

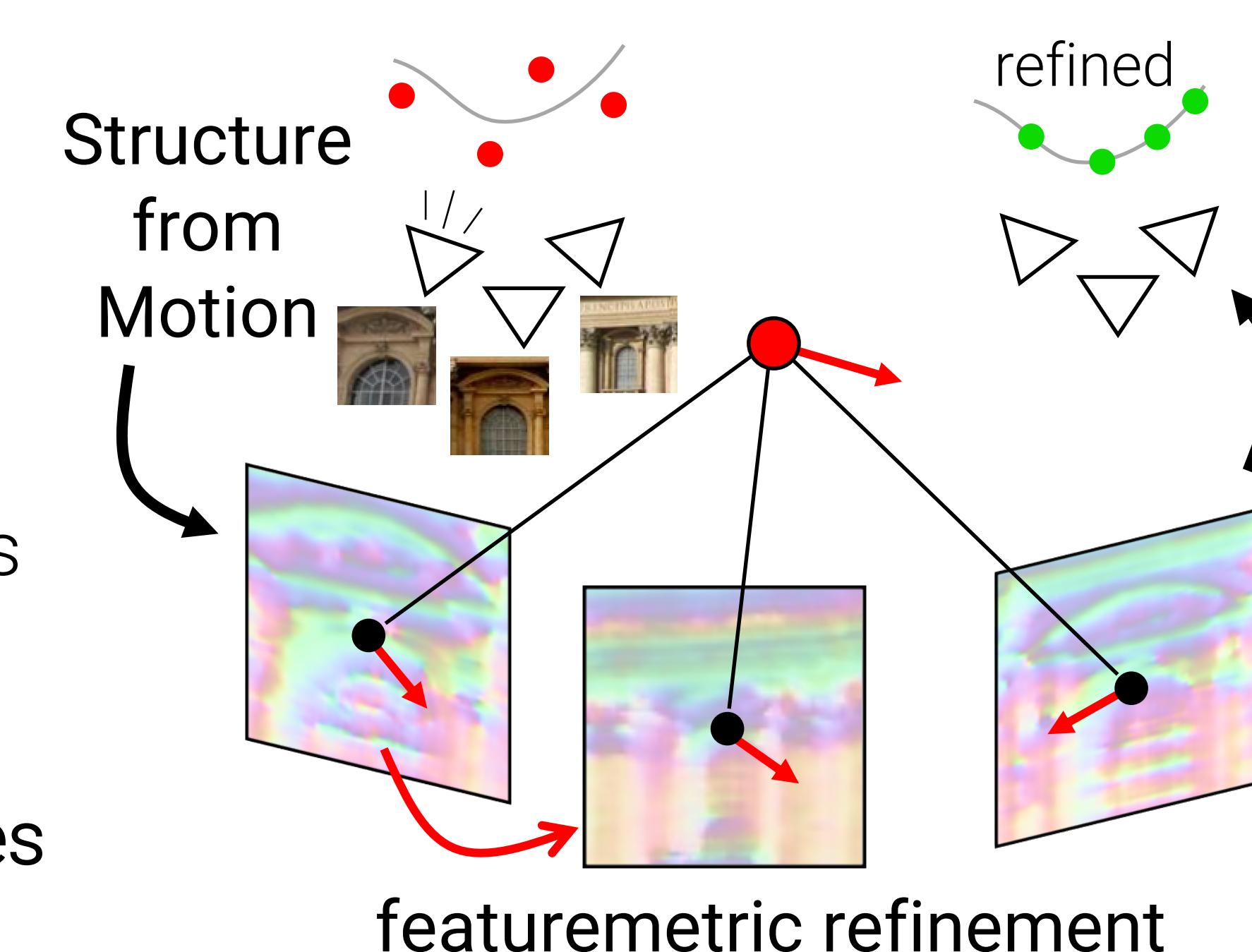
Pixel-Perfect Structure-from-Motion with Featuremetric Refinement

Philipp Lindenberger^{1*} Paul-Edouard Sarlin^{1*} Viktor Larsson¹ Marc Pollefeys^{1,2}

1. Improving the accuracy of SfM via refinement

Our problem:

- Sparse SfM: detect → match → SfM
- Detect once and trust the 2D keypoints
- But: single-view detection is not subpixel-accurate
- ⇒ errors in camera poses & 3D points when few images



