

Photometric Bundle Adjustment for Vision-Based SLAM



Hatem Alismail, Brett Browning, Simon Lucey

Carnegie Mellon University

Summary & Contributions

- Fundamentally different approach to VSLAM in unstructured scenes
- Direct VSLAM for the joint estimation of structure & motion
- Does not require precise correspondences
- Correspondences are estimated automatically to maximize precision of the estimated parameters
- Improves on results obtained using the *Gold Standard* geometric Bundle Adjustment and loop closure
- Almost no parameters to tune

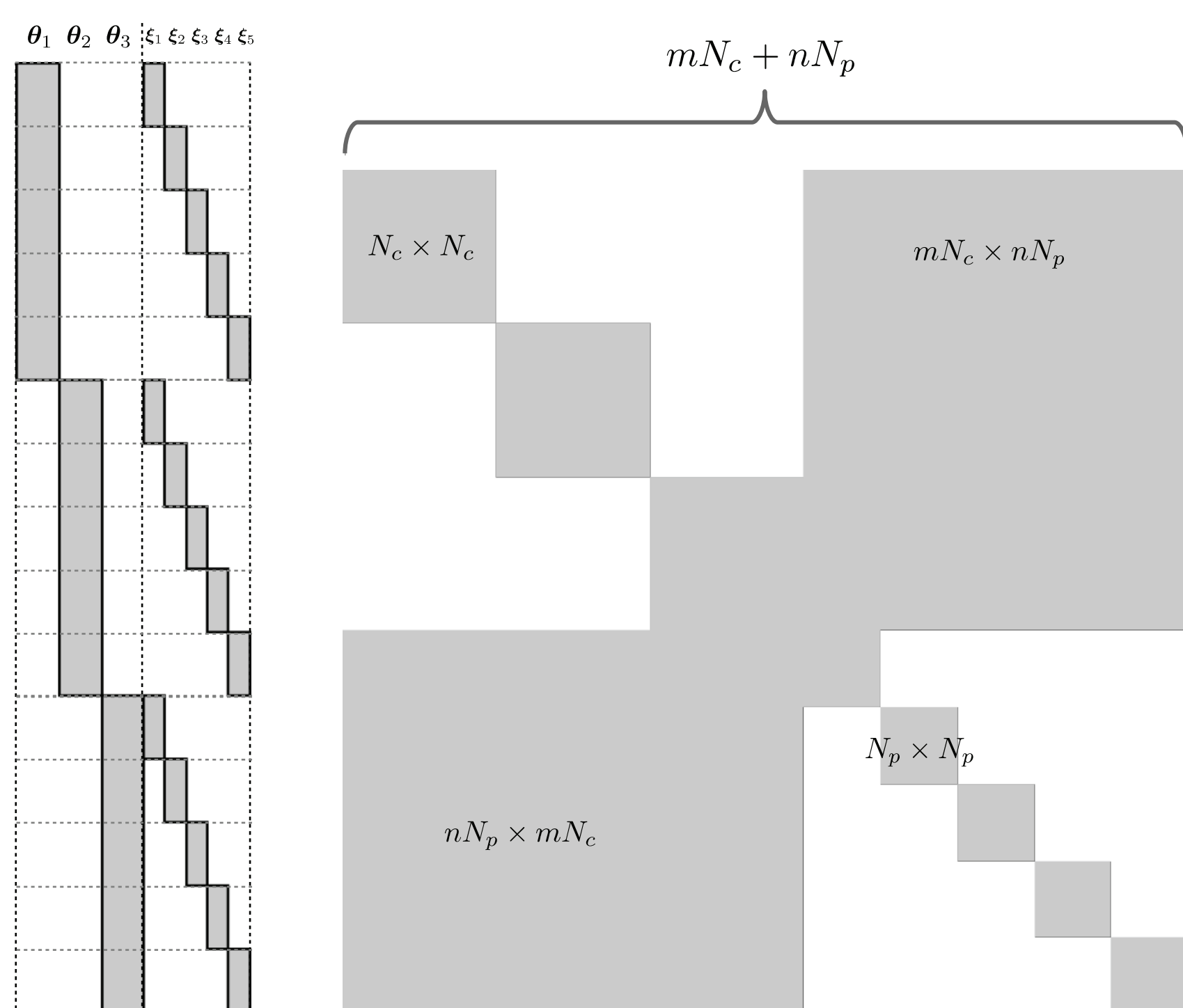
Objective

