MP-SfM: Monocular Surface Priors for Robust Structure-from-Motion

Zador Pataki[1] Paul-Edouard Sarlin[2] Johannes Schönberger[1,3] Marc Pollefeys[1,3]

[1] ETH zürich [2] Google [3] Microsoft

CVPR Nashville JUNE 11-15, 2025

github.com/cvg/mpsfm

Overview

Why do classical Structure-from-Motion pipelines break in low-overlap?

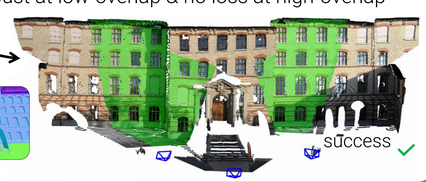
insufficient salient point observations yield **unconstrained relative scales!**

classical Structure-from-Motion

broken reconstruction ✗

MP-SfM leverages monocular surface priors
→ Robust at low-overlap & no loss at high-overlap

monocular depth & normals



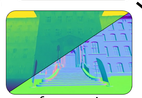
success ✓

Classical SfM pipeline + monocular surface priors

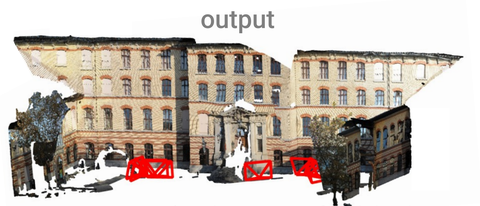
inputs

correspondences

surface priors

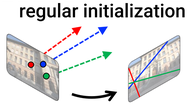


output



camera poses, triangulated points & refined depth maps

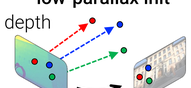
regular initialization



relative pose

low-parallax init

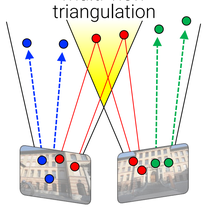
depth



absolute pose

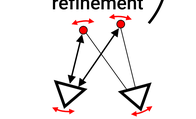
3D lifting

multi-view triangulation



single-view lifting

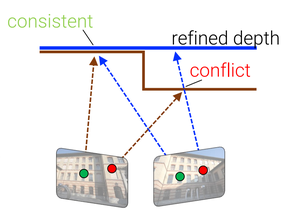
refinement



depth-constrained bundle adjustment

depth refinement

depth consistency check



our SfM pipeline

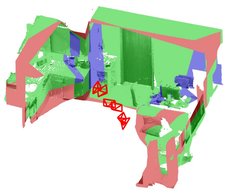
Robust low-overlap reconstruction

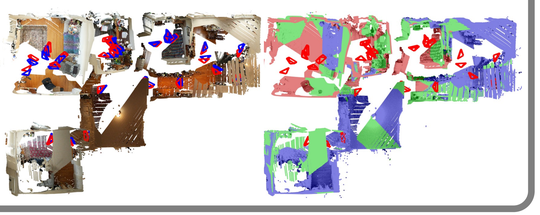
MP-SfM can register images without three-view overlap thanks to monocular priors

input images

estimated vs **ground-truth** camera poses

3 view overlap
2 view overlap
1 view overlap





MP-SfM optimization problem

BA extended using monocular depth and normal constraints

$$\arg\min_{\mathcal{P},\mathcal{X},\mathcal{D}^*} C_{BA} + C_{reg} + C_{int}$$

under constrained points $\mathcal{X}$ regularized by refined depth maps $\mathcal{D}^*$

$$C_{reg} = \sum_{i\in\mathcal{R}}\sum_{j,k} \rho_{reg}\left(\left\|\widehat{D}_i(X_k) - D_i^*(x_j)\right\|^2\right)$$

3D point depth regularization

$$C_{int} = \sum_{i\in\mathcal{R}}\sum_{u,v} \Big[\rho_{prior}\Big(\|D_i^*(u,v) - D_i(u,v)\|^2_{\Sigma_{D_i(u,v)}}\Big) + \rho_{int}\Big(\|N_i(u,v) - \Delta D_i^*(u,v)\|^2_{\Sigma_{N_i(u,v)}}\Big)\Big]$$

depth prior constraint

normals prior constraint

depth refinement via normal integration

efficient solving

$$\arg\min_{\mathcal{D}^*} C_{reg} + C_{int} \;\rightleftharpoons\; \arg\min_{\mathcal{P},\mathcal{X}} C_{BA} + C_{reg}$$

Hessian of BA objective is dense → solve using block coordinate descent

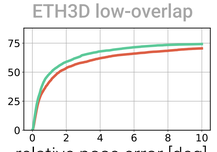
robust depth refinement: Normal integration solved using uncertainties and Cao et. al (2022) to preserve discontinuities

Support for dense correspondences

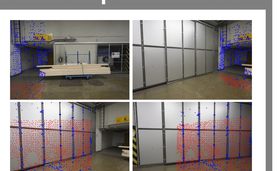
In addition to building 3+ view tracks at salient points

MP-SfM builds depth constrained 2 view tracks

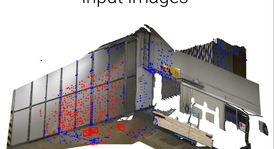
ETH3D low-overlap



relative pose error [deg]



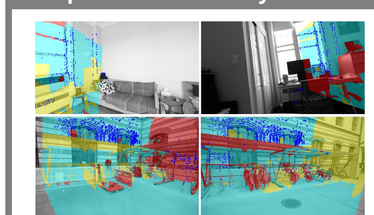
input images



output reconstruction

MP-SfM +
RoMA-sparse
RoMA dense

Depth consistency check

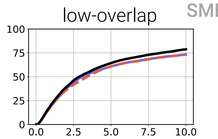


robust registration despite symmetries

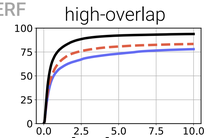
consistent
occlusion
conflict

SMERF

low-overlap



high-overlap



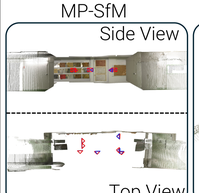
relative pose error [deg]

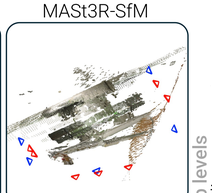
MP-SfM + doppelganger | no DC-check | **with DC-check**

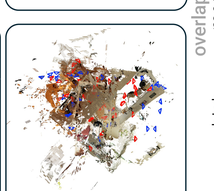
Results: State-of-the-art Structure-from-Motion

MP-SfM

MASt3R-SfM



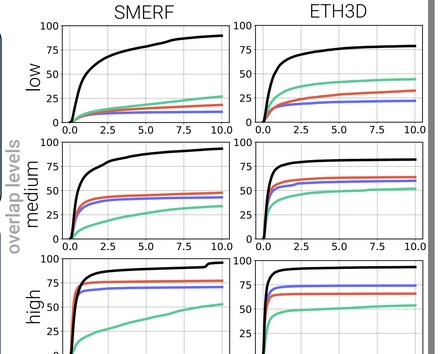




estimated vs **ground-truth** camera poses

SMERF

ETH3D



relative pose error [deg]

COLMAP GLOMAP MASt3R-SfM
MASt3R+MP-SfM