



Robotics Lab
Institute of Cyber-Systems & Control
Zhejiang University



工业控制技术国家重点实验室
State Key Laboratory of
Industrial Control Technology



地面无人平台中的SLAM技术

A tutorial on SLAM for ground unmanned platform

刘 勇

(yongliu@iipc.zju.edu.cn)

<http://www.csc.zju.edu.cn/yliu/>

智能系统与amp;控制研究所



浙江大学 控制科学与工程学院

College of Control Science and Engineering, Zhejiang University

提纲

- 1、背景介绍
- 2、SLAM 理论与方法
- 3、常用开源系统与算法
- 4、总结与展望

ORB-SLAM



Universidad
Zaragoza



instituto de investigación
en ingeniería de Aragón
Universidad de Zaragoza

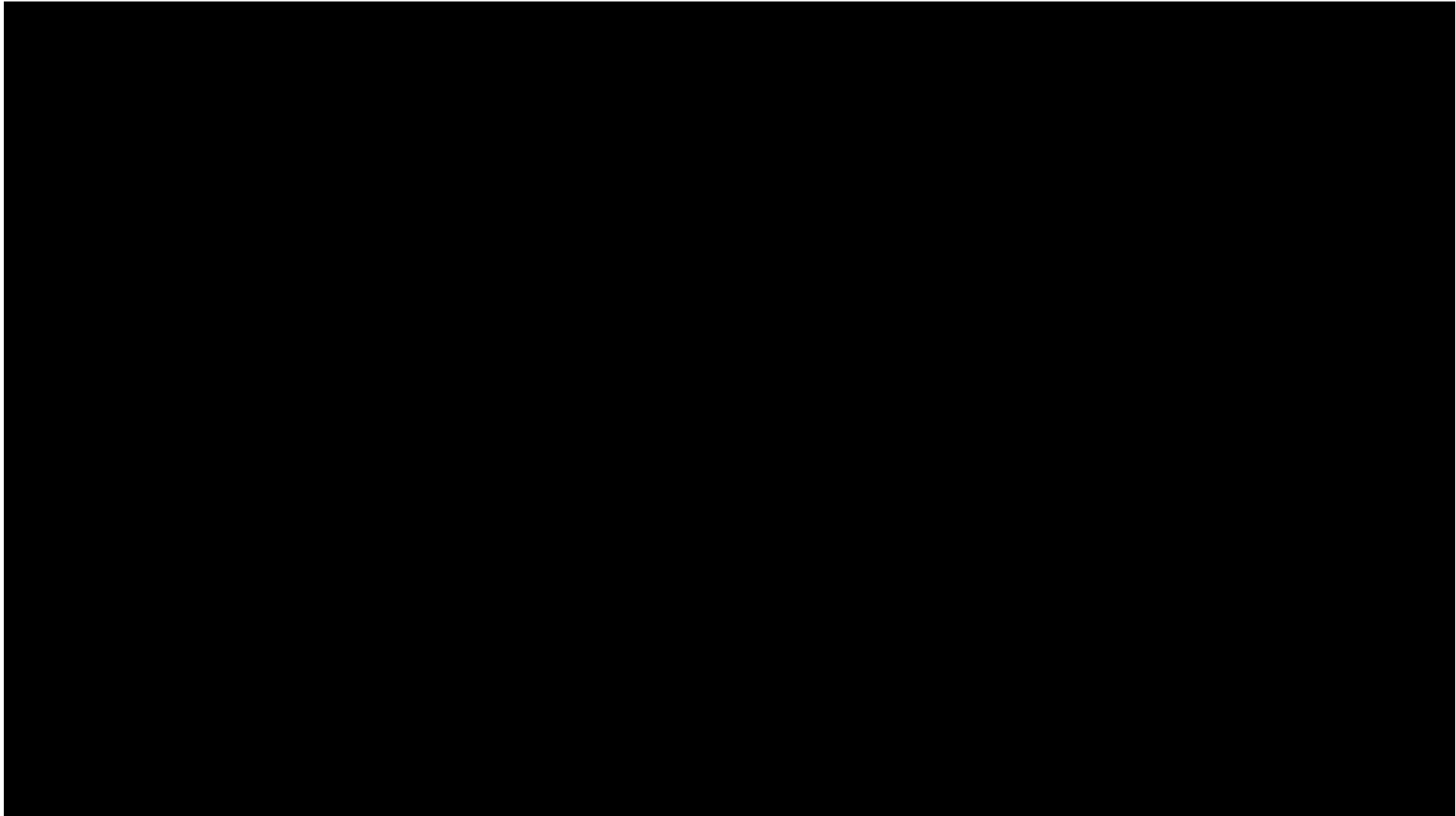


STEREO ORB-SLAM

Raúl Mur Artal and Juan D. Tardós

KITTI 01

Wide-Area Indoor and Outdoor Real-Time 3D SLAM



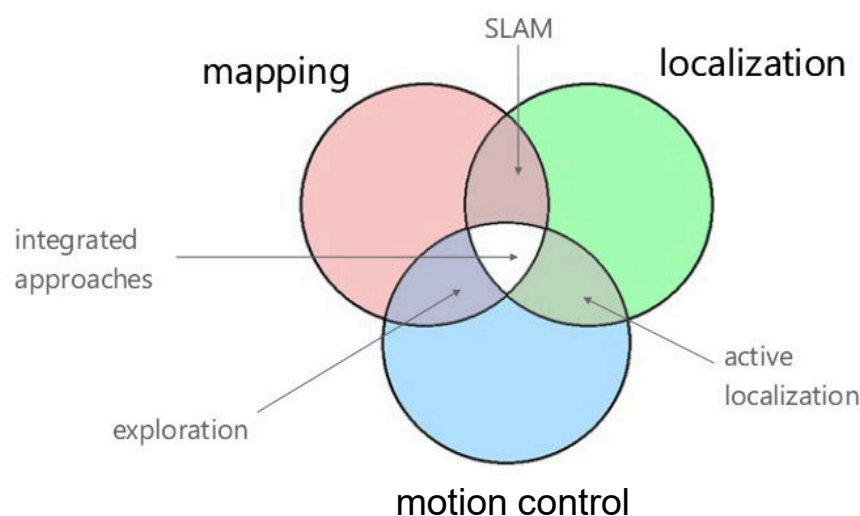
地图构建与定位 (SLAM)

- 机器人在未知环境中从一个未知位置开始移动，在移动过程中根据位置估计和地图进行自身定位，同时在自身定位的基础上建造增量式地图，实现机器人的自主定位和导航。

定位 $\begin{matrix} \leftarrow \text{需要精确的地图} \\ \leftarrow \text{需要精确的定位} \end{matrix}$ 构图
Where am I? What is around me?

SLAM

Simultaneously minimize both error



应用：
机器人、无人机、汽车、AR/VR、智能手机等



SLAM历史（30年）

1. Classical age (1986-2004)

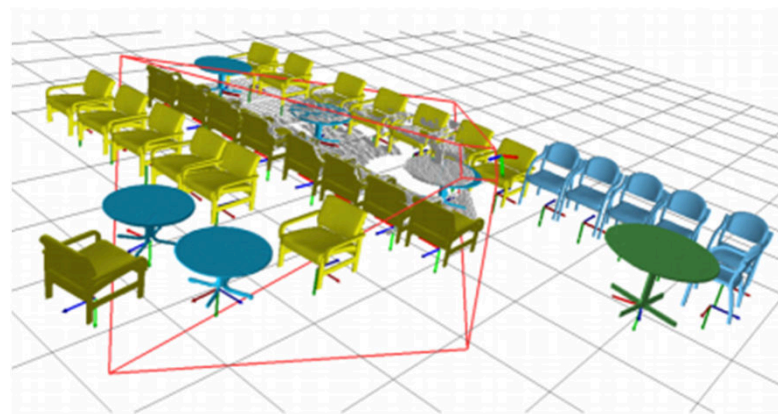
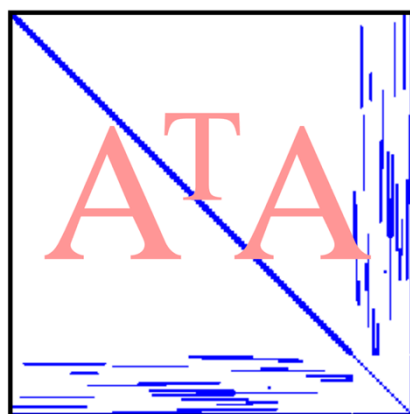
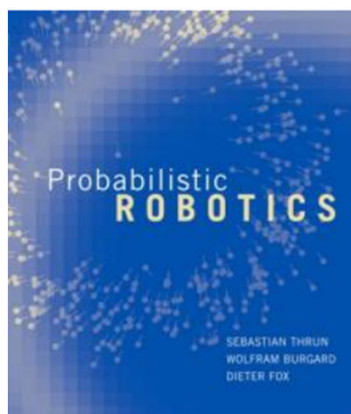
概率公式（EKF、RBPF、MLE），数据关联（data association）

