视觉SLAM与大尺度增强现实

章国锋

浙江大学CAD&CG国家重点实验室

视觉SLAM面临的关键挑战

挑战一

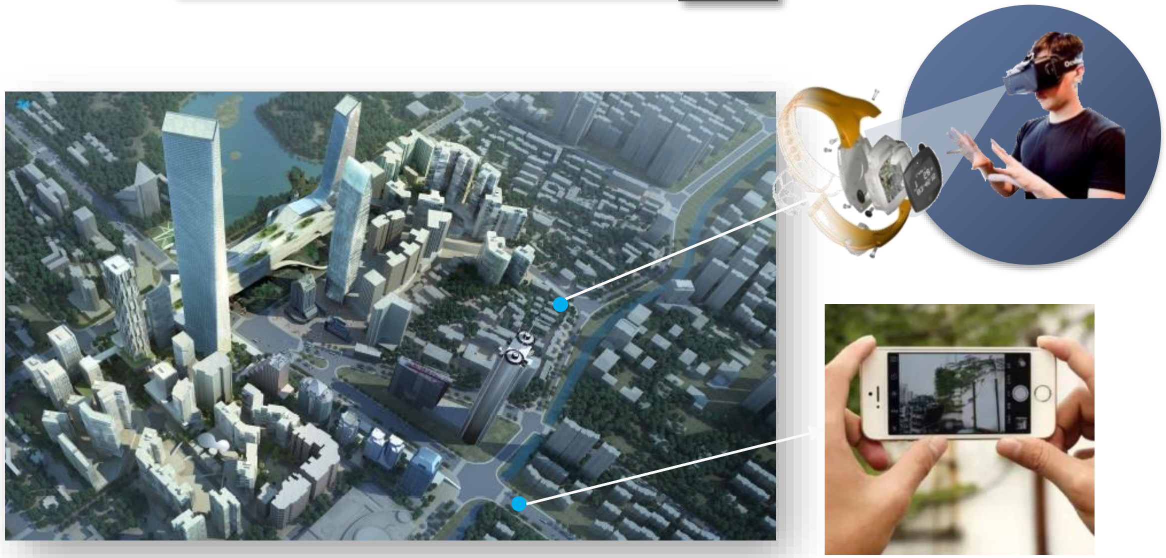
精度和稳定性

- 动态变化、快速运动
- 弱纹理、重复纹理
- 优化计算不稳定

挑战二

实时性

- 场景规模大
- 计算维度高
- 低功耗设备计算能力有限



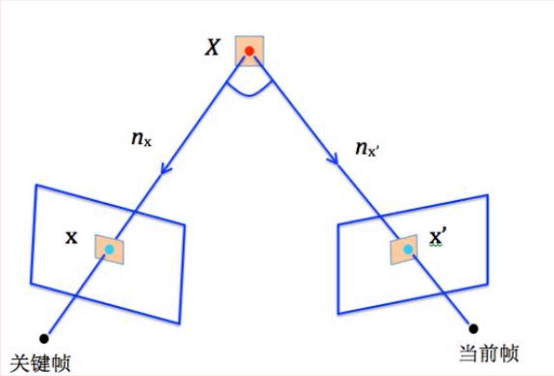
如何提升稳定性?

稳定求解

约束的**正确性**

高效准确的匹配，剔除外点

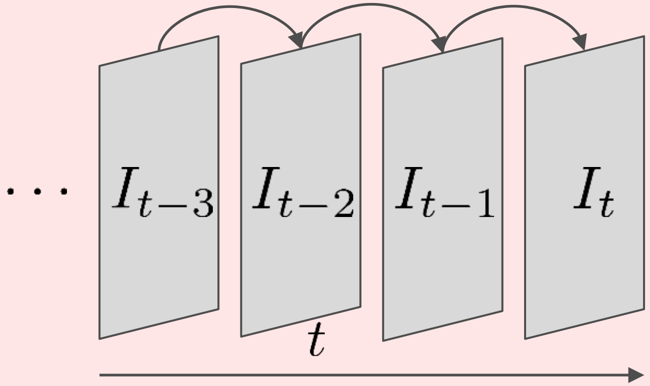
分布先验



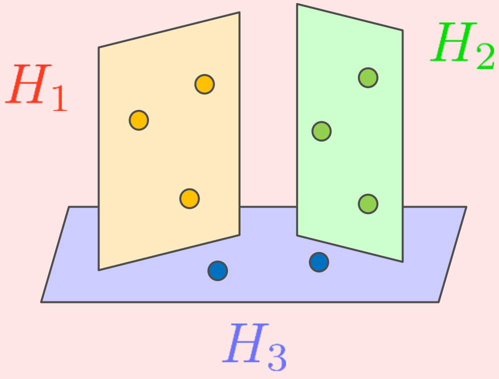
变化检测

约束的**充分性**

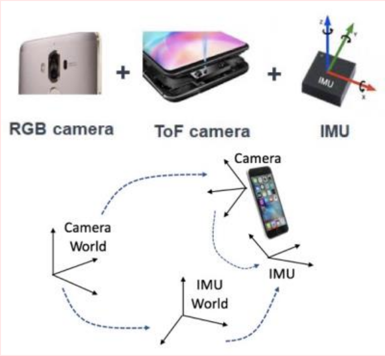
运动先验约束



结构先验约束



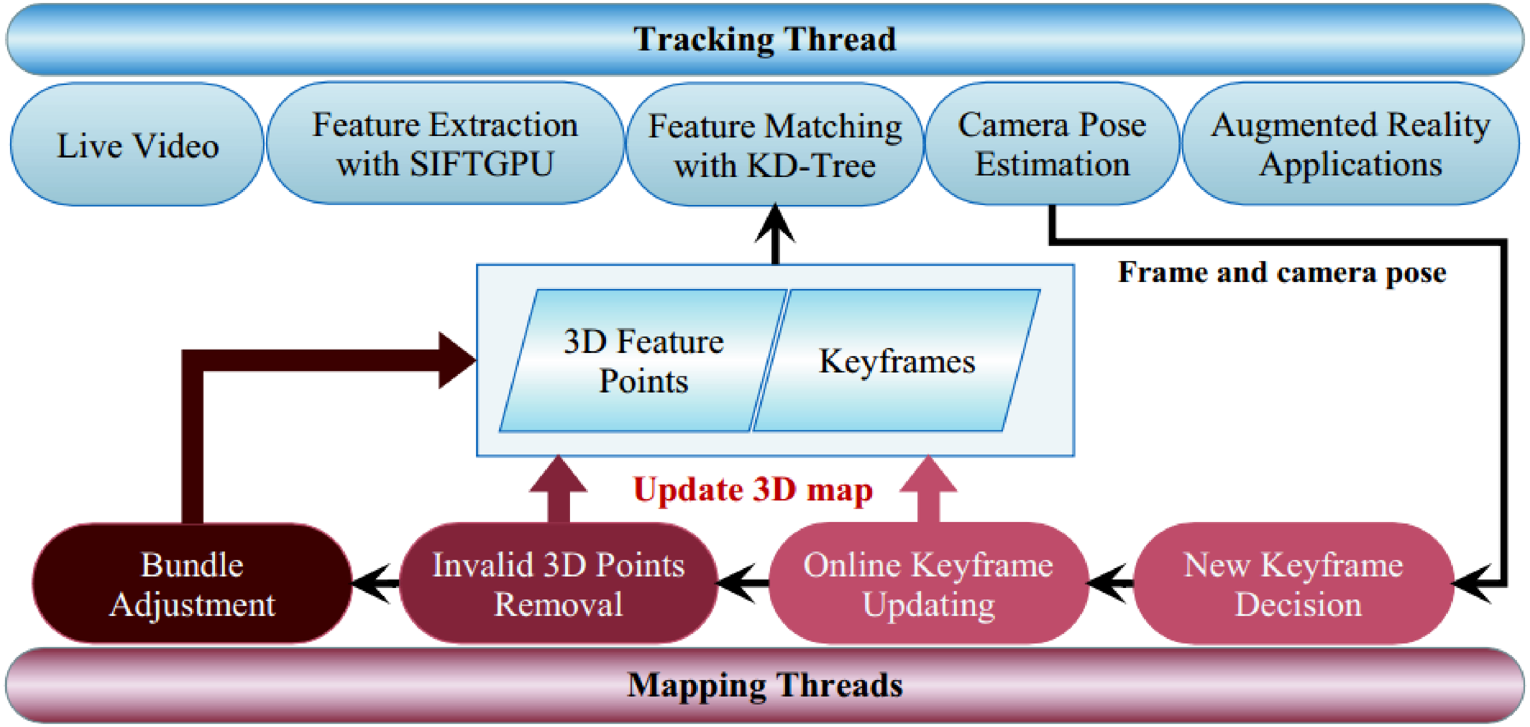
多传感器融合



面向动态场景的视觉SLAM框架

关键思路

- 在线检测环境中变化的特征点，在地图中及时去除失效的三维点并替换相应的关键帧；
- 利用时序连惯性预测内点分布，显著提高内点选中概率。



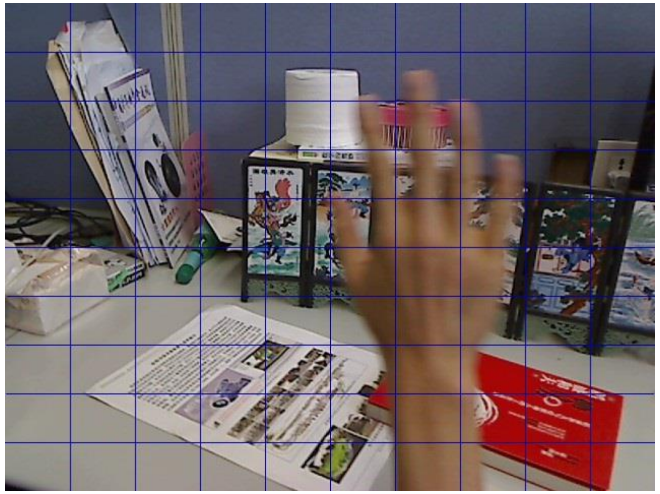
Wei Tan, Haomin Liu, Zilong Dong, Guofeng Zhang and Hujun Bao. Robust Monocular SLAM in Dynamic Environments. International Symposium on Mixed and Augmented Reality (ISMAR), 2013.*

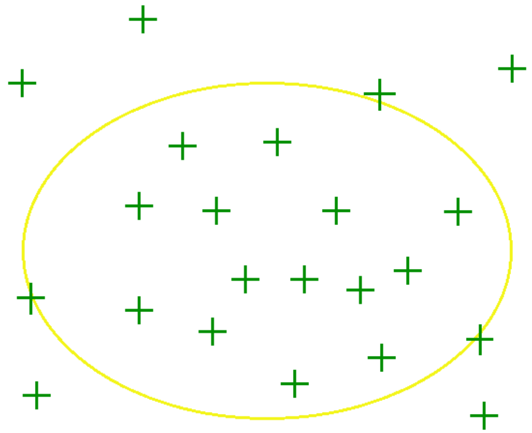
Prior-based Adaptive RANSAC

- Sample generation
  - 10x10 bins
  - Prior probability $p_i = \varepsilon_i^* / \sum_j \varepsilon_j^*$
- Hypothesis evaluation

$$s = \left(\sum_i \varepsilon_i\right) \frac{\pi \sqrt{\det(C)}}{A}$$

  - Inliers number $N \approx \sum_i \varepsilon_i$
  - Inliers distribution, i.e., distribution ellipse C

