



Robust SFM & SLAM in Challenging Environments

章国锋

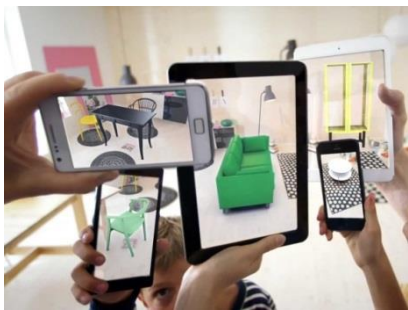
浙江大学**CAD&CG**国家重点实验室

<http://www.zjucvg.net>

<http://www.cad.zju.edu.cn/home/gfzhang/>

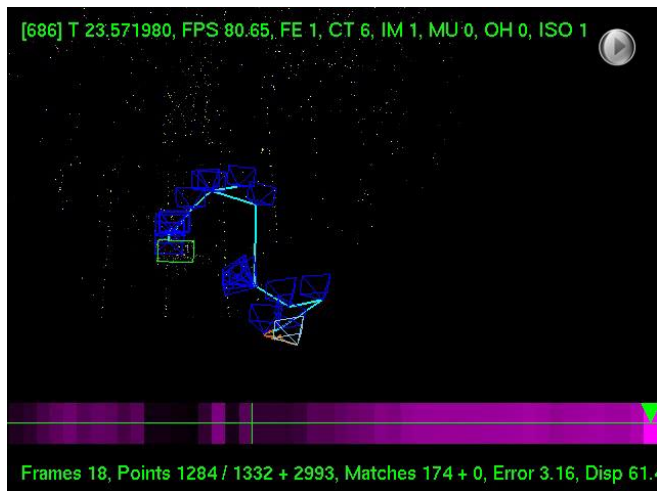
SLAM: 同时定位与地图构建

- 机器人和计算机视觉领域的基本问题
 - 在未知环境中定位自身方位并同时构建环境三维地图
- 广泛的应用
 - 增强现实、虚拟现实
 - 机器人、无人驾驶

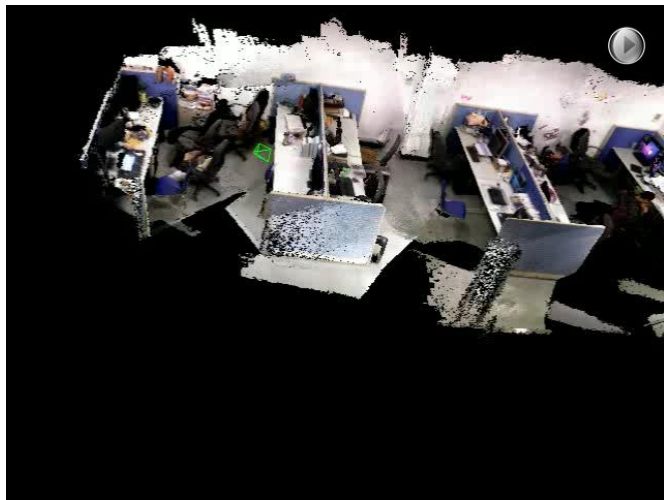


SLAM的运行结果

- 设备根据传感器的信息
 - 计算自身位置（在空间中的位置和朝向）
 - 构建环境地图（稀疏或者稠密的三维点云）



稀疏SLAM



稠密SLAM

SLAM系统常用的框架

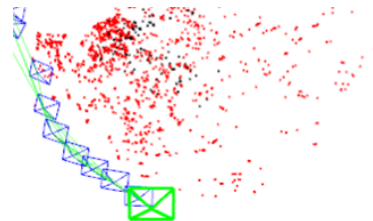
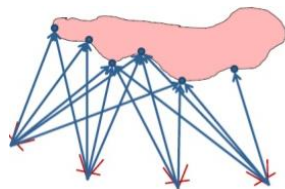
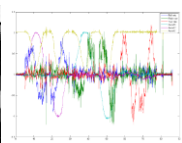
RGB图