2. Algorithmic-analysis age(2004-2015)

基础特性研究（可观性、收敛性、一致性、稀疏性），开源SLAM

3. Robust-perception age(2015~)

鲁棒性、高层次感知、自适应



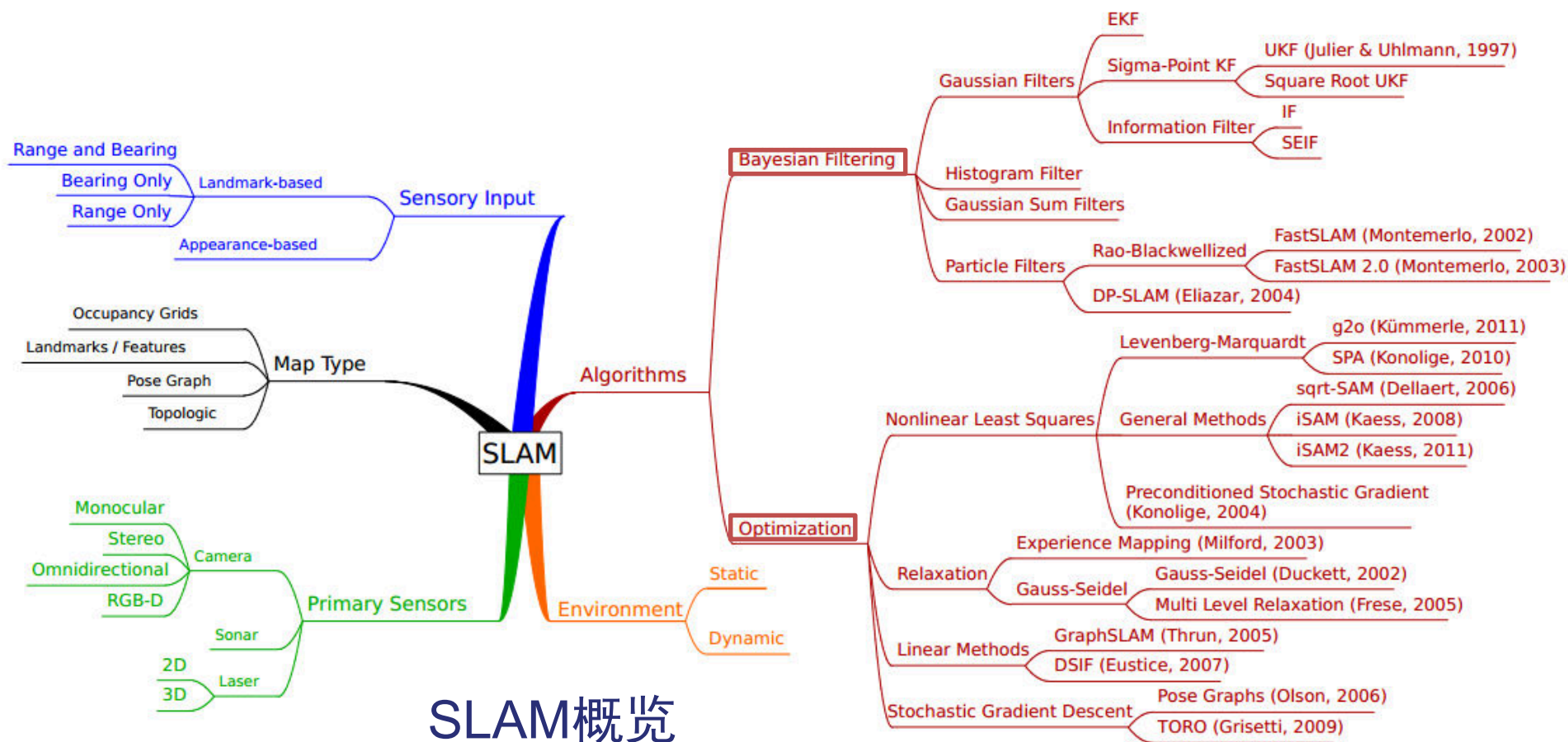
推荐文献

- Durrant-Whyte, Hugh, and Tim Bailey. **"Simultaneous localization and mapping: part I."** IEEE robotics & automation magazine 13.2 (2006): 99-110.
- Bailey, Tim, and Hugh Durrant-Whyte. **"Simultaneous localization and mapping (SLAM): Part II."** IEEE Robotics & Automation Magazine 13.3 (2006)
- Dissanayake, Gamini, et al. **"A review of recent developments in simultaneous localization and mapping."** 2011 6th International Conference on Industrial and Information Systems. IEEE, 2011.
- Fuentes-Pacheco, Jorge, José Ruiz-Ascencio, and Juan Manuel Rendón-Mancha. **"Visual simultaneous localization and mapping: a survey."** Artificial Intelligence Review 43.1 (2015): 55-81.
- Cadena, Cesar, et al. **"Simultaneous Localization And Mapping: Present, Future, and the Robust-Perception Age."** arXiv preprint arXiv:1606.05830 (2016).

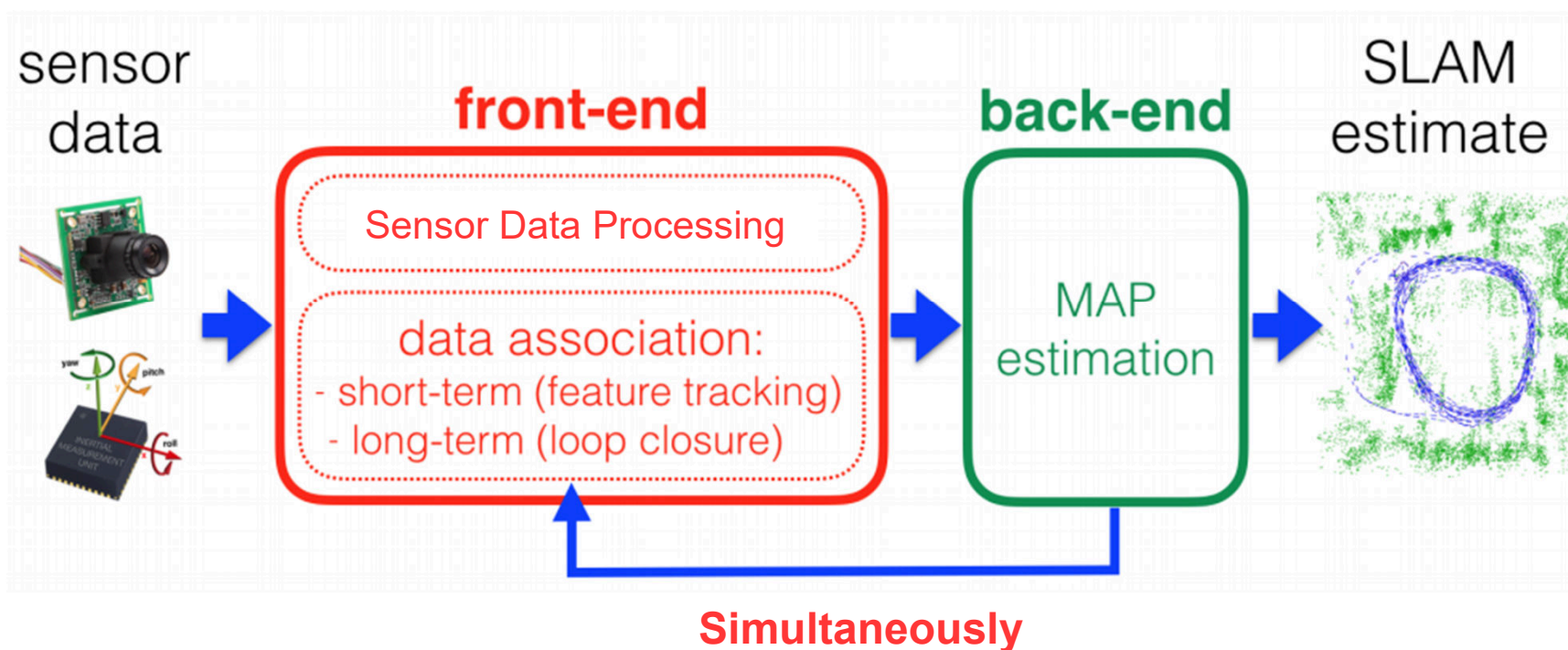
提纲

- 1、背景介绍
- 2、SLAM 理论与方法
- 3、常用开源系统与算法
- 4、总结与展望

SLAM框架



SLAM框架



前端：

- 提取传感器采集的数据
- 进行数据关联

后端：

- 对前端产生的模型进行参数估计

前端 – 传感器

传感器相关

depend heavily on the equipped sensors

激光

Accurate

Fast

Long history in research

Heavy

Expensive

Categories: 2D, 3D



相机

Cheap

Light-weight

Rich information

High computation cost

Work under assumptions

Categories: monocular, stereo, RGBD

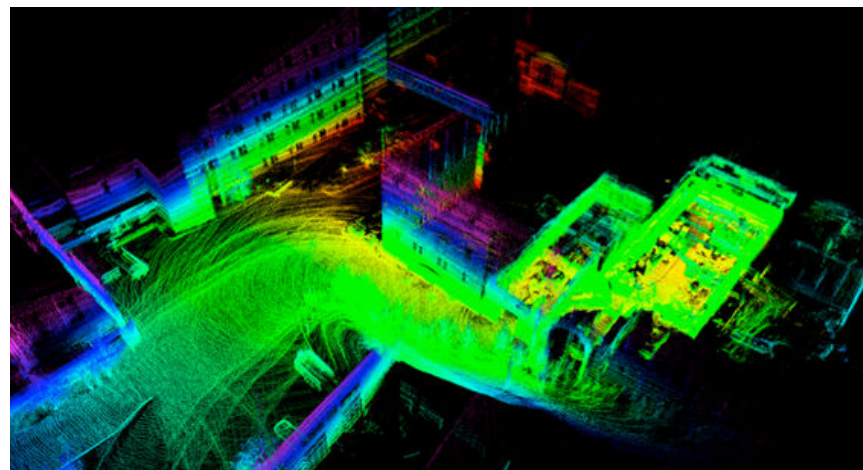


前端 – 传感器

2D Laser



3D Laser

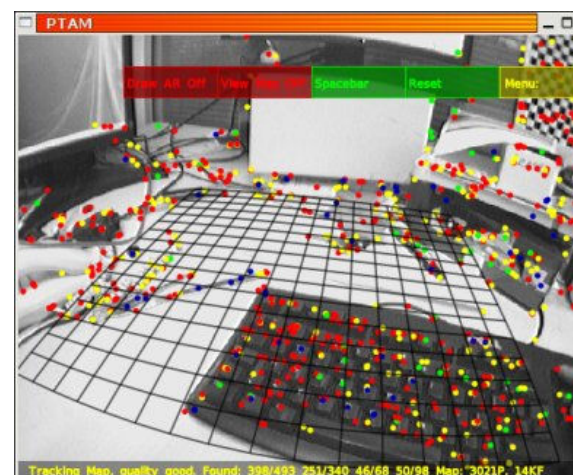
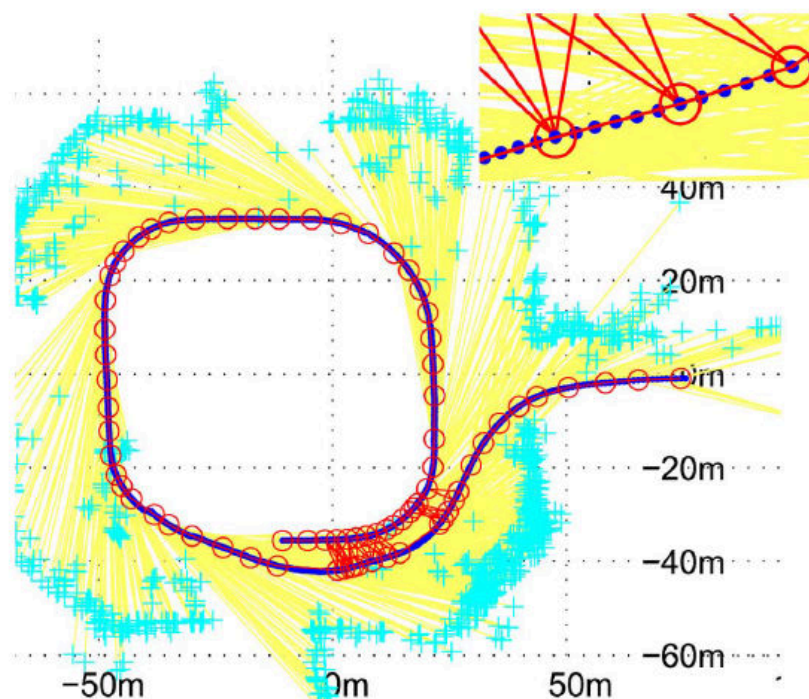