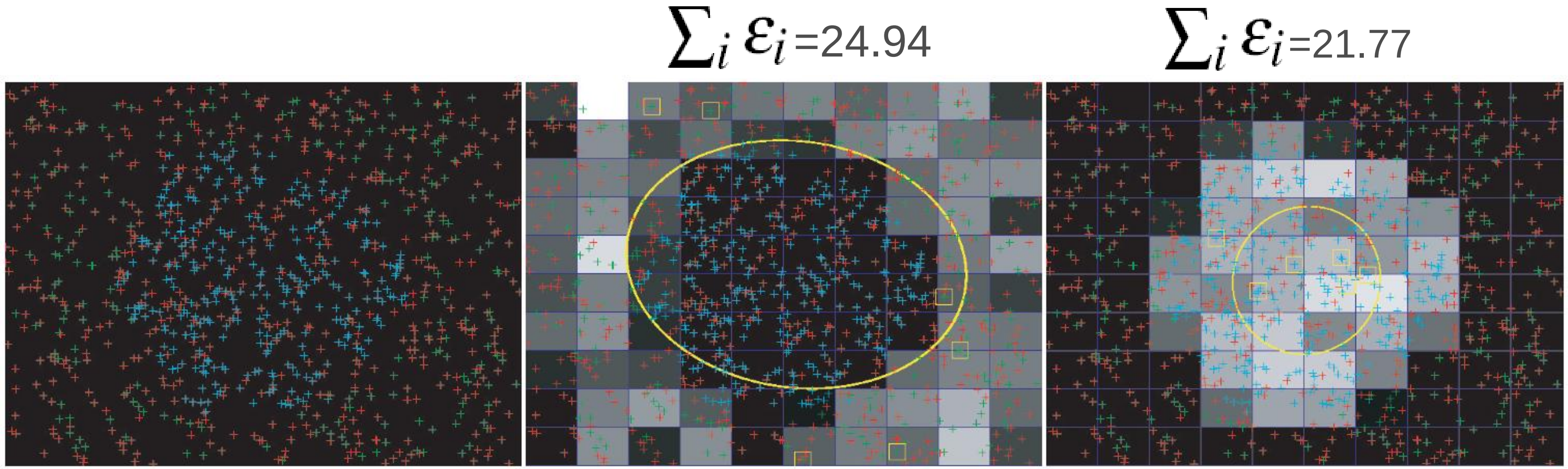




Prior-based Adaptive RANSAC

- Hypothesis evaluation

$$s = \left(\sum_i \varepsilon_i\right)\frac{\pi\sqrt{\det(\Sigma)}}{A}$$



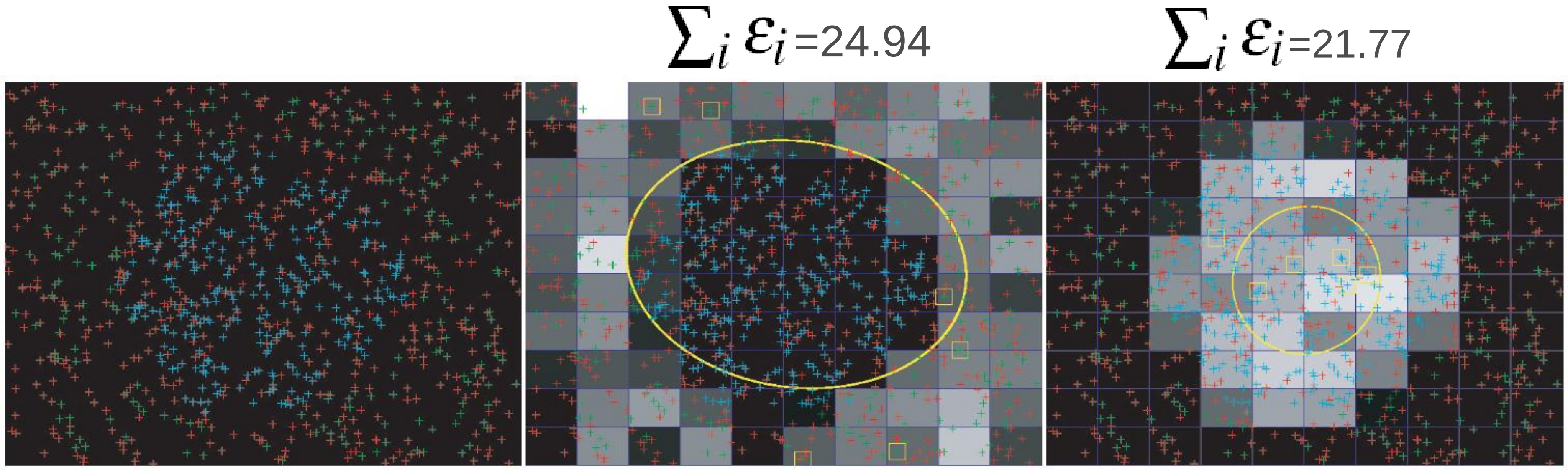
200 green points on the static background, 300 cyan points on the rigidly moving object, 500 red points are randomly moving.

Prior-based Adaptive RANSAC

- Hypothesis evaluation

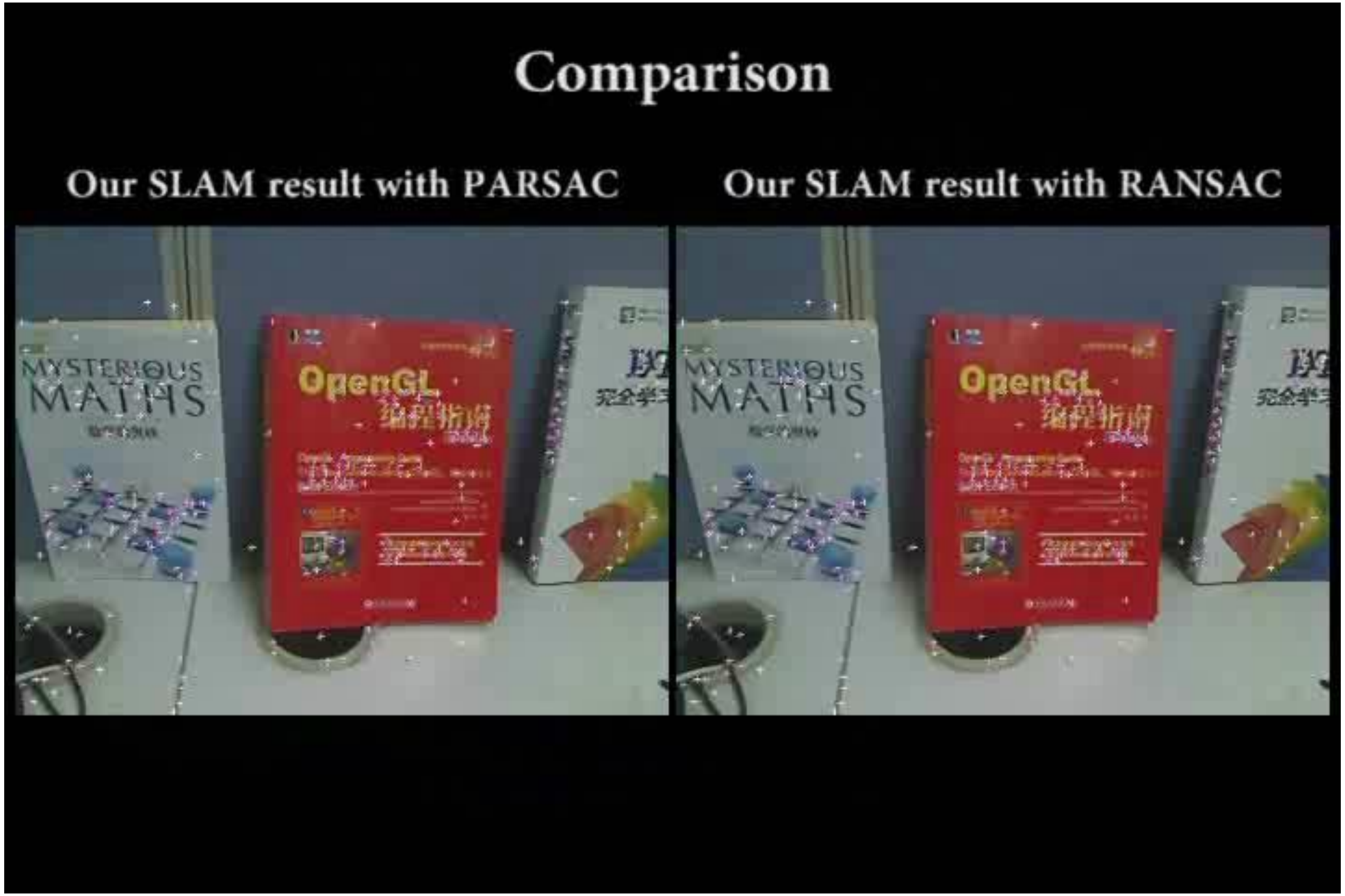
$$S = \left(\sum_i \varepsilon_i\right) \frac{\pi\sqrt{\det(\Sigma)}}{A}$$

S1 = 8.31 > S2 = 1.98



200 green points on the static background, 300 cyan points on the rigidly moving object, 500 red points are randomly moving.

Results Comparison

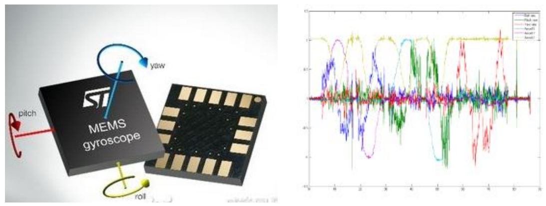


Optimization with Motion Priors

- **Sliding-window based pose optimization**
  - Motion constraints with IMU measurements

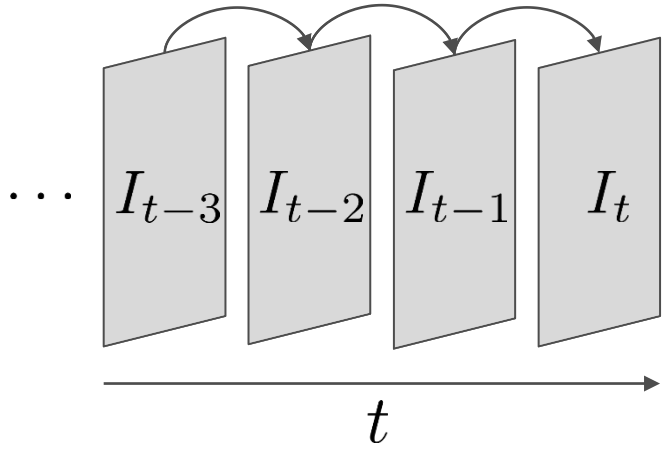
$$\mathcal{X}^{\star} = \arg\min_{\mathcal{X}} \sum_{i,j} \|\mathbf{r}_{\mathcal{I}_{ij}}\|^2_{\Sigma_{ij}} + \sum_{i} \sum_{l} \|\mathbf{r}_{\mathcal{C}_{il}}\|^2_{\Sigma_C}$$

IMU error　　reprojection error

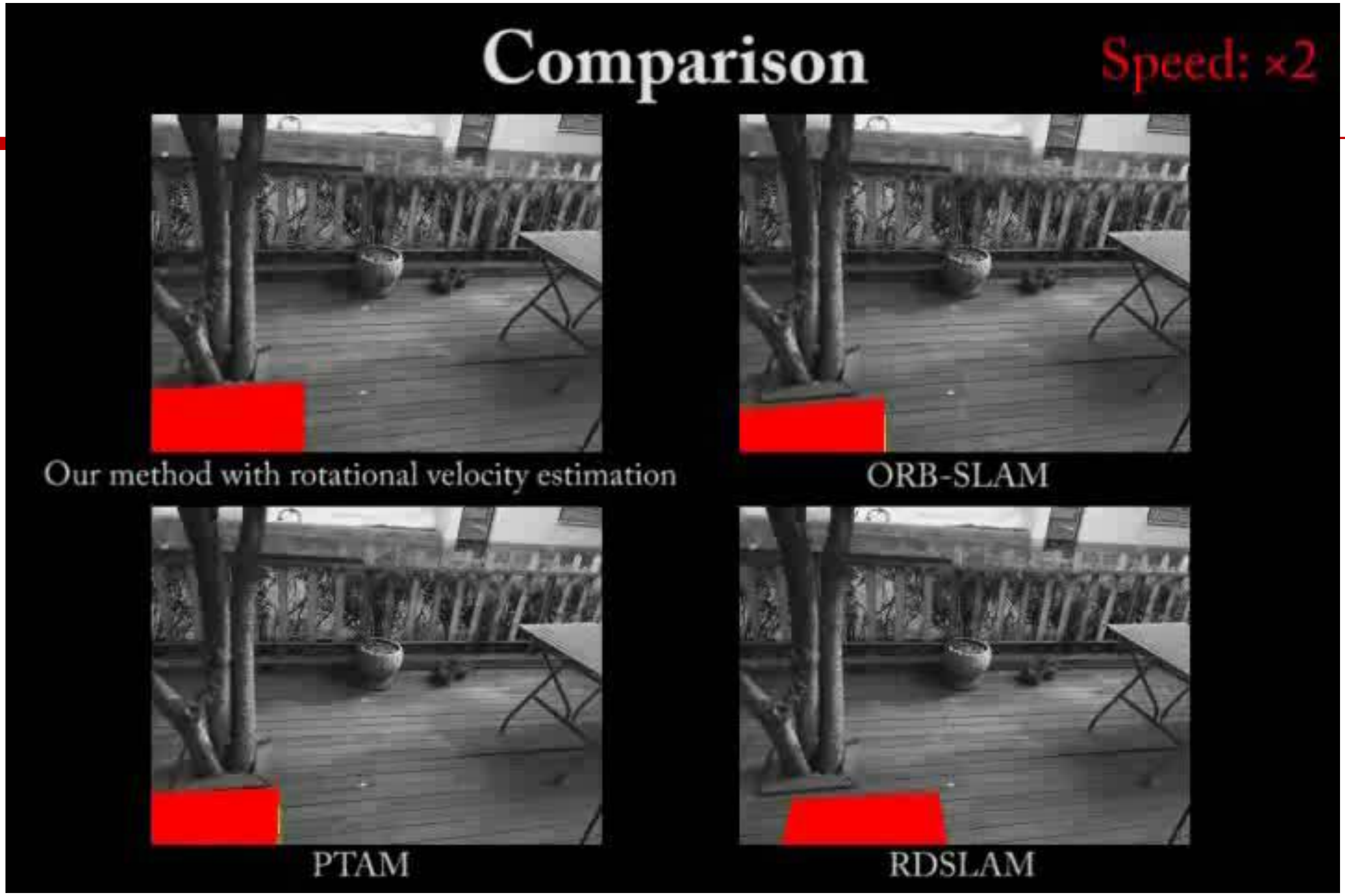


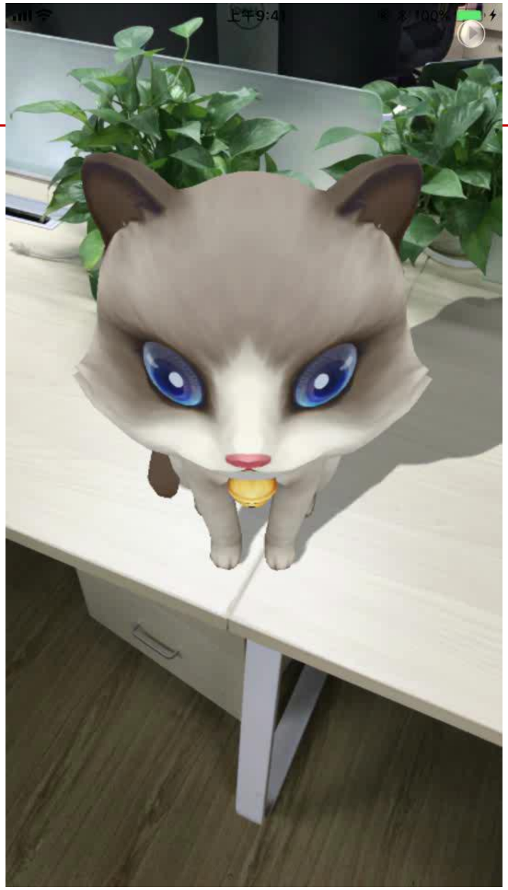
IMU measurements

- **Motion constraints without IMU measurements**
  - Acceleration: is generally small and assumed to be zero
  - Rotational velocity: combine feature matching and global image alignment to estimate it



$$\hat{\omega}_i = \arg\min_{\omega} \Big(\sum_{x \in \Omega} ||\tilde{I}_i(\mathbf{x}) - \tilde{I}_{i+1}(\pi(\mathbf{K}\mathbf{R}_{\Delta}(\boldsymbol{\omega}, t_{\Delta_i})\mathbf{K}^{-1}\mathbf{x}^h))||_{\delta_{\mathbf{I}}} + \sum_{(\mathbf{x}_i, \mathbf{x}_{i+1}) \in M_{i,i+1}} \frac{1}{\delta_{\mathbf{X}}} ||\pi(\mathbf{K}\mathbf{R}_{\Delta}(\boldsymbol{\omega}, t_{\Delta_i})\mathbf{K}^{-1}\mathbf{x}_i^h) - \mathbf{x}_{i+1}||_2^2 \Big)$$