$$\{\Delta\theta_i^*, \Delta\xi_j^*\} = \underset{\Delta\theta_i, \Delta\xi_j}{\operatorname{argmin}} \sum_{j=1}^N \sum_{k \in V(j)} \mathcal{E}(\phi_j, \mathbf{I}_k; \Delta\theta_k, \Delta\xi_j),$$

motion & structure
visibility list for the j-th scene point

$$\mathcal{E}(\phi, \mathbf{I}'; \theta, \xi) = \sum_{\mathbf{u} \in \mathcal{N}} \frac{1}{2} \|\phi(\mathbf{u}) - \mathbf{I}'(\pi(\theta, \xi) + \mathbf{u})\|^2$$

image patch (descriptor)
correspondence given current estimates of motion & structure

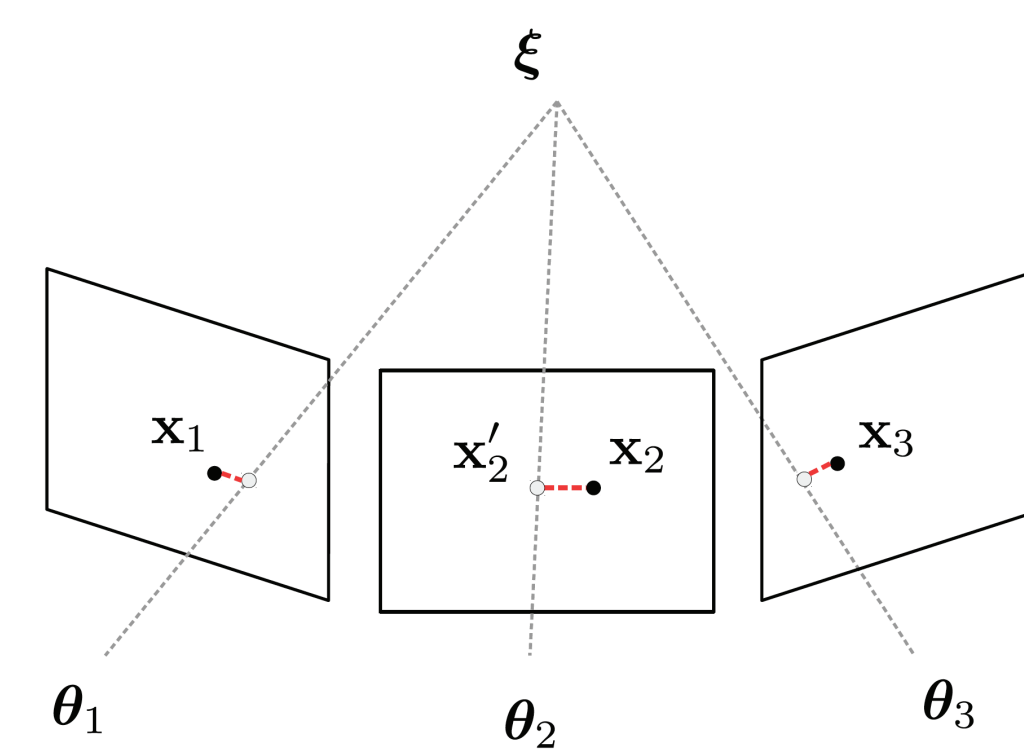


The same **sparsity** structure as in traditional feature-based Bundle Adjustment

