Robust Tightly-Coupled Visual-Inertial Odometry with Pre-built Maps in High Latency Situations

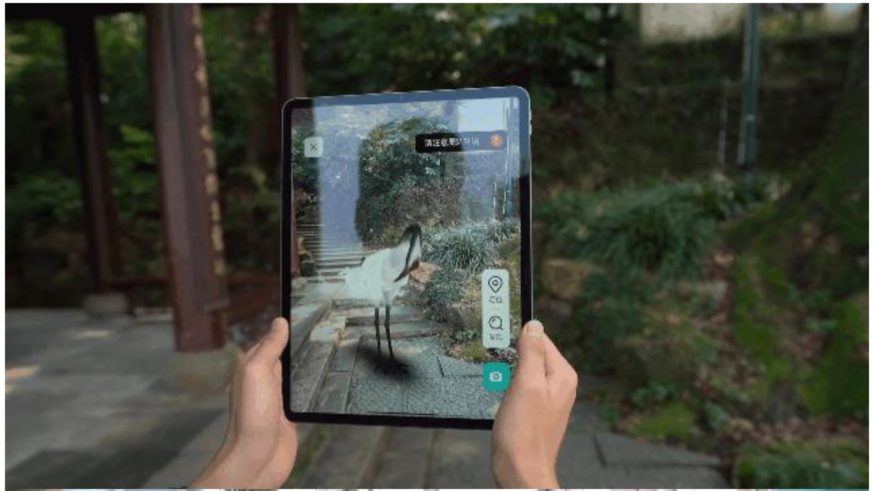
Hujun Bao[1], Weijian Xie[1,2], Quanhao Qian[2], Danpeng Chen[1,2,3],

Shangjin Zhai[2], Nan Wang[2,3], and Guofeng Zhang[1*]

[1]State Key Lab of CAD&CG, Zhejiang University [2]SenseTime Research [3]Tetras.AI

❑Motivation

With the rise of the digital twin and high-precision maps, the demand for AR and VR of large scenes combined with high-precision maps gradually becomes prosperous.



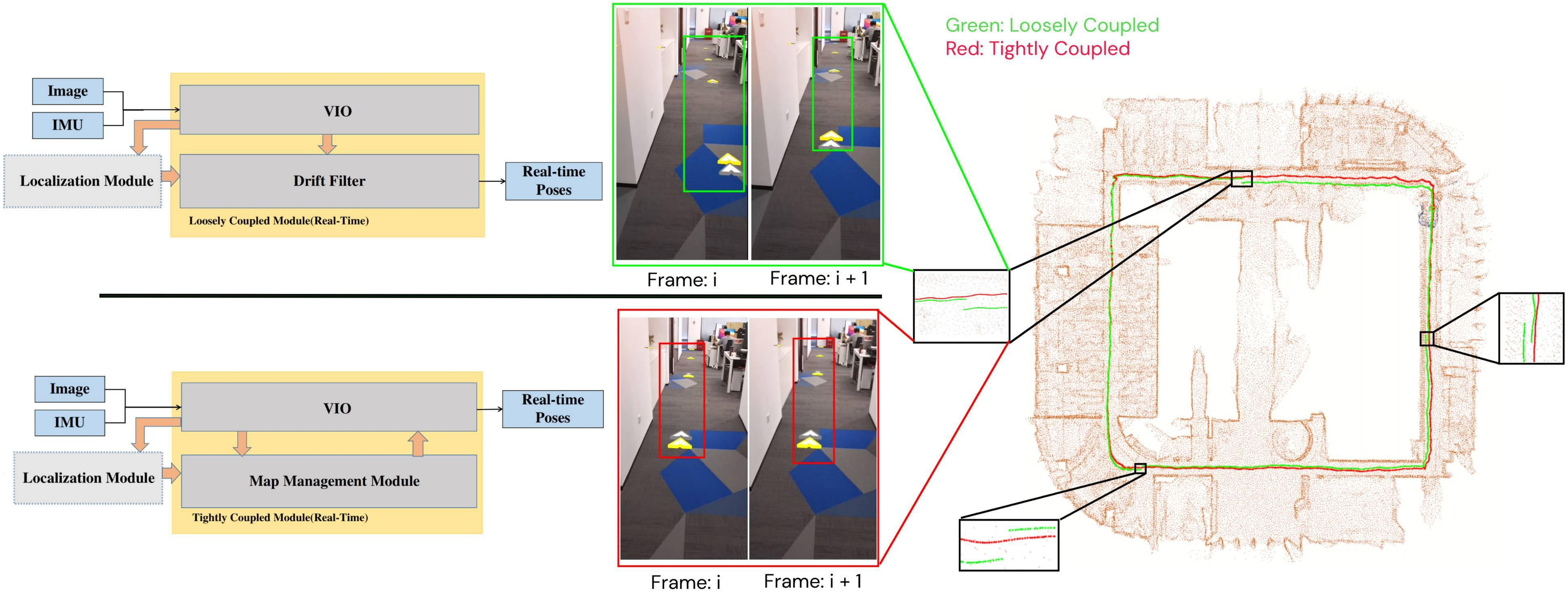
❑Motivation

	GNSS	VIO/SLAM	Global Localization Algorithm based on the pre-built map
Advantage	• Global position • No need for the pre-built map	• Smooth trajectory • Local high-precision pose	• Global high-precision pose
Disadvantage	• Cannot work in indoor scenes • Low-precision	• Accumulate drift	• High algorithm complexity • Unsmooth trajectory

- An affordable way to combine the advantages of VIO and pre-built maps is to fuse the pre-built map into the VIO tracking process.