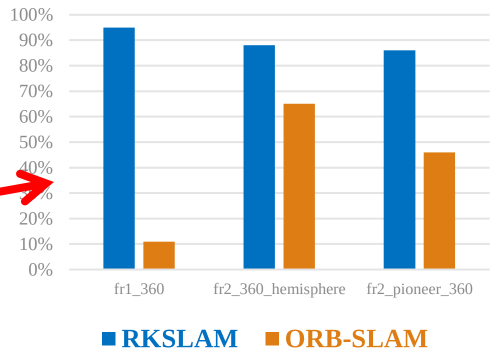




TUM Benchmark

Group	Sequence	RKSLAM	ORB-SLAM	PTAM	LSD-SLAM
A	fr1_xyz	0.61/0%/100%	1.05/0%/100%	1.29/0%/100%	7.64/0%/100%
A	fr2_xyz	0.43/0%/100%	0.23/0%/100%	0.29/0%/100%	6.32/0%/100%
A	fr3_sitting_xyz	1.98/0%/92%	1.31/5%/100%	X	9.12/0%/100%
B	fr1_desk	1.69/0%/100%	1.40/12%/100%	2.71/0%/44%	3.86/27%/100%
B	fr2_desk	10.10/0%/97%	0.78/6%/100%	0.55/0%/20%	17.41/0%/100%
B	fr3_long_office	2.48/0%/100%	2.17/0%/100%	0.82/0%/31%	36.04/30%/100%
C	fr1_rpy	1.26/0%/100%	5.53/4%/84%	X	3.26/0%/11%
C	fr2_rpy	0.41/0%/100%	0.23/32%/100%	0.56/0%/100%	3.71/0%/25%
C	fr3_sitting_rpy	1.44/0%/100%	0.19/93%/100%	2.44/0%/93%	3.36/0%/89%
D	fr1_360	11.81/0%/95%	8.16/5%/11%	X	8.25/0%/5%
D	fr2_360_hemisphere	17.48/0%/88%	12.27/1%/65%	76.50/0%/33%	25.64/0%/19%
D	fr2_pioneer_360	20.24/0%/86%	1.40/69%/46%	59.09/0%/98%	30.62/0%/41%

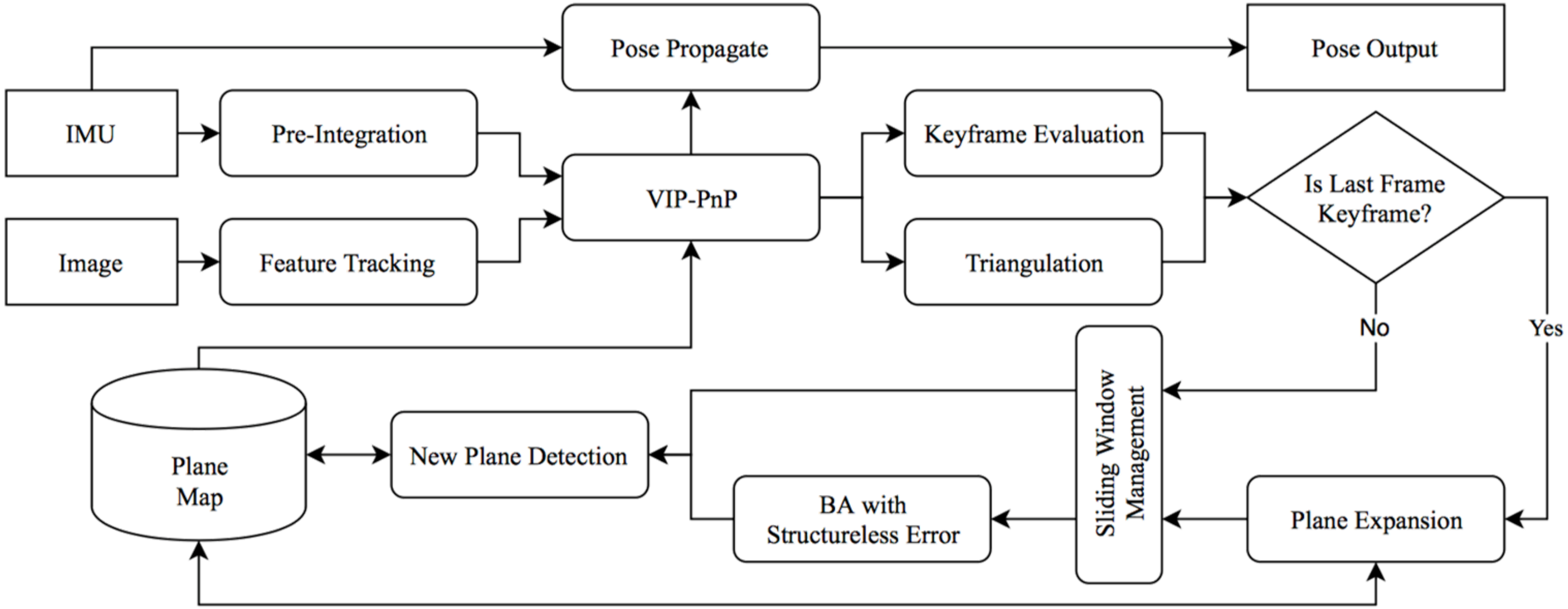
Tracking Success Ratio



Speed	PC
RKSLAM	~150fps
ORB-SLAM	~30fps

Robust and Efficient Visual-Inertial Odometry with Multi-plane Priors

- Motivation
  - Planes commonly exist in human-made scene
  - Planes should be useful for robust localization



System Overview

Robust and Efficient Visual-Inertial Odometry with Multi-Plane Priors. Jinyu Li, Bangbang Yang, Kai Huang, Guofeng Zhang, Hujun Bao*. Chinese Conference on Pattern Recognition and Computer Vision 2019*

Plane Detection and Expansion

- Plane Detection
  - 3-point RANSAC
  - Add planes with inlier points larger than certain threshold

- Plane Expansion
  - Reprojection Consensus

$$\epsilon_k = \sum_i \| u_{ik}(\lambda_k) - \tilde{u}_{ik} \|^2, \quad \epsilon_k^{\perp} = \sum_i \| u_{ik}(\lambda_k^{\perp}) - \tilde{u}_{ik} \|^2.$$

Original RPE   Point On Plane RPE

Add point to the plane when

- On plane RPE not greater than 0.5 pixel
- Point consistent with plane via RPE consensus (RPE not increase more than 1.2 times)

Remove point from the plane when $\epsilon_k^{\perp} \leq \max\{\alpha \epsilon_k, \gamma\}$

- Sufficient observation and triangulated position far from the plane

Plane Constraints

- VIP-PnP
  - Solving the BA as if some points lying on planes

$$\underset{{}^w_b p_i, {}^w_b q_i}{\arg\min} \sum_{k=1}^{M} \|u_{ik} - \tilde{u}_{ik}\|_{\Psi}^2 + \|r_{\mathrm{IMU}}({}^w_b p_i, {}^w_b q_i)\|_{\Phi}^2 + \sum_{s=1}^{P}\sum_{k=1}^{M_s} \|u_{i,sk}^{\perp} - \tilde{u}_{i,sk}\|_{\Psi}^2$$

Ordinary RPE   IMU Constraints   Point-on-Plane RPE

- Structureless Plane-Distance Error

$$A_k = \begin{pmatrix} \vdots \\ \tilde{u}_{ikx} r_{i3} - r_{i1} \\ \tilde{u}_{iky} r_{i3} - r_{i2} \\ \vdots \end{pmatrix}, \quad b_k = \begin{pmatrix} \vdots \\ \tilde{u}_{ikx} p_{i3} - p_{i1} \\ \tilde{u}_{iky} p_{i3} - p_{i2} \\ \vdots \end{pmatrix} \Rightarrow A_{sk} = \begin{pmatrix} A_k \\ w_k n_s^{\top} \end{pmatrix}, \quad b_{sk} = \begin{pmatrix} b_k \\ w_k d_s \end{pmatrix}$$

Original Triangulation   Augmented with Plane Constraint

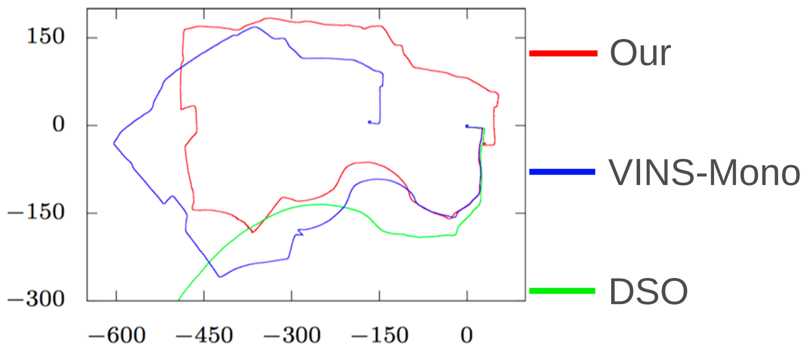
$$x_{sk} = (A_{sk}^{\top} A_{sk})^{-1} A_{sk}^{\top} b_{sk} \qquad r_P(\{{}^w_b p_i, {}^w_b q_i\}, n_s, d_s) = |n_s^{\top} x_{sk} - d_s|$$

Minimize Point to Plane Distance Error

- Augment for degenerated constraints (insufficient parallax/observations)
- For 1 landmark, *m* reprojection error → 1 structureless error

Results

Trajectories on TUM-VI Outdoors1

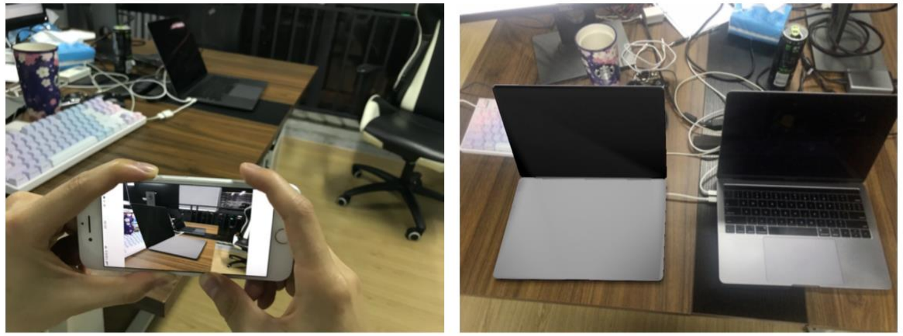


The RMSE(m) of localization

	Dataset	ORB-SLAM2 −Loop	ORB-SLAM2 +Loop	SVO2 E+P	SVO2 BA	DSO	VINS-Mono −Loop	VINS-Mono +Loop	PVIO −Plane	PVIO +Plane
EuRoC [2]	MH_01	0.02	0.03	0.10	0.06	0.05	0.16	0.15	0.19	**0.13**
	MH_02	0.03	0.03	0.12	0.07	0.05	0.18	0.26	**0.16**	0.21
	MH_03	0.17	0.05	0.41	×	0.18	0.20	**0.11**	0.31	0.16
	MH_04	0.15	0.37	0.43	0.40	2.50	0.35	0.37	0.29	**0.29**
	MH_05	0.06	0.04	0.30	×	0.11	0.30	**0.28**	0.79	0.34
	V1_01	0.03	0.03	0.07	0.05	0.12	0.09	0.10	0.10	**0.08**
	V1_02	0.15	0.03	0.21	×	0.11	0.11	0.09	×	**0.09**
	V1_03	(0.49)	0.10	×	×	0.93	0.19	0.18	×	**0.16**
	V2_01	0.03	0.03	0.11	×	0.04	0.09	0.08	0.11	**0.05**
	V2_02	0.15	0.03	0.11	×	0.13	**0.16**	0.17	×	0.20
	V2_03	(0.73)	(0.40)	1.08	×	1.16	0.29	0.37	×	**0.29**
TUM-VI [21]	Room1	×	0.10	×	×	0.06	0.07	**0.07**	1.65	0.26
	Room2	×	0.12	×	×	0.11	0.07	**0.07**	0.12	0.15
	Room3	×	(0.04)	×	×	0.12	0.12	**0.12**	0.18	0.18
	Corridor1	×	×	×	×	5.43	0.59	0.59	×	**0.23**
	Outdoors1	×	×	×	×	×	74.55	81.57	×	**22.26**

System Output Trajectory Preview

Real-time AR Effect on iPhone 7



The source code will be released soon at https://www.github.com/zju3dv/

如何提高全局优化的效率?

