expOSE: Accurate Initialization-Free Projective Factorization using Exponential Regularization

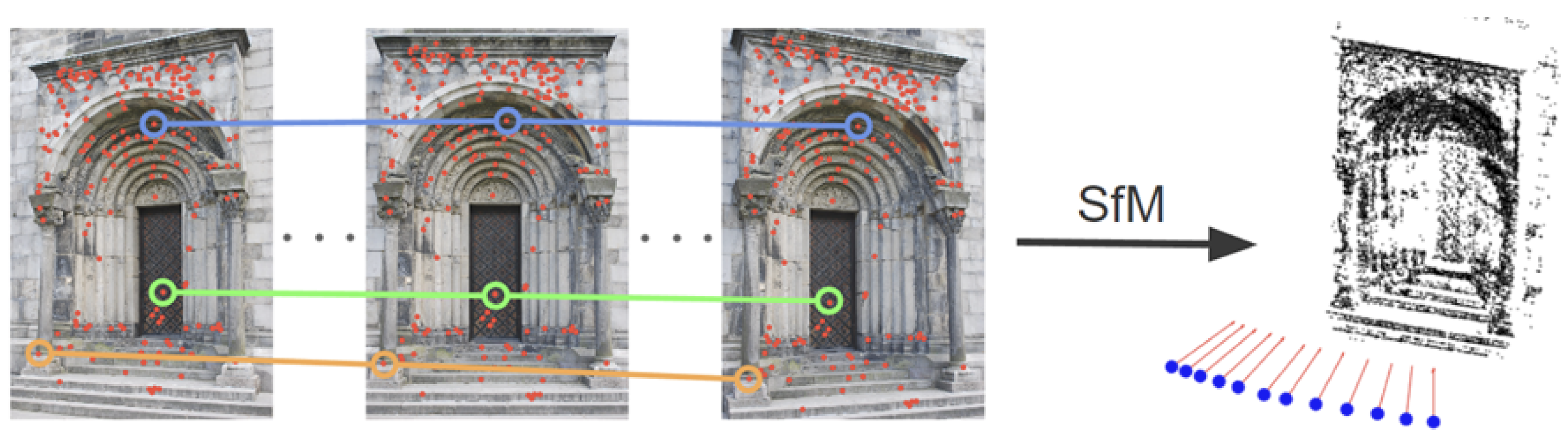
José Pedro Iglesias[1], Amanda Nilsson[2], Carl Olsson[1,2]
[1]Chalmers University of Technology, [2]Lund University

LUND UNIVERSITY

CHALMERS UNIVERSITY OF TECHNOLOGY

Introduction and Motivation

Structure-from-Motion consists of estimating 3D points and camera poses from sets of 2D points tracked amongst several views.



Bundle adjustment (BA) is the preferred method to solve the problem

$$\min_{\mathbf{P},\bar{\mathbf{X}}} \sum_{i,j} \|\mathbf{m}_{ij} - \frac{1}{\mathbf{P}_i^{(3)}\bar{\mathbf{X}}_j}\mathbf{P}_i^{(1:2)}\bar{\mathbf{X}}_j\|^2$$

but its non-convexity only guarantees convergence to nearest minimum.

Pseudo Object Space Error (pOSE) [1] has been suggested to obtain a good approximate solution with VarPro [2] (2nd order method):

$$\ell_{\text{OSE}} = \sum_{ij} \|z_{ij}\mathbf{m}_{ij} - \mathbf{x}_{ij}\|^2$$

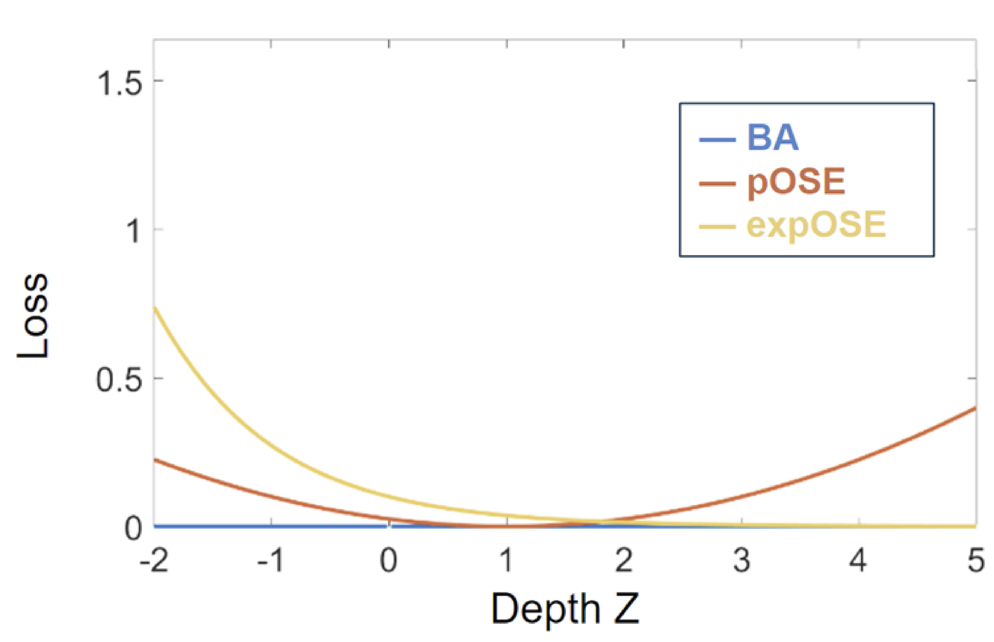
$$\ell_{\text{Affine}} = \sum_{ij} \|\mathbf{m}_{ij} - \mathbf{x}_{ij}\|^2$$