❑Motivation

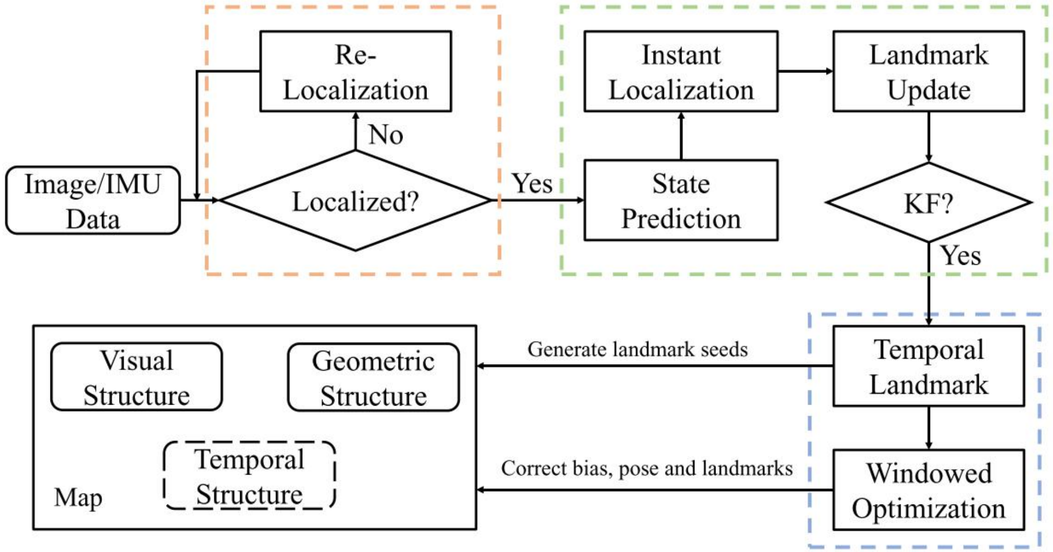
Loosely coupled methods easily lead to jumpness of trajectory, while tightly coupled methods do not.



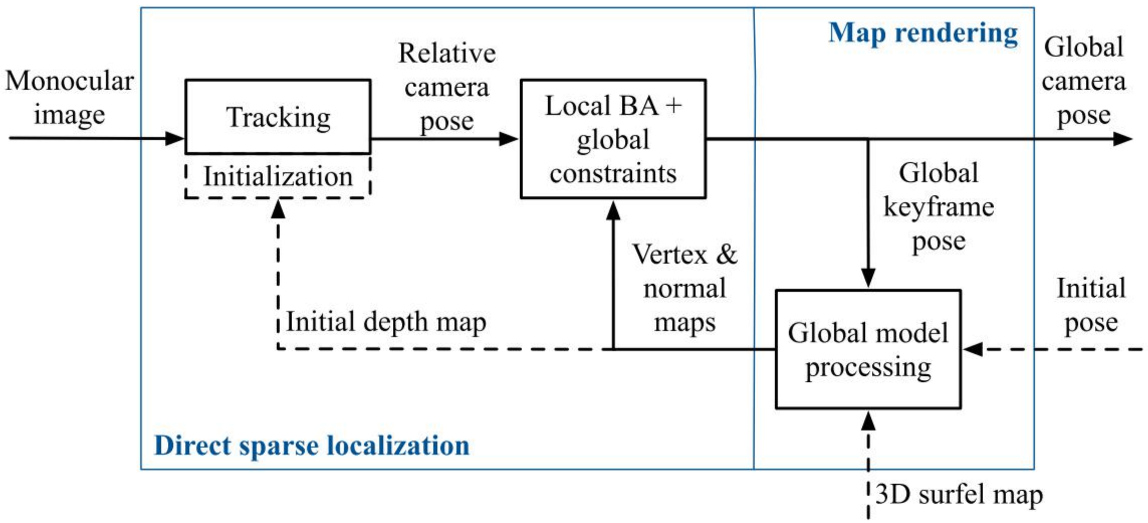
❑Motivation

The weakness of previous methods:

- Sensitive to map noise and scenarios changes.
- It is challenging to achieve good performance under the condition of localization with time delay and low frequency.



