Supplemental Material (3): Visual Localization Application

Chi Xu, Tingrui Guo, Yuan Huang, Li Cheng, Senior Member, IEEE

Abstract

It is the supplemental material of paper "A Hough Voting based 2-Point RANSAC Solution to the Perspective-n-Point Problem". Detailed information of the visual localization application is provided in this document.

1 Overview

We evaluate our solution in a visual localization application. A total of 1751 images of the star-river complex (a landmark of Wuhan city) are obtained, including 812 reference images and 939 query images; some images are collected by cameras from the field, and some images are retrieved from the Internet. The longer side (width or height) of the image is uniformly resized to 1080 pixels. The SfM model reconstructed by COLMAP are shown in Fig. 1.

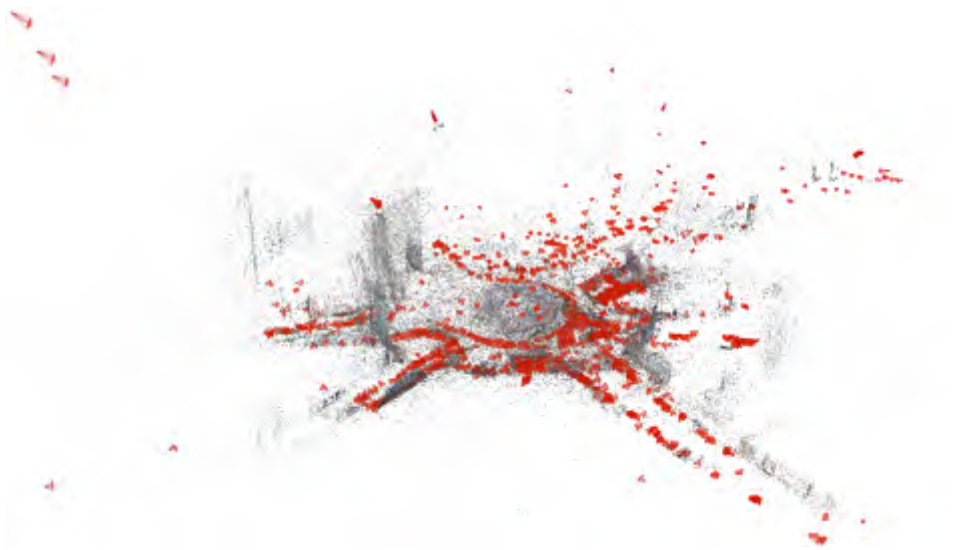
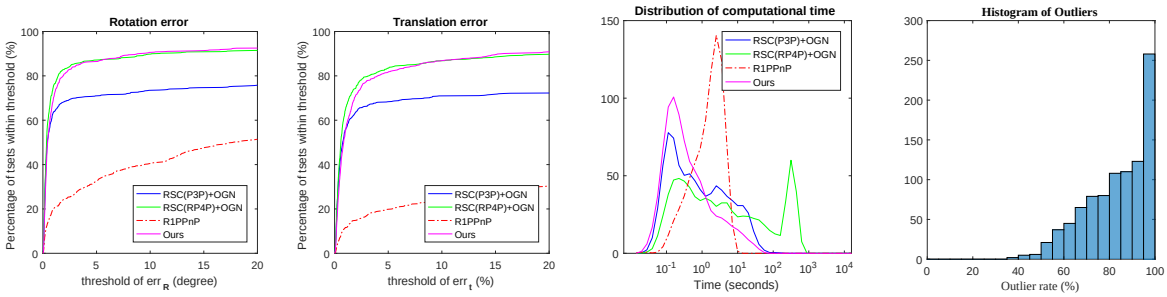


Figure 1: SfM reconstruction of star-river complex using COLMAP. The black point cloud represents the reconstructed 3D model points, and the red cones represent the cameras.

2 Experimental results

Deep learning based methods are used in this experiment for feature extracting and matching. Firstly, given a query image, EigenPlaces [1] (viewpoint robust image retrieval method) is used to retrieve top-k images from the database as references; k is set as 10 in this application. And then, SuperPoint+LightGlue [2] (which is widely used in the community) is employed for feature extracting and matching. The 2D-3D correspondences retrieved from the top-k image pairs are aggregated and fed into the PnP solvers for camera pose estimation. Compared to estimating camera pose using 2D-3D correspondences of individual image pairs, it is more accurate and efficient to aggregate the correspondences of top-k image pairs together [3, 4] for pose estimation, because by aggregation we can obtain more inliers which enhance the accuracy, and the PnP solver can be efficiently executed only once instead of k times. The experimental results are shown in Fig. 2(a)-(c).



(a) (b) (c) (d)

Figure 2: Experimental results of star-river application.

In the experiment, there exist many challenging factors (such as occlusions, snow & rain, motion blur, and reflections from window) which significantly affect the correctness of feature matching, and the histogram of the outlier rate is illustrated in Fig. 2(d). We investigate the relationship between the parameter k and the outlier rate, as can be seen in Table 1. When $k = 1$, the percentage of query images with outliers more than 95% is 32.8%, and the average number of inliers is only 1.3. As k increases from 1 to 10, the percentage of query images with high outlier rates (i.e. >95%) first drops slightly and then increases slightly. When $k = 10$, the percentage of query images with outliers more than 95% is 27.5%, and the average number of inliers reaches 22.5. Overall, an increase in k does not significantly affect the percentage of high outlier rates, but effectively increase the number of inliers.

Table 1: The relationship between k and outlier rate.

top-k	outlier>90%		outlier>95%	
	percentage	nInlier	percentage	nInlier
1	37.2%	2.4	32.8%	1.3
2	33.6%	6.4	28.3%	3.2
3	32.5%	10.8	25.8%	4.8
4	31.8%	16.2	24.6%	7.3
5	33.7%	22.0	25.0%	9.5
6	34.1%	27.6	25.7%	11.9
7	35.1%	33.6	26.0%	14.5
8	37.4%	41.9	26.0%	16.1
9	39.1%	49.2	26.3%	18.8
10	40.5%	54.4	27.5%	22.5

The number of extracted feature points is closely related to the resolution of the image. As the resolution of modern image acquisition equipment increases, the number of feature points extracted increases accordingly. Taking a pair of images with a width of 1080 pixels as an example, the number of matched point pairs is about 1000˜2000, and more point pairs can be obtained by aggregating the results from the top-k images. Therefore, the number of inliers is normally sufficient for the PnP solution in case of high outlier rate.