激光传感器

- 直接获得相对于环境的距离信息
- 通过Scan match实现直接定位
- 其定位及轨迹优化可以直接进行

前端 – 传感器

vSLAM



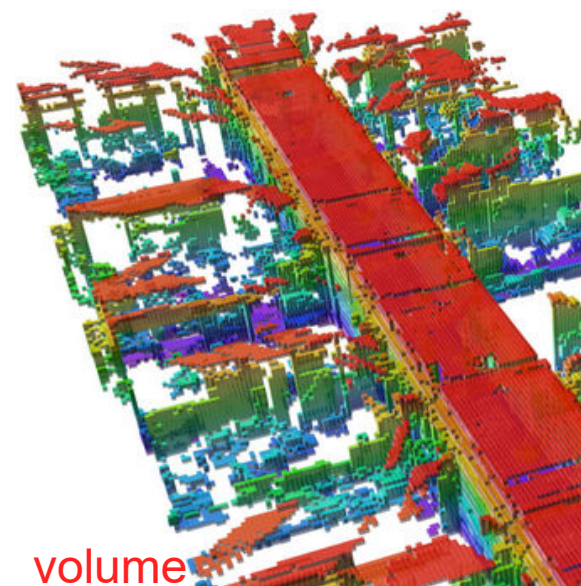
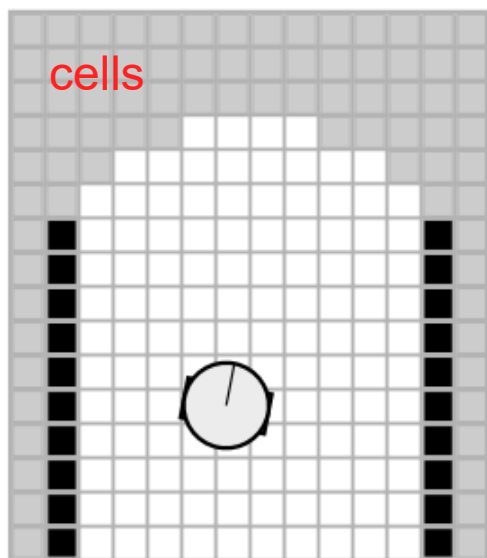
视觉传感器

- 能够获得比激光更丰富的信息
- 需要经过复杂的图像处理已经运算

地图表示

种类繁多，与选择的传感器和特定的任务有关

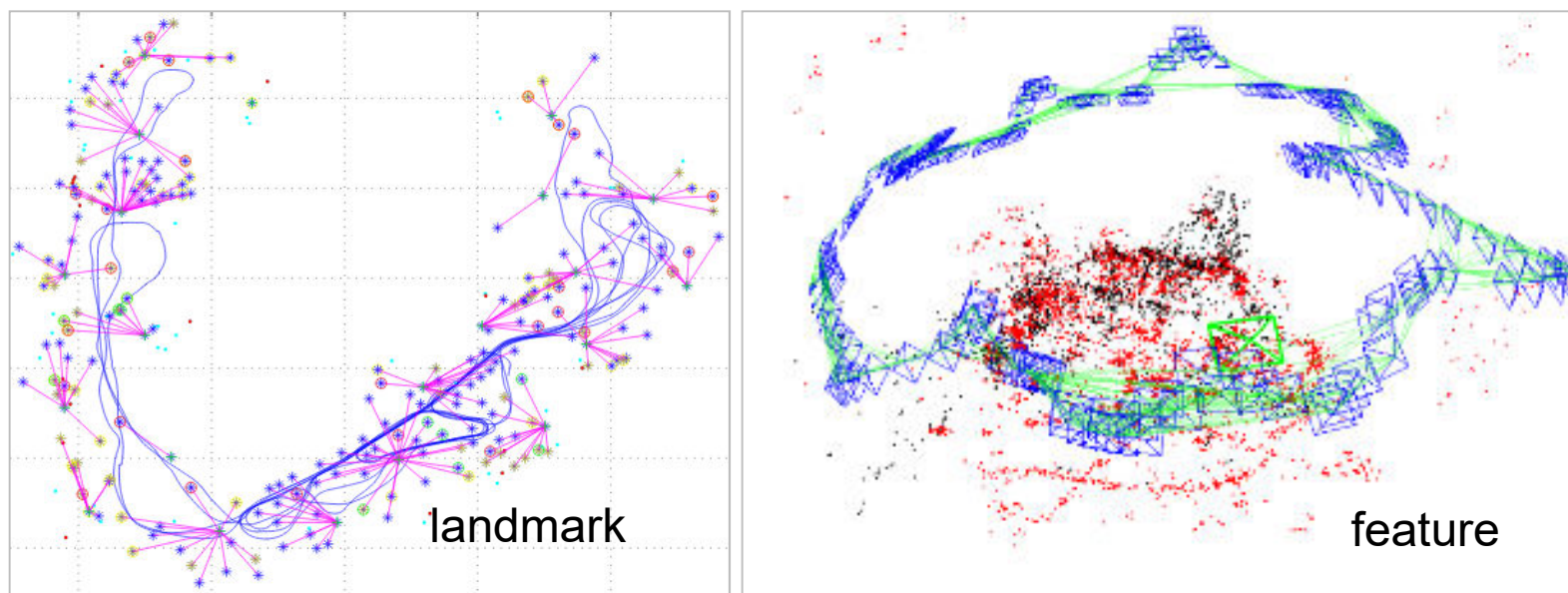
1. 栅格地图



- 每个栅格以概率的形式表示
- 3D的栅格地图采用八叉树存储
- 适用于路径规划与避障

地图表示

2. 特征地图



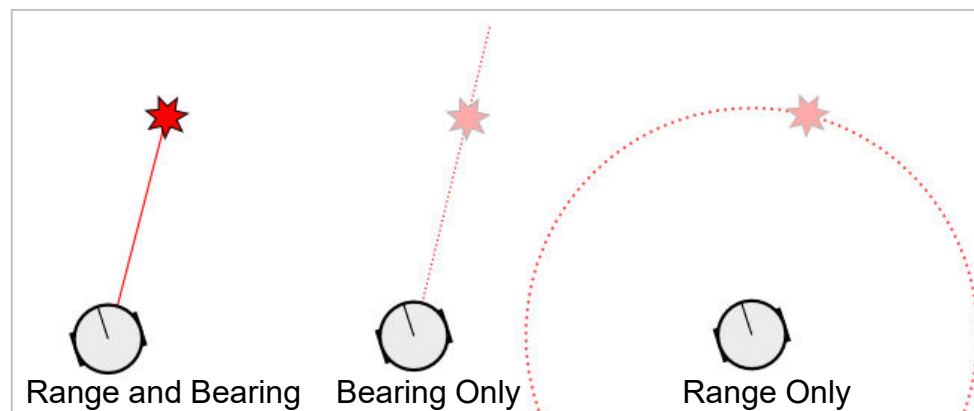
- 地图比较稀疏
- 适用于存在较多特征的场景

地图表示

2. 特征地图

根据传感器类型，特征分为3类

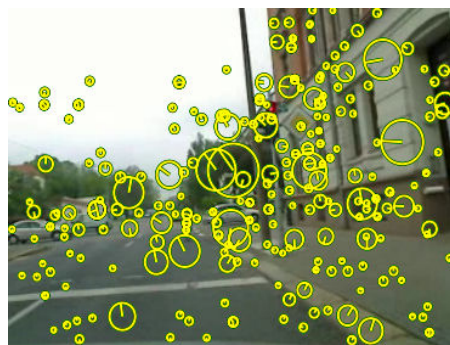
- a) Range and Bearing(激光, 双目、RGBD)
- b) Bearing Only (单目)
- c) Range Only (GPS、WiFi、UWB)



vSLAM特征

点: SIFT, SURF, CenSurE, Harris, MSER, FAST, ORB

线: LSD, EdLine, LBD, MSLD



特征点参数化 (parametrizations)

Euclidean points:

$$P = [x \ y \ z]^T$$

Homogeneous points:

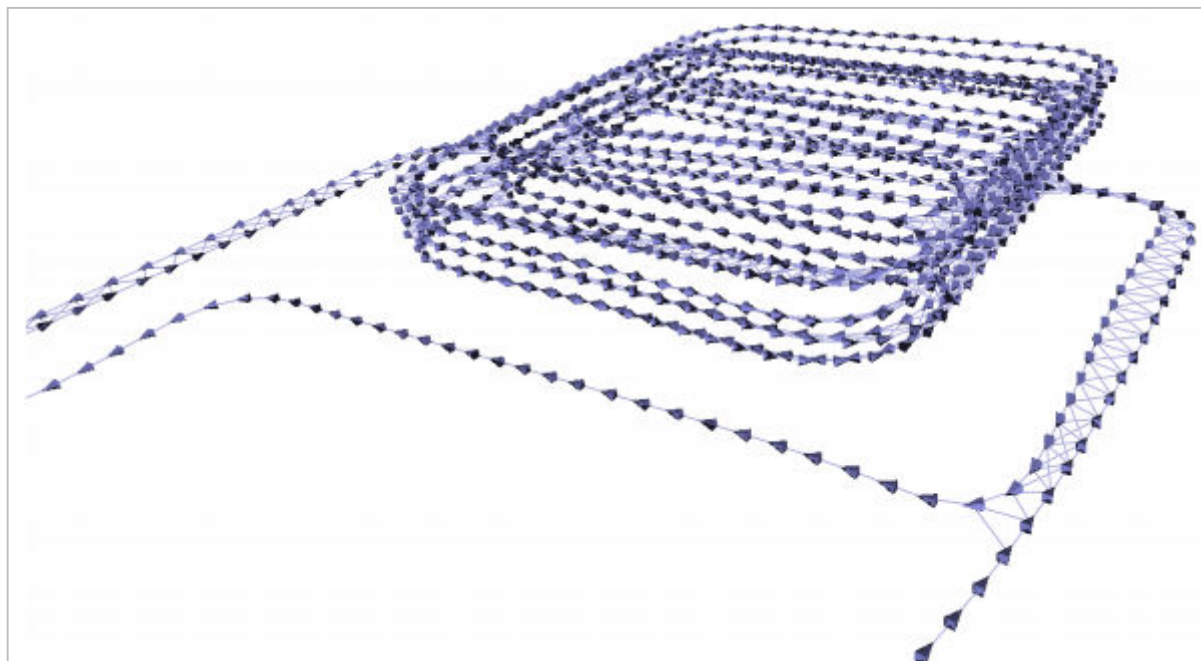
$$P = [x \ y \ z \ \rho]^T$$

Inverse depth parametrization

$$\begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} + \frac{1}{\rho_i} \mathbf{m}(\theta_i, \phi_i)$$