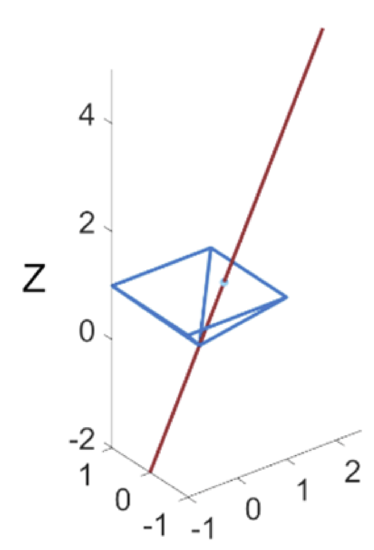
$$\ell_{\text{pOSE}} = (1-\eta)\ell_{\text{OSE}} + \eta\ell_{\text{Affine}}$$

However, we observe that the affine regularization penalizes large depths too much, and propose to use instead an exponential regularization.

$$\ell_{\text{exp}} = \sum_{ij} e^{-\left(\frac{\mathbf{m}_{ij}\mathbf{x}_{ij}+z_{ij}}{\sqrt{\|\mathbf{m}_{ij}\|^2+1}}\right)}$$

$$\ell_{\text{expOSE}} = (1-\eta)\ell_{\text{OSE}} + \eta\ell_{\text{exp}}$$



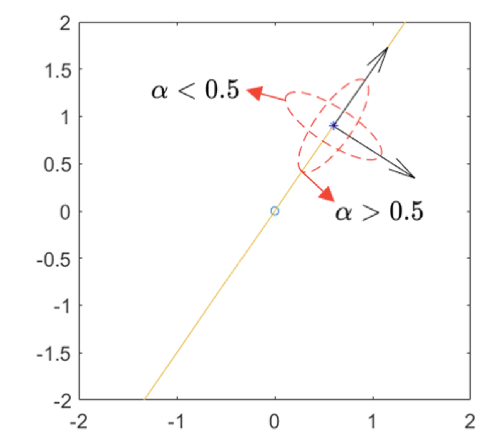


Proposed Method

To achieve robustness to lens distortion and to be able to optimize with VarPro, we propose:

- Objective function: Object Space Error with radial and tangential components weighting (wOSE):

$$\ell_{\text{wOSE}} = \sum_{ij}\left((1-\alpha)\left(\frac{\mathbf{m}_{ij}^T}{\|\mathbf{m}_{ij}\|}\mathbf{x}_{ij} - \|\mathbf{m}_{ij}\|z_{ij}\right)^2 + \alpha\left(\frac{\mathbf{m}_{ij\perp}^T}{\|\mathbf{m}_{ij}\|}\mathbf{x}_{ij}\right)^2\right)$$



Tuning α allows to model lens distortion:

- For $\alpha = 0.5$, wOSE is equivalent to OSE;
- For $\alpha = 1$, wOSE is equivalent to radial OSE in RpOSE[3];
- Values between 0.5 and 1 allow trade-off between the methods stability and accuracy.

- Regularization: quadratic approximation of exponential regularization, needed to solve the problem with VarPro:

$$\tilde{\ell}_{\text{exp}} = \begin{cases} \sum_{ij} \frac{\ell_{\text{exp}}(\bar{\mathbf{x}}_{ij},\bar{z}_{ij})}{2}\left(\frac{\mathbf{m}_{ij}^T\Delta\mathbf{x}_{ij}+\Delta z_{ij}}{\sqrt{\|\mathbf{m}_{ij}\|^2+1}} - 1\right)^2, & \alpha \in [0,1[\\ \sum_{ij} \frac{\ell_{\text{exp}}(\bar{\mathbf{x}}_{ij})}{2}\left(\frac{\mathbf{m}_{ij}^T\Delta\mathbf{x}_{ij}}{\|\mathbf{m}_{ij}\|} - 1\right)^2, & \alpha = 1 \end{cases}$$

