

Depth-Guided Sparse Structure-from-Motion for Movies and TV Shows

Sheng Liu^{*1,2} Xiaohan Nie² Raffay Hamid²

¹ *work done when Sheng was an intern at Amazon

Structure from motion (SfM) for movies and TV shows

Input

- a set of video frames observing a scene 100%

Goal

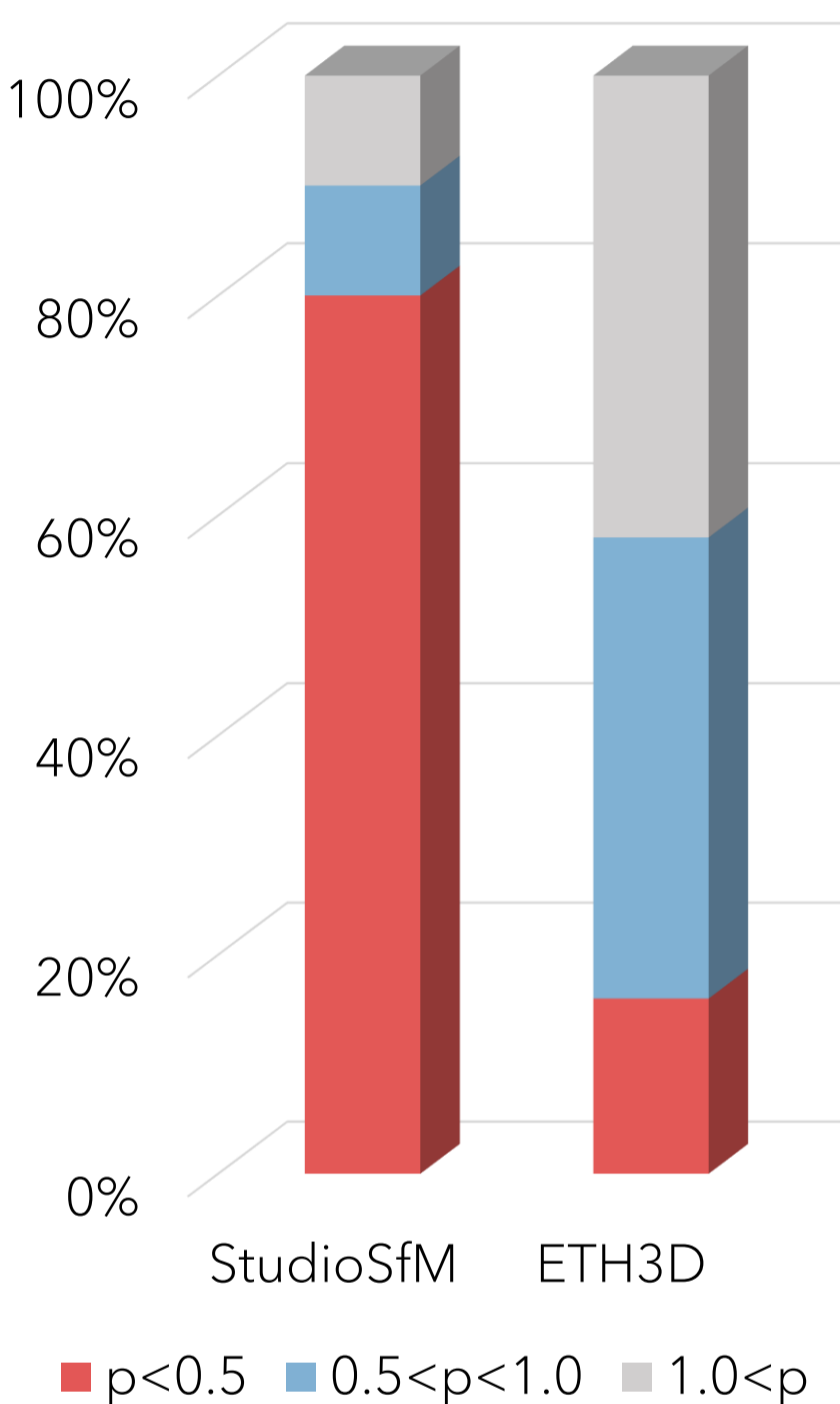
- estimate camera poses
- reconstruct 3D structure of the scene

