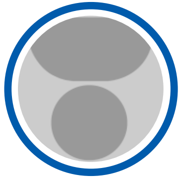
深蓝学院
shenlanxueyuan.com

视觉SLAM开源算法

ORB-SLAM3
原理与代码解析



主讲人 刘国庆

上海交通大学 感知与导航研究所 科研助理

研究方向：多模态定位，动态场景VSLAM，语义SLAM

Github: https://github.com/DreamWaterFound/
知乎：刘国庆
Email：guoqing_liu@sjtu.edu.cn

2021.11.21

课题组简介

郁文贤老师　　**裴凌老师**

吴奇

- 泡泡机器人源代码组组长
- 基于激光/视觉/惯性的紧耦合SLAM算法开发
- robotics_qi@sjtu.edu.cn

李涛

- P3-LOAM作者
- 研究挑战环境下高精度卫星导航/INS/SLAM组合导航算法
- tao_li@sjtu.edu.cn

孙振

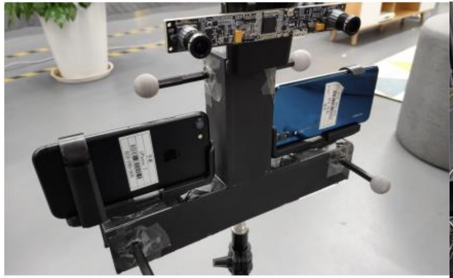
- 面向高效率/低能耗/强鲁棒应用场景的认知导航研究
- 虚拟现实结合的仿真平台研发
- zhensun@sjtu.edu.cn

夏宋鹏程

- 研究Sensor-based 人体行为识别
- 研究状态估计与定位
- songpengchengxia@sjtu.edu.cn

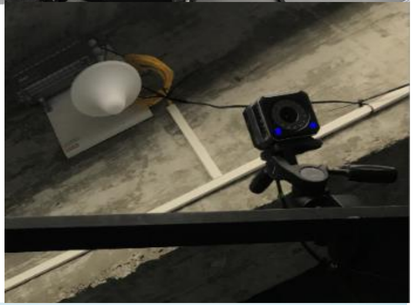
课题组简介

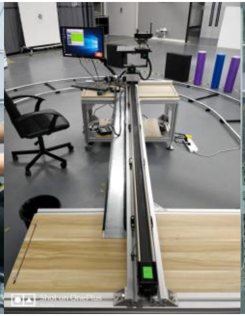
(a) Living area (b) Table area



(a) Typical rig

- NEAR：The NetEase AR Oriented Visual Inertial Dataset
- P3-LOAM: PPP/LiDAR Loosely Coupled SLAM With Accurate Covariance Estimation and Robust RAIM in Urban Canyon Environment
- 泡泡传感器测评





ORB-SLAM历史与基本概念回顾

抽象相机模型

VISLAM实现与IMU初始化

改进的回环检测与多地图融合

总结

ORB-SLAM Series

		ORB-SLAM1	ORB-SLAM2	ORB-SLAM3
时间	arXiv 正式发表	2015(RSS 2014) TRO 2015	2016 TRO 2017	2020 TRO 2021
传感器模态	单目	✓	✓	✓
	双目		✓	✓
	RGB-D		✓	✓
	IMU+单目			✓
	IMU+双目			✓
相机模型	针孔相机成像模型	✓	✓	✓
	鱼眼相机成像模型			✓
地图	单地图	✓	✓	✓
	Altas多地图			✓
追踪线程	Close & Far Points		✓	✓
	VI-SLAM			✓
回环检测线程	Essential Graph优化	✓	✓	✓
	Full Global BA		✓	✓
	Welding BA			✓

ORB-SLAM3 Abstract

- 支持视觉-惯性融合

Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM3 Abstract

- 支持视觉-惯性融合
- 多地图

Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM3 Abstract

- 支持视觉-惯性融合
- 多地图
- 支持针孔+鱼眼相机模型

Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM3 Abstract

- 支持视觉-惯性融合
- 多地图
- 支持针孔+鱼眼相机模型

- 基于MAP的VI-SLAM和IMU初始化

Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM3 Abstract

- 支持视觉-惯性融合
- 多地图
- 支持针孔+鱼眼相机模型

- 基于MAP的VI-SLAM和IMU初始化
- 基于改善后场景识别的多地图SLAM

Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM3 Abstract

- 支持视觉-惯性融合
- 多地图
- 支持针孔+鱼眼相机模型

- 基于MAP的VI-SLAM和IMU初始化
- 基于改善后场景识别的多地图SLAM

- 第一个在所有阶段均可重用先前信息的SLAM

Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM3 Abstract

- 支持视觉-惯性融合
- 多地图
- 支持针孔+鱼眼相机模型

- 基于MAP的VI-SLAM和IMU初始化
- 基于改善后场景识别的多地图SLAM

- 第一个在所有阶段均可重用先前信息的SLAM

- 实验结果：SOTA

Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM3 Abstract

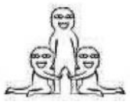
- 支持视觉-惯性融合
- 多地图
- 支持针孔+鱼眼相机模型

- 基于MAP的VI-SLAM和IMU初始化
- 基于改善后场景识别的多地图SLAM

- 第一个在所有阶段均可重用先前信息的SLAM

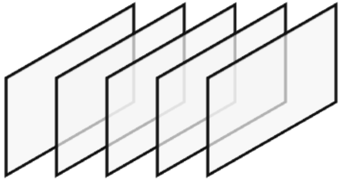
- 实验结果：SOTA

- 好消息：开源



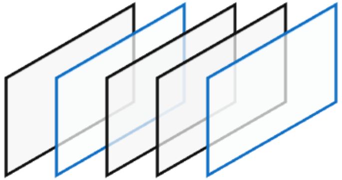
Abstract—This article presents ORB-SLAM3, the first system able to perform visual, visual-inertial and multimap SLAM with monocular, stereo and RGB-D cameras, using pin-hole and fisheye lens models. The first main novelty is a tightly integrated visual-inertial SLAM system that fully relies on maximum *a posteriori* (MAP) estimation, even during IMU initialization, resulting in real-time robust operation in small and large, indoor and outdoor environments, being two to ten times more accurate than previous approaches. The second main novelty is a multiple map system relying on a new place recognition method with improved recall that lets ORB-SLAM3 survive to long periods of poor visual information: when it gets lost, it starts a new map that will be seamlessly merged with previous maps when revisiting them. Compared with visual odometry systems that only use information from the last few seconds, ORB-SLAM3 is the first system able to reuse in all the algorithm stages all previous information from high parallax co-visible keyframes, even if they are widely separated in time or come from previous mapping sessions, boosting accuracy. Our experiments show that, in all sensor configurations, ORB-SLAM3 is as robust as the best systems available in the literature and significantly more accurate. Notably, our stereo-inertial SLAM achieves an average accuracy of 3.5 cm in the EuRoC drone and 9 mm under quick hand-held motions in the room of TUM-VI dataset, representative of AR/VR scenarios. For the benefit of the community we make public the source code.

