城市空间建模与仿真

第十五讲　激光与视觉SLAM

任课教师：汤圣君
建筑与城市规划学院　城市空间信息工程系

目录
CONTENTS

SLAM

- Simultaneous Localization and Mapping
  - Localization: estimating the sensor's pose (location and orientation)
  - Mapping: building a map
  - SLAM: building a map and locating the sensor at the same time

- A chicken-and-egg problem
  - A map is needed for localization
  - A pose estimate is needed for mapping

Visual SLAM Demos

Universidad Zaragoza 1542

ROBOTICS UNIVERSIDAD DE ZARAGOZA

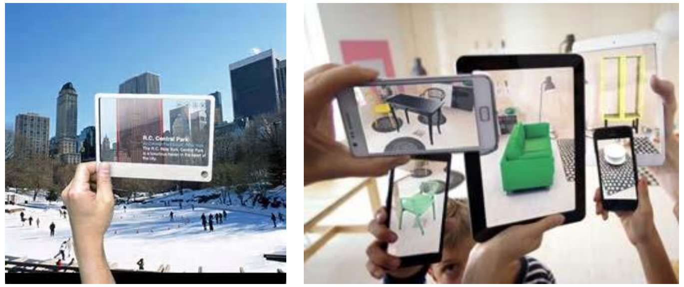
Instituto Universitario de Investigación en Ingeniería de Aragón Universidad Zaragoza

ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras

Raúl Mur-Artal and Juan D. Tardós

raulmur@unizar.es tardos@unizar.es

Visual SLAM Applications



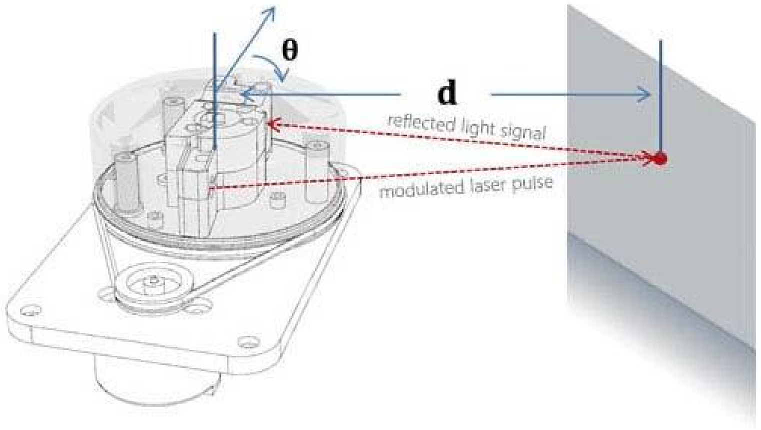
Augmented Reality

Microsoft Hololens

5

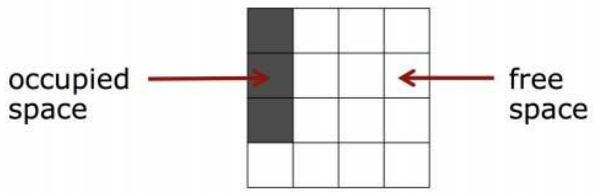
2D LiDAR SLAM

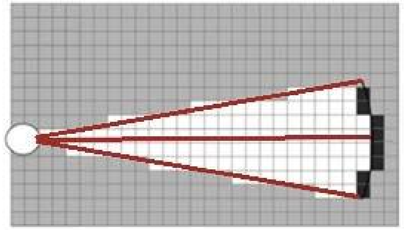
- Widely used for ground robots
- Use a 2D LiDAR sensor to track planar motion
- 2D LiDAR sensor
  - With a rotating laser beam
  - Return distances to obstacles in a plane
  - E.g. 10×360 measurements per second

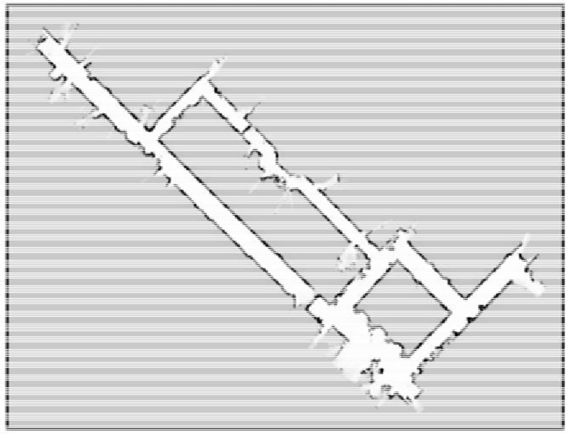


Map Representation (Occupancy Grid Map)