地图表示

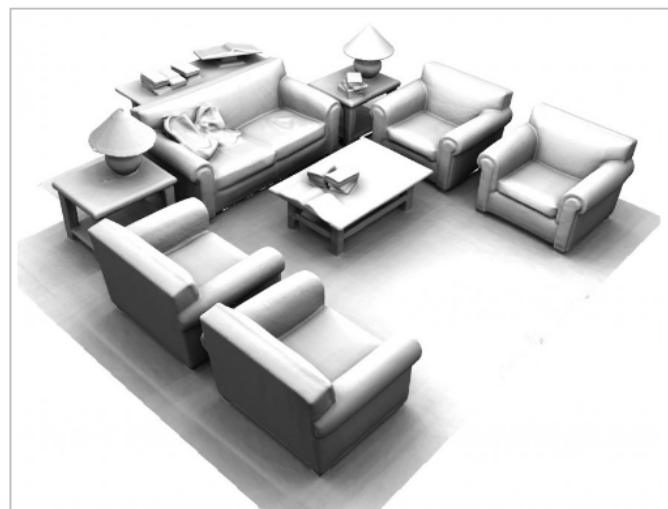
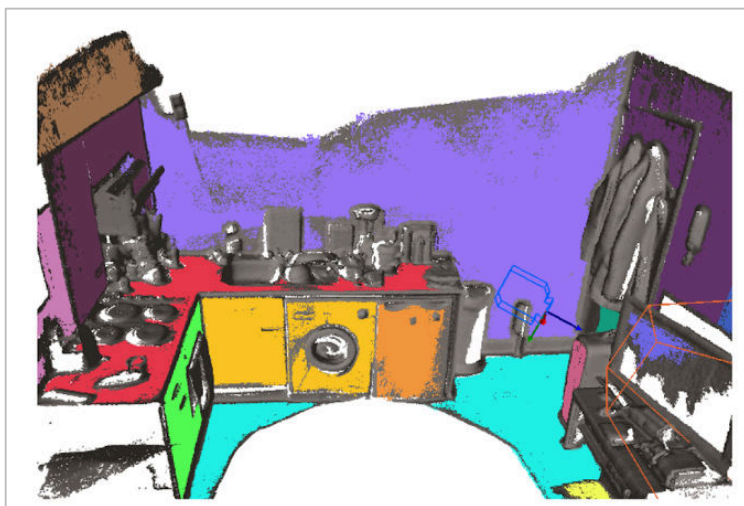
3. 位姿图 (Pose Graph)



- 仅包含机器人的位姿信息
- 有效表达位姿之间的约束
- 可以与栅格地图和特征地图结合

地图表示

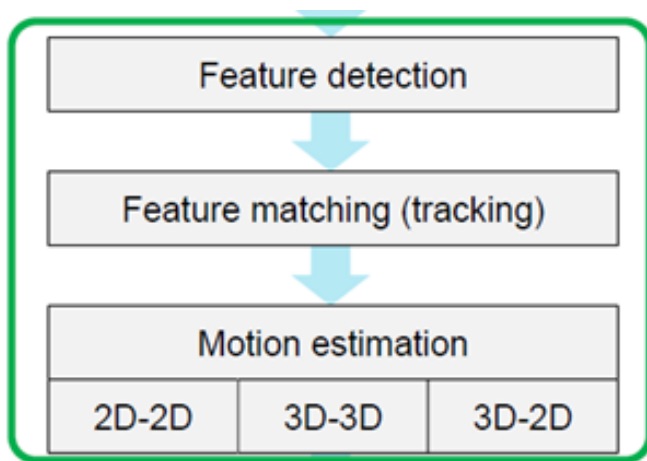
4. 其他地图



- 用于场景构建
- TSDF (Truncated signed distance function), Surfels

前端 – 数据关联

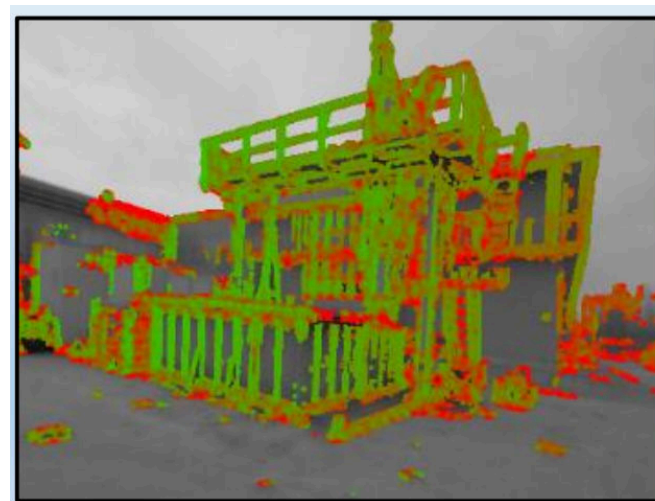
1. Short term (feature tracking)



(1) 提取两帧的特征（RANSAC剔除外点）

(2) 最小化重投影误差

$$T_{k,k-1} = \arg \min_T \sum_i \| \mathbf{u}'_i - \pi(\mathbf{p}_i) \|_{\Sigma}^2$$



最小化光度测量误差（photometric error）

$$T_{k,k-1} = \arg \min_T \sum_i \| I_k(\mathbf{u}'_i) - I_{k-1}(\mathbf{u}_i) \|_{\sigma}^2$$

$$\text{where } \mathbf{u}'_i = \pi(T \cdot (\pi^{-1}(\mathbf{u}_i) \cdot d))$$

前端 – 视觉里程计

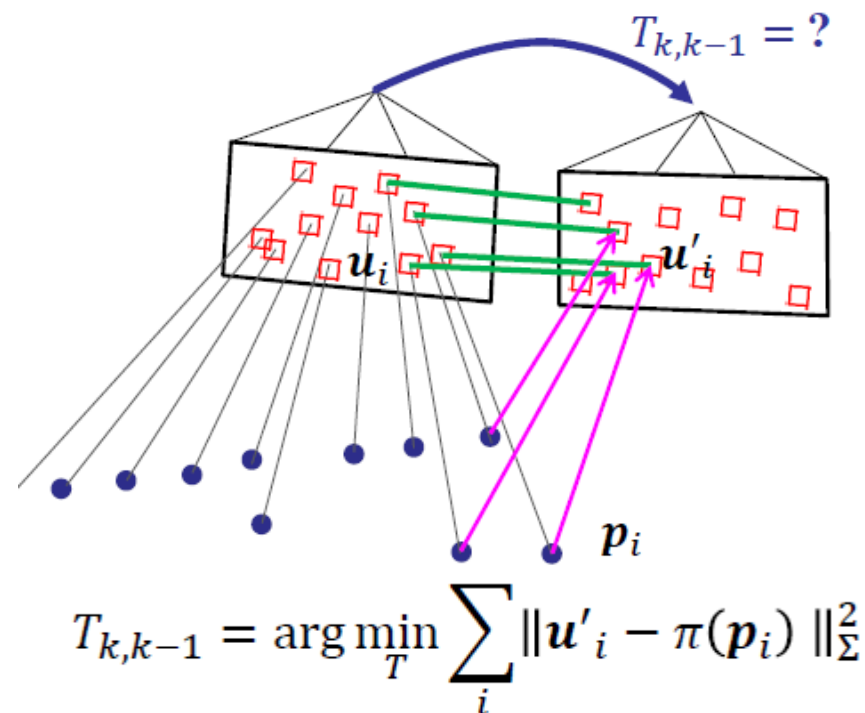
基于特征的方法

步骤:

1. 提取特征和描述子
■ FAST, SIFT, SURF, ORB
2. 找到对应的匹配对
■ Brute-force or kNN match
3. 运动估计
■ PnP+RANSAC、MLE

注意:

1. 特征匹配并不能保证是正确的。
2. 帧间的运动过大可能会导致跟踪失败。
3. 运动估计的结果可能不存在或非全局最优。



前端 – 视觉里程计

基于直接的方法

对于图像 $k-1$ 中的任意像素 u_i

根据深度反投影到空间点 p_i

投影到图像 k 中，其位置为 u'_i

Photometric Error: $e_i = I_1(u_i) - I_2(u'_i)$

最小化误差:

$$\min J = \sum_{i=1}^N e_i^T e_i$$

基本假设:

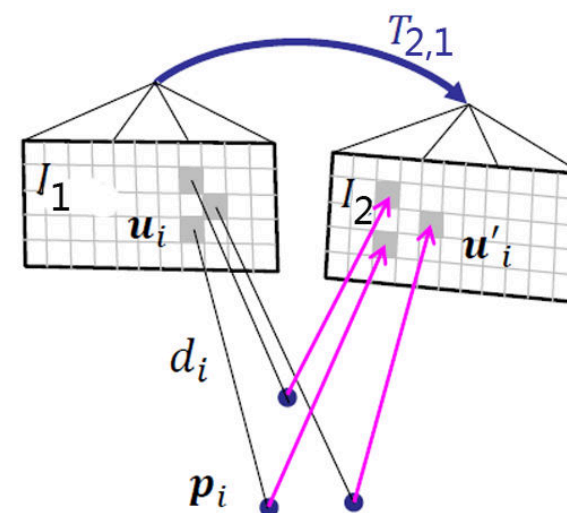
同一个空间点在各视角下，测到的灰度不变。

使用哪些像素？

使用所有像素。DTAM

使用fast特征点处所在的4*4像素块。SVO

使用梯度明显的所有像素。LSD-SLAM



$$T_{k,k-1} = \arg \min_T \sum_i \|I_k(u'_i) - I_{k-1}(u_i)\|_{\sigma}^2$$

$$\text{where } u'_i = \pi(T \cdot (\pi^{-1}(u_i) \cdot d))$$

I : Image gray scale value

d : Depth map of pixels

π : Projection function

前端 – 视觉里程计

基于特征的方法

- ✓支持大帧间运动
- ✓高效率实时优化获得结构与运动
- ✗特征提取和匹配导致效率低
- ✗可能匹配到外点

基于直接的方法

- ✓图像中的大量信息都会被用到（精确，鲁棒）
- ✓提高帧率可降低平均每帧的计算代价
- ✗仅支持有限的帧间运动
- ✗联合优化稠密的结构运动序列代价太高

视觉里程计：

1. 跟踪特征点（线、面）；
2. 估计帧间的增量位姿
3. 构建局部地图

缺点

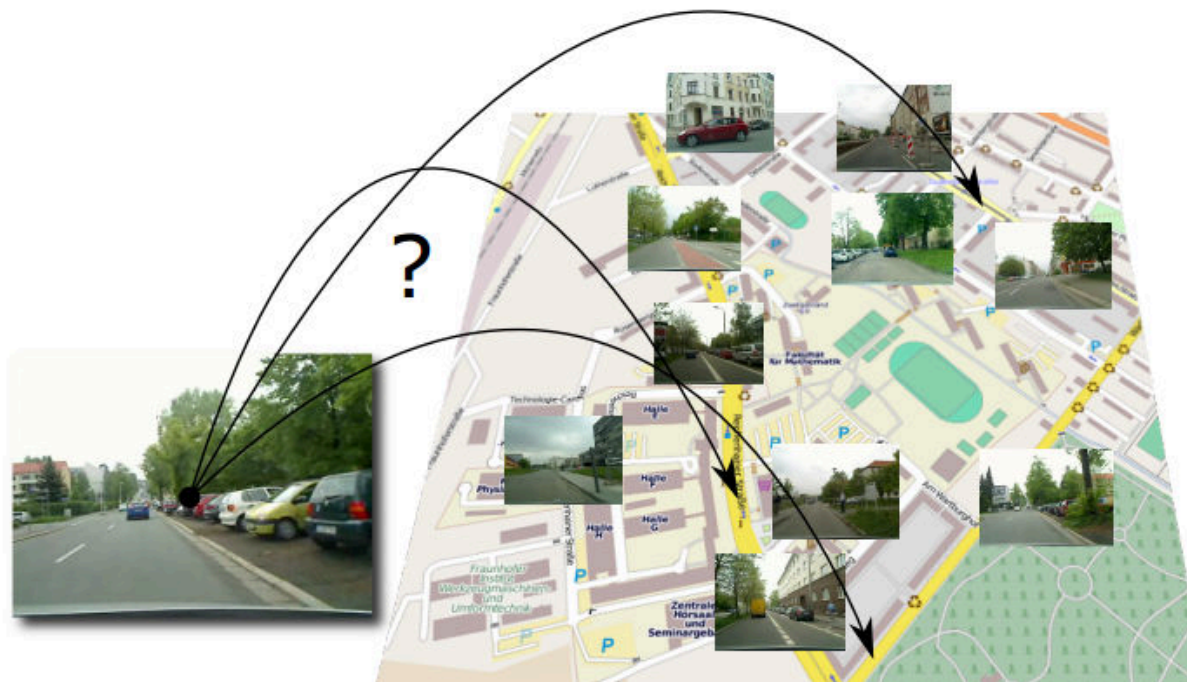
- Drift** 存在累计误差
- Inconsistent** 存在不一致性
- Lost** 由于过快的运动等造成失败

Solutions

- Global optimization
- Loop closure
- Re-localization

前端 – 数据关联

2. Long term (Loop close)



To recognize **visited** places.

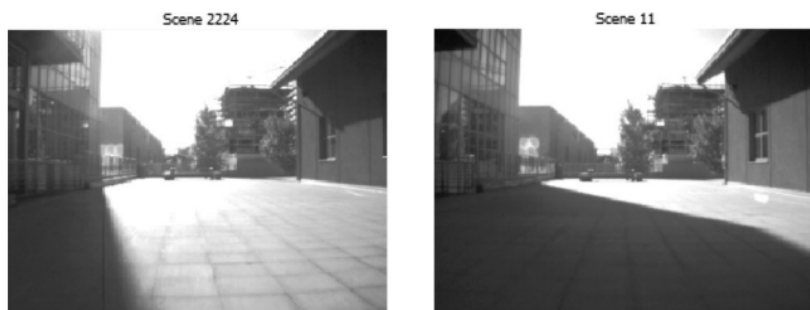


前端 – 闭环

2. Long term (Loop close)



False Positive



False Negative

闭环的关键

- 相机的位置相邻?
- 图像具有相似性?



Appearance-based Approaches

- 仅考虑图像的相似性
- State of the art in vSLAM.

前端 – 闭环

TFIDF 如果某个词比较^{IDF}少见，但是它在这篇文章中^{TF}多次出现，那么它很可能就反映了这篇文章的特性。

词频TF = $\frac{\text{某单词在某文章中出现的次数}}{\text{该文章中单词的总数}}$

逆向文件频率IDF = $\log\left(\frac{\text{资料库中文章的总数}}{\text{包含该单词的文章数量}}\right)$

TFIDF = TF*IDF

- 与一个词在文档中的出现次数成正比
- 与该词在整个资料库中的出现次数成反比

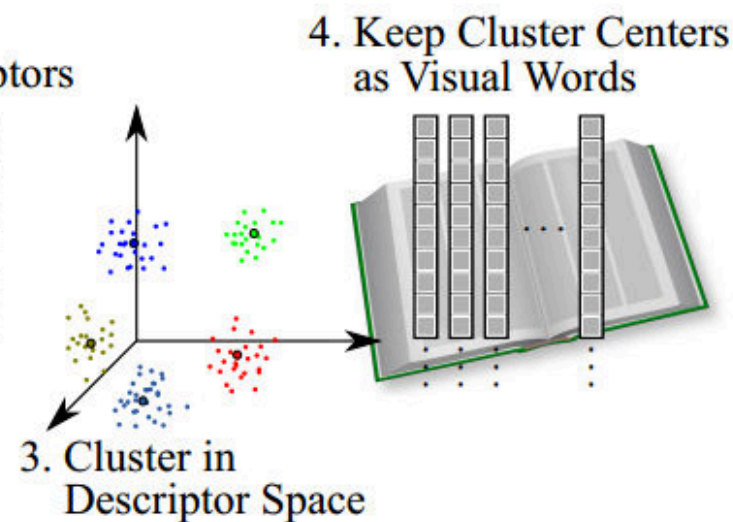
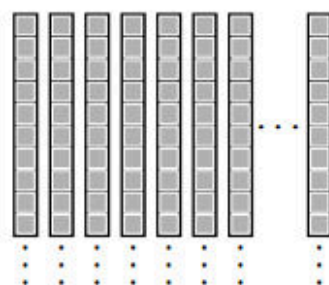
前端 – 闭环

词包模型 (BoW)



1. Collect Keypoints from Training Set

2. Calculate Keypoint Descriptors



Kmeans++

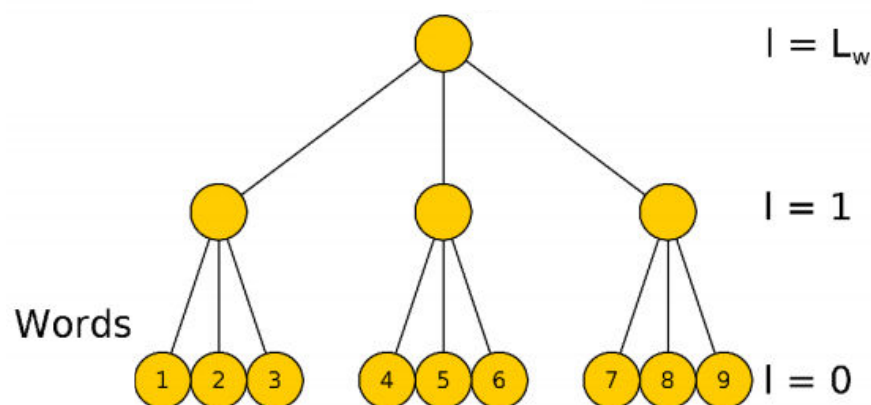
字典的训练 (离线)

1. 收集数据集中的特征点
2. 计算特征点的描述子
3. 对描述子进行聚类
4. 将聚类中心作为单词

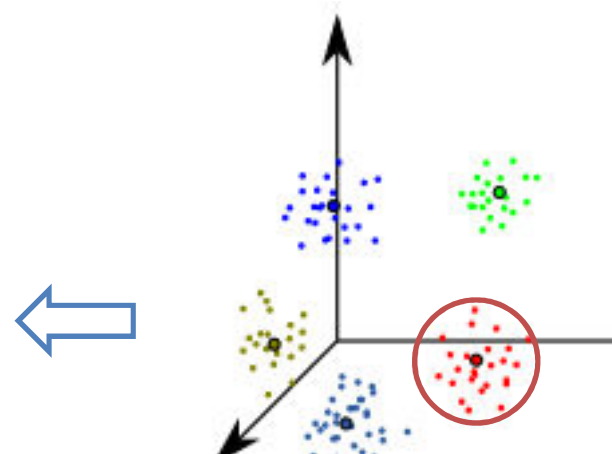
前端 - 闭环

词包模型 (BoW)

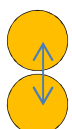
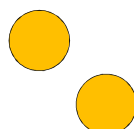
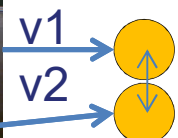
Vocabulary Tree



- 每个word记录了在数据集中出现的频率 **IDF**



对每一类再进行聚类
构建树模型



Bag of word **TFIDF**

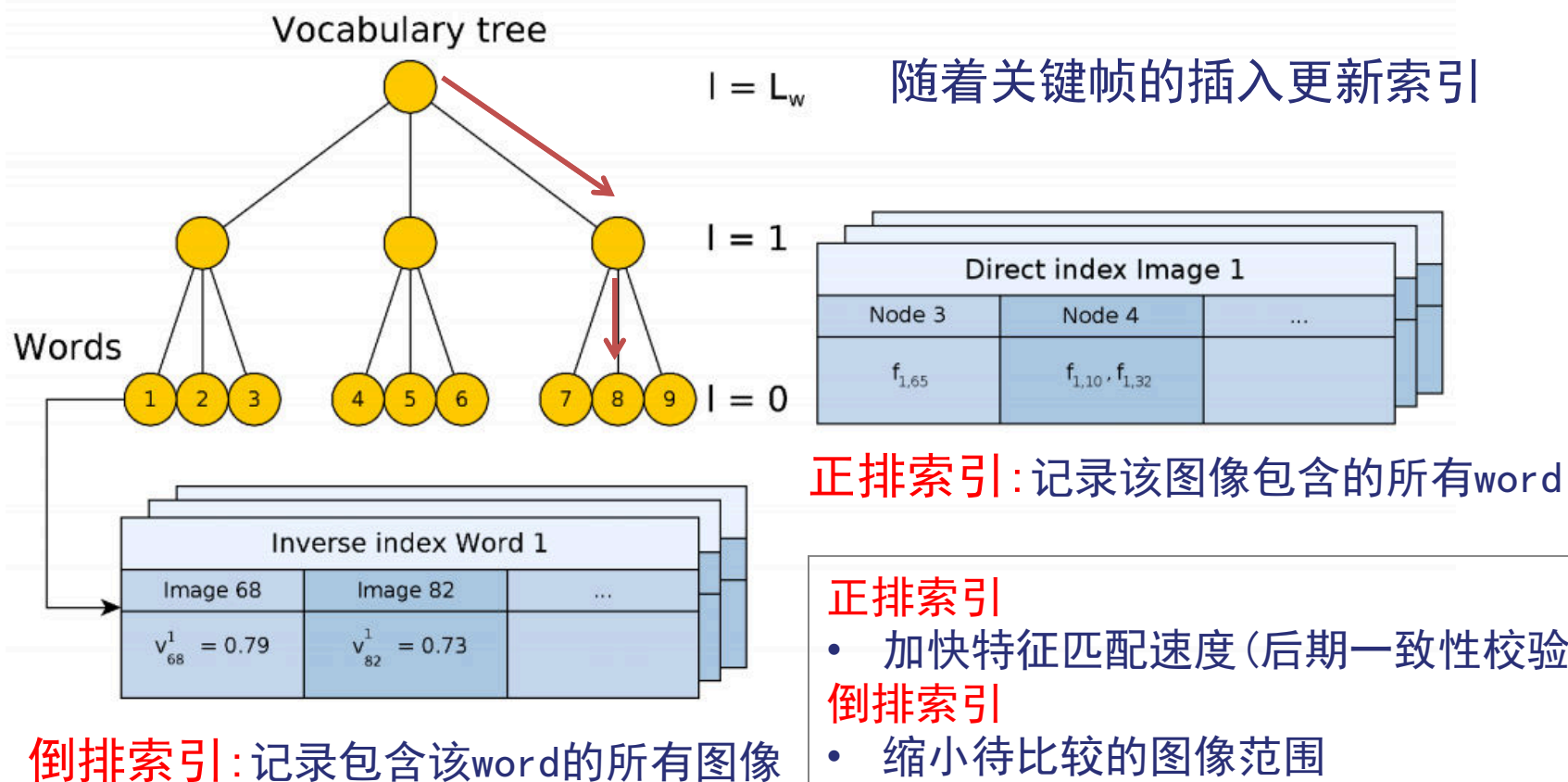
L1 score $s(\mathbf{v}_1, \mathbf{v}_2) = 1 - \frac{1}{2} \left| \frac{\mathbf{v}_1}{|\mathbf{v}_1|} - \frac{\mathbf{v}_2}{|\mathbf{v}_2|} \right|$