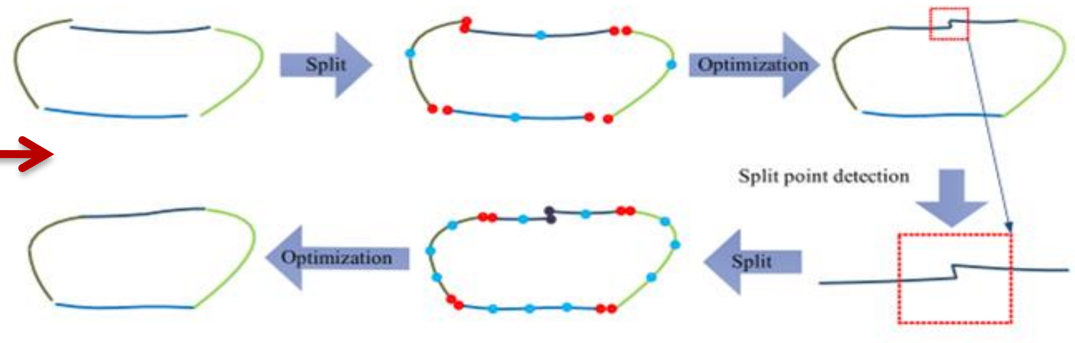
集束调整（Bundle Adjustment）

- 联合优化相机参数和三维点

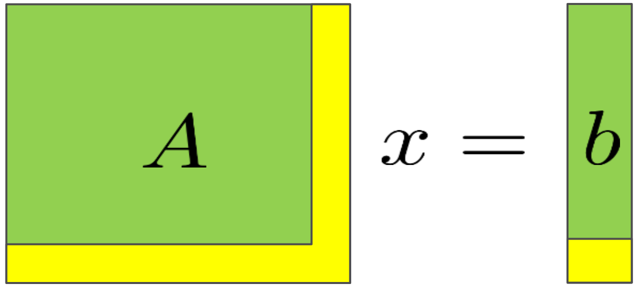
$$\underset{C_1,\dots,C_{N_c},X_1,\dots,X_{N_p}}{\operatorname{argmin}} \sum \left\| \pi(C_i, X_j) - x_{ij} \right\|^2$$

高效求解

- Coarse-to-Fine 求解策略



- 增量式计算



如何实现大尺度场景的精准跟踪定位与增强现实?

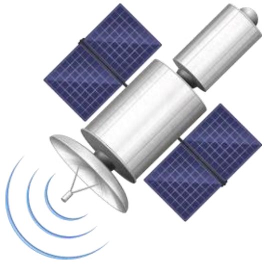
SenseMARS: 端-云协同的混合现实平台

定位与AR导航

进入
距离终点：
北四环西路
入口
出口
路况
渤海银行
善缘街
海淀西街
理想国际大厦
辉煌时代
方正国际大厦
中国银行
海淀中街
蒿房胡同
海淀北一街
西
东
南
中信银行
招商银行
天天易家
便利
爱奇艺创新大厦
高德地图
中关村SOHO
继续导航
开始导航
全览