ORB-SLAM Basic Concepts



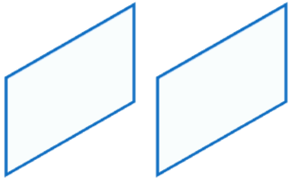
帧

ORB-SLAM Basic Concepts

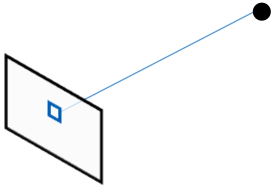


帧 & 关键帧

ORB-SLAM Basic Concepts

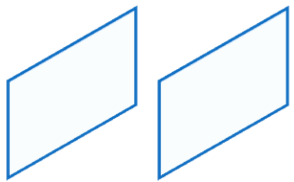


帧 & 关键帧

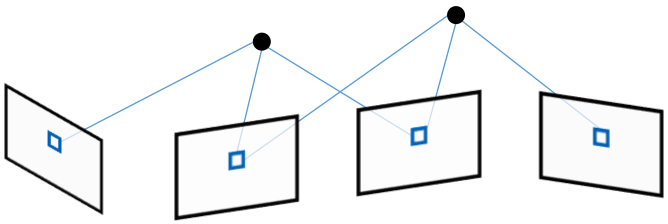


地图点

ORB-SLAM Basic Concepts

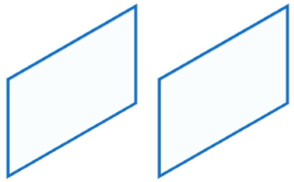


帧 & 关键帧

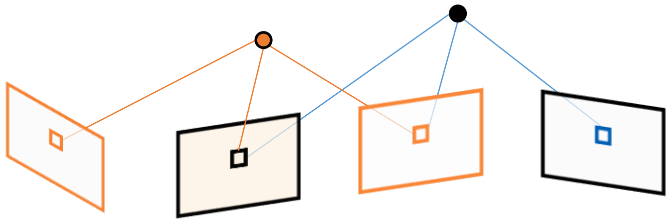


地图点

ORB-SLAM Basic Concepts

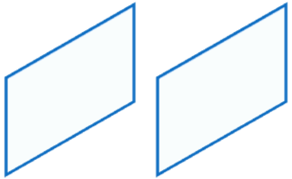


帧 & 关键帧

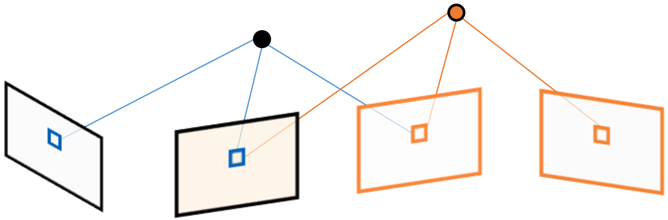


地图点 & 共视

ORB-SLAM Basic Concepts

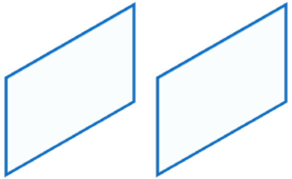


帧 & 关键帧

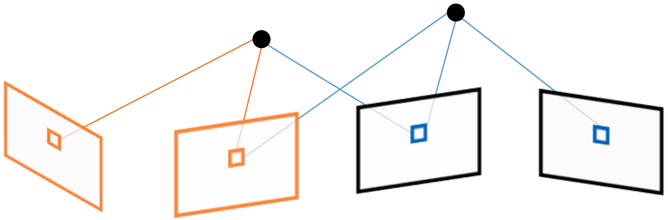


地图点 & 共视

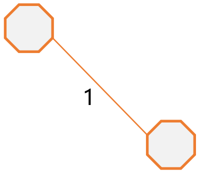
ORB-SLAM Basic Concepts



帧 & 关键帧

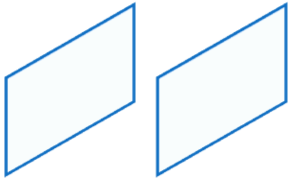


地图点 & 共视

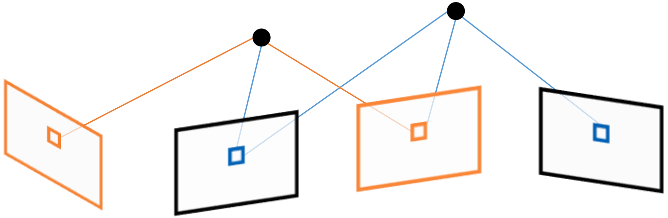


共视图

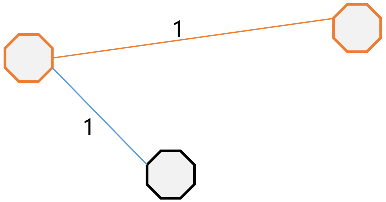
ORB-SLAM Basic Concepts



帧 & 关键帧

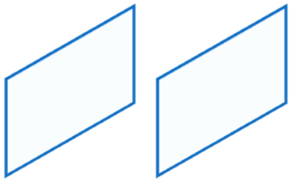


地图点 & 共视

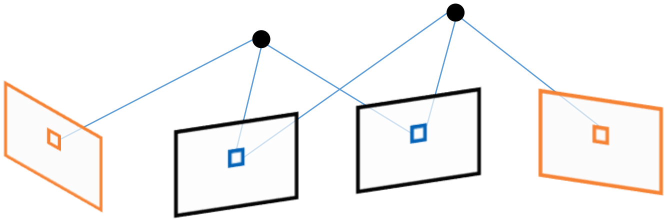


共视图

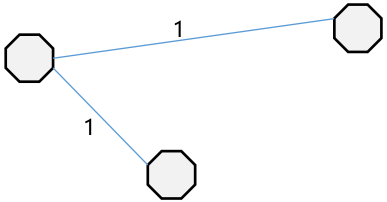
ORB-SLAM Basic Concepts



帧 & 关键帧

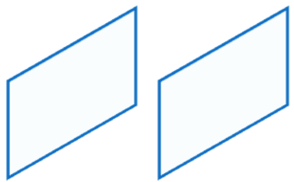


地图点 & 共视

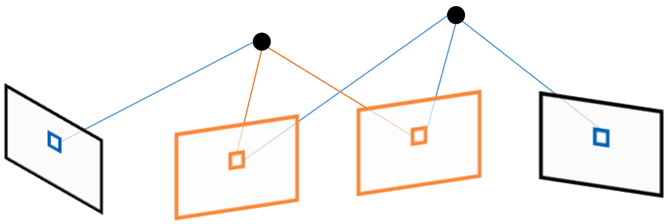


共视图

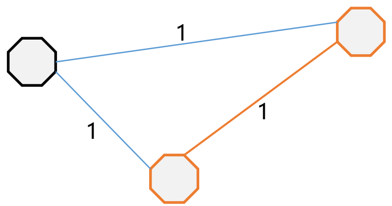
ORB-SLAM Basic Concepts



帧 & 关键帧

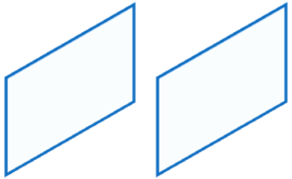


地图点 & 共视

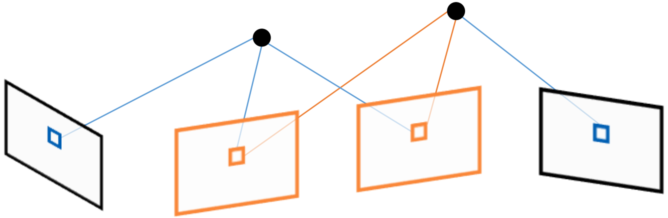


共视图

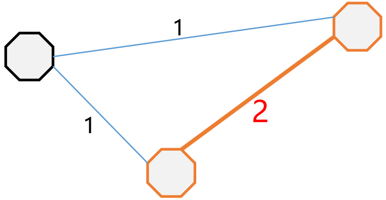
ORB-SLAM Basic Concepts



帧 & 关键帧

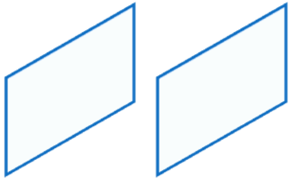


地图点 & 共视

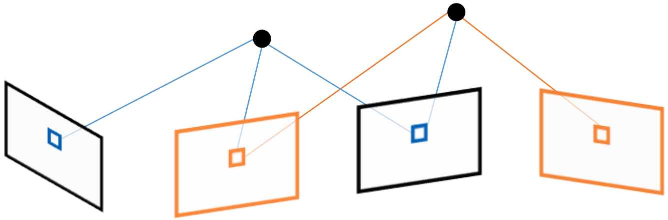


共视图

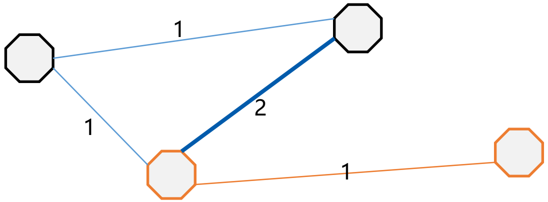
ORB-SLAM Basic Concepts



帧 & 关键帧

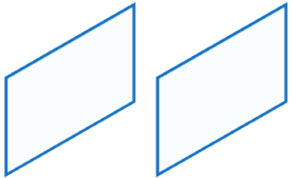


地图点 & 共视

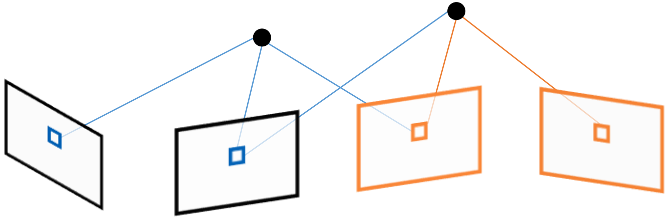


共视图

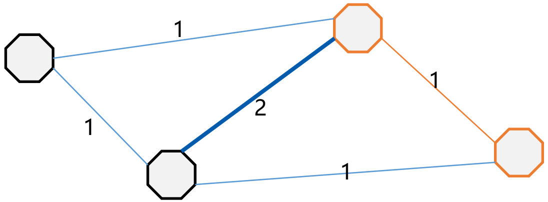
ORB-SLAM Basic Concepts



帧 & 关键帧

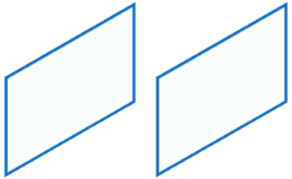


地图点 & 共视

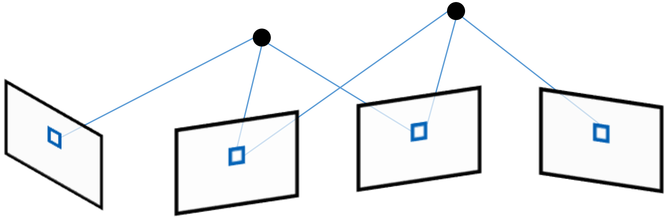


共视图

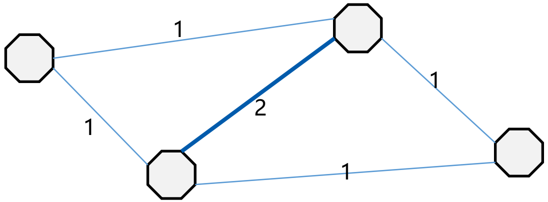
ORB-SLAM Basic Concepts



帧 & 关键帧

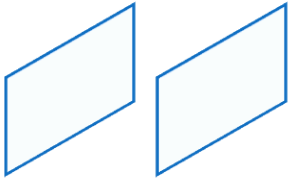


地图点 & 共视

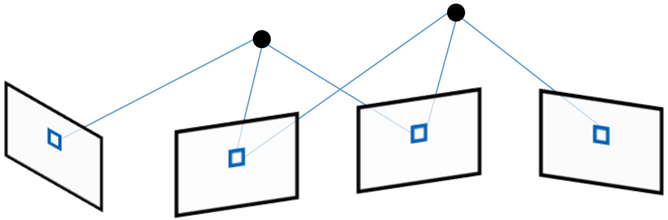


共视图

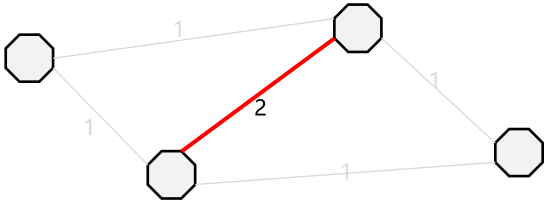
ORB-SLAM Basic Concepts



帧 & 关键帧



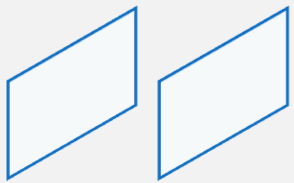
地图点 & 共视



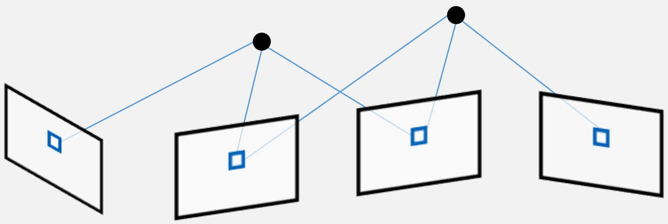
共视图 & 本质图

ORB-SLAM Basic Concepts

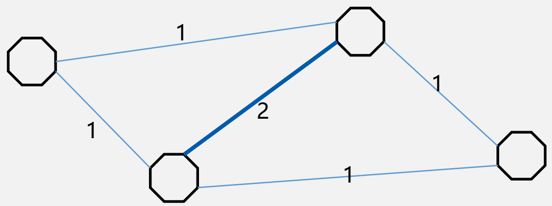
地图



帧 & 关键帧



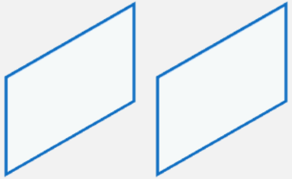
地图点 & 共视



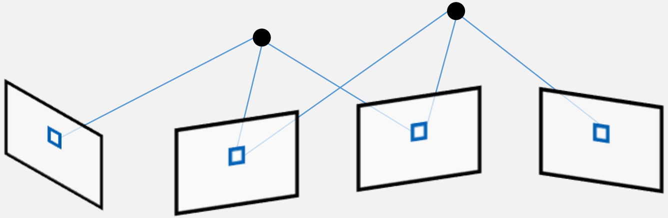
共视图 & 本质图

ORB-SLAM Basic Concepts

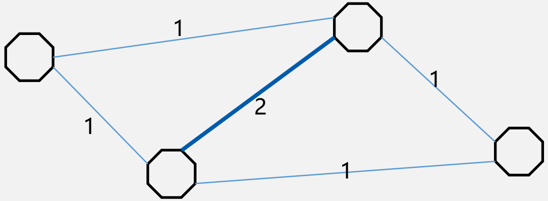
地图



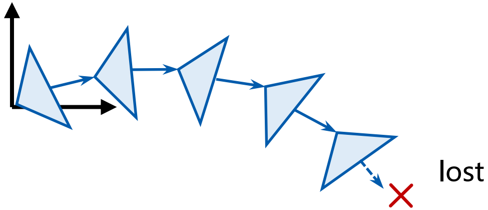
帧 & 关键帧



地图点 & 共视



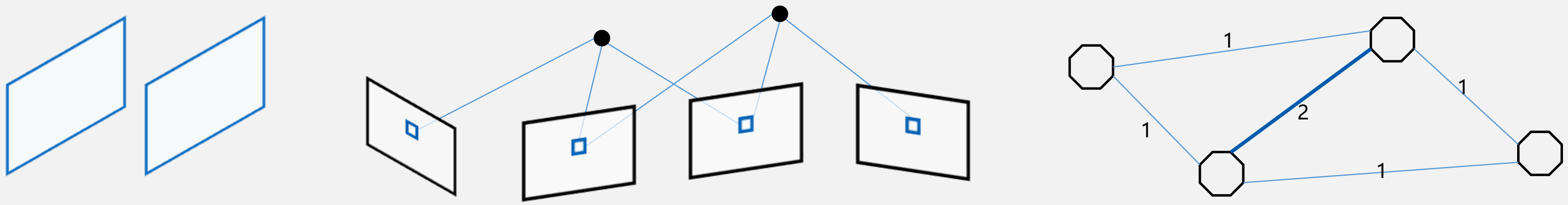
共视图 & 本质图



Altas（地图集）

ORB-SLAM Basic Concepts

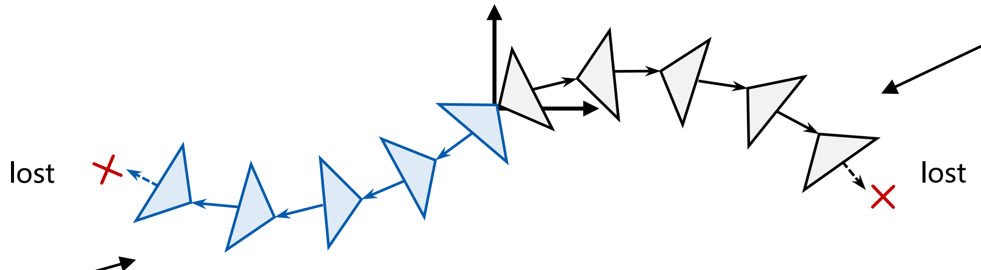
地图



帧 & 关键帧

地图点 & 共视

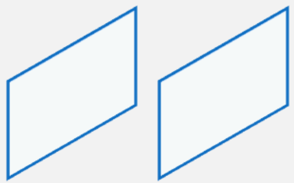
共视图 & 本质图



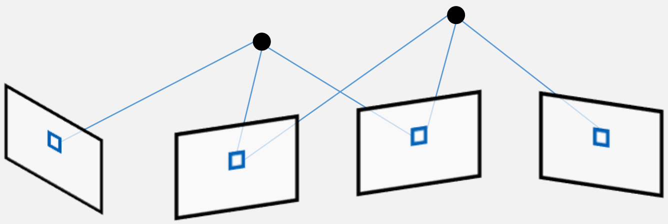
Altas（地图集）

ORB-SLAM Basic Concepts

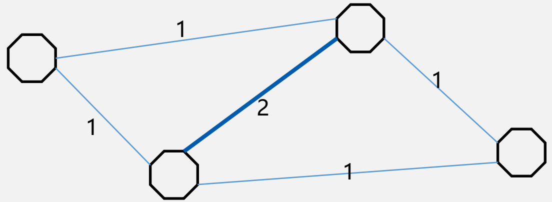
地图



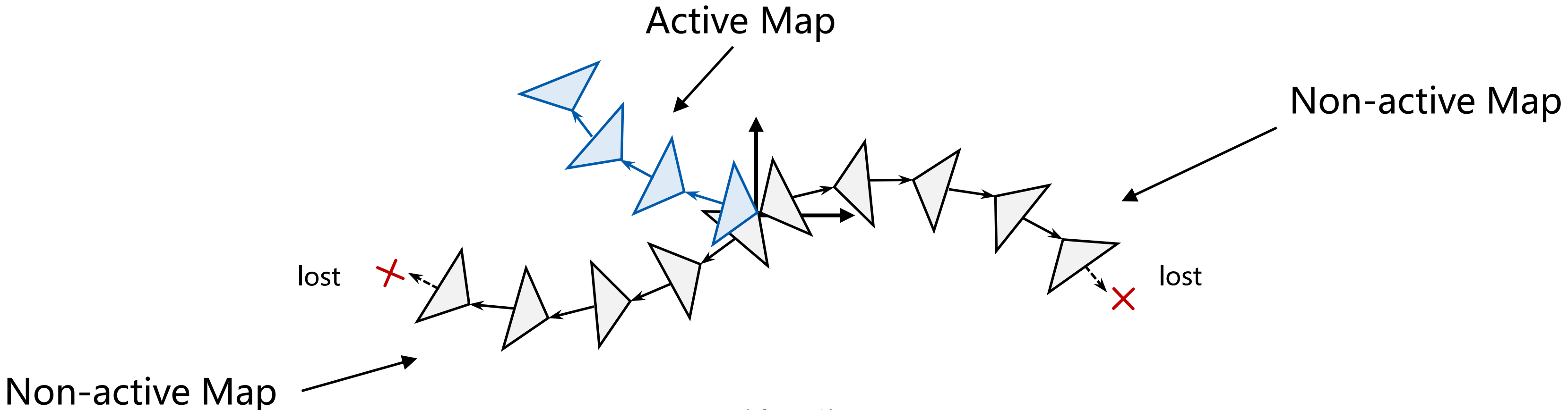
帧 & 关键帧



地图点 & 共视



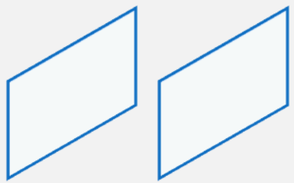
共视图 & 本质图



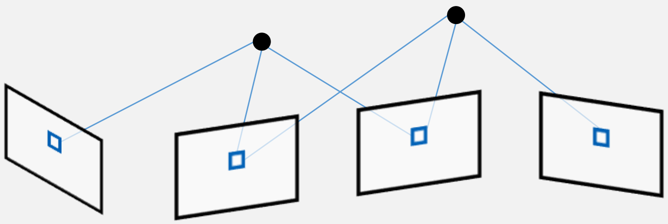
Altas（地图集）

ORB-SLAM Basic Concepts

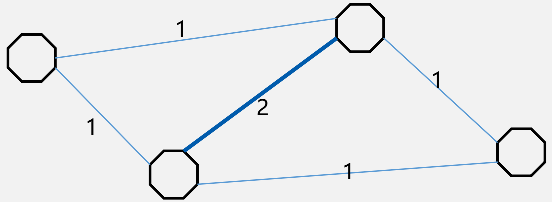
地图



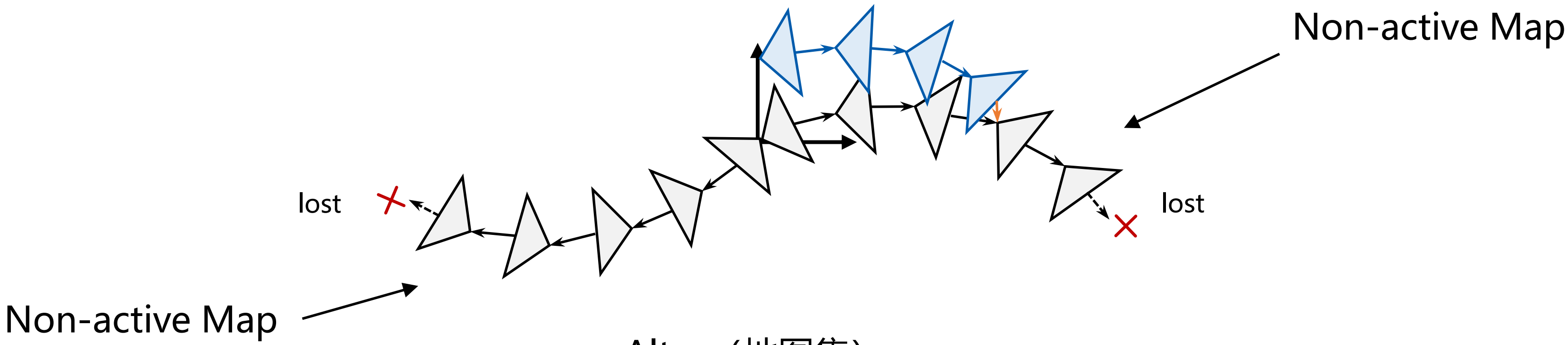
帧 & 关键帧



地图点 & 共视



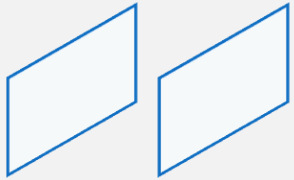
共视图 & 本质图



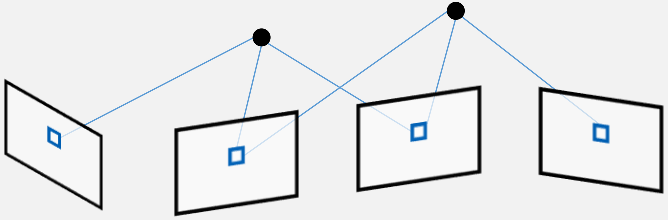
Altas（地图集）

ORB-SLAM Basic Concepts

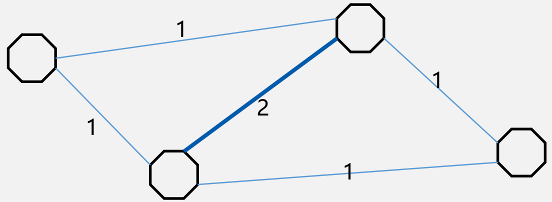
地图



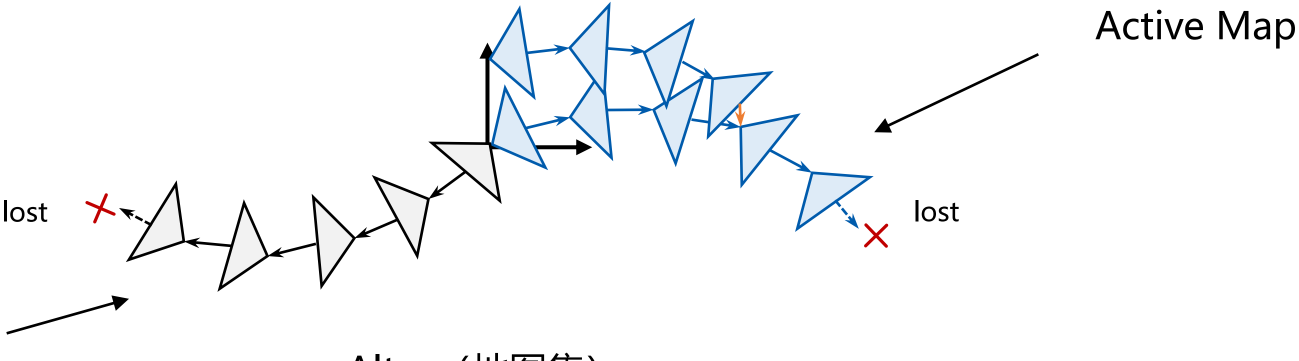
帧 & 关键帧



地图点 & 共视



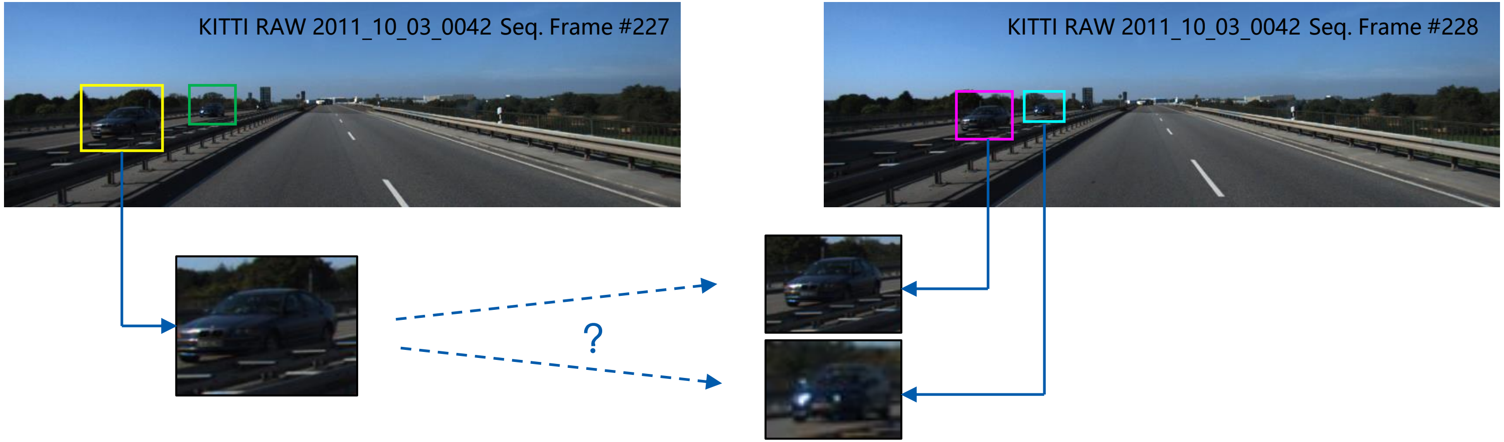
共视图 & 本质图



Altas（地图集）

ORB-SLAM3 Data Association (I)

