UNIVERSITY OF CAMBRIDGE [1] KIST [2]
github.com/jhh37/pose

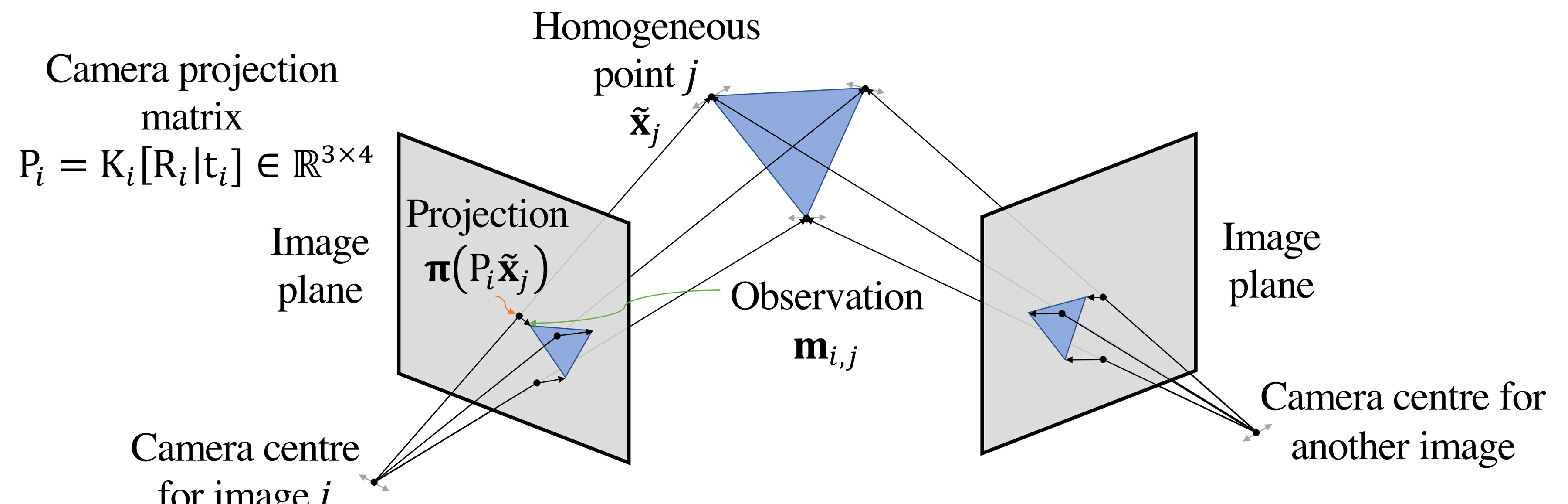
pOSE: Pseudo Objective Space Error for Initialization-Free Bundle Adjustment

Je Hyeong Hong[1→2] Christopher Zach[3]

TOSHIBA Leading Innovation [3]

1. Motivation

Bundle adjustment (**BA**) is a nonlinear optimization process simultaneously refining camera poses and 3D points given a set of 2D image point tracks.



$$\min_{\{P_i\},\{\tilde{\mathbf{x}}_j\}} \sum_{(i,j)\in\Omega} \rho\left(\left\|\boldsymbol{\pi}(P_i\tilde{\mathbf{x}}_j) - \mathbf{m}_{i,j}\right\|_2^2\right) \qquad (1)$$

where $\rho\colon \mathbb{R} \to \mathbb{R}$ is an isotropic loss function.

- BA is known to require good initialization due to its nonlinear objective.
- Various SfM pipelines utilize different initialization schemes. i.e. no gold standard rule.
- A related problem, matrix factorization, does not require careful initialization [1, 2, 3].
- What is the major cause of difficulty in BA? *Can we improve its convergence basin*?

Sources of non-linearity	Potential ways to treat them
1. Bilinear term $\mathbf{P}_i\mathbf{x}_j$	Variable projection (**VarPro**) algorithm [4, 5]
2. Perspective projection $\boldsymbol{\pi}(\cdot)$	*Use of proposed pOSE objective*
3. SO(3) / SE(3) constraints in metric BA	Stratified approach starting from projective BA
4. Robust formulation $\rho(\cdot)$	Not much one can do about this. This work subsequently focuses on the non-robust case.

2. Related work

A. Affine bundle adjustment from arbitrary initialization

- Since affine cameras have centres at infinity, no perspective division and (1) becomes

$$\min_{\{P_i\},\{\tilde{\mathbf{x}}_j\}} \sum_{(i,j)\in\Omega} \left\|P_{i,1:2}\tilde{\mathbf{x}}_j - \mathbf{m}_{i,j}\right\|_2^2 \qquad (2)$$

- (2) is a bilinear problem, which can be efficiently solved using VarPro [1, 2, 3].

