1 Mapping Ablation study

We first conduct ablation study experiments to validate the effectiveness of each component in the proposed mapping module.

Multi-camera loop closure. The need for multi-camera loop closure detection arises from the limitations of single-camera methods, especially noticeable as the number of feature points decreases. We structure experiments into three groups based on reduced feature counts and compared monocular and multi-camera detections. Results in Table 1 show that the success rate of loop closure detections diminishes in correlation with the decrease in feature points. This effect is markedly pronounced when the count of feature points falls below 100. Hence, multi-camera loop closure detection is essential for multi-camera VINS, ensuring reliability when individual cameras capture fewer features.

Table 1: Comparison of the number of successful loop closure detection by monocular and multi-camera Algorithms.

Dataset		Number of feature points		
		50	100	150
ZJG	Mono	113	554	706
	MulCam	498	878	954
NC	Mono	6	71	140
	MulCam	151	186	213

"Mono" means using a single camera for loop closure detection.
"MulCam" means using multiple cameras for loop closure detection.

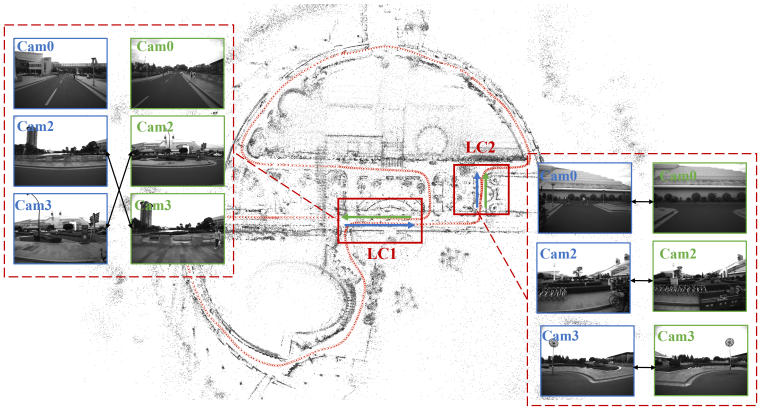


Figure 1: Visualization results of loop-closure observations in ZJG dataset.

To further illustrate the advantages of multi-camera closure over monocular closure, we visualize some loop-closure results using a scenario from the ZJG dataset as an example, as shown in Fig. 1. The blue and green arrows indicate the actual orientation of the robot at the two moments when the loop closure (denoted as LC1 and LC2) were detected. The pictures in the boxes of different colors correspond to the images observed at the moment of the arrows of different colors respectively. Cam0, Cam1 and Cam2 means respectively stereo left, around left and around right camera. We can see that when the robots are oriented in the same direction in LC2, the loop-closure can be detected by both the monocular camera in the forward view (cam0) and the surround view camera (cam2 and cam3). However, when encountering a situation like the one in LC1 where the robots are facing the opposite direction, the front-view camera (cam0) with a large difference in viewing angle cannot detect the loop-closure, while the left and right side of the surround view cameras can detect the loop-closure. This explains the large increase in the success rate of the multi-camera compared to the mono-camera loop closure approach.

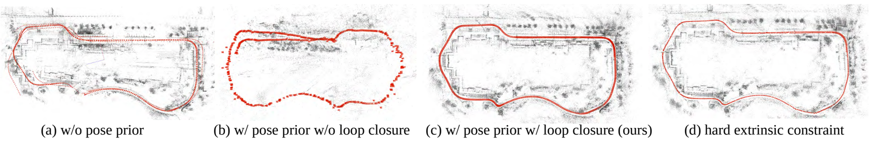

(a) w/o pose prior
(b) w/ pose prior w/o loop closure
(c) w/ pose prior w/ loop closure (ours)
(d) hard extrinsic constraint

Figure 2: Initial map reconstruction results. (a) generate initial map without using pose prior. (b) and (c) use different pose priors in the matching process. (d) use extrinsic parameters as hard constraints.

Multi-camera VINS pose prior. To assess the impact of pose priors on initial map construction, we utilize the qsdjt scene from the ZJG dataset as an example. We first implement a pose-independent method, sequential matching, which involves matching consecutively captured images and conducting loop closure detection, to generate an initial map without pose prior as shown in Fig. 2 (a). Then we operate multi-camera VINS both without (as shown in (b)) and with (as shwon in (c)) loop closure to derive different pose priors. Subsequently, we apply a spatial matching method to pair each image with its closest spatial neighbors based on the given pose prior. The results indicate that a reliable pose prior significantly enhances feature matching, especially in commercial street scenes, resulting in a clearer 3D reconstruction in (c) compared to (a) and (b).