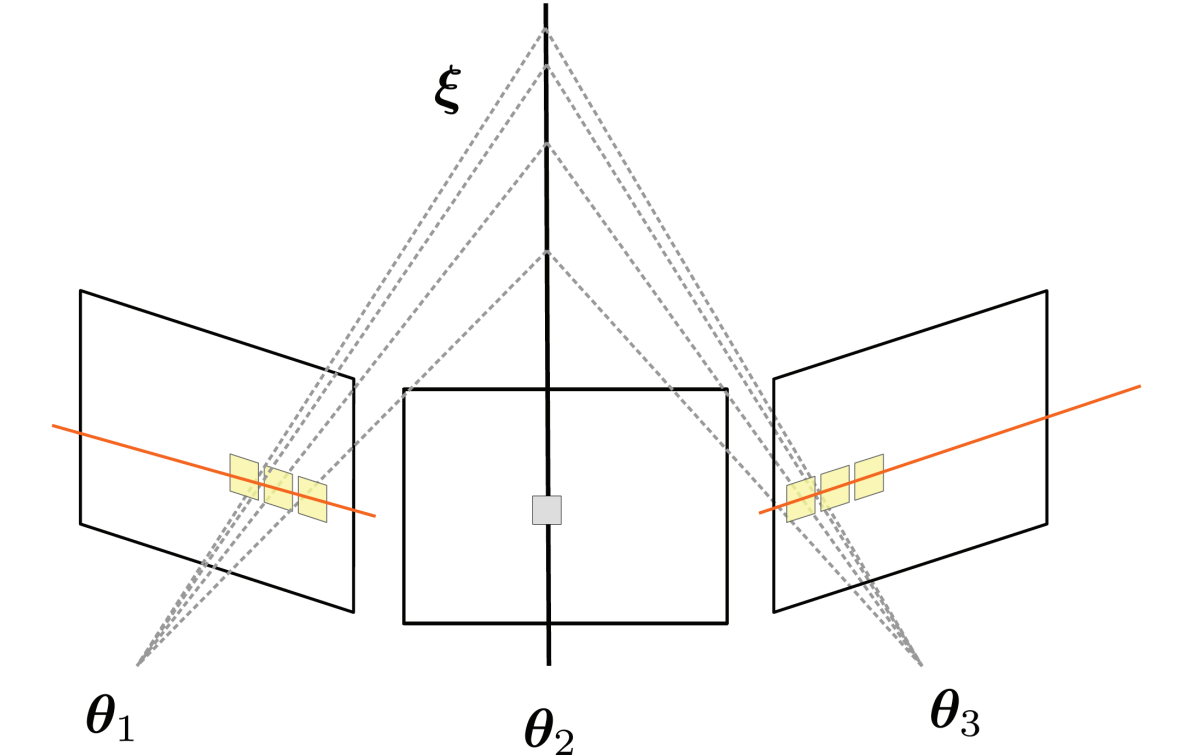
Slightly more expensive to evaluate due to repeated interpolation and requiring images to remain in memory

