Large Scale Photometric Bundle Adjustment

Interdisciplinary Project

Alexander Gaul

Technical University Munich

September 23, 2021

Main Reference

Large Scale Photometric Bundle Adjustment [5]

Oliver J. Woodford, Edward Rosten

- Joint, photometric optimization over dense geometry and camera parameters
- Variety of lighting conditions and camera intrinsics
- Memory efficient implementation of the Variable Projection optimizer [1]

Project Goals

- Manual implementation of the presented framework
- Evaluation on the Tanks and Temples training datasets [4]
- Comparison of the Variable Projection optimizer with standard Levenberg-Marquardt

Problem Parameters

- Camera intrinsics $f_x, f_y, c_x, c_y, l_1, l_2$

$$\kappa(\mathbf{x}) = \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \mathbf{x} + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \qquad \varphi(\mathbf{x}) = \mathbf{x}(1 + l_1\|\mathbf{x}\|^2 + l_2\|\mathbf{x}\|^4)$$

- Camera poses

$$\{\mathbf{R}_i, \mathbf{t}_i\}_{i=1}^{P}, \qquad \mathbf{R}_i \in \mathbb{SO}(3), \mathbf{t}_i \in \mathbb{R}^3$$

- Landmark parameters

$$\{\mathbf{n}_k\}_{k=1}^{L}, \qquad \mathbf{n}_k \in \mathbb{R}^3$$

- $\bar{\mathbf{\Theta}} = \{\{\mathbf{R}_i, \mathbf{t}_i\}_{i=1}^{P}, f_x, f_y, c_x, c_y, l_1, l_2\} \qquad \mathbf{\Theta} = \bar{\mathbf{\Theta}} \cup \{\mathbf{n}_k\}_{k=1}^{L}$

Landmark Patches

- Visibilities $\mathcal{V}_k$ and source view I_k for landmark k
- 4×4 grid centered on the landmark

$$\{\mathbf{P}_k\}_{k=1}^{L}, \qquad \mathbf{P}_k \in \mathbb{R}^{2\times N}, N = 16$$

- Projection of landmark k from frame i to j

$$\Pi_{ijk}^{\mathbf{\Theta}} : \mathbb{R}^{2\times N} \rightarrow \mathbb{R}^{2\times N}$$

- Normalized Cross Correlation for sampled patch $\bar{\mathsf{I}} = \mathsf{I}(\mathbf{P})$

$$\Psi(\bar{\mathsf{I}}) = \frac{\bar{\mathsf{I}} - \mu_{\bar{\mathsf{I}}}}{\sigma_{\bar{\mathsf{I}}}} \qquad \sigma_{\bar{\mathsf{I}}} = \|\bar{\mathsf{I}} - \mu_{\bar{\mathsf{I}}}\| \qquad \mu_{\bar{\mathsf{I}}} = \frac{\mathbf{1}^\top \bar{\mathsf{I}}}{N}$$

Landmark Patches

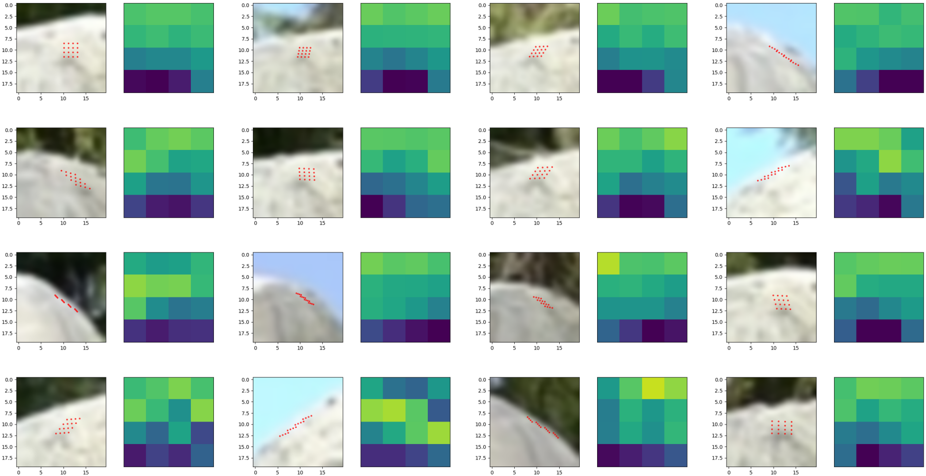


Figure: Grids and normalized patches from different views

Cost Formulation

- Patch Residual

$$\mathcal{E}_{jk} = \Psi(\mathsf{I}_j(\Pi^{\boldsymbol{\Theta}}_{ijk}(\mathbf{P}_k))) - \Psi(\mathsf{I}_i(\mathbf{P}_k)), \qquad i = I_k$$