References

[1] G. Berton, G. Trivigno, B. Caputo, and C. Masone, "Eigenplaces: Training viewpoint robust models for visual place recognition," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 11 080–11 090.

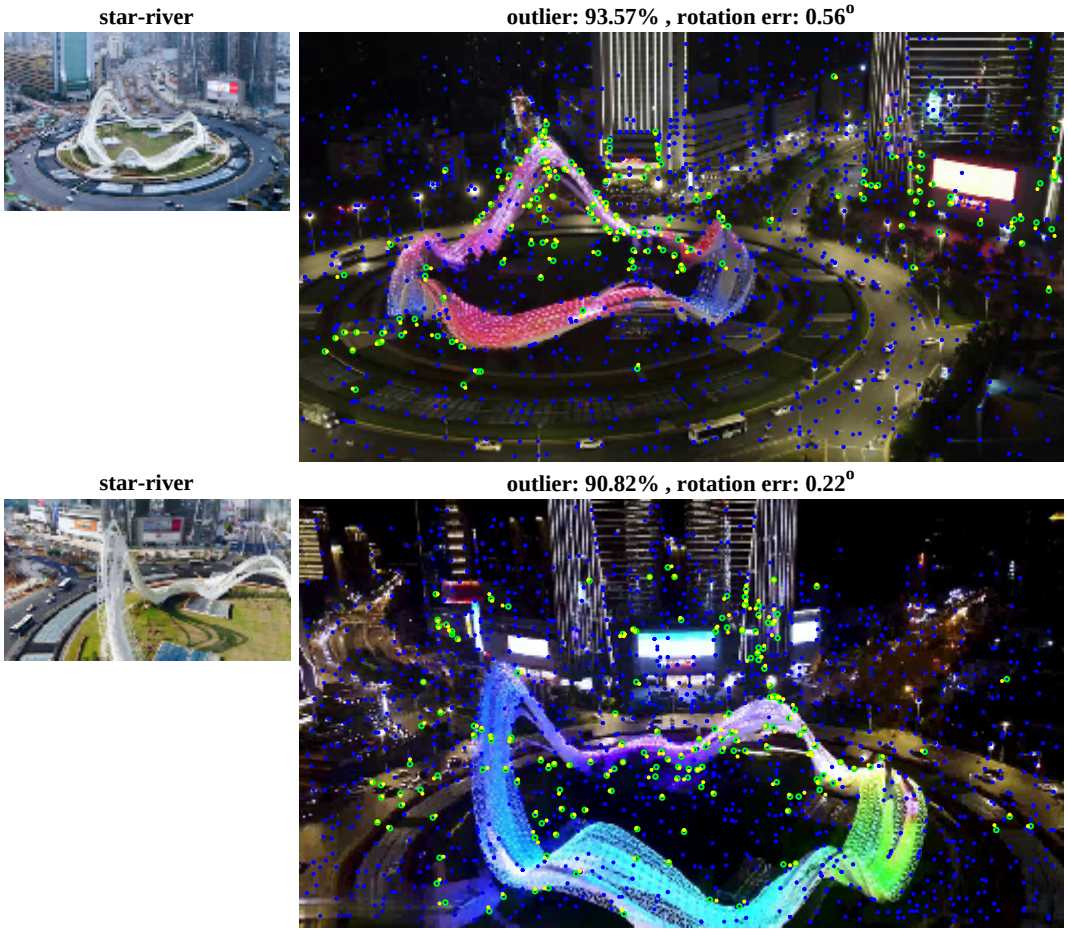
[2] P. Lindenberger, P.-E. Sarlin, and M. Pollefeys, "LightGlue: Local Feature Matching at Light Speed," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023.

[3] P.-E. Sarlin, F. Debraine, M. Dymczyk, R. Siegwart, and C. Cadena, "Leveraging deep visual descriptors for hierarchical efficient localization," in *Conference on Robot Learning (CoRL)*, 2018.

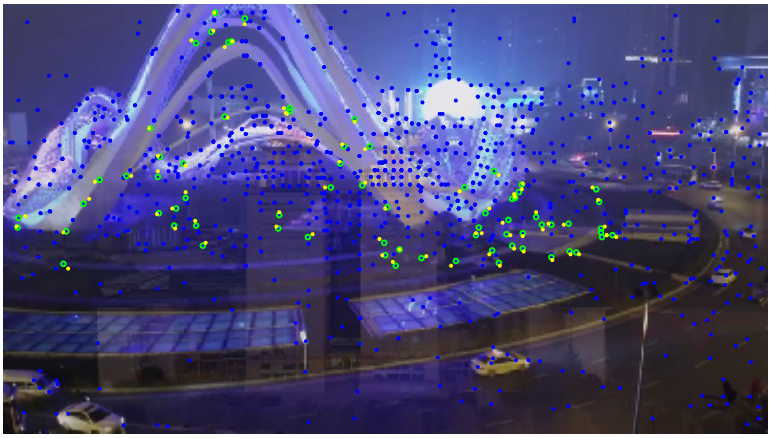
[4] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, "From coarse to fine: Robust hierarchical localization at large scale," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 12 716–12 725.

Visual Examples

Exemplar pose estimation results of our solution are provided in this section. top-k images are retrieved by EigenPlaces for pose estimation. In each figure, on the left shows the top-1 matched reference image, and on the right is the query image. Outlier rate and estimated rotation error of our solution are shown on top of each target image. **Blue dots** denote 2D points matched, **green dots** denote inliers detected, and **yellow dots** denote re-projection of the inliers using the estimated pose.