深度图



IMU测量值



输入

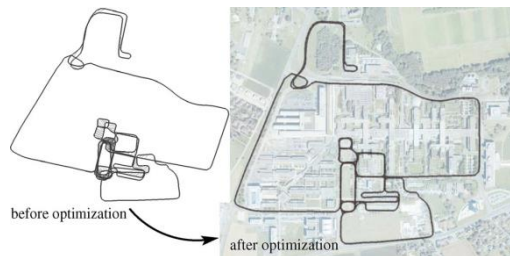
- 传感器数据

前台线程

- 根据传感器数据进行跟踪求解，实时恢复每个时刻的位姿

输出

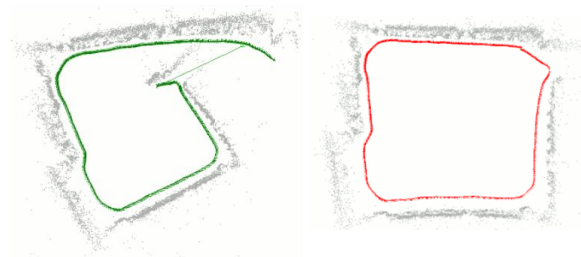
- 设备实时位姿
- 三维点云



优化以减少误差累积

后台线程

- 进行局部或全局优化，减少误差累积
- 场景回路检测



回路检测

SfM & Visual SLAM

■ 主要传感器

- 单目摄像头
- 双目摄像头
- 多目摄像头

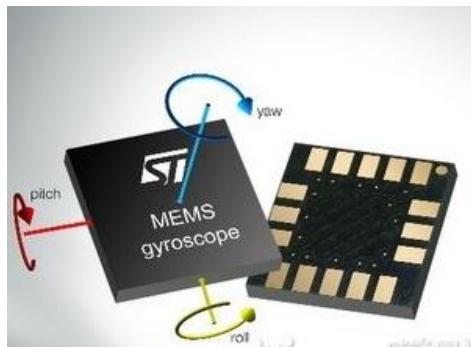


■ 其它辅助传感器

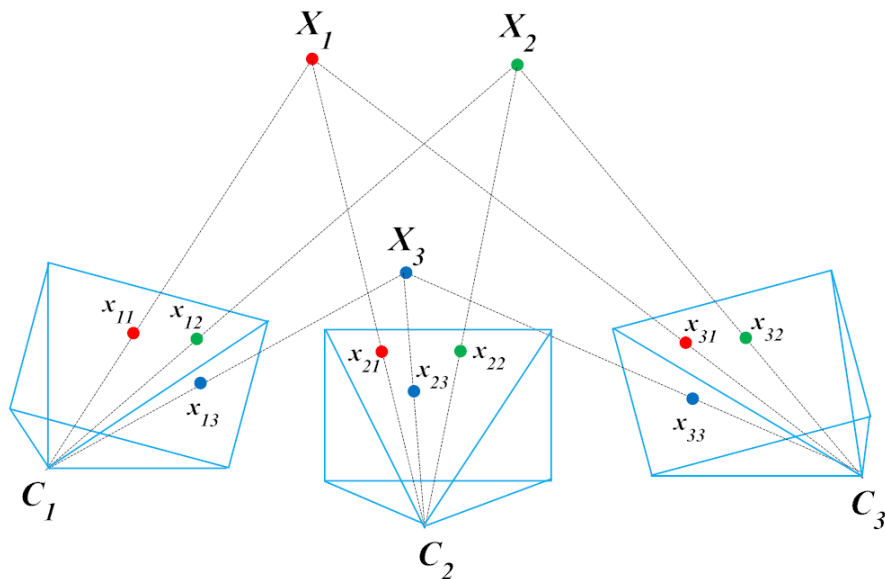
- 廉价IMU、GPS
- 深度摄像头

■ 优势

- 硬件成本低廉
- 小范围内定位精度较高
- 无需预先布置场景



基本原理：多视图几何



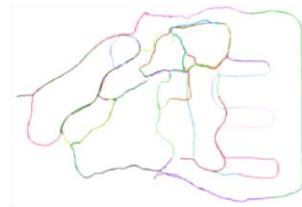
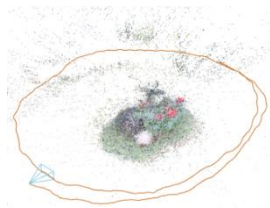
$$\mathbf{x}_{ij} = \pi(\mathbf{P}_i X_j)$$

投影函数 $\pi(x, y, z) = (x/z, y/z)$

$$\mathbf{P}_i = \mathbf{K}_i[\mathbf{R}_i | \mathbf{T}_i]$$

主要挑战

- 如何处理循环回路序列和多视频序列？
- 如何高效高精度地处理大尺度场景？
- 如何处理动态场景？
- 如何处理快速运动和强旋转？



我们课题组的SfM & SLAM工作

■ ENFT-SFM

- 能够高效地处理循环回路和多视频序列

■ 单目SLAM系统

- ENFT-SLAM: 能在大尺度场景下实时运行
- RD-SLAM: 能在动态场景下稳定工作
- RK-SLAM: 可以实时运行在移动设备上

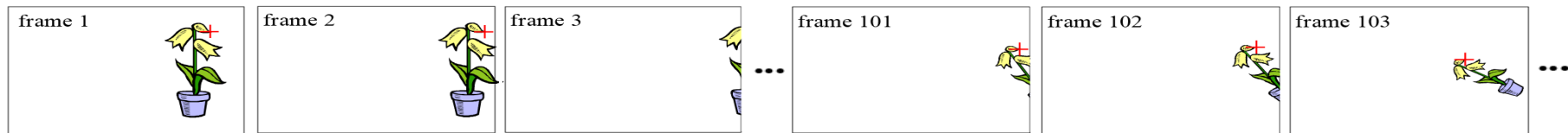
■ RGB-D SLAM: RKD-SLAM



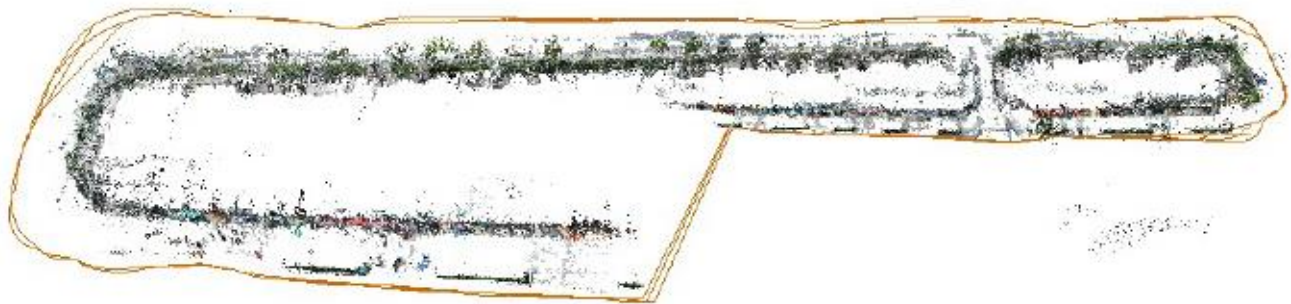
ENFT-SFM: Efficient Non- Consecutive Feature Tracking for Robust SFM