- Total cost

$$E(\Theta) = \frac{1}{2}\|\mathcal{E}_{\text{reg}}\|^2 + \frac{1}{2}\sum_k \sum_{j \in V_k} \rho(\|\mathcal{E}_{jk}\|^2), \qquad \rho(s) = \frac{s}{s+\tau^2}$$

- Camera regularization

$$\mathcal{E}_{\text{reg}} = 10^5 \left[\frac{f_x - f_y}{f_x + f_y}, \frac{c_x - W/2}{\max(W,H)}, \frac{c_y - H/2}{\max(W,H)}\right]^\top$$

Optimization

- Weighted jacobians

$$\bar{\mathbf{J}}_k = \left[\sqrt{\rho'(\|\mathcal{E}_{jk}\|^2)} \frac{\partial \mathcal{E}_{jk}}{\partial \bar{\boldsymbol{\Theta}}} \right]_{\forall j \in \mathcal{V}_k} \qquad \hat{\mathbf{J}}_k = \left[\sqrt{\rho'(\|\mathcal{E}_{jk}\|^2)} \frac{\partial \mathcal{E}_{jk}}{\partial \mathbf{n}_k} \right]_{\forall j \in \mathcal{V}_k}$$

- Weighted residual

$$\mathcal{E}_k = \left[\sqrt{\rho'(\|\mathcal{E}_{jk}\|^2)} \mathcal{E}_{jk} \right]_{\forall j \in \mathcal{V}_k}$$

Optimization

- Hessian with damping

$$
\begin{aligned}
\mathbf{H}_{\bar{\boldsymbol{\Theta}}} &= \sum_{k=0}^{L} \bar{\mathbf{J}}_k^\top \bar{\mathbf{J}}_k + \lambda \,\mathrm{diag}(\bar{\mathbf{J}}_k^\top \bar{\mathbf{J}}_k) \\
\mathbf{H}_{\mathbf{n}_k} &= \hat{\mathbf{J}}_k^\top \hat{\mathbf{J}}_k + \lambda \,\mathrm{diag}(\hat{\mathbf{J}}_k^\top \hat{\mathbf{J}}_k) \\
\mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}_k} &= \hat{\mathbf{J}}_k^\top \bar{\mathbf{J}}_k
\end{aligned}
$$

- Gradient

$$
\mathbf{g}_{\bar{\boldsymbol{\Theta}}} = \sum_{k=0}^{L} \bar{\mathbf{J}}_k^\top \mathcal{E}_k \qquad \mathbf{g}_{\mathbf{n}_k} = \hat{\mathbf{J}}_k^\top \mathcal{E}_k
$$

Schur Complement

- Linear System

$$\begin{bmatrix} \mathbf{H}_{\bar{\boldsymbol{\Theta}}} & \mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}} \\ \mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}}^\top & \mathbf{H}_{\mathbf{n}} \end{bmatrix} \begin{bmatrix} \delta\bar{\boldsymbol{\Theta}} \\ \delta\mathbf{n} \end{bmatrix} = - \begin{bmatrix} \mathbf{g}_{\bar{\boldsymbol{\Theta}}} \\ \mathbf{g}_{\mathbf{n}} \end{bmatrix}$$

- Solving for $\delta\bar{\boldsymbol{\Theta}}$ we obtain the reduced camera system

$$\delta\bar{\boldsymbol{\Theta}} = -\mathbf{H}_{\mathrm{rcs}}^{-1}\mathbf{g}_{\mathrm{rcs}}$$

$$\mathbf{H}_{\mathrm{rcs}} = \mathbf{H}_{\bar{\boldsymbol{\Theta}}} - \mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}}\mathbf{H}_{\mathbf{n}}^{-1}\mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}}^\top \qquad \mathbf{g}_{\mathrm{rcs}} = \mathbf{g}_{\bar{\boldsymbol{\Theta}}} - \mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}}\mathbf{H}_{\mathbf{n}}^{-1}\mathbf{g}_{\mathbf{n}}$$

- Back-substitution

$$\delta\mathbf{n} = -\mathbf{H}_{\mathbf{n}}^{-1}(\mathbf{g}_{\mathbf{n}} - \mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}}^\top\delta\bar{\boldsymbol{\Theta}})$$