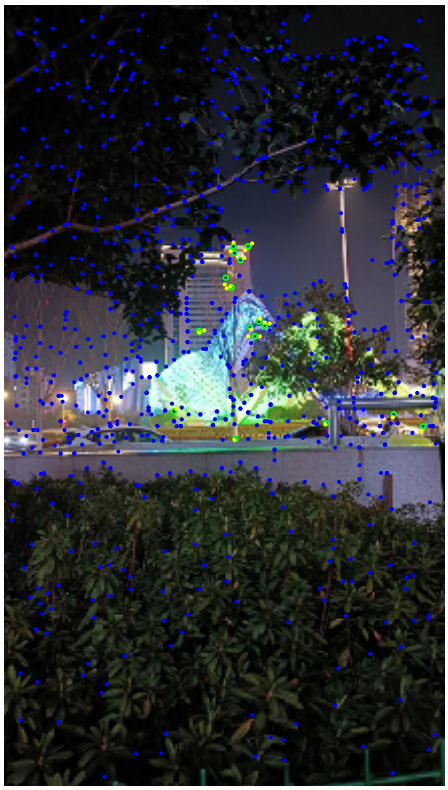


star-river **outlier: 95.51% , rotation err: 9.80^{o}**



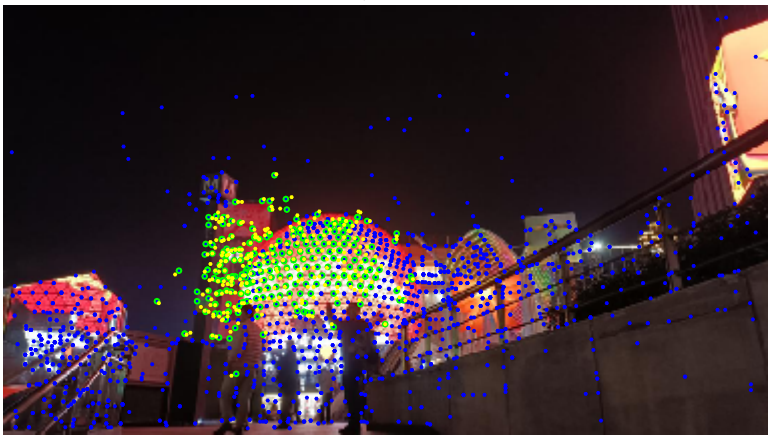
star-river **outlier: 94.00% , rotation err: 0.40^{o}**

star-river **outlier: 98.44% , rotation err: 2.78^{o}**



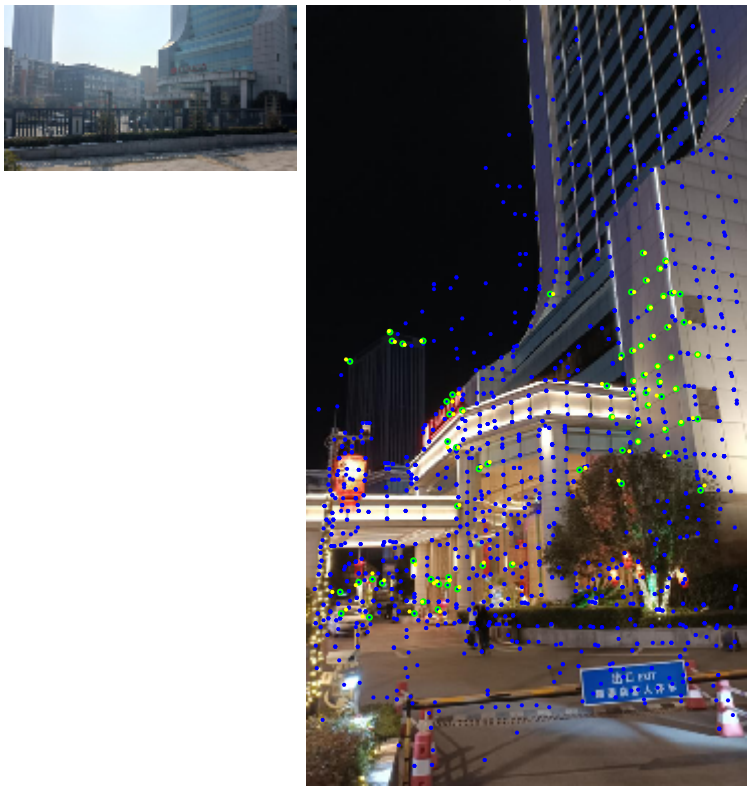
star-river **outlier: 67.65% , rotation err: 0.65^{o}**