Our solution: featuremetric optimization

- Refinement of keypoints, camera poses & 3D points
- Accurate: multi-view, before and after SfM
- Robust to real-world changes: using deep feature alignment
- Scalable: with dense but local image information

4. Geometric vs featuremetric costs

Keypoint adjustment

$$\sum_{(u,v) \in \mathcal{M}(j)} \|\mathbf{p}_v + \mathbf{T}_{v \rightarrow u}[\mathbf{p}_v] - \mathbf{p}_u\|_\gamma$$

→ Patch Flow, Dusmanu et al 2020

- ✗ quadratic complexity of flow \mathbf{T}
- ✗ approximate optimization

→ Our KA with features \mathbf{F}

- ✓ linear complexity (# images)
- ✓ exact interpolation

Bundle adjustment

$$\sum_j \sum_{(i,u) \in \mathcal{T}(j)} \|\Pi(\mathbf{R}_i \mathbf{P}_j + \mathbf{t}_i, \mathbf{C}_i) - \mathbf{p}_u\|_\gamma$$

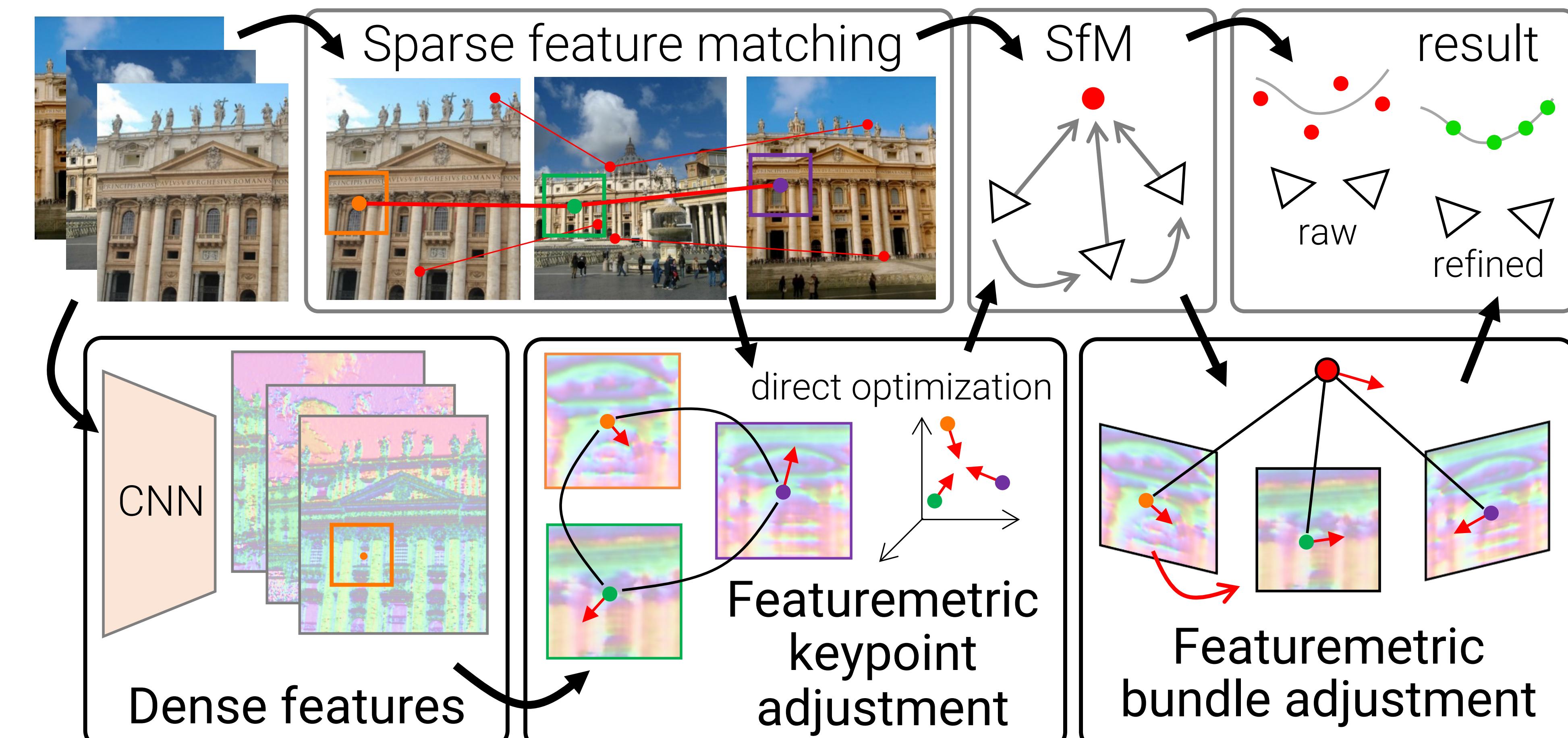
→ Gold standard BA

- ✗ ignores keypoint uncertainties

→ Our BA with features \mathbf{F}

- ✓ uses uncertainties from image information

2. Two-step refinement compatible with any SfM pipeline



Keypoint adjustment:

- More inlier matches for SfM/localization
- Robust to incorrect matches
- Parallel computation over all tracks

Bundle adjustment:

- Optimize camera poses & 3D points
- Use constraints of the known geometry
- Cost pre-computation ⇒ memory-efficient