定位与导航

常用的解决方案



GPS

- 10米误差
- 室内不能用

WIFI, Blue Tooth

- 需要布置设备
- 工程量大
- 费用较高

基于视觉的方案

优势

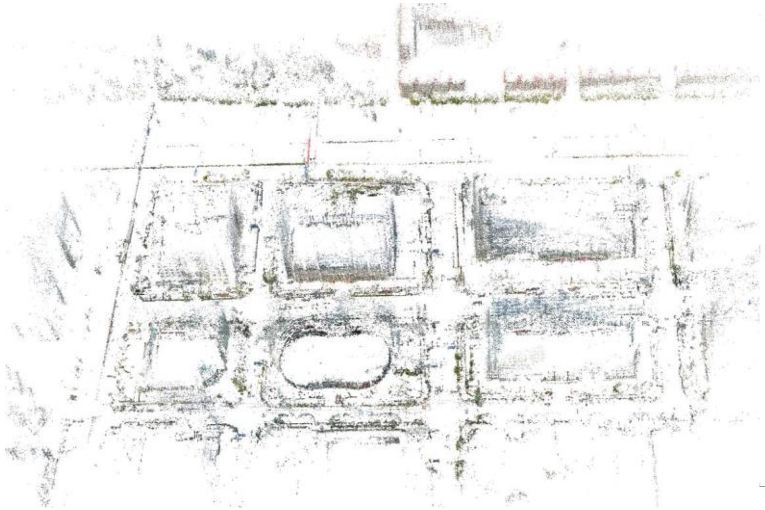
- 成本低
- 不需要额外布置传感器
- 分米/厘米级定位精度

挑战性

- 缺乏视觉特征
- 环境改变
- 计算量大

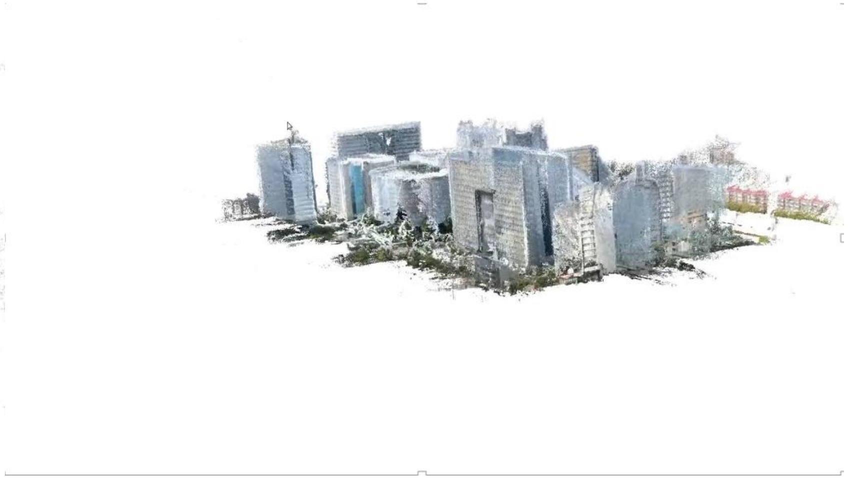
视觉定位与AR导航

稀疏地图重建



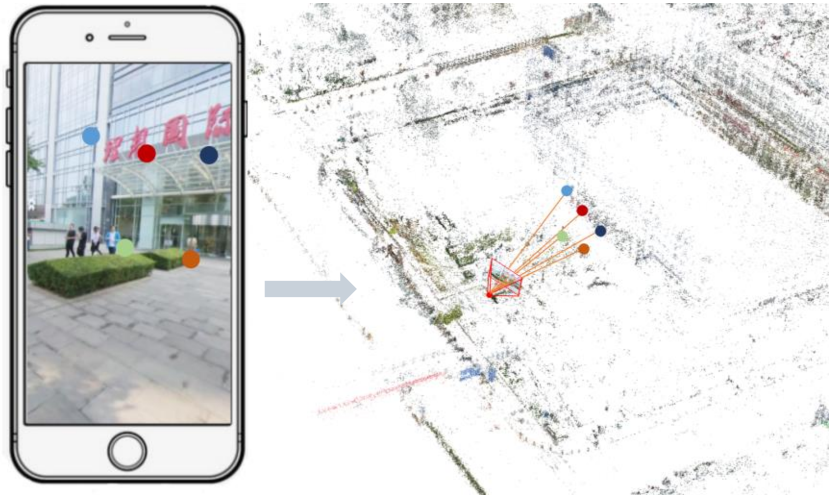
- 抽取视觉特征
- 恢复三维结构

稠密地图重建



- 处理遮挡和碰撞
- 自由视角浏览

视觉定位与跟踪



- 全局重定位与实时6DoF相机位姿恢复

地图重建

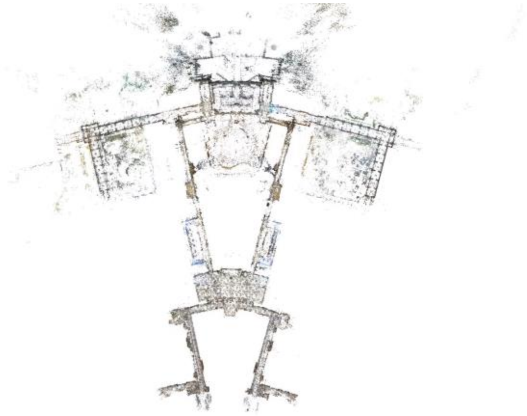
挑战

- 大量弱纹理区域
- 存在视觉歧义
- 场景规模大




关键思路

- 拍摄全景视频
- SLAM与SfM结合
- 分而治之的求解策略
- 精准的稠密深度图估计和融合
- Out-of-Core策略

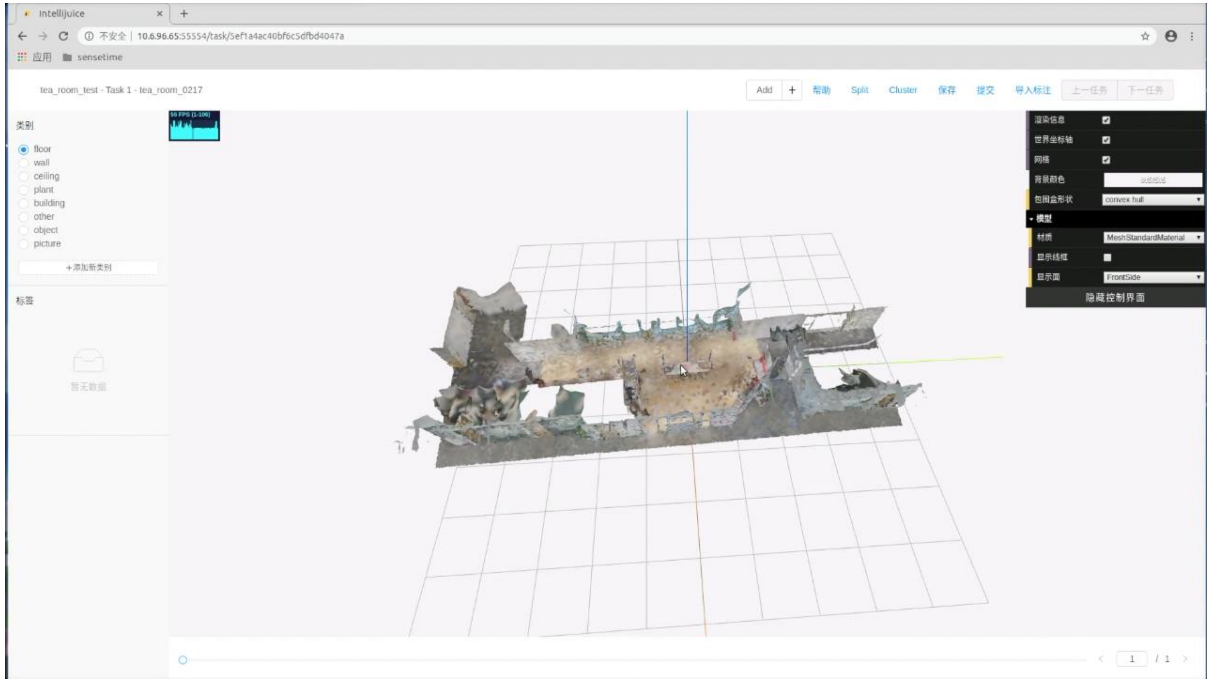

三维地图语义化

挑战

- 标注数据量巨大
- 2D标注效率低

关键思路

- 3D标注
- 半自动标注
  - 算法+手工修正
  - 算法在线训练

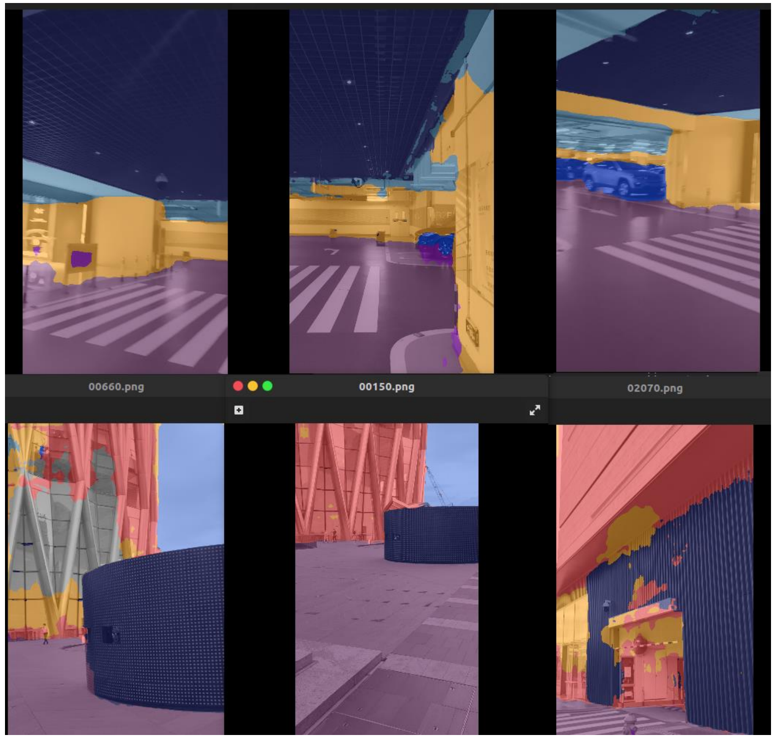


标注速度提升： 1500平米办公室场景，两万张透视图

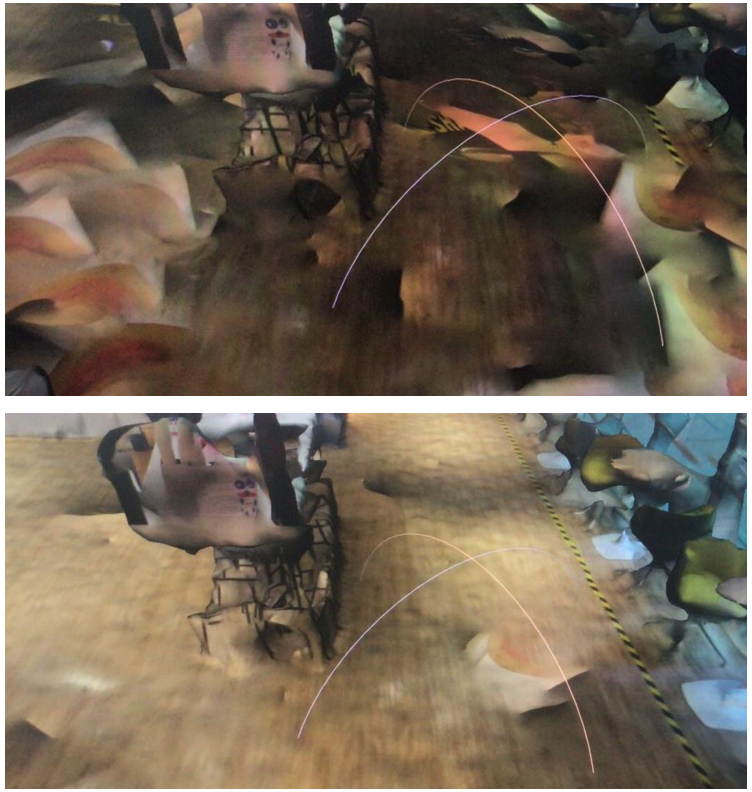
	传统2D标注	3D标注	半自动3D标注
标注时间	14天	4小时	1小时

语义分割提升建图和定位精度

训练语义分割网络实现实时语义分割，可帮助定位与重建



视觉定位重复纹理去除效果



纹理贴图运动人体去除效果

基于语义的弱纹理地面三维补全效果

大尺度场景跟踪定位

松耦合

紧耦合

挑战

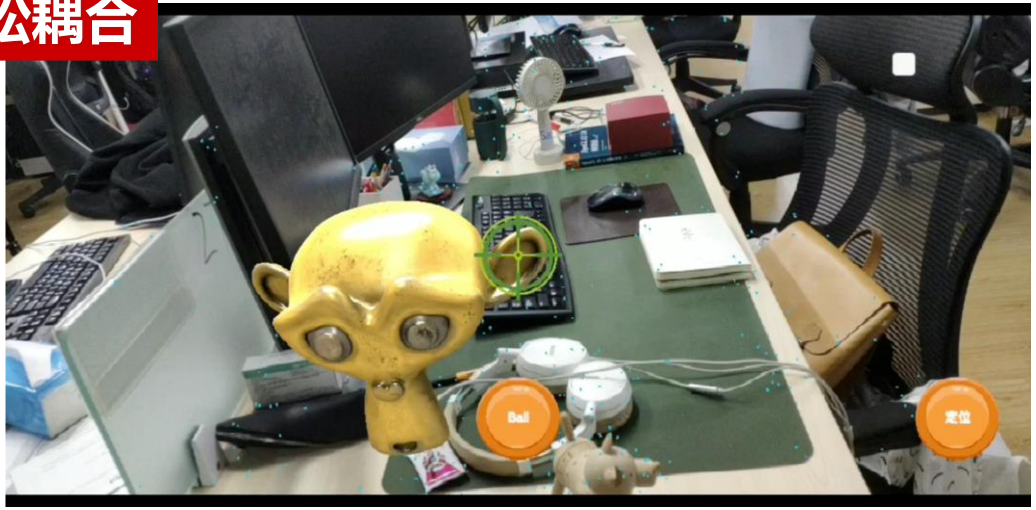
- 在线定位成功率高
- 长距离稳定跟踪
- 视点、光照、外观变化带来的影响

关键思路

- 云和端结合
- 基于高精地图的重定位与SLAM紧耦合
- 基于学习的视觉特征

大尺度场景跟踪定位

松耦合



紧耦合

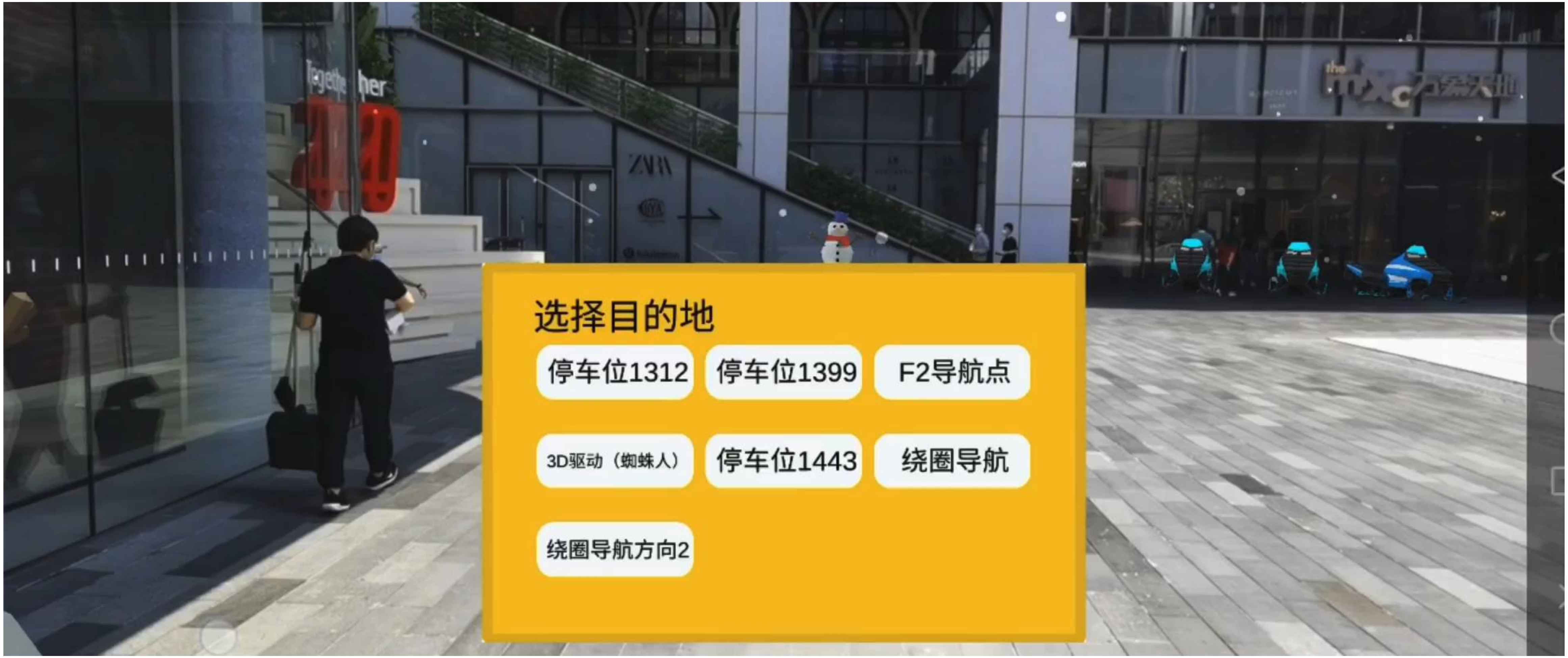
挑战

- 在线定位成功率高
- 长距离稳定跟踪
- 视点、光照、外观变化带来的影响

关键思路

- 云和端结合
- 基于高精地图的重定位与SLAM紧耦合
- 基于学习的视觉特征

定位与AR导航应用


选择目的地
停车位1312
停车位1399
F2导航点
3D驱动（蜘蛛人）
停车位1443
绕圈导航
绕圈导航方向2
ZARA

面向移动AR应用的SLAM性能评测

- 常见的视觉惯性数据集(e.g. EuRoC, TUM VI)
  - 严格时间同步的传感器.
  - 全局快门相机以及高精度IMU.
- 实际手机数据
  - 不可靠的时间同步.
  - 卷帘快门相机以及廉价IMU.
- 难以衡量实际AR应用中的效果.

EuRoC

TUM VI

Real AR Application

视觉惯性数据集

常用VISLAM数据集比较

Dataset	KITTI	EuRoC	TUM VI	ADVIO
Hardware	Car	MAV	Custom Handheld	iPhone 6s
Camera	2×1392×512 10FPS Global Shutter	2×768×480 20FPS Global Shutter	2×1024×1024 20FPS Global Shutter	1×1280×720 60FPS RollingShutter
IMU	OXTS RT 3003 10Hz	ADIS 16488 200Hz	BMI160 200Hz	The IMU of iPhone 6s 100Hz
Ground- truth	OXTS RT 3003 10Hz	VICON/Leica 200Hz	OptiTrack 120Hz (Partially)	Sensor Fusion 100Hz
Environment	Outdoors	Indoors	In-/outdoors	In-/outdoors
Total Distance	39.2 km	0.9 km	20 km	4.5 km
Accuracy	~10 cm	~1 mm	~1 mm	~few dm
Sync	Software	Hardware	Hardware	Software

我们需要一个更为合适的数据集来衡量SLAM在实际AR场景下的表现，同时也希望能有高精度的参考轨迹 (ground-truth).

视觉惯性数据集

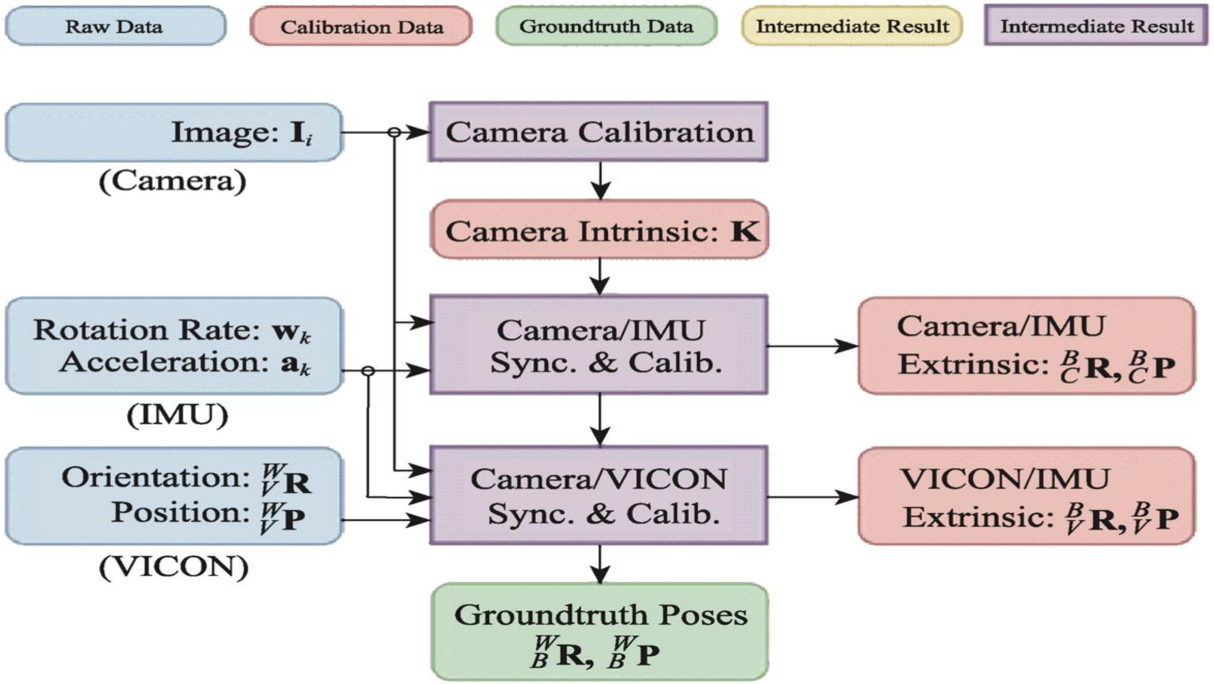
常用VISLAM数据集比较

Dataset	KITTI	EuRoC	TUM VI	ADVIO	Ours
Hardware	Car	MAV	Custom Handheld	iPhone 6s	iPhone X/ Xiaomi Mi 8
Camera	2×1392×512 10FPS Global Shutter	2×768×480 20FPS Global Shutter	2×1024×1024 20FPS Global Shutter	1×1280×720 60FPS RollingShutter	1×640×480 30FPS RollingShutter
IMU	OXTS RT 3003 10Hz	ADIS 16488 200Hz	BMI160 200Hz	The IMU of iPhone 6s 100Hz	The IMU of iPhoneX/ The IMU of Xiaomi Mi 8 100Hz/400Hz
Ground- truth	OXTS RT 3003 10Hz	VICON/Leica 200Hz	OptiTrack 120Hz (Partially)	Sensor Fusion 100Hz	VICON 400Hz
Environment	Outdoors	Indoors	In-/outdoors	In-/outdoors	Indoors
Total Distance	39.2 km	0.9 km	20 km	4.5 km	377 m
Accuracy	~10 cm	~1 mm	~1 mm	~few dm	~1 mm
Sync	Software	Hardware	Hardware	Software	Software

