VIO-Aided Structure from Motion Under Challenging Environments

ICIT 2021

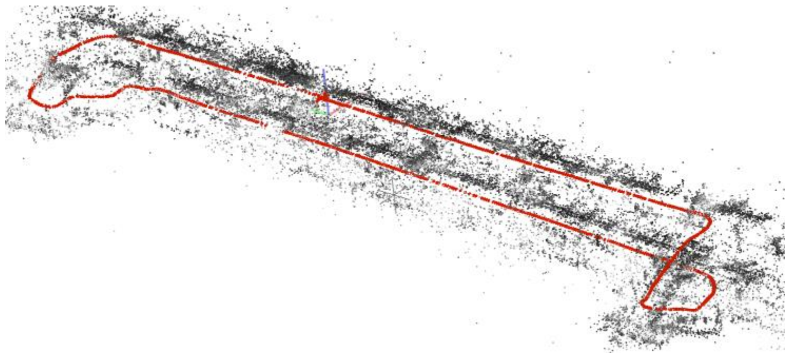
Author: *Zijie Jiang*[1], *Hajime Taira*[1], *Naoyuki Miyashita*[2], *Masatoshi Okutomi*[1]

[1]*Tokyo Institute of Technology* [2]*Olympus Coorporation*

3D Reconstruction from multiple images

Image Sequence

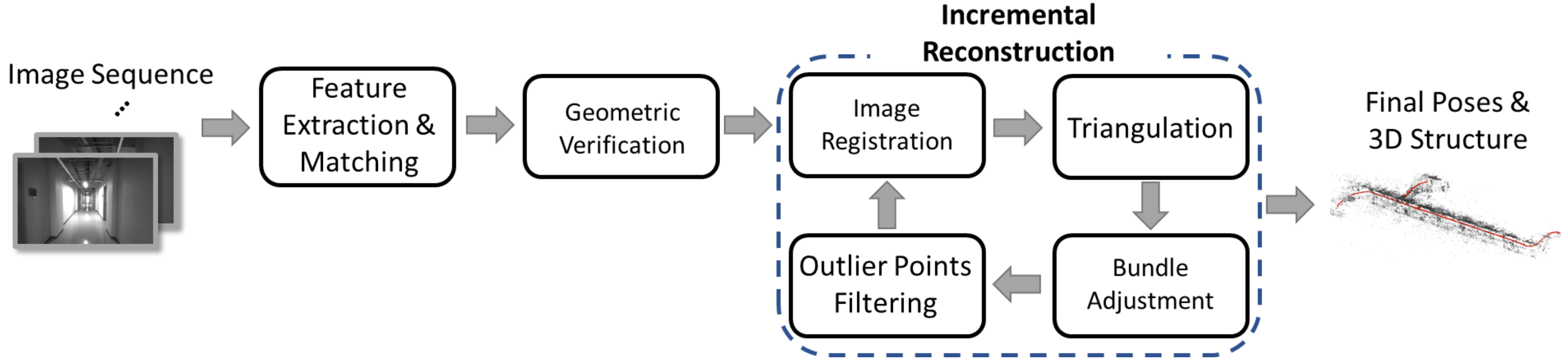
Camera Poses (red) &
3D Structure (black)



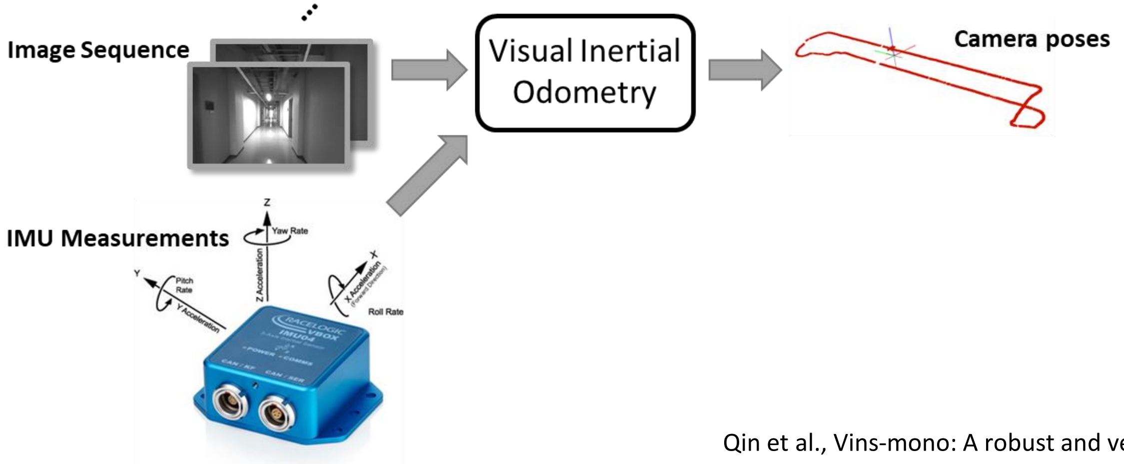
Recover 3D geometry and camera poses from sets of images of a common scene.