什么是数据关联（Data Association）？

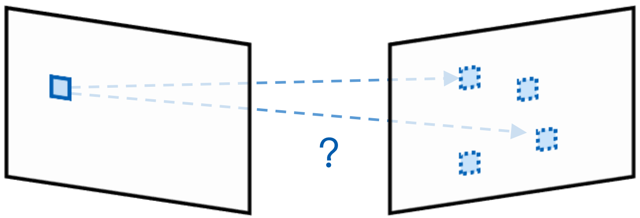


和#228中哪辆车是同一辆车?

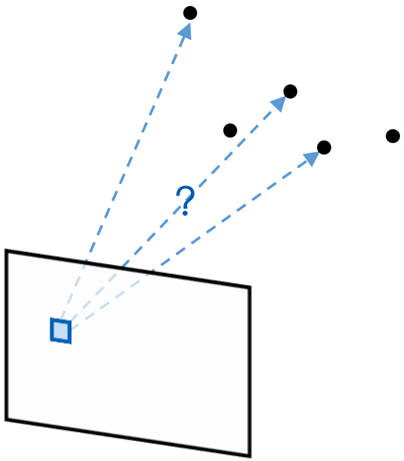
和#247中哪个标志是同一个标志?

ORB-SLAM3 Data Association (I)

什么是数据关联（Data Association）？



和哪个特征是匹配的?



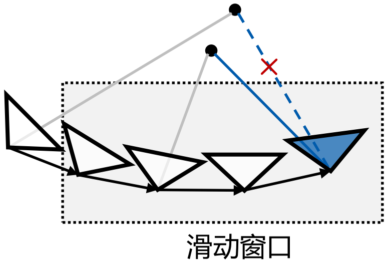
是哪个的路标点的观测?

确定上述关系的过程称为数据关联。

ORB-SLAM3 Data Association (I)

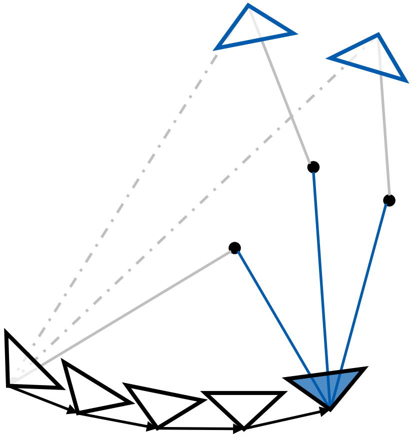
短期

1) **Short-term data association**: matching map elements obtained during the last few seconds. This is the only data association type used by most VO systems, which forget environment elements once they get out of view, resulting in continuous estimation drift even when the system moves in the same area.



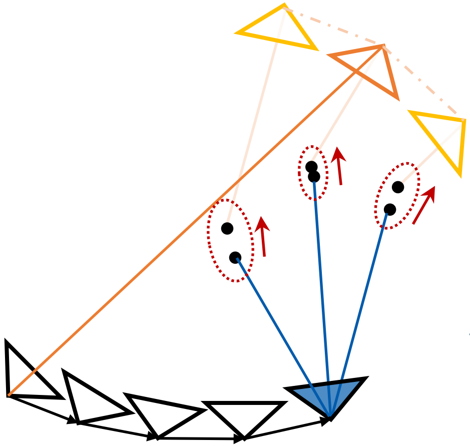
中期

2) **Mid-term data association**: matching map elements that are close to the camera whose accumulated drift is still small. These can be matched and used in BA in the same way than short-term observations and allow to reach zero drift when the systems move in mapped areas. They are the key to the better accuracy obtained by our system compared against VO systems with loop detection.



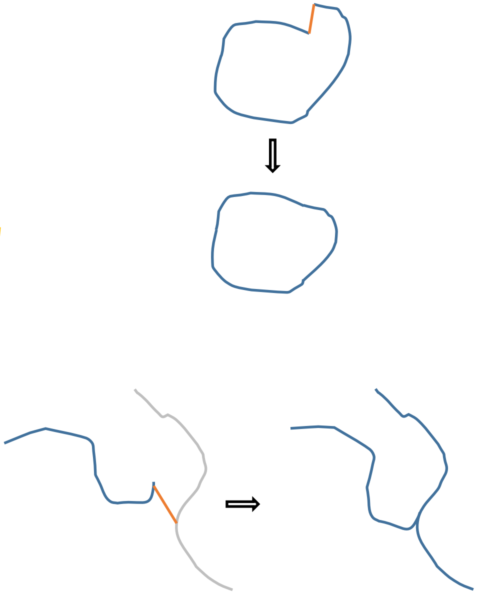
长期

3) **Long-term data association**: matching observations with elements in previously visited areas using a place recognition technique, regardless of the accumulated drift (loop detection), the current area being previously mapped in a disconnected map (map merging), or the tracking being lost (relocalization). Long-term matching allows to reset the drift and to correct the map using pose-graph (PG) optimization or, more accurately, using BA. This is the key to SLAM accuracy in medium and large loopy environments.



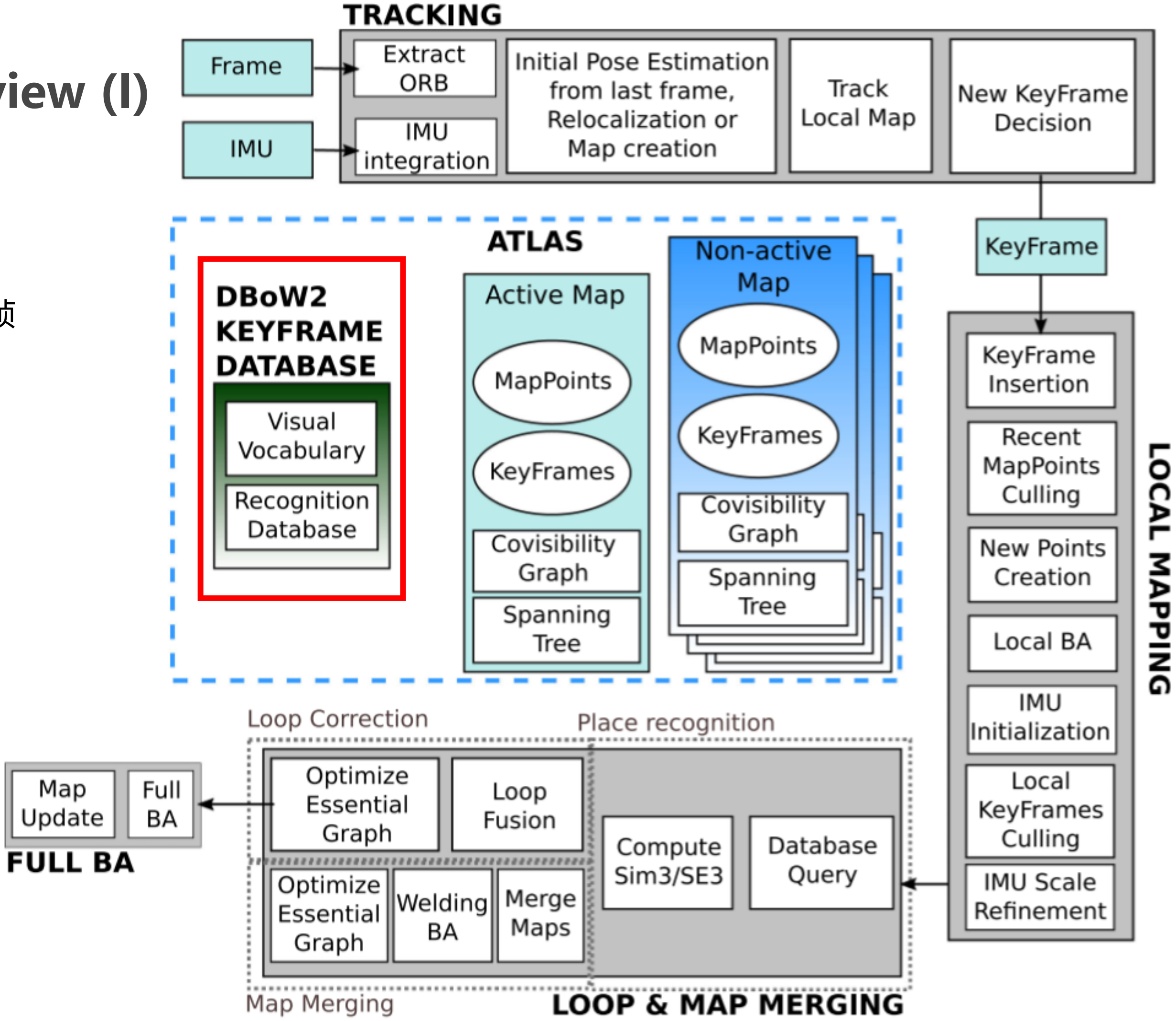
多地图

term data association, reaching zero drift in mapped areas. Here, we go one step further providing **multimap data association**, which allows us to match and use in BA map elements coming from previous mapping sessions, achieving the true goal of a SLAM system: building a map that can be used later to provide accurate localization.



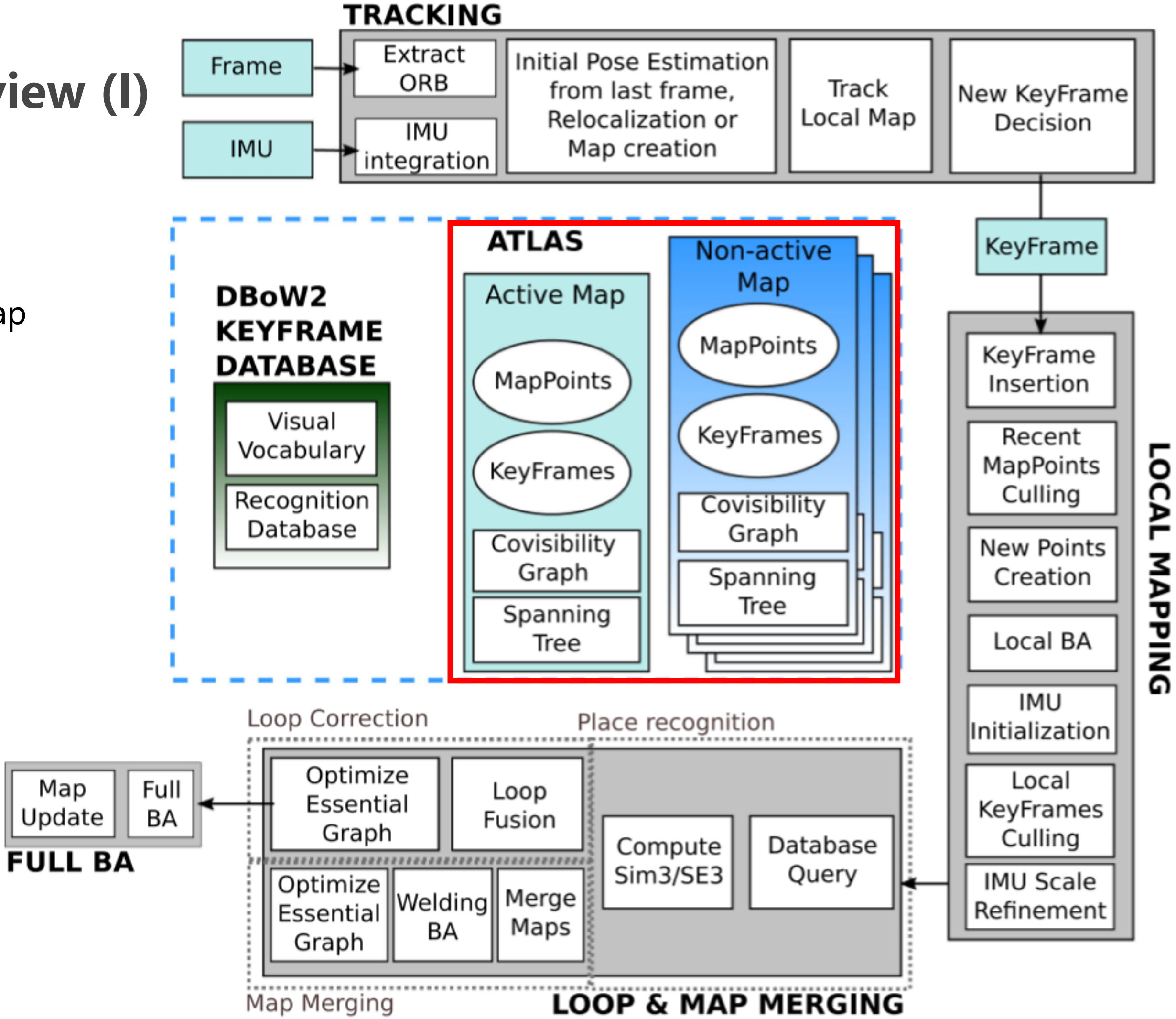
ORB-SLAM3 Overview (I)