Possible to run in real-time, but current implementation is unoptimized

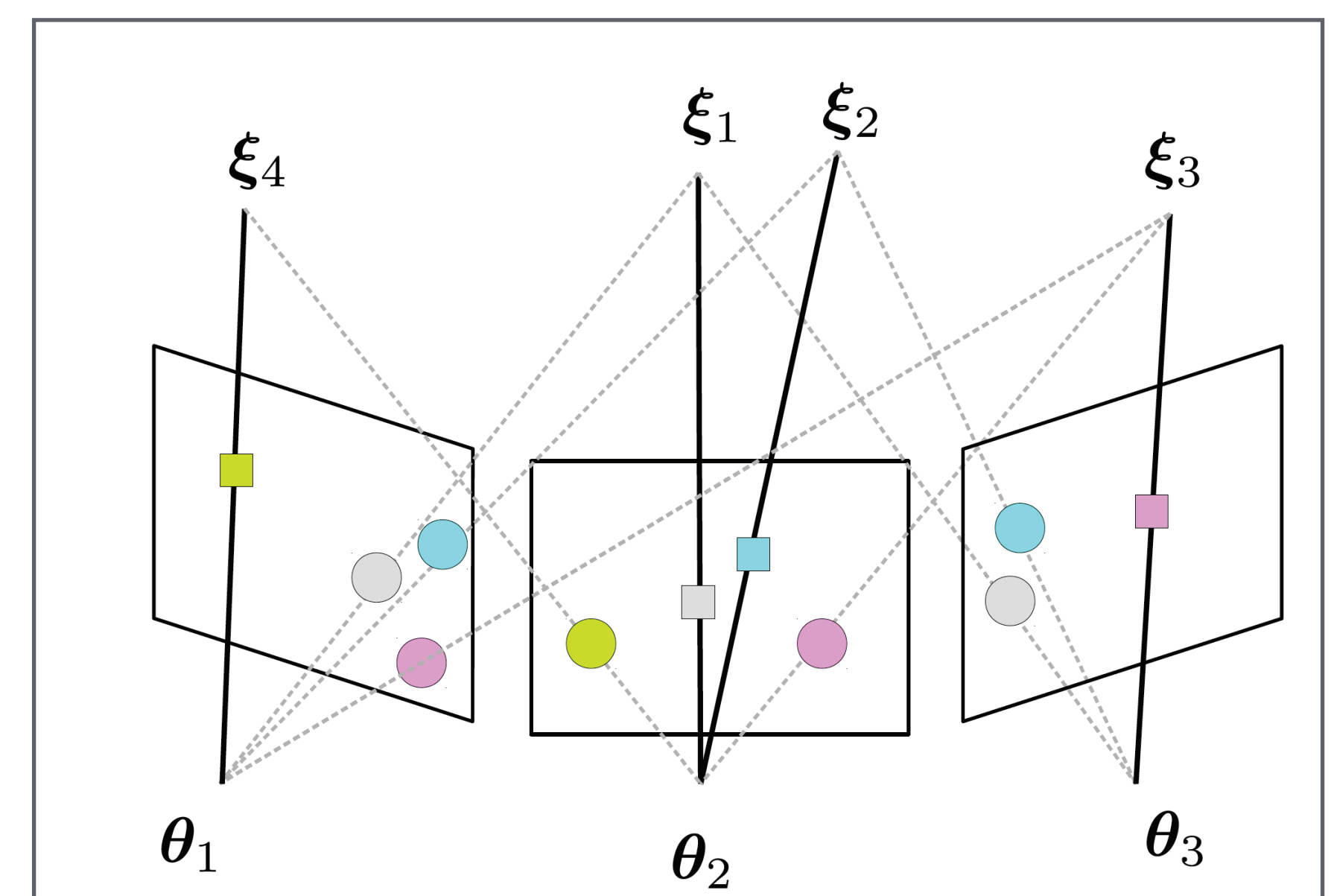
Overview



Traditional BA: Given fixed correspondences refine motion & structure to minimize the re-projection error



Multi-view Stereo: Given fixed poses, refine structure by finding correspondences



Proposed: Given initialization, **simultaneously** estimate motion & structure to maximize the photometric consistency
implicitly estimate the correspondences