- Discretize the environment by a grid
  - E.g. $10 \text{ m} \times 20 \text{ m}$ space, 5 cm resolution → 200×400 map
  - Large maps require substantial memory resources
- Each grid cell can be empty, occupied, or unknown
  - E.g. white is empty, black is occupied, and grey is unknown
- Each laser beam tells the occupancy of some cells
  - Mark empty for cells on its path
  - Mark occupied for the cell at its end







Mapping with Known Poses

- Suppose the robot pose is known at all time
  - E.g. by some SLAM algorithms
- Accumulate empty/occupied votes from the LiDAR sensor over time
  - A cell is occupied, if the number of occupied votes is larger (by a threshold)
  - A cell is empty, if the number of empty votes is larger (by a threshold)
  - Otherwise, a cell is unknown

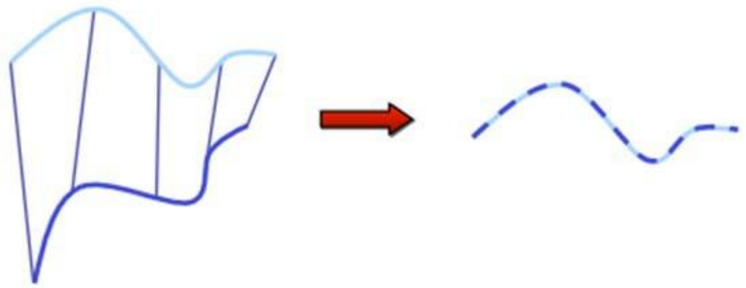
A sample map built from known poses along the red trajectory

8

Pose Estimation

- Find the sensor pose according to its scan and a map
- It might be solved by the ICP (iterative closest point) algorithm
  - A widely used algorithm to register two sets of points

- ICP iterates the following two steps till converge
  - Find correspondence as nearest neighbors
  - Solve sensor motion from the found correspondences



Registration with Known Correspondence

- Given two sets of corresponding points

$$X = \{x_1, x_2, \dots, x_n\} \text{ and } P = \{p_1, p_2, \dots, p_n\}$$

- Want: translation t and rotation R that register these two sets
- Mathematically, that means to minimize the following error

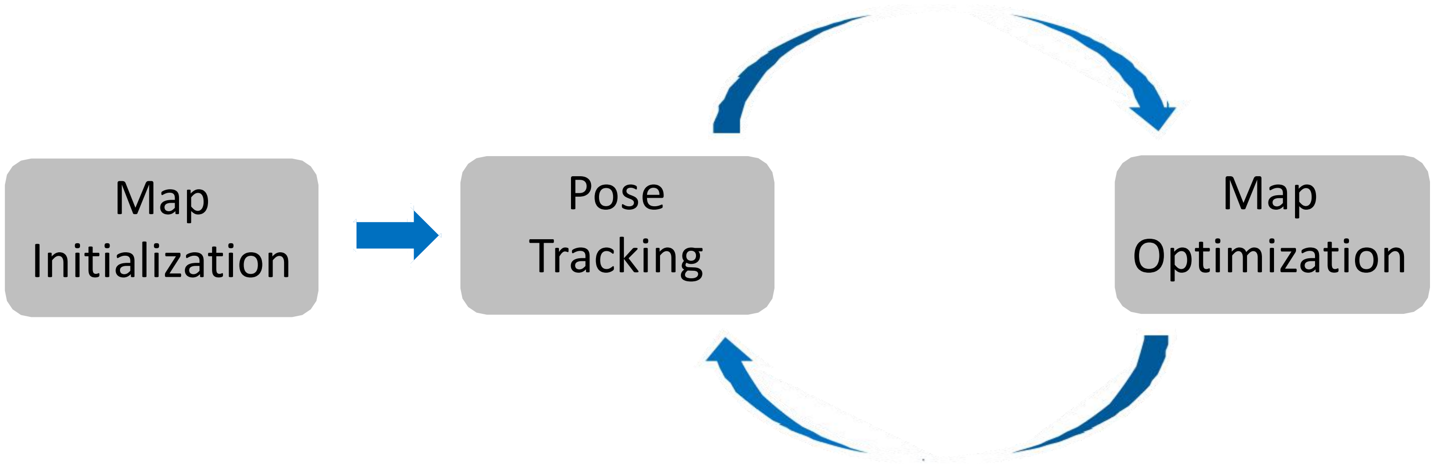
$$E(R,t) = \sum_{i=1}^{n} |x_i - Rp_i - t|^2$$