- 为每一个关键帧计算词袋向量
- 在数据库中索引相似度高的关键帧



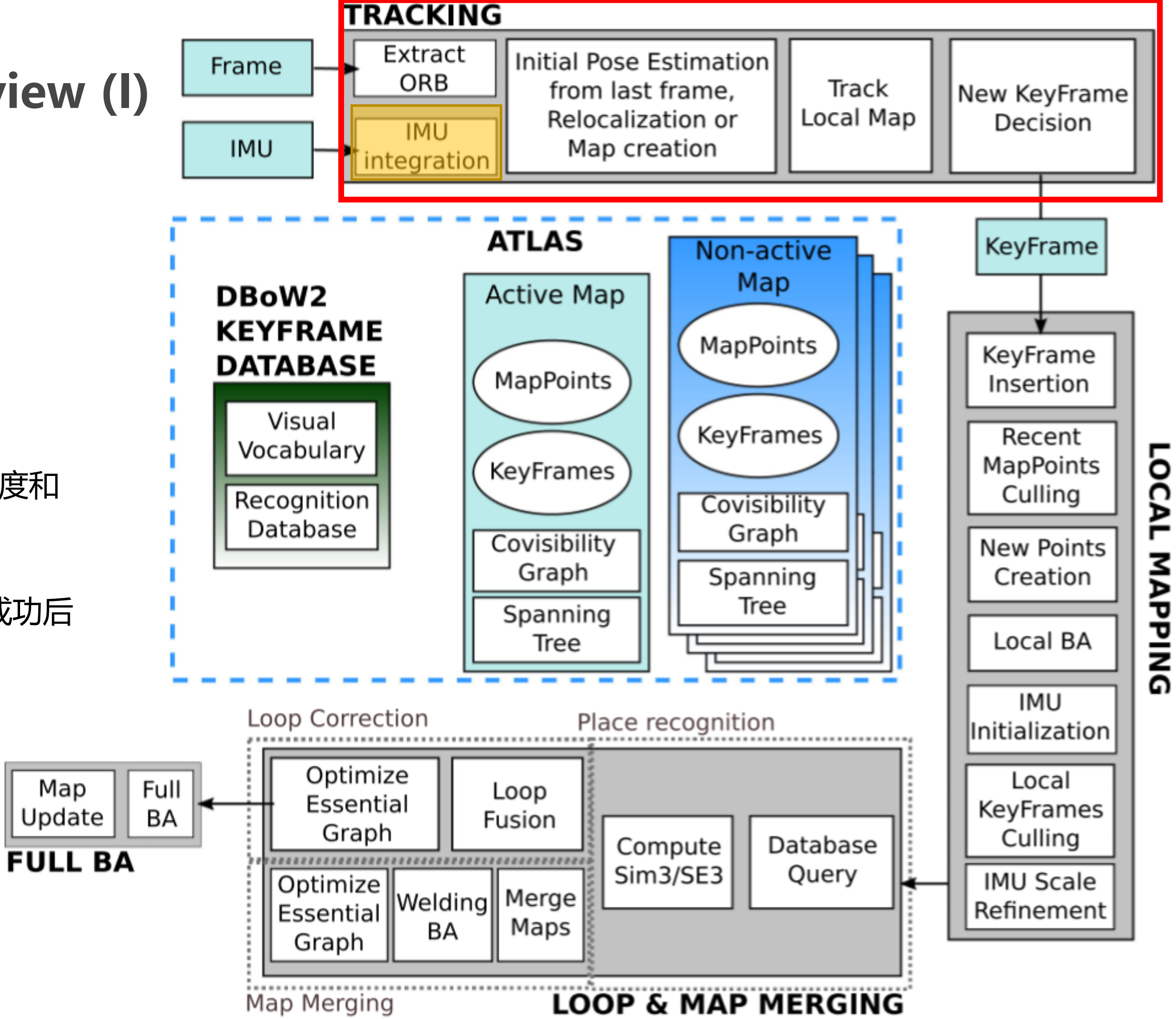
ORB-SLAM3 Overview (I)

- 多个地图
- Tracking线程使用的为Active Map



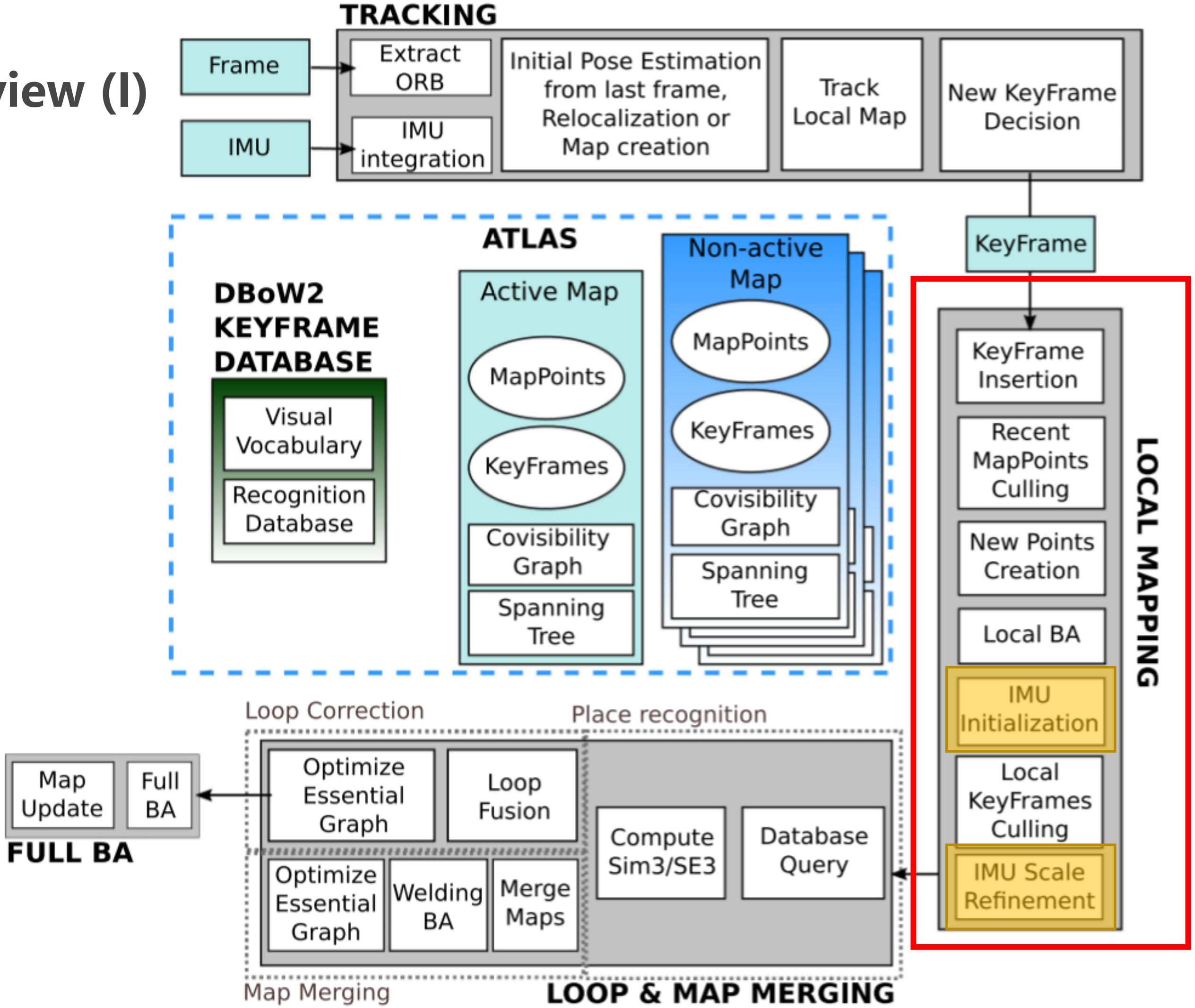
ORB-SLAM3 Overview (I)

- +IMU预积分
- 后续LocalMapping、LoopAndMapMerging线程只处理关键帧
- V+I 输入下估计Body系位姿、速度和IMU bias
- Tracking跟丢、Relocalization 成功后可能切换Active Map



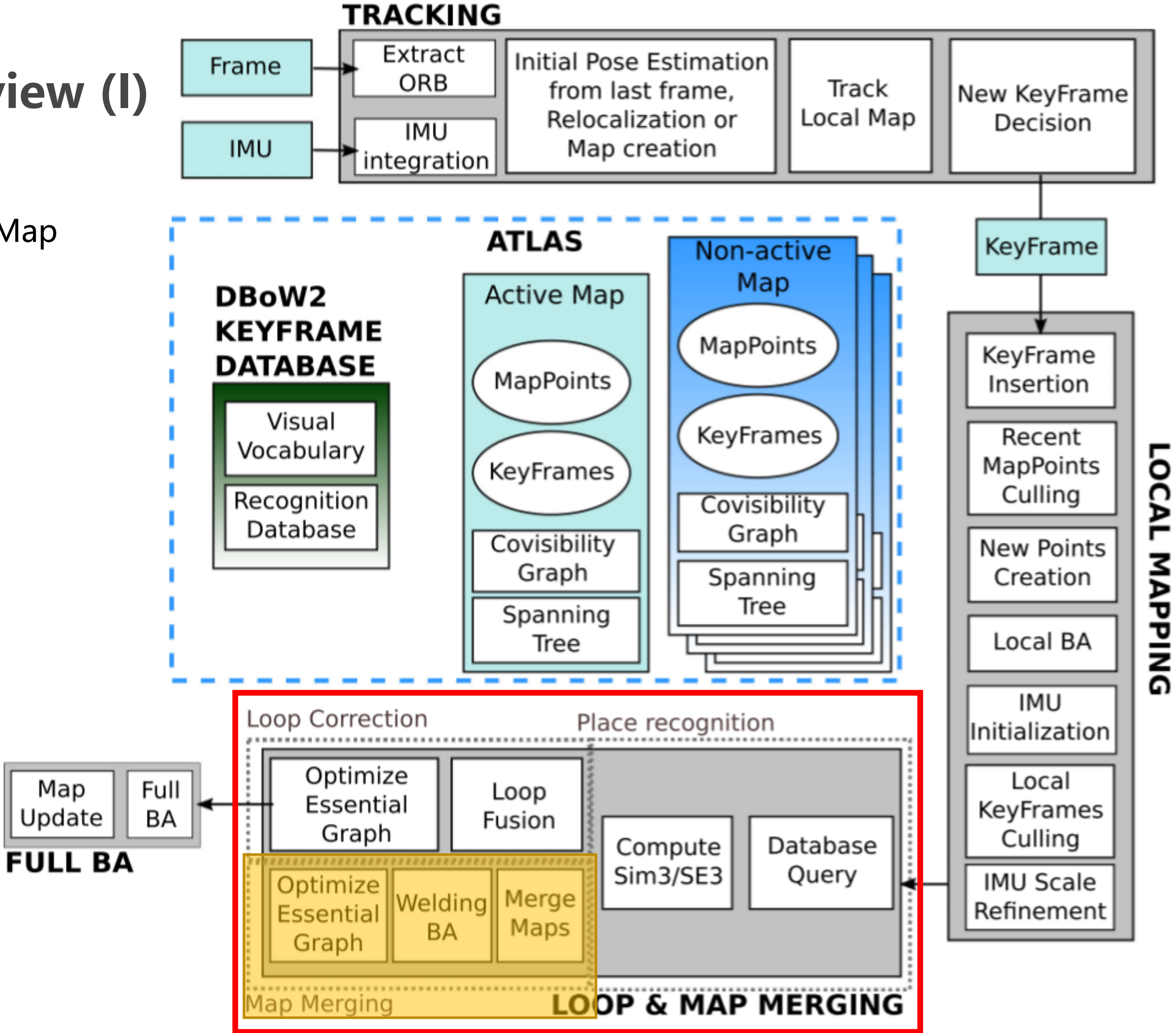
ORB-SLAM3 Overview (I)

- +IMU初始化
- +IMU辅助下的尺度求精



ORB-SLAM3 Overview (I)

- +多地图融合，融合后为 Active Map



ORB-SLAM3 Contributions & Novelties (I)

- **A monocular and stereo visual–inertial SLAM system**

  单双目VI-SLAM

- **Improved-recall place recognition.**

  改善召回率（查全率）的场景识别技术

- **ORB-SLAM Atlas.**

  多地图

- **An abstract camera representation**

  抽象的相机表示

ORB-SLAM历史与基本概念回顾

抽象相机模型

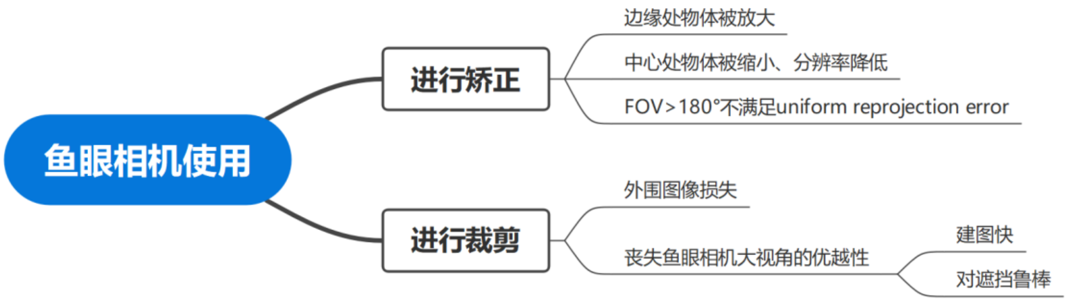
VISLAM实现与IMU初始化

改进的回环检测与多地图融合

总结

Abstract Camera Model (IV.A)

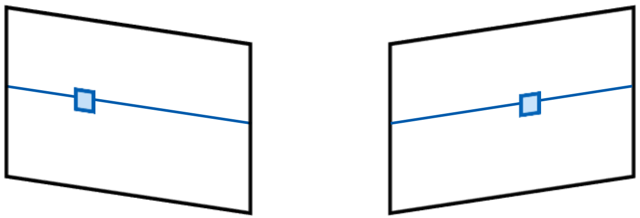
为什么要这样做?



- 相机成像模型与SLAM部分解耦
- 相机成像模型提供投影、反投影、雅可比等函数
- SLAM中原EPnP更换为 MAP-PNP
- 可拓展为任意类型的相机

Abstract Camera Model (IV.B)

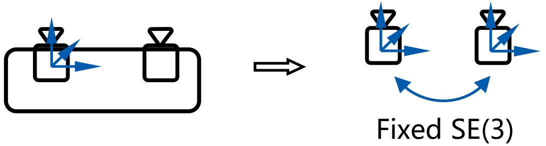
关于双目相机：



极线矫正对两台相机安装要求很高……

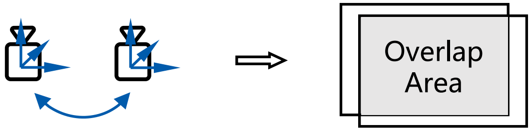
两台相机成像模型和内参需要尽可能相似……

- 将双目“掰开”，看作为两个外参固定的单目



此时等同于对传感器组成的刚体进行位姿估计

- 如果两个相机具有重叠图像区域，则相当于正常的双目相机



Overlapping区域可恢复深度 ✓

无共视区域，也能用 ✓

Visual-Inertial SLAM (V.A) - 问题定义

相比ORB-SLAM-VI的改进：

- 提供快速、准确的IMU初始化
- 支持单双目VI-SLAM
- 支持针孔/鱼眼相机模型

定义VI-SLAM第 i 时刻状态向量（论文式1）：

$$\mathcal{S}_i \doteq \{\underbrace{\mathbf{T}_i}_{\text{位姿}}, \underbrace{\mathbf{v}_i}_{\text{速度}}, \overbrace{\mathbf{b}_i^g}^{\text{加速度计Bias}}, \overbrace{\mathbf{b}_i^a}^{\text{角速度计Bias}}\}.$$

记第i到$i+1$时刻IMU预积分得到的旋转： $\Delta\mathbf{R}_{i,i+1}$ 速度: $\Delta\mathbf{v}_{i,i+1}$ 位置: $\Delta\mathbf{p}_{i,i+1}$ 协方差: $\Sigma_{\mathcal{I}_{i,i+1}}$

定义第i到$i+1$时刻IMU残差项（论文式2）

$$\mathbf{r}_{\mathcal{I}_{i,i+1}} = [\underbrace{\mathbf{r}_{\Delta\mathrm{R}_{i,i+1}}}_{\text{旋转}}, \underbrace{\mathbf{r}_{\Delta\mathrm{v}_{i,i+1}}}_{\text{速度}}, \underbrace{\mathbf{r}_{\Delta\mathrm{p}_{i,i+1}}}_{\text{位置}}]$$

$$\mathbf{r}_{\Delta\mathrm{R}_{i,i+1}} = \mathrm{Log}\left(\Delta\mathbf{R}_{i,i+1}^{\mathrm{T}}\mathbf{R}_i^{\mathrm{T}}\mathbf{R}_{i+1}\right) \qquad \mathrm{Log}:\mathrm{SO}(3)\rightarrow\mathbb{R}^3$$

$$\mathbf{r}_{\Delta\mathrm{v}_{i,i+1}} = \mathbf{R}_i^{\mathrm{T}}\left(\mathbf{v}_{i+1} - \mathbf{v}_i - \mathbf{g}\Delta t_{i,i+1}\right) - \Delta\mathbf{v}_{i,i+1}$$

$$\mathbf{r}_{\Delta\mathrm{p}_{i,i+1}} = \mathbf{R}_i^{\mathrm{T}}\left(\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i\Delta t_{i,i+1} - \frac{1}{2}\mathbf{g}\Delta t^2\right) - \Delta\mathbf{p}_{i,i+1}$$

定义第i到$i+1$时刻第j点视觉残差项（论文式3）：

$$\underbrace{\mathbf{r}_{ij}}_{\text{重投影误差}} = \underbrace{\mathbf{u}_{ij}}_{\text{观测}} - \underbrace{\Pi}_{\text{相机投影函数}}\left(\mathbf{T}_{\mathrm{CB}}\mathbf{T}_i^{-1}\oplus\mathbf{x}_j\right)$$