循环回路序列和多视频序列

- 如何将不同子序列上的相同特征点高效地匹配上？

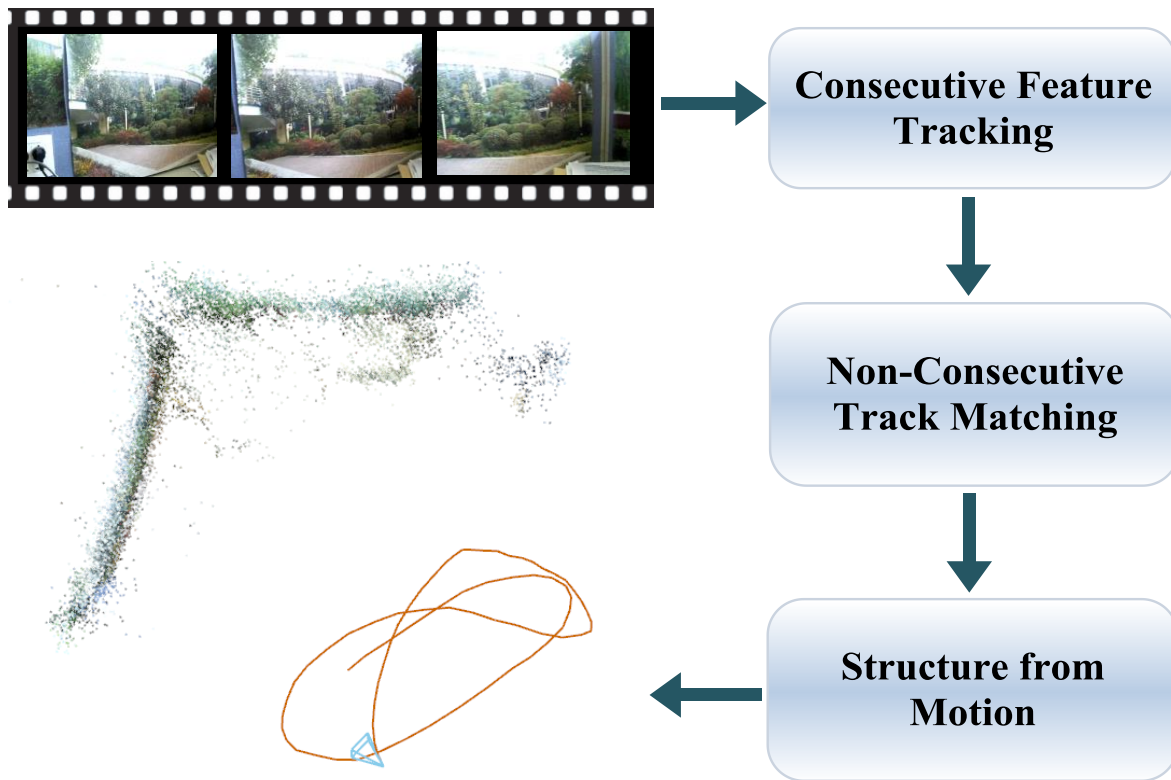


- 如何高效地进行全局优化，消除重建漂移问题？



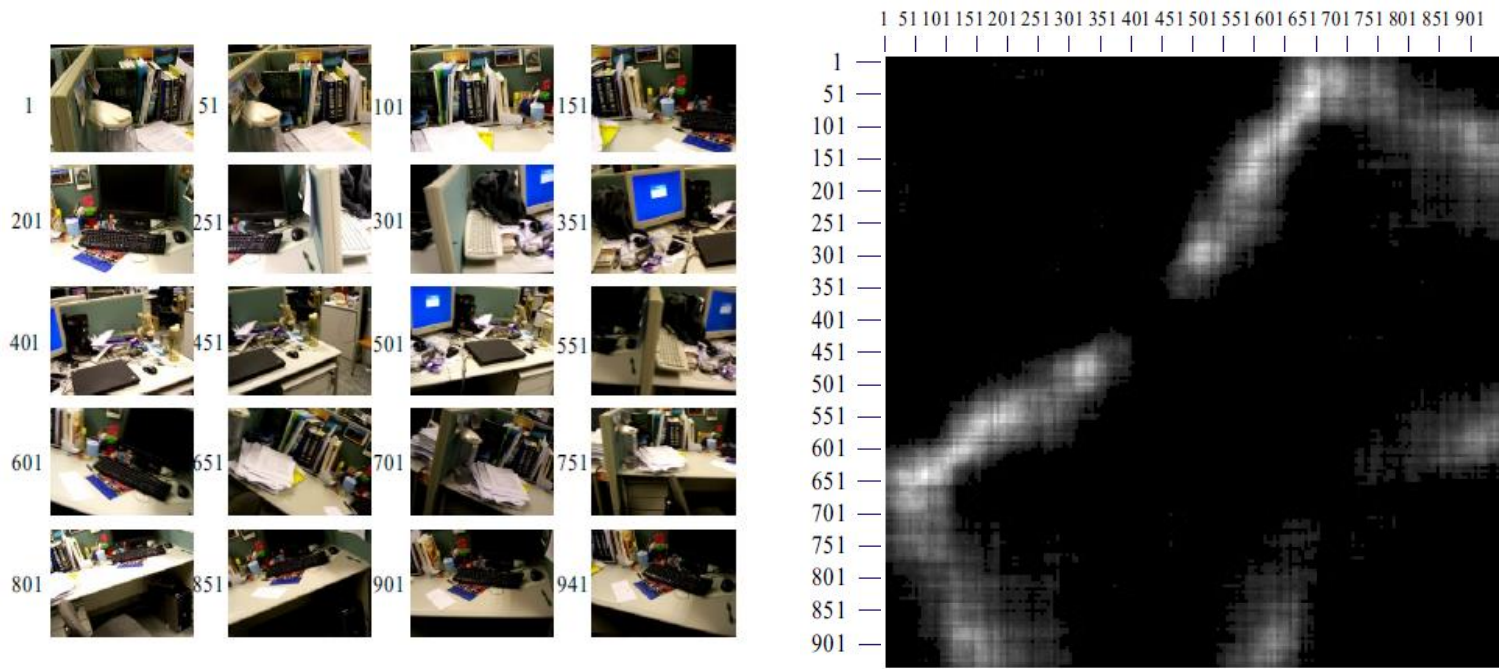
VisualSFM
结果

ENFT: Efficient Non-Consecutive Feature Tracking



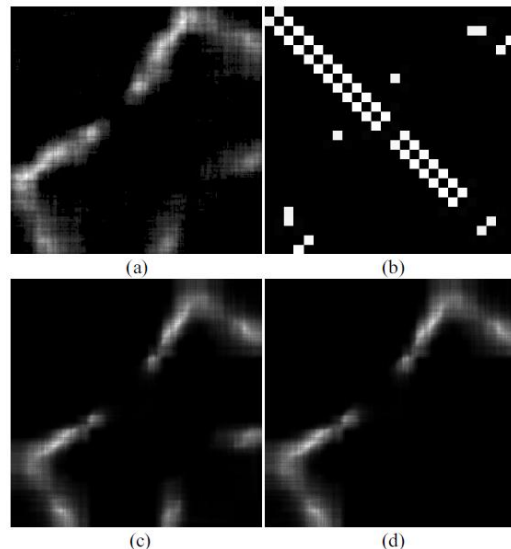
非连续帧上的特征点轨迹匹配

■ 快速匹配矩阵估计



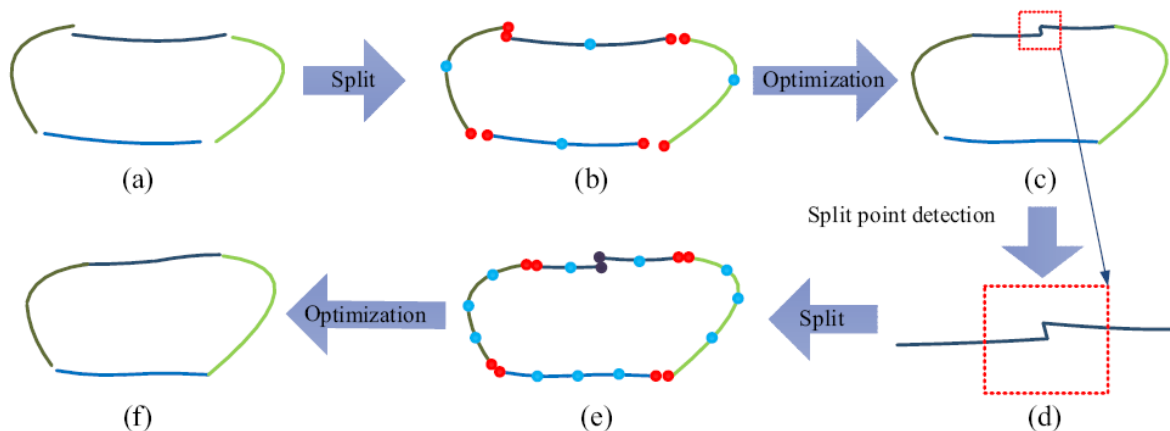
非连续特征轨迹匹配

- 同时进行图像对的特征匹配和优化匹配矩阵
 - 根据选择的图像对的特征匹配结果对匹配矩阵进行优化;
 - 根据更新的匹配矩阵更可靠地选择出有公共内容的图像对进行特征匹配。



基于自适应分段的集束调整

- 将长序列分成若干段短序列，每个短序列进行独立的**SfM**并根据公共匹配对进行对齐，每个段由7个自由度的相似变换控制；
- 如果投影误差比较大，检测分裂点将序列分段，然后优化；
- 重复上述步骤直至投影误差小于阈值或不能再分裂为止。