- A closed-form solution can be derived easily (try it yourself)

2D LiDAR SLAM Summary

- Initialize at t=0
  - The raw LiDAR scan is the initial map
- Start from t=1, iterate the following to steps:
  - Solve sensor pose at time t by the ICP algorithm
  - Update map according to the new scan at time t

Typical SLAM Systems Architecture



- LiDAR SLAM
  - **Initialization**: the first scan
  - **Pose Tracking**: ICP
  - **Map Optimization**: occupancy grid map
- Visual SLAM
  - **Initialization**: (essential matrix, triangulation, etc)
  - **Pose Tracking**:
    1. Feature tracking
    2. Pose-only BA
  - **Map Optimization**:
    1. Triangulation, BA
    2. Loop closure, pose-graph

Questions?

目录
CONTENTS

02 Visual SLAM

Visual SLAM by SfM

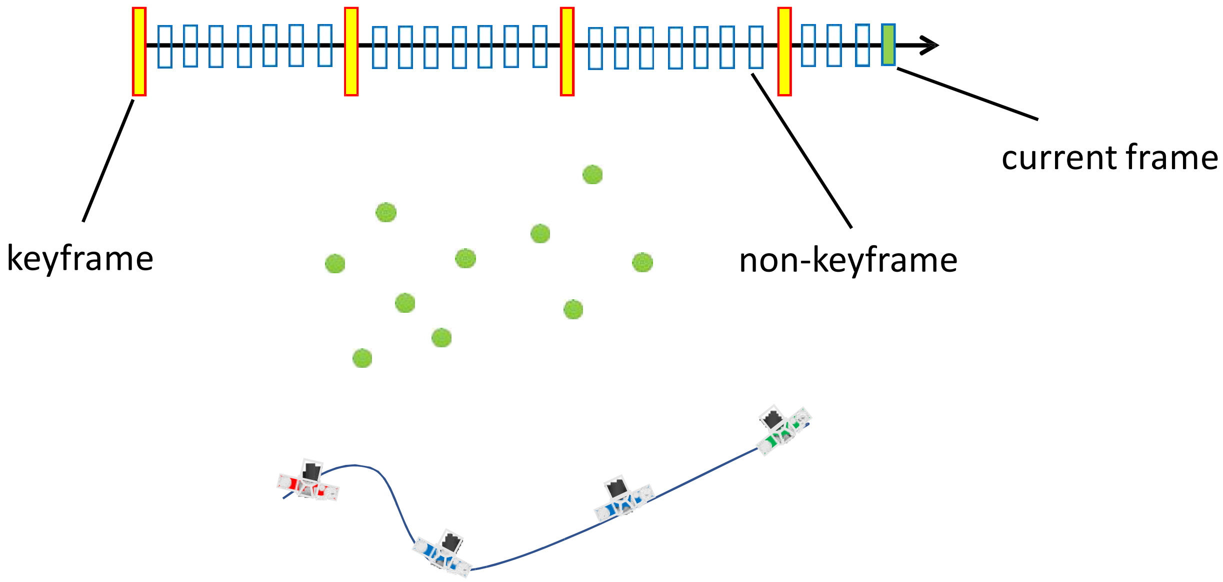
- Solve an incremental SfM at every new frame

- Realtime constraint (many trade-offs for better efficiency)
  - Keyframe based mapping (only a subset of frames are used for mapping)
  - Local BA (bundle adjustment with only nearby video frames)

- Sequential video input
  - Sorted input images (match each image to its previous frame)
  - Regular time interval between frames (motion model to facilitate matching)