Algorithm Details

Input image

Gradient magnitude

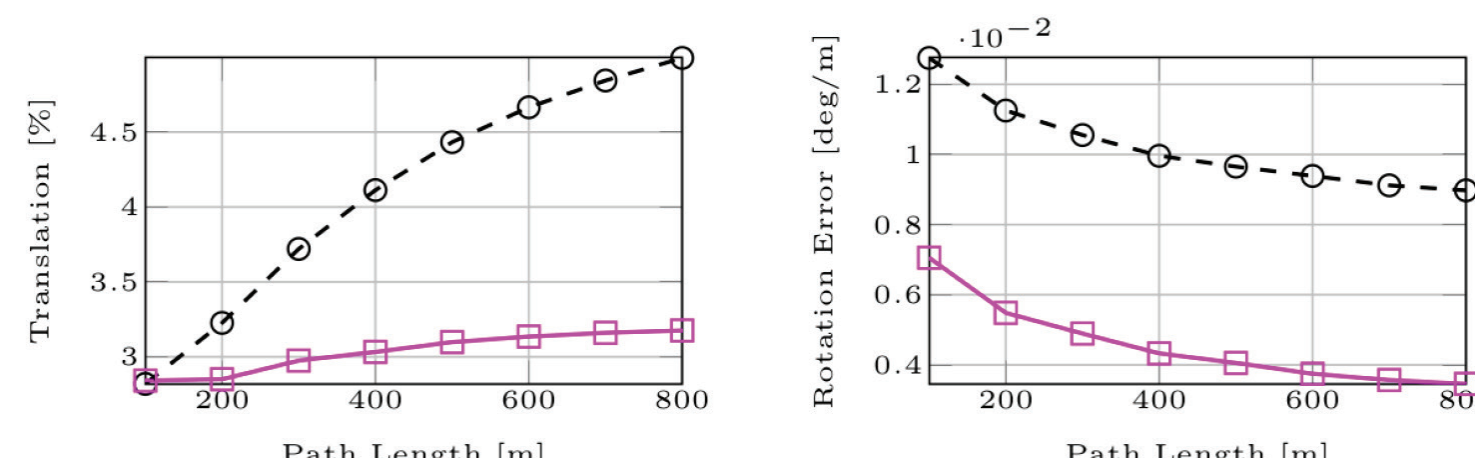
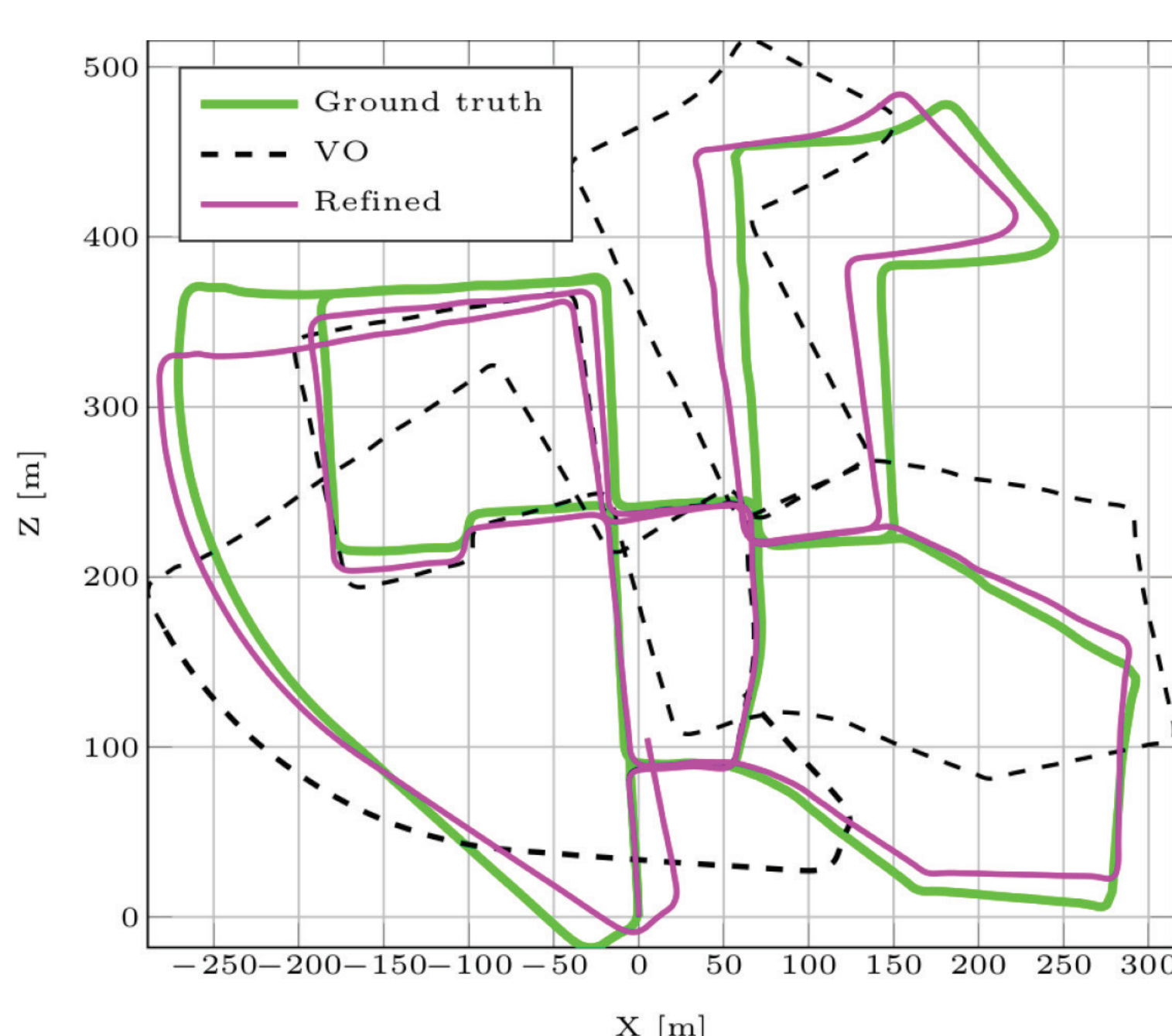
Depth initialization