B. Object space error

- The projective depth $\mathbf{p}_{i,3}^\top\tilde{\mathbf{x}}_j$ is multiplied to both the projection and observation terms.

$$\min_{\{P_i\},\{\tilde{\mathbf{x}}_j\}} \sum_{(i,j)\in\Omega} \left\|P_{i,1:2}\tilde{\mathbf{x}}_j - (\mathbf{p}_{i,3}^\top\tilde{\mathbf{x}}_j)\mathbf{m}_{i,j}\right\|_2^2 \qquad (3)$$

- (3) is bilinear but is degenerate when $\{P_i\} = \{0\}$.

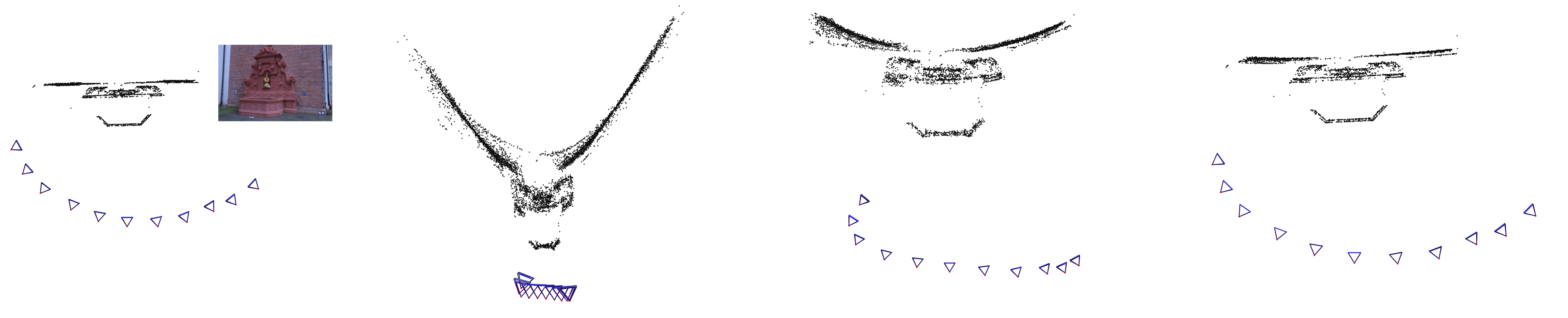
3. Pseudo object space error (pOSE)

We combine (2) and (3) to obtain pseudo object space error (pOSE):

$$\min_{\{P_i\},\{\tilde{\mathbf{x}}_j\}} \sum_{(i,j)\in\Omega} (1-\eta)\left\|P_{i,1:2}\tilde{\mathbf{x}}_j - (\mathbf{p}_{i,3}^\top\tilde{\mathbf{x}}_j)\mathbf{m}_{i,j}\right\|_2^2 + \eta\left\|P_{i,1:2}\tilde{\mathbf{x}}_j - \mathbf{m}_{i,j}\right\|_2^2 \qquad (6)$$

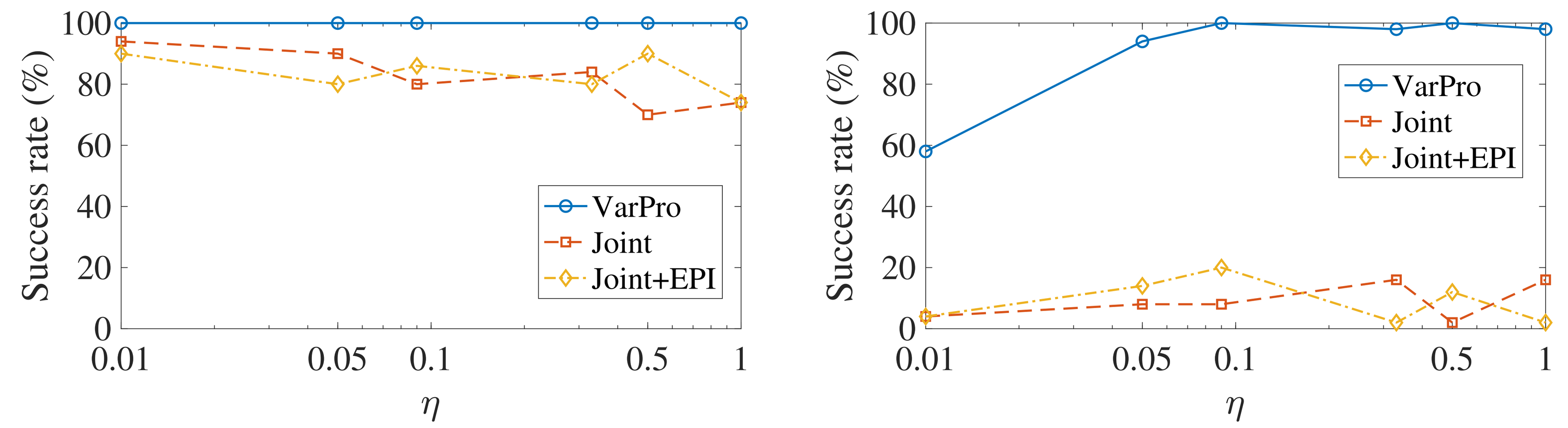
- (6) is bilinear, which can be efficiently solved by VarPro.
- (6) does not have a degenerate solution like (3) does.