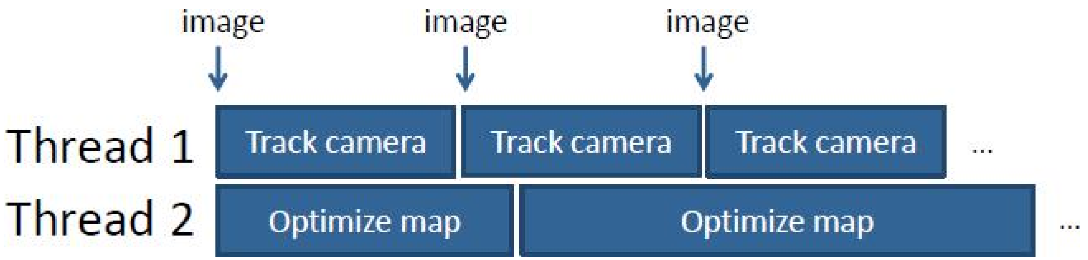
Key-frame based Visual SLAM

- Solving camera pose for every input frame
- Only use some “keyframes” to triangulate/optimize map points



Parallel Tracking and Mapping (PTAM)

- Parallel tracking and mapping
  - A real time tracking thread runs in real-time (30Hz)
  - An offline mapping thread for map maintenance



Parallel Tracking and Mapping for Small AR Workspaces

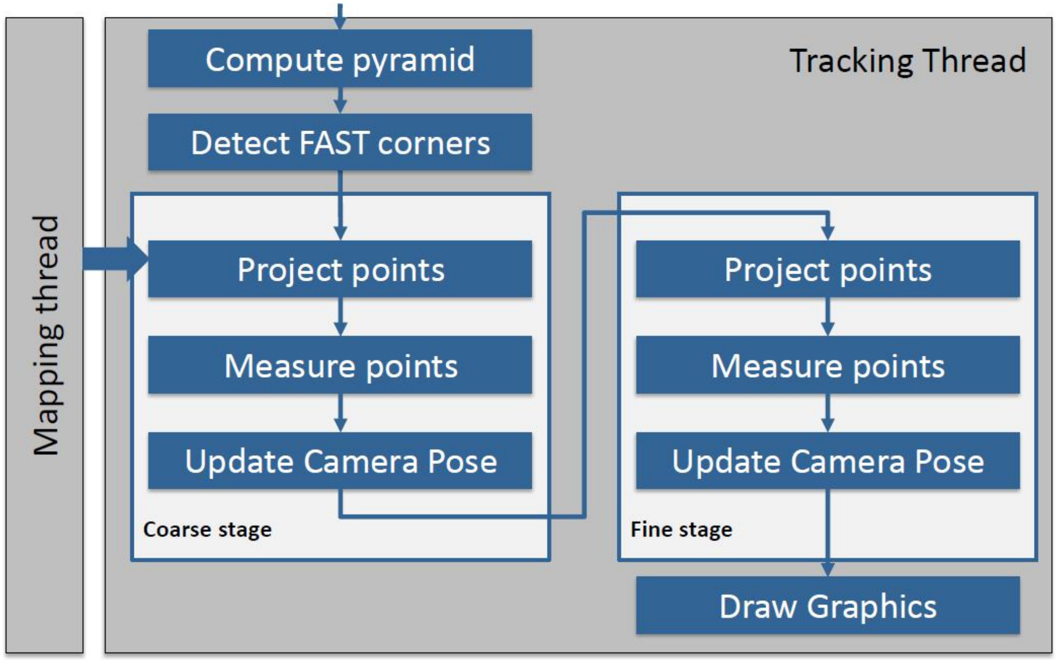
Georg Klein* David Murray†

Active Vision Laboratory
Department of Engineering Science
University of Oxford

[ISMAR 2007]

Tracking

- Step 1: feature correspondence
  - Option 1: KLT feature tracking
  - Option 2: feature detection & matching (within a nearby neighborhood)



[PTAM 2007]

feature correspondence

- Generate 8x8 matching template (warped from keyframe)
- Search for correspondence in a fixed radius around projected position
  - Using SUM OF SQUARE DISTANCE(SSD)
  - Only search at pre-detected corner points (e.g. FAST points)