Selected pixels

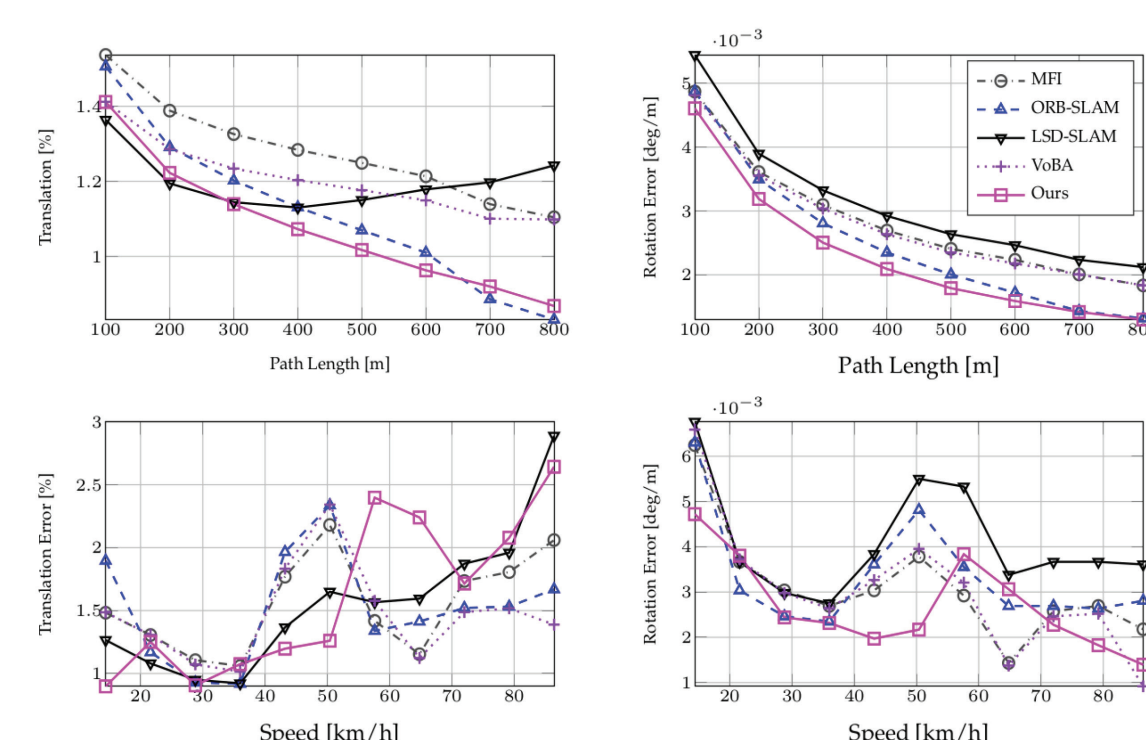
- 1 Select pixels at integer locations based on the gradient magnitude
- 2 Determine visibility based on initialization; filter outliers with ZNCC
- 3 Prevent re-initializing the same scene points; keep track of expected projection location and disallow close points