→ 空间点先变换到第i帧Body系，再转换到第i帧相机坐标系下

Visual-Inertial SLAM (V.A) - 问题定义

定义VI-SLAM第 i 时刻状态向量（论文式1）：

加速度计Bias　　角速度计Bias

$$\mathcal{S}_i \doteq \{\mathbf{T}_i, \mathbf{v}_i, \mathbf{b}_i^g, \mathbf{b}_i^a\}.$$

位姿　速度

定义第i到$i+1$时刻IMU残差项（论文式2）

$$\mathbf{r}_{\mathcal{I}_{i,i+1}} = \left[\mathbf{r}_{\Delta \mathrm{R}_{i,i+1}}, \mathbf{r}_{\Delta \mathrm{v}_{i,i+1}}, \mathbf{r}_{\Delta \mathrm{p}_{i,i+1}}\right]$$

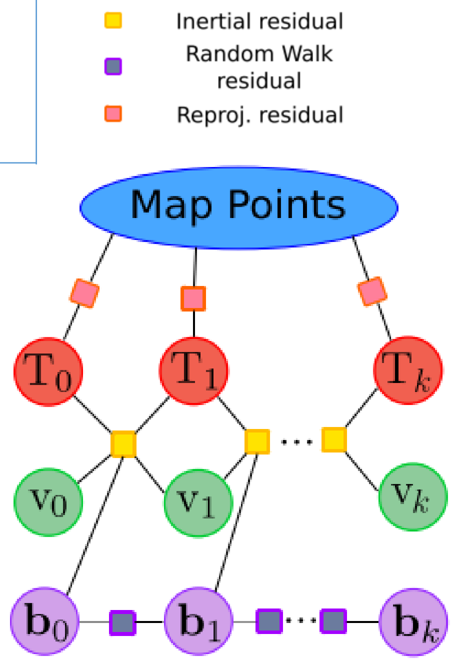
旋转　速度　位置

定义第i到$i+1$时刻第j点视觉残差项（论文式3）：

$$\mathbf{r}_{ij} = \mathbf{u}_{ij} - \Pi\left(\mathbf{T}_{\mathrm{CB}}\mathbf{T}_i^{-1} \oplus \mathbf{x}_j\right)$$

重投影误差　观测　相机投影函数

使用因子图表示：



定义VI-SLAM的优化问题：　（论文式4）

$$\min_{\bar{\mathcal{S}}_k, \mathcal{X}} \left(\sum_{i=1}^{k} \left\| \mathbf{r}_{\mathcal{I}_{i-1,i}} \right\|^2_{\Sigma^{-1}_{\mathcal{I}_{i,i+1}}} + \sum_{j=0}^{l-1} \sum_{i \in \mathcal{K}^j} \rho_{\mathrm{Hub}}\left(\left\| \mathbf{r}_{ij} \right\|_{\Sigma^{-1}_{ij}} \right) \right)$$

所有关键帧和上一帧IMU残差项

所有路标点观测的视觉误差项

视觉的特征点数据关联可能出错，使用鲁棒核函数缓解错误匹配的负面影响

Visual-Inertial SLAM (V.B) - IMU初始化

目的：得到Body系速度、重力方向、IMU bias

作者对该问题的几点认识和思考：

- ORB-SLAM纯单目已经可以初始化得到精确的地图，尺度问题可以通过IMU信息解决
- 已有文献证明，如果将尺度单独作为优化变量进行表示、优化，将比BA中的隐式表达收敛更快
- IMU初始化过程中必须考虑传感器的不确定性，否则会产生难以预测的巨大误差

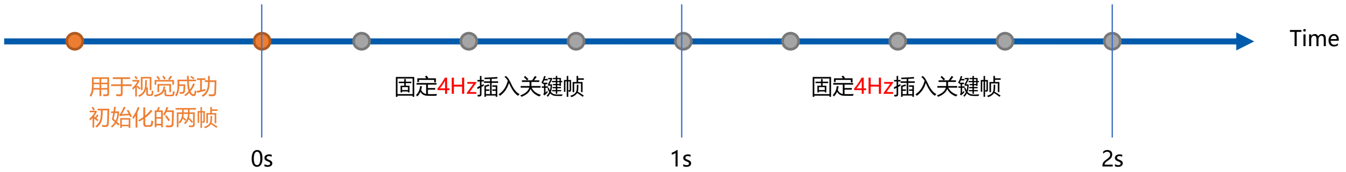
设计了三个IMU初始化步骤：

- 纯视觉MAP估计（Vision-Only MAP Estimation）
- 纯惯性 MAP估计（Inertial-Only MAP Estimation）
- 视觉+惯性 MAP估计（Visual-Inertial MAP Estimation）

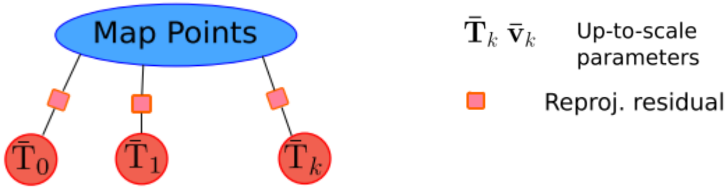
除非特别指明，后面传感器配置均理解为单目+IMU

Visual-Inertial SLAM (V.B) - IMU初始化 - 1/3 Vision-Only MAP Est.

取视觉初始化后续2s内的10帧图像进行纯视觉BA，得到求精后的：相机位姿、路标点



进行纯视觉BA，得到求精后的：相机位姿、路标点。使用因子图表示该过程：



数学符号带上横线表示无尺度

Visual-Inertial SLAM (V.B) - IMU初始化 - 2/3 Inertial-Only MAP Est.

定义IMU初始化阶段的状态向量（论文式5）： $\mathcal{Y}_k = \{s, \mathbf{R}_{wg}, \mathbf{b}, \bar{\mathbf{v}}_{0:k}\}$

IMU初始化期间的帧帧间速度，初始值通过上一阶段位姿得到

显式表示的尺度因子

认为在IMU初始化过程中不变的IMU bias

重力向量到世界坐标系的旋转

记这参与当前阶段初始化的k帧IMU测量为： $\mathcal{I}_{0:k} \doteq \{\mathcal{I}_{0,1} \dots \mathcal{I}_{k-1,k}\}$

则IMU初始化问题可以定义为（论文式6）： $p(\mathcal{Y}_k|\mathcal{I}_{0:k}) \propto p(\mathcal{I}_{0:k}|\mathcal{Y}_k)p(\mathcal{Y}_k)$

贝叶斯公式

有什么样的yk才能使得后验概率最大　已知IMU观测的前提下　似然(likelihood)　先验(prior)

注意IMU每次的观测是独立的，MAP问题可以写为（论文式7）：

$$\mathcal{Y}_k^* = \arg\max_{\mathcal{Y}_k} \left(p(\mathcal{Y}_k) \prod_{i=1}^{k} p(\mathcal{I}_{i-1,i}|s, \mathbf{R}_{wg}, \mathbf{b}, \bar{\mathbf{v}}_{i-1}, \bar{\mathbf{v}}_i) \right)$$

此时y_k^*就是IMU初始化的结果。

？ 但是，有问题：连乘、求最大值，这种优化怎么做？

Visual-Inertial SLAM (V.B) - IMU初始化 - 2/3 Inertial-Only MAP Est.

定义IMU初始化阶段的状态向量（论文式5）：

$$\mathcal{Y}_k = \{s, \mathbf{R}_{wg}, \mathbf{b}, \bar{\mathbf{v}}_{0:k}\}$$

显式表示的尺度因子

重力向量到世界坐标系的旋转

认为在IMU初始化过程中不变的IMU bias

IMU初始化期间的帧帧间速度，初始值通过上一阶段位姿得到

MAP问题可以写为（论文式7）：

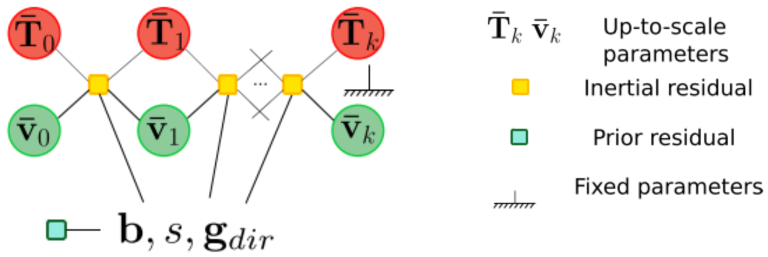
$$\mathcal{Y}_k^* = \arg\max_{\mathcal{Y}_k} \left(p(\mathcal{Y}_k) \prod_{i=1}^{k} p(\mathcal{I}_{i-1,i} | s, \mathbf{R}_{wg}, \mathbf{b}, \bar{\mathbf{v}}_{i-1}, \bar{\mathbf{v}}_i) \right)$$

利用IMU预积分误差服从高斯分布的假设，取负对数转换为“和的最小”问题就好做了（论文式8）：

$$\mathcal{Y}_k^* = \arg\min_{\mathcal{Y}_k} \left(\|\mathbf{b}\|^2_{\Sigma_b^{-1}} + \sum_{i=1}^{k} \|\mathbf{r}_{\mathcal{I}_{i-1,i}}\|^2_{\Sigma^{-1}_{\mathcal{I}_{i-1,i}}} \right).$$

在IMU bias的影响下使IMU误差项最小

因子图表示：



Visual-Inertial SLAM (V.B) - IMU初始化 - 2/3 Inertial-Only MAP Est.

定义IMU初始化阶段的状态向量（论文式5）：

$$\mathcal{Y}_k = \{s, \mathbf{R}_{wg}, \mathbf{b}, \bar{\mathbf{v}}_{0:k}\}$$

IMU初始化期间的帧帧间速度，初始值通过上一阶段位姿得到

显式表示的尺度因子

认为在IMU初始化过程中不变的IMU bias

重力向量到世界坐标系的旋转

MAP问题可以写为（论文式7）：

$$\mathcal{Y}_k^* = \underset{\mathcal{Y}_k}{\arg\max}\left(p(\mathcal{Y}_k)\prod_{i=1}^{k} p(\mathcal{I}_{i-1,i}|s, \mathbf{R}_{wg}, \mathbf{b}, \bar{\mathbf{v}}_{i-1}, \bar{\mathbf{v}}_i)\right)$$

利用IMU预积分误差服从高斯分布的假设，取负对数转换为“和的最小”问题就好做了（论文式8）：

$$\mathcal{Y}_k^* = \underset{\mathcal{Y}_k}{\arg\min}\left(\|\mathbf{b}\|^2_{\Sigma_b^{-1}} + \sum_{i=1}^{k}\|\mathbf{r}_{\mathcal{I}_{i-1,i}}\|^2_{\Sigma^{-1}_{\mathcal{I}_{i-1,i}}}\right).$$

在IMU bias的影响下使IMU误差项最小

新问题：优化过程中，$\boldsymbol{R}_{wg}$的旋转轴如果碰巧和重力向量方向接近重合，则不怎么会改变重力向量方向，此次迭代更新了个寂寞

解决：用两个和重力向量方向正交的旋转进行更新（论文式9），避开绕着重力向量方向的旋转：

$$\mathbf{R}_{\mathrm{wg}}^{\mathrm{new}} = \mathbf{R}_{\mathrm{wg}}^{\mathrm{old}}\mathrm{Exp}(\delta\alpha_{\mathbf{g}}, \delta\beta_{\mathbf{g}}, 0)$$

新问题：优化过程中对尺度因子s的加减乘除很容易导致s出现负数的情况

解决：用指数的方式进行更新，管你更新量 δs 是正是负、是大是小，最终反映在 s 上的更新都是不会改变符号的（论文式10）：

$$s^{\mathrm{new}} = s^{\mathrm{old}}\exp(\delta s).$$

Visual-Inertial SLAM (V.B) - IMU初始化 - 2/3 Inertial-Only MAP Est.

定义IMU初始化阶段的状态向量（论文式5）：

$$\mathcal{Y}_k = \{s, \mathbf{R}_{wg}, \mathbf{b}, \bar{\mathbf{v}}_{0:k}\}$$

IMU初始化期间的帧帧间速度，初始值通过上一阶段位姿得到

显式表示的尺度因子

认为在IMU初始化过程中不变的IMU bias

重力向量到世界坐标系的旋转

…利用IMU预积分误差服从高斯分布的假设，取负对数转换为“和的最小”问题就好做了（论文式8）：

$$\mathcal{Y}_k^* = \underset{\mathcal{Y}_k}{\arg\min} \left(\|\mathbf{b}\|^2_{\Sigma_b^{-1}} + \sum_{i=1}^{k} \|\mathbf{r}_{\mathcal{I}_{i-1,i}}\|^2_{\Sigma^{-1}_{\mathcal{I}_{i-1,i}}} \right).$$

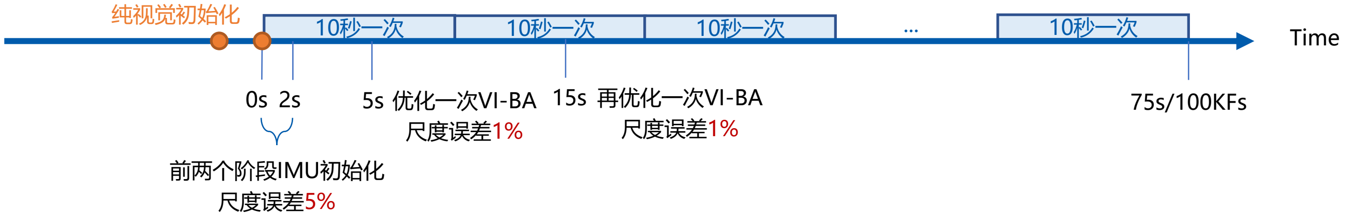
在IMU bias的影响下使IMU误差项最小

一顿操作后，得到的优化结果：

- 帧位姿、速度和地图点位置都具有正确的尺度
- Body系Z轴将被旋转到和重力方向一致
- IMU的Bias被更新

Visual-Inertial SLAM (V.B) - IMU初始化 - 3/3 Visual-Inertial MAP Est.

在第5s和15s再进行Visual-Inertial BA。IMU初始化后的地图点称为mature(成熟的)

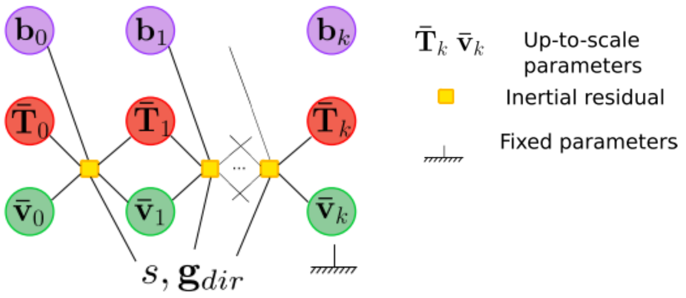


这也是本文IMU初始化速度快并且准确的原因

新问题：传感器运动缓慢，IMU激励不够怎么办？

只进行尺度因子和重力方向的优化（bias不变假设不成立），10s一次

因子图表示：

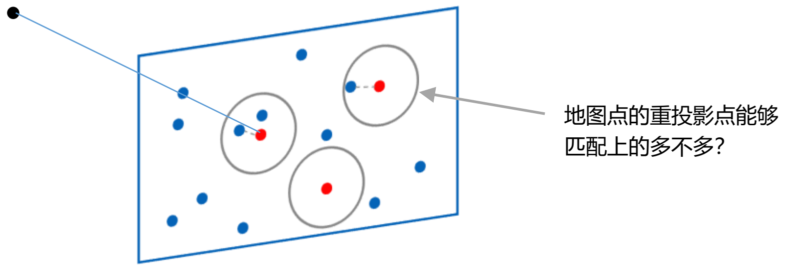


Visual-Inertial SLAM (V.C & V.D)

VI-SLAM的Tracking：和ORB-SLAM-VI相同，仅优化最近两帧，地图点fixed

VI-SLAM的Mapping：为了实时性考虑，设置的滑窗内关键帧和观测到的地图点被优化，共视帧仅提供约束

VI-SLAM处理Track lost：短期跟丢：IMU提供相机位姿估计，地图点重投影搜索窗口变大、尝试匹配



跟丢持续好一段时间了？放弃治疗，重新初始化，创建新的Active Map

ORB-SLAM历史与基本概念回顾

抽象相机模型

VISLAM实现与IMU初始化

改进的回环检测与多地图融合

总结

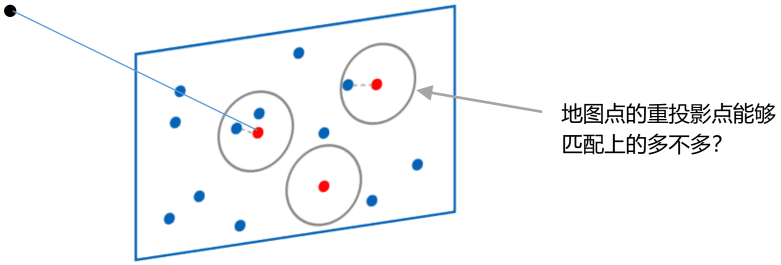
Map Merging & Loop Closing (VI)