Tracking

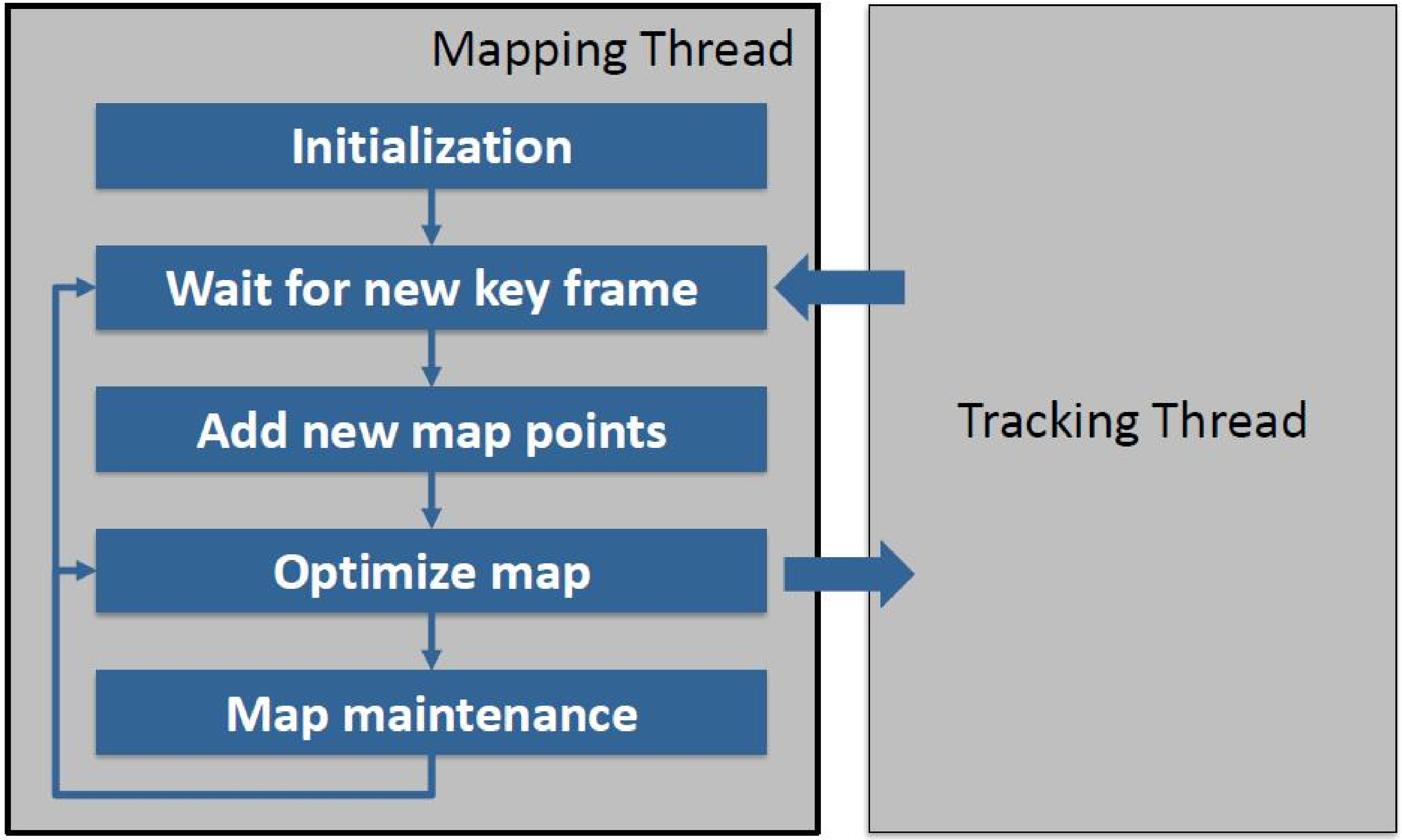
- Step 2: solve camera motion
  - With camera pose initialized by extrapolation, e.g. constant velocity motion
  - With 3D map points fixed
  - Perform local camera only BA (BA with map points fixed)(camera only BA)

  - Typically, use a robust cost function ρ on the re-projection error
  - Camera might also be initialized by PnP

Mapping

- Triggered by the insertion of a new keyframe
- Triangulate any tracked image corners that are not reconstructed
- Run corner detection (e.g. Harris) to generate more points for tracking
- Call local BA to optimize both points and poses
  - BA is slow, may be called after several keyframe insertions

