

N L P R

# 面向无人系统的大规模场景三维感知

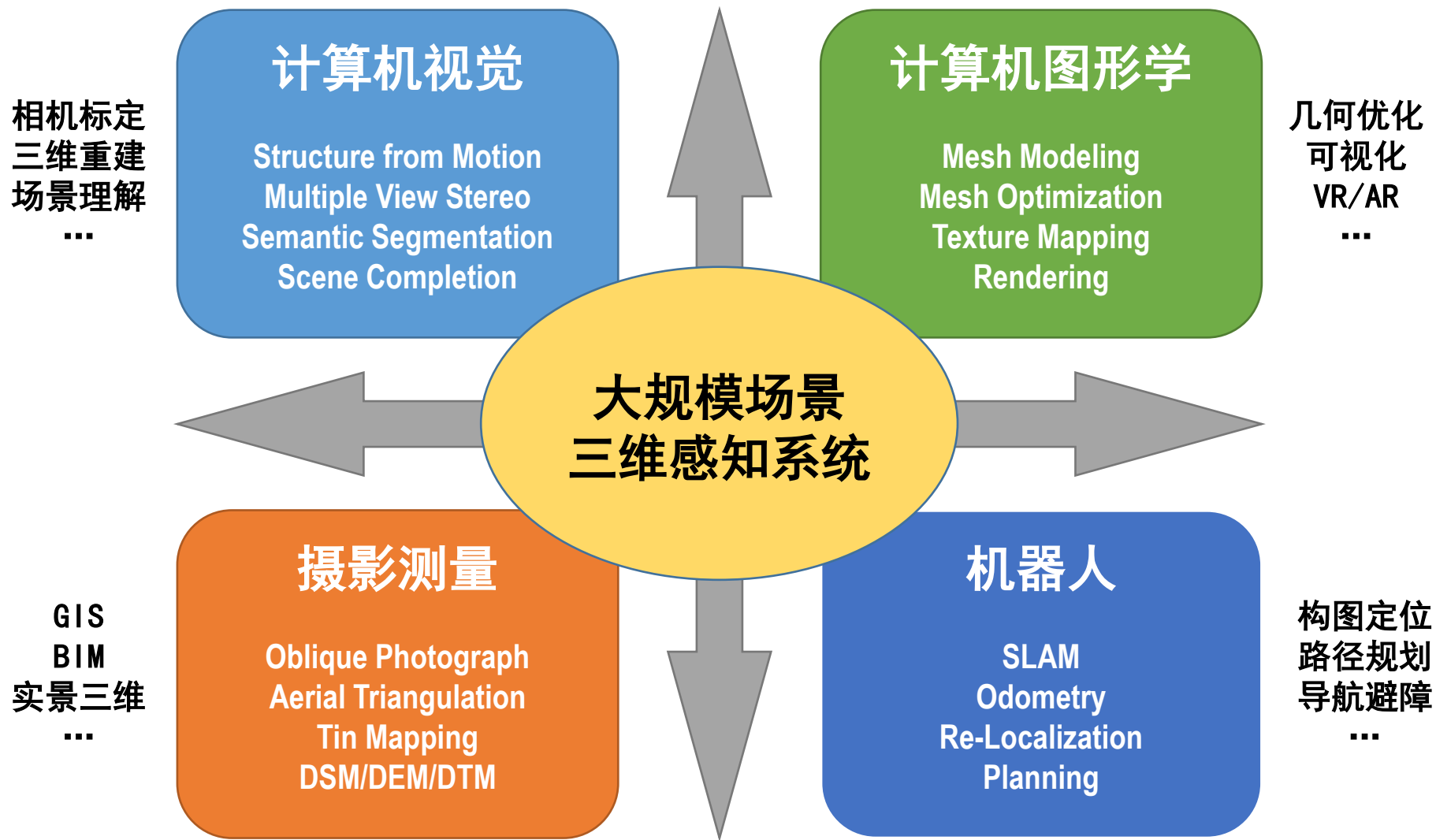
申抒含

中国科学院自动化研究所  
模式识别国家重点实验室

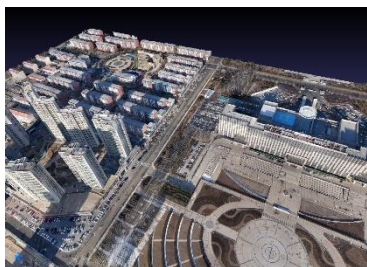
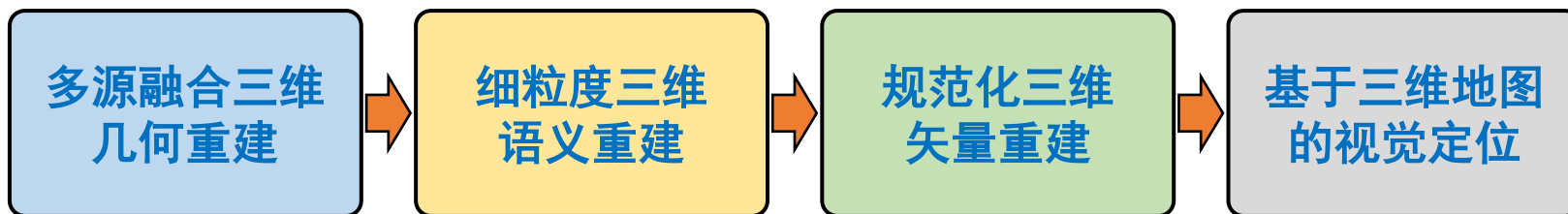
<http://vision.ia.ac.cn/Faculty/shshen/index.htm>



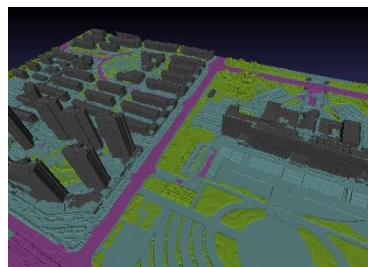
# 大规模场景三维感知系统



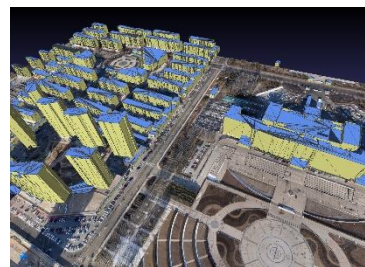
# 大规模场景三维感知系统



3D几何模型



3D语义模型

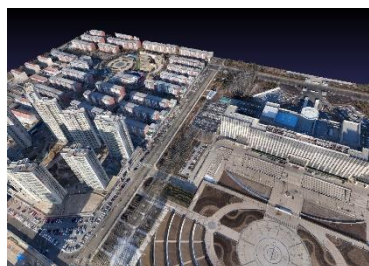
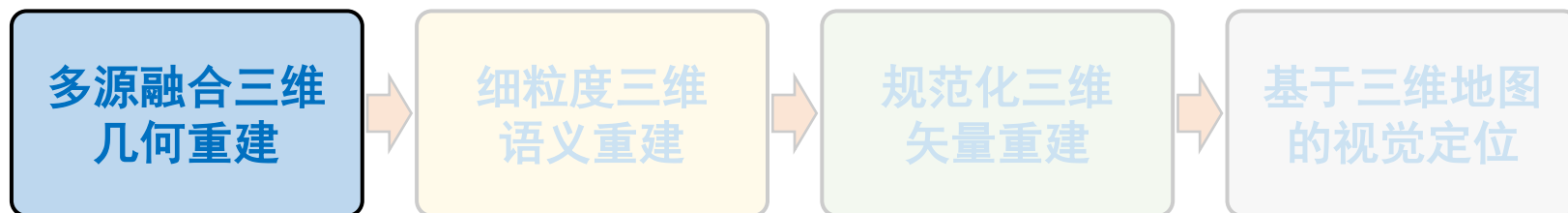


3D矢量模型

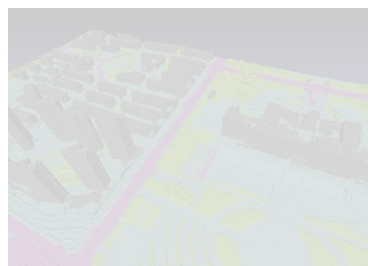


6DoF视觉定位

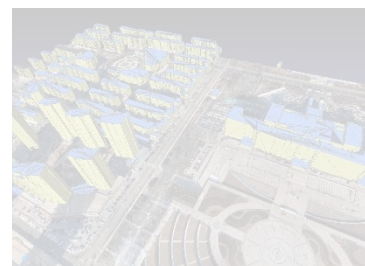
# 多源融合三维几何重建



3D几何模型



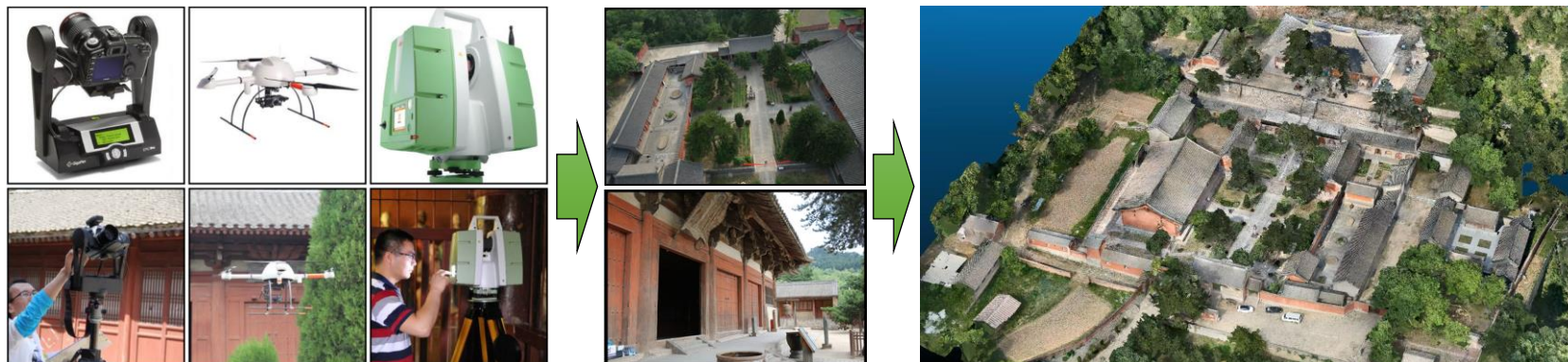
3D语义模型



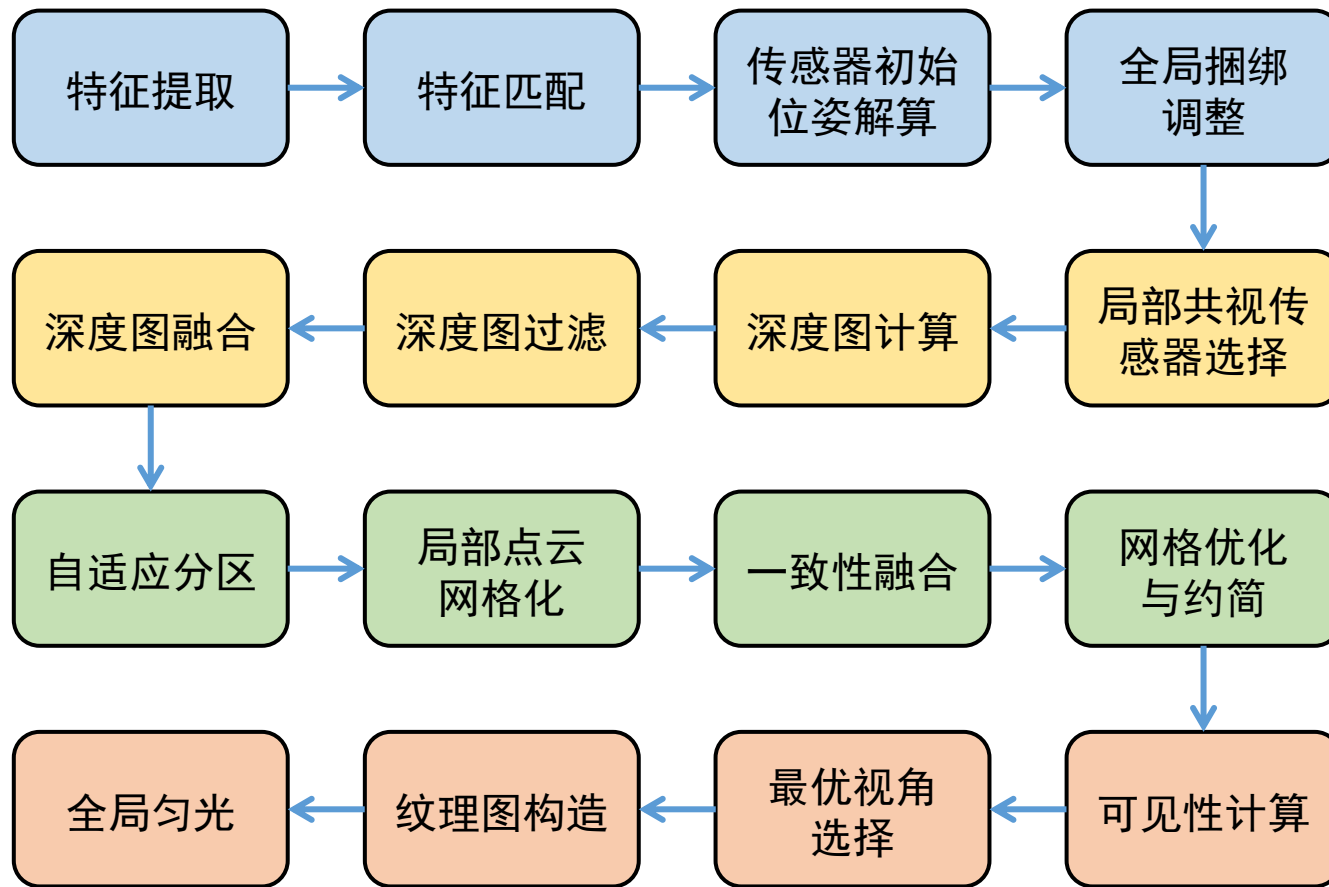
3D矢量模型



6DoF视觉定位



# 多源融合三维几何重建

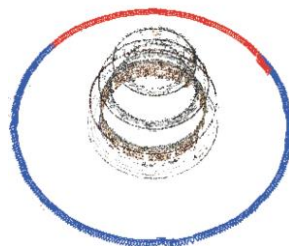


Pipeline of 3D Geometry Reconstruction

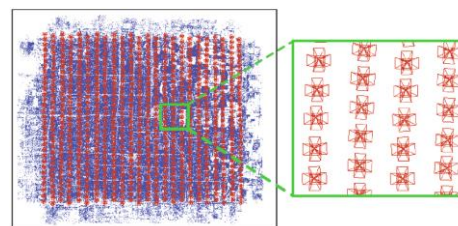
# 多源融合三维几何重建



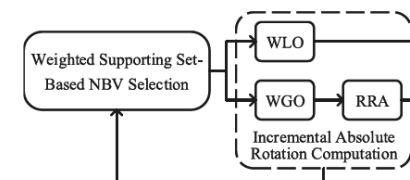
**Hybrid SfM**  
CVPR 2017  
混合稀疏重建



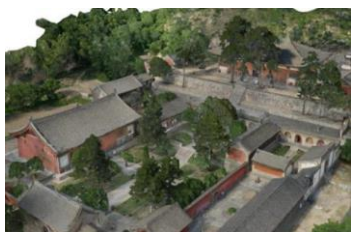
**Progressive SfM**  
3DV 2018  
增量稀疏重建



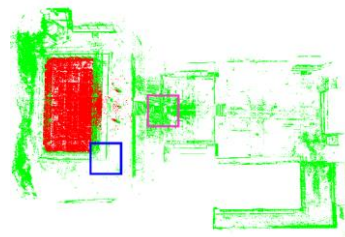
**Tracks Selection in SfM**  
ISPRS JPRS 2019  
增量稀疏重建