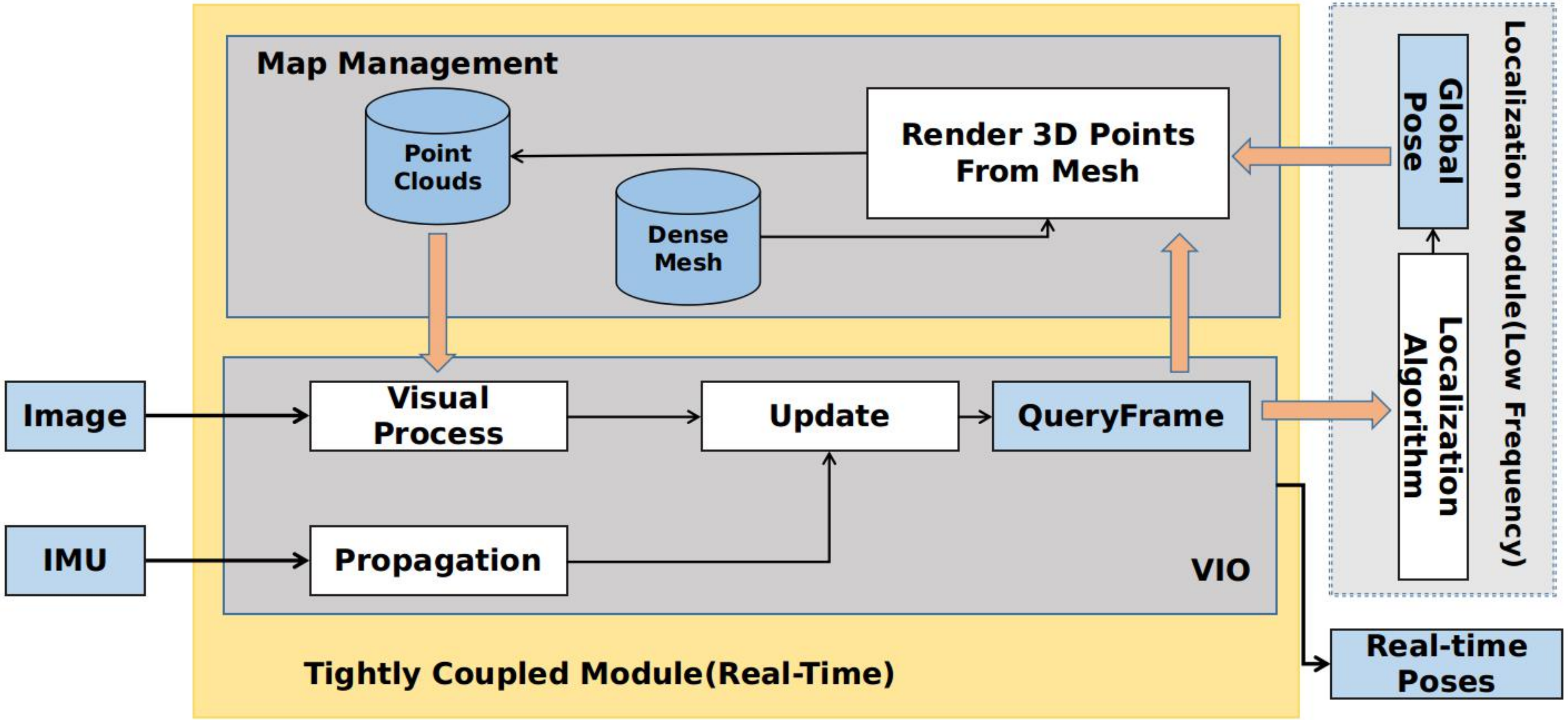
GMM w/map(Huang et al., 2020)



DSL(Ye et al., 2020)

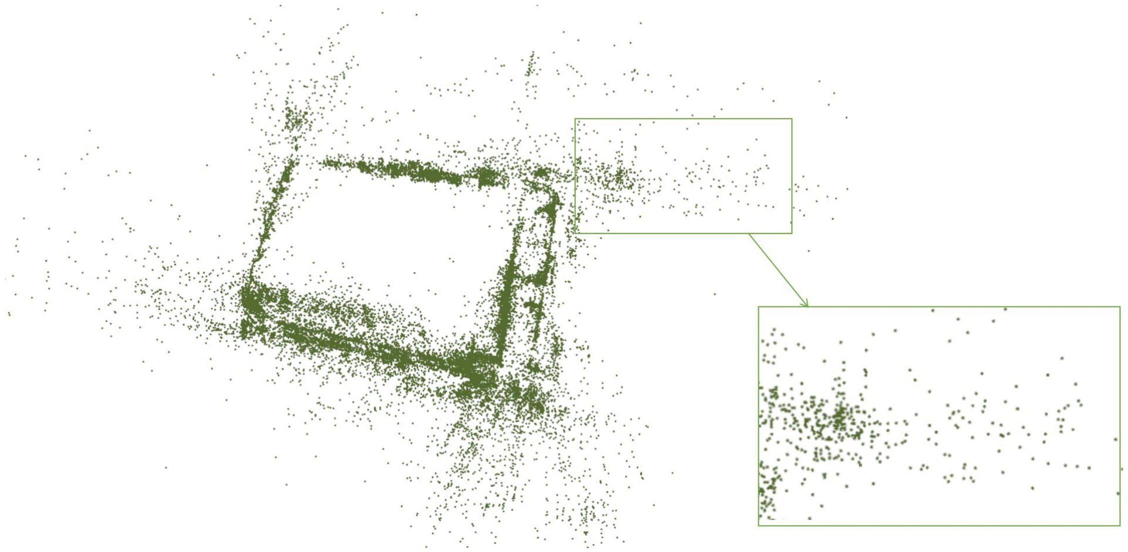
❑Contribution

- A complete scheme to generate the association between the pre-built map and local features
- Different constraints for different types of map points
- A degeneration state recovery strategy