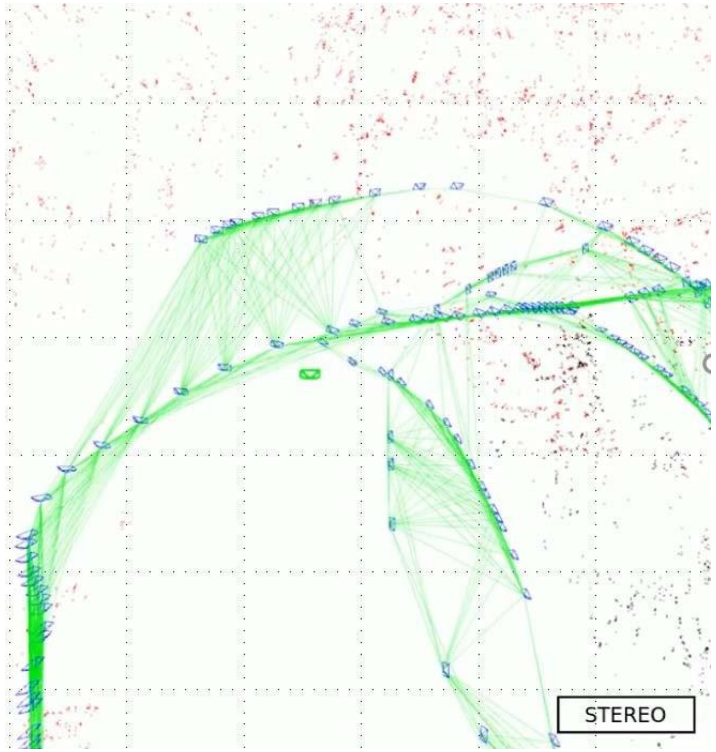
Mapping



[PTAM 2007]

Keyframes

- Defined heuristically, various from system to system
- A keyframe is typically inserted when:
  - Camera tracking is robust
  - There are insufficient corner points to track
  - There is a large camera motion (rotation or translation
  - There is no keyframe inserted for a while (e.g. more than 30 frames)
  - Etc..



PTAM – Example Timings

- Tracking thread

Total	19.2 ms
Key frame preparation	2.2 ms
Feature Projection	3.5 ms
Patch search	9.8 ms
Iterative pose update	3.7 ms

- Mapping thread

Key frames	2-49	50-99	100-149
Local Bundle Adjustment	170 ms	270 ms	440 ms
Global Bundle Adjustment	380 ms	1.7 s	6.9 s

Questions?

目录
CONTENTS

Re-localization

- Tracking can lose due to various reasons
  - Motion blurs
  - Moving objects
  - Large occlusion
  - Sudden fast motion
  - Sudden illumination change

- Re-localization is to recover from such a sudden tracking failure

Re-localization

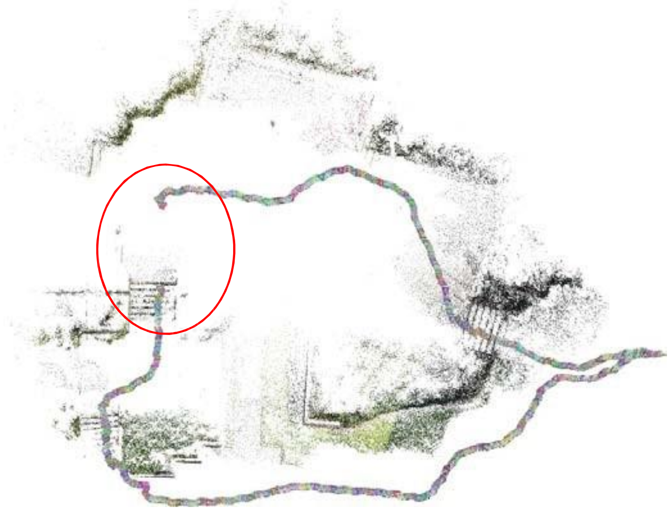
- Re-localization typically includes the following steps
  - Image search: search the current frame among the pre-indexed keyframes
    - E.g. by bag-of-words models
    - It returns a keyframe with sufficient view overlap with the current frame
  - Feature matching between these two frames
  - PnP and local BA to register the current frame to the map
  - Continue the original SLAM