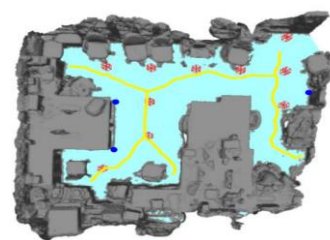
**Incremental RA**  
IJCV 2021  
全局稀疏重建



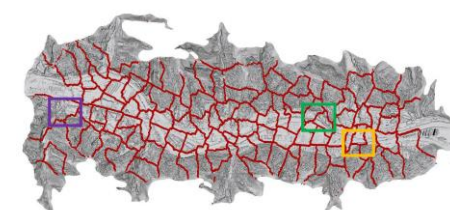
**Aerial and Ground Fusion**  
ISPRS JPRS 2018  
天地图像融合重建



**Image and Laser Fusion**  
IEEE T-CSVT 2020  
图像-激光融合重建

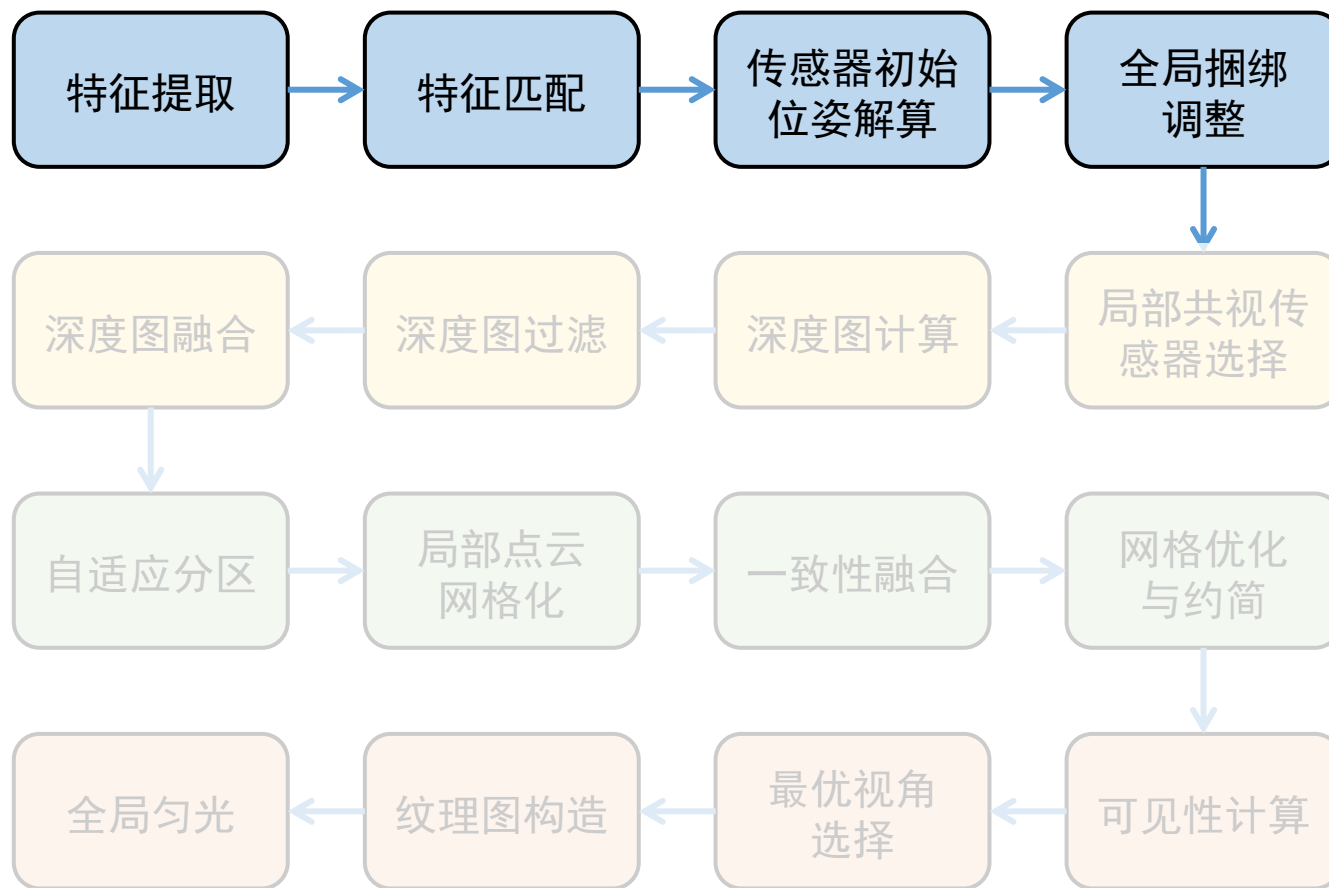


**Mini Drone and Robot Fusion**  
IEEE Sensor Journal 2021  
无人机-机器人融合重建



**Large Meshing**  
3DV 2019  
分布式点云网格化

# Structure-from-Motion



**Pipeline of 3D Geometry Reconstruction**







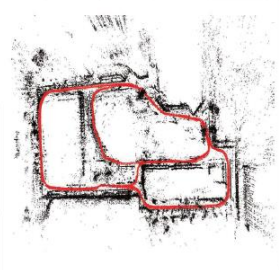
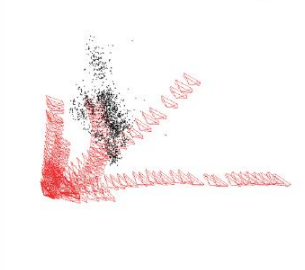
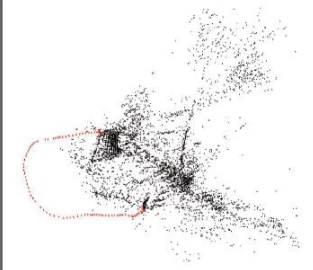
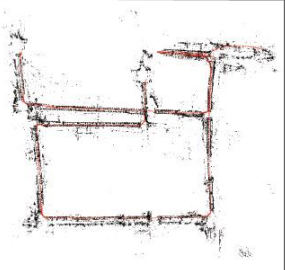

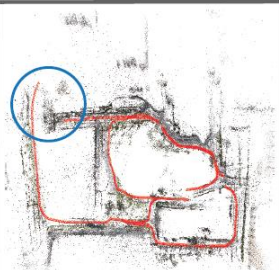

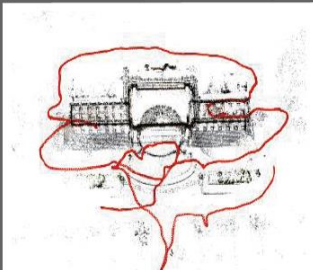
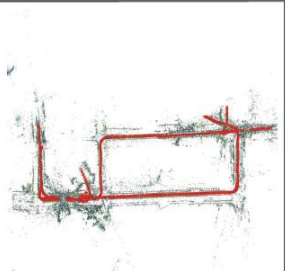
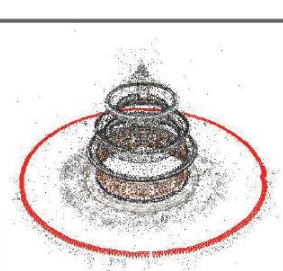
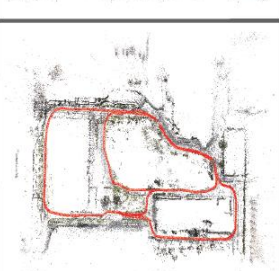
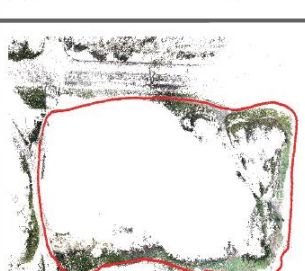
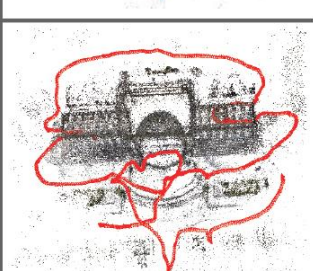
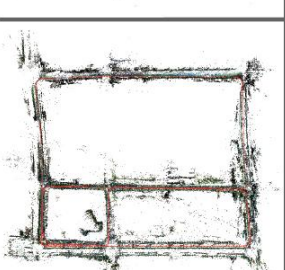
# Incremental Structure-from-Motion

	data00	data02	data05	data08	data09
ORB SLAM3					
COL MAP					
VidSfM					

SLAM v.s. SfM (easy)



# Incremental Structure-from-Motion

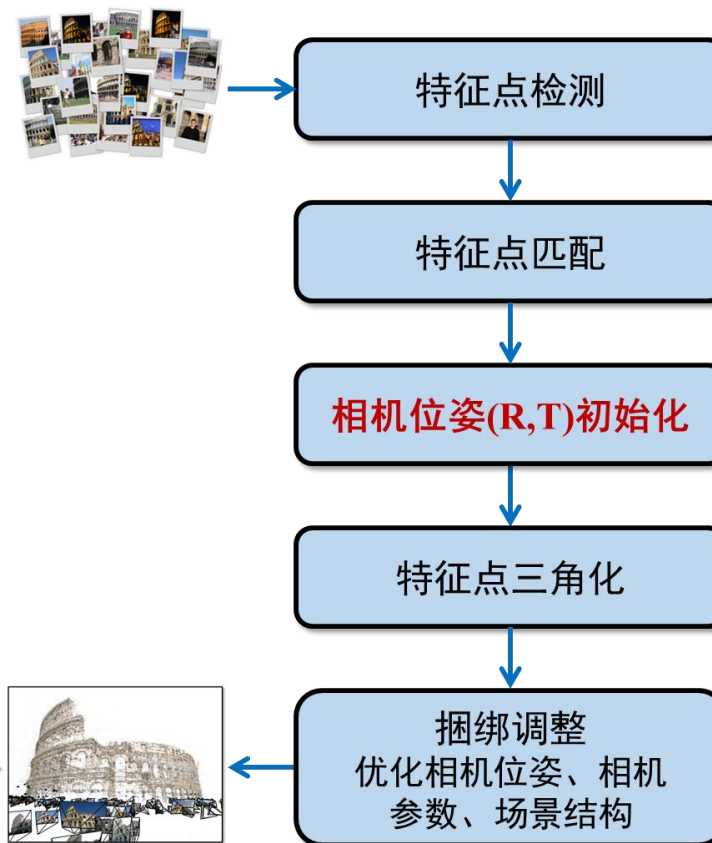
	Temple-of-Heaven 	Institute 	Campus 	Temple 	StreetView 
ORB SLAM3					
COL MAP					
VidSfM					

SLAM v.s. SfM (difficult)

# Incremental Structure-from-Motion

## Incremental SfM误差消除策略:

- Camera seeds selection
- Tracks selection
- Next best view selection
- Prioritized camera registration
- Global bundle adjustment

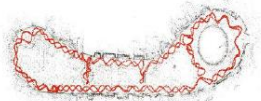
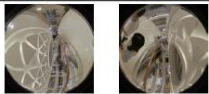


## Incremental SfM主要瓶颈:

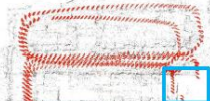
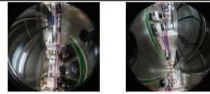
- 误差累积可减弱但不能消除
- 反复调用BA耗时

# Incremental Structure-from-Motion

Data		COLMAP [41]			BATA [56]			MIS		
Name	N	$R_e$	$C_e$	$T$	$R_e$	$C_e$	$T$	$R_e$	$C_e$	$T$
data00	9182	1.7	2.8	9498	1.3	36.8	7.0	<b>0.3</b>	<b>0.5</b>	286
data01	2202	<b>0.2</b>	<b>1.0</b>	649	1.8	29.3	0.3	0.4	<b>1.0</b>	47
data02	9322	1.4	11.4	2415	78.3	259.5	6.0	<b>0.3</b>	<b>1.0</b>	355
data03	1602	0.3	<b>0.2</b>	577	0.7	10.0	0.3	<b>0.2</b>	<b>0.2</b>	77
data04	542	<b>0.1</b>	<b>0.1</b>	58	0.4	87.8	0.1	0.2	<b>0.1</b>	4
data05	5522	3.6	2.9	3764	1.8	19.7	1.7	<b>0.3</b>	<b>0.2</b>	116
data06	2202	0.3	0.7	1238	57.8	65.1	0.3	<b>0.1</b>	<b>0.2</b>	34
data07	2202	<b>0.7</b>	1.2	1284	2.9	8.9	0.3	<b>0.7</b>	<b>0.6</b>	62
data08	8142	3.1	8.0	4032	0.8	24.8	3.5	<b>0.4</b>	<b>1.2</b>	276
data09	3182	0.4	1.7	764	1.4	38.5	0.5	<b>0.3</b>	<b>0.5</b>	74
data10	2402	0.9	1.5	886	1.2	26.0	0.3	<b>0.4</b>	<b>0.4</b>	53