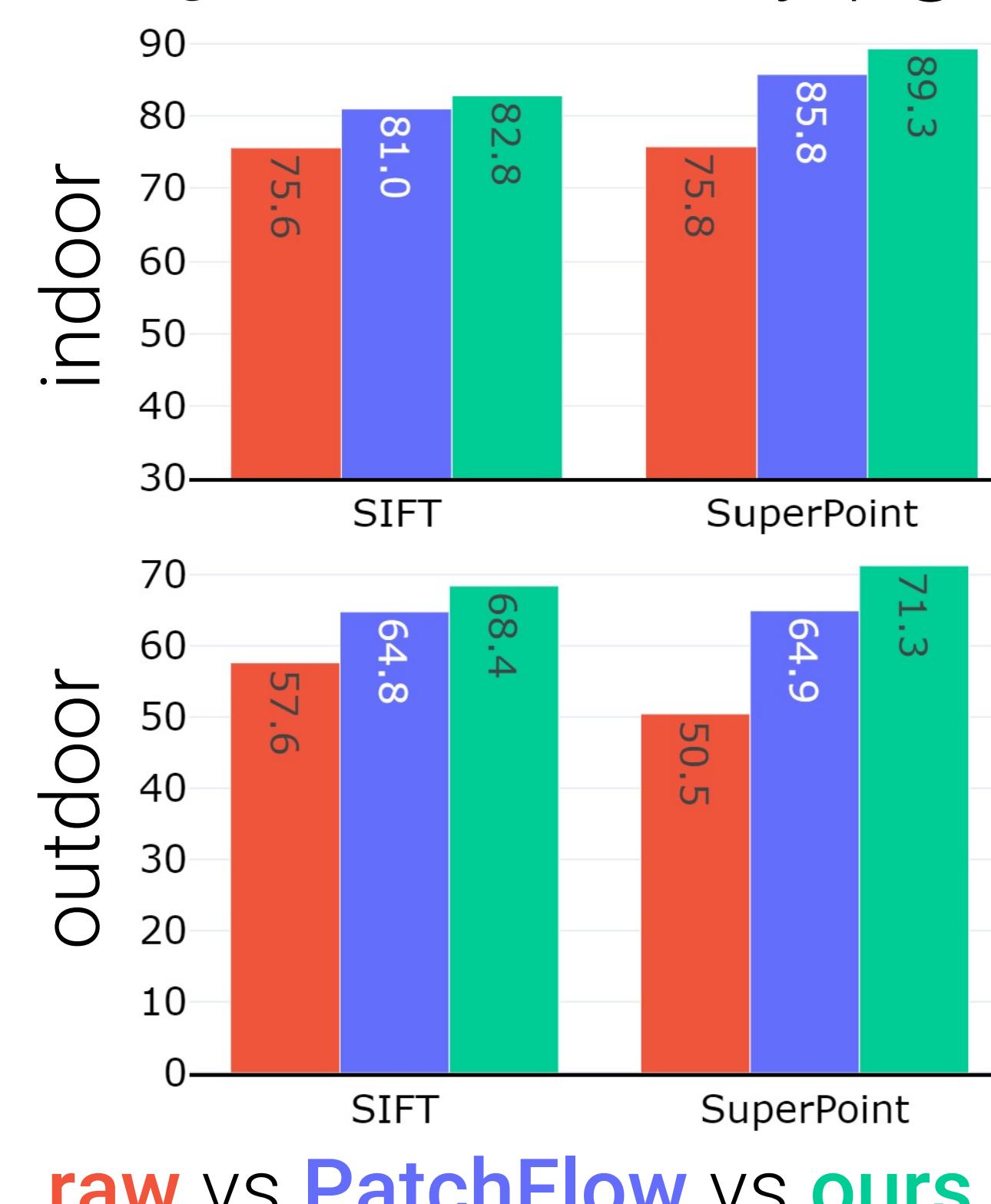
3. Scalable to large scenes

- x20+ faster than PatchFlow
- Only 20% time overhead over COLMAP
- 10GB RAM for 1k images ⇒ runs on desktop PC
- Local image patches: 50GB on disk for 1k images