Related Work

- Structure from Motion (COLMAP, Schonberger et al., 2016)



- Visual-Inertial Odometry (VINS-Mono, Qin et al. 2018)



Schonberger et al., Structure-from-Motion Revisited. CVPR, 2016
Qin et al., Vins-mono: A robust and versatile monocular visual-inertial state estimator. IEEE Trans. Robot., 2018

Challenging Environments

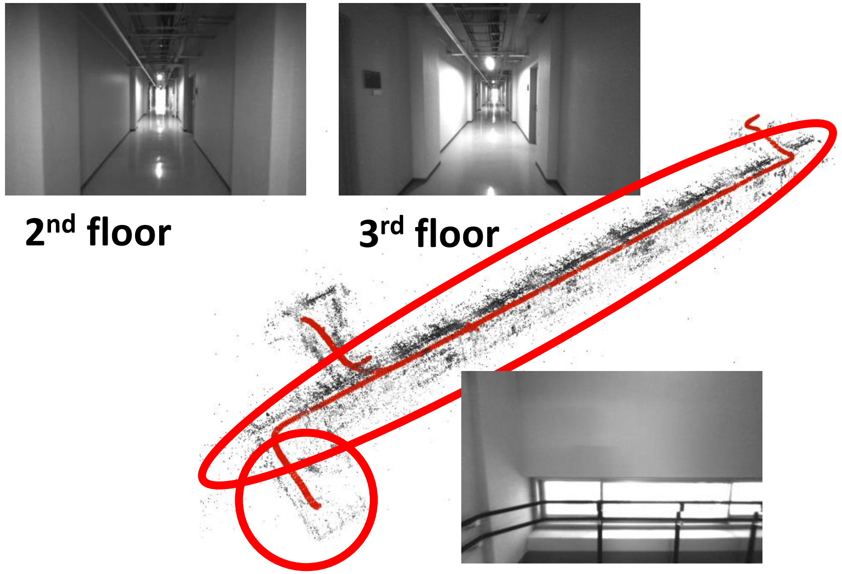
- Degradation of visual information

Texture-less area

Duplicated structure