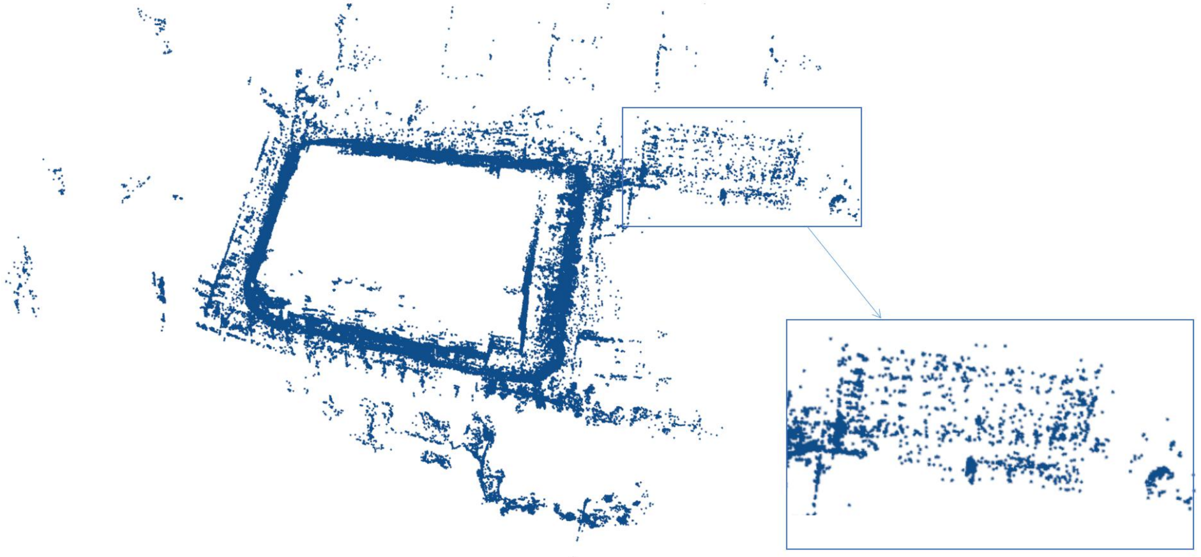


❑Structure Extraction

- Extract the position and normal of ORB feature from the dense mesh.
- The descriptor of 3D point comes from the query image.
- The points obtained by the real-time pose are defined as local map points.
- The points obtained by the global localization pose are defined as global map points.