A. Diversity of solutions for different η values



(a) Ground truth (b) $\eta = 0.5$ (c) $\eta = 0.1$ (d) $\eta = 0.01$

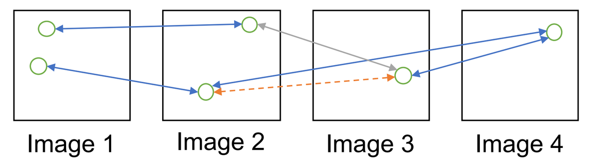
B. Effect of varying η on success rates



(a) Fountain-P11 (b) Small dinosaur

- Some datasets are more sensitive to η than others.
- By performing a grid search of η across datasets listed in **5.D**, we set η to 0.05.

4. Stratified BA strategy



- First minimize pOSE and use this solution to bootstrap BA.
- Since pOSE only has wide convergence basin for L2-norm, inlier point tracks are generated using two-view geometric verification and revisiting newly connected tracks.

A. Stratified BA algorithm

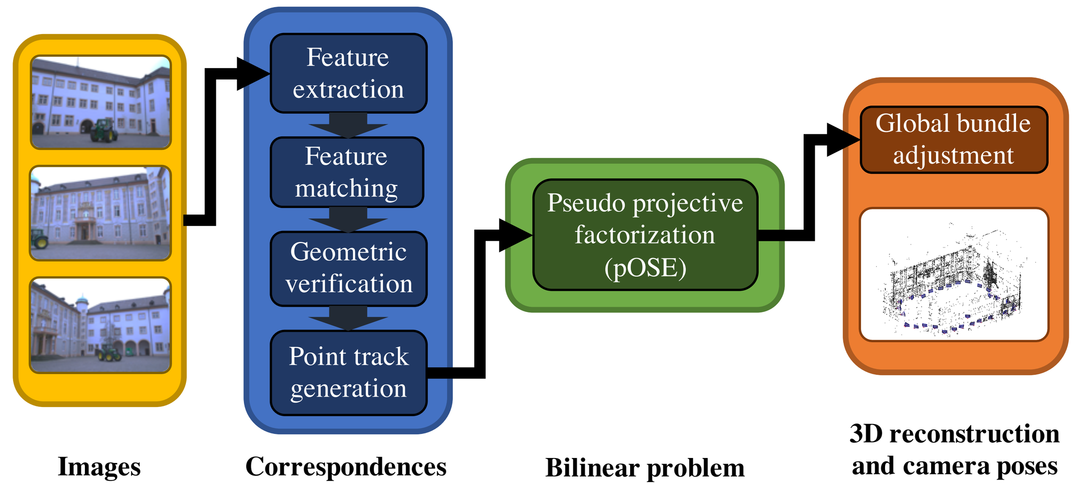
Algorithm 1 Our initialization-free BA pipeline

Input: a set of geometrically-verified point tracks

1. Solve the L^2-norm pOSE problem (see §3) from arbitrarily sampled cameras and points using VarPro.
2. Refine the solution using a robustified projective BA algorithm incorporating nonlinear VarPro, and discard points with large reprojection errors.
3. Upgrade the above solution to metric and discard 3D points which do not satisfy the cheirality constraint.
4. Refine the solution in the metric space.