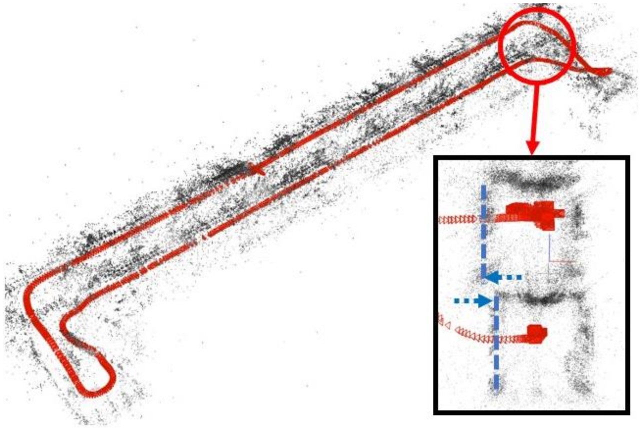
Challenging Environments

- Limitation of existing methods



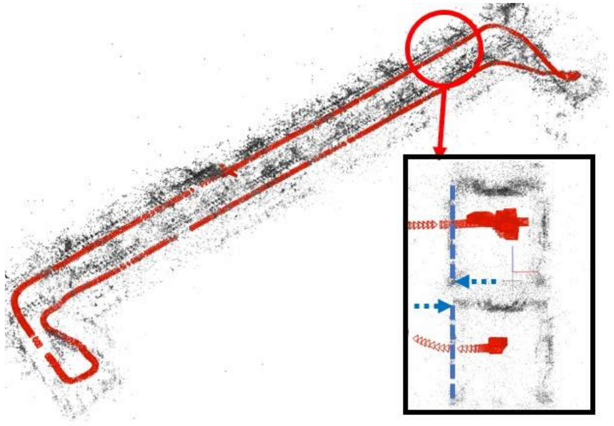
COLMAP

✗ **Robustness**



VINS-Mono

✗ **Accuracy**



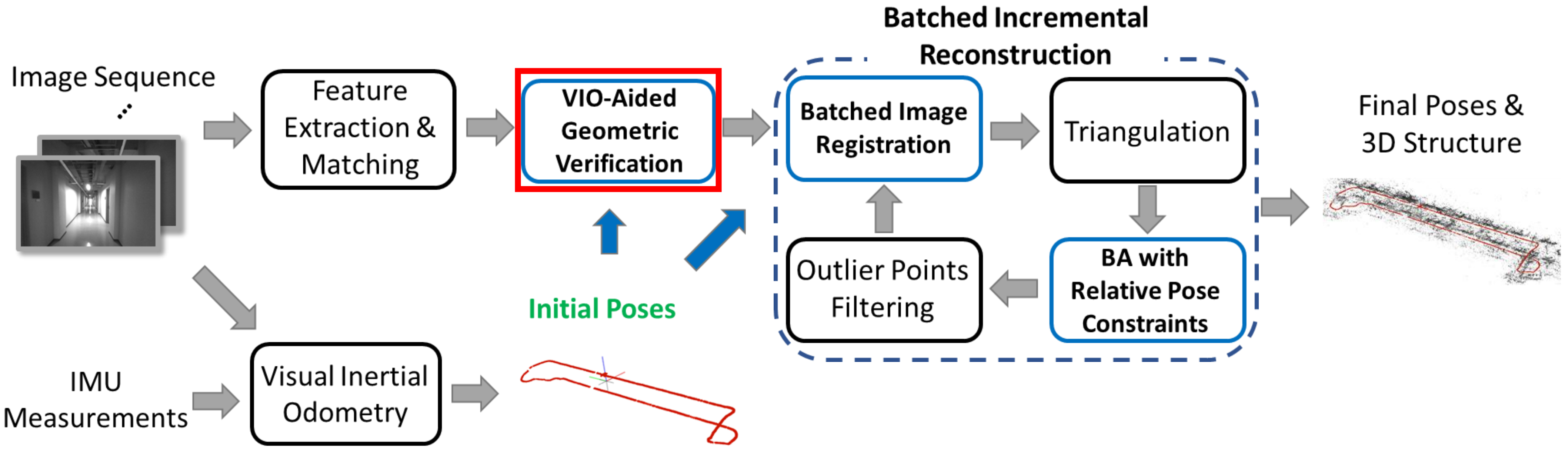
Proposed Method

✓ **Robustness**

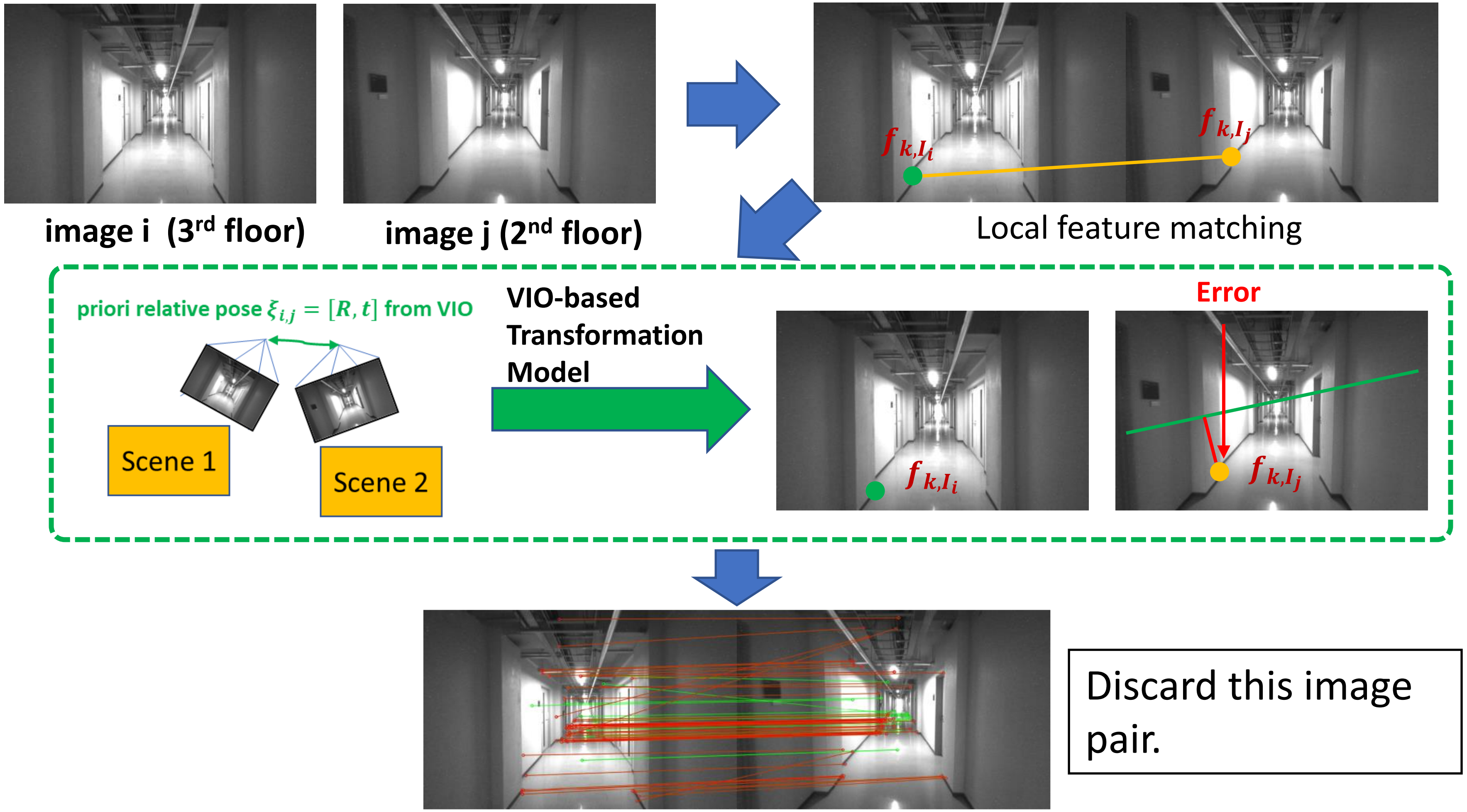
✓ **Accuracy**

Our Pipeline

- We adopt the camera pose estimation of visual-inertial odometry (VIO) into each step of the SfM pipeline.