两幅图像越相似得分越高。

给定一幅图，如何
找到相似的另一幅？
索引！

前端 - 闭环

词包模型 (BoW) 字典的使用



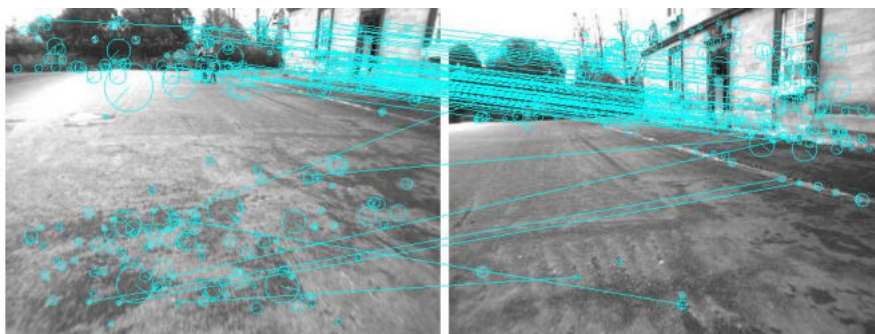
前端 – 闭环

时间一致性校验

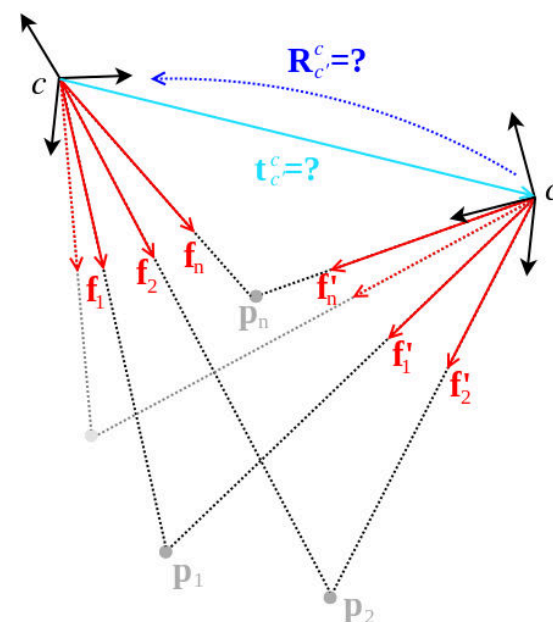


Matching group

空间一致性校验



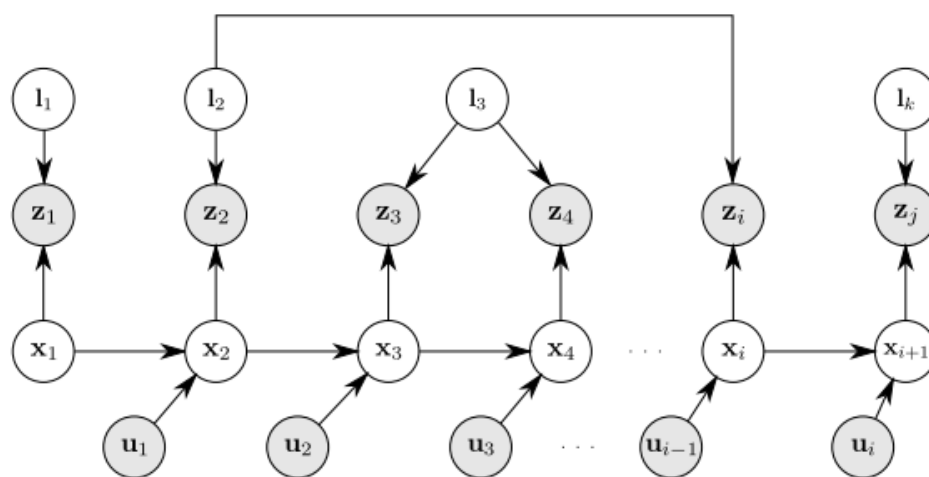
2D: Fundamental Matrix



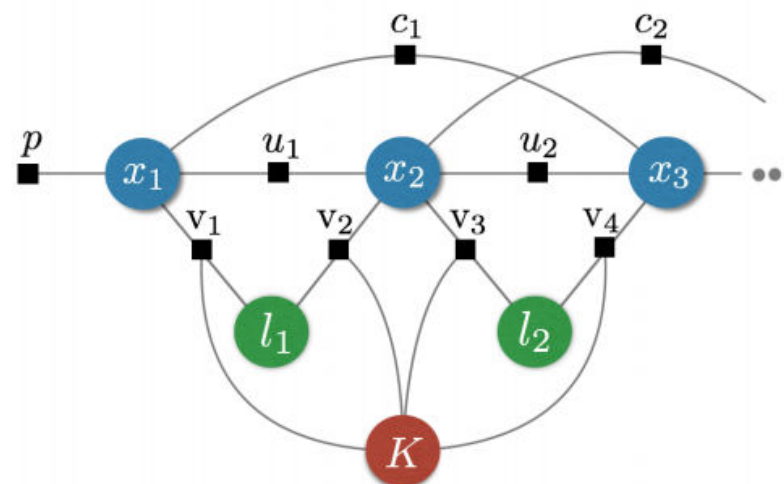
3D: PNP+RANSAC

前端

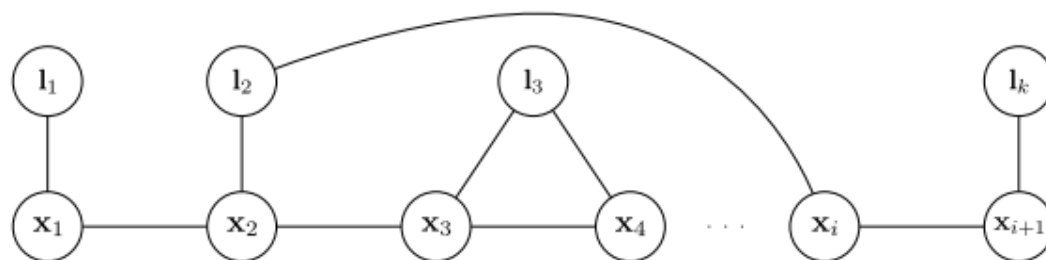
构建图模型



Dynamic Bayesian Network



Factor Graph

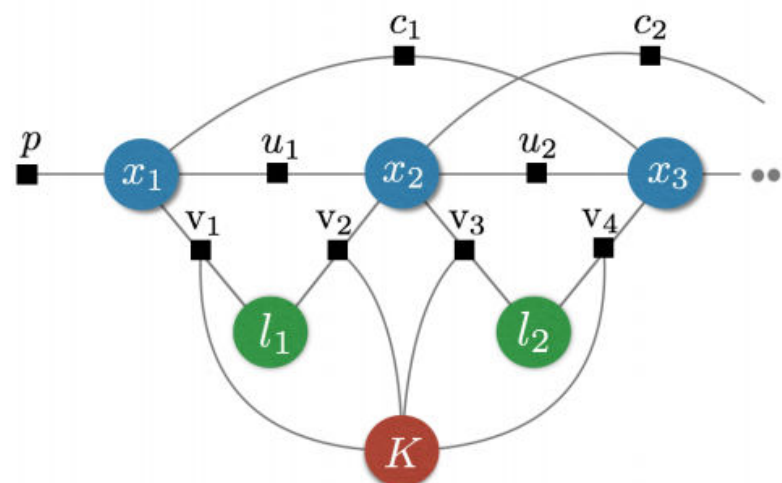


Markov Random Fields

三者可以相互转换

前端

- o x 机器人位姿
- o l 特征点位置
- o K 相机参数
- p prior factor: IMU, GPS...
- v vision factor: 相机观测
- u odometry factor: 里程观测
- c close loop factor: 闭环



Factor Graph

Motion:

$$x_i = f(x_{i-1}, u_i, \varepsilon_i^u)$$

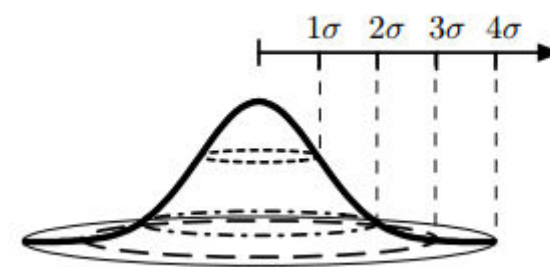
Observation:

$$z_{i,j}^v = h^v(x_i, l_j, K) + \varepsilon_{i,j}^v$$

$$z_i^p = x_i \oplus \varepsilon_i^p \quad \text{非线性SE(3)}$$

$$z_{ij}^c = h^c(x_i, x_j) \oplus \varepsilon_{ij}^c$$

Sensor model



后端

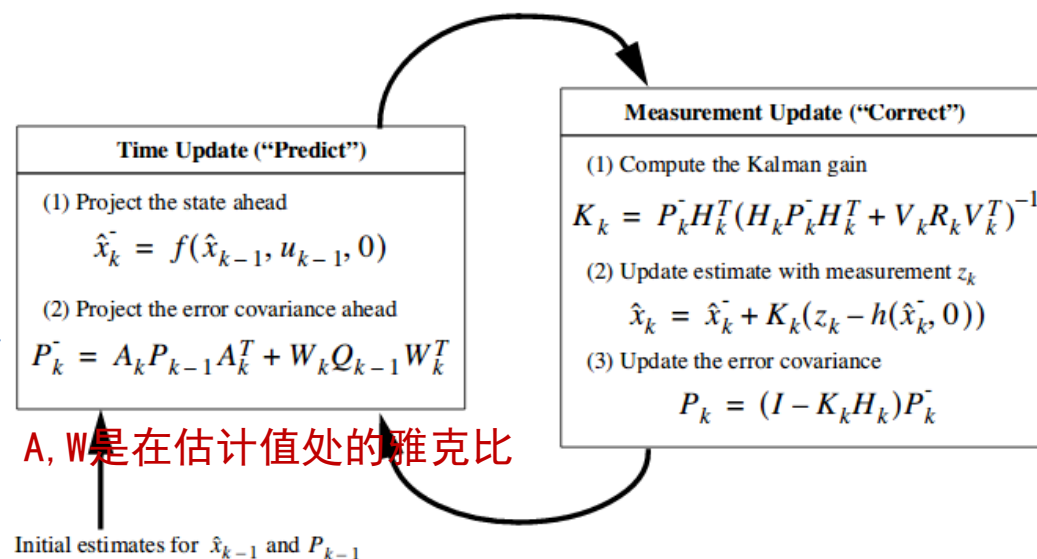
滤波方法

$$x_i = f(x_{i-1}, u_i, \varepsilon_i^u)$$

$$z_{i,j}^v = h^v(x_i, l_j, K) + \varepsilon_{i,j}^v$$

$$X = [x_i, l_1, \dots, l_j]$$

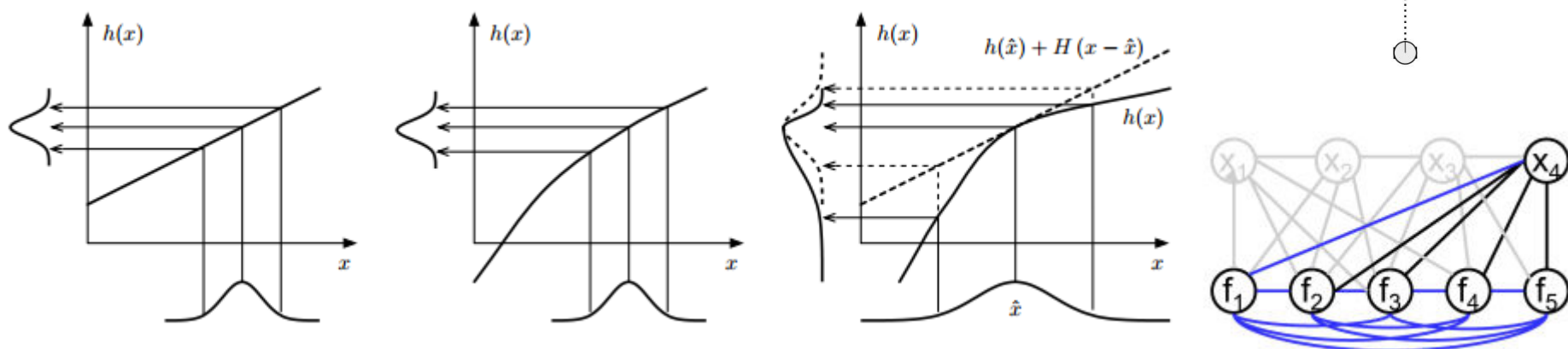
需要维护当前姿态以及观测到 landmark



Event	SLAM	EKF
Robot moves	Robot motion	EKF prediction
Sensor detects new landmark	Landmark initialization	State augmentation
Sensor observes known landmark	Map correction	EKF correction
Mapped landmark is corrupted	Landmark deletion	State reduction

后端

滤波方法



1. 误差（高斯分布）经非线性变换后并非高斯分布。
2. 系统线性化过程中，丢掉了高阶项。
3. 线性化的**工作点**往往不是状态真实的均值，而是一个估计的均值。IEKF
4. EKF的更新后的均值和协方差并不能代表真实的分布。UKF、PF（计算量大）
5. EKF的工作在上一次估计值处。Delayed state KF

后端

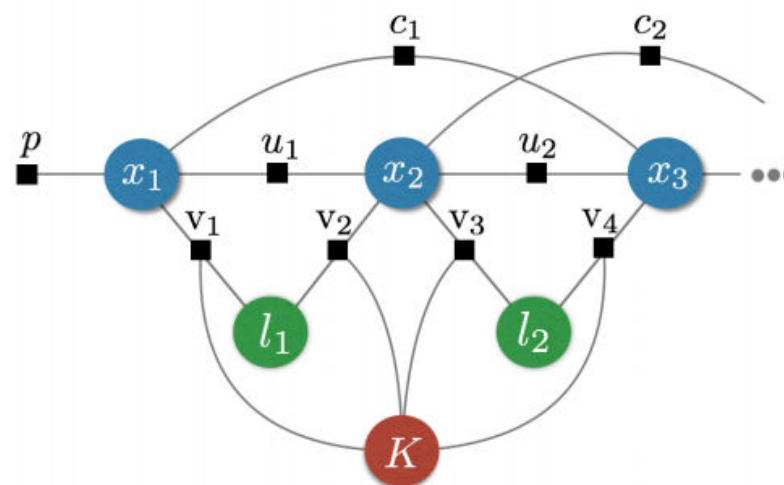
优化方法

$$z_i^u = h^u(x_{i-1}, x_i) \oplus \varepsilon_i^u$$

$$z_{i,j}^v = h^v(x_i, l_j, K) + \varepsilon_{i,j}^v$$

$$z_i^p = x_i \oplus \varepsilon_i^p$$

$$z_{ij}^c = h^c(x_i, x_j) \oplus \varepsilon_{ij}^c$$



最大后验概率MAP

$$\begin{aligned} \chi^* &= \underset{\chi}{\operatorname{argmax}} P(\chi | Z) \\ &= \underset{\chi}{\operatorname{argmax}} P(Z | \chi) P(\chi) \end{aligned}$$

先验未知则退化为极大似然估计MLE

后端

优化方法

最大后验概率MAP

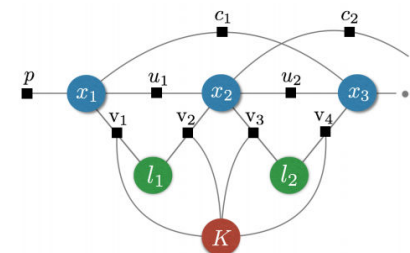
$$\chi^* = \underset{\chi}{\operatorname{argmax}} P(\chi | Z)$$

观测相互独立

$$\propto \underset{\chi}{\operatorname{argmax}} P(Z | \chi) P(\chi)$$

$$\chi^* = \underset{\chi}{\operatorname{argmax}} P(\chi) \prod_{k=1}^m P(z_k | \chi_k)$$

factor



假设服从高斯分布

$$P(z_k | \chi_k) \propto \exp\left(-\frac{1}{2} \|h_k(\chi_k) - z_k\|_{\Omega_k}^2\right) \quad \chi^* = \underset{\chi}{\operatorname{argmin}} -\log \left(P(\chi) \prod_{k=1}^m P(z_k | \chi_k) \right)$$

$$P(\chi_0) \propto \exp\left(-\frac{1}{2} \|h^p(\chi_0) - z_0^p\|_{\Omega_0^p}^2\right)$$

假设只有一个prior factor

$$= \underset{\chi}{\operatorname{argmin}} \sum_{k=0}^m \frac{1}{2} \|h_k(\chi_k) - z_k\|_{\Omega_k}^2 +$$

$$\frac{1}{2} \|h^p(\chi_0) - z_0^p\|_{\Omega_0^p}^2$$

后端

优化方法

非线性最小二乘

$$\chi^* = \operatorname{argmin}_{\chi} \left(\sum_{k=0}^m \|h_k(\chi_k) - z_k\|_{\Omega_k}^2 + \|h^p(\chi_0) - z_0^p\|_{\Omega_0^p}^2 + \dots \right)$$

简单解法

$$\begin{aligned} \delta_{\chi}^* &= \operatorname{argmin}_{\delta_{\chi}} \frac{1}{2} \sum_{k=0}^m \|A_k \delta_{\chi} - b_k\|_{\Omega_k}^2 = \operatorname{argmin}_{\delta_{\chi}} \frac{1}{2} \|A \delta_{\chi} - b\|_{\Omega}^2 \\ \delta_{\chi}^* &= (A^T \Omega A)^{-1} A^T \Omega b \\ \hat{\chi} &\leftarrow \hat{\chi} \oplus \delta_{\chi}^* \end{aligned}$$

- $A^T \Omega A$ 可能非奇异
- 求逆的计算量大

后端

优化方法

非线性最小二乘

$$\delta^* = \operatorname{argmin}_{\delta} \|A\delta - b\|_{\Omega}^2$$

GN

$$(A^T \Omega A) \delta = A^T \Omega b$$

LM

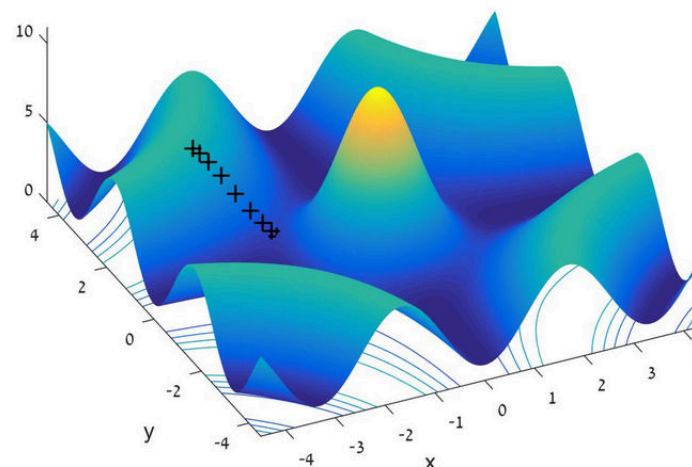
$$(A^T \Omega A + \mu I) \delta = A^T \Omega b$$

DogLeg(最速下降+高斯牛顿)



$$H\delta x = b$$

Linear solver



后端

优化方法

Linear solver $H\delta x = b$

1. Cholesky 分解 $H = LL^T$

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

求解 $Ly = b$ 得到 y

将 y 回代求解 $L^T x = y$ 得到 x

2. QR 分解 $A = Q \begin{bmatrix} R \\ \mathbf{0} \end{bmatrix}$

$$\delta^* = \underset{\delta}{\operatorname{argmin}} \|A\delta - b\|^2$$

$$\begin{aligned} \|A\delta - b\|^2 &= \left\| Q \begin{bmatrix} R \\ \mathbf{0} \end{bmatrix} \delta - b \right\|^2 \\ &= \left\| Q^T Q \begin{bmatrix} R \\ \mathbf{0} \end{bmatrix} \delta - Q^T b \right\|^2 \\ &= \left\| \begin{bmatrix} R \\ \mathbf{0} \end{bmatrix} \delta - \begin{bmatrix} d \\ c \end{bmatrix} \right\|^2 \\ &= \|R\delta - d\|^2 + \|c\|^2 \\ R\delta^* &= d \end{aligned}$$

后端

优化方法 Bundle Adjustment

$$\chi^* = \underset{\chi}{\operatorname{argmin}} \sum_{i,j} \|h(x_i, l_j) - z_{ij}\|_{\Omega_k}^2$$

具有稀疏性

每个相机只观测到部分点

在相机和landmarks较多时仍十分耗时

KeyFrame base, Local BA, Incremental (iSAM)