- 回顾

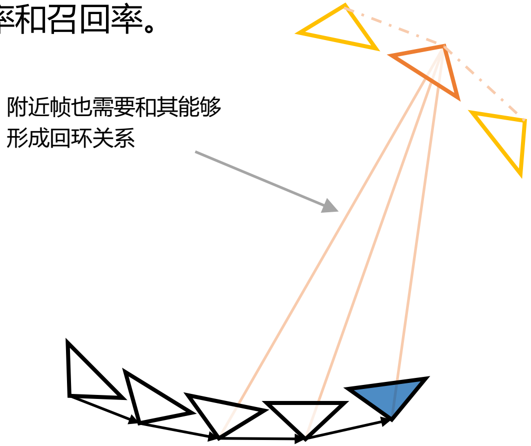
PrecisionRate：准确率/查准率，检测到的回环确实是回环的比率——检测到的**准不准**　　SLAM对它更敏感

RecallRate：召回率/查全率，检测到的回环数占总真实回环数的比率——检测到的**全不全**

ORB-SLAM1/2仅使用DBow词袋数据库即可实现 50%~80% 的准确率和召回率。
增加几何一致性和时间一致性检验，牺牲召回率增加准确率



几何一致性检验



时间一致性检验

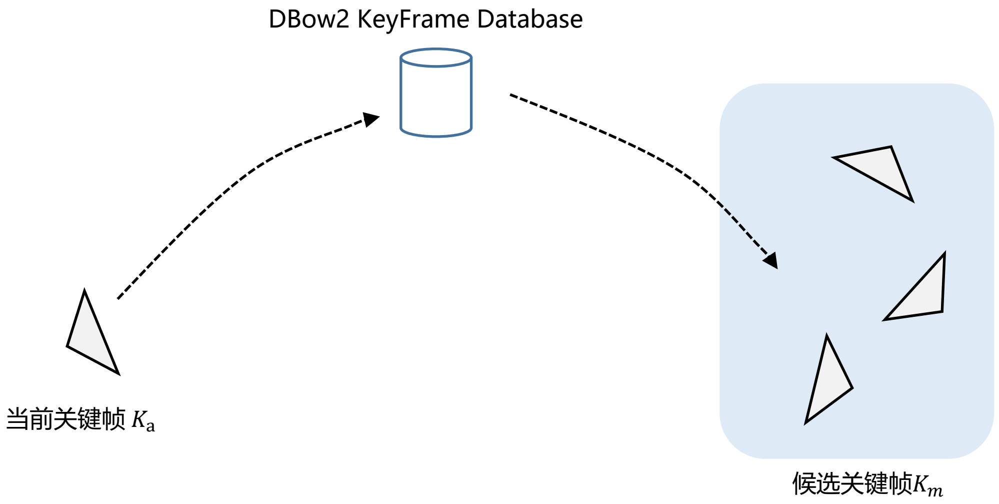
- 改进：更快的时间一致性检验、多地图的融合

Map Merging & Loop Closing (VI.A)

场景识别(1/6)：当前关键帧K_a

- 查询数据库，得到三个最相似的关键帧，记为K_m
- K_m 不是当前关键帧的共视关键帧

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



Map Merging & Loop Closing (VI.A)

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

- 构造Local window:

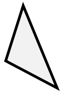
  K_a& K_a共视关键帧 + K_m& K_m共视关键帧 + 它们看到的地图点

- 分别进行2D-2D、3D-3D的点匹配

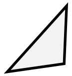
1) **DBoW2 candidate keyframes.**
2) **Local window.**
3) **3-D aligning transformation.**

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

5) **Verification in three covisible keyframes.**
6) **VI gravity direction verification.**



当前关键帧 K_a



一个候选关键帧K_m

Map Merging & Loop Closing (VI.A)

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

- 构造Local window:

  K_a& K_a共视关键帧 + K_m& K_m共视关键帧 + 它们看到的地图点

- 分别进行2D-2D、3D-3D的点匹配

1) DBoW2 candidate keyframes.

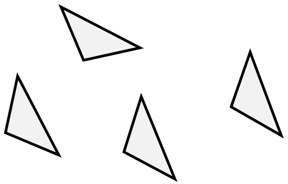
2) Local window.

3) 3-D aligning transformation.

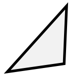
场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

5) Verification in three covisible keyframes.

6) VI gravity direction verification.



当前关键帧 K_a



一个候选关键帧K_m

Map Merging & Loop Closing (VI.A)

场景识别(2/6)：当前关键帧K_{a}，候选关键帧记为K_m

- 构造Local window:

  K_{a}& K_{a}共视关键帧 + K_m& K_m共视关键帧 + 它们看到的地图点

- 分别进行2D-2D、3D-3D的点匹配

1) **DBoW2 candidate keyframes.**

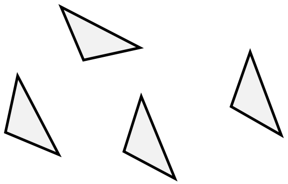
2) **Local window.**

3) **3-D aligning transformation.**

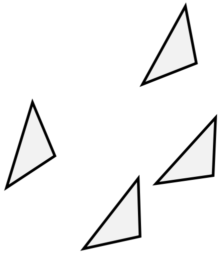
场景识别(2/6)：当前关键帧K_{a}，候选关键帧记为K_m

5) **Verification in three covisible keyframes.**

6) **VI gravity direction verification.**



当前关键帧 K_{a}



一个候选关键帧K_m

Map Merging & Loop Closing (VI.A)

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

- 构造Local window:

  K_a& K_a共视关键帧 + K_m& K_m共视关键帧 + 它们看到的地图点

- 分别进行2D-2D、3D-3D的点匹配

1) DBoW2 candidate keyframes.

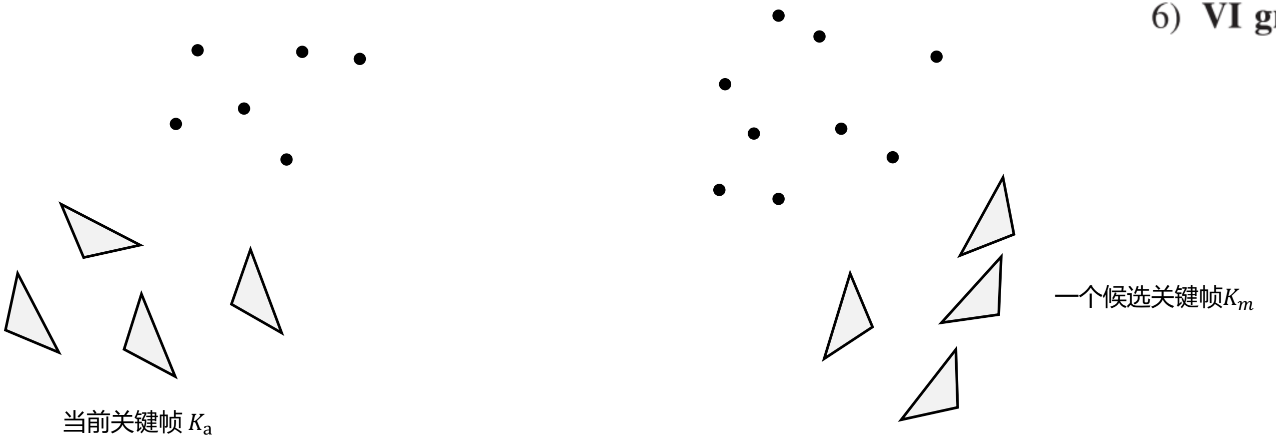
2) Local window.

3) 3-D aligning transformation.

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

5) Verification in three covisible keyframes.

6) VI gravity direction verification.



Map Merging & Loop Closing (VI.A)

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

- 构造Local window:

  K_a& K_a共视关键帧 + K_m& K_m共视关键帧 + 它们看到的地图点

- 分别进行2D-2D、3D-3D的点匹配

1) DBoW2 candidate keyframes.

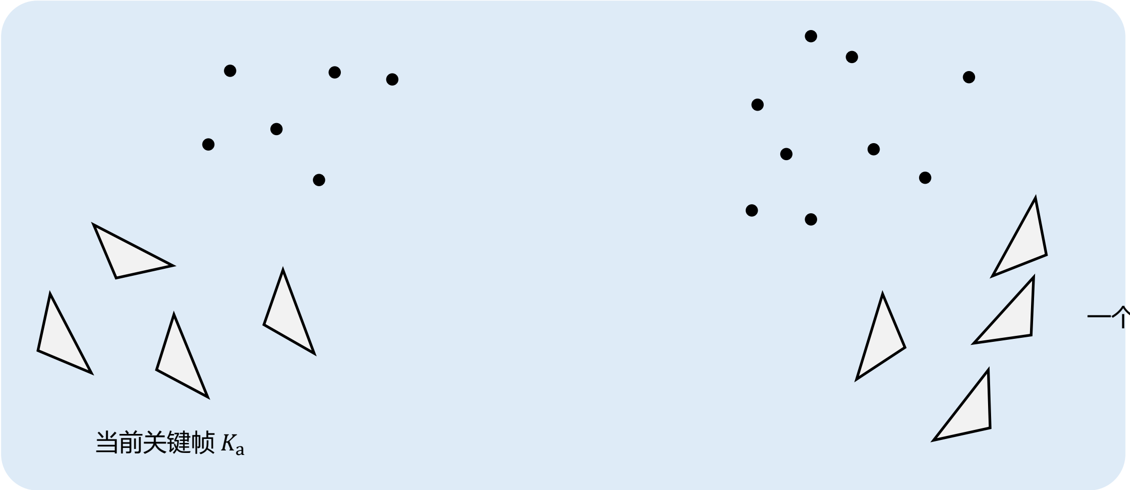
2) Local window.

3) 3-D aligning transformation.

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

5) Verification in three covisible keyframes.

6) VI gravity direction verification.



一个候选关键帧K_m

Map Merging & Loop Closing (VI.A)

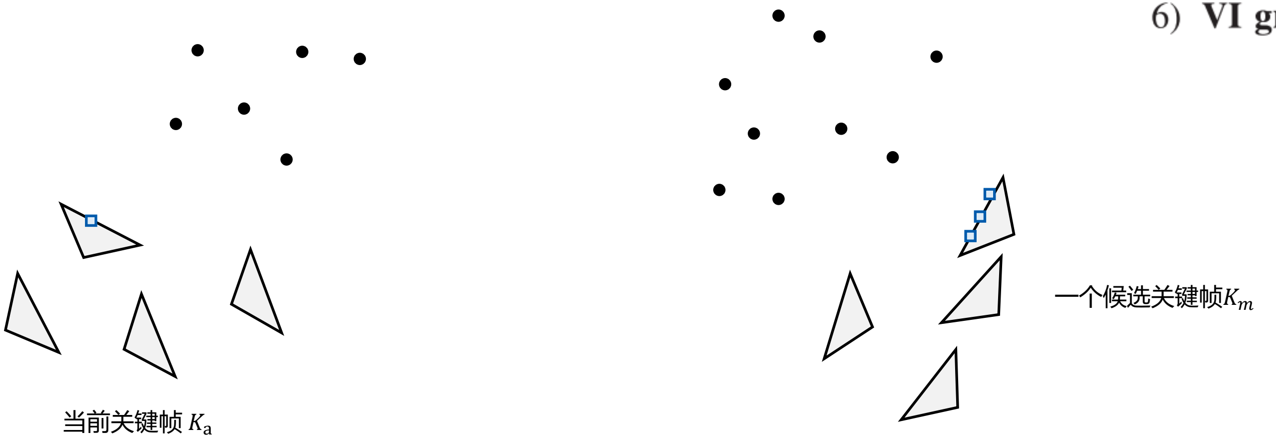
场景识别(2/6)：当前关键帧K_{a}，候选关键帧记为K_m

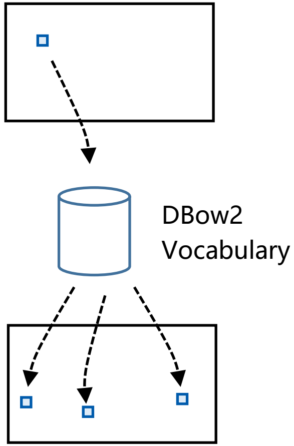
- 构造Local window:

  K_{a}& K_{a}共视关键帧 + K_m& K_m共视关键帧 + 它们看到的地图点

- 分别进行2D-2D、3D-3D的点匹配

1) **DBoW2 candidate keyframes.**
2) **Local window.**
3) **3-D aligning transformation.**
4) **Guided matching refinement.**
5) **Verification in three covisible keyframes.**
6) **VI gravity direction verification.**





Map Merging & Loop Closing (VI.A)

场景识别(2/6)：当前关键帧K_a，候选关键帧记为K_m

- 构造Local window:

  K_a& K_a共视关键帧 + K_m& K_m共视关键帧 + 它们看到的地图点

- 分别进行2D-2D、3D-3D的点匹配

1) **DBoW2 candidate keyframes.**

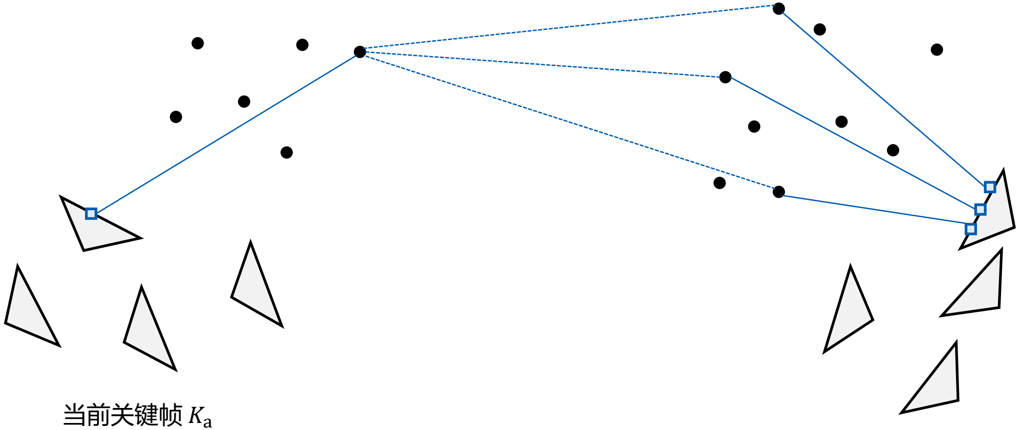
2) **Local window.**

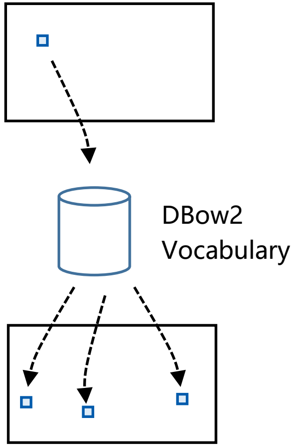
3) **3-D aligning transformation.**

4) **Guided matching refinement.**

分别进行2D-2D、3D-3D的点匹配

6) **VI gravity direction verification.**



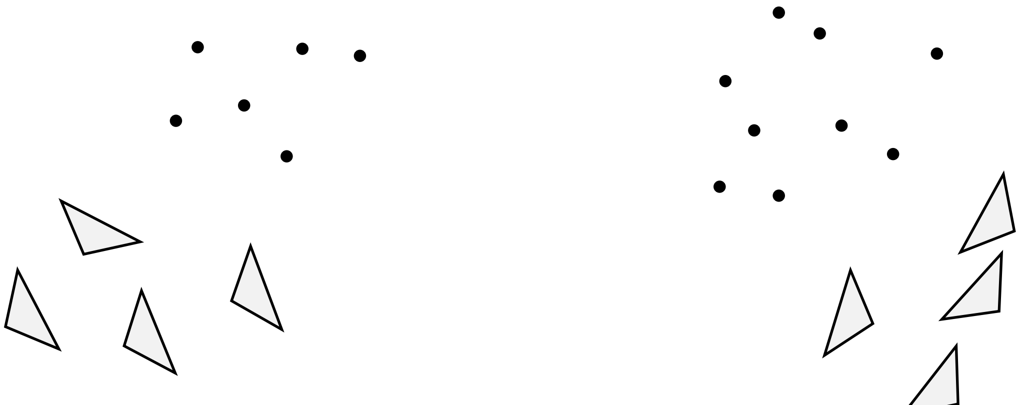


Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_a，候选关键帧记为K_m

- 使用RANSAC得到一系列K_a <=> K_m的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



一个候选关键帧K_m

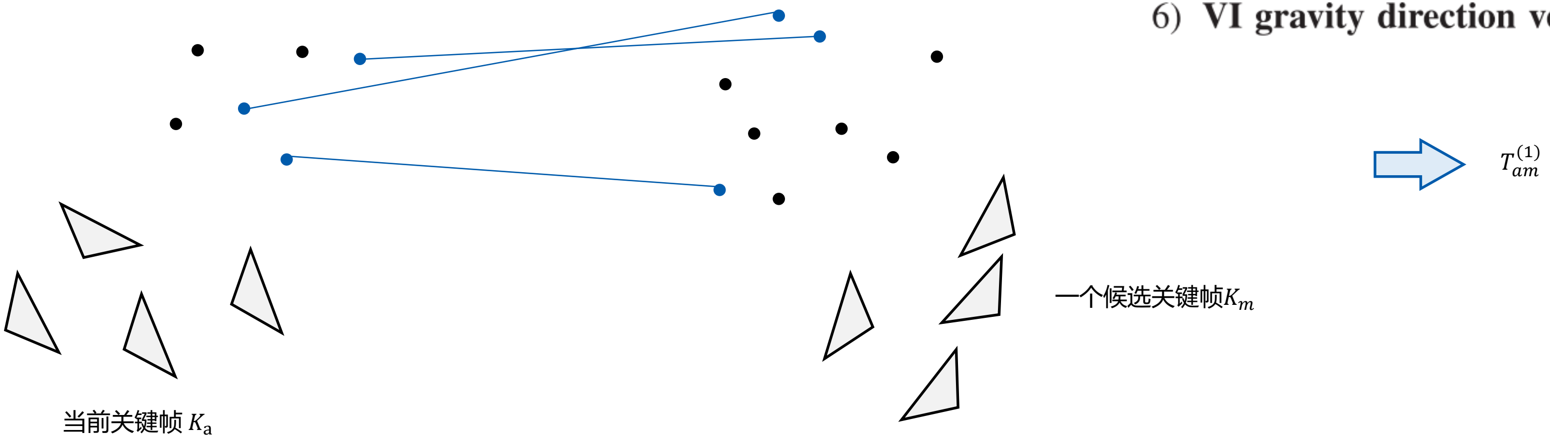
当前关键帧 K_a

Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_a，候选关键帧记为K_m

- 使用RANSAC得到一系列$K_a <=> K_m$的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.

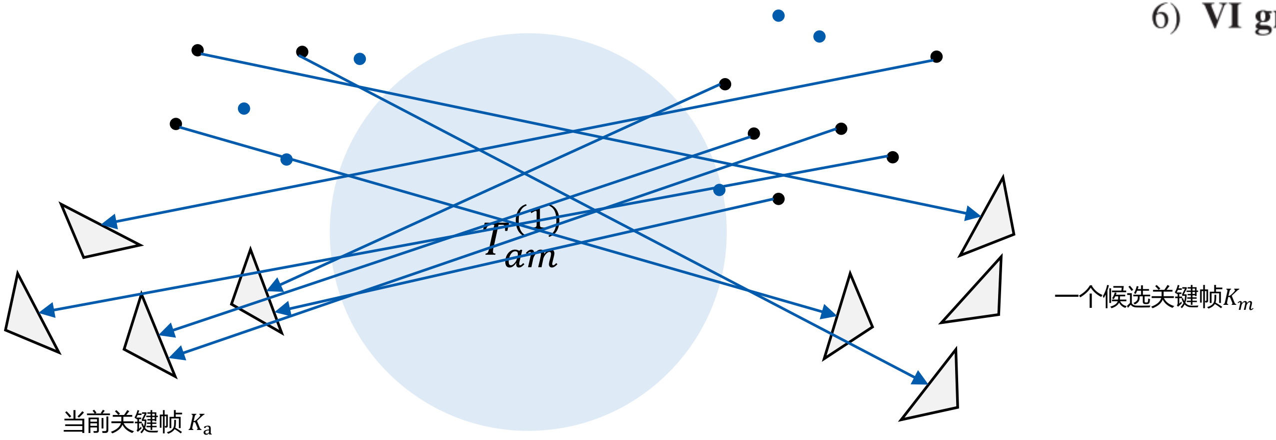


Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_{a}，候选关键帧记为K_m

- 使用RANSAC得到一系列$K_{\mathrm{a}} <=> K_m$的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



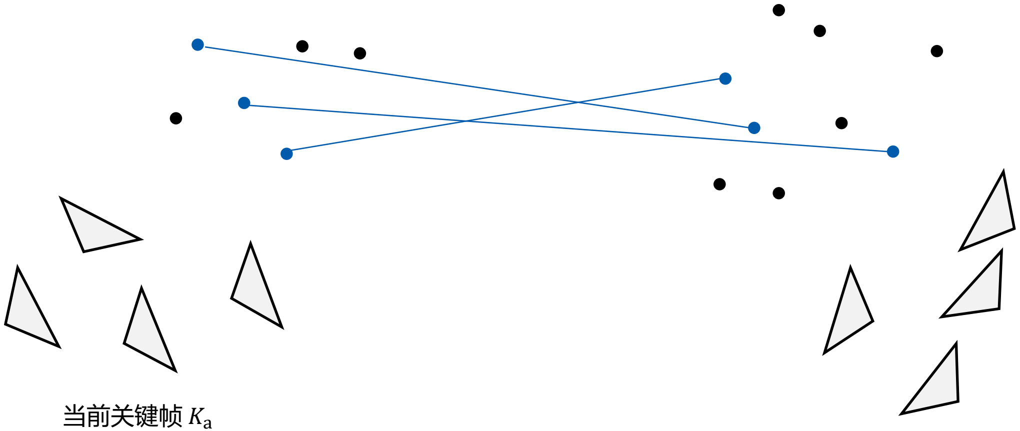
$T_{am}^{(1)}$ Inliers: 30

Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_a，候选关键帧记为K_m

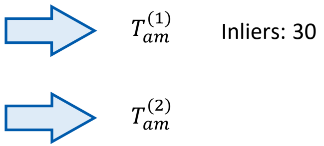
- 使用RANSAC得到一系列$K_a <=> K_m$的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



一个候选关键帧K_m

当前关键帧 K_a

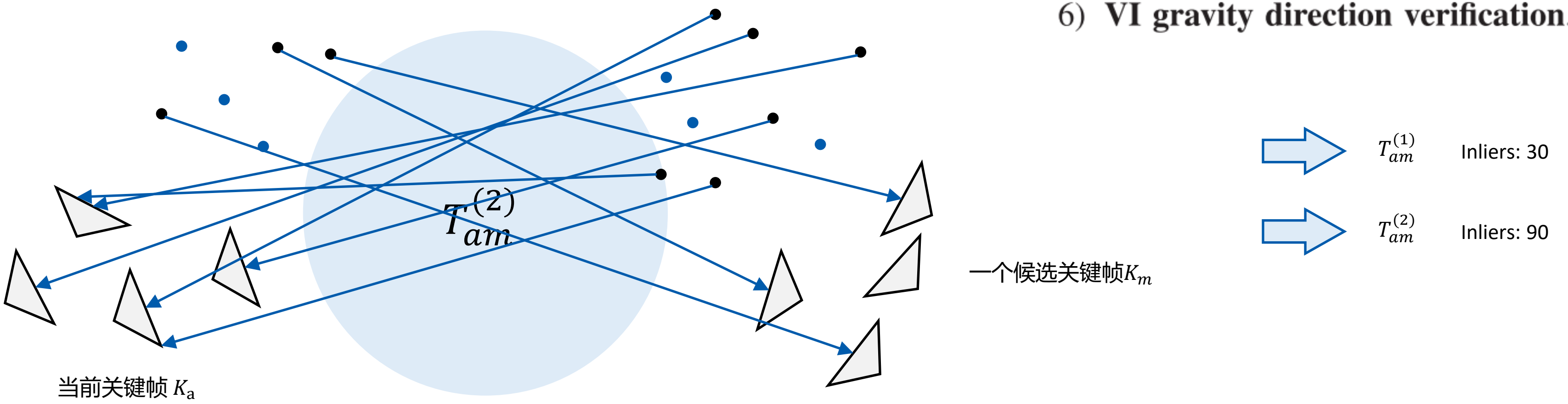


Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_a，候选关键帧记为K_m

- 使用RANSAC得到一系列$K_a <=> K_m$的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.

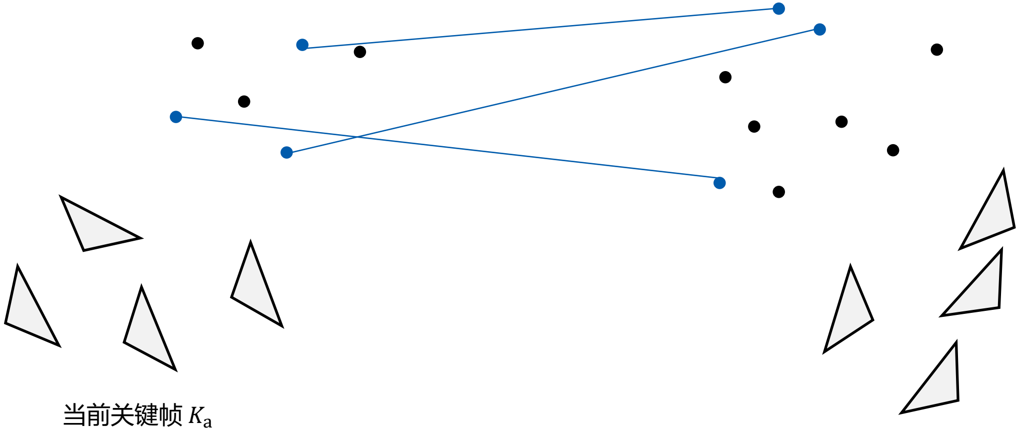


Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_{a}，候选关键帧记为K_m

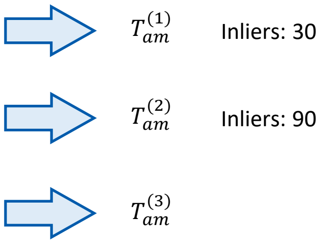
- 使用RANSAC得到一系列$K_{\mathrm{a}} <=> K_m$的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) **DBoW2 candidate keyframes.**
2) **Local window.**
3) **3-D aligning transformation.**
4) **Guided matching refinement.**
5) **Verification in three covisible keyframes.**
6) **VI gravity direction verification.**



当前关键帧 K_{a}

一个候选关键帧K_m

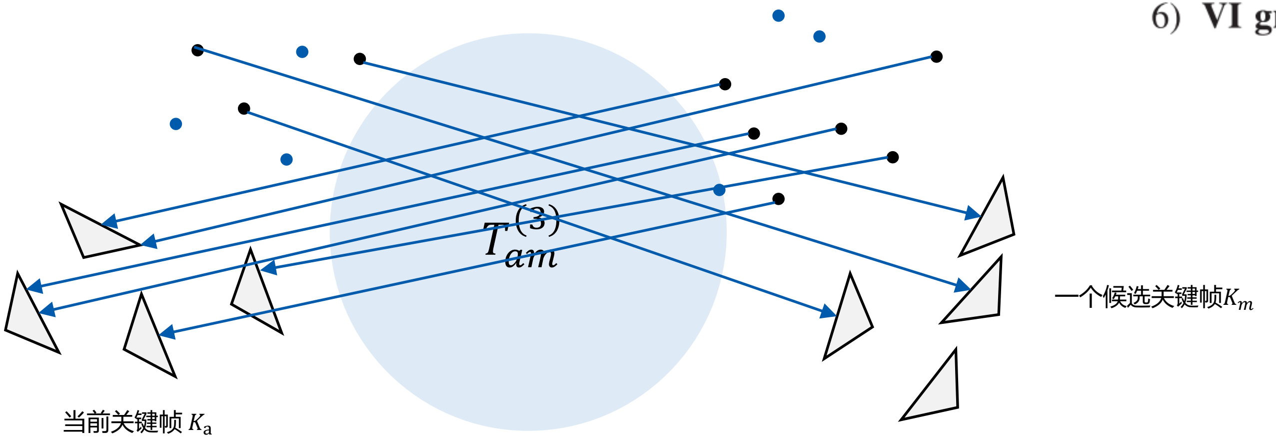


Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_a，候选关键帧记为K_m

- 使用RANSAC得到一系列K_a <=> K_m的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



$T_{am}^{(1)}$ Inliers: 30

$T_{am}^{(2)}$ Inliers: 90

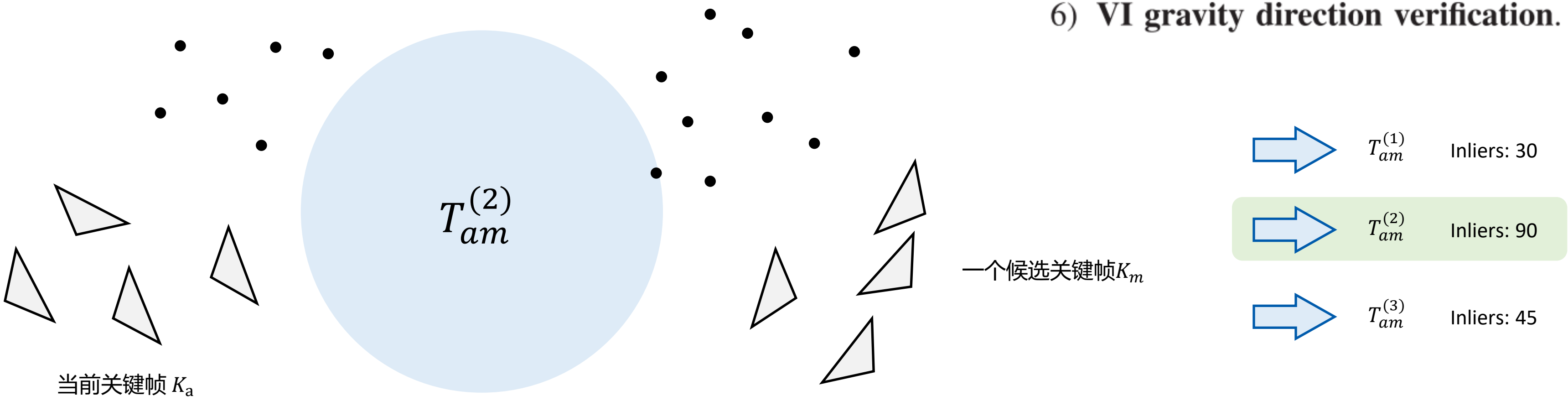
$T_{am}^{(3)}$ Inliers: 45

Map Merging & Loop Closing (VI.A)

场景识别(3/6)：当前关键帧K_a，候选关键帧记为K_m

- 使用RANSAC得到一系列$K_a <=> K_m$的变换T_{am}
- 纯单目/未成熟地图的单目+IMU使用Sim3, 否则SE3
- 重投影误差进行投票得到最佳T_{am}

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.