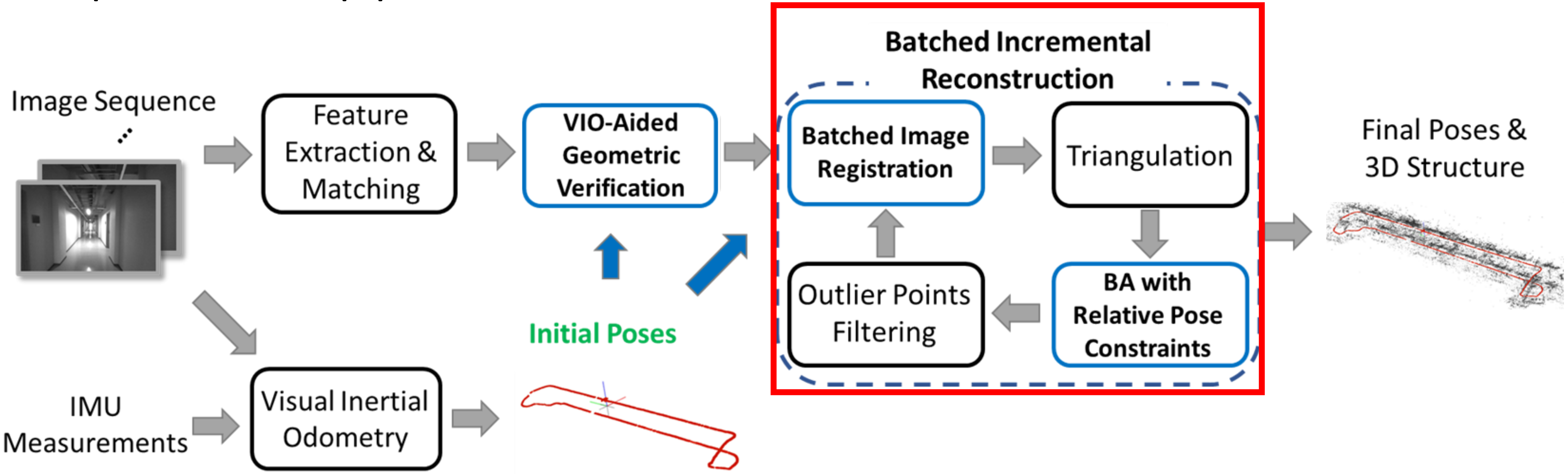


Our Approach: VIO-Aided Geometric Verification

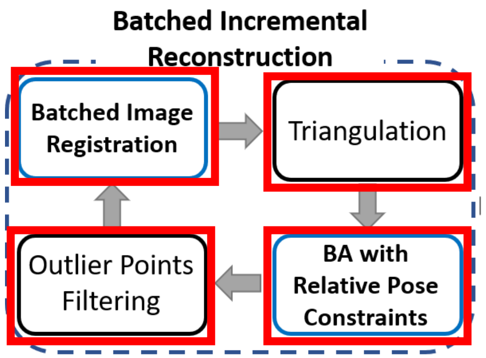


Our Pipeline

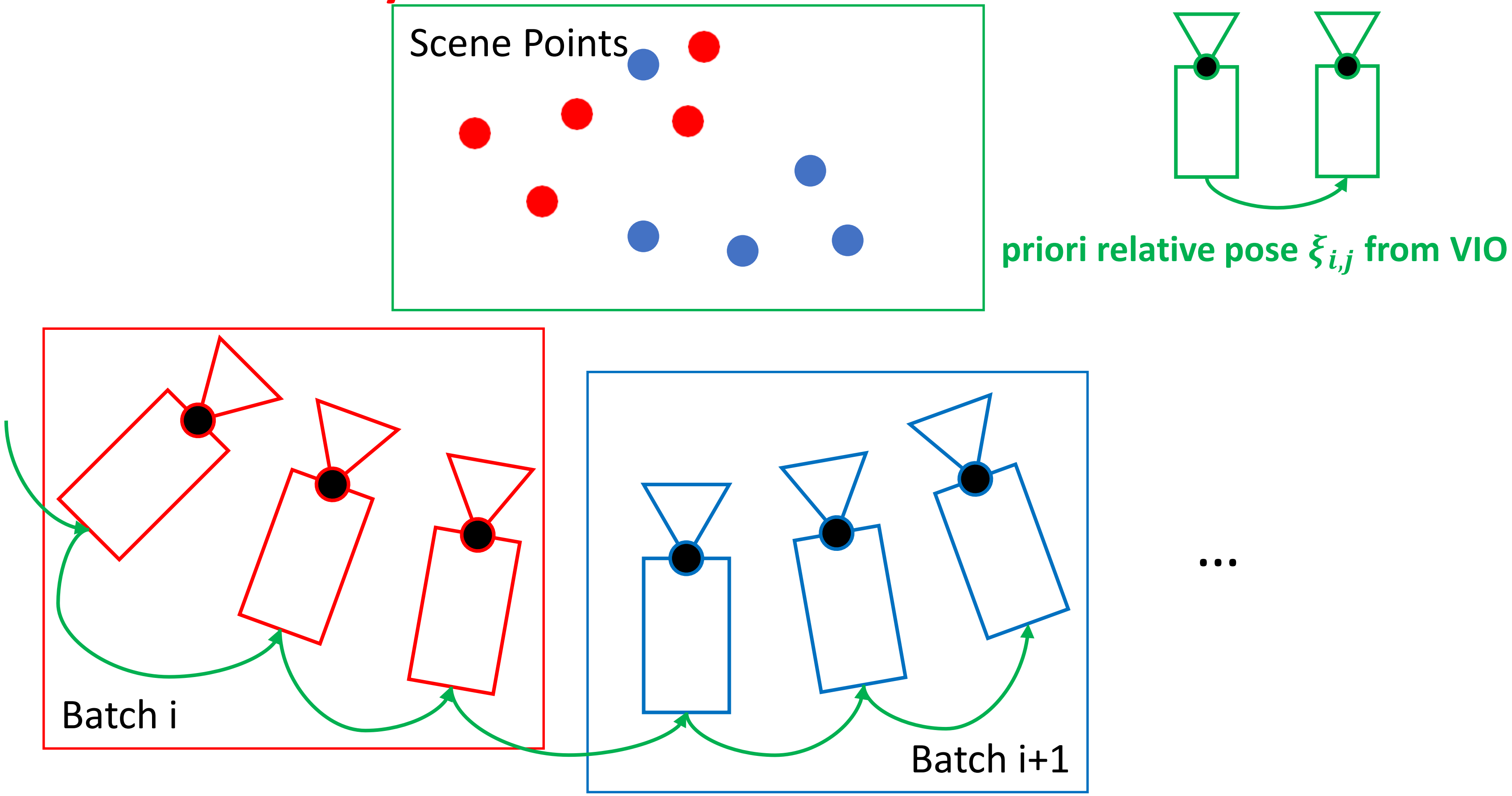
- We adopt the camera pose estimation of visual-inertial