star-river **outlier: 90.46% , rotation err: 11.35^{o}**

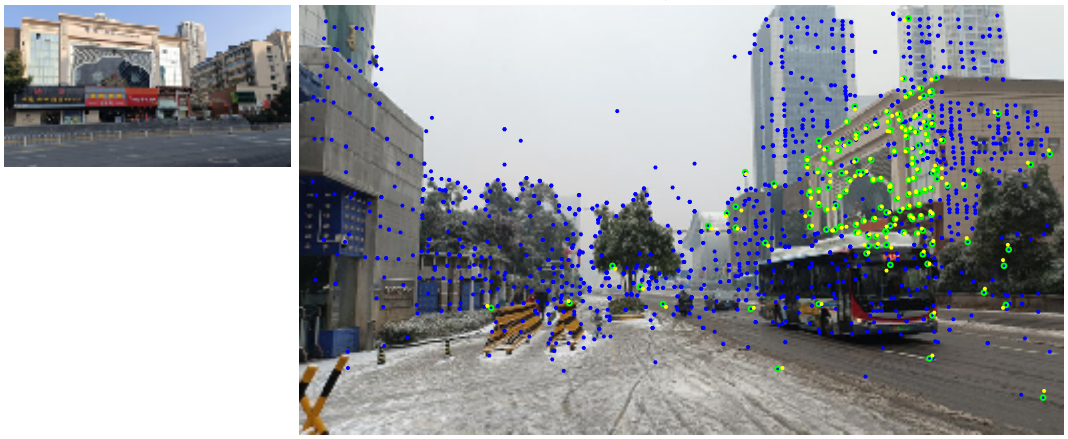


star-river **outlier: 84.12% , rotation err: 0.11^{o}**

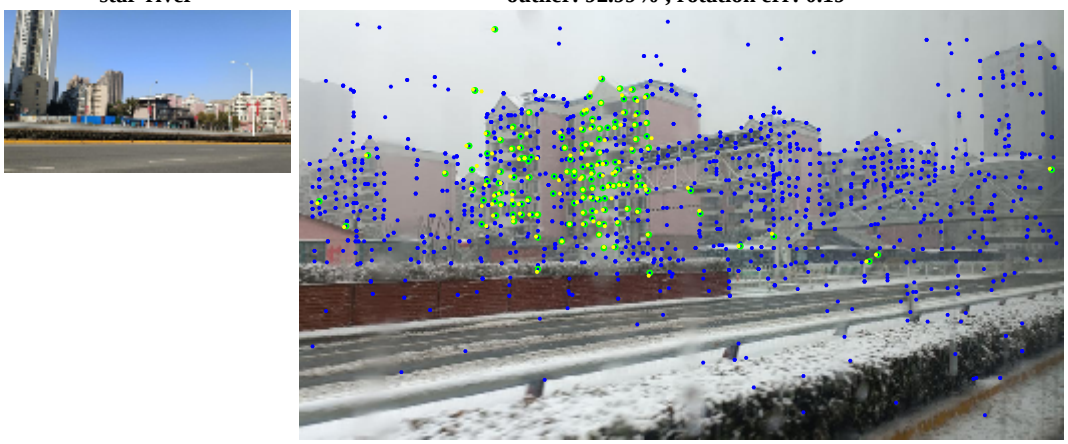
star-river **outlier: 96.76% , rotation err: 1.38^{o}**



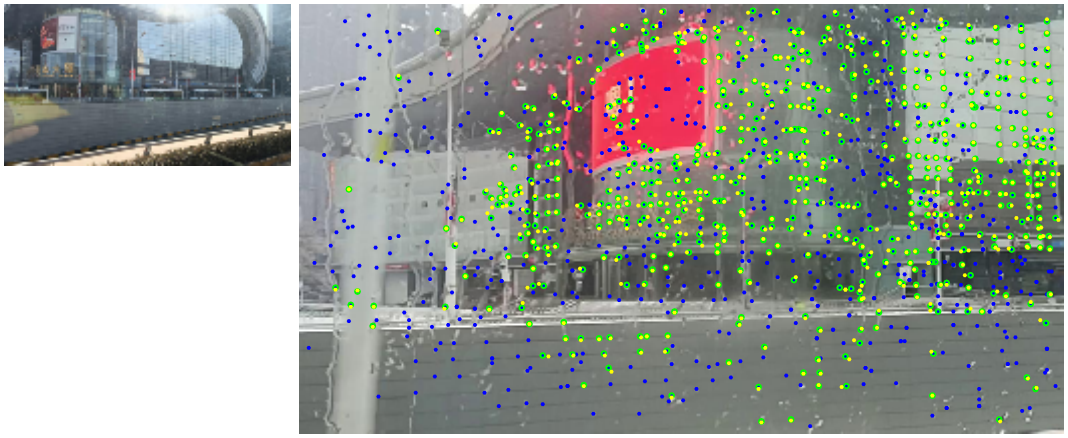
star-river **outlier: 91.85% , rotation err: 0.99^{o}**



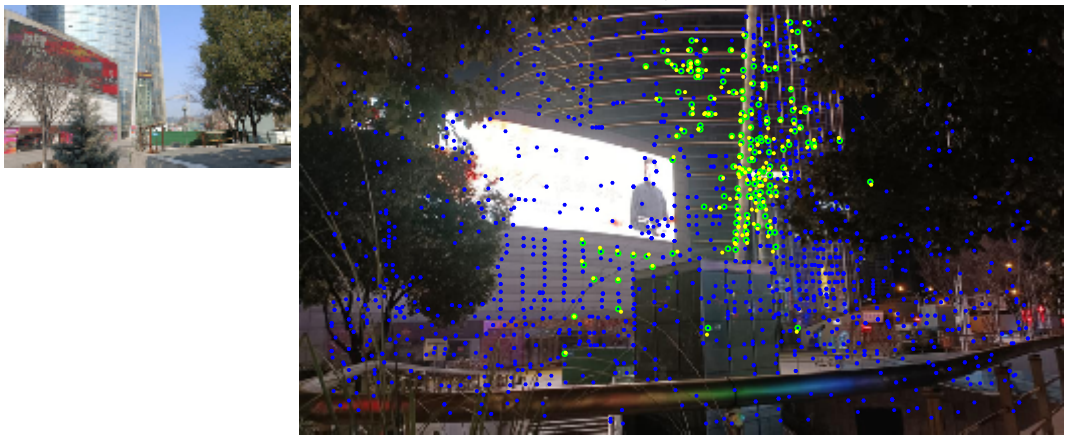
star-river **outlier: 92.99% , rotation err: 0.19^{o}**



star-river **outlier: 60.52% , rotation err: 0.08^{o}**



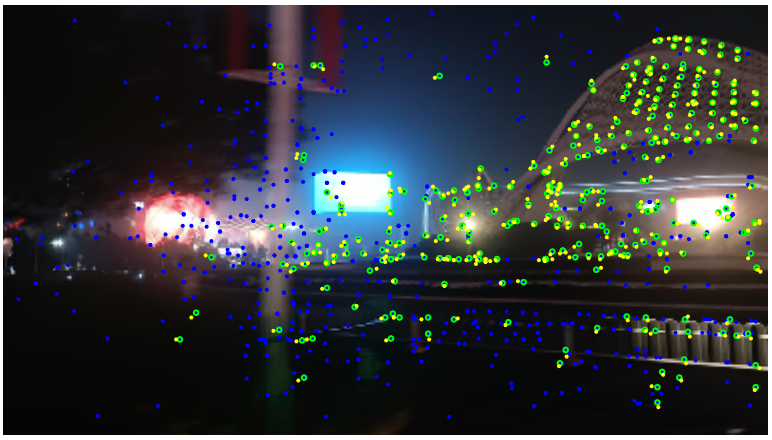
star-river **outlier: 92.09% , rotation err: 5.97^{o}**