硬件设置 & 数据处理

- 两只不同类型的手机
  iPhone X (相机 640x480 30fps, IMU 100Hz)
  小米 Mi 8 (相机 640x480 30fps, IMU 400Hz)
- 使用VICON系统（400Hz）获得轨迹真值（ground-truth）

手机固定在粘有反光球的支架(VICON Marker)上，用于VICON的运动捕捉



数据集运动和场景类别

- 5 种运动模式：手持静止(hold), 挥动(wave), 瞄准目标(aiming)，左右观测(inspect), 来回走动(petrol)
- 5 种场景类型：杂乱(mess), 整洁(clean), 桌面(desktop), 带高光的地板(floor)
- 3 种时序划分：静止(static), 初始化(initialization), 正常运动(main)
- B0~B7 专门用于衡量重定位等特殊指标

Sequence		Motion	Scene	Description
Xiaomi	A0	inspect+patrol	floor	Walking and looking around the glossy floor.
	A1	inspect+patrol	clean	Walking around some texture-less areas.
	A2	inspect+patrol	mess	Walking around some random objects.
	A3	aiming+inspect	mess+floor	Random objects first, and then glossy floor.
	A4	aiming+inspect	desktop+clean	From a small scene to a texture-less area.
	A5	wave+inspect	desktop+mess	From a small scene to a texture-rich area.
	A6	hold+inspect	desktop	Looking at a small desktop scene.
	A7	inspect+aiming	desktop	Looking at a small desktop scene.
iPhone	B0	rapid-rotation	desktop	Rotating the phone rapidly at some time.
	B1	rapid-translation	desktop	Moving the phone rapidly at some time.
	B2	rapid-shaking	desktop	Shaking the phone violently at some time.
	B3	inspect	moving people	A person walks in and out.
	B4	inspect	covering camera	An object occasionally occluding the camera.
	B5	inspect	desktop	Similar to A6 but with black frames.
	B6	inspect	desktop	Similar to A6 but with black frames.
	B7	inspect	desktop	Similar to A6 but with black frames.

数据集预览

衡量指标

- 跟踪精度
- 初始化质量
- 跟踪鲁棒性
- 重定位时间

跟踪精度

- **4种常见的指标:**

绝对位置误差 (APE)

$$\epsilon_{\mathrm{APE}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \| \mathbf{p}_{\mathrm{SLAM}}[i] - \mathbf{p}_{\mathrm{GT}}[i] \|^2}$$

绝对旋转误差 (ARE)

相对位置误差 (RPE)

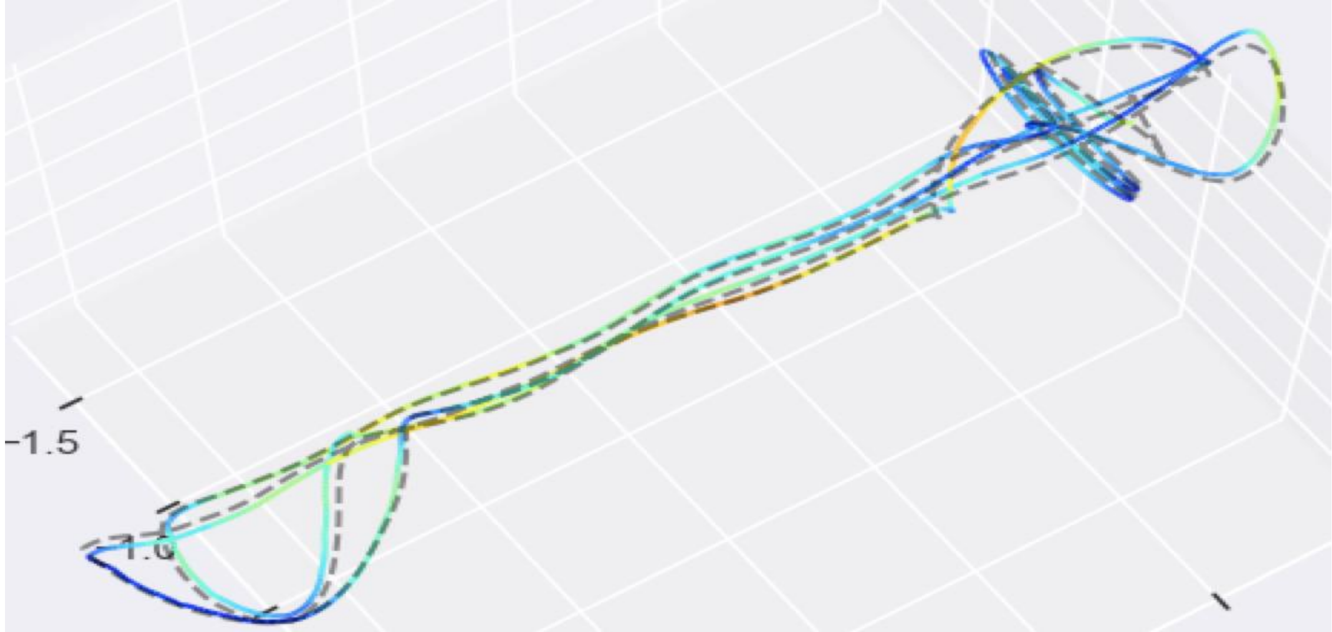
$$\epsilon_{\mathrm{ARE}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m} \| \log(\mathbf{R}_{\mathrm{SLAM}}^{-1}[i]\cdot \mathbf{R}_{\mathrm{GT}}[i]) \|^2}$$

相对旋转误差 (RRE)

- **完整度**

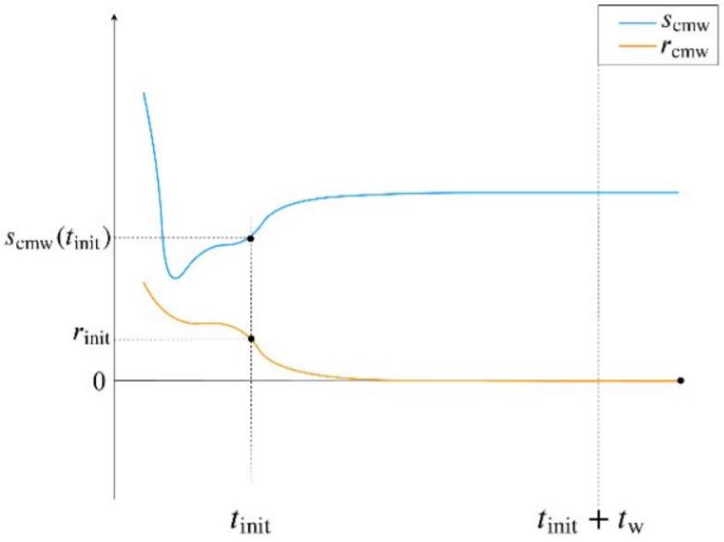
正确位姿占总位姿的百分比

初始化前的位姿未包含在内



初始化质量

- 尺度收敛时间 t_{init}
  - 准确的尺度对于一些AR应用来说十分关键；
  - 在初始状态下，尺度通常会波动，通常在尺度收敛之后放入AR物体。
- 收敛尺度的质量 ϵ_{scale}
  - 对于像AR尺子这样的应用来说非常关键。
- 对于VSLAM而言, 尺度信息是未知的
  - 通过将轨迹与参考轨迹(ground-truth)对齐来获得全局的尺度缩放因子。



$$\epsilon_{\text{scale}} = \frac{1}{2}\left(\left|\frac{s_{\text{cmw}}(t_{\text{init}})}{s_{\text{g}}} - 1\right| + \left|\frac{s_{\text{g}}}{s_{\text{cmw}}(t_{\text{init}})} - 1\right|\right) \times 100\%$$

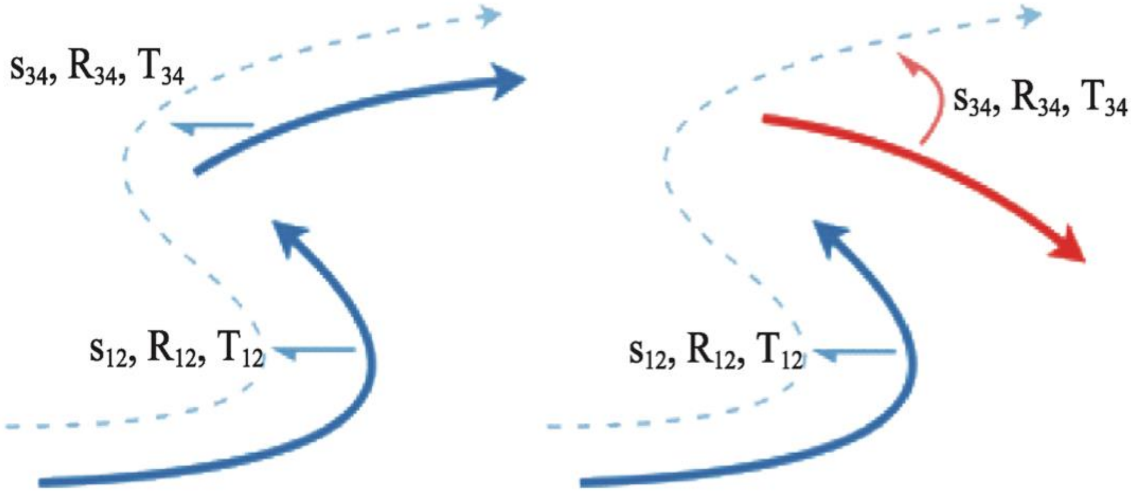
$$\epsilon_{\text{init}} = t_{\text{init}}(\epsilon_{\text{scale}} + \beta)^{\alpha}$$

跟踪质量

- **重定位误差**

  在从跟丢状态恢复回正常状态之后，跟踪结果应该保持一致。

$$\epsilon_{RL} = \sum_{i=1}^{n-1} \| \log_{Sim(3)}(\xi_i^{-1}\xi_{i+1}) \|$$



- 跟丢时间：越短越好
- 越小的跟踪误差越好

$$\epsilon_R = (\alpha_{lost} + \eta_{lost})(\epsilon_{RL} + \eta_{APE}\epsilon_{APE})$$

跟丢时间百分比

绝对位置误差(APE)

重定位时间

- **强制进入跟丢状态**

  手动加入一些纯黑色的帧.

- **重定位时间测量**

  即使没有足够的特征匹配，VISLAM也能根据IMU的积分结果继续跟踪.
  根据轨迹中的跳跃来检测重定位时刻.

$$t_{SLAM[i]} = \min\{t_k > t_{K[i]} \mid \|\mathbf{p}_{SLAM}[k+1] - \mathbf{p}_{SLAM}[k]\| > \delta\}$$

选择8个VSLAM/VISLAM系统进行实验

- **VSLAM**
  - PTAM : http://wiki.ros.org/ethzasl_ptam
  - ORB-SLAM2 : https://github.com/raulmur/ORB_SLAM2
  - LSD-SLAM : https://github.com/tum-vision/lsd_slam
  - DSO : https://github.com/JakobEngel/dso
- **VISLAM**
  - MSCKF : https://github.com/daniilidis-group/msckf_mono
  - OKVIS : https://github.com/ethz-asl/okvis
  - VINS-Mono : https://github.com/HKUST-Aerial-Robotics/VINS-Mono
  - SenseSLAM : http://www.zjucvg.net/senseslam

实验结果

- VSLAM跟踪精度

Sequence		PTAM		ORB-SLAM2		LSD-SLAM		DSO	
APE/RPE (mm)	A0	**75.442**	6.696	96.777	**5.965**	105.963	11.761	231.860	10.456
	A1	113.406	16.344	**95.379**	**10.285**	221.643	23.833	431.929	12.555
	A2	**67.099**	6.833	69.486	5.706	310.963	8.156	216.893	**5.337**
	A3	**10.913**	4.627	15.310	7.386	199.445	10.872	188.989	**4.294**
	A4	21.007	4.773	**10.061**	**2.995**	155.692	10.756	115.477	4.595
	A5	40.403	8.926	**29.653**	11.717	249.644	12.302	323.482	**7.978**
	A6	19.483	3.051	**12.145**	6.741	49.805	3.018	14.864	**2.561**
	A7	13.503	2.462	**5.832**	**1.557**	38.673	2.662	27.142	2.213
ARE/RRE (deg)	A0	12.051	**0.257**	**5.119**	0.342	20.589	0.371	9.983	0.401
	A1	53.954	0.291	**8.534**	**0.242**	51.122	0.288	39.007	0.524
	A2	8.789	0.301	**5.550**	0.255	30.282	0.296	10.584	**0.253**
	A3	6.225	0.293	**1.431**	0.264	31.370	0.475	20.580	**0.241**
	A4	6.295	0.255	**1.015**	**0.157**	9.592	0.498	5.217	0.180
	A5	14.030	0.452	**1.963**	0.546	36.789	0.810	40.939	**0.324**
	A6	2.348	0.217	**0.892**	**0.169**	5.012	0.207	1.435	0.189
	A7	1.218	0.153	**0.569**	**0.115**	3.052	0.147	2.239	0.135
Completeness (%)	A0	**79.386**		65.175		49.513		14.476	
	A1	60.893		**68.303**		11.511		0.869	
	A2	**85.348**		79.263		21.804		22.878	
	A3	71.635		**98.497**		27.112		43.493	
	A4	95.418		**100.000**		64.283		80.371	
	A5	87.399		**97.785**		25.033		2.059	
	A6	97.399		99.786		94.883		**100.000**	
	A7	**100.000**		**100.000**		98.663		**100.000**	