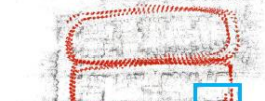
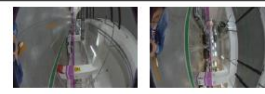
data11



data12



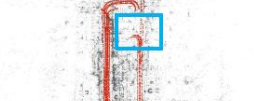
data13



data14



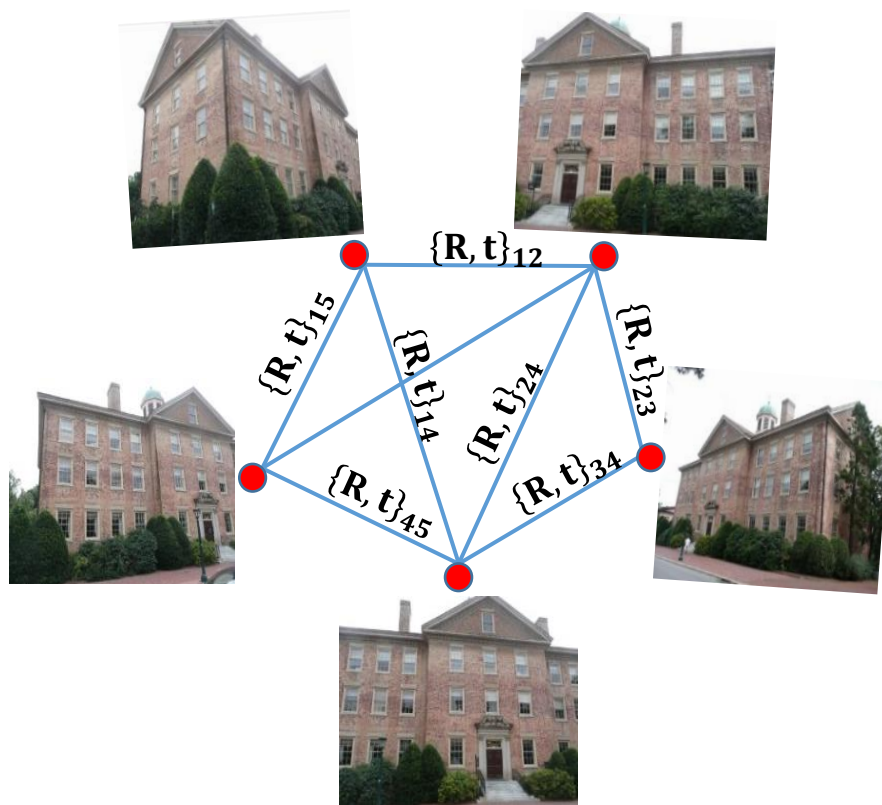
data15



data16

MIS : Multi-camera based Incremental Structure-from-Motion

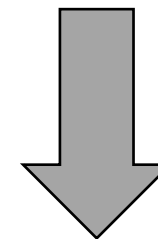
# Global Structure-from-Motion



**View Graph**

$$R_{ij} = R_j R_i^T$$

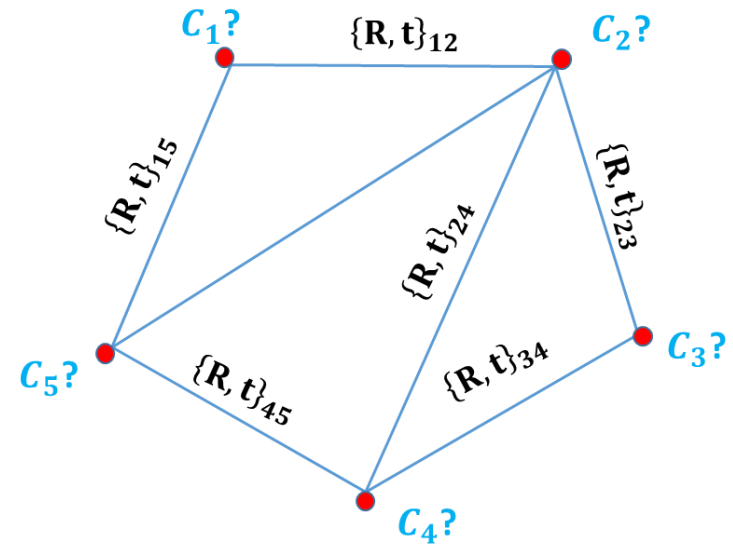
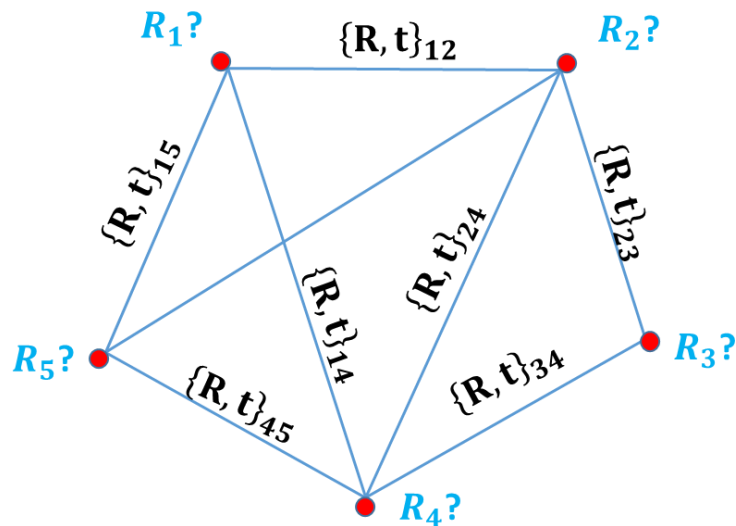
$$\lambda_{ij} t_{ij} = R_j (C_i - C_j)$$



1. 估计所有相机的旋转矩阵
2. 估计所有相机的空间位置
3. 三角化初始场景点
4. Bundle adjustment

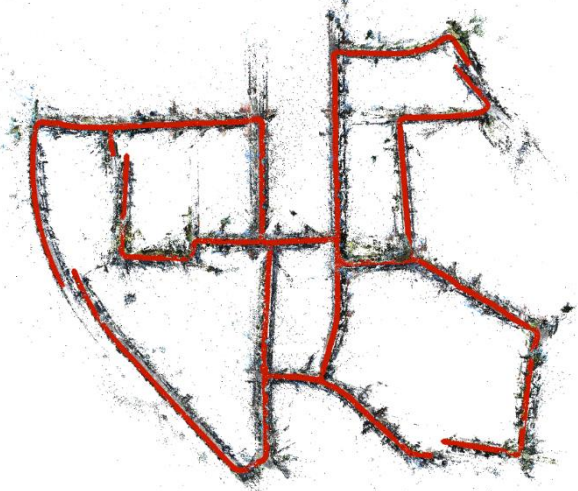
# Global Structure-from-Motion

- Global Translation Averaging



- IRA: Incremental Rotation Averaging (*IJCV 2021*)
- IRA++: Distributed Incremental Rotation Averaging (*IEEE TCSVT minor revision*)
- ITA: Incremental Translation Averaging (*submitted to IJCV*)
- MMA: Multi-camera Motion Averaging (*submitted to AAAI*)

# Global Structure-from-Motion



Colmap-9498mins

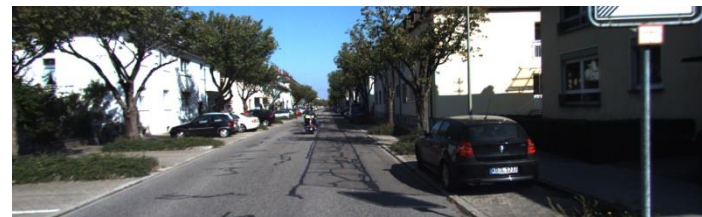


VidSfM-360mins

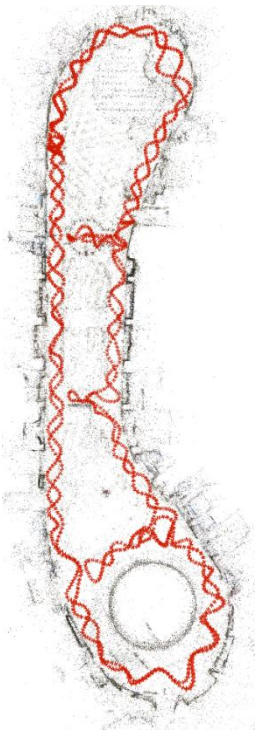


GMSfM-50mins

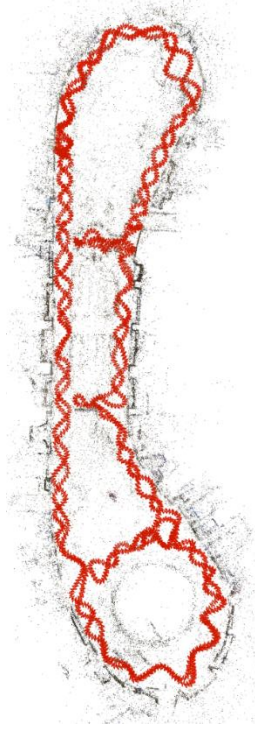
KITTI-00 (4541\*2 images)



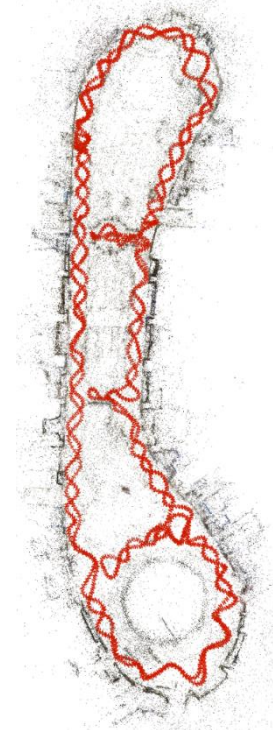
# Global Structure-from-Motion



Colmap-590mins



VidSfM-50mins

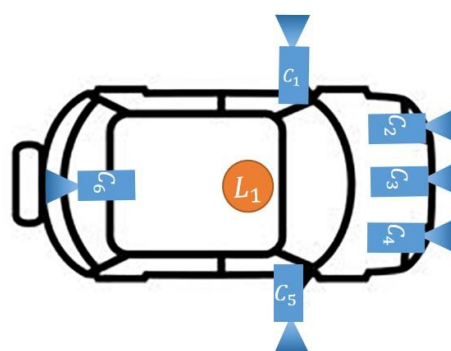
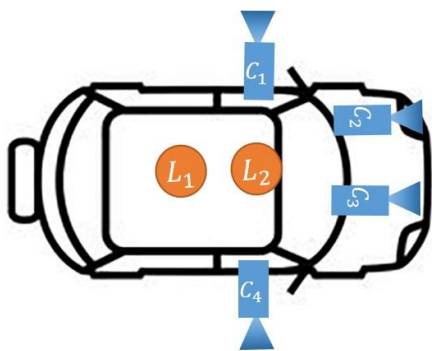
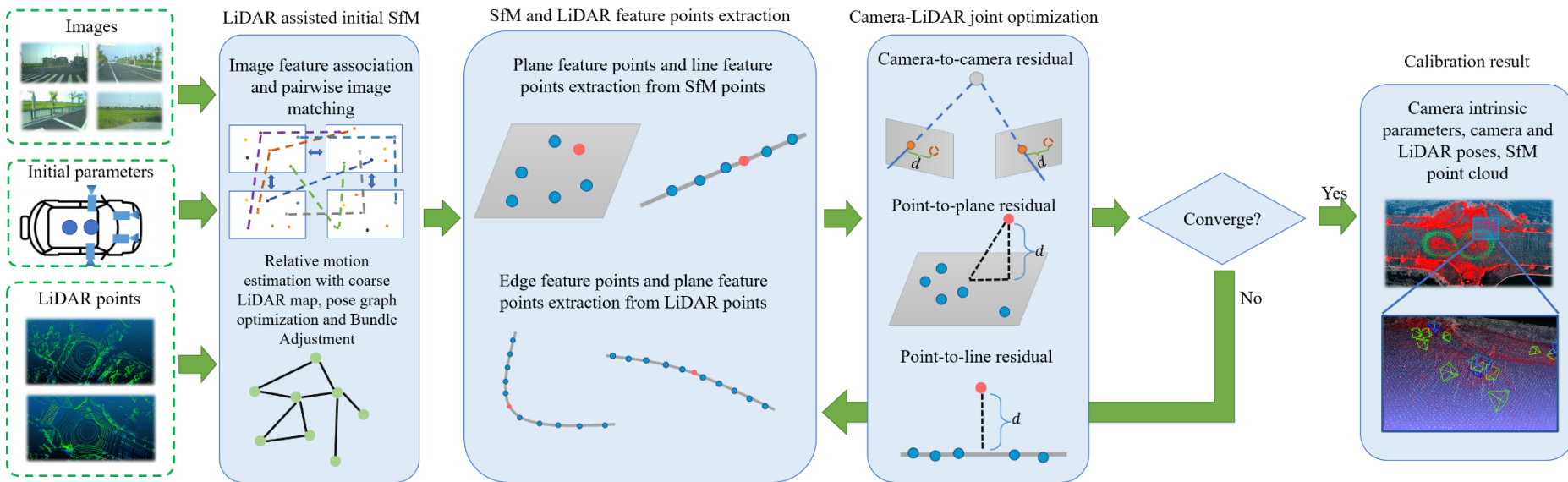


GMSfM-15mins

Store (1276\*2 images)