Localization

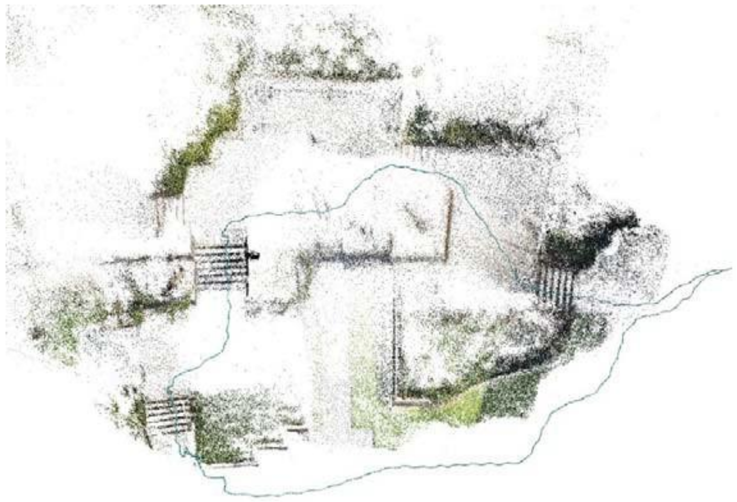
- The re-localization technique can be used in a different scenario:
  - Solve the mapping beforehand (offline)
  - Solve only the tracking online
    - E.g. solve every frame by re-localization
    - Frame-to-frame constraint might also be included
- The advantages of separate mapping and tracking
  - A high quality map is guaranteed
  - The most time consuming step (i.e. global BA) is moved to offline

Drifting

- Small error accumulates to large map distortions
- The technical to reduce drifting is called loop closure



result with drifting error

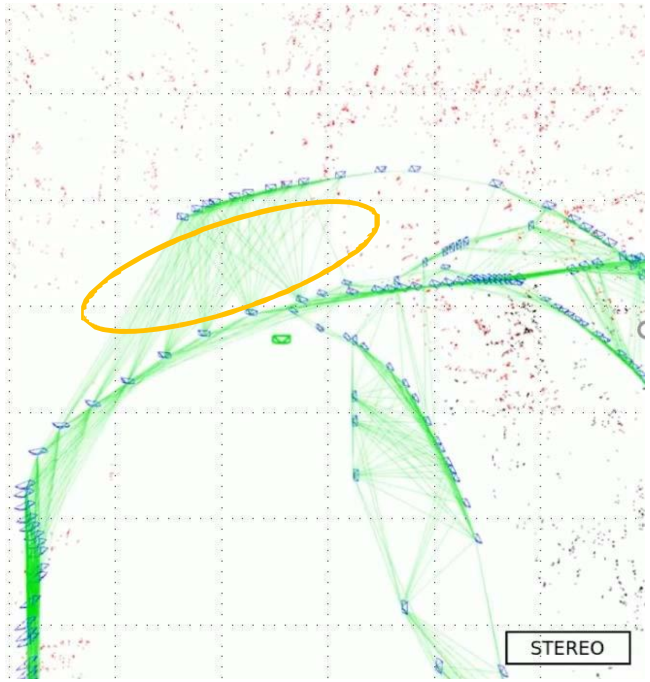


result after loop closure

Loop Closure

- **Loop detection**
  - Identify if a loop exist
  - Again, by the same image search technical in re-localization
  - Typically, applied to every newly inserted keyframe

- **Construct a pose-graph for the next step**
  - Where a keyframe is a vertex
  - Two keyframes are connected if they have view overlap



connection due to loop detection

Loop Closure

- Loop optimization
  - In the simplest case, a direct BA can generate good result
    - In many cases, the drifting error is too large to be corrected by BA

- Most of the time, a global SfM is desirable
  - Solving all keyframe poses from input pairwise relative motion constraints
  - Update map points afterwards
  - Referred as ‘pose-graph optimization’ in robotics
    - the keyframe graph is a graph with camera poses (no 3D points)
  - The most challenging problem is to deal with wrong loops (due to repetitive structures)
    - Studied in both computer vision and robotics community

Questions?

A Brief History of Visual SLAM

- MonoSLAM, Andrew Davison [ICCV 2003] [PAMI 2007]
  - The first work of visual SLAM with a single camera
- Visual Odometry, David Nister [CVPR 2004]
  - Visual slam by SfM
- PTAM, Klein & Murray [ISMAR 2007] (open source)
  - Separating tracking and mapping
- LSD-SLAM, Engel et al. [ECCV 2014] (open source)
  - Direct method
- ORB-SLAM, Mur-Artal et al. [PAMI 2015] (open source)
  - A stronger version than PTAM, with re-localization, pose-graph, etc
- DSO, Engel et al. [PAMI 2017] (open source)
  - A stronger version of LSD-SLAM, with photometric auto-calibration, etc