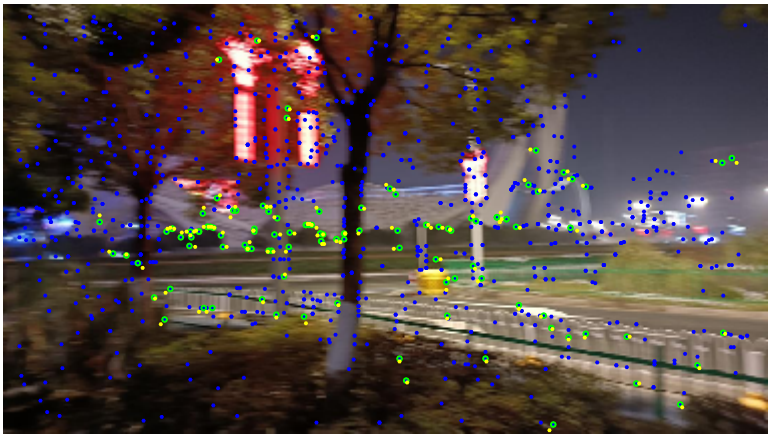
star-river **outlier: 86.71% , rotation err: 3.80^{o}**

star-river **outlier: 94.84% , rotation err: 1.55^{o}**

star-river **outlier: 75.61% , rotation err: 0.12^{o}**

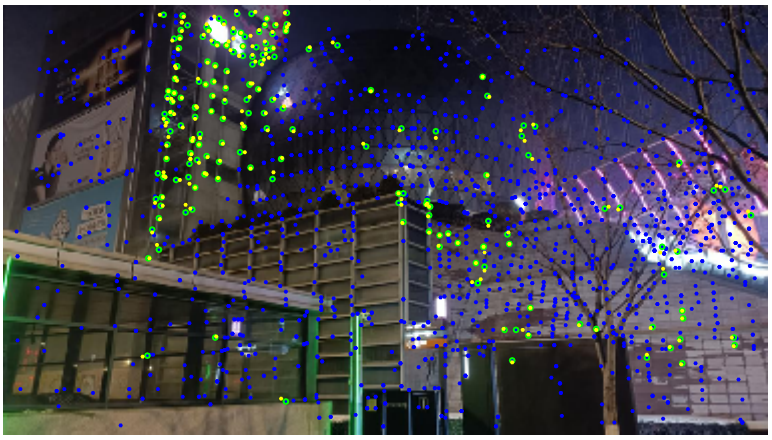


star-river **outlier: 92.39% , rotation err: 0.33^{o}**

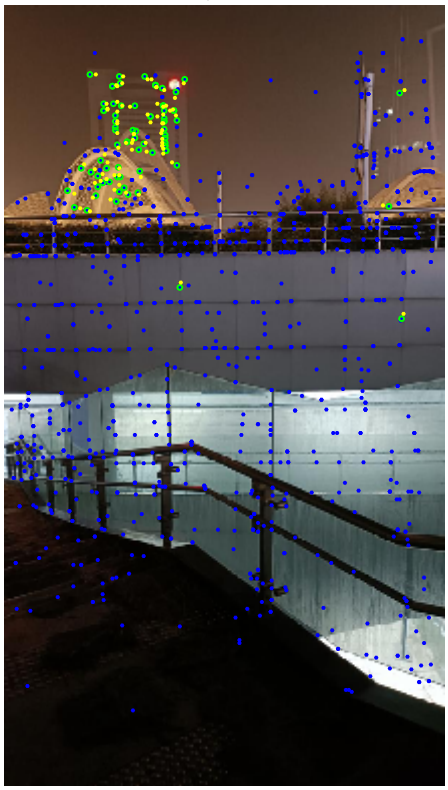


star-river **outlier: 93.74% , rotation err: 0.88^o**

star-river **outlier: 93.21% , rotation err: 0.11^o**

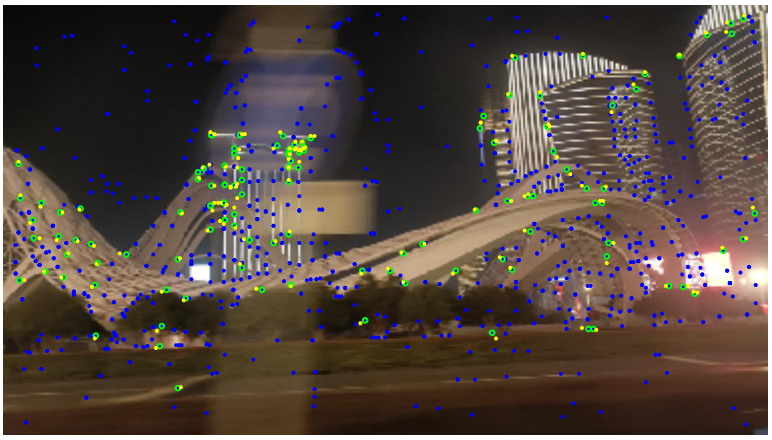


star-river **outlier: 95.46% , rotation err: 3.02^o**

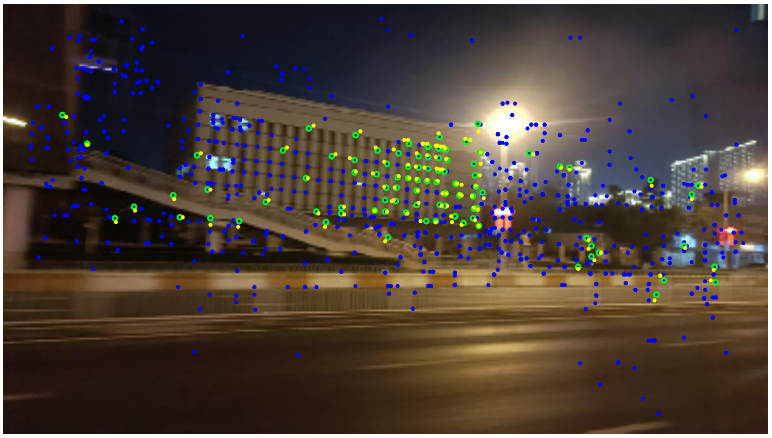