# Multi-Camera-Multi-LiDAR Auto-Calibration

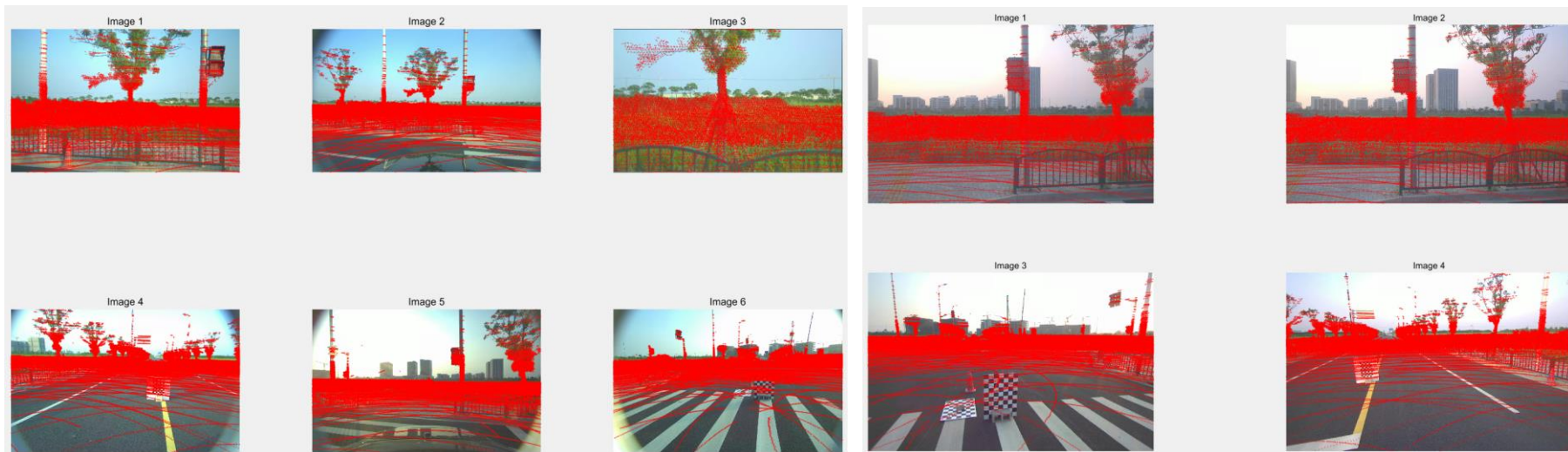
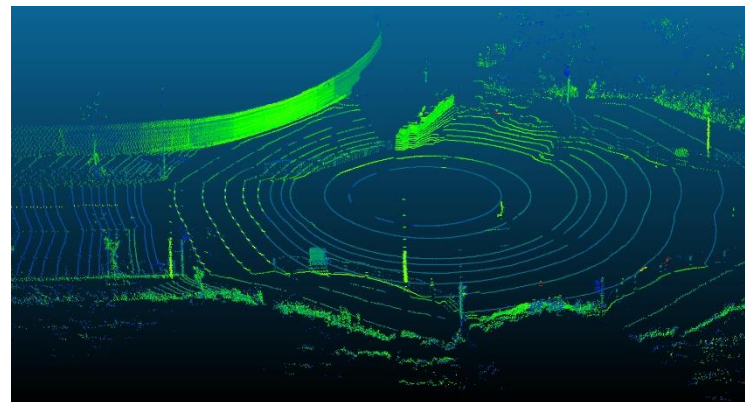


$C$  : Camera

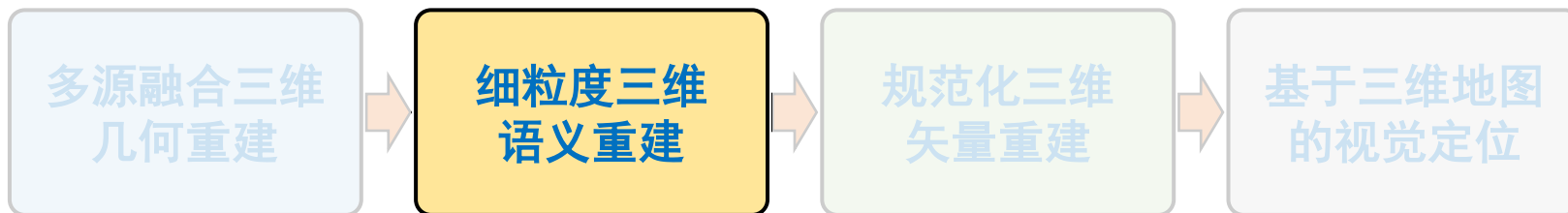
$L$  : LiDAR



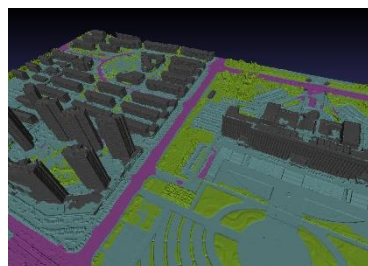
# Multi-Camera-Multi-LiDAR Auto-Calibration



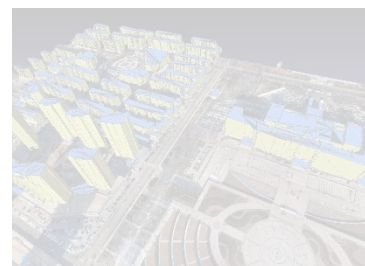
# 细粒度三维语义重建



3D几何模型



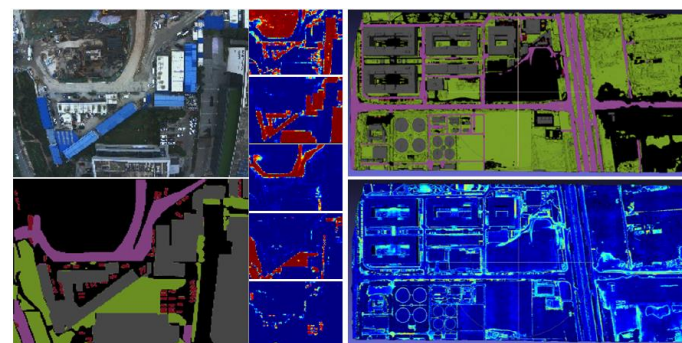
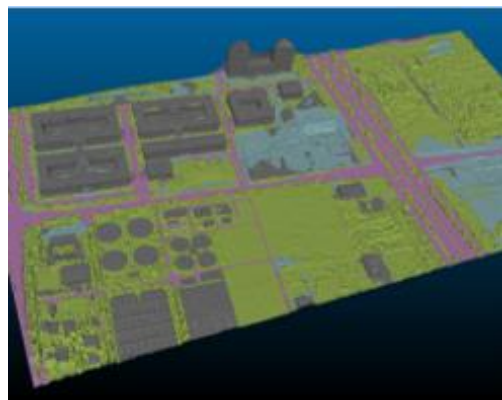
3D语义模型



3D矢量模型

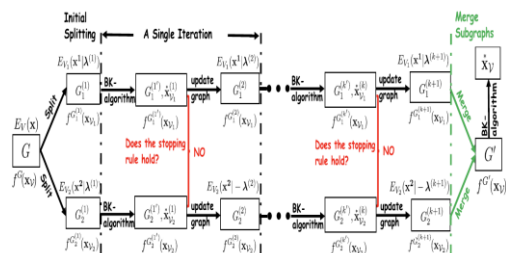


6DoF视觉定位

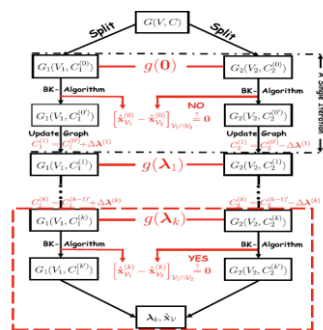


unlabeled road vegetable building car Low confidence High confidence

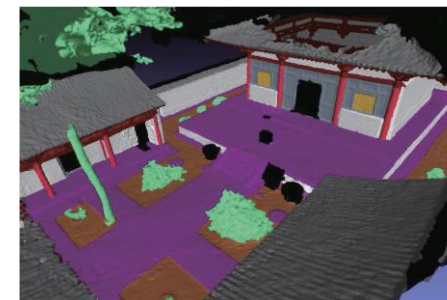
# 细粒度三维语义重建



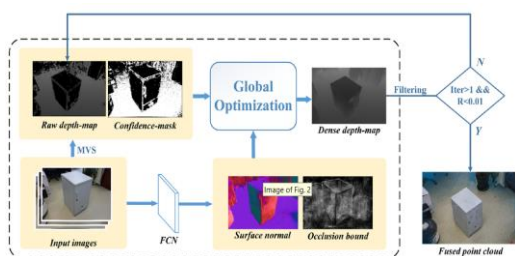
**Distributed Graph Cuts**  
*IEEE TIP 2016*  
 分布式图割优化



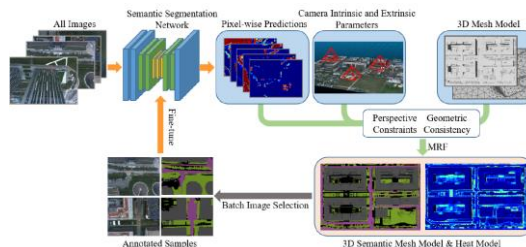
**Parallel Graph Cuts**  
*IEEE TIP 2017*  
 并行图割优化



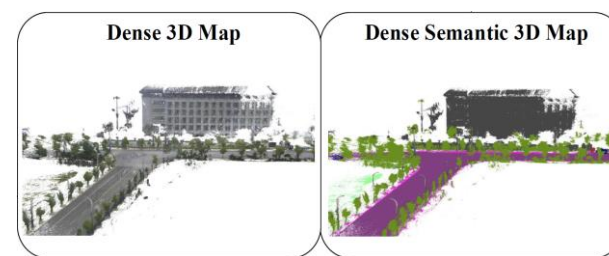
**AL for Fine-Level Scene Parsing**  
*3DV 2018*  
 三维场景细粒度分割



**Semantic Depth Completion**  
*PR 2020*  
 场景语义补全



**AL for Large 3D Scenes**  
*IEEE T-CSVT 2021*  
 主动学习场景分割



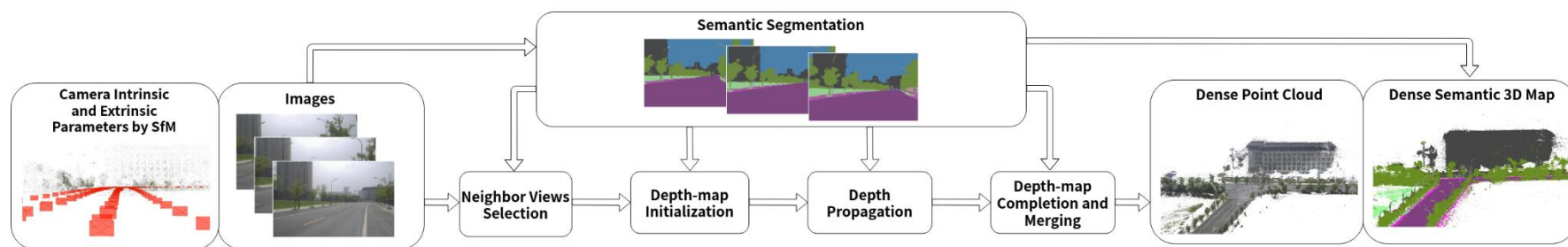
**Semantic Road Mapping**  
*ICRA 2021*  
 道路三维语义地图

# Semantic 3D Mapping

## Semantically Guided Multi-View Stereo for Dense 3D Road Mapping

*Mingzhe Lv, Diantao Tu, Xincheng Tang, Yuqian Liu and Shuhan Shen*

*ICRA 2021*

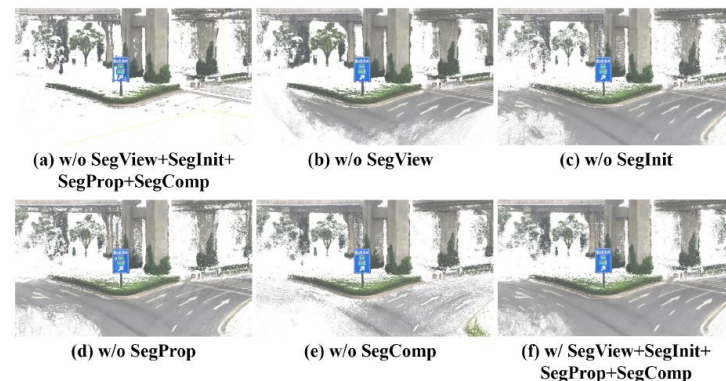


A) Semantically Guided Neighbor Views Selection

B) Semantically Guided Depth Map Initialization

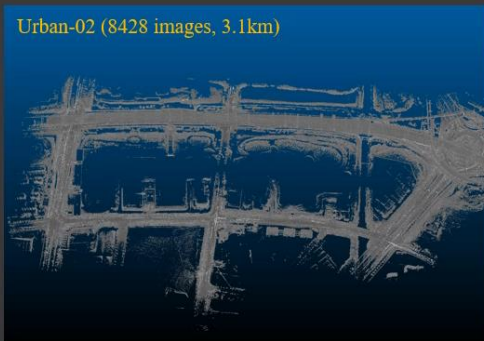
C) Scale-Adaptive Depth Propagation

D) Semantically Guided Depth Completion and Merging



# Semantic 3D Mapping

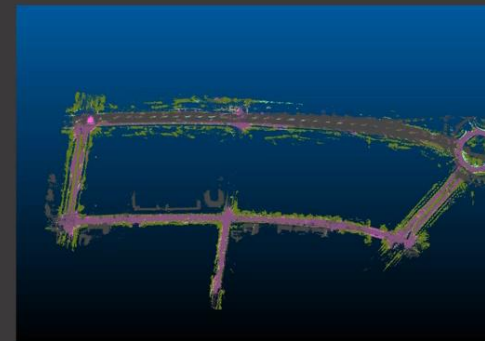
Urban-02 (8428 images, 3.1km)



LiDAR Map



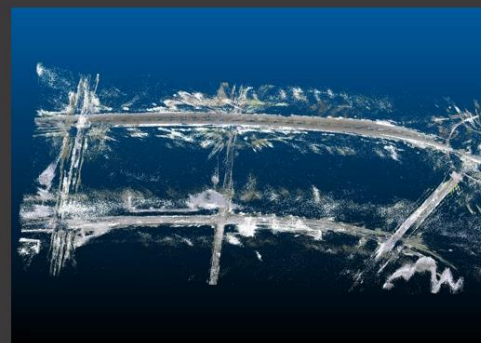
Ours (Dense 3D Map)



Ours (Dense Semantic 3D Map)



OpenMVS [42]