Landmark Updates

- Back-Substitution in Levenberg-Marquardt (LM)

$$\delta \mathbf{n}_k = -\mathbf{H}_{\mathbf{n}_k}^{-1}(\mathbf{g}_{\mathbf{n}_k} - \mathbf{H}_{\bar{\boldsymbol{\Theta}}\mathbf{n}_\mathrm{k}}^{\top} \delta \bar{\boldsymbol{\Theta}})$$

- Embedded Point Iterations (EPIs) for Variable Projection

$$\delta \mathbf{n}_k = -\mathbf{H}_{\mathbf{n}_k}^{-1} \mathbf{g}_{\mathbf{n}_k}$$

COLMAP SfM

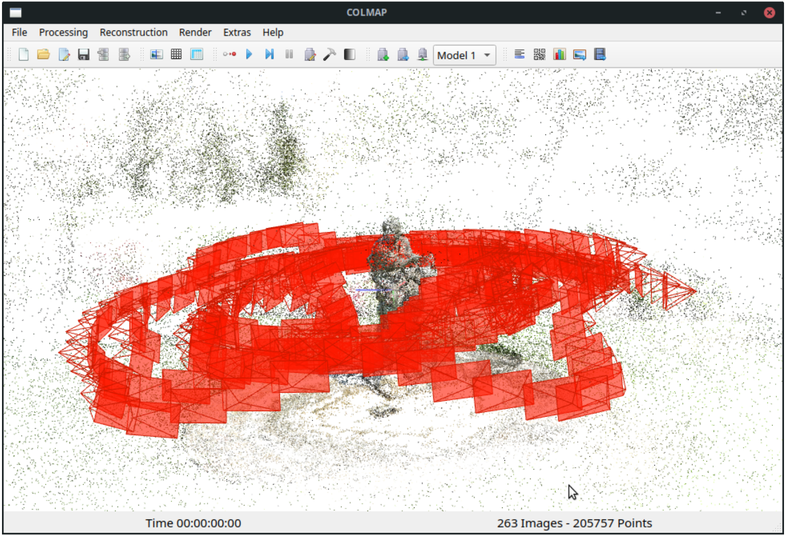


Figure: Colmap sparse reconstruction [2]

COLMAP MVS

Figure: Dense input cloud with background [3]

COLMAP Mesh

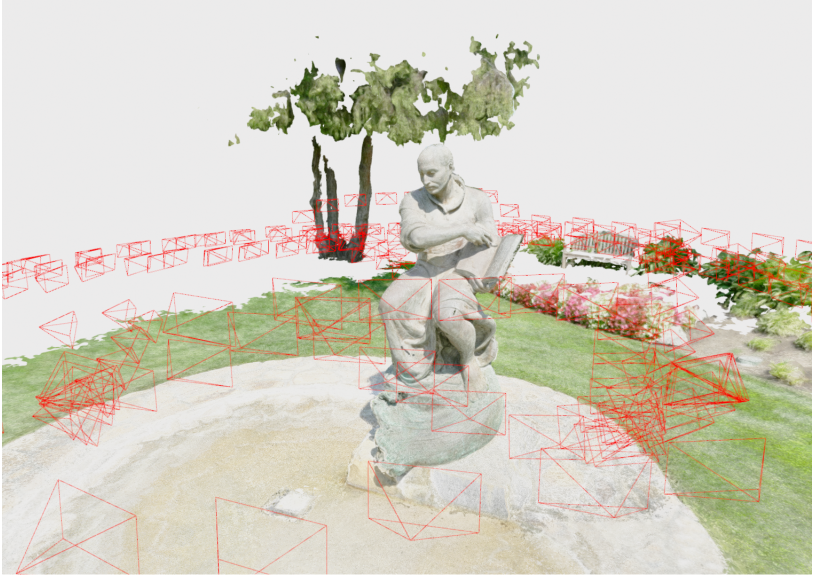


Figure: Mesh to compute visibilities

Preprocessing

- Colmap `automatic_reconstructor` (SfM + MVS)
- Poisson mesh to compute visibilities
- Patch closest to the robust mean determines source view I_k

$$I_k = \arg\min_{j \in \mathcal{V}_k} \|\bar{\mathbf{I}} - \mu\|^2$$

$$\mu = \arg\min_{\hat{\mu} \in \mathbb{R}^N} \sum_{j \in \mathcal{V}_k} \rho(\|\bar{\mathbf{I}} - \hat{\mu}\|^2)$$