star-river **outlier: 90.51% , rotation err: 0.11^{o}**

star-river **outlier: 91.82% , rotation err: 0.37^{o}**

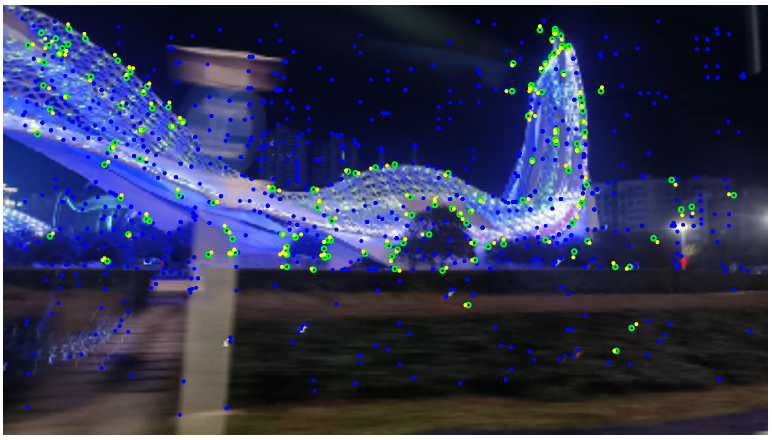


star-river **outlier: 92.03% , rotation err: 0.97^{o}**



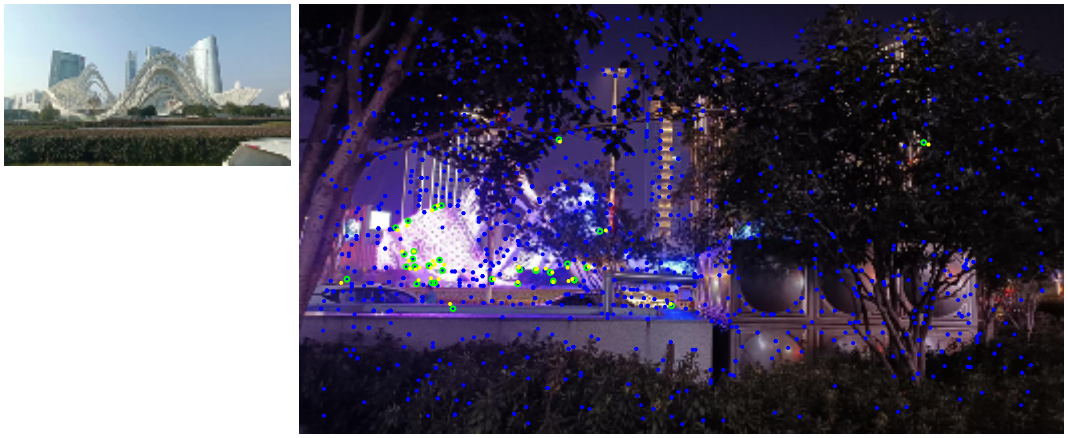
star-river **outlier: 97.98% , rotation err: 1.34^{o}**

star-river **outlier: 91.14% , rotation err: 0.63^{o}**

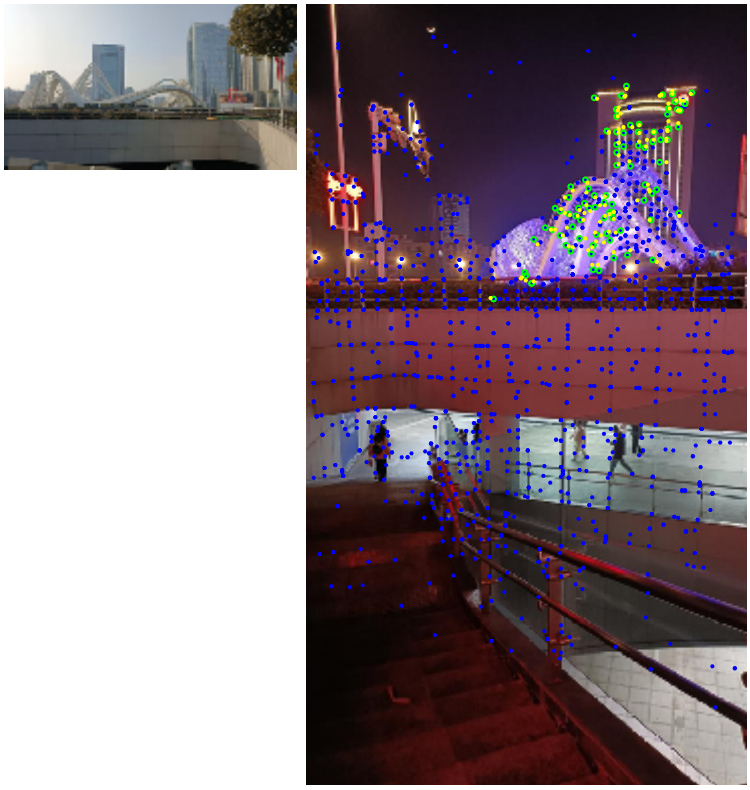


star-river **outlier: 90.60% , rotation err: 0.22^{o}**

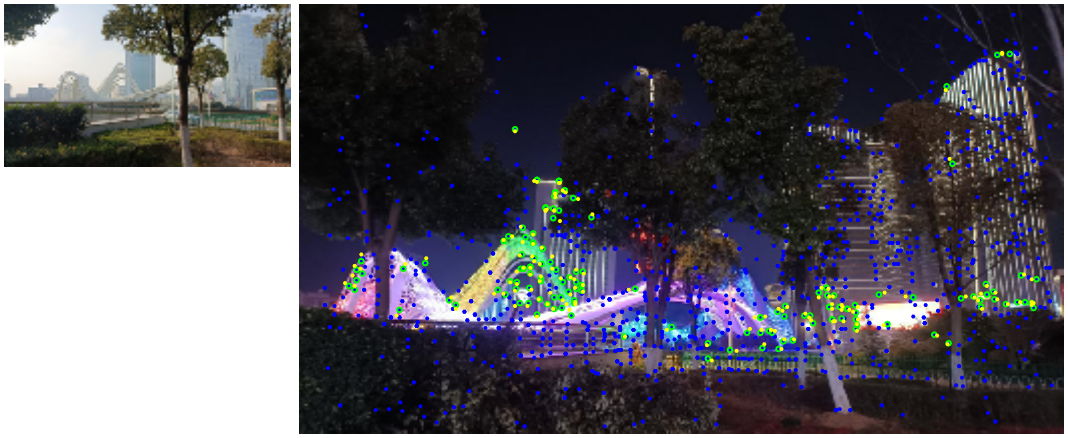
star-river **outlier: 98.02% , rotation err: 0.97°**