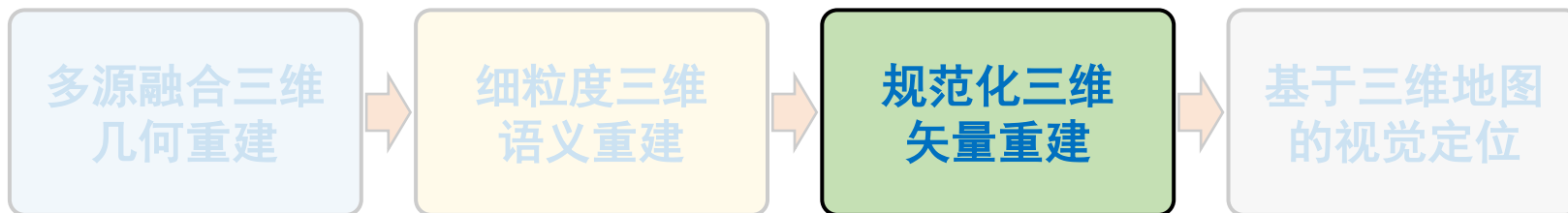
CasMVSNet [32]



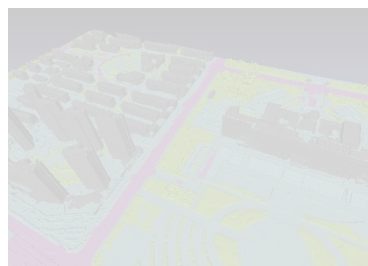
VisMVSNet [33]

杭州萧山区无人驾驶测试路段三维语义地图  
(8428幅前向车载图像, 3.1公里路段)

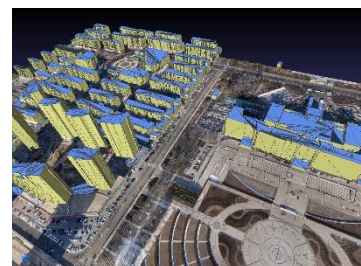
# 规范化三维矢量重建



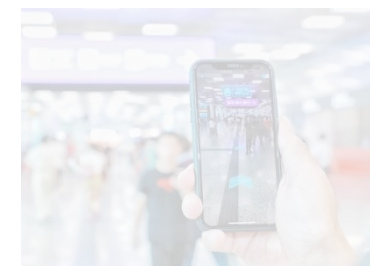
3D几何模型



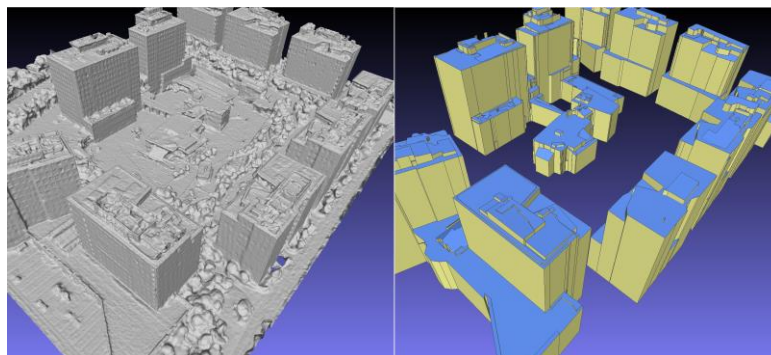
3D语义模型



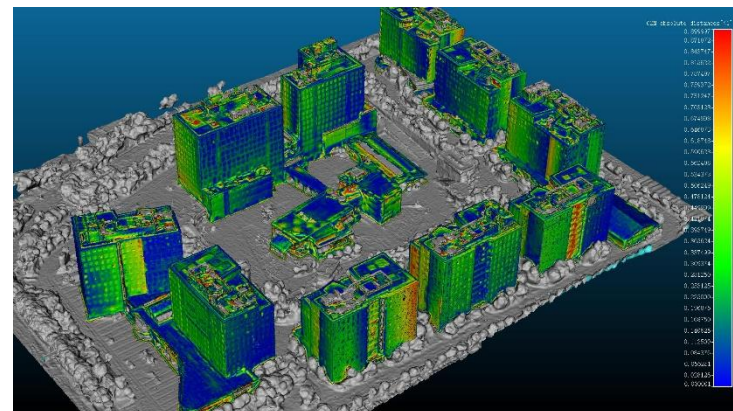
3D矢量模型



6DoF视觉定位



Mesh Model (9M facets)    LOD2 Model (8K facets)

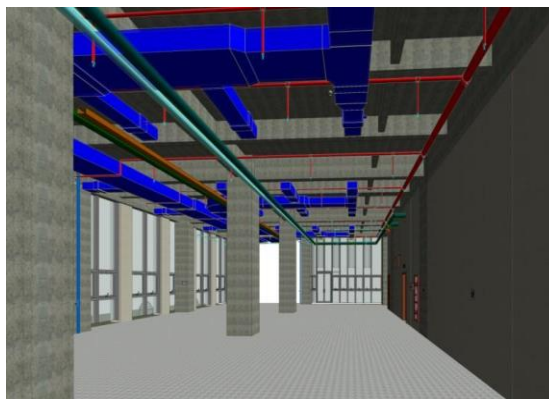


# 规范化三维矢量重建



三维地理信息  
(3D GIS)

地物矢量模型  
(CityGML)



建筑信息建模  
(BIM)

建筑物矢量模型  
(IFC)

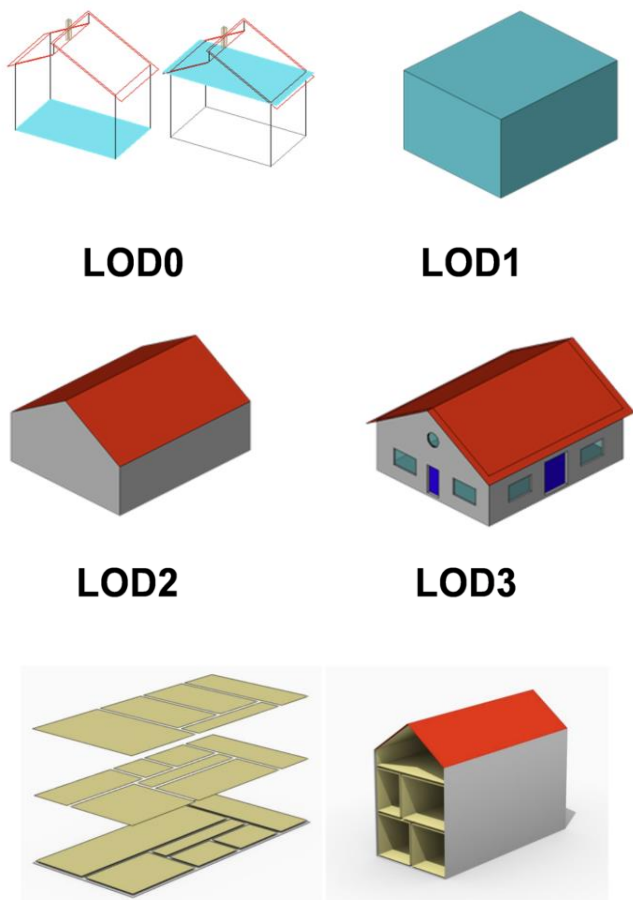


无人系统高精地图  
(HD Map)

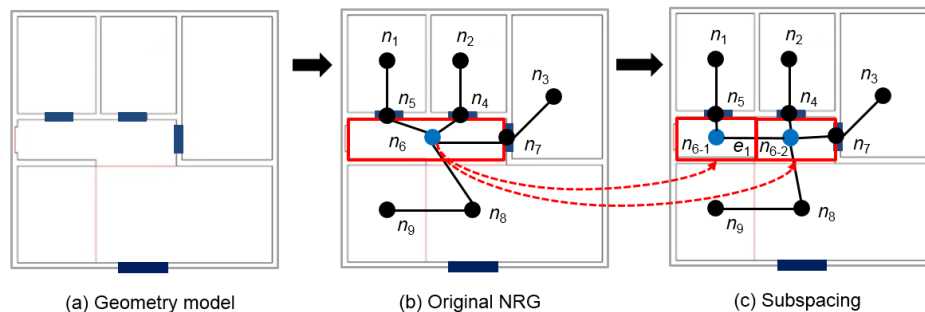
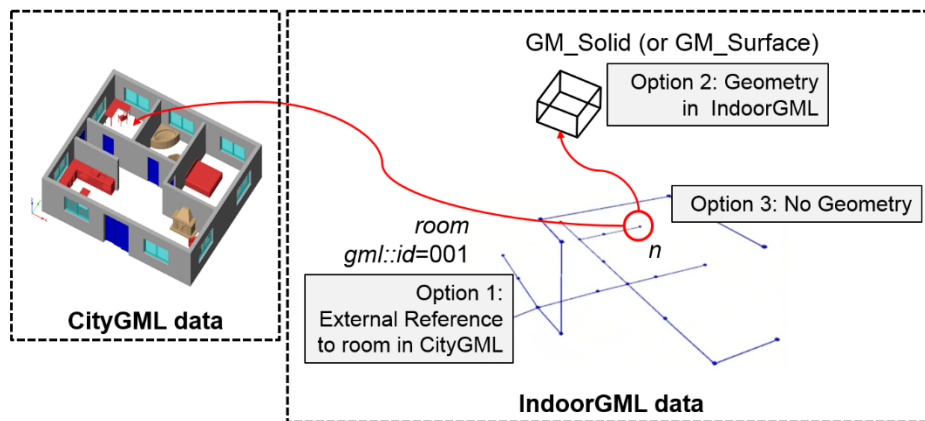
车道线级矢量地图  
(OpenDrive)