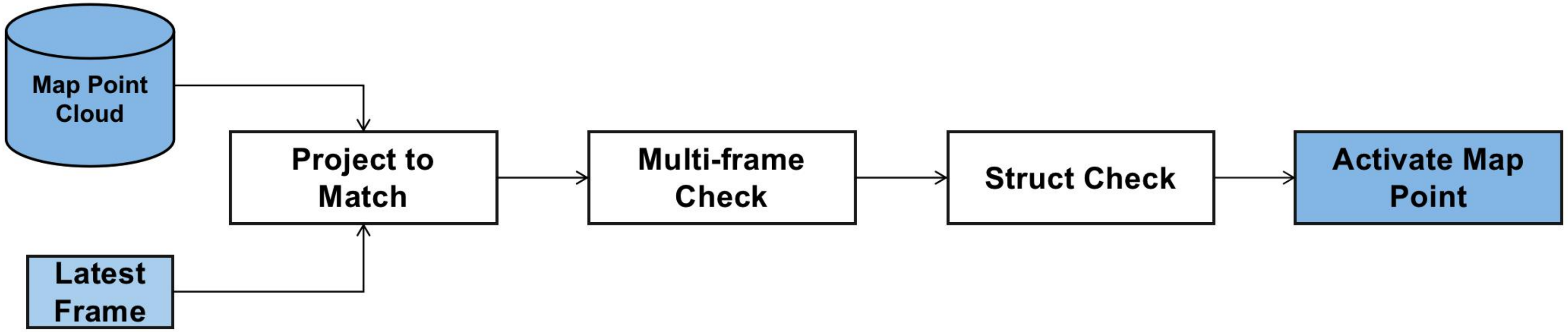


Point clouds generated by **2D-3D** matches

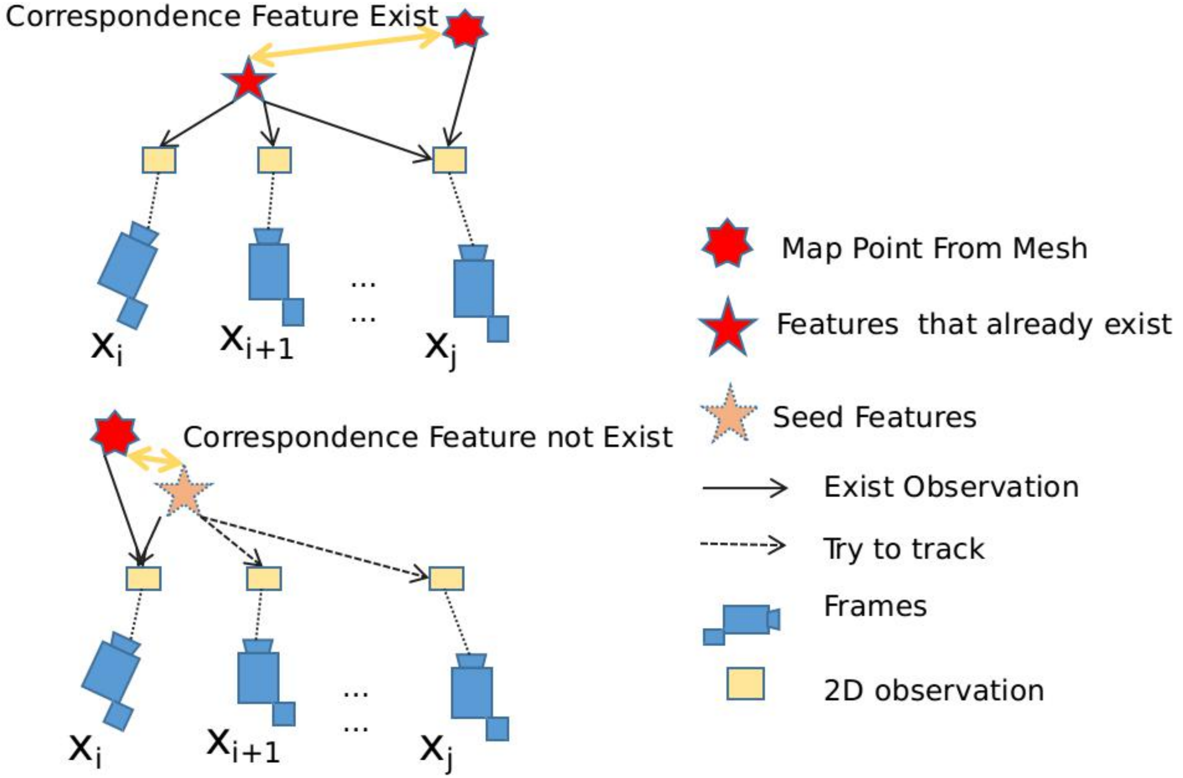


Point clouds generated by **ray casting**

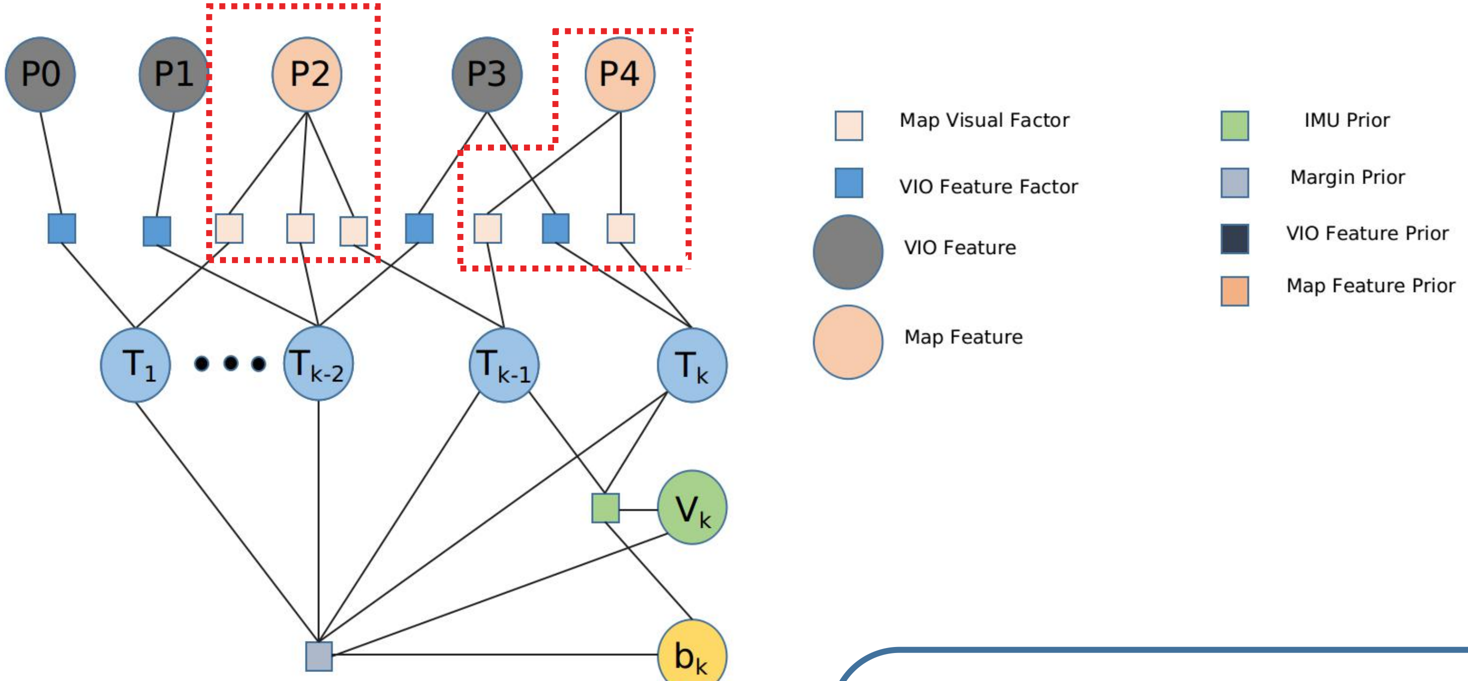
❑Visual Processing



- Each map point will correspond to a local feature.
- Only when there are enough activate map points will we add them to VIO’s status updates as constraints.
- We will re-check activate map points in every frame.



❑Visual-Inertial State Estimate



- Global Map Point Cost:

$$\tilde{\boldsymbol{r}}_{gm}^{j} = \sum_i \omega_j \tilde{\boldsymbol{r}}_{gm}^{j,i} = \sum_i \omega_j \left(\pi(\boldsymbol{R}_G^{C_i} \boldsymbol{p}_{m_j}^G + \boldsymbol{p}_G^{C_i}) - \mu_j^{C_i} \right),$$

$$\boldsymbol{p}_{m_j}^G = \boldsymbol{R}_M^G \boldsymbol{p}_{m_j}^M + \boldsymbol{p}_M^G,$$

$$\omega_j = \begin{cases} 0 & \text{if } l_j < \alpha \\ b^{(l_j-\alpha)} & \text{if } l_j < \beta \\ b^{(\beta-\alpha)} & \text{other} \end{cases}$$

$$C_z\left(\tilde{\boldsymbol{X}}_{k+1}\right) = \left\| \omega_j \left(\boldsymbol{H}_x^j \tilde{\boldsymbol{X}}_{k+1} - \tilde{\boldsymbol{r}}_{gm}^j \right) \right\|_{\sigma \boldsymbol{I}_2}^2,$$

- Local Map Point Cost:

$$\tilde{\boldsymbol{r}}_n^j = \sum_i \tilde{\boldsymbol{r}}_n^{j,i} = \sum_i \boldsymbol{n}_{m_j}^{G\top} \left(\boldsymbol{p}_{m_j}^G - \left(\boldsymbol{R}_{C_i}^G \boldsymbol{p}_{f_j}^{C_i} + \boldsymbol{p}_{C_i}^G \right) \right),$$

$$\boldsymbol{n}_{m_j}^G = \boldsymbol{R}_M^G \boldsymbol{n}_{m_j}^M,$$

$$\boldsymbol{p}_{f_j}^{C_i} = \boldsymbol{R}_G^{C_i} \left(\boldsymbol{R}_{C_k}^G \left(d_k K^{-1} \pi^{-1} \left(\mu_j^{C_k} \right) \right) + \boldsymbol{p}_{C_k}^G \right) + \boldsymbol{p}_G^{C_i},$$

$$C_n\left(\tilde{\boldsymbol{X}}_{k+1}\right) = \left\| \boldsymbol{H}_x \tilde{\boldsymbol{X}}_{k+1} - \tilde{\boldsymbol{r}}_n^j \right\|_{\sigma}^2,$$

❑Degeneration Analysis

- System Recovery

$$x_{m_j}^{C_i} = \pi\left(K\left(R_G^{C_i}\left(\boxed{R_M^G} p_{m_j}^M + \boxed{p_M^G}\right) + p_G^{C_i}\right)\right),$$

$$\min_{\hat{q}_M^G, \hat{p}_M^G} \left\{ \sum \left\| q_i^G \otimes q_i^{M^{-1}} \otimes \hat{q}_G^M \right\|_2 + \sum \left\| p_i^G - \hat{q}_M^G p_i^M - \hat{p}_M^G \right\|_2 \right\}, \quad (16)$$

$$entropy = 0.5 \times \log\left((2\pi e)^k det\left(H_{rel}\right)\right).$$

- $norm\left(p_M^G - \hat{p}_M^G\right) > p_{threshold}$
- $entropy > \lambda_e$

- The VIO module needs to be initialized independently before using the constraints of the pre-built map. After the initialization of the VIO, the system will directly fall into a degeneration state.