odometry (VIO) into each step of the SfM pipeline.



Our Approach: Batched Incremental Reconstruction



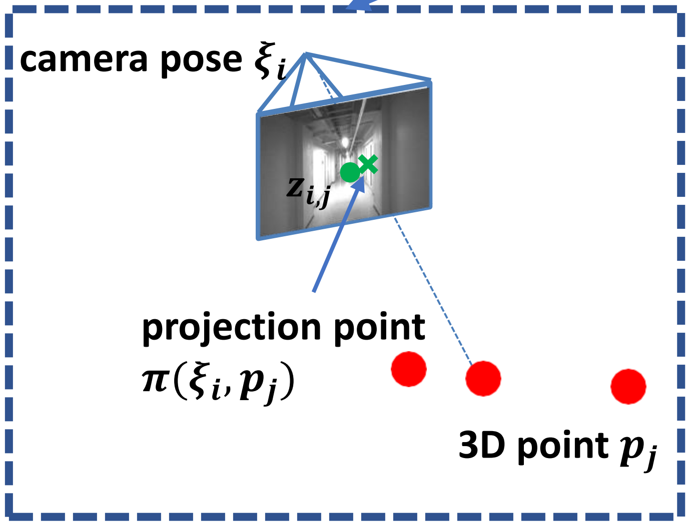
- We split the image sequence into several batches and register them **incrementally**.