不同行业使用的矢量化三维模型规范

# 规范化三维矢量重建



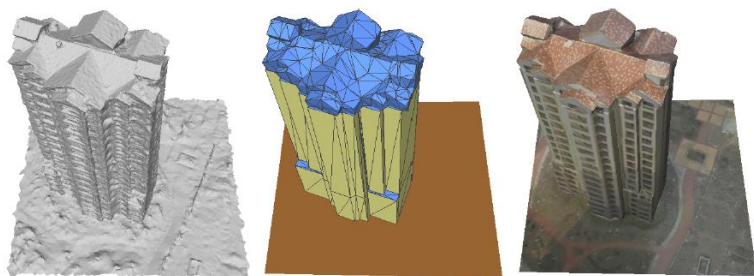
CityGML 3.0



IndoorGML



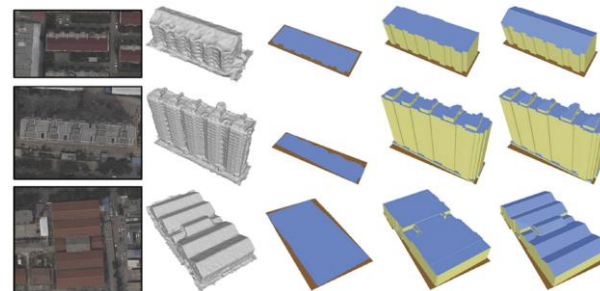
# 规范化三维矢量重建



**Variational Building Modeling**

*3DV 2017*

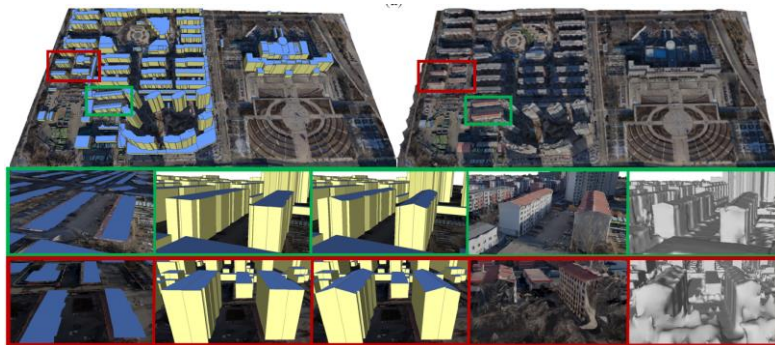
*LOD2室外矢量建模*



**Urban Scene Modeling**

*ECCV 2018*

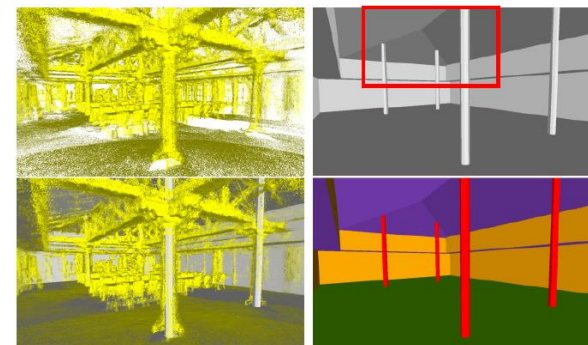
*LOD2室外矢量建模*



**Non Manhattan LOD Modeling**

*IEEE TIP 2021*

*LOD2室外矢量建模*



**Multistep Indoor Modeling**

*ISPRS JPRS 2021*

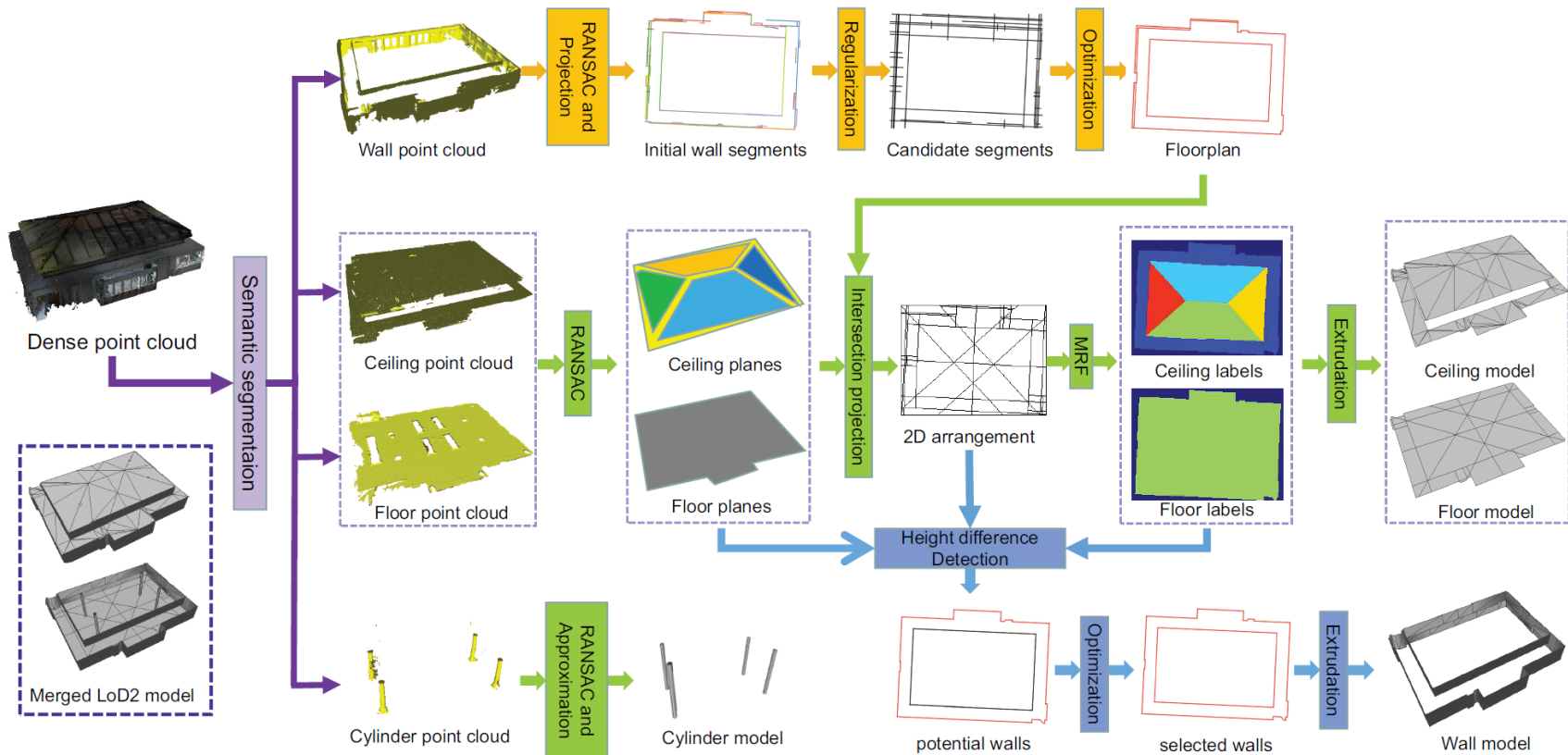
*LOD2室内矢量建模*

# Vectorized Indoor Modeling

## Vectorized Indoor Surface Reconstruction from 3D Point Cloud with Multistep 2D Optimization

*Jiali Han, Mengqi Rong, Hanqing Jiang, Hongmin Liu, Shuhan Shen*

*ISPRS JPRS 2021*



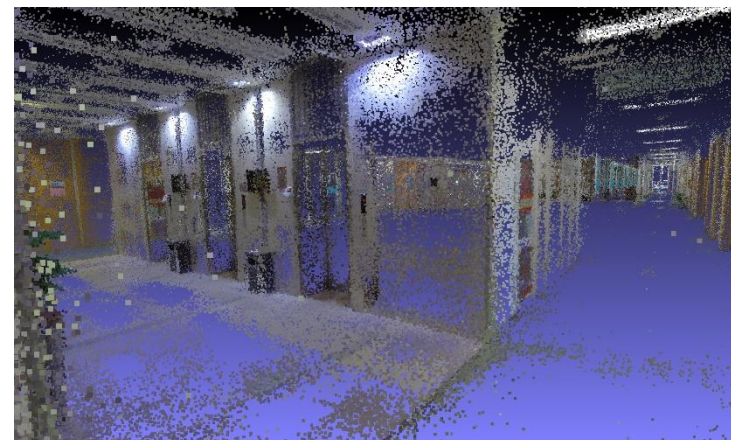
# Vectorized Indoor Modeling



Insta360 Pro2



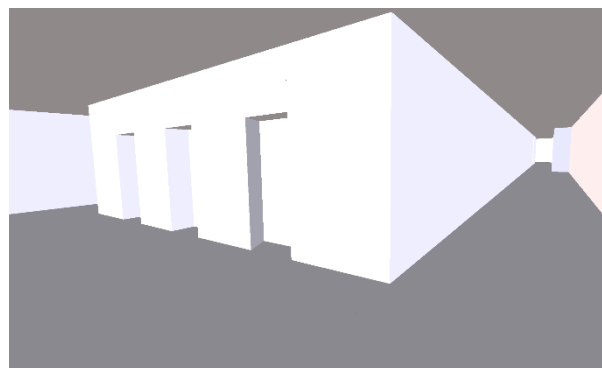
6分钟全景相机视频, 960平方米



稠密三维点云, 1000万三维点



三角网格模型, 300万面片



LOD2级矢量模型, 930面片



LOD2级纹理矢量模型

# Vectorized Indoor Modeling



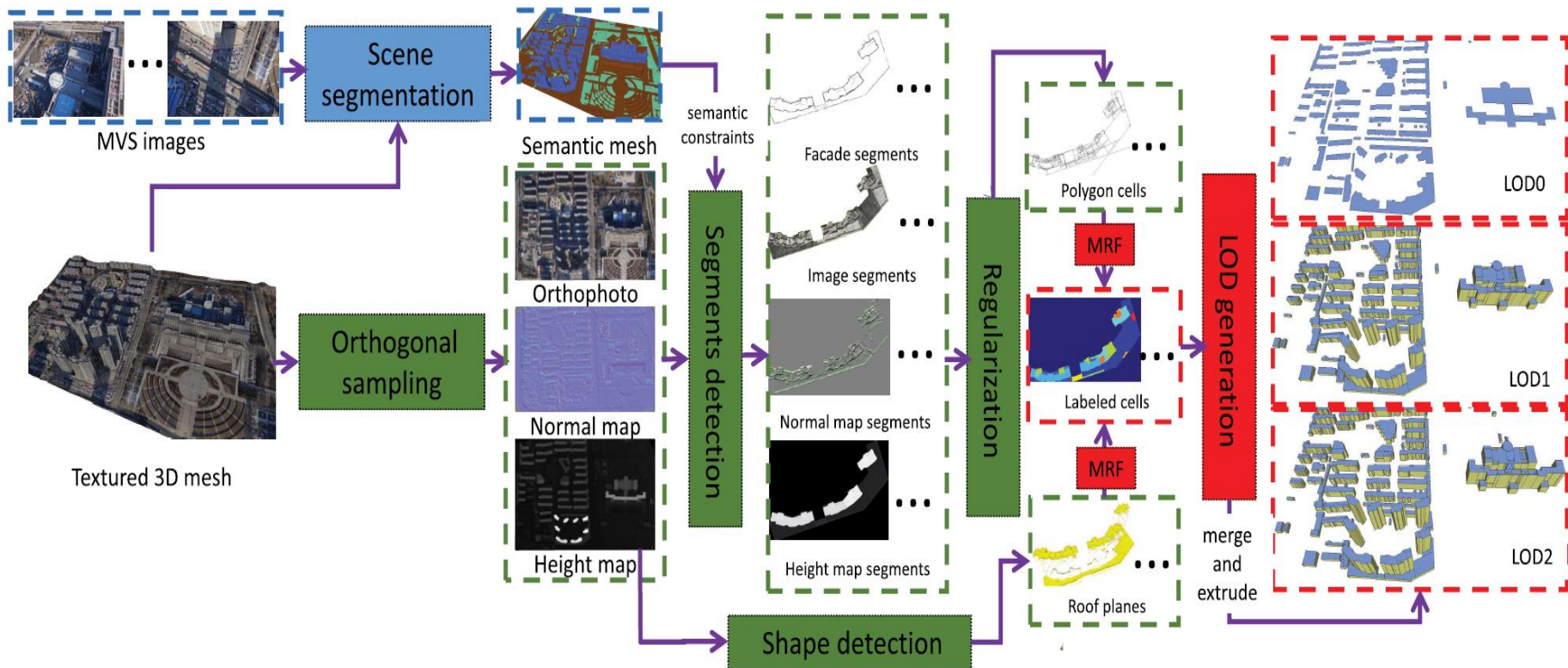
6分钟全景相机视频, Insta Pro2 6镜头全景

# Vectorized Outdoor Modeling

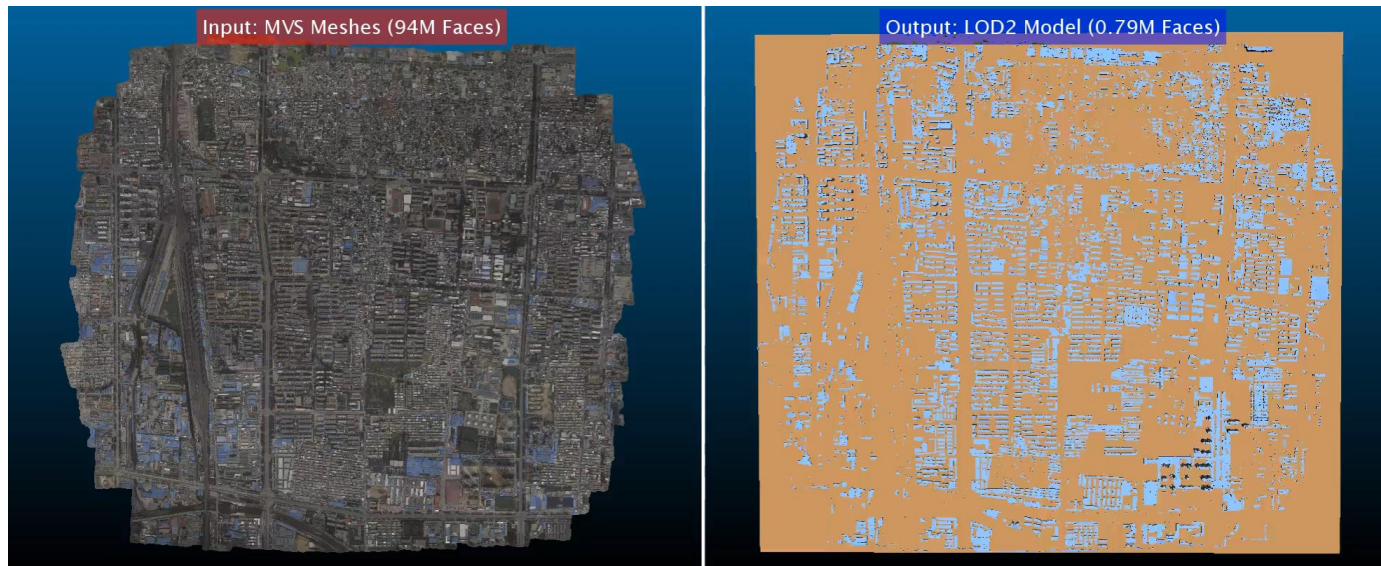
## Urban Scene LOD Vectorized Modeling From Photogrammetry Meshes

Jiali Han, Lingjie Zhu, Xiang Gao, Zhanyi Hu, Liyang Zhou, Hongmin Liu, Shuhan Shen

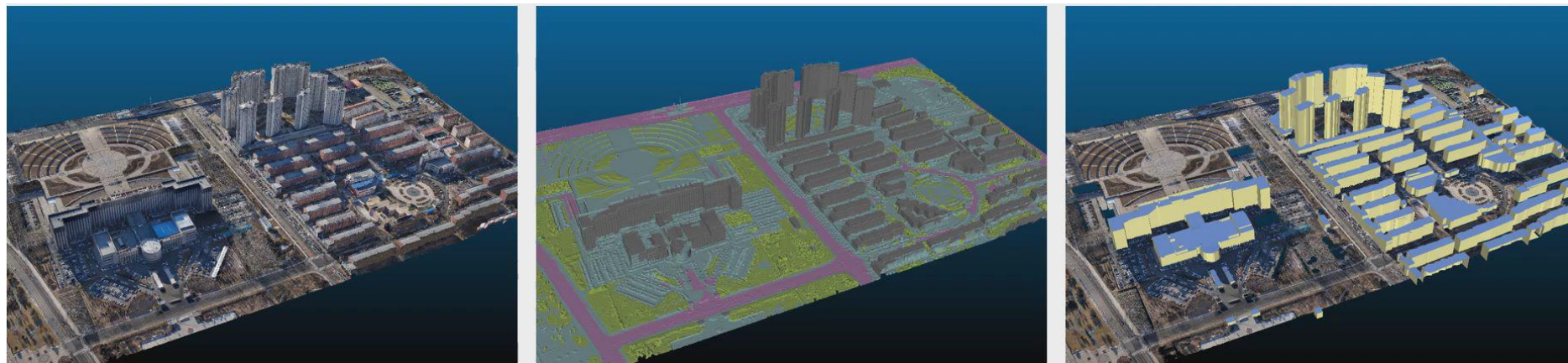
IEEE TIP 2021



# Vectorized Outdoor Modeling

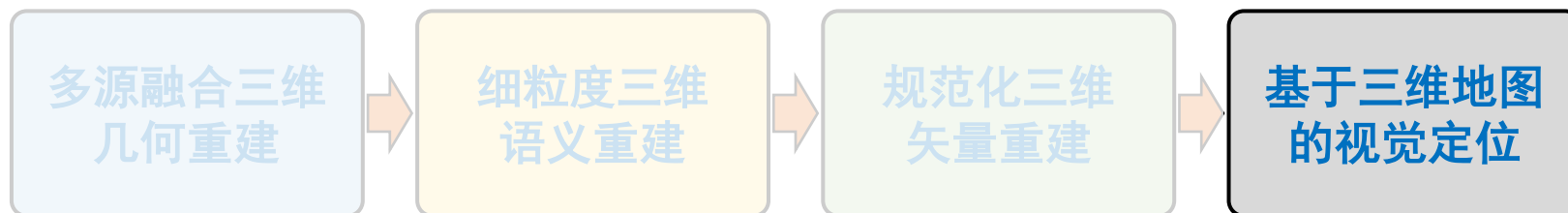


安阳市城区倾斜摄影LOD2矢量建模（15平方公里）

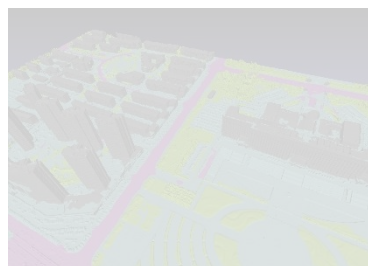


呼和浩特政府广场倾斜摄影LOD2矢量建模（1平方公里）

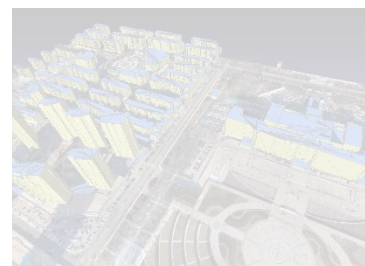
# 基于三维地图的视觉定位



3D几何模型



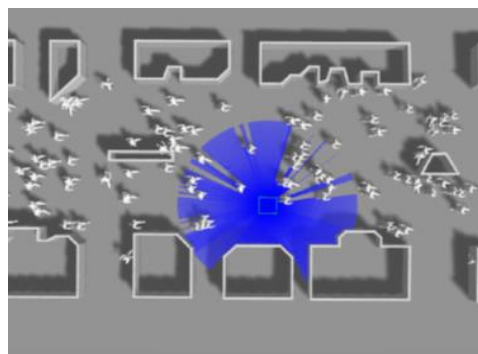
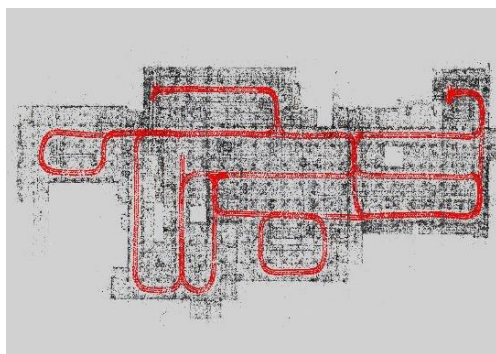
3D语义模型