Challenge

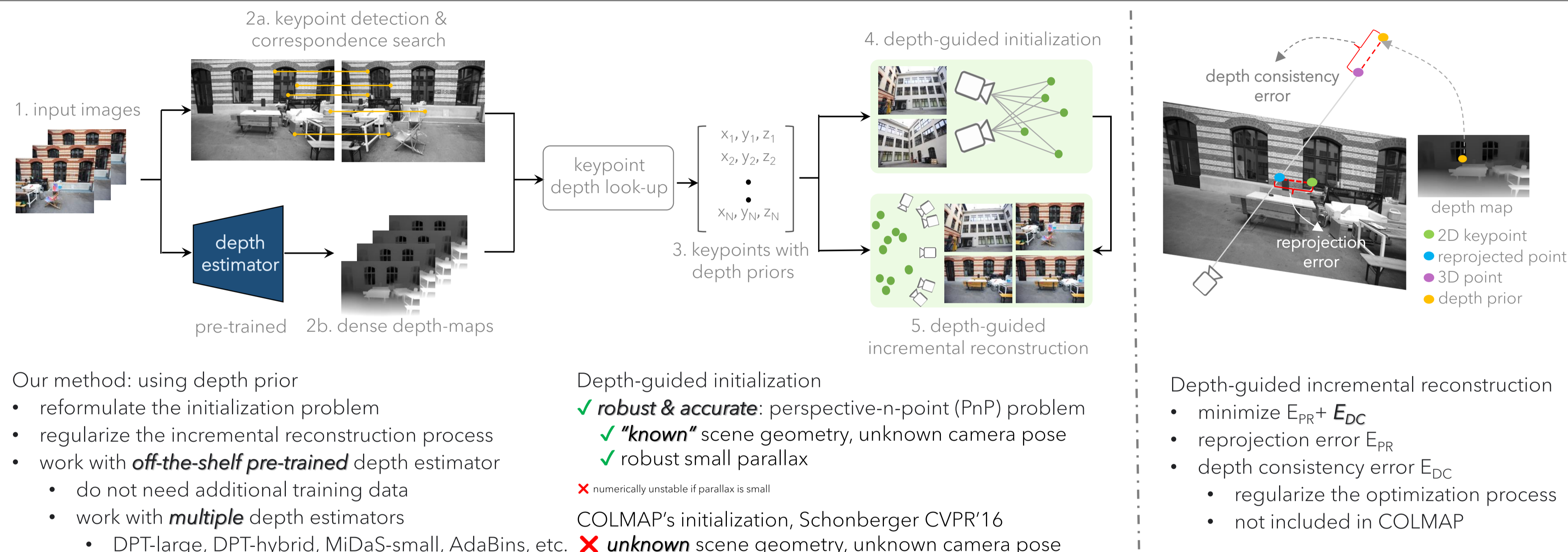
- small parallax (p)**
- structure from motion motion \rightarrow **parallax** \rightarrow 3D structure
- 3x more videos with parallax < 0.5
- 3x less videos with parallax > 1.0

StudioSfM dataset

- accurate:** annotated by professional visual effects artists
 - camera poses
 - sparse 3D point cloud
- diverse:** 130 shots with 21K frames from 15 TV episodes