star-river **outlier: 95.68% , rotation err: 0.60°**

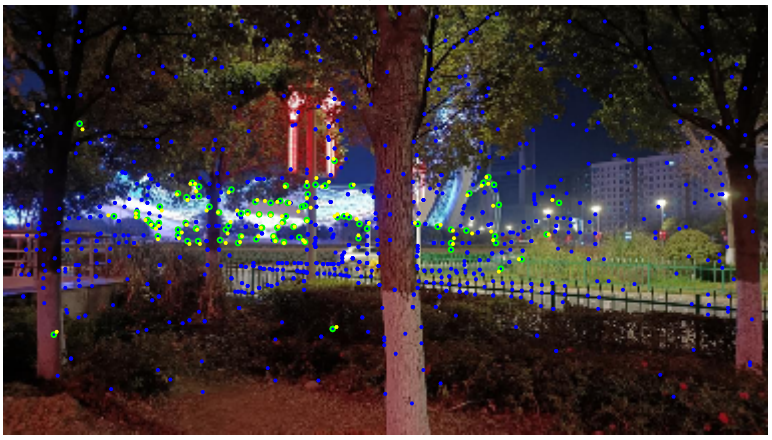


star-river **outlier: 91.87% , rotation err: 0.52°**

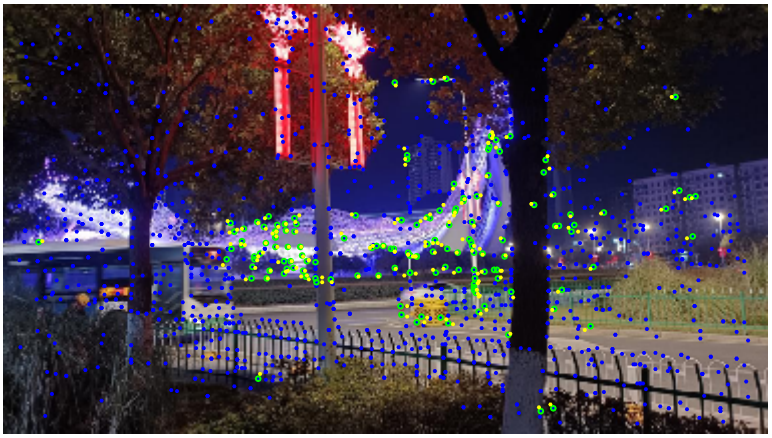


star-river **outlier: 93.75% , rotation err: 1.00^{o}**

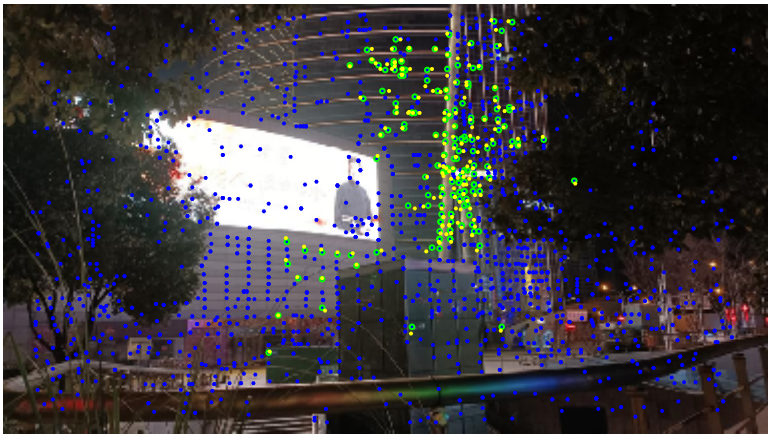
star-river **outlier: 93.70% , rotation err: 1.33^{o}**



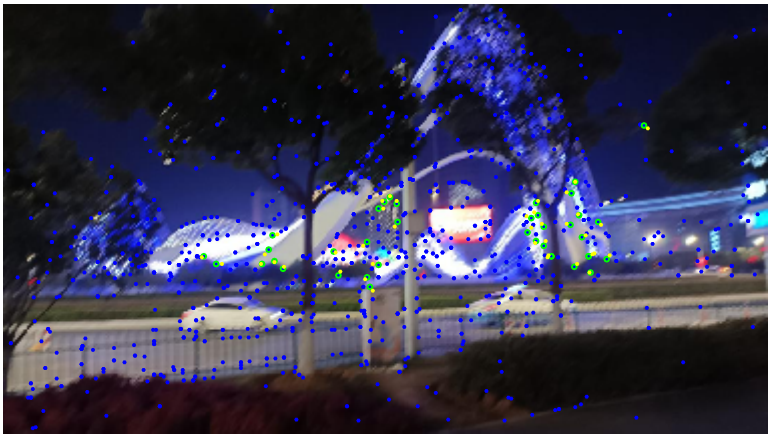
star-river **outlier: 92.13% , rotation err: 0.42^{o}**