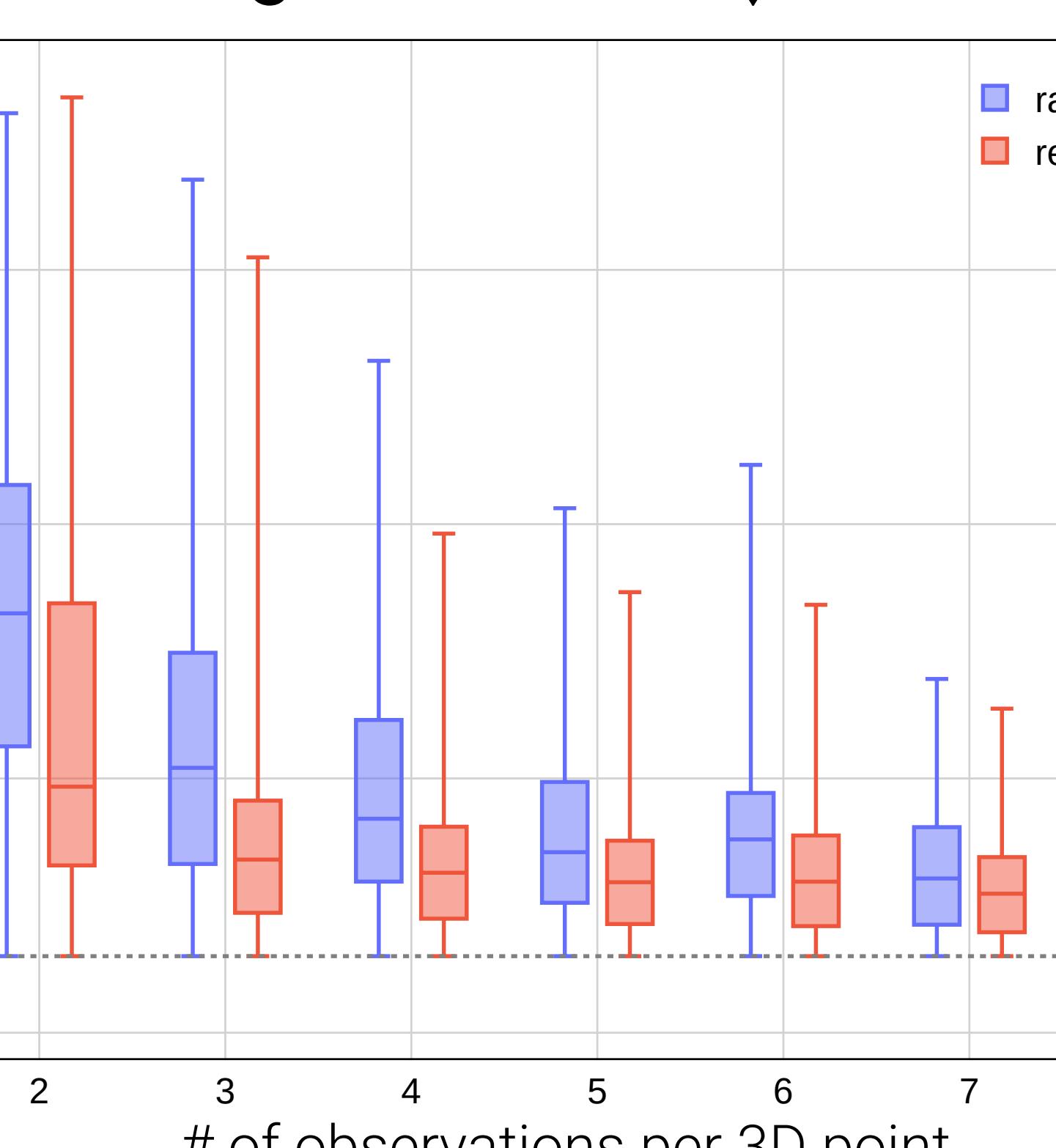


5. Results: pixel-accurate SfM & visual localization

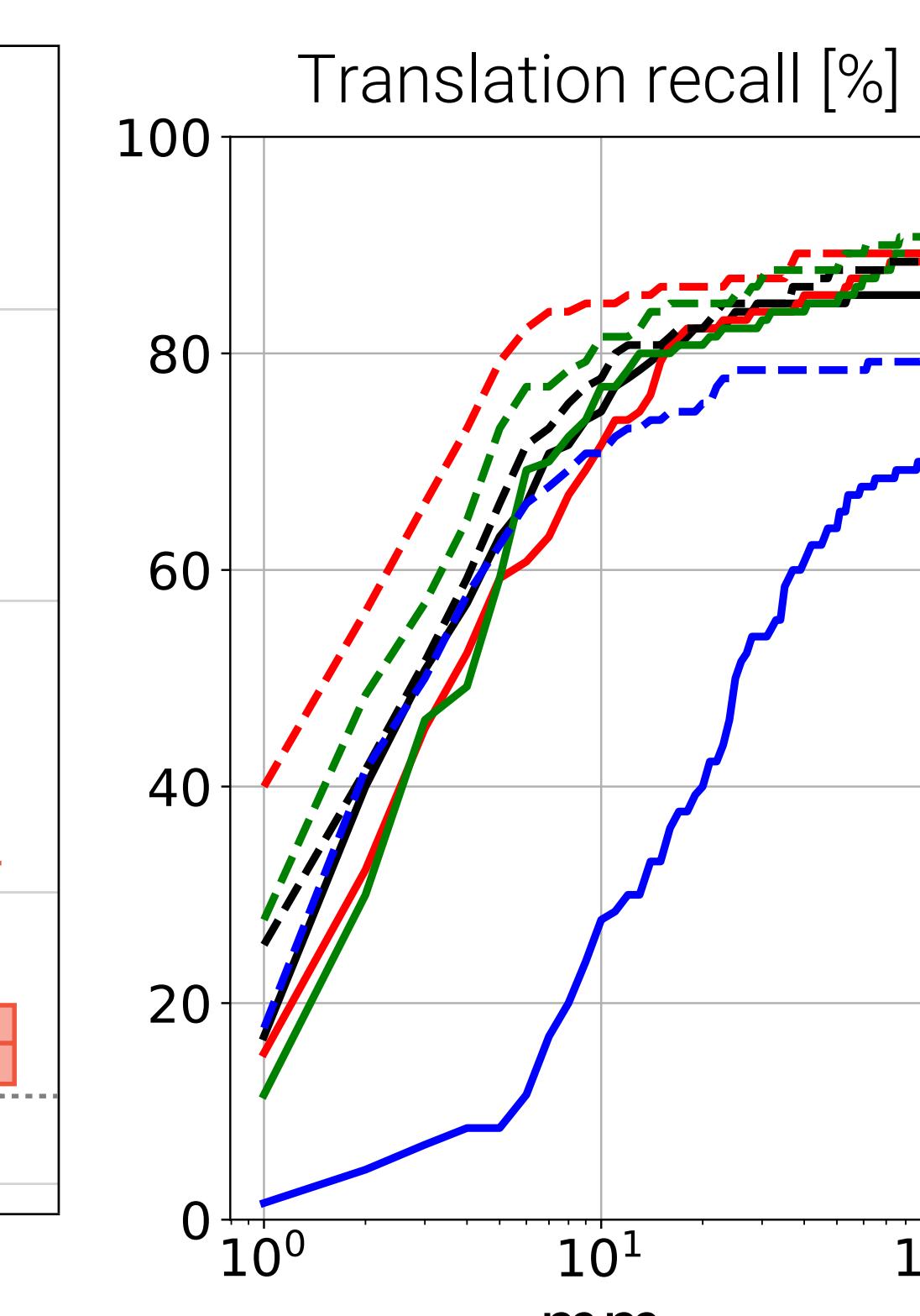
triangulation accuracy ↑ @ 1cm



triangulation error ↓ on ETH3D



Camera localization



End-to-end SfM on crowd-sourced data



More accurate sparse point cloud