- Discard landmarks with insufficient texture $\|\bar{\mathbf{I}} - \mu_{\bar{\mathbf{I}}}\| < 8$

Algorithm Overview

Algorithm 1: Optimization

$\lambda \leftarrow L$; $\omega \leftarrow 10$;
$S_0 \leftarrow E(\mathbf{\Theta}_0)$;
for $t = 1 : 10$ **do**
    $\mathbf{\Theta}_t \leftarrow \mathbf{\Theta}_{t-1}$;
    for $k = 1 : L$ **do**
        Add Landmark k to $\mathbf{H}_{\text{rcs}}$ and $\mathbf{g}_{\text{rcs}}$;
    $\bar{\mathbf{\Theta}}_t \leftarrow \bar{\mathbf{\Theta}}_t \oplus \delta\bar{\mathbf{\Theta}}$; `/* Update camera parameters */`
    for $k = 1 : L$ **do**
        $\mathbf{n}_{kt} \leftarrow \mathbf{n}_{kt} + \delta\mathbf{n}_k$; `/* Update landmark parameters */`
    $S_t \leftarrow E(\mathbf{\Theta_t})$;
    if $S_t < S_{t-1}$ **then**
        $\lambda \leftarrow \lambda/10$; $\omega \leftarrow 10$; `/* Reduce damping */`
    else
        $\lambda \leftarrow \max(\lambda\omega, 10^{-6})$; $\omega \leftarrow 2\omega$; `/* Increase damping */`
        goto 7 ; `/* Retry */`

Tanks and Temples Dataset

- Indoor and outdoor scenes
- Training Data with ground-truth provided

	Truck	Ignatius	Meetingroom	Barn	Caterpillar	Church
Cameras	251	263	371	410	383	507
Landmarks [M]	1.51	1.69	5.35	4.40	4.77	2.92
Observations [M]	122.42	148.56	325.95	358.94	462.09	373.75

Table: Dataset properties after preprocessing

Evaluation Results

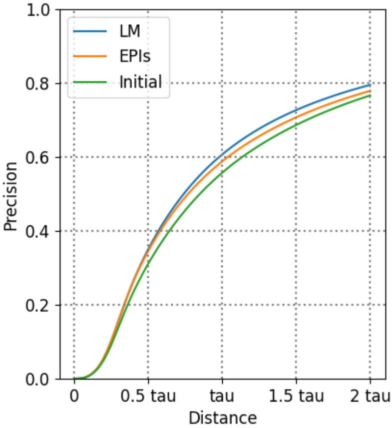


Figure: Average Precision across all datasets

Evaluation Results

	Truck	Ignatius	Meetingroom	Barn	Caterpillar	Church
Initialization	0.555	0.613	0.426	0.385	0.348	0.555
LM	0.612	0.694	0.451	0.408	0.372	0.556
EPIs	0.579	0.691	0.435	0.395	0.357	0.560

Table: Precison-AUC score

Memory and Runtime

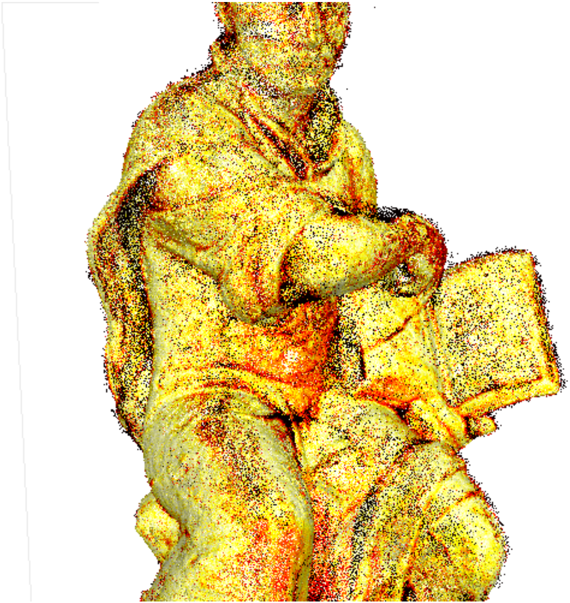
	Truck	Ignatius	Meetingroom	Barn	Caterpillar	Church
LM	33.23	38.41	79.74	87.39	109.00	101.54
EPIs	16.99	19.26	36.41	39.88	47.49	51.73

Table: Memory [GB]