Input Sequences

Speed: $\times 2$

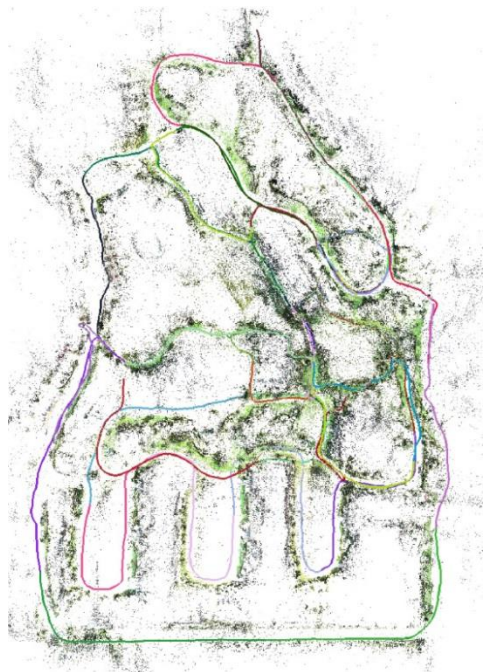


...

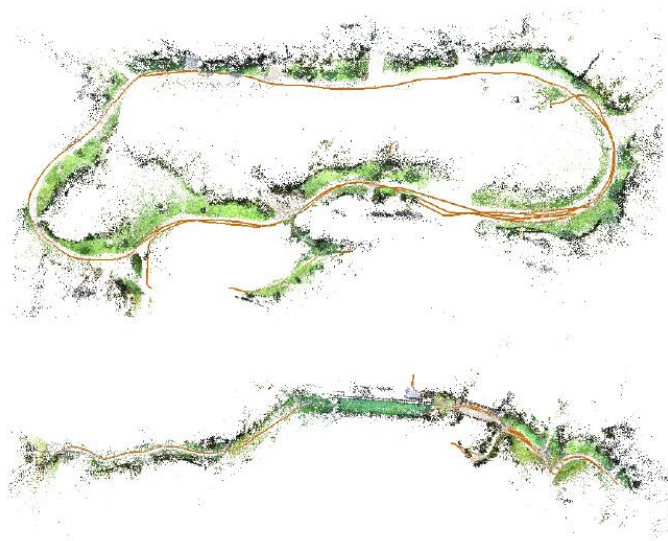
6段长视频序列，将近
10万帧，特征匹配74分
钟，SfM求解16分钟
(单线程)，平均
17.7fps

VisualSfM: SfM求解
57 分钟 (GPU加速)

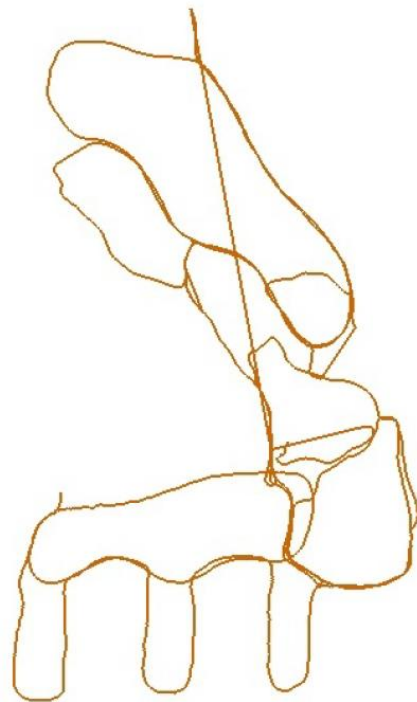
Garden数据集上的比较



ENFT-SFM



VisualSFM



ORB-SLAM

KITTI数据集上的定量比较


TABLE V
LOCALIZATION ERROR (RMSE (M)/COMPLETENESS) COMPARISON IN
KITTI DATASET.

Seq.	ENFT-SFM	ENFT-SFM (Keyframes)	ORB- SLAM	VisualSFM (Keyframes)	OpenMVG (Keyframes)
00	4.58 / 100%	4.76 / 100%	5.33	2.78 / 3.71%	5.83 / 0.7%
01	57.20 / 100%	53.96 / 100%	X	52.34 / 12.46%	8.79 / 2.08%
02	28.13 / 100%	28.26 / 100%	21.28	1.77 / 4.53%	50.36 / 3.74%
03	2.82 / 100%	2.94 / 100%	1.51	0.28 / 12.05%	3.53 / 8.43%
04	0.66 / 100%	0.66 / 100%	1.62	0.76 / 23.44%	5.14 / 14.06%
05	2.88 / 100%	3.48 / 100%	4.85	9.77 / 7.42%	22.42 / 9.07%
06	14.24 / 100%	14.43 / 100%	12.34	8.58 / 7.41%	3.16 / 3.37%
07	1.83 / 100%	2.03 / 100%	2.26	3.85 / 7.78%	7.75 / 5%
08	30.74 / 100%	28.32 / 100%	46.68	0.81 / 0.90%	17.82 / 2.58%
09	5.63 / 100%	5.88 / 100%	6.62	0.90 / 4.92%	14.26 / 3.36%
10	19.53 / 100%	18.49 / 100%	8.8	5.70 / 6.05%	27.06 / 7.01%

TUM数据集上的定量比较

TABLE VI
LOCALIZATION ERROR (RMSE (CM)/COMPLETENESS) COMPARISON
IN TUM RGB-D BENCHMARK.

Sequence	ENFT-SFM	ENFT-SFM (Keyframes)	ORB- SLAM	VisualSFM (Keyframes)	OpenMVG (Keyframes)
fr1_desk	2.71/99.84%	2.96/100%	1.69	2.74 / 100%	X
fr1_floor	4.08/96.70%	3.93/100%	2.99	53.11/69.23%	0.52/6.92%
fr1_xyz	1.25/100%	1.59/100%	0.9	1.43/100%	X
fr2_360_kidnap	13.57/91.47%	15.31/100%	3.81	10.08/50.91%	5.21/14.55%
fr2_desk	2.43/100%	2.27/100%	0.88	1.79/100%	1.38/13.95%
fr2_desk_person	2.46/100%	2.55/100%	0.63	1.92/100%	2.16/97.01%
fr2_xyz	0.81/100%	0.73/100%	0.3	0.71/100%	5.74/97.6%
fr3_long_office	1.21/100%	1.44/100%	3.45	1.15/100%	2.94/32.74%
fr3_nst_tex_far	3.60/86.58%	7.76/100%	X	7.29/100%	35.64/3.79%
fr3_nst_tex_near	1.87/100%	1.66/100%	1.39	1.13/100%	3.4/39.13%
fr3_sit_half	1.50/100%	1.55/100%	1.34	2.30/100%	0.68/9.3%
fr3_sit_xyz	0.84/100%	1.39/100%	0.79	1.28/100%	1.03/100%
fr3_str_tex_far	0.94/100%	0.95/100%	0.77	2.15/100%	1.12/100%
fr3_str_tex_near	1.86/100%	1.82/100%	1.58	0.95/100%	0.97/19.74%
fr3_walk_half	2.08/100%	2.21/100%	1.74	1.88/100%	X
fr3_walk_xyz	1.30/100%	1.74/100%	1.24	1.62/100%	X



