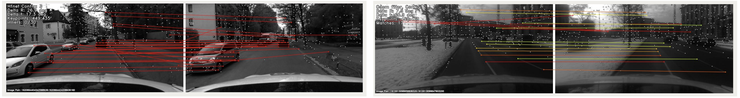
# AdaSfM: From Coarse Global to Fine Incremental Adaptive Structure from Motion

Yu Chen[1] Zihao Yu[2] Shu Song[3] Tianning Yu[4] Jianming Li[4] Gim Hee Lee[1]
[1]National University of Singapore [2]ByteDance [3]Nreal [4]Segway-Navimow

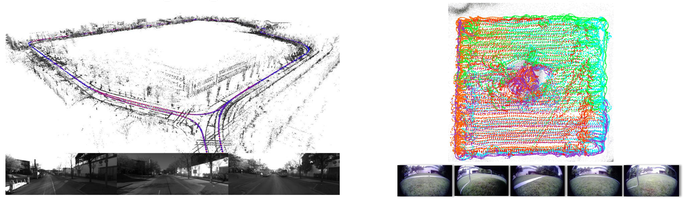
Project Page

## Challenges in Visual Reconstruction

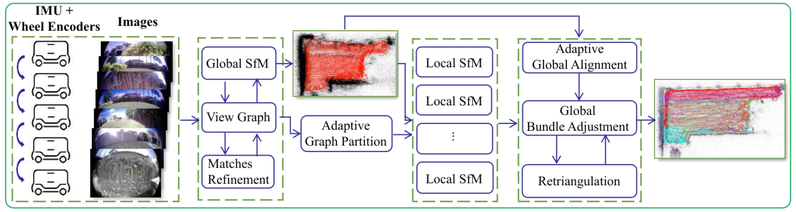
- Wrong two-view geometries



- Reconstructing large-scale scenes is time-consuming



## System Pipeline of AdaSfM



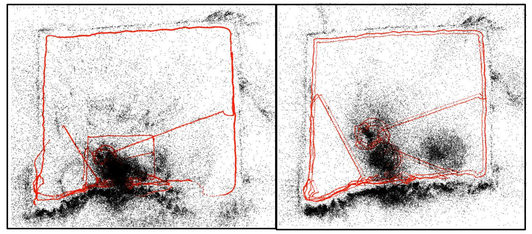

## Challenges of Global SfM

- Wrong two-view geometries (edge outliers)
- Scale ambiguities of translation averaging
- Determine a solvable view graph is time-consuming
  - Involve solving a system of polynomial equations with a large number of unknowns
  - The practical solvabilitytesting up to minimal graphs with up to 90 nodes

## Global SfM with Augmented View Graph

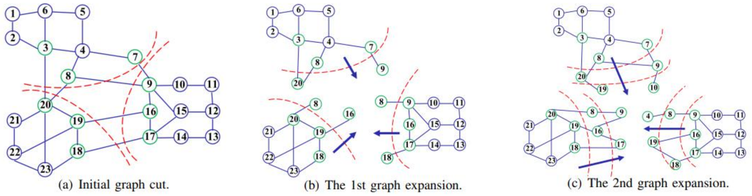
- Adding extra edges (from consecutive IMU and wheel encoder data) can link nodes to the maximally parallel rigid components
- Filter wrong relative translations by 1DSfM
- Replace the potentially wrong two view geometries
- Solving the RA problem & TA problem within the augmented view graph

**Left:** Scale ambiguities hinder the loop-closure of global SfM



**Right:** Augmented view graph fixed the scale ambiguities issue

## Adaptive Graph Partition



(a) Initial graph cut. (b) The 1st graph expansion. (c) The 2nd graph expansion.

## Local Incremental SfM

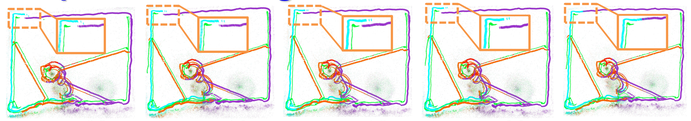
- Image Registration
  - Perspective-n-point (i.e. P3P):
  - Single Rotation Averaging + Median($\{\mathbf{t}_i\}$)
  - The best camera pose for current registered image is selected by reprojecting 3D point to image plane and then checking the pixel residuals
- Global Refinement

$$\min \sum_i \sum_k \boxed{\|\mathbf{\Pi}(\mathbf{P}_i, \mathbf{X}_k) - \mathbf{u}_{ik}\|} + \boxed{\sum_{(i,j)\in\mathcal{E}_{\text{aug}}} \|\log(\mathbf{R}_j \mathbf{R}_i^T, \hat{\mathbf{R}}_j \hat{\mathbf{R}}_i^T)\|} + \boxed{\sum_{(i,j)\in\mathcal{E}_{\text{aug}}} d_\angle(\mathbf{C}_i - \mathbf{C}_j, \hat{\mathbf{C}}_i - \hat{\mathbf{C}}_j)}$$

**Traditional BA term**

**camera pose regularization term**

## Adaptive Global Alignment



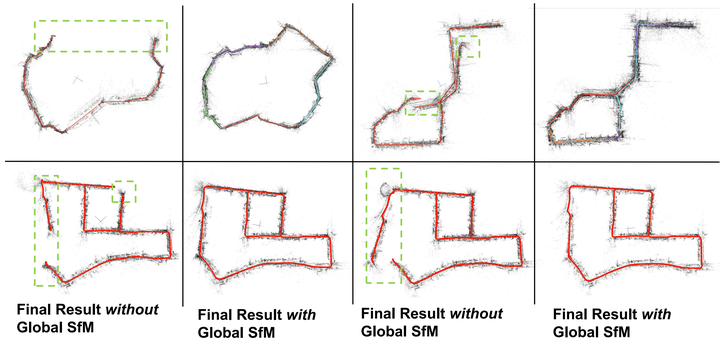
(a) $\tau_{\text{inlier}} = 0.5$ (b) $\tau_{\text{inlier}} = 1.0$ (c) $\tau_{\text{inlier}} = 1.5$ (d) $\tau_{\text{inlier}} = 2.0$ (e) $\tau_{\text{init}} = 1.0$

**(a)-(d)** are alignment results **by using different fixed inlier threshold within RANSAC**; **(e)** is the result **with our adaptive global alignment algorithm** with an initial inlier threshold 1.0.