	Truck	Ignatius	Meetingroom	Barn	Caterpillar	Church
LM	272	424	749	928	1408	2171
EPIs	709	1415	1488	1920	2888	3526

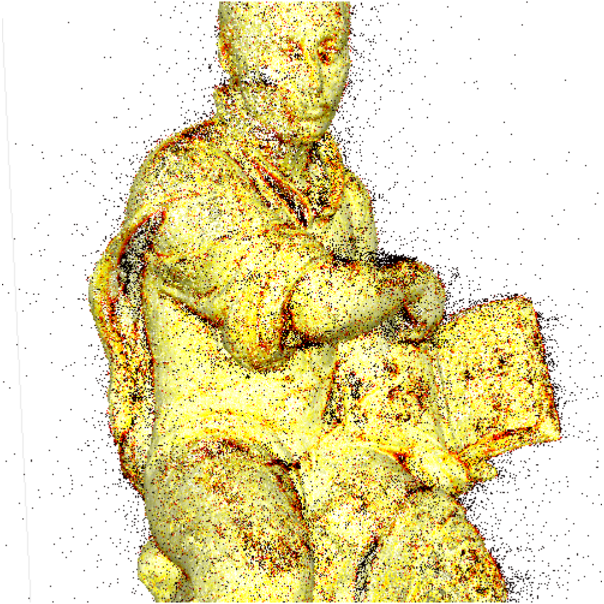
Table: Runtime [min]