ENFT-SLAM: ENFT-based Large-Scale Monocular SLAM

ENFT-SLAM

- 特征跟踪

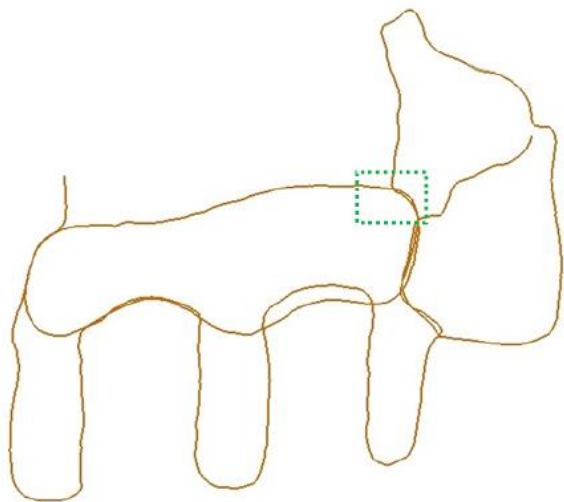
- 直接采用ENFT特征跟踪

- 回路检测与闭合

- 对原来的非连续特征轨迹匹配进行修改
 - 计算当前帧与历史关键帧的相似度，并选择相似度高的关键帧进行匹配
 - 采用基于分段的集束调整进行优化

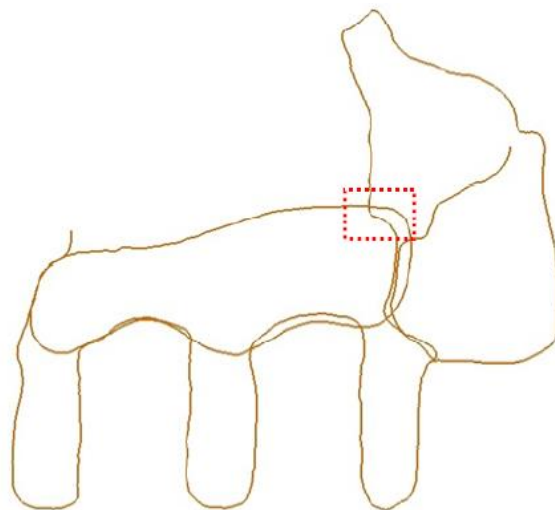


ENFT-SLAM vs. ORB-SLAM



ENFT-SLAM

Non-consecutive Track Matching
Segment-based BA



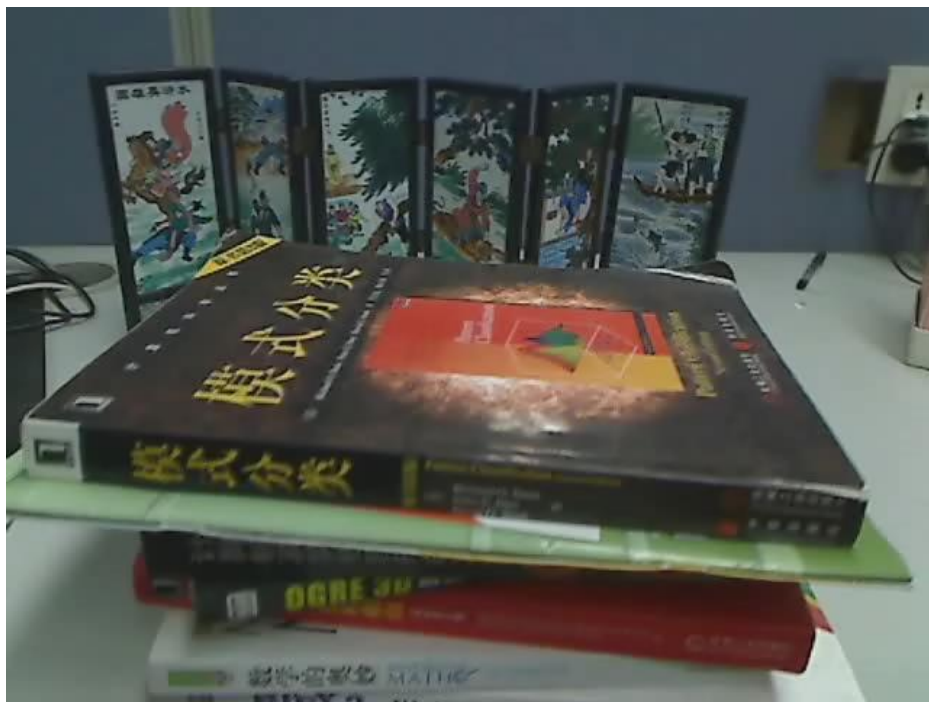
ORB-SLAM

Bag-of-words Place Recognition
Pose Graph Optimization + Traditional BA

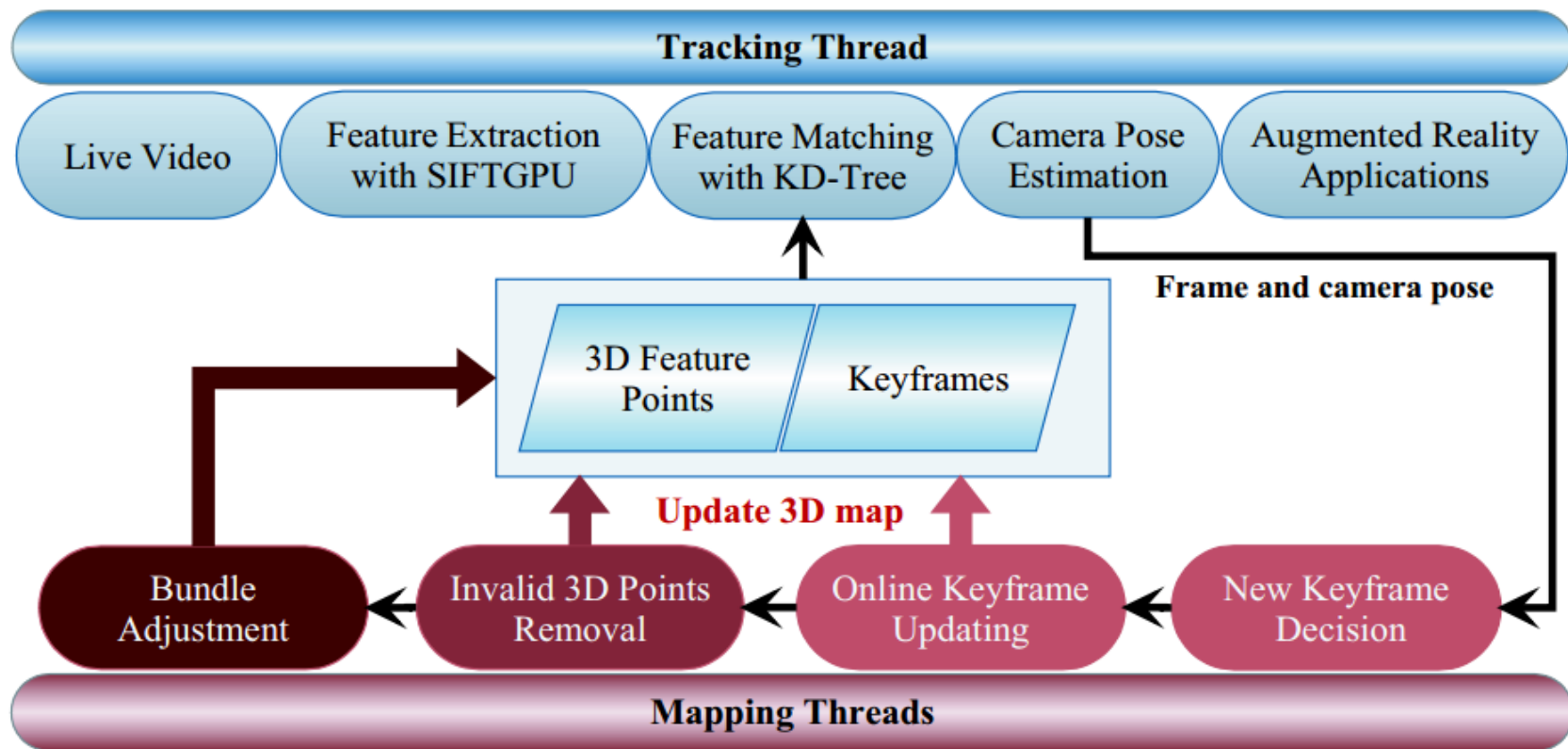
动态场景SLAM的主要问题（1）



动态场景SLAM的主要问题（2）



RDSLAM系统框架



Our SLAM Result





Robust Keyframe-based Monocular SLAM for Augmented Reality

Keyframe-based SLAM vs Filtering-based SLAM

■ 优点

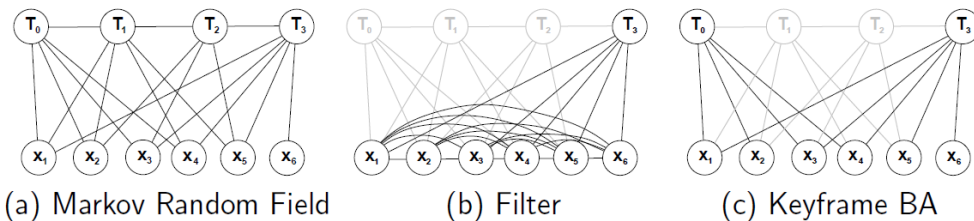
- 精度高
- 效率高
- 扩展性好

■ 缺点

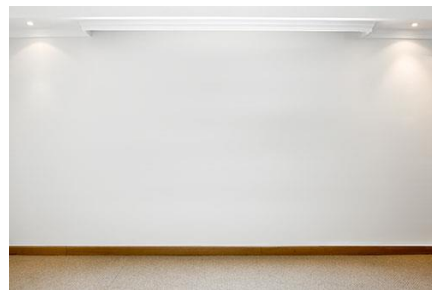
- 对强旋转很敏感

■ 共同的挑战

- 快速运动
- 运动模糊
- 特征不够丰富



H. Strasdat, J. Montiel, and A. J. Davison. Visual SLAM: Why filter? Image and Vision Computing, 30:65-77, 2012.

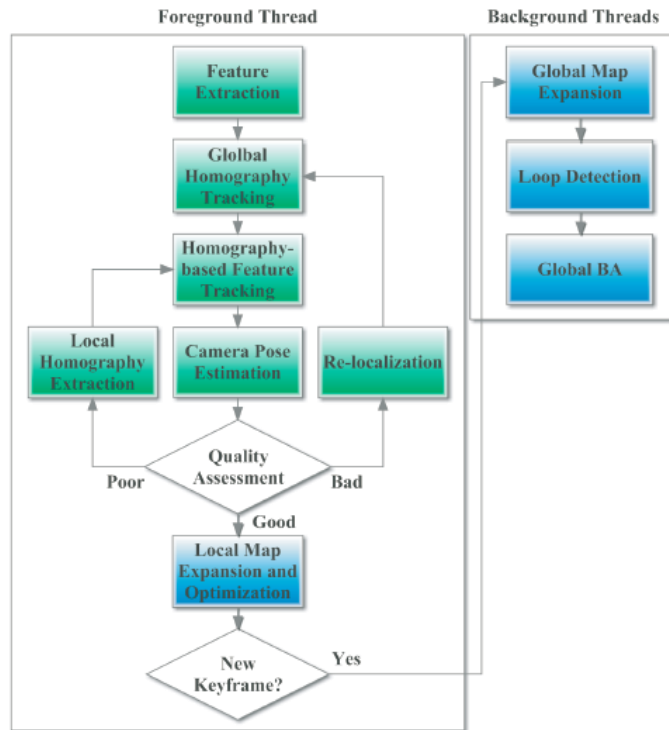


Visual-Inertial SLAM

- 使用IMU数据提高鲁棒性
 - 基于滤波的方法
 - MSCKF, SLAM in Project Tango, ...
 - 基于非线性优化的方法
 - OKVIS, ...
- 没有真实IMU数据的情况下，能否工作？

RKSLAM系统框架