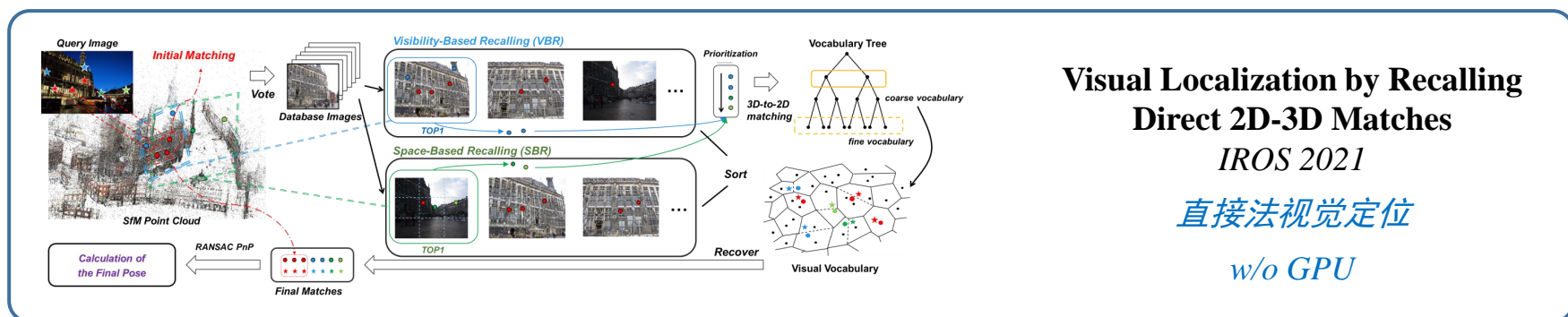
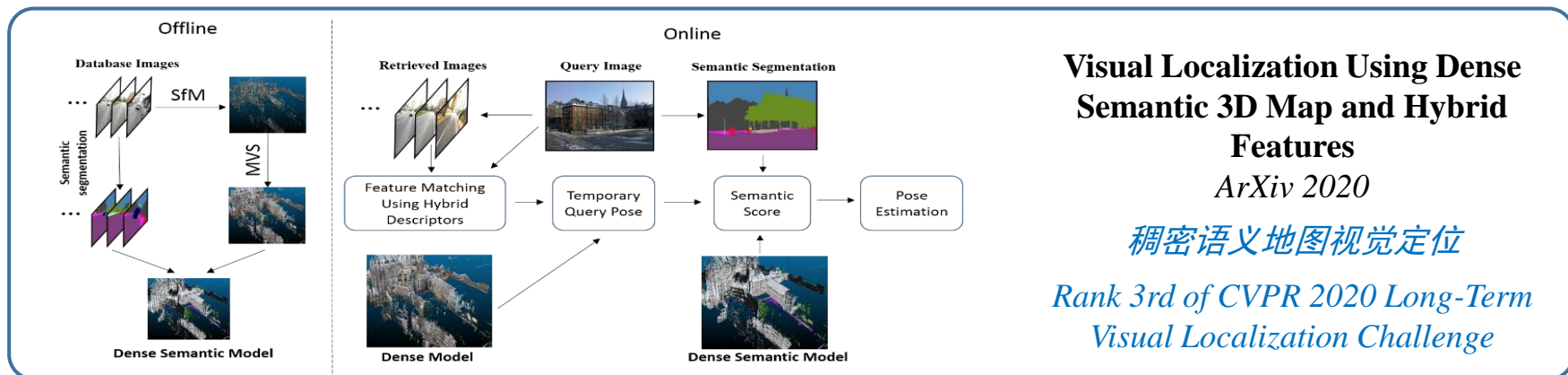
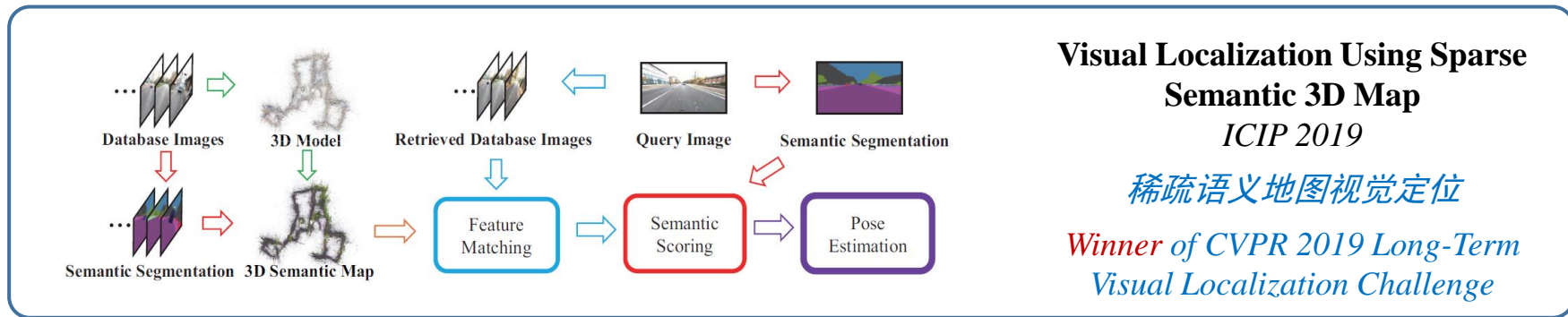
3D矢量模型



6DoF视觉定位



# 基于三维地图的视觉定位



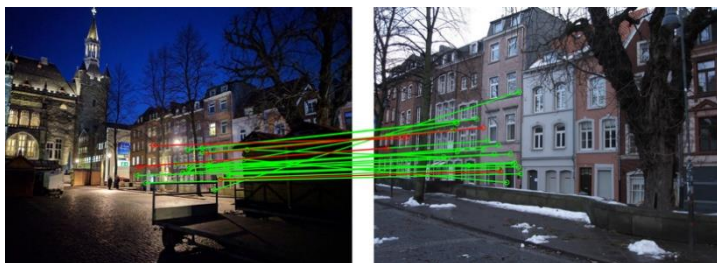
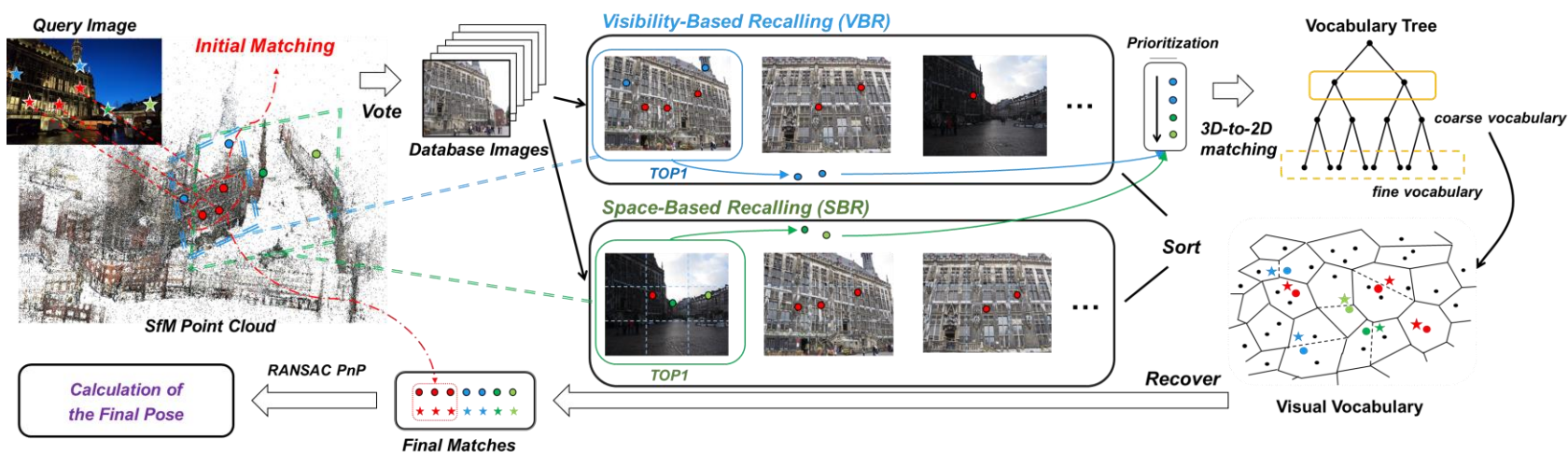


# Direct Visual Localization

## Recalling Direct 2D-3D Matches for Large-Scale Visual Localization

Zhuo Song, Chuting Wang, Yuqian Liu, Shuhan Shen

IROS 2021



Visibility-Based Recalling



Space-Based Recalling

# Direct Visual Localization

RobotCar Seasons		
Precision	All Day	All Night
	high / medium / coarse	high / medium / coarse
AS v1.1 [1]	43.6 / 76.0 / 94.0	1.8 / 7.4 / 14.2
CPF [8]	48.0 / 78.0 / 94.2	3.4 / 9.5 / 17.0
CSL [9]	45.3 / 73.5 / 90.1	0.6 / 2.6 / 7.2
SMC [10]	50.3 / 79.3 / 95.2	7.1 / <b>22.4</b> / <b>45.3</b>
NetVLAD [15]	6.4 / 26.3 / 90.9	0.3 / 2.3 / 15.9
DenseVLAD [14]	7.6 / 31.2 / 91.2	1.0 / 4.4 / 22.7
Ours ( $N_t=100$ )	52.7 / 79.4 / 94.0	6.9 / 13.1 / 20.6
Ours ( $N_t=200$ )	<b>53.3</b> / 81.0 / 95.6	9.1 / 17.8 / 29.4
Ours ( $N_t=500$ )	<b>53.3</b> / <b>81.2</b> / <b>96.4</b>	<b>10.3</b> / 20.1 / 32.7

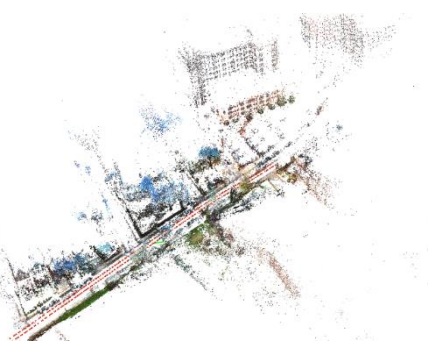
Aachen Day-Night		
Precision	Day	Night
	high / medium / coarse	high / medium / coarse
AS v1.1 [1]	85.3 / 92.2 / 97.9	39.8 / 49.0 / 64.3
CPF [8]	76.7 / 88.6 / 95.8	33.7 / 48.0 / 62.2
CSL [9]	52.3 / 80.0 / 94.3	29.6 / 40.8 / 56.1
SMC [10]	- / - / -	- / - / -
NetVLAD [15]	0.0 / 0.2 / 18.9	0.0 / 0.0 / 14.3
DenseVLAD [14]	0.0 / 0.1 / 22.8	0.0 / 1.0 / 19.4
Ours ( $N_t=100$ )	87.9 / 94.8 / <b>98.2</b>	61.2 / 72.4 / <b>80.6</b>
Ours ( $N_t=200$ )	88.6 / 94.9 / <b>98.2</b>	<b>63.3</b> / <b>73.5</b> / <b>80.6</b>
Ours ( $N_t=500$ )	<b>90.2</b> / <b>95.6</b> / <b>98.2</b>	<b>63.3</b> / <b>73.5</b> / <b>80.6</b>