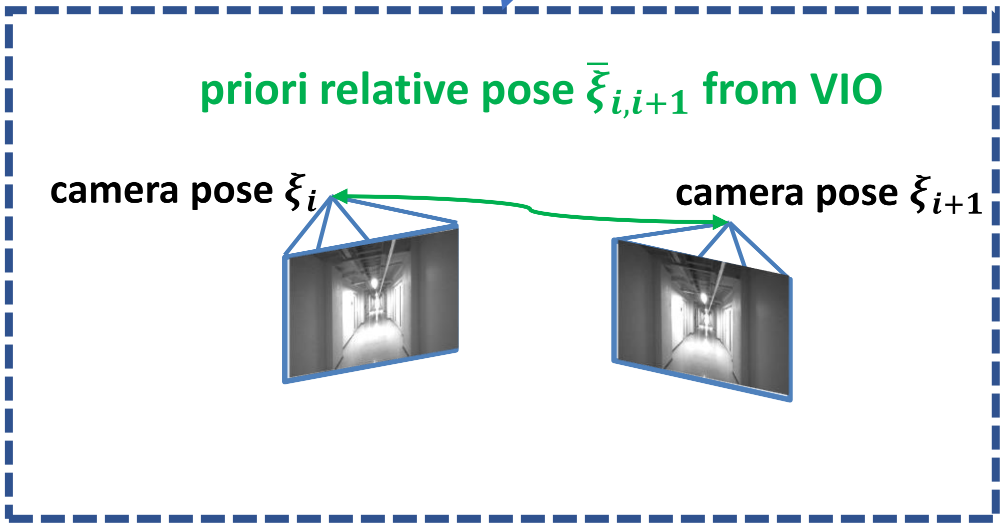


Our Approach: Bundle adjustment with Relative Pose Constraint

- Incremental optimization using relative poses
  - ❑ cost function

$$\arg\min_{\xi_i, p_j} \underbrace{\sum_i \sum_j \left\| z_{ij} - \pi\left(\xi_i, p_j\right) \right\|_2^2}_{\text{Reprojection Error}} + \sum_i w\left(c_{i,i+1}\right) \underbrace{\left\| \xi_i^{-1} \otimes \xi_{i+1} - \bar{\xi}_{i,i+1} \right\|^2}_{\text{Relative pose constraint}}$$





Our Approach: Bundle adjustment with Relative Pose Constraint

- Incremental optimization using relative poses
  - ❑ cost function

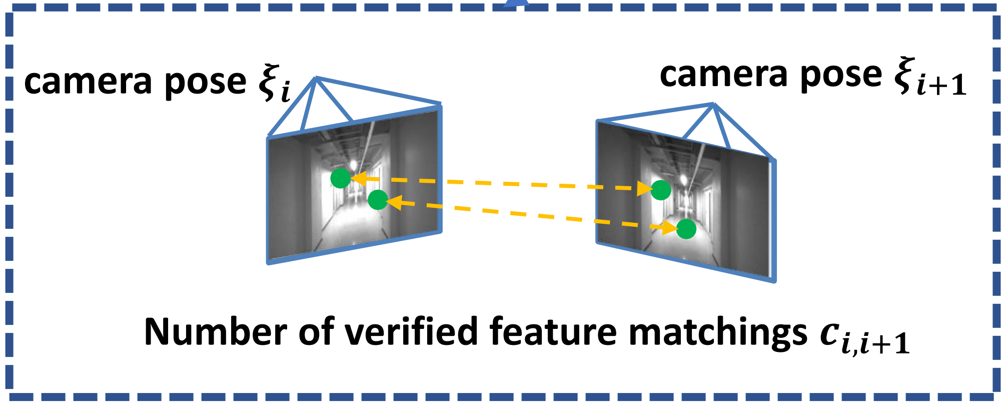
$$\arg\min_{\xi_i, p_j} \sum_i \sum_j \left\| z_{ij} - \pi\left(\xi_i, p_j\right) \right\|_2^2 + \sum_i w\left(c_{i,i+1}\right) \left\| \xi_i^{-1} \otimes \xi_{i+1} - \bar{\xi}_{i,i+1} \right\|^2$$

Self-Adaptive weighting factor

$$w = \alpha e^{-\beta c_{i,i+1}}$$

$c_{i,i+1} \downarrow, \omega \uparrow$

$c_{i,i+1} \uparrow, \omega \downarrow$



Results

- Quantitative evaluation of recovered camera poses (EuRoC Dataset)

Sequence	Visual methods						Visual-inertial methods					
	COLMAP		ORB-SLAM2		DSO		OKVIS		VINS-Mono		Ours	
Name	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE	ME
V1_02_medium	0.043	0.040	0.064	0.063	0.598	0.213	0.067	0.062	0.060	0.057	**0.022**	**0.019**
V1_03_difficult	0.054	0.051	0.531	0.235	0.925	0.935	0.105	0.089	0.173	0.131	**0.043**	**0.032**
V2_02_medium	0.029	0.032	0.056	0.056	0.092	0.080	0.081	0.058	0.124	0.103	**0.014**	**0.012**
V2_03_difficult	0.041	0.036	0.079	0.073	1.386	1.008	-	-	0.191	0.153	**0.029**	**0.021**
MH_03_medium	0.040	0.034	0.038	0.032	0.172	0.135	0.146	0.143	0.080	0.067	**0.035**	**0.029**
MH_04_difficult	0.095	0.078	**0.059**	**0.049**	0.172	0.171	0.138	0.131	0.124	0.123	0.092	0.077
MH_05_difficult	0.084	0.064	**0.068**	**0.055**	0.102	0.093	0.261	0.227	0.133	0.110	0.083	0.072