star-river **outlier: 92.09% , rotation err: 5.97^{o}**

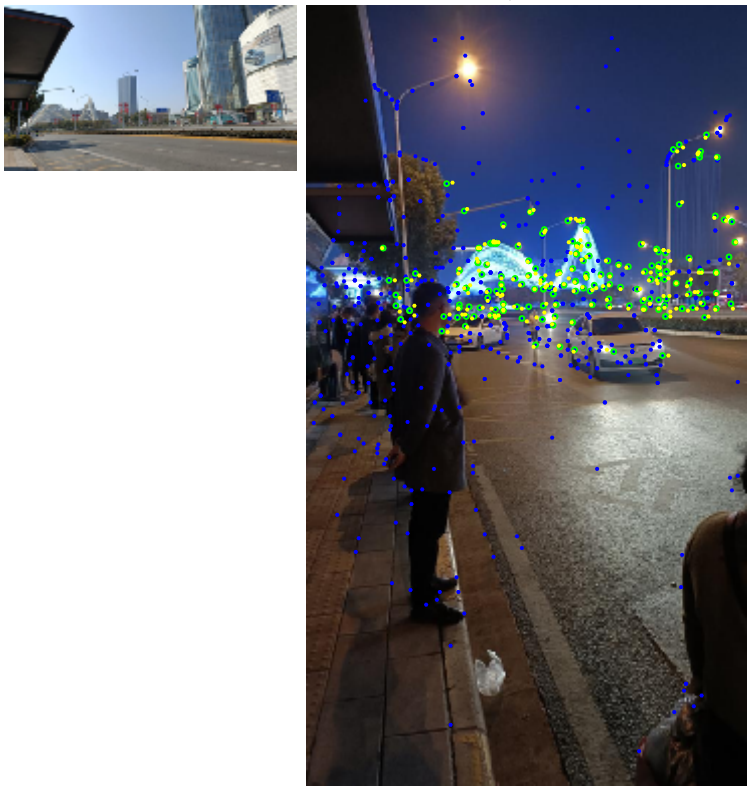


star-river **outlier: 96.99% , rotation err: 7.38^{o}**



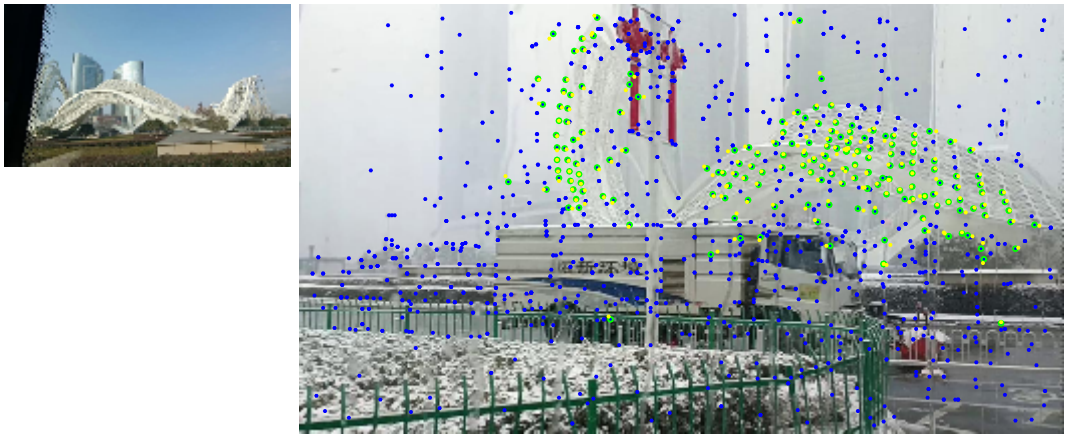
star-river **outlier: 88.72% , rotation err: 1.54^{o}**

star-river **outlier: 84.08% , rotation err: 0.40°**

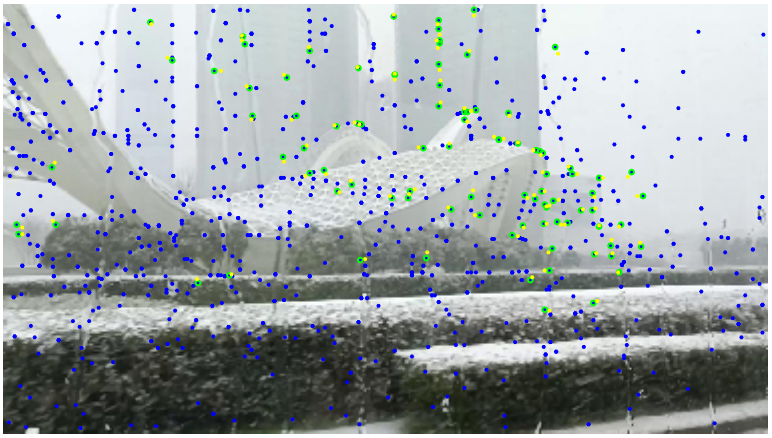


star-river **outlier: 98.24% , rotation err: 5.53°**

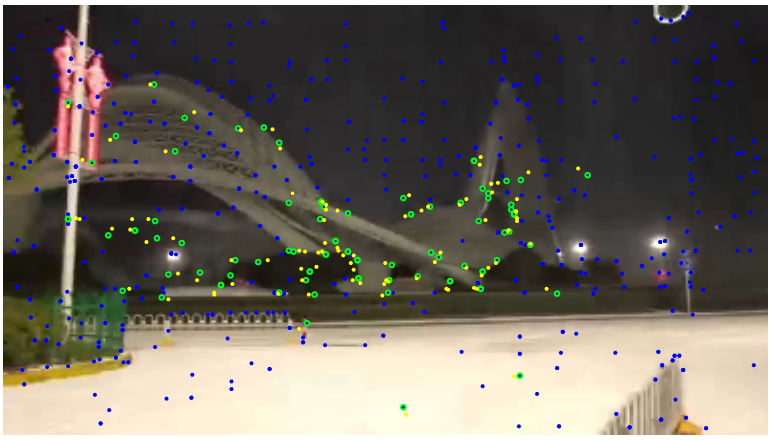
star-river **outlier: 89.50% , rotation err: 0.97°**



star-river **outlier: 92.97% , rotation err: 0.80^{o}**



star-river **outlier: 90.63% , rotation err: 1.44^{o}**



star-river **outlier: 93.90% , rotation err: 3.10^{o}**