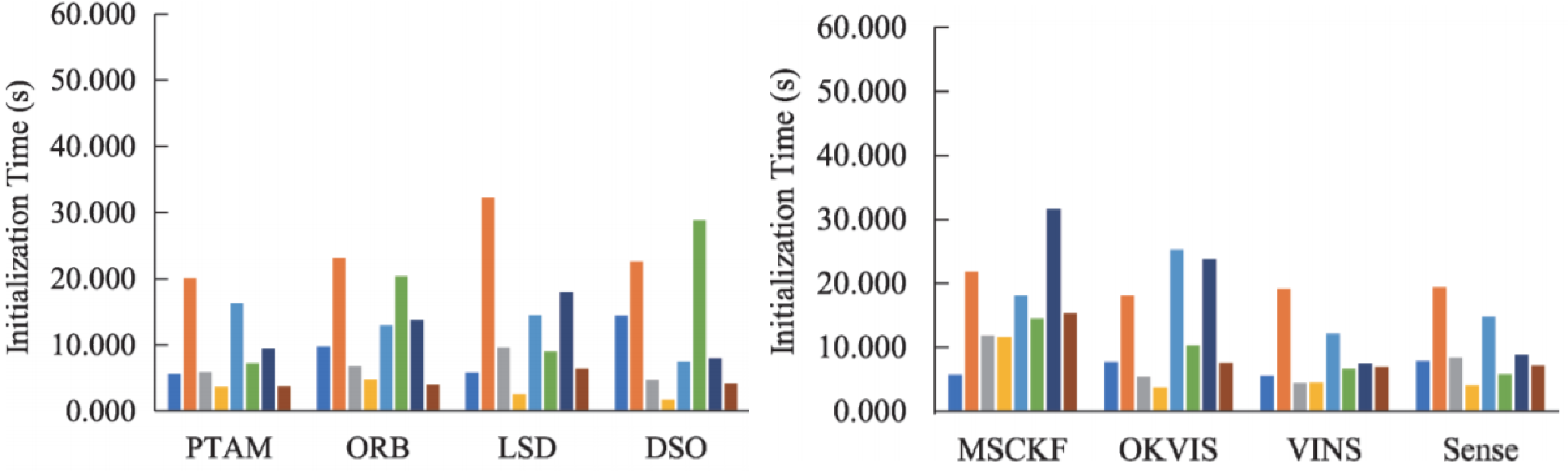
实验结果

- VISLAM跟踪精度

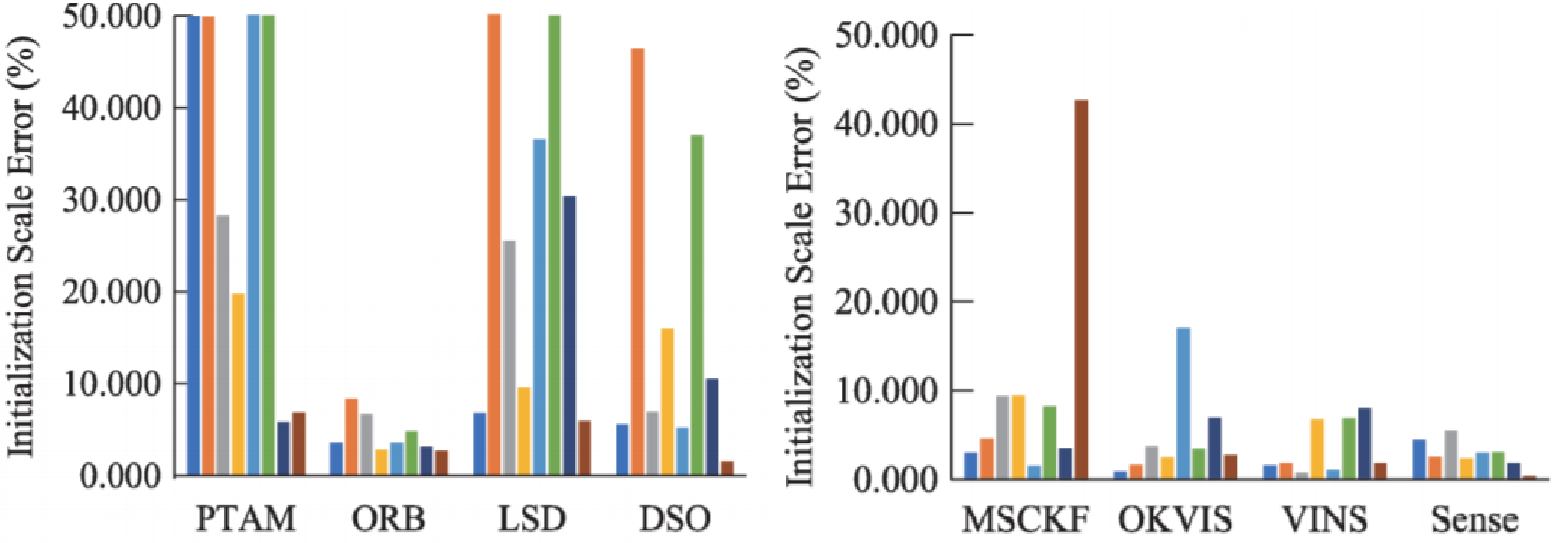
	Sequence	MSCKF		OKVIS		VINS-Mono		SenseSLAM	
APE/RPE(mm)	A0	156.018	7.436	71.677	7.064	63.395	3.510	**58.995**	**2.525**
	A1	294.091	14.580	87.730	4.283	80.687	3.472	**55.097**	**2.876**
	A2	102.657	10.151	68.381	5.412	74.842	8.605	**36.370**	**1.560**
	A3	44.493	3.780	22.949	8.739	19.964	1.234	**17.792**	**0.779**
	A4	114.845	8.338	146.890	12.460	18.691	1.091	**15.558**	**0.930**
	A5	82.885	8.388	77.924	7.588	42.451	2.964	**34.810**	**1.954**
	A6	66.001	6.761	63.895	6.860	26.240	1.167	**20.467**	**0.569**
	A7	105.492	4.576	47.465	6.352	18.226	1.465	**10.777**	**0.831**
ARE/RRE(deg)	A0	6.584	0.203	3.637	0.741	**3.441**	0.205	3.660	**0.197**
	A1	8.703	0.135	5.140	1.098	**1.518**	**0.088**	2.676	0.092
	A2	3.324	0.195	2.493	0.869	1.775	0.201	**1.674**	**0.181**
	A3	6.952	0.186	2.459	0.825	2.121	**0.176**	**1.642**	0.182
	A4	4.031	0.104	3.765	0.603	1.185	**0.063**	**1.129**	0.071
	A5	4.928	0.167	8.843	0.360	3.000	**0.040**	**2.041**	0.089
	A6	2.625	0.170	2.275	0.629	**1.478**	**0.131**	1.656	0.134
	A7	6.810	0.120	3.536	0.602	1.248	**0.073**	**0.502**	0.082
Completeness(%)	A0	40.186		94.255		92.546		**97.317**	
	A1	1.646		**98.235**		86.508		95.072	
	A2	61.423		94.959		88.301		**99.707**	
	A3	97.814		95.972		**100.000**		**100.000**	
	A4	76.629		97.429		**100.000**		**100.000**	
	A5	76.738		98.162		98.795		**99.143**	
	A6	94.128		97.805		**100.000**		**100.000**	
	A7	68.341		96.690		**100.000**		**100.000**	

实验结果

- 初始化时间



- 初始化尺度



实验结果

- 初始化质量

$$\epsilon_{\text{init}} = t_{\text{init}}(\epsilon_{\text{scale}} + \beta)^{\alpha}$$

Sequence	VSLAM				VISLAM			
	PTAM	ORB-SLAM2	LSD-SLAM	DSO	MSCKF	OKVIS	VINS-Mono	SenseSLAM
A0	13.914	2.040	**1.615**	3.783	1.154	1.067	**0.895**	1.840
A1	18.334	**6.930**	25.578	15.598	5.182	**2.892**	3.220	3.674
A2	3.087	1.945	4.980	**1.321**	3.820	1.155	**0.584**	2.154
A3	1.667	0.974	0.810	**0.683**	3.730	**0.690**	1.254	0.764
A4	12.059	2.777	6.404	**1.793**	2.872	10.997	**1.751**	2.967
A5	18.743	**4.062**	12.934	17.815	4.366	2.119	1.866	**1.183**
A6	**2.415**	2.794	5.655	2.699	6.712	6.696	2.246	**1.484**
A7	1.037	0.772	1.624	**0.671**	9.532	1.413	1.164	**0.835**
Average	8.907	**2.787**	7.450	5.545	4.671	3.379	**1.622**	1.863
Max	18.743	**6.930**	25.578	17.815	9.532	10.997	**3.220**	3.674

实验结果

- 跟踪鲁棒性

Sequence	PTAM	ORB-SLAM2	LSD-SLAM	DSO	MSCKF	OKVIS	VINS-Mono	SenseSLAM
B0 (Rapid Rotation)	4.730	0.844	1.911	6.991	–	1.071	2.789	**0.306**
B1 (Rapid Translation)	4.971	0.231	1.090	2.636	–	0.597	1.211	**0.199**
B2 (Rapid Shaking)	5.475	**0.294**	1.387	–	–	3.917	13.403	2.013
B3 (Moving People)	7.455	0.600	0.897	6.399	–	0.673	0.785	**0.465**
B4 (Covering Camera)	16.033	2.702	0.727	–	–	1.976	0.714	**0.326**

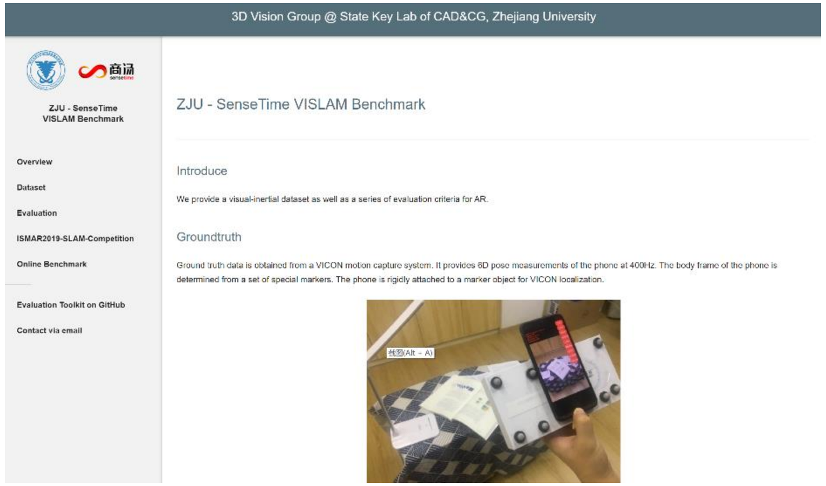
实验结果

- 重定位耗时

Sequence	PTAM	ORB SLAM2	LSD-SLAM	VINS-Mono	SenseSLAM
B5 (1s black-out)	1.032	**0.077**	1.082	5.274	0.592
B6 (2s black-out)	**0.366**	0.465	5.413	3.755	1.567
B7 (3s black-out)	0.651	**0.118**	1.834	1.282	0.332
Average	0.683	**0.220**	2.776	3.437	0.830

ZJU-SenseTime VISLAM Benchmark

- Benchmark网站
  - http://www.zjucvg.net/eval-vislam/
- 评测工具
  - https://github.com/zju3dv/eval-vislam
- 论文



Jinyu Li, Bangbang Yang, Danpeng Chen, Nan Wang, Guofeng Zhang*, Hujun Bao*. Survey and Evaluation of Monocular Visual-Inertial SLAM Algorithms for Augmented Reality. Journal of Virtual Reality & Intelligent Hardware, 1(4): 386 –410, 2019. http://www.vr-ih.com/vrih/html/CN/10.1016/j.vrih.2019.07.002