❑Experiments

Dataset	V101	V102	V103	V201	V202	V203
BVIO	0.055	0.064	0.086	0.054	0.106	0.129
RTC-VIO	**0.020**	**0.023**	**0.035**	**0.021**	0.027	**0.047**
OpenVINS	0.050	0.084	0.078	0.068	0.064	0.081
VINS-Mono (loop)	0.039	0.037	0.087	0.076	0.105	0.330
ORB (online)	0.427	1.176	0.985	0.417	0.864	2.308
GMM W/ Map	0.023	0.057	0.058	0.047	0.040	0.392
DSL (left cam)	0.035	0.034	0.045	0.026	**0.023**	0.103
MSCKF (w/Map)	0.056	0.055	0.087	0.069	0.089	0.149
ORB (offline)	0.041	0.017	0.029	0.051	0.017	0.030

Dataset	BVIO	RTC-VIO	ORB (online)	VINS-Mono (loop)	GMM	DSL
indoor	0.230	**0.023**	0.826	0.159	-	0.050
indoor patial	0.062	**0.020**	3.300	0.078	0.068	0.060
outdoor	2.253	**0.19**	14.702	2.963	25.463	0.383

- The experiments on EurocMav datasets and simulation datasets show that our method can achieve higher accuracy compared with state-of-the-art methods.

❑Experiments

Dataset	V101	V102	V103	V201	V202	V203
GMM w/ wrong map	0.469	0.366	0.413	0.851	0.831	1.987
RTC-VIO w/ wrong map	**0.029**	**0.039**	**0.037**	**0.022**	**0.032**	**0.052**
GMM w/ wrong map & GT	0.422	0.399	0.309	0.758	0.803	0.596
RTC-VIO w/ wrong map & GT	**0.023**	**0.023**	**0.033**	**0.019**	**0.028**	**0.041**

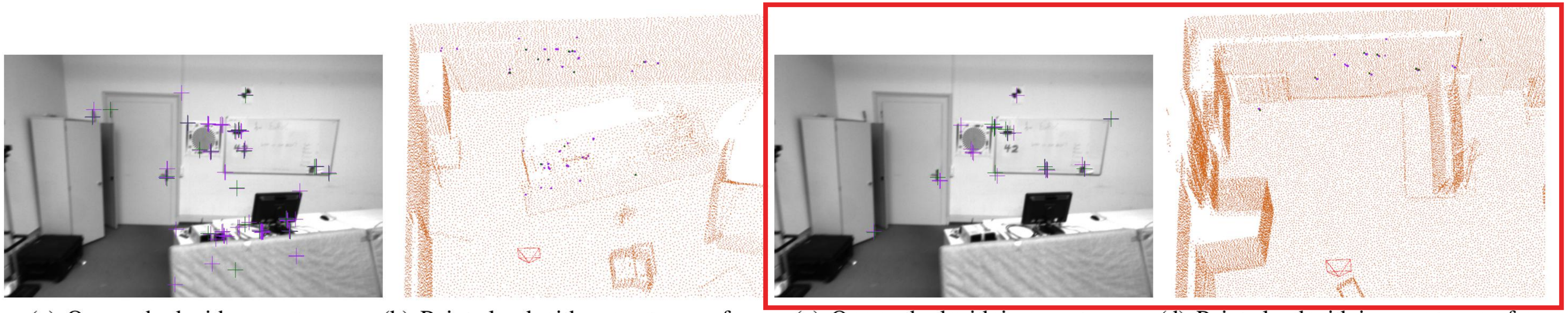
Dataset	σ:0	σ: 0.1	σ: 0.3	σ: 0.5	σ: 1.0
indoor	0.023	0.049	0.076	0.084	0.148
indoor patial	0.020	0.026	0.036	0.047	0.054
outdoor	0.190	0.212	0.215	0.217	0.258

DSL fails in all three synthetic datasets when we add standard deviation σ = 0.1m to the pre-built map.

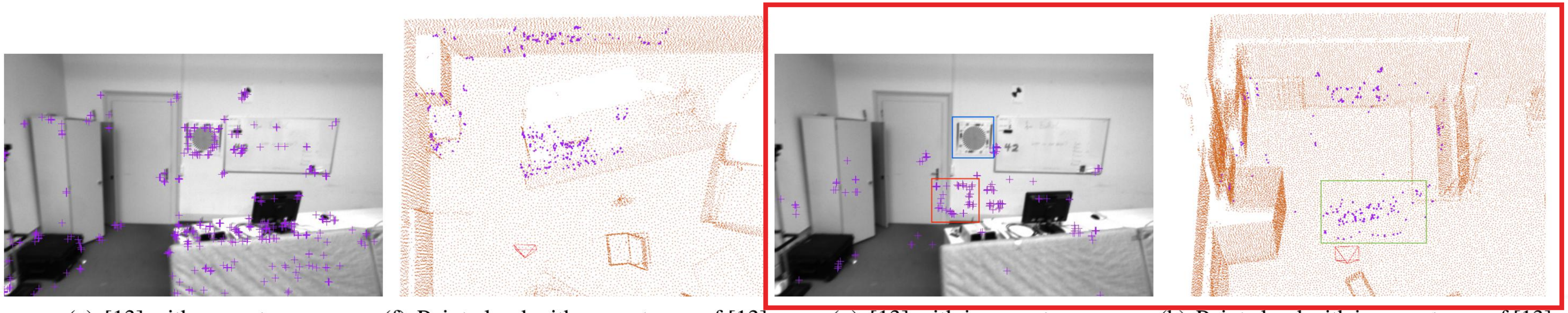
- Compared with GMM and DSL, our method is more robust to the changes of scenarios and the noise of the pre-built maps.