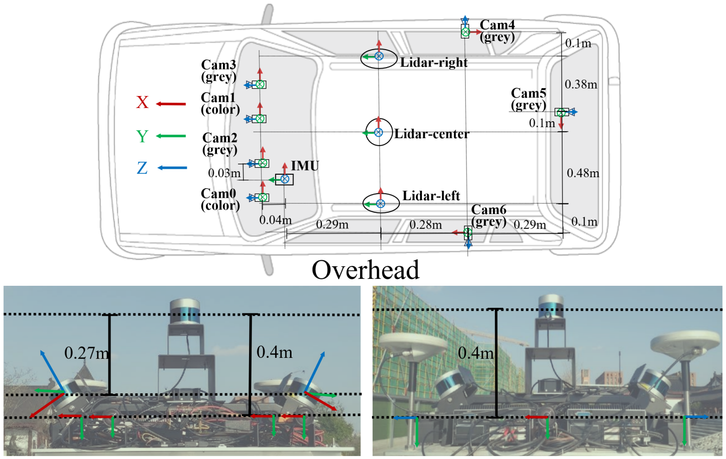
Additionally, since our proposed method optimizes the extrinsic parameters alongside the system as soft constraints, we also incorporate extrinsic parameters as hard constraints to refine camera poses at each moment as comparison. The results are shown in (d), which reflects that using extrinsic parameters as a hard constraint introduced more noise into the map compared to the result in (c).

2 Evaluation of multi-camera configuration

We first verify the effect of multi-camera configuration on the localization algorithm. We run the algorithm for localization in different scenarios and increase the number of cameras used for localization from 1 to 4, and the results obtained are shown in Table 2.

Table 2: The translation part (m) of map-based trajectory error under different number of camera observations.

	1 cam	2 cam	3 cam	4 cam
ZJG-qsdjt-1112	6.77	6.76	4.85	4.28
ZJG-yyl-0304	6.5	6.45	4.18	3.62
NC-math-hard	0.56	0.54	0.64	0.66
NC-quad-hard	0.44	0.49	0.45	0.26
Average	3.57	3.56	2.53	2.21



Overhead

Front

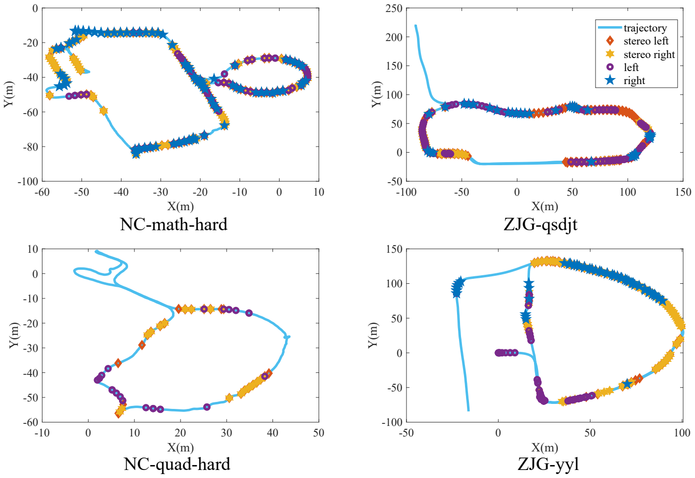
Rear

Figure 3: Multi-sensor data collection vehicle.

As shown in Fig.3 ’1cam’ means use cam2 observation only; ’2cam’ means cam2 and cam3 observation; ’3cam’ means use cam2; cam3 and cam6 observation, ’4cam’ means use cam2, cam3, cam4 and cam6 observation. Since the observation areas of cam2 and cam3 have a large overlap, so the localization observations obtained are basically the same, and the accuracy of the localization results is not much improved when the number of cameras is changed from 1 to 2. And when the number of cameras changes from 2 to 3 and 4, the localization accuracy varies more because the observation difference between cam4 and cam6 is large. But the variation of this variation varies on different datasets.

To further illustrate this, we plot the localization observations of different cameras (the number of feature points on a match greater than 40 is considered to be a valid observation) on each dataset on a trajectory, as shown in Fig. 4.

On the ZJG dataset, since the number of observations from the two binocular cameras is much smaller than the number of observations from the left and right cameras, there is a large improvement in localization accuracy when the number of cameras is increased to 3 and 4. On the NC dataset, the number of localization observations of the



NC-math-hard

ZJG-qsdjt

NC-quad-hard

ZJG-yyl

Figure 4: Visualization results of localization observations from different datasets.

two binocular cameras is approximately the same as that of the left and right cameras, so there is not much improvement in localization accuracy when the number of cameras is increased from 2 to 3 and 4. This shows that multi-camera localization observation has a significant effect on the improvement of localization accuracy.