- 基于多单应矩阵的跟踪
 - **FAST corners**
 - 基于滑动窗口的姿态优化
 - 用整张图像对齐来估计旋转角速度
- $$\hat{\omega}_i = \arg \min_{\omega} \left(\sum_{\mathbf{x} \in \Omega} \|\tilde{I}_i(\mathbf{x}) - \tilde{I}_{i+1}(\pi(\mathbf{K}\mathbf{R}_{\Delta}(\omega, t_{\Delta_i})\mathbf{K}^{-1}\mathbf{x}^h))\|_{\delta_I} \right. \\ \left. + \sum_{(\mathbf{x}_i, \mathbf{x}_{i+1}) \in M_{i,i+1}} \frac{1}{\delta_x} \|\pi(\mathbf{K}\mathbf{R}_{\Delta}(\omega, t_{\Delta_i})\mathbf{K}^{-1}\mathbf{x}_i^h) - \mathbf{x}_{i+1}\|_2^2 \right)$$
- 采用模拟的IMU数据进行姿态优化



TUM RGB-D数据集上的定量比较

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%

From left to right: RMSE (cm) of keyframes, the starting ratio (i.e. dividing the initialization frame index by the total frame number), and the tracking success ratio after initialization.

Group A: simple translation

Group B: there are loops

Group C: slow and nearly pure rotation

Group D: fast motion with strong rotation

时间统计

■ 台式机上的计算时间

Module	Time per frame
Feature extraction	~ 2 ms
Feature tracking	2 ~ 8 ms
Local map expansion and optimization	2 ~ 4 ms

Table 1: Process time per frame with a single thread.

■ 移动终端上

□ 20~50 fps on an iPhone 6.