Results

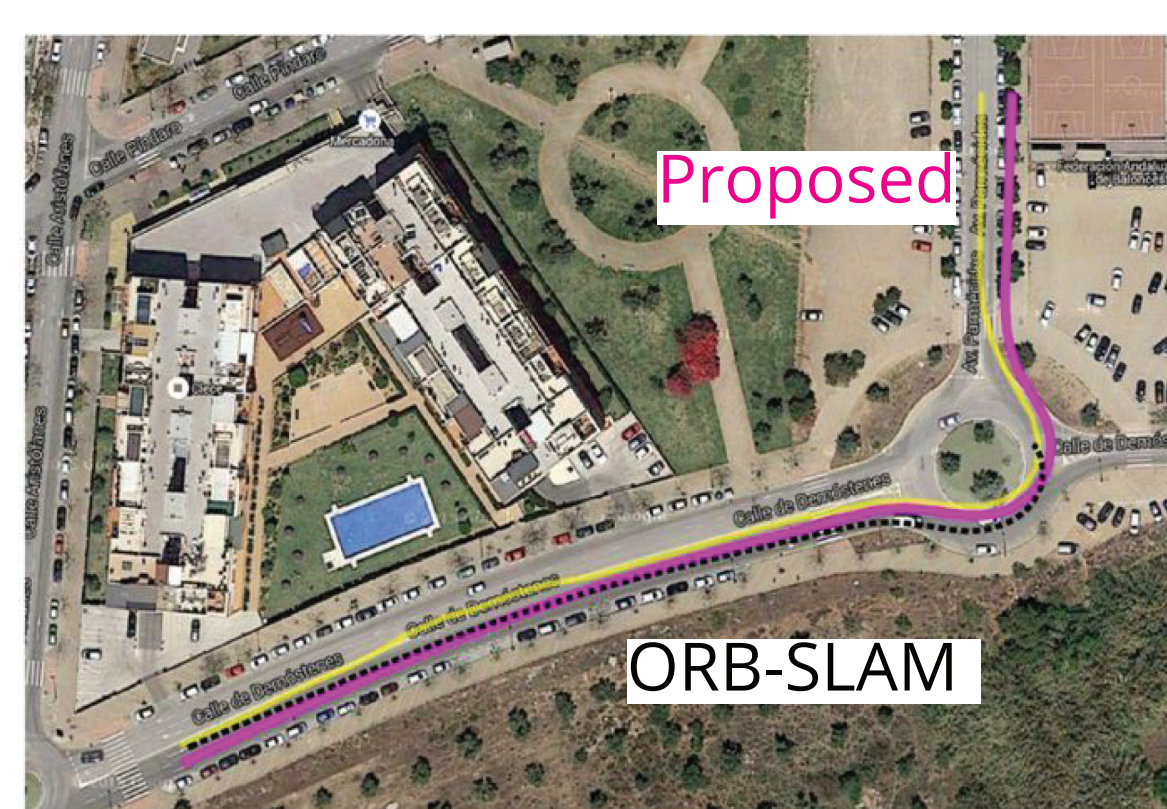
Starting from poor initialization



Pose from ORB-SLAM; structure from block matching



High frame rate data, no need for pose initialization



Conclusions & Future Work

- New correspondence-free approach to VSLAM based on maximizing the photometric consistency directly
- Correspondences are implicitly determined to maximize precision of pose & structure
- Improves on state-of-the-art feature-based Bundle Adjustment & loop closure
- Future: Better pixel selection & real-time implementation
- Future: Theoretical analysis of convergence & optimality
- Is minimizing the re-projection error the last word on visual reconstruction? Can we obtain more precision by bring back the image in the loop?

For more info. contact:
hatem@cmu.edu