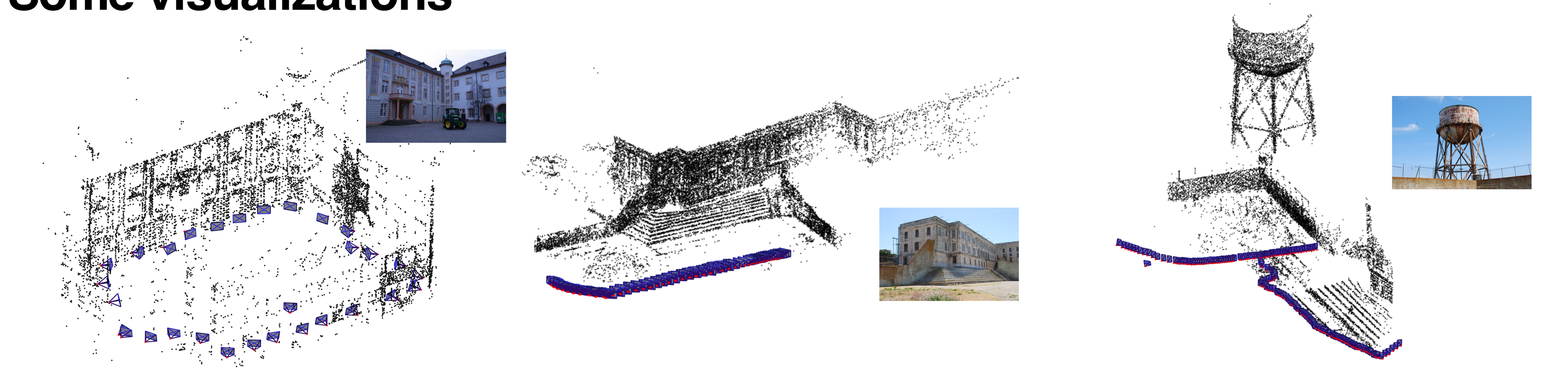
Output: metric camera poses and 3D reconstruction

B. Overall SfM pipeline



5. Experimental results

A. Some visualizations



B. Accuracy comparisons [6]

Seq. Pipeline	pOSE	GPRT	Theia	COLMAP
Fountain-P11	2.8	2.8–4.5	**2.4**	2.8
Entry-P10	7.1	6.4–7.0	**6.0**	6.3
Herz-Jesu-P8	3.4–3.9	3.8–4.5	**3.1**	4.1
Herz-Jesu-P25	5.2	5.1	**5.1**	5.2
Castle-P19	**24.5**–41.1	N/A	25.3	24.9
Castle-P30	**21.7**–26.0	N/A	**21.7**	23.2

C. Small classic datasets

			Fill	pOSE		GPRT	
Seq.	# img	# pts	(%)	SRI	$\bar{t}(s)$	SRI	$\bar{t}(s)$
House	10	672	42.4	100	4.2	35	5.8
Corridor	11	737	49.8	100	1.7	15	9.0
Dino (S)	36	319	23.1	96	3.0	0	10.1
Dino (L)	36	4983	9.2	98	6.9	0	32.3
Wilshire	190	411	39.3	100	38.1	0	272.2
Blue bear	196	2480	19.3	80	70.7	0	337.7

D. Other small to medium-sized datasets

				pOSE			GPRT		
Sequence	# img	# pts	Fill (%)	SRI (%)	SRF (%)	$\bar{t}$ (s)	SRI (%)	SRF (%)	$\bar{t}$ (s)
Fountain-P11	11	9181	50.3	100	100	6.5	15	30	70.3
Entry-P10	10	4270	55.5	100	98	5.2	25	5	27.5
Herz-Jesu-P8	8	3553	60.7	100	100	3.4	30	15	16.1
Herz-Jesu-P25	25	12469	27.3	100	100	12.9	10	5	86.2
Castle-P19	19	5144	27.2	94	88	21.4	0	0	21.3
Castle-P30	30	11531	20.4	94	100	32.1	0	0	73.6
House Martenstorget	12	5934	53.2	100	100	11.3	20	15	47.8
Lund Cathedral (small)	17	9400	30.6	92	96	19.2	10	15	75.6
Gustav II Adolf	57	9562	12.3	100	86	23.7	20	35	82.0
Univ. of West. Ontario	57	6742	14.4	100	98	25.2	5	35	92.6
Vercingetorix	69	5231	10.3	78	92	15.0	10	15	60.0
Lund University Sphinx	70	22770	9.9	96	76	74.7	0	5	150.3
Alcatraz courtyard	133	31558	9.7	94	100	145.5	0	0	1653.8
Water tower	173	25531	8.0	90	76	274.6	40	35	1932.1
Pumpkin	209	25962	4.1	100	100	147.5	0	0	869.7

7. Conclusions

- pOSE improves the basin of convergence for bundle adjustment when using the variable projection (VarPro) method given inlier 2D point tracks.
- pOSE allows to avoid poor local minima in structure-from-motion (SfM) computation, allowing one to focus on feature matching and point track generation.

Acknowledgement

- Christ's College and Roberto Cipolla for travel support.

References

[1] T. Okatani et al., ICCV 2011
[2] P. F. Gotardo and A. Martinez., TPAMI
[3] J.H. Hong and A.W. Fitzgibbon, ICCV 2015
[4] G. Golub and V. Pereyra, SINUM 1973
[5] J.H. Hong et al., CVPR 2017
[6] Strecha et al., CVPR 2008