Qualitative Results - Ignatius Initial

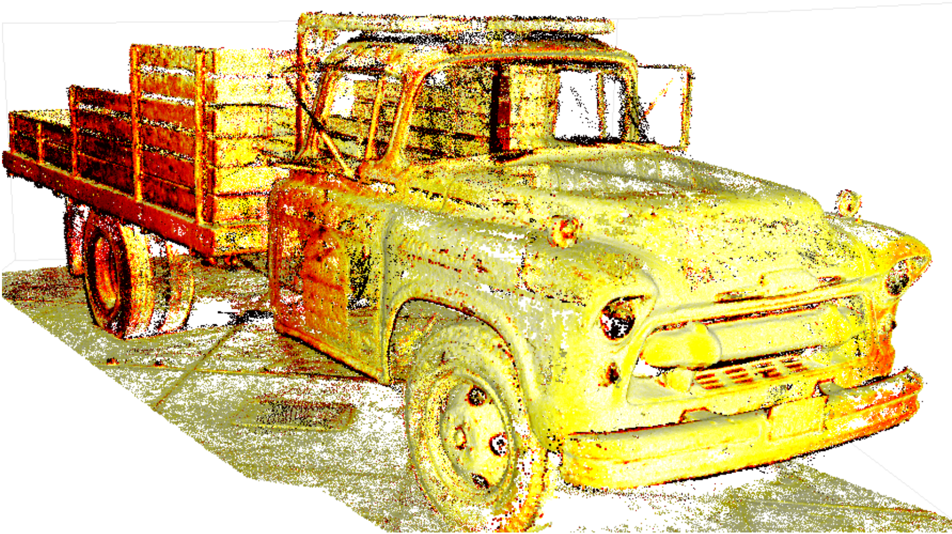


Qualitative Results - Ignatius LM

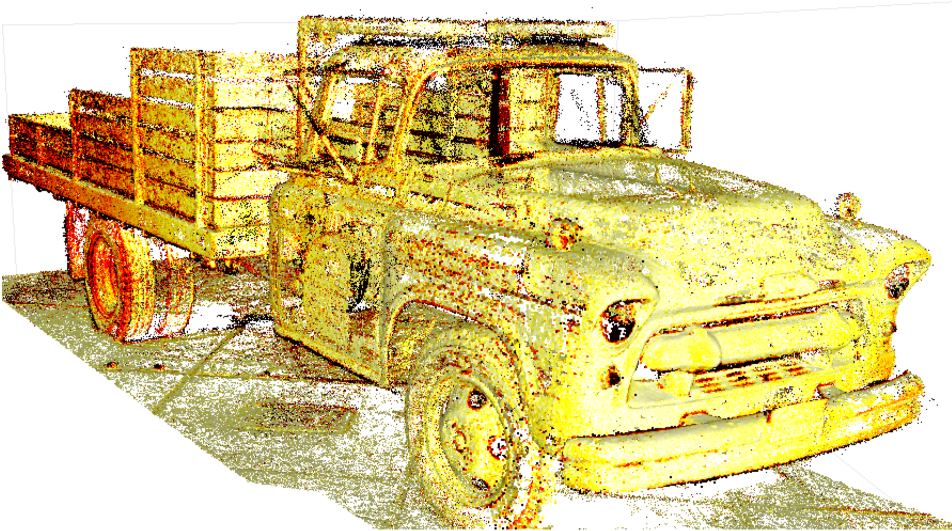
Qualitative Results - Ignatius EPIs



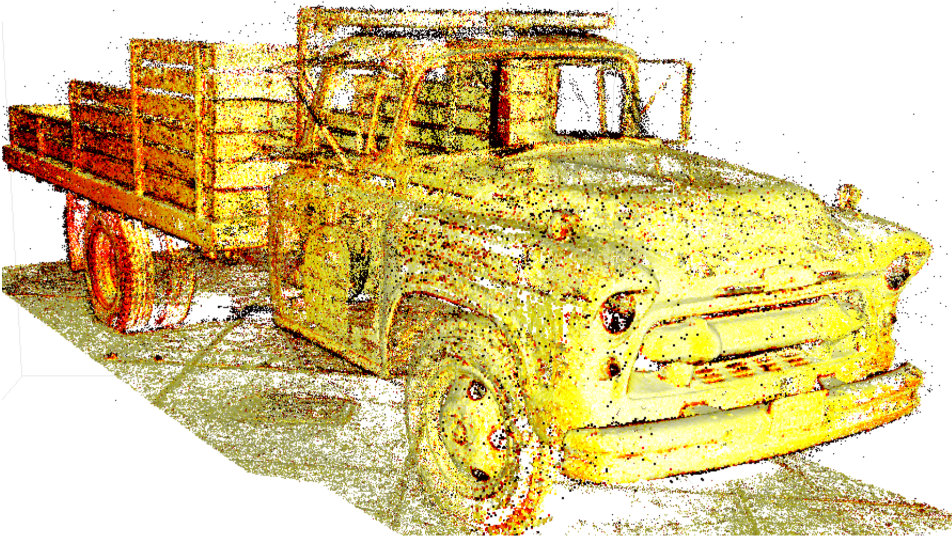
Qualitative Results - Truck Initial



Qualitative Results - Truck LM



Qualitative Results - Truck EPIs



Ablation Study - Levenberg-Marquardt

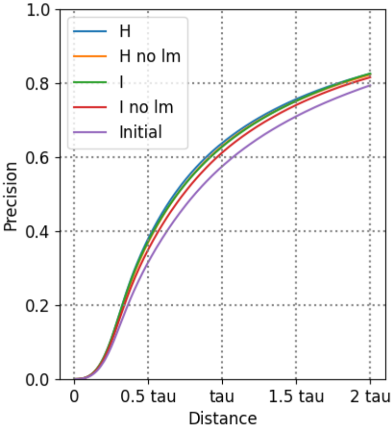


Figure: Damping settings

Ablation Study - EPIs

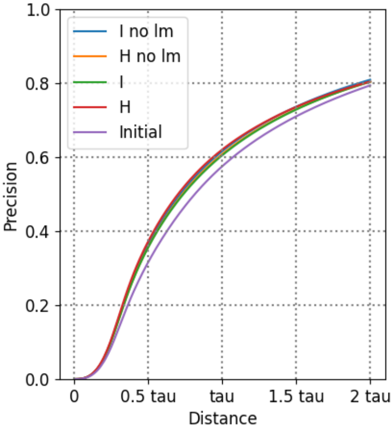


Figure: Damping settings

Conclusion

- Levenberg-Marquardt delivers slightly better quantitative and qualitative results
- Requires about half the time and double the memory
- Further memory reduction possible

References

- J. Hong et al. "Projective Bundle Adjustment from Arbitrary Initialization Using the Variable Projection Method". In: *Lecture Notes in Computer Science* 9905 (2016), pp. 477–493.
- Johannes Lutz Schönberger and Jan-Michael Frahm. "Structure-from-Motion Revisited". In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.
- Johannes Lutz Schönberger et al. "Pixelwise View Selection for Unstructured Multi-View Stereo". In: *European Conference on Computer Vision (ECCV)*. 2016.
- Arno Knapitsch et al. "Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction". In: *ACM Transactions on Graphics* 36.4 (2017).
- Oliver J. Woodford and Edward Rosten. "Large Scale Photometric Bundle Adjustment". In: (2020).