❑Experiments

- Our method can filter out the outliers introduced by environmental changes, while the GMM's algorithm cannot.



(a) Our method with correct map (b) Point cloud with correct map of our method (c) Our method with incorrect map (d) Point cloud with incorrect map of our method



(e) [13] with correct map (f) Point cloud with correct map of [13] (g) [13] with incorrect map (h) Point cloud with incorrect map of [13]

❑Experiments

- ①② show that the number of point clouds from the pre-built map will significantly affect the coupled result.

	stages						indoor	indoor partial	outdoor
	LP	GP	M	S	R	LR			
①	✓	✓	✓	✓	-	-	0.089	0.060	0.226
②	✓	✓	✓	✓	✓	-	0.307	0.135	0.456
③	✓	✓	-	✓	-	-	0.096	0.085	0.227
④	✓	✓	✓	-	-	-	0.098	0.078	0.239
⑤	✓	✓	-	-	-	-	0.102	0.082	0.253
⑥	-	✓	✓	✓	-	-	0.102	0.051	0.270
⑦	✓	-	✓	✓	-	-	0.162	0.094	1.680
⑧	✓	✓	✓	✓	-	✓	0.105	0.244	0.331

❑Experiments

- ①⑥⑦ show that both local map points and global map points can improve the accuracy of the coupled algorithm, among which global map points are more helpful for improving the accuracy.

	stages						indoor	indoor partial	outdoor
	LP	GP	M	S	R	LR			
①	✓	✓	✓	✓	-	-	0.089	0.060	0.226
②	✓	✓	✓	✓	✓	-	0.307	0.135	0.456
③	✓	✓	-	✓	-	-	0.096	0.085	0.227
④	✓	✓	✓	-	-	-	0.098	0.078	0.239
⑤	✓	✓	-	-	-	-	0.102	0.082	0.253
⑥	-	✓	✓	✓	-	-	0.102	0.051	0.270
⑦	✓	-	✓	✓	-	-	0.162	0.094	1.680
⑧	✓	✓	✓	✓	-	✓	0.105	0.244	0.331

❑Experiments

- ①⑥⑧ show that using reprojection constraints for local map points will take a negative effect.

	stages						indoor	indoor partial	outdoor
	LP	GP	M	S	R	LR			
①	✓	✓	✓	✓	-	-	0.089	0.060	0.226
②	✓	✓	✓	✓	✓	-	0.307	0.135	0.456
③	✓	✓	-	✓	-	-	0.096	0.085	0.227
④	✓	✓	✓	-	-	-	0.098	0.078	0.239
⑤	✓	✓	-	-	-	-	0.102	0.082	0.253
⑥	-	✓	✓	✓	-	-	0.102	0.051	0.270
⑦	✓	-	✓	✓	-	-	0.162	0.094	1.680
⑧	✓	✓	✓	✓	-	✓	0.105	0.244	0.331

❑Experiments

Table 8: Evaluation on general localization performanceon synthetic dataset with APE (m) under different time delay. The interval of sending localization request is set to 1000ms.

Delay (ms)	indoor		indoor partial		outdoor	
	w	w/o	w	w/o	w	w/o
200	**0.035**	0.037	0.037	**0.036**	**0.267**	0.300
400	**0.041**	0.049	0.041	**0.037**	**0.284**	0.327
800	**0.077**	0.088	0.049	**0.046**	0.378	**0.330**
1200	**0.121**	0.146	**0.047**	0.054	**0.458**	0.567

Table 9: Evaluation on ablation of local map point constraints with APE (m) under different localization frequencies. The latency of localization pose is set to 400ms.

Interval (ms)	indoor		indoor partial		outdoor	
	w	w/o	w	w/o	w	w/o
1000	**0.041**	0.049	0.041	**0.037**	**0.284**	0.327
2000	**0.060**	0.065	0.043	**0.042**	**0.319**	0.368
4000	**0.064**	0.068	**0.043**	0.052	**0.528**	0.541
8000	**0.076**	0.105	**0.052**	0.055	**0.528**	0.541
12000	**0.092**	0.129	0.061	**0.056**	**0.527**	0.588

Still better than BVIO!

❑Demo

Video 1: Our method vs ARCore-LC on indoor scene
Video 2: Our method vs ARCore-LC on outdoor scene
Video 3: Our method on V103_difficult
Video 4: Comparations on V103_difficult
Video 5: Our method on synthetic indoor scene

Thank you!

xieweijian@sensetime.com