SfM image



Query image



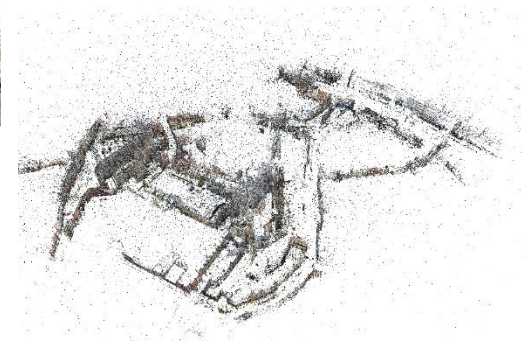
RobotCat Seasons



SfM image



Query image

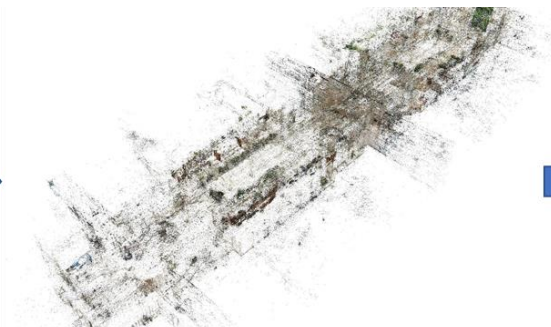


Aachen Day-Night

# 综合示例：室内视觉导航地图



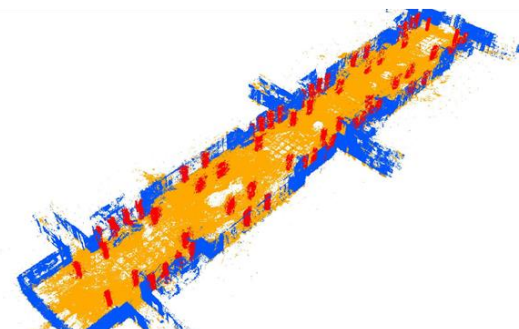
众包图像



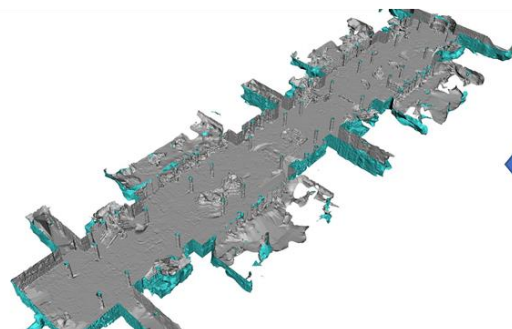
稀疏点云



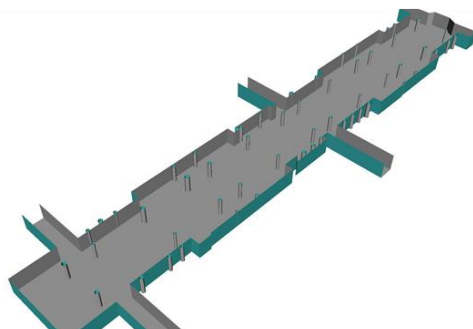
密集点云



点云分割



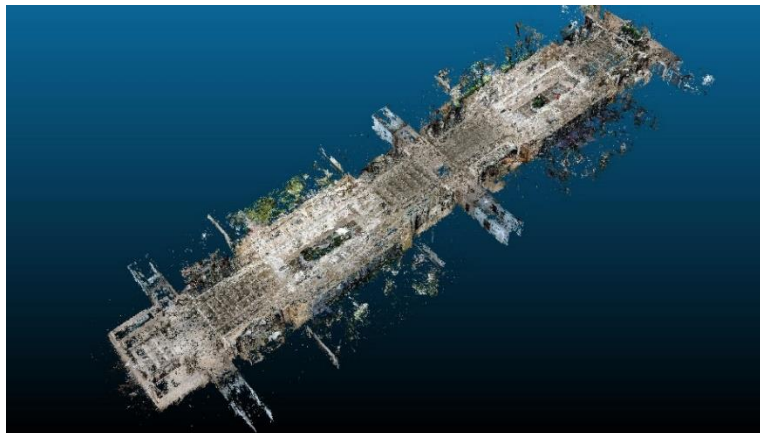
粗糙网格



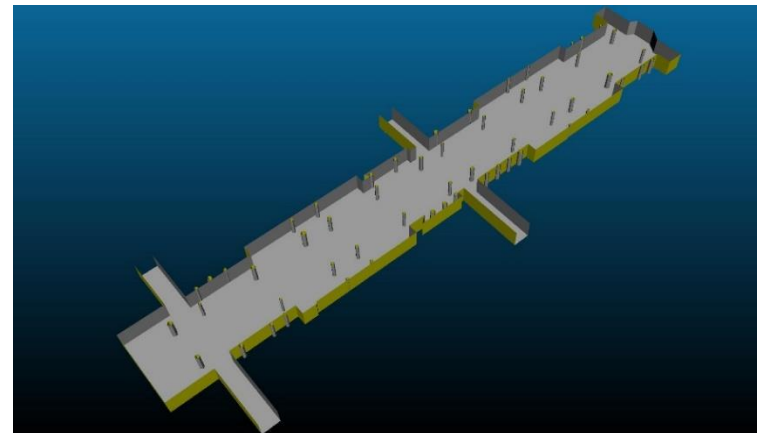
矢量网格

室内场景点云地图、矢量地图与单目视觉定位

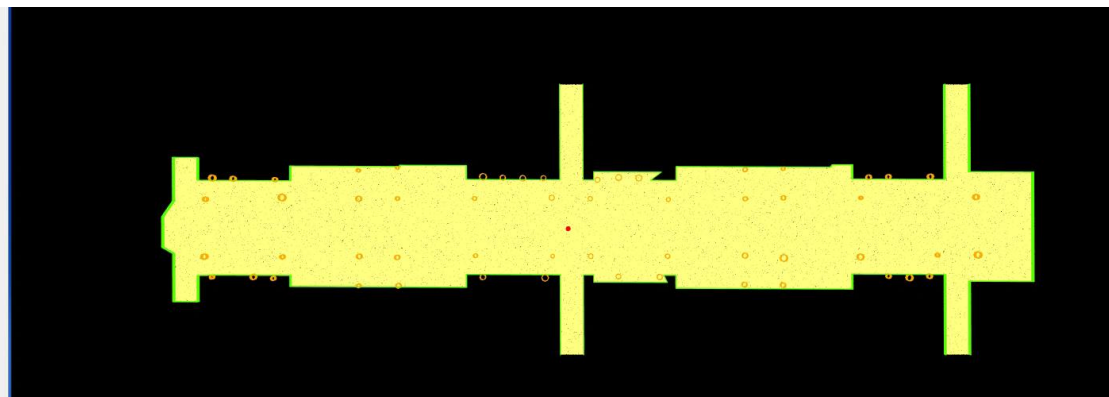
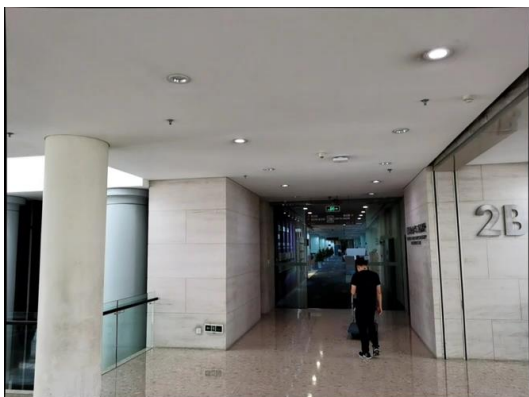
# 综合示例：室内视觉导航地图



点云地图



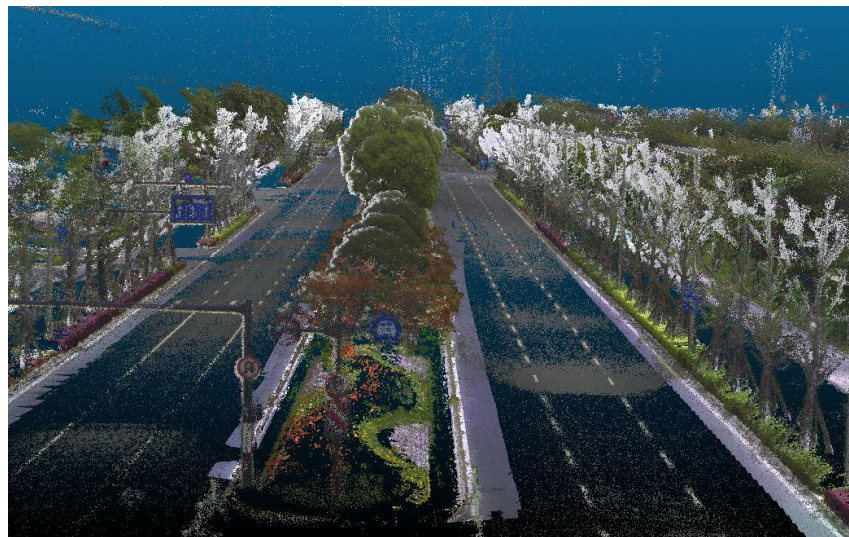
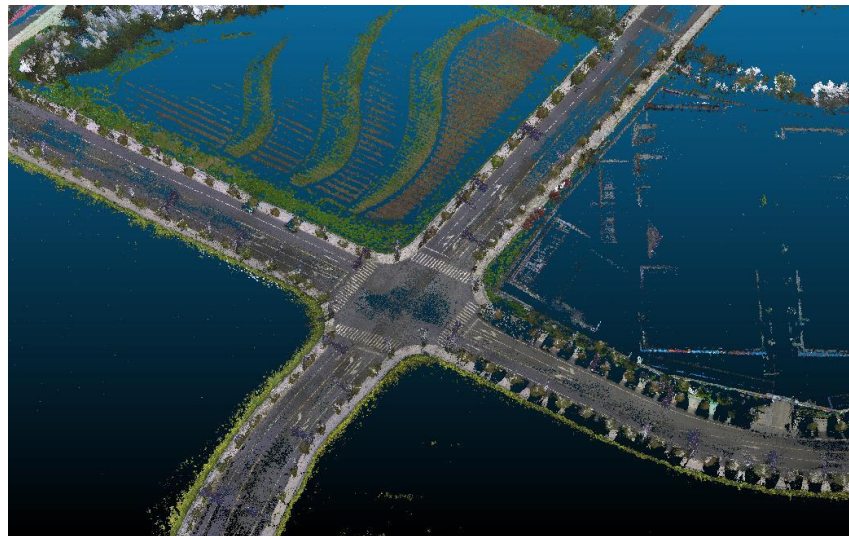
矢量地图



单目视觉定位

室内场景点云地图、矢量地图与单目视觉定位

# 综合示例：道路视觉导航地图



Dense Road Maps (>30000 images)

# 综合示例：道路视觉导航地图



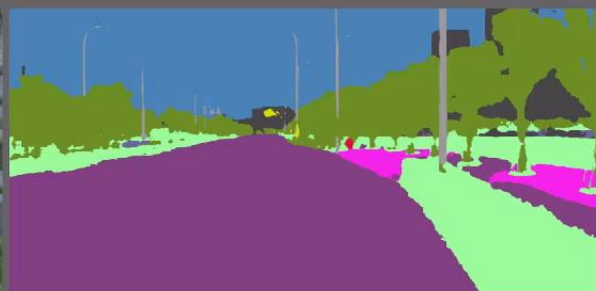
Qurey image



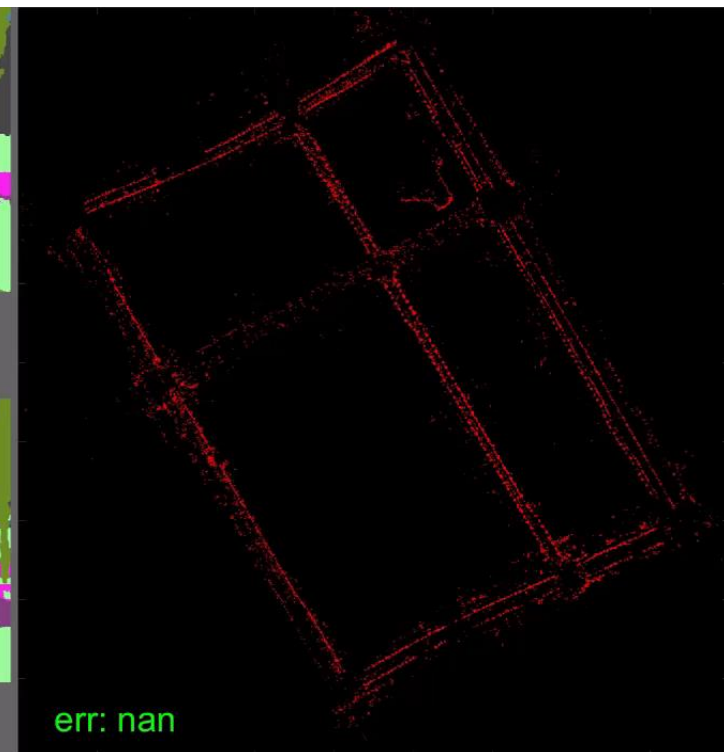
Qurey semantic



Database image



Database semantic

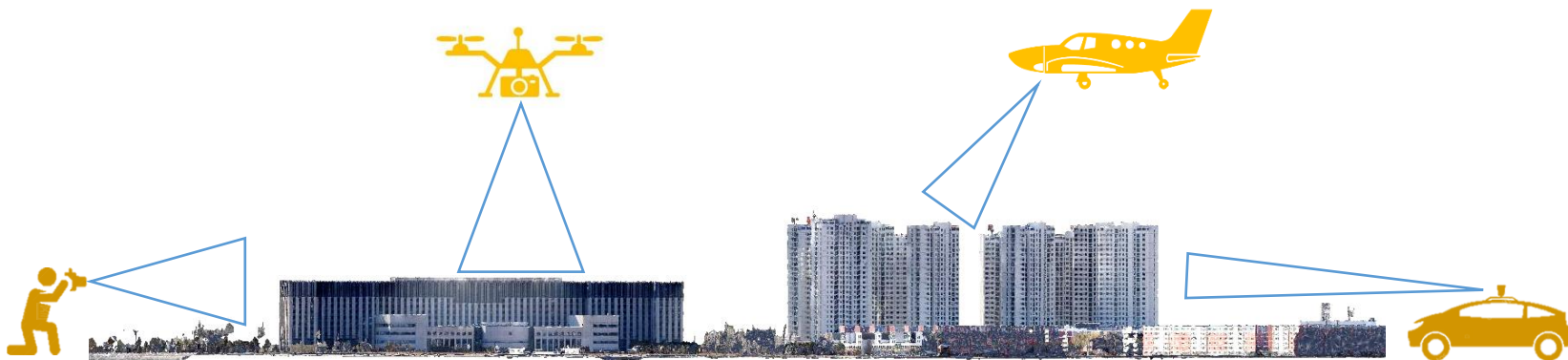


无人车单目视觉定位（语义地图：2018年8月，定位图像：2019年3月）

单幅图像定位成功率：92.1%，定位中值误差：0.41m

杭州萧山无人驾驶测试场无人车单目视觉定位  
（前向单目相机，2600幅定位图像）

# 总结



以航拍、车载、全景、手机图像为主要数据源，低成本、高精度、全自动、高效率的获取从天空到地面、从室外到室内的城市场景三维表达。



大兴航天星汉小区实景三维



五台山佛光寺三维VR漫游



丰台小屯路三维道路地图