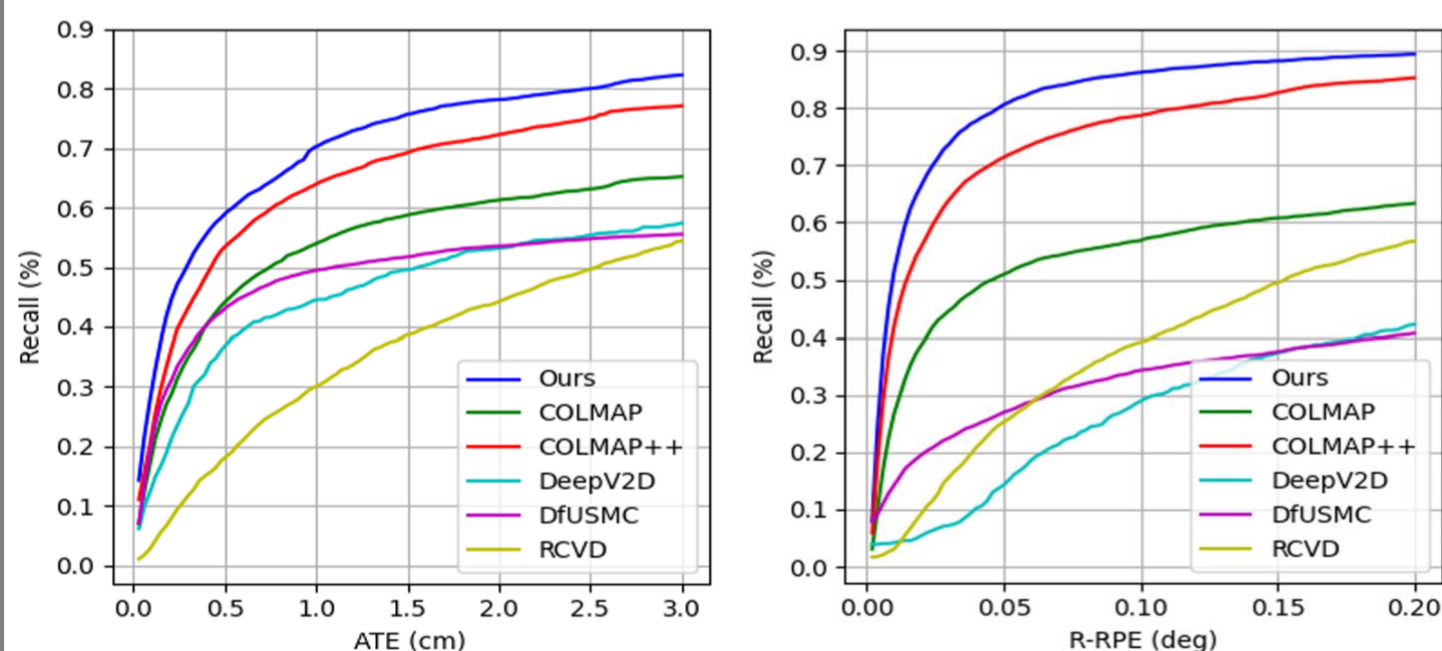
Depth-guided structure from motion



Experiment results

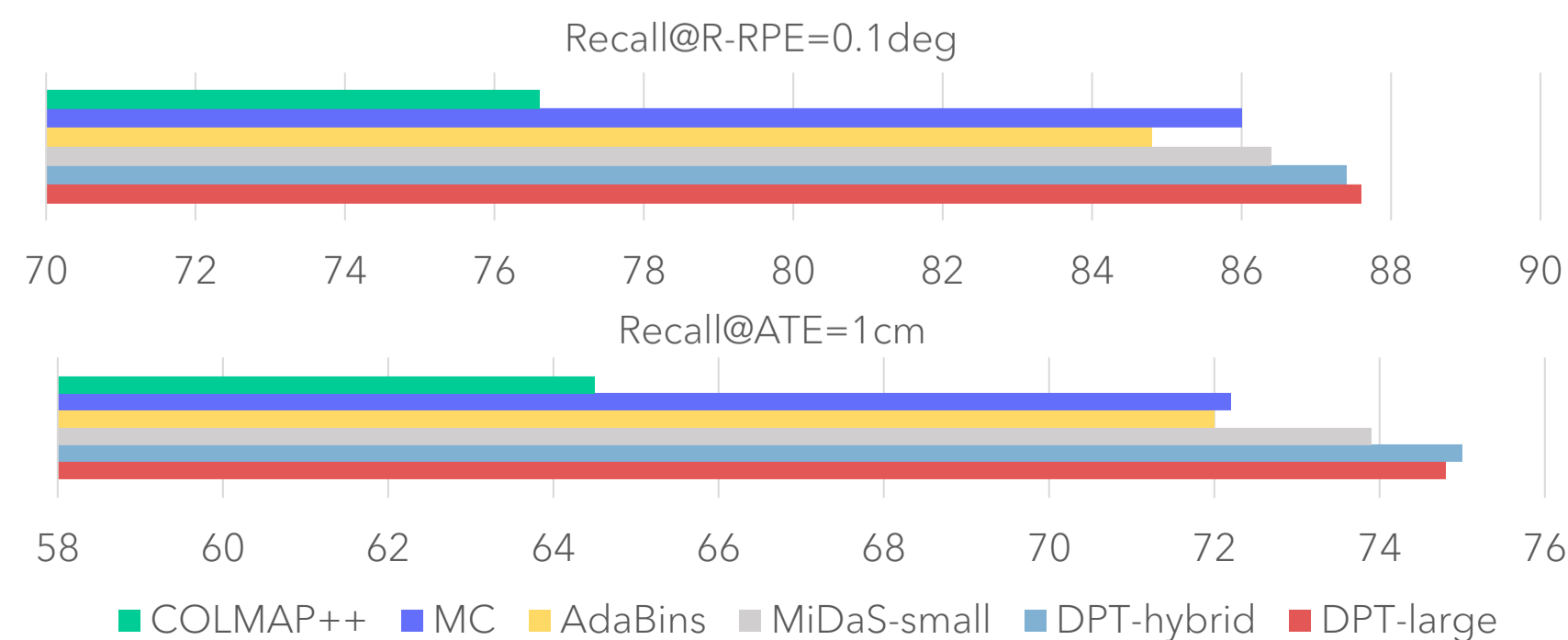
Recall \uparrow our method outperforms existing methods on StudioSfM

- COLMAP, COLMAP++, DeepV2D, DfUSMC, RCVD



Recall \uparrow , our method work with **multiple off-the-shelf** depth estimators

- DPT-large, DPT-hybrid, MiDaS-small, AdaBins, MC



Qualitative results