视觉与激光SLAM相关项目

1	基于室内高精度三维测图的BIM关键部件自动化重建方法(KF-2019-04-010), 自然资源部城市自然资源监测与仿真重点实验室开放基金, 负责人-汤圣君
2	基于多 RGB-D 数据的 FCN 场景语义分类与 BIM 模型自动化三维重建技术(000002110335), 深圳大学高水平大学建设2期, 负责人-汤圣君
3	多RGB-D传感器集成的在线室内高精度三维测图方法(41801392), 国家自然科学基金委, 负责人-汤圣君
4	基于便携式深度传感器的城市封闭/半封闭空间快速三维测图技术研究(JCYJ20180305125131482), 深圳市科技创新委员会, 负责人-汤圣君
5	多元特征混合优化的RGB-D室内高精度三维测图方法(2018M633133), 中国博士后科学基金, 负责人-汤圣君
6	集成视觉与几何特征的深度传感器高精度SLAM方法(17E04), 武汉大学测绘遥感信息工程国家重点实验室开放基金, 负责人-汤圣君
7	联合视觉SLAM与深度神经网络的室内场景语义分类与自动建模技术，广东省面上基金，负责人汤圣君
8	室内全空间三维测图与标准化建模技术研究，促进高校科技成果服务产业发展扶持项目，负责人汤圣君

产业化获奖情况

- 创新南山2020“创业之星”大赛初创组一等奖
- 创新南山2020“创业之星”大赛总决赛优胜奖
- 2020深创赛互联网与移动互联网总决赛优胜奖
- 广东省高校科技成果转化大赛电子信息组二等奖

获奖20万+80万政府创业资助

凤凰新闻　立即打开

乘新基建东风 三维数字化技术助力城市智慧发展

凤凰网广东综合
11/2/ 14:36

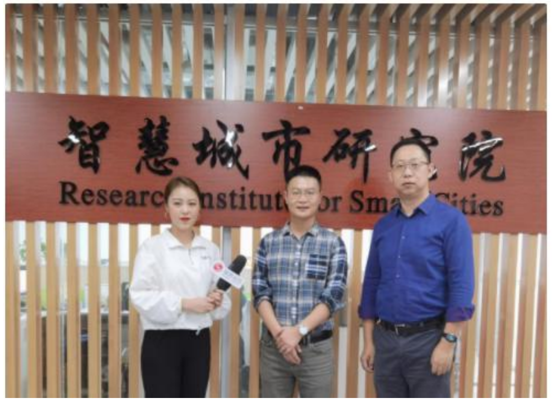
人类自定居以来始终处于城市化的过程，《大国大城》中提到“大国需要大城”，大城市群主导着世界的经济，如何管理好城市也就成为21世纪最重要的发展挑战之一。而作为伴随物联网、云计算等信息技术发展应运而生的全新城市建设模式——智慧城市，面临城市发展问题提出的新发展思路，正成为世界城市发展新的制高点。

随着我国城市化进程的不断加快，对城市规划建设管理工作也提出了更高的要求，仅仅基于二维GIS技术的城市规划与分析，已无法满足城市规划与监督管理中多维动态空间分析的需求，必须利用三维GIS技术，对整个城市的三维立体空间进行统一描述，并充分准确地集成表达地下的地质、管线、构筑物，地上的土地、交通、建筑、植被,以及室内的设施、房产、人口等，形成与现实世界一致的三维立体空间框架。

产品化落地，瞄准智慧城市室内空间建模

深圳大学智慧城市研究院副院长贺彪认为，城市有三个空间，即物理空间、社会空间、信息空间。物理空间是指人们生活的空间。社会空间是指政府管理和人类社会经济活动的人文社会空间。信息空间是指信息化带来的网络虚拟空间。按照三元空间理论，捕获三个空间相应实体、相应活动、相应规律产生的数据，将其归纳为城市基础的时空数据、城市管理对象数据和城市运行状态的感知数据，将这些数据捕捉之后，可以生成数字化城市，而深圳大学智慧城市研究院正致力于为用户提供GIS服务，赋能传统建造行业实现数字化、智能化、可视化的产业升级，进一步推动智慧城市的发展。

深圳大学智慧城市研究院汤圣君助理教授表示，从城市的角度来讲，外部三维空间的建模手段丰富且进入成熟期，而室内建模精细化还需要进一步的发展，室内空间具有场景分散、空间小的特点，智慧城市研究院团队针对这一痛点问题，系统性研发了一套低成本的室内三维全自动数字化软硬件产品“Gho3D三维扫描与VR可视化系统”。



凤凰網廣東 gd.ifeng.com

www.news.cn 新华网 NEWS 广东频道 gd.xinhuanet.com

南方都市报

深圳特區報

深圳晚报

谢谢