1. Cholesky 分解 Schur complement

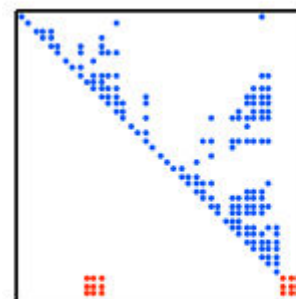
$$\mathbf{X} = \begin{bmatrix} \mathbf{U}_1 & & & & \\ & \mathbf{U}_2 & & & \\ & & \mathbf{U}_3 & & \\ & & & \mathbf{V}_1 & \\ & \mathbf{W}^T & & & \mathbf{V}_2 \\ & & & & & \mathbf{V}_3 \\ & & & & & & \mathbf{V}_4 \end{bmatrix} \quad \mathbf{X} \begin{bmatrix} \Delta(\mathbf{P}^1) \\ \Delta(\mathbf{P}^2) \\ \Delta(\mathbf{P}^3) \\ \Delta(\mathbf{X}_1) \\ \Delta(\mathbf{X}_2) \\ \Delta(\mathbf{X}_3) \\ \Delta(\mathbf{X}_4) \end{bmatrix} = \begin{bmatrix} \varepsilon(\mathbf{P}^1) \\ \varepsilon(\mathbf{P}^2) \\ \varepsilon(\mathbf{P}^3) \\ \varepsilon(\mathbf{X}_1) \\ \varepsilon(\mathbf{X}_2) \\ \varepsilon(\mathbf{X}_3) \\ \varepsilon(\mathbf{X}_4) \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_P \\ \mathbf{x}_L \end{bmatrix}$$

$$\begin{bmatrix} \mathbf{H}_{PP} & \mathbf{H}_{PL} \\ \mathbf{H}_{PL}^T & \mathbf{H}_{LL} \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}_P^* \\ \Delta \mathbf{x}_L^* \end{bmatrix} = - \begin{bmatrix} \mathbf{b}_P \\ \mathbf{b}_L \end{bmatrix}$$

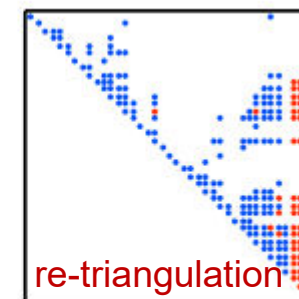
$$\mathbf{S}_{PP} \Delta \mathbf{x}_P^* = -\mathbf{b}_P + \mathbf{H}_{PL} \mathbf{H}_{LL}^{-1} \mathbf{b}_L$$

2. QR 分解 Incremental

新的观测只影响到部分变量，不需要每次都重新构建整个优化问题。



nz = 173



re-triangulation

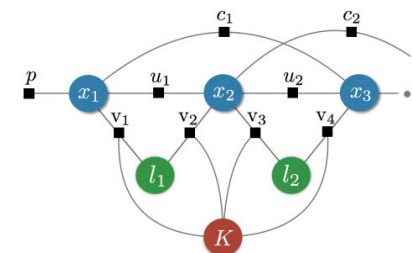
nz = 218

后端

优化方法

非线性最小二乘

$$\chi^* = \underset{\chi}{\operatorname{argmin}} \left(\sum_{k=0}^m \|h_k(\chi_k) - z_k\|_{\Omega_k}^2 + \|h^p(\chi_0) - z_0^p\|_{\Omega_0^p}^2 + \dots \right)$$



优点

- 相对于滤波，使用了更多的信息
- 更容易添加闭环
- 更容易添加各种类型的传感器
- 稀疏的图结构有利于降低计算

缺点

- 需要合理的初值保证收敛
- 很难维护图模型
- 全局优化相比于滤波的方法仍十分耗时

后端

优化方法

存在问题

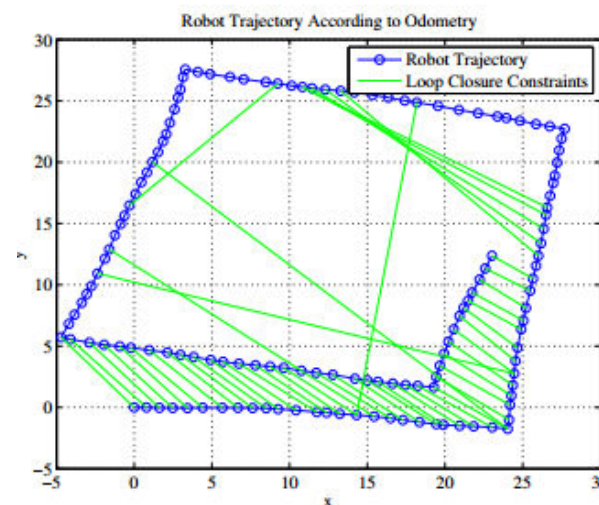
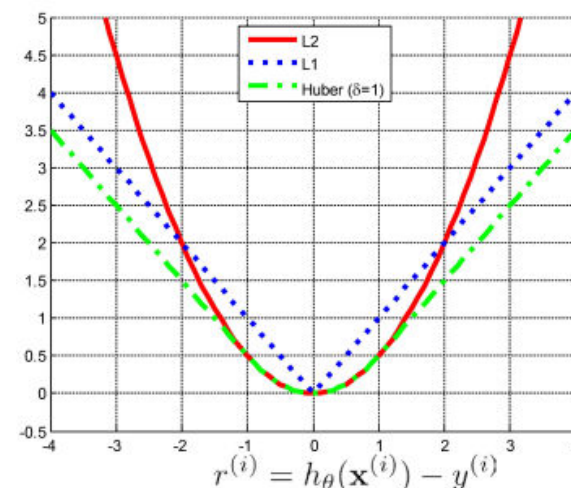
- 图模型中存在错误的观测?
- 鲁棒函数, 动态剔除, Pose graph
- 前端跟踪失败?

测量图的关联性

- 图模型随着时间的增长越来越大
- 长期SLAM, 需要对图进行修剪
- 融合两个图模型?

多机器人SLAM

...



提纲

- 1、背景介绍
- 2、SLAM 理论与方法
- 3、常用开源系统与算法
- 4、总结与展望

4、常用开源系统与算法

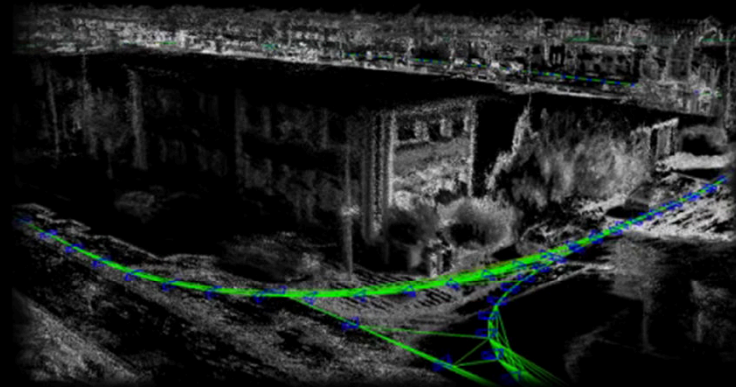
Name	Site
Rgbd-slam-v2	https://github.com/felixendres/rgbdslam_v2
ORB-slam	https://github.com/raulmur/ORB_SLAM
LSD-slam	https://github.com/tum-vision/lsd_slam
Hector-slam	https://github.com/tu-darmstadt-ros-pkg/hector_slam
SVO	https://github.com/uzh-rpg/rpg_svo
R-TAB slam	https://github.com/introlab/rtabmap_ros#rtabmap_ros
DVO-slam	https://github.com/tum-vision/dvo_slam
Kinect Fusion	http://research.microsoft.com/en-us/projects/surfacerecon/
Kinfu_large_scale	http://pointclouds.org/documentation/tutorials/using_kinfu_large_scale.php
DTAM	https://github.com/anuranbaka/OpenDTAM

Large-Scale Direct SLAM with Stereo Cameras



Large-Scale Direct SLAM with Stereo Cameras

Jakob Engel, Jörg Stückler, Daniel Cremers
IROS 2015, Hamburg



Computer Vision Group
Technical University Munich



ElasticFusion Dense SLAM Without A Pose Graph



ElasticFusion: Dense SLAM Without A Pose Graph

Thomas Whelan, Stefan Leutenegger, Renato Salas-Moreno, Ben Glocker, Andrew Davison

Imperial College London

提纲

- 1、背景介绍
- 2、SLAM 理论与方法
- 3、常用开源系统与算法
- 4、总结与展望

总结与展望

Robust-perception age(2015~)

Long Term Autonomy

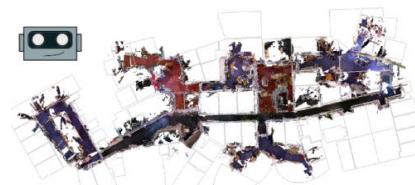
1. 鲁棒性

SLAM系统在任意环境、长时间的正常工作。

- 对算法的鲁棒性（动态环境中数据关联）
- 对硬件的鲁棒性（故障、老化）
- 可度量的重定位(不同视角、天气等)
- 适应季节和地形变化
- 自适应参数调整（特征提取、匹配、关键帧选取、闭环选取）

2. 尺度

- 地图维护（云服务、压缩）
- 分布式构图（多机器人）
- 地图的更新（频率、哪一部分）
- 多平台(计算能力、传感器数量)



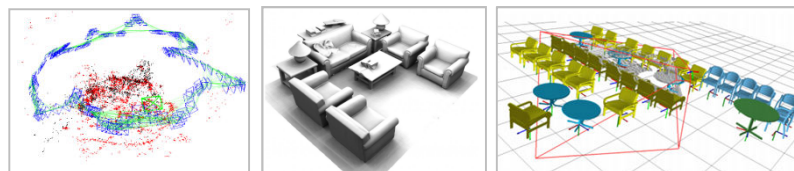
总结与展望

Robust-perception age(2015~) Representation

1. 度量

特征地图表示？稠密地图表示？基于物体的表示？

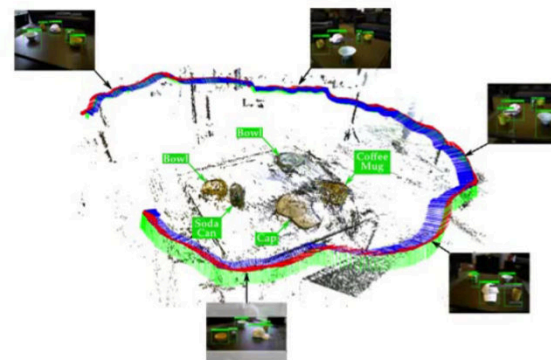
- 高层次，可表达的表示
- 最优表示的判断
- 自适应地选择表示



2. 语义

SLAM和语义相互增强

- 语义构图不仅仅是分类问题
- 根据语义的自主学习与认知
- 基于语义的推断(用于定位和构图)



总结与展望

谢谢观看！

Why is SLAM Difficult?