Comparison

Speed: $\times 2$



Our method with rotational velocity estimation



ORB-SLAM



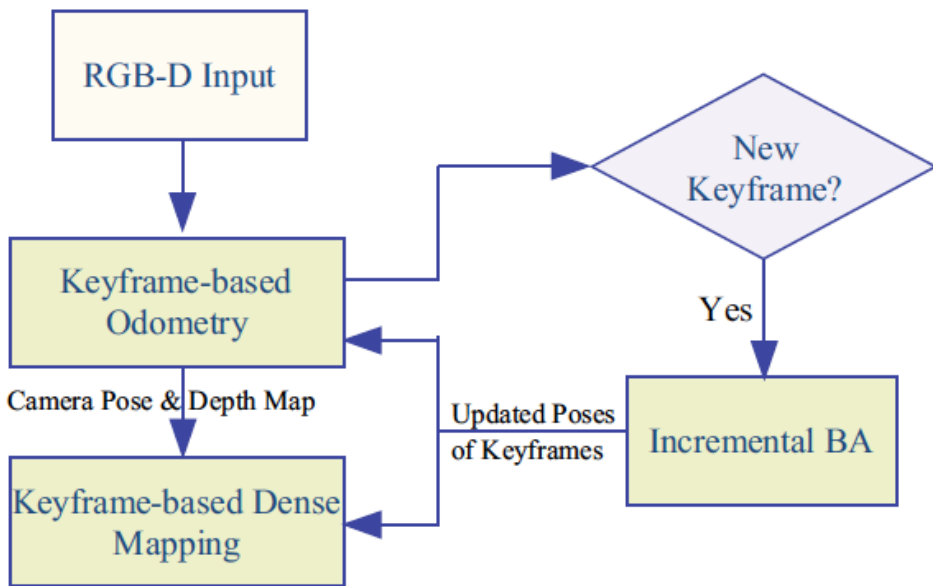
PTAM



RDSLAM

RKD-SLAM系统框架

- 非常快速鲁棒的基于RGB-D的跟踪方法(CPU下约100 fps)
- 非常快速的增量集束调整算法
- 非常高效的基于关键帧的深度表达和融合方法

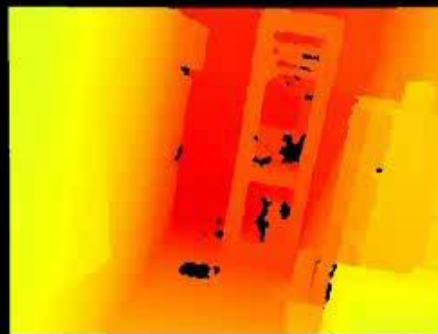


支持快速运动、回路闭合、重定位和长时间运行

鲁棒处理快速运动

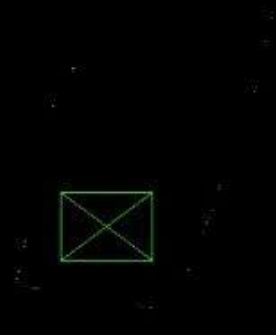
Our Result

Speed 1.2

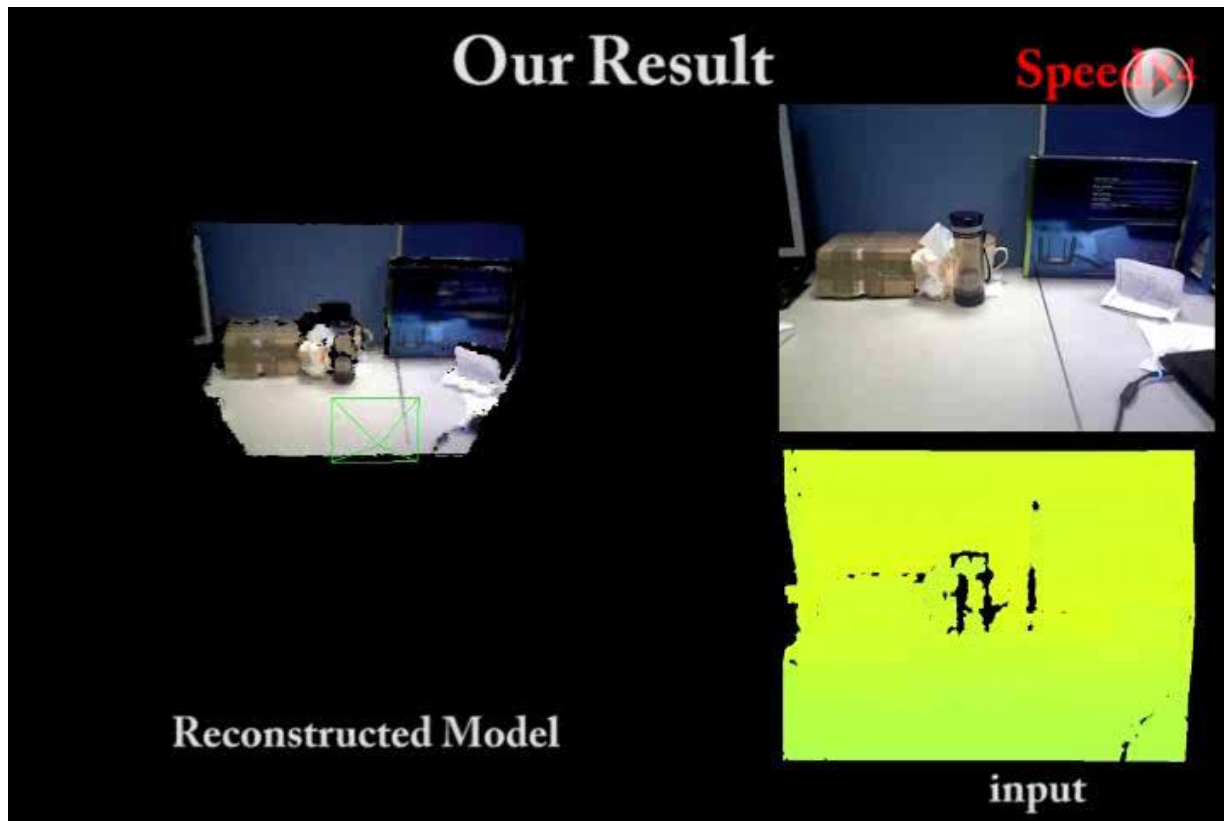


input

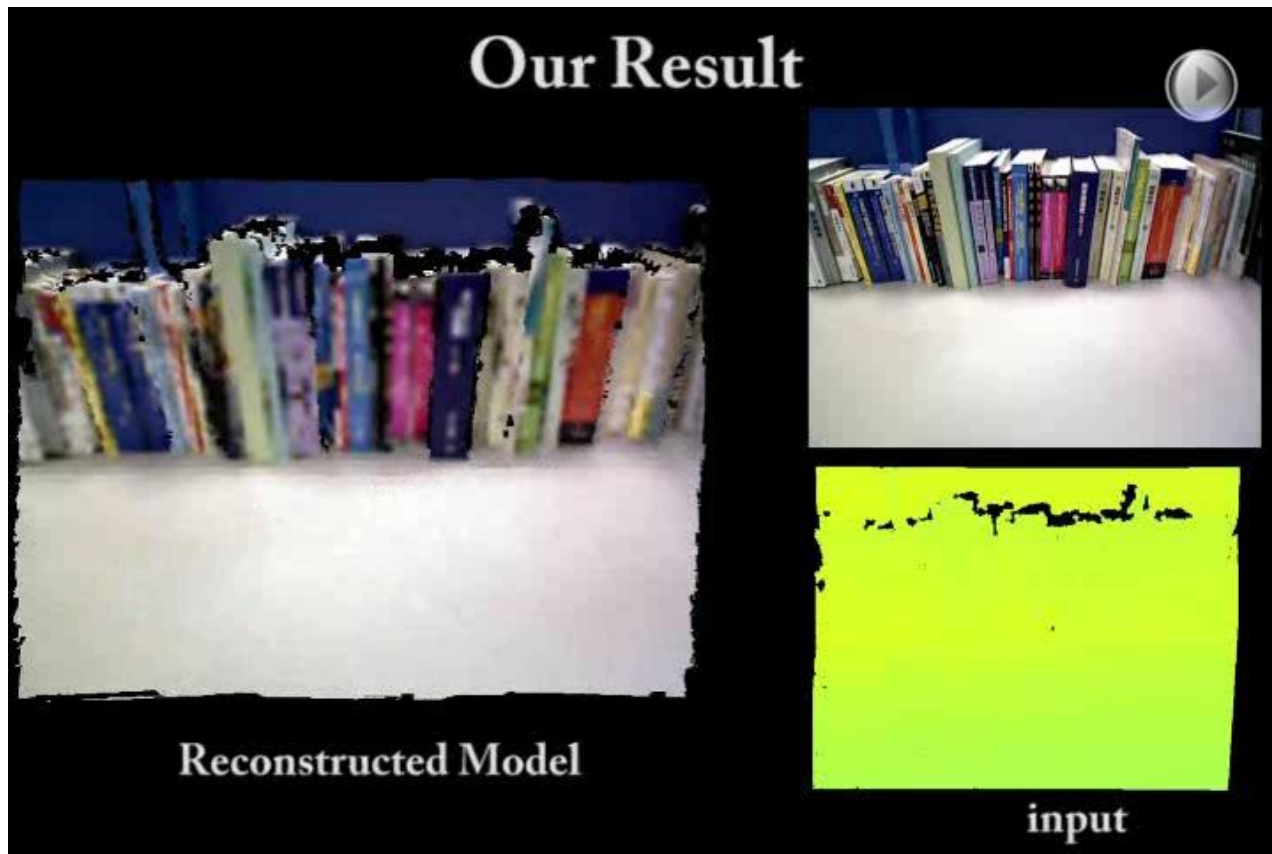
Reconstructed Model



在线的回路闭合和三维表面调整



支持重定位



时空一致性深度恢复

Input Sequence



Recovered Depth Video



视频场景编辑

Video Scene Editing

软件或源代码

- ENFT-SFM or LS-ACTS

