建图

1. 总结及理解

* 整体架构如下图所示，agent负责处理实时性较高的计算，如定位。Server负责处理计算量大且实时性不高的计算，如冗余检测、闭环检测、地图融合、BA。这样可以利用server强大计算力，也可以保持agent相对独立，但是这对两者之间的通信速度和质量要求比较高，图像的数据量比较大。从本文实验结果来看，该方法在agent数量上具有较好的可拓展性，对网速的需求可以接受。



* 下图为不同agent在frame tracking, local mapping of the VIO, communication所用时间，通信时间相对处理时间还是比较少的，但随着agent的增加，agent的处理量不变，但通信压力会变大，server的处理数据量会线性增大
* 下图是在4个agent情况下有无冗余检测模块随着时间增长待处理关键帧的数量变化图，由图可见冗余检测可以有效减少待处理关键帧的数量，减少server的处理压力
* 从本文大致了解了CVI-SLAM的方法，里面涉及的很多公式还需以后继续弄明白

1. 《Keyframe-based monocular SLAM: design, survey, and future directions》总结
2. 内容摘要

本文主要概述基于关键帧的方法。指出了各种方法的优缺点，为选择合适的方法提供了指导，包含了开源和闭源的方法。可以用于查阅文献

1. 总结及理解

* 未来研究方向：
* illumination changes：光照变化。
* initialization:初始化。
* highly dynamic motion：自身的高速运动或者动态物体较多。
* poorly textured scenes：如白墙。
* repetitive textures：一般由相似物体产生。
* map maintenance：全局优化。保证地图的一致性，减少误差，去除异常值。
* failure recovery：最小化失败或者数据关联失败后会触发
* 关键帧方法的重要环节：
* data association：设计选择有直接法、关键帧方法（均衡速度和精度）和混合方法（半直接半关键帧）。像素关联的方法有2D-2D、2D-3D、3D-3D。robust data association against illumination changes, dynamic scenes, and occluded environments
* visual initialization：a robust initialization method that can operate without an initial scene assumption
* pose estimation：an accurate camera pose estimate that is not affected by  
  sudden movements, blur, noise, large depth variations, nor moving objects.
* topological/metric map generation：a map making module capable of generating an efficient dense scene representation in regions of little texture, while incorporating a higher level of perception
* BA/Pose Graph Optimization(PGO)/map maintenance：a map maintenance method that improves the map, with resilience against dynamic, changing environments
* global localization ：在failure recovery和loop closure时使用。a failure recovery procedure capable of reviving the system from significantly large changes in camera viewpoints
* 直接法和关键帧法的比较见图16
* 重要文献：
* 其他综述：1、2、3、4、5、6
* 直接法FAIA,FCIA,ICIA,IAIA比较：8
* Random depth initialization：
* Topological map：36
* Triangulation：43
* Overcome triangulation delay：45

1. 启发