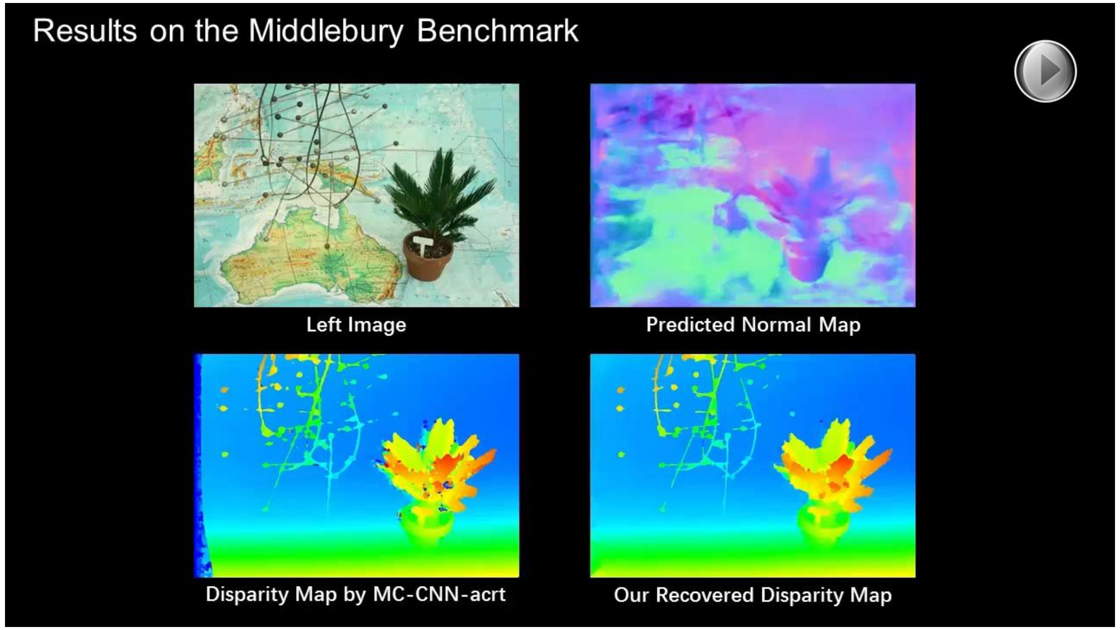
ZJU3DV开源开放计划

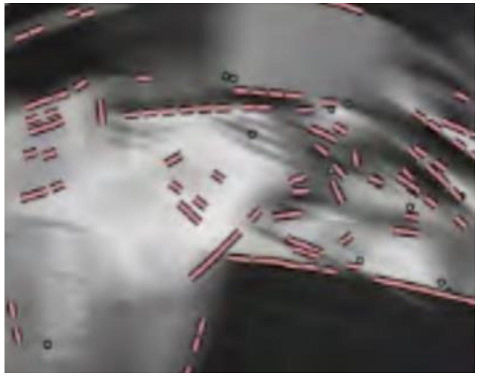
- 已开源或开放的SfM/SLAM系统、算法或数据集
  - ENFT-SfM: https://github.com/zju3dv/ENFT-SfM
  - Segment-based Bundle Adjustment: https://github.com/zju3dv/SegmentBA
  - RKSLAM: http://www.zjucvg.net/rkslam/rkslam.html
  - RDSLAM: http://www.zjucvg.net/rdslam/rdslam.html
  - SenseSLAM v1.0: http://www.zjucvg.net/senseslam/
  - EIBA： https://github.com/zju3dv/EIBA
  - VIG-Init: https://github.com/zju3dv/vig-init
  - RVL-Dynamic: https://github.com/zju3dv/RVL-Dynamic
  - ZJU - SenseTime VISLAM Benchmark: http://www.zjucvg.net/eval-vislam/
- 未来将开源开放更多SLAM算法/系统或数据集
  - http://www.zjucvg.net
  - http://github.com/zju3dv

视觉SLAM技术发展趋势（1）

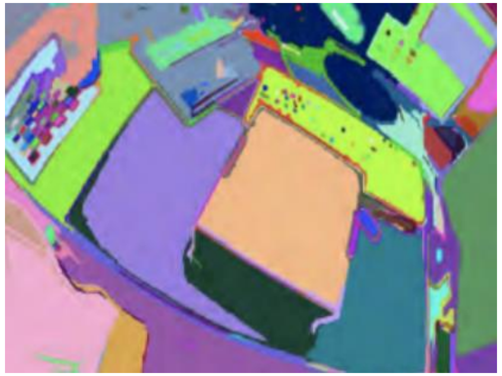
- 缓解特征依赖
  - 基于边、面特征的跟踪
  - 直接图像跟踪或半稠密跟踪
  - 结合机器学习和先验/语义信息

（预测空间布局/语义信息、深度图和法向图等）

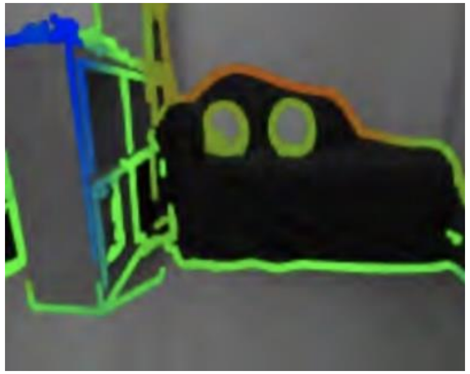




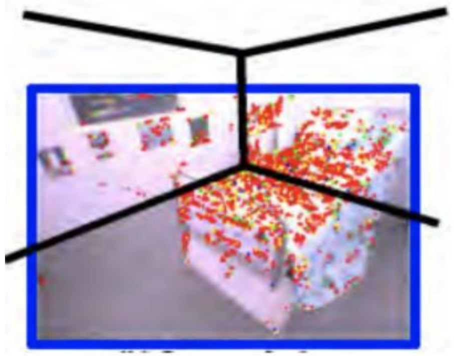
边特征
(Klein et al., 2008)



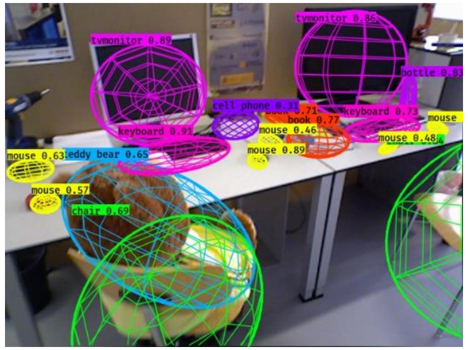
面特征
(Concha et al., 2014)



半稠密区域
(Engel et al.， 2014)



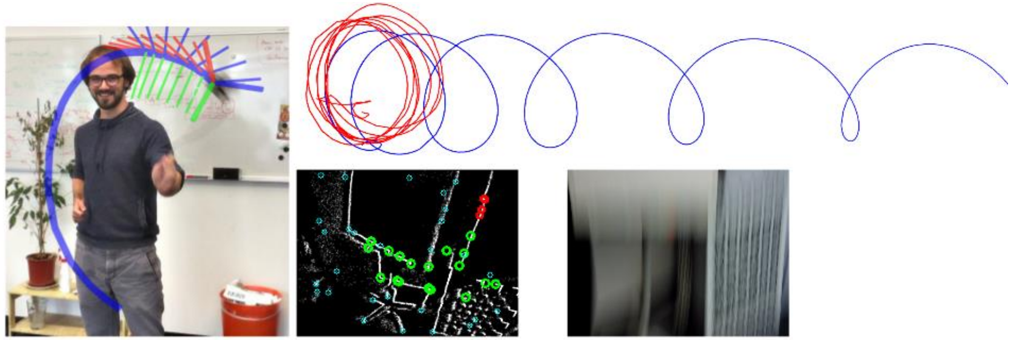
空间布局信息
(Salas et al., 2015)



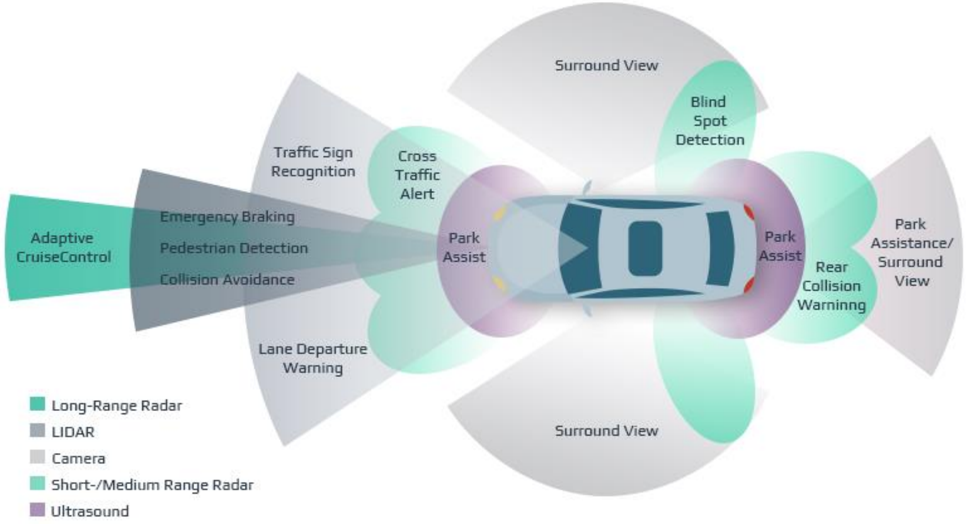
语义SLAM
(Nicholson et al., 2018)

视觉SLAM技术发展趋势（2）

- 多传感器融合
  - 结合IMU、GPS、深度相机、事件相机、光流计、里程计、WiFi、5G

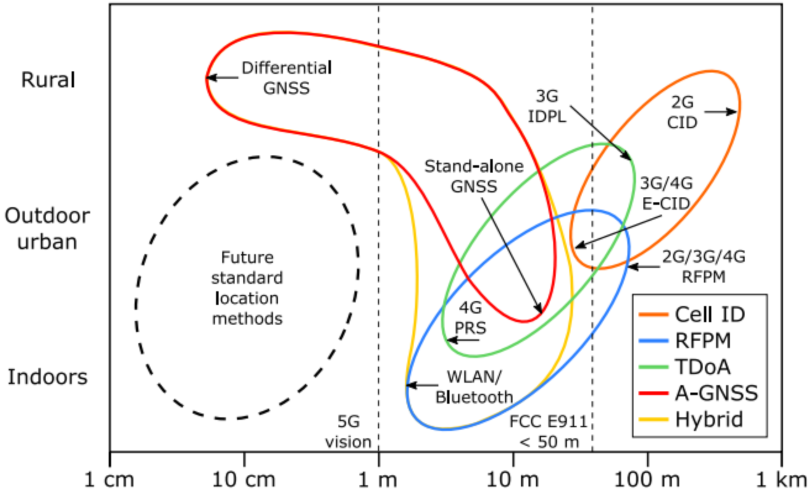


基于事件相机的VIO（Rebecq et al., 2017）



多传感器融合

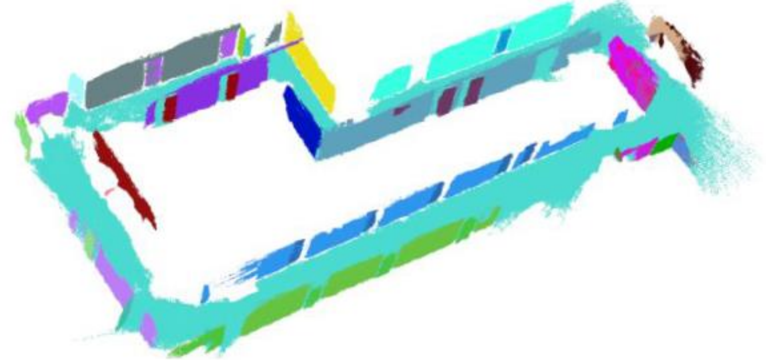
https://www.intellias.com/sensor-fusion-autonomous-cars-helps-avoid-deaths-road/



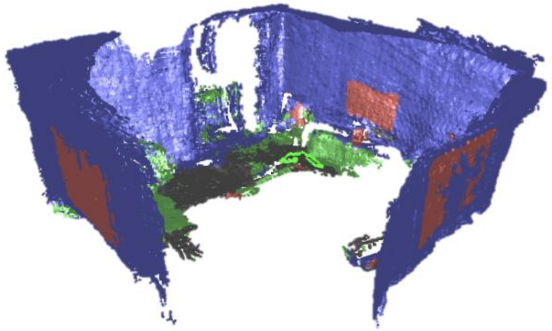
手机无线电定位精度调研（Peral-Rosado et al., 2017）

视觉SLAM技术发展趋势（3）

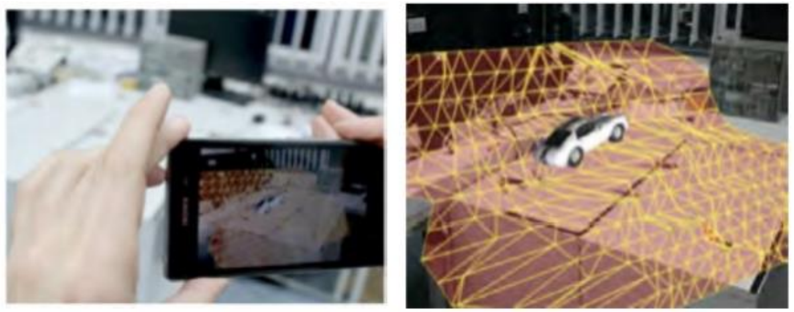
- 稠密三维重建
  - 单 / 多目实时三维重建
  - 基于深度相机的实时三维重建
  - 非刚性物体的实时三维重建



Keyframe-based Dense Planar SLAM (Hsiao et al., 2017)

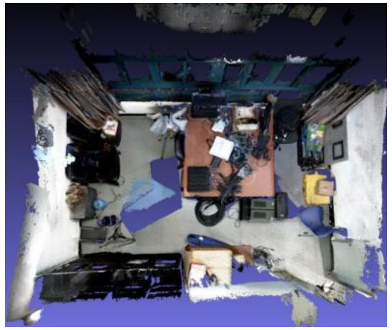


CNN-SLAM (Tateno et al., 2017)



基于稠密三维重建的 AR 应用
(Schöps et al., 2014)

MobileFusion
(Ondrúška et al., 2015)



RKD-SLAM
(Liu et al., 2017)

DynamicFusion
(Newcombe et al., 2015)

谢 谢！