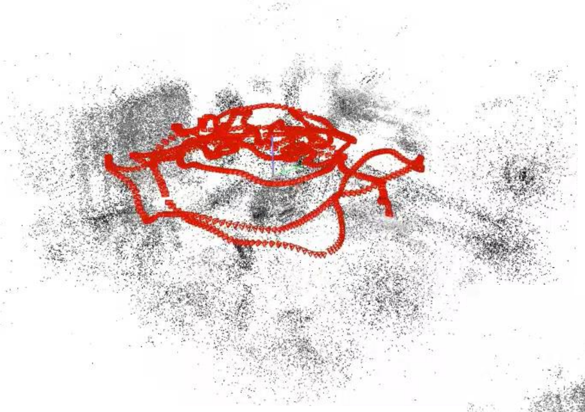
- RMSE : the root mean square position error in meters
- ME : the median position error in meters

- Red : Best
- Blue : Second best

Qualitative Results

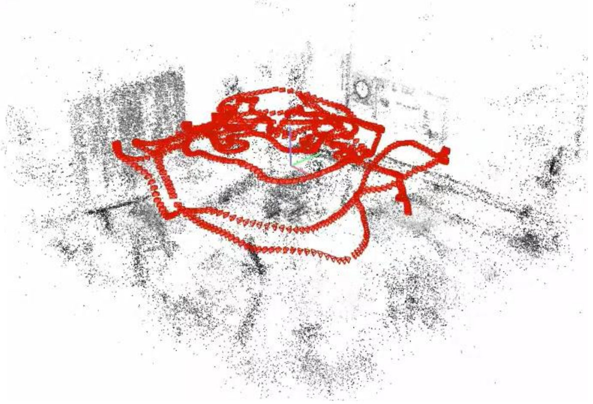
- Indoor scene (EuRoC Dataset)

Input images



VINS-Mono
(VIO)

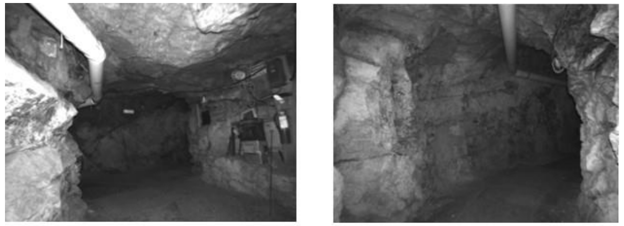
COLMAP
(SfM)



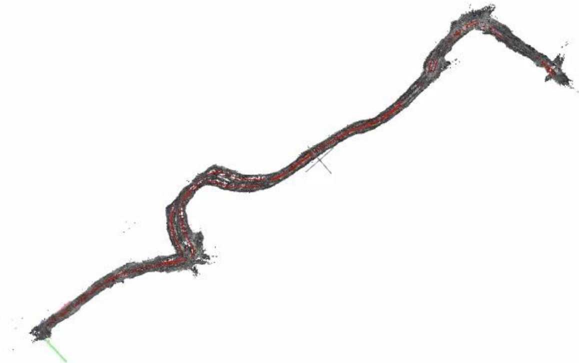
Proposed method

Qualitative Results

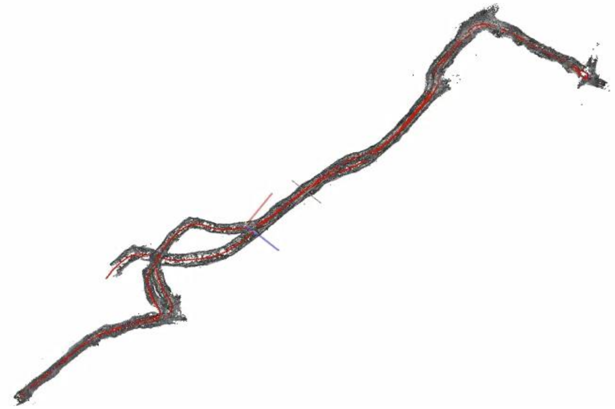
- Mine scene (OIVIO Dataset)



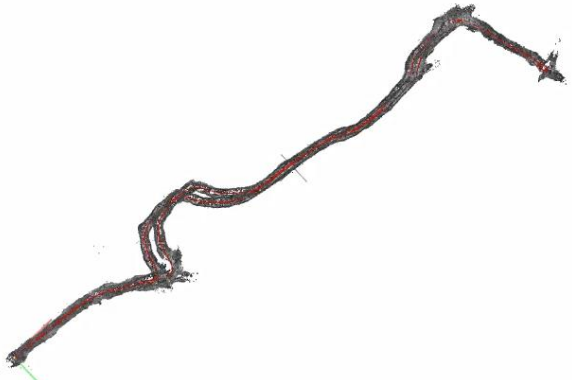
Input images



VINS-Mono
(VIO)