- <http://www.zjucvg.net/ls-acts/ls-acts.html>

- RKSLAM:

- <http://www.zjucvg.net/rkslam/rkslam.html>

- RDSLAM:

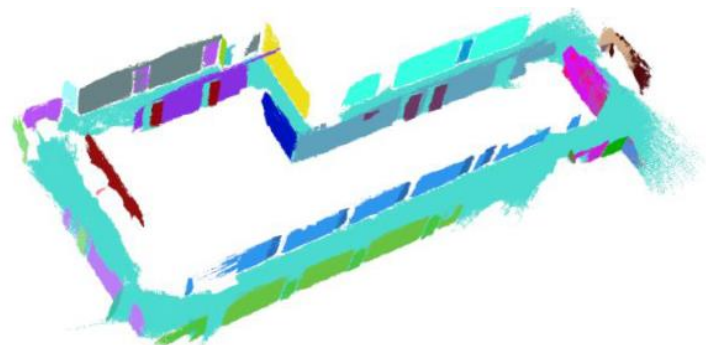
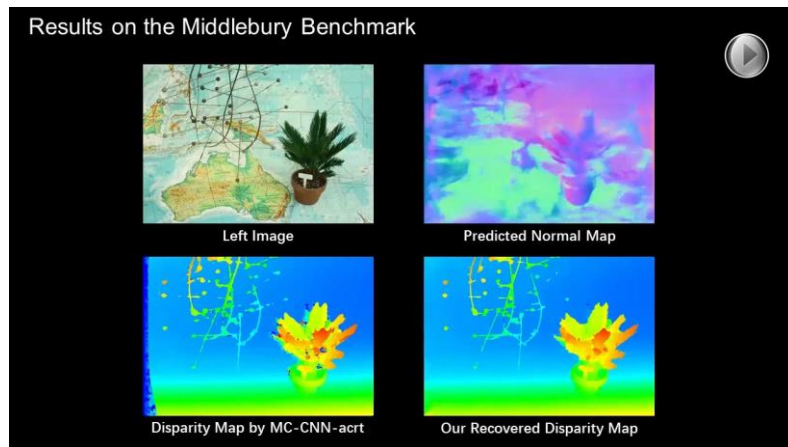
- <http://www.zjucvg.net/rdslam/rdslam.html>

- ACTS:

- <http://www.zjucvg.net/acts/acts.html>

Visual SLAM技术发展趋势（1）

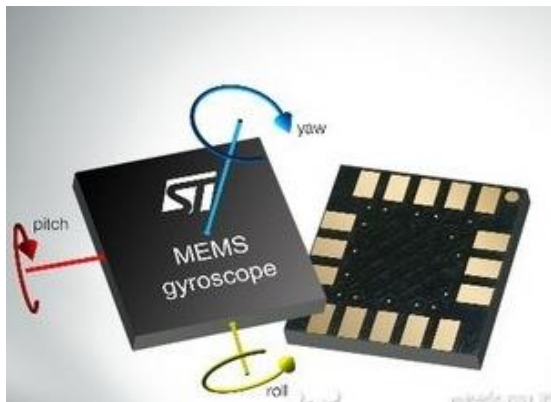
- 缓解特征依赖
 - 基于边的跟踪
 - 直接图像跟踪或半稠密跟踪
 - 结合机器学习和先验/语义信息
- 稠密三维重建
 - 单 / 多目实时三维重建
 - 基于深度相机的实时三维重建
 - 平面表达和模型自适应简化



Visual SLAM技术发展趋势（2）

■ 多传感器融合

□ 结合IMU、GPS、深度相机、光流计、里程计



未来工作展望

- 协同SLAM
- 稠密SLAM
- 场景分析和理解
- 在VR/AR、机器人和无人驾驶领域进行应用





谢谢！

ZJUCVG Group: <http://www.zjucvg.net/>

Email: zhangguofeng@cad.zju.edu.cn