## Experiments

| Scene | Sequence | COLMAP [8] | | | | | Ours (Global SfM) | | | | | Ours (final) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $N_c$ | $N_p$ | $\Delta\mathbf{R}$ | $\Delta\mathbf{t}$ | $T$ | $N_c$ | $N_p$ | $\Delta\mathbf{R}$ | $\Delta\mathbf{t}$ | $T$ | $N_c$ | $N_p$ | $\Delta\mathbf{R}$ | $\Delta\mathbf{t}$ | $T$ |
| Neighborhood | recording_2020-10-07_14-53-52 | 6,326 | 137,135 | **0.65** | 1.78 | 334.90 | 6,036 | 66,777 | 2.52 | 1.17 | 14.68 | 6,033 | 109,483 | 0.74 | **0.52** | 123.96 |
| | recording_2020-12-22_11-54-24 | 6,518 | 127,892 | 0.55 | 3.68 | 354.35 | 6,144 | 64,405 | 1.10 | 0.86 | 15.83 | 6,144 | 102,857 | **0.51** | **0.62** | 151.88 |
| | recording_2020-03-26_13-32-55 | 7,414 | 148,848 | **0.61** | 1.24 | 603.13 | 5,982 | 70,066 | 0.92 | **0.79** | 17.10 | 5,982 | 111,807 | 1.11 | 0.98 | 157.76 |
| | recording_2020-10-07_14-47-51 | 6,688 | 152,307 | **0.56** | 1.67 | 359.03 | 6,248 | 76,305 | 2.20 | 1.17 | 15.70 | 6,248 | 121,657 | 0.75 | **0.74** | 152.85 |
| | recording_2021-02-25_13-25-15 | 6,174 | 138,807 | 0.75 | 1.05 | 325.65 | 5,238 | 62,879 | 1.00 | 1.14 | 15.12 | 5,238 | 106,609 | **0.46** | **0.81** | 202.85 |
| | recording_2021-05-10_18-02-12 | 7,784 | 149,528 | 3.04 | 9.57 | 444.85 | 5,834 | 61,889 | 1.49 | 1.38 | 12.76 | 5,834 | 101,102 | **0.47** | **0.59** | 153.36 |
| | recording_2021-05-10_18-32-32 | 7,174 | 141,864 | 2.77 | 19.15 | 416.34 | 6,046 | 89,010 | **1.14** | **1.03** | 23.81 | 6,046 | 142,430 | 1.49 | 1.34 | 264.75 |
| Business Park | recording_2021-01-07_13-12-23 | 8,016 | 109,399 | 0.72 | 0.75 | 643.22 | 9,010 | 72,096 | 1.76 | 1.60 | 56.16 | 9,010 | 100,057 | **0.66** | **0.51** | 465.34 |
| | recording_2020-10-08_09-30-57 | 11,520 | 127,013 | **0.37** | 1.57 | 1284.44 | 8,278 | 66,087 | 1.59 | 1.51 | 48.72 | 8,278 | 108,000 | 0.63 | **0.45** | 366.81 |
| | recording_2021-02-25_14-16-43 | 7,414 | 148,848 | **0.61** | 1.24 | 603.13 | 5,982 | 70,066 | 0.92 | **0.79** | 17.10 | 5,982 | 111,807 | 1.11 | 0.98 | 157.76 |
| Old Town | recording_2020-10-08_11-53-41 | 19,332 | 279,989 | - | - | 2454 | 12,910 | 181,569 | 2.23 | 2.81 | 45.72 | 12,048 | 279,127 | **0.55** | **0.56** | 254.71 |
| | recording_2021-01-07_10-49-45 | 16,420 | 307,383 | 8.63 | 360.51 | 1496.6 | 12,728 | 194,340 | 2.56 | 3.14 | 53.18 | 12,728 | 327,348 | **1.55** | **1.03** | 238.82 |
| | recording_2021-02-25_12-34-08 | 18,950 | 305,461 | - | - | 2392.98 | 12,387 | 182,940 | 2.02 | 3.14 | 40.97 | 12,387 | 302,833 | **0.63** | **0.74** | 683.97 |
| Office Loop | recording_2020-03-24_17-36-22 | 10,188 | 209,942 | 1.17 | 3.40 | 822.38 | 9,522 | 126,680 | 2.28 | 2.38 | 31.87 | 9,377 | 214,285 | **0.97** | **0.98** | 166.54 |
| | recording_2020-03-24_17-45-31 | 8,582 | 195,738 | 0.92 | 3.04 | 865.48 | 9,186 | 122,713 | 2.79 | 2.20 | 33.91 | 8,940 | 205,790 | **0.84** | **0.85** | 209.06 |
| | recording_2020-04-07_10-20-31 | 10,350 | 223,649 | 4.22 | 42.44 | 795.68 | 10,184 | 138,446 | 2.53 | 1.78 | 39.83 | 10,184 | 224,499 | **1.47** | **1.14** | 253.24 |
| | recording_2020-06-12_10-10-57 | 9,990 | 236,593 | 18.97 | 83.94 | 705.93 | 10,150 | 164,062 | 1.92 | 1.61 | 37.32 | 10,150 | 246,516 | **0.76** | **0.87** | 206.48 |
| | recording_2021-01-07_12-04-03 | 9,164 | 475,950 | **0.71** | 2.58 | 1000.75 | 10,300 | 143,715 | 3.32 | 2.39 | 48.68 | 10,300 | 223,676 | 1.08 | **0.67** | 249.42 |
| | recording_2021-02-25_13-51-57 | 9,574 | 214,695 | **0.84** | 2.84 | 773.32 | 9,426 | 122,746 | 3.80 | 2.68 | 28.96 | 9,426 | 204,289 | 1.01 | **0.91** | 173.29 |

**Comparison of runtime and accuracy on the 4Seasons datasets**. T denotes the runtime (in minutes). $N_c$, $N_p$ denote the number of registered images and 3D points, respectively. R denotes the mean rotation error (in degrees) and translation error (in meters), respectively, and we highlight the best results in bold.



**Final Result *without* Global SfM** | **Final Result *with* Global SfM** | **Final Result *without* Global SfM** | **Final Result *with* Global SfM**