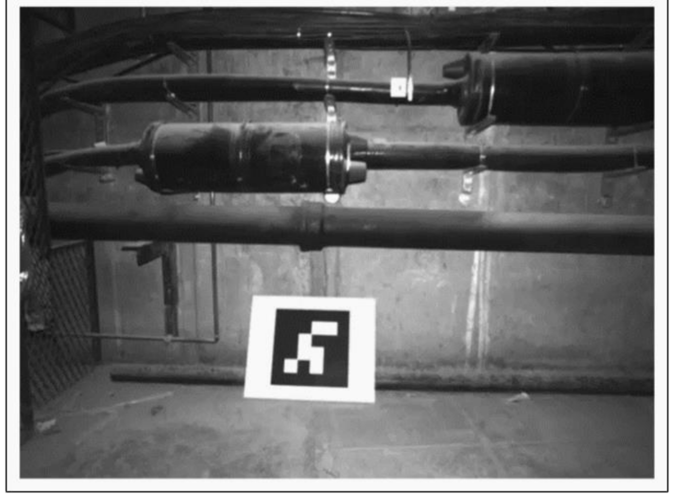
COLMAP
(SfM)



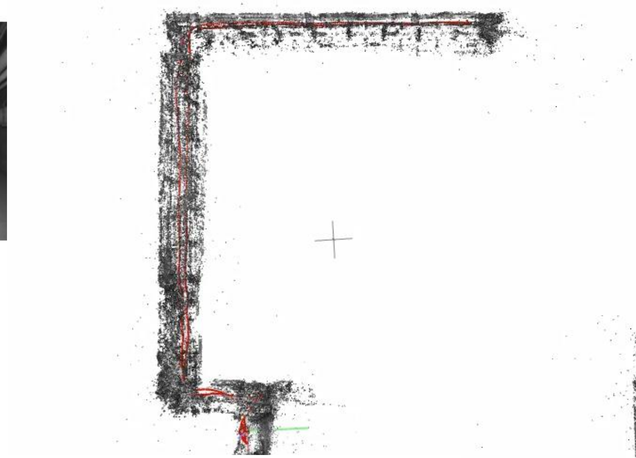
Proposed method

Qualitative Results

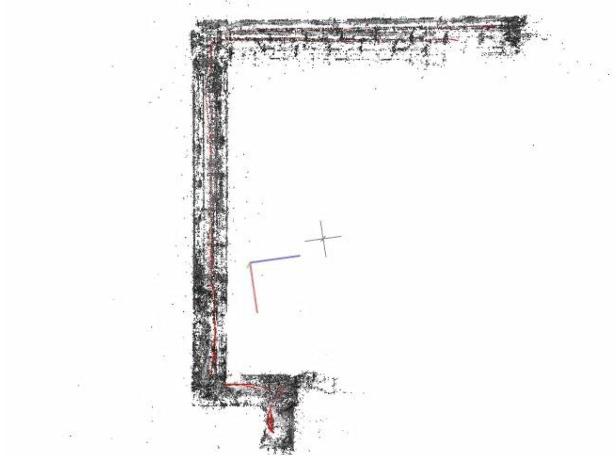
- Tunnel scene (OIVIO Dataset)



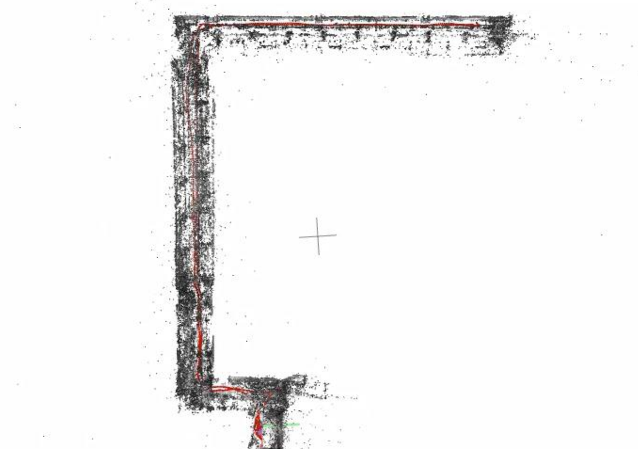
Input images



VINS-Mono
(VIO)



COLMAP
(SfM)



Proposed method

Conclusion

- We propose an SfM-based 3D reconstruction pipeline that effectively takes advantage of the camera pose information from a VIO.

- Our method provides accurate camera poses and 3D points by adopting the camera pose estimation of VIO into each step of the SfM pipeline.

- The experiments on publicly available datasets demonstrate that our system can achieve an accurate and robust 3D reconstruction in challenging environments containing less visual information.

Thank you for listening!