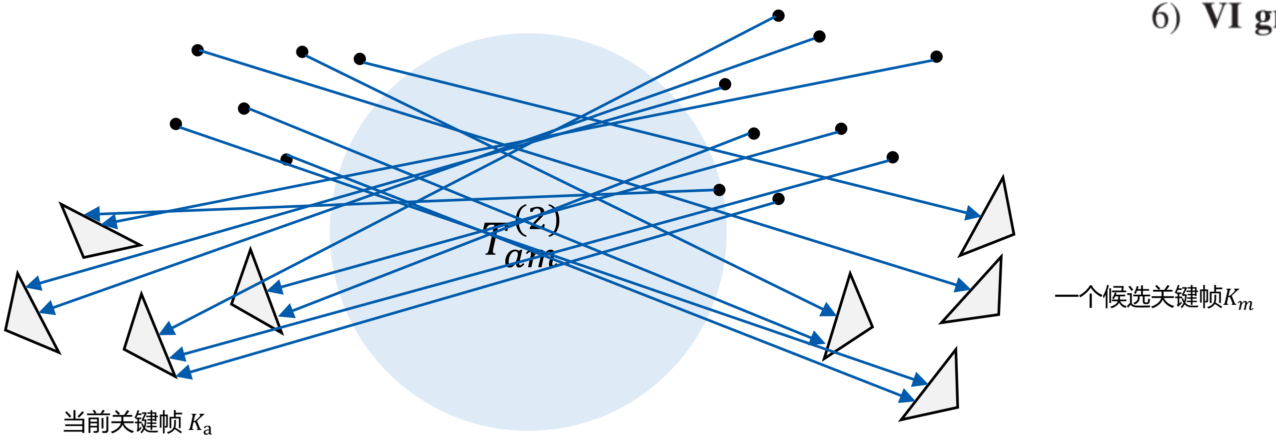


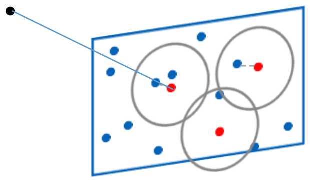
Map Merging & Loop Closing (VI.A)

场景识别(4/6)：当前关键帧K_a，候选关键帧记为K_m

- 基于变换T_{am}的初始估计， K_a、 K_m点云分别投影到对方坐标系下寻找新的匹配
- 优化进一步求精
- 若Inliers超过阈值，则减小投影搜索窗口再优化一次

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



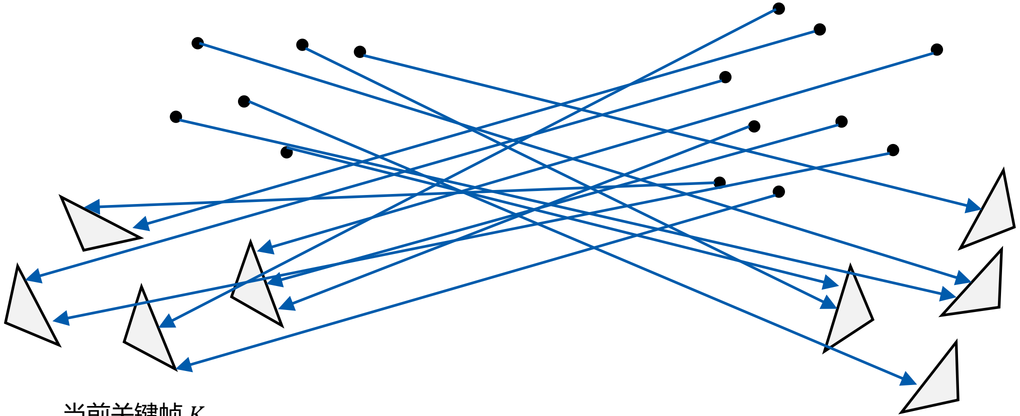


Map Merging & Loop Closing (VI.A)

场景识别(4/6)：当前关键帧K_a，候选关键帧记为K_m

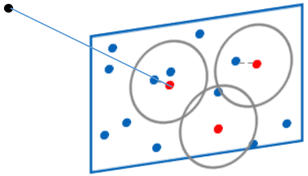
- 基于变换T_{am}的初始估计， K_a、 K_m点云分别投影到对方坐标系下寻找新的匹配
- 优化进一步求精
- 若Inliers超过阈值，则减小投影搜索窗口再优化一次

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



一个候选关键帧K_m

当前关键帧 K_a

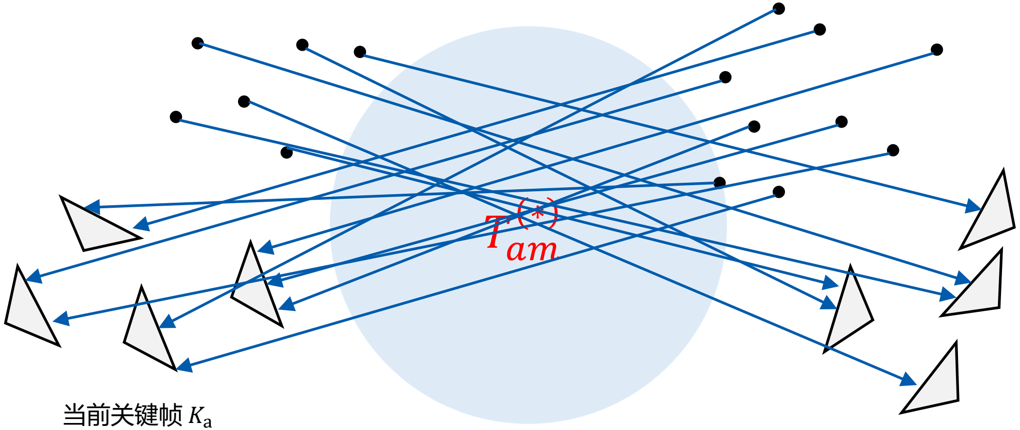


Map Merging & Loop Closing (VI.A)

场景识别(4/6)：当前关键帧K_{a}，候选关键帧记为K_m

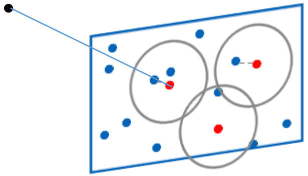
- 基于变换T_{am}的初始估计， K_{a}、 K_m点云分别投影到对方坐标系下寻找新的匹配
- 优化进一步求精
- 若Inliers超过阈值，则减小投影搜索窗口再优化一次

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) **Guided matching refinement.**
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



一个候选关键帧K_m

当前关键帧 K_{a}

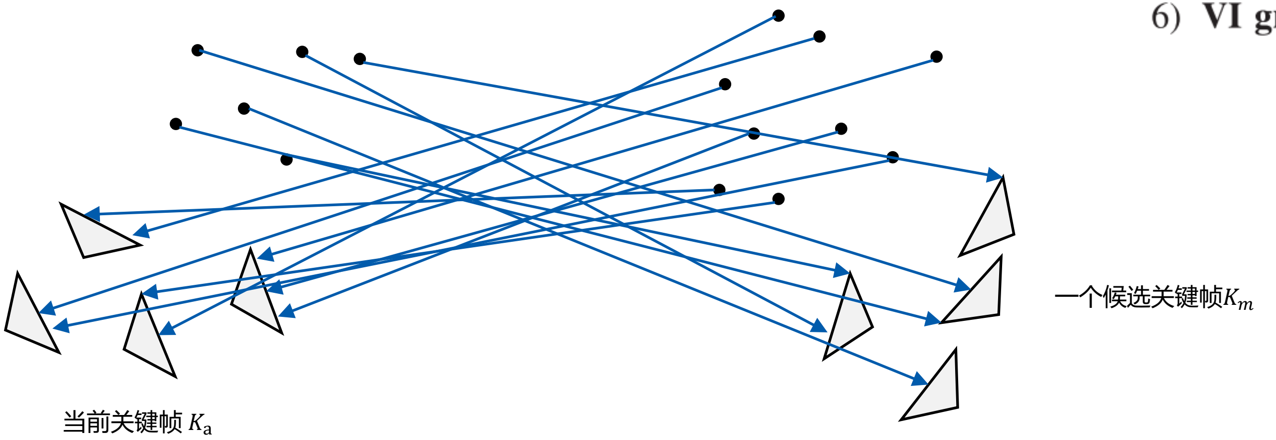


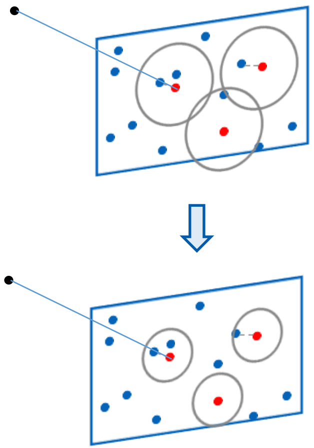
Map Merging & Loop Closing (VI.A)

场景识别(4/6)：当前关键帧K_a，候选关键帧记为K_m

- 基于变换T_{am}的初始估计， K_a、 K_m点云分别投影到对方坐标系下寻找新的匹配
- 优化进一步求精
- 若Inliers超过阈值，则减小投影搜索窗口再优化一次

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.



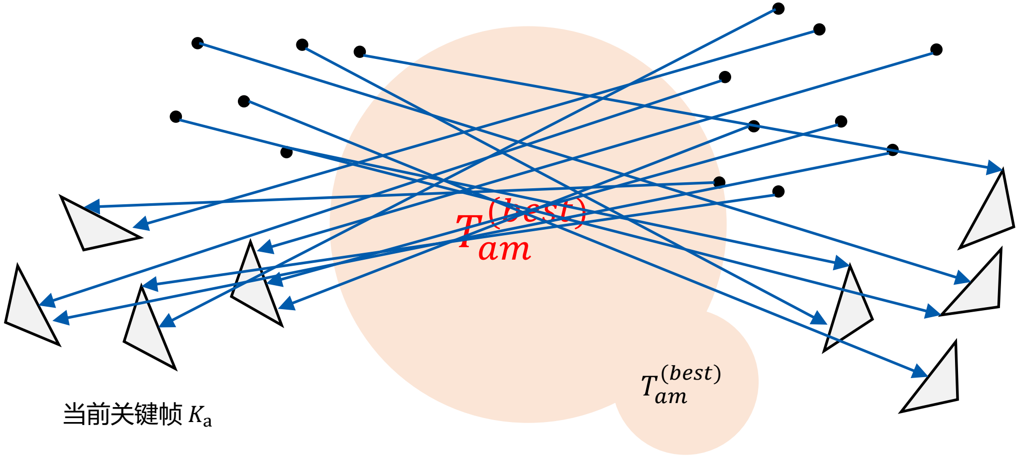


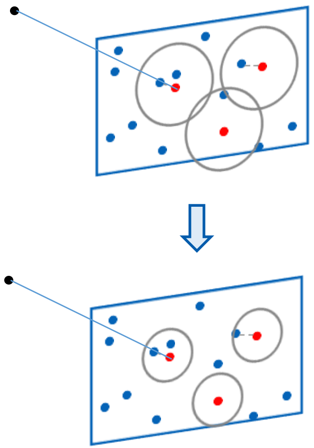
Map Merging & Loop Closing (VI.A)

场景识别(4/6)：当前关键帧K_a，候选关键帧记为K_m

- 基于变换T_{am}的初始估计， K_a、 K_m点云分别投影到对方坐标系下寻找新的匹配
- 优化进一步求精
- 若Inliers超过阈值，则减小投影搜索窗口再优化一次

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) **Guided matching refinement.**
5) Verification in three covisible keyframes.
6) VI gravity direction verification.

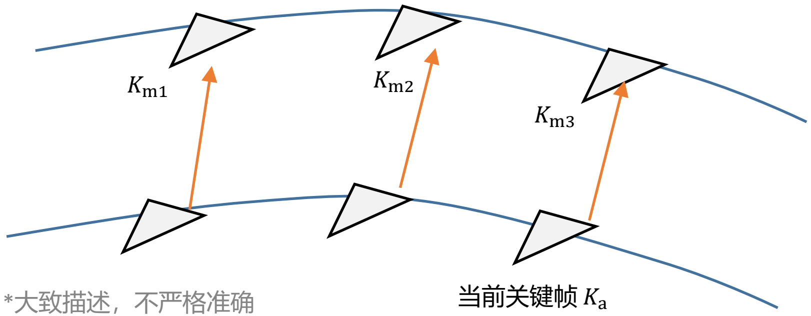




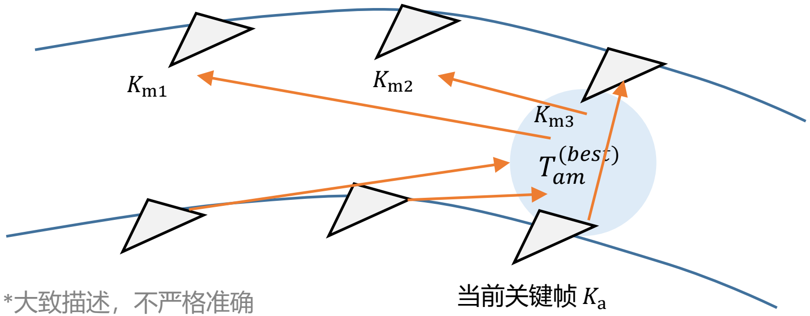
Map Merging & Loop Closing (VI.A)

场景识别(5/6)：当前关键帧K_a，候选关键帧记为K_m

- 改进：在Active Map中找特征匹配数超过阈值的共视关键帧，验证T_{am}



*大致描述，不严格准确



*大致描述，不严格准确

1) DBoW2 candidate keyframes.
2) Local window.
3) 3-D aligning transformation.
4) Guided matching refinement.
5) Verification in three covisible keyframes.
6) VI gravity direction verification.

Map Merging & Loop Closing (VI.A)

场景识别(6/6)：当前关键帧K_a，候选关键帧记为K_m

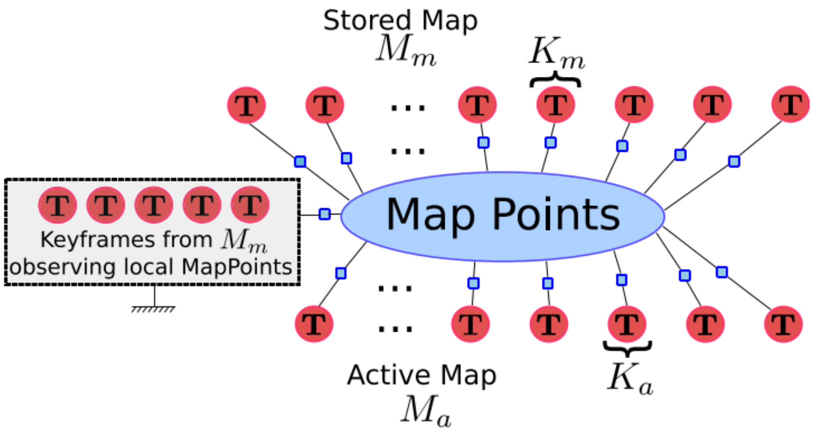
- VI-SLAM模式下将使用重力方向对回环检测结果进行验证，如果 Roll / Pitch 角度差过大，则认为计算结果不合法

1) **DBoW2 candidate keyframes.**
2) **Local window.**
3) **3-D aligning transformation.**
4) **Guided matching refinement.**
5) **Verification in three covisible keyframes.**
6) **VI gravity direction verification.**

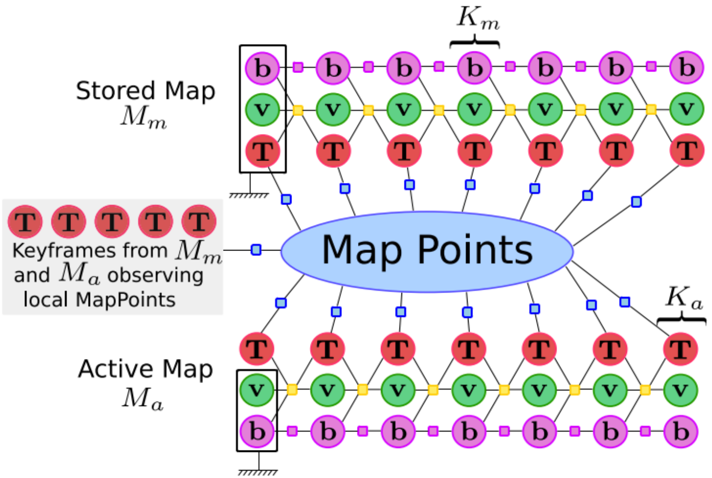
Map Merging & Loop Closing (VI.B & VI.C)

是回环还是地图合并，取决于当前关键帧检测到的回环关键帧是在当前active map还是在其他map

- M_a下的关键帧和地图点将被转换到M_m坐标系下, 组成Welding Windows。纯单目/未成熟单目VIO地图时使用Sim3变换
- 地图融合， M_a中和M_m匹配的地图点，其观测信息将被追加到M_m，共视图和本质图同步更新。最后M_m称为新的Active Map
- Welding windows 中进行 Welding BA



纯视觉



VI-SLAM

- 本质图优化

ORB-SLAM历史与基本概念回顾

抽象相机模型

VISLAM实现与IMU初始化

改进的回环检测与多地图融合

总结

总结

- ORB-SLAM3总体流程和前几代非常相似，对ORB-SLAM系列熟悉的同学能很快上手
- 相机模型的抽象处理，使得SLAM和相机模型解耦，理论上支持绝大多数成像模型的相机
- ORB-SLAM3加入了VI-SLAM的大家庭，也表明多传感器融合的SLAM是目前一大发展趋势
- 多地图的机制有利于在跟丢后保留尽可能多的信息用于后续补救；同时也为实现多机器协同的SLAM、多会话的SLAM的研究提供了基础工作

在线问答

Q&A