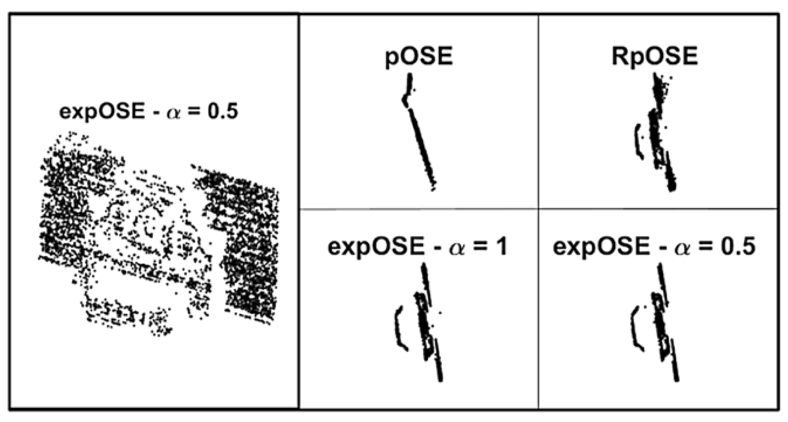
The proposed method consists of the following objective function and quadratic optimization problem:

$$\ell_{\text{expOSE}} = (1-\eta)\ell_{\text{wOSE}} + \eta\tilde{\ell}_{\text{exp}} \implies \min_{\mathbf{P},\bar{\mathbf{X}}} \|\mathcal{A}(\mathbf{P},\bar{\mathbf{X}}) - \mathbf{b}\|^2$$

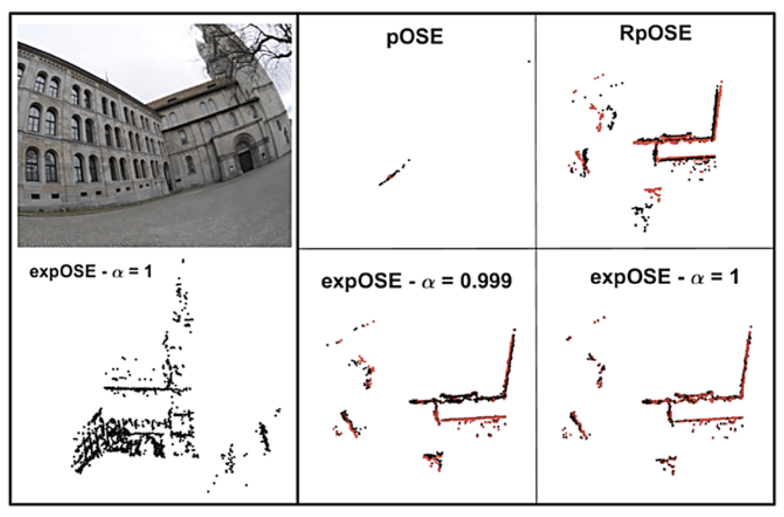
which can be efficiently optimized with VarPro, while keeping a large basin of convergence.

Experimental Results

Use in Structure-from-Motion pipeline and compare against other factorization methods (pOSE, RpOSE), with and without BA refinement.



fountain-p11		Convergence Rate	2D reproj. [pix]
pOSE		100%	3.05
	+ BA	100%	0.52
RpOSE		90%	53.13
	+ BA	90%	0.52
ExpOSE	$\alpha = 0.5$	100%	0.52
	$\alpha = 0.5$ + BA	100%	0.52
	$\alpha = 1$	100%	4.97
	$\alpha = 1$ + BA	100%	0.52



Grossmunster		Conv. Rate	Rot. [deg]	3D [unit]	2D [pix]
pOSE		50%	148.25	0.762	18.48
	+ BA	50%	27.61	0.293	1.50
RpOSE		90%	2.24	0.082	2.91
	+ BA	90%	0.53	0.011	1.48
ExpOSE	α=0.999	100%	44.74	0.227	41.51
	α=0.999+BA	100%	0.43	0.007	1.48
	α=1	100%	0.18	0.004	1.86
	α=1+BA	100%	0.42	0.006	1.48

References

[1] pOSE: Pseudo object space error for initialization-free bundle adjustment. Je Hyeong Hong and Christopher Zach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018

[2] Projective bundle adjustment from arbitrary initialization using the variable projection method. Je Hyeong Hong, Christopher Zach, Andrew W. Fitzgibbon, and Roberto Cipolla. In European Conf. on Computer Vision, 2016

[3] Radial distortion invariant factorization for structure from motion. Jose Iglesias and Carl Olsson. Proceedings of the IEEE International Conference on Computer Vision, 2021

Summary of Contributions and Conclusions

- Exponential regularization that penalizes less large depths than pOSE
- Large basin of convergence, like pOSE
- OSE decomposed into radial and tangential terms enables robustness to lens distortion
- Experimental results show that the proposed method results in more accurate reconstructions

WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM