expOSE: Accurate Initialization-Free Projective Factorization using Exponential Regularization (Supplementary Material)

José Pedro Iglesias[1], Amanda Nilsson[2], Carl Olsson[1,2]

[1]Chalmers University of Technology, Sweden
[2]Lund University, Sweden

1. pOSE and expOSE Level sets

In Figure 1 (a)-(c) we illustrate what level sets of the proposed expOSE formulation look like. Here we have drawn a two-dimensional version where the x-axis corresponds to an image coordinate and the y-axis corresponds to depth. We can think of the camera as being located in the origin facing the y-direction, as shown by the black arrow in the center of the image. The image measurement $\boldsymbol{m}$ is the blue star located at $(0.25, 1)$ and the red line shows all points with zero reprojection error. The solid blue curves in (a)-(c) are level sets of ℓ_{expOSE} for the settings $\eta = 0.5, 0.1$ and 0.05. For comparison, we also plot the level sets of the quadratic approximation, around $(0.25, 1)$, of ℓ_{expOSE}. In Figure 1 (d)-(f) we plot corresponding level sets for ℓ_{pOSE}. Since pOSE is quadratic these are ellipses. Note that with the same setting η ℓ_{pOSE} and ℓ_{expOSE} have the same weight on the OSE term, which means that, when η is small as in (c) and (f), they should have roughly the same size in the direction perpendicular to the red zero-reprojection-line. It is clear that expOSE generally enforces a higher penalty for points that are behind the camera than pOSE does in all cases. Additionally, for large values of η, as in (d), the half axis of the ellipse does generally not align with the line of zero reprojection error. This is due to the fact that the affine term does not measure a distance that is perpendicular to the distance measured by the pOSE term, which our exponential term and its approximation do.

2. Weighting radial and tangential OSE components

The main reason behind introducing different weights for the radial and tangential components of the OSE is, as explained in the main paper, to increase the robustness of the proposed method to radial distortion. In the ideal case, the weight α is set to 1 and radial distortion invariance is achieved since the tangent component of the OSE is unbiased. This is what is done in Section 5 of the paper, with high-quality reconstruction being retrieved by expOSE. These results raise the question of what could be the motivation to use $1/2 < \alpha < 1$.

Even though $\alpha = 1$ results in an unbiased estimation of the factors, it is also true that by doing so fewer data is used for the estimation of those same factors, since half of it, i.e. the radial component of the OSE, is dropped from the loss. This might not only result in lower-quality reconstructions but also in a less stable algorithm in cases where not a lot of data is available.

In this experiment, we investigate the trade-off between reconstruction accuracy and convergence rate of expOSE for different values of α, as a function of the amount of data available for estimation. We use modified versions of the Fountain dataset by setting the maximum amount of viewpoints per point tracked to a certain value, thus controlling the amount of available data per point, as well as synthetically adding radial distortion for further evaluation of the effect of weight α. The procedure to generate the modified sequences is as follows

1. Randomly select a subset of 750 points from the original dataset (for faster optimization);

2. Generate 5 versions of the dataset where each point is seen at most 11, 8, 6, 5, and 4 times. If in the original sequence, a point is seen more than K times, then we randomly remove image measurements of that point until only K measurements are available;

3. For each of the 5 previously mentioned versions of the dataset, generate 2 new versions with synthetic radial distortion, where distortion is applied as

$$\boldsymbol{m}_{ij}^d = (1 + k\|\boldsymbol{m}_{ij}\|^2)\boldsymbol{m}_{ij}, \tag{1}$$

where $\boldsymbol{m}_{ij}^d$ and $\boldsymbol{m}_{ij}$ are the distorted and undistorted (original) image measurement, and $k = \{-10^{-7}, -3 \times 10^{-7}\}$. Examples of the distortion synthetically added to the sequence are shown in Figure 2.

From this, we get 15 versions (5 max views per point $\times$ 3 distortion levels) of the Fountain dataset, for which we

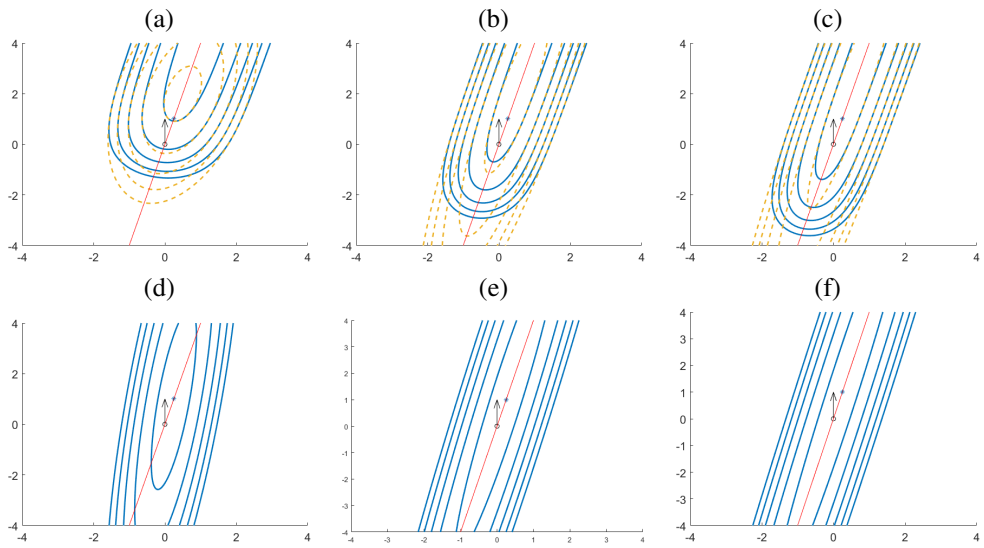


Figure 1. (a)-(c): Level sets of a two dimensional version of ℓ_{expOSE} (solid blue curves) and its approximation (dached orange curves) at $\boldsymbol{m} = (0.25, 0)$. (a): $\eta = 0.5$, (b): $\eta = 0.1$ and (c): $\eta = 0.05$.
(d)-(f): Level sets of a two dimensional version of ℓ_{pOSE} (solid blue curves) with $\boldsymbol{m} = (0.25, 0)$. (d): $\eta = 0.5$, (e): $\eta = 0.1$ and (f): $\eta = 0.05$.

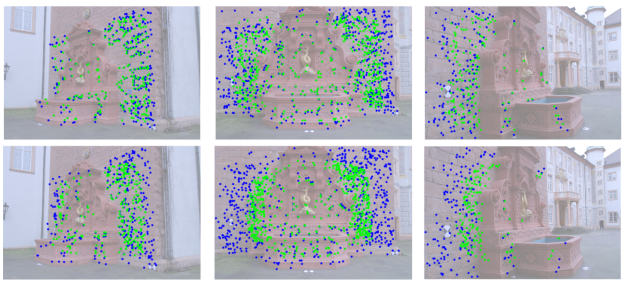


Figure 2. Examples of 3 images with radially distorted image points overlapping for $k = -10^{-7}$ (top) and $k = -3 \times 10^{-7}$ (bottom). The distorted points are shown in green, while the undistorted ones are shown in blue.

Figure 3. Four examples of images in the Grossmunster (top), Kirchenge (middle), and Munsterhof (bottom) sequences.

calculate convergence rates and 3D reconstruction errors through projective registration to GT as done in Section 3.2 of the paper. For each sequence, method, and α, 100 problem instances starting from random solutions are evaluated. The results are presented in Table 1.

As the amount of image data per points is reduced, the convergence rate of the purely radial model $\alpha = 1$ is highly affected, while for other values of α the effect is not as severe. For the sequences without radial distortion, we see the effect of using fewer data for the estimation of factors on the higher 3D error for $\alpha = 1$. When there is radial distortion, the bias added by the radial component of the OSE results in higher errors for $\alpha < 1$, with the optimal trade-off between convergence rate and 3D error being located within the range $\alpha = 0.9$ to 0.99.

3. Additional reconstructions using expOSE

In this section, we present some additional reconstruction results on several other datasets using expOSE and some competing factorization methods.

3.1. Grossmunster, Kirchenge and Munsterhof

We start by showing in Figure 3 some of the images in the Grossmunster, Kirchenge, and Munsterhof sequences used in Section 5.1 of the paper. In Figure 4 we show reconstruction visualizations for the Kirchenge and Munsterhof sequences, as done in the paper for the Grossmunster sequence. In Table 2 we an extended version of the results

Table 1. Values of convergence rate and relative 3D error for evaluation of the effect of different values of α under different amounts of available data. Three levels of radial distortion are tested: $k = 0$, $k = -1 \times 10^{-7}$, and $k = -3 \times 10^{-7}$.

	Convergence Rate [%]					3D error [%]				
$k = 0$	11views	8views	6views	5views	4views	11views	8views	6views	5views	4views
$\alpha = 0.5$	100	100	100	99	100	0.44	0.44	0.44	0.44	0.44
$\alpha = 0.7$	100	99	100	100	100	0.44	0.44	0.44	0.44	0.44
$\alpha = 0.9$	100	100	100	96	99	0.44	0.44	0.44	0.44	0.44
$\alpha = 0.99$	100	98	99	97	95	0.43	0.43	0.43	0.43	0.43
$\alpha = 0.999$	71	67	68	64	62	0.45	0.45	0.45	0.44	0.44
$\alpha = 1$	77	83	82	47	10	0.68	0.68	0.69	0.68	0.76
	Convergence Rate [%]					3D error [%]				
$k = -1 \times 10^{-7}$	11views	8views	6views	5views	4views	11views	8views	6views	5views	4views
$\alpha = 0.5$	100	100	100	100	99	0.52	0.52	0.52	0.52	0.52
$\alpha = 0.7$	100	100	100	100	99	0.52	0.52	0.52	0.52	0.52
$\alpha = 0.9$	100	100	99	100	99	0.52	0.52	0.52	0.52	0.52
$\alpha = 0.99$	92	96	100	97	94	0.51	0.51	0.51	0.51	0.51
$\alpha = 0.999$	84	79	70	58	33	0.47	0.47	0.47	0.52	0.48
$\alpha = 1$	94	95	86	46	12	0.82	0.88	0.80	2.08	1.31
	Convergence Rate [%]					3D error [%]				
$k = -3 \times 10^{-7}$	11views	8views	6views	5views	4views	11views	8views	6views	5views	4views
$\alpha = 0.5$	100	100	100	98	99	0.81	0.81	0.82	0.81	0.82
$\alpha = 0.7$	98	100	100	100	100	0.81	0.81	0.82	0.81	0.82
$\alpha = 0.9$	99	100	100	100	99	0.81	0.81	0.81	0.81	0.82
$\alpha = 0.99$	99	100	100	100	99	0.75	0.75	0.76	0.77	0.78
$\alpha = 0.999$	82	82	71	47	59	0.58	0.58	0.60	0.61	0.65
$\alpha = 1$	95	96	87	82	38	0.73	0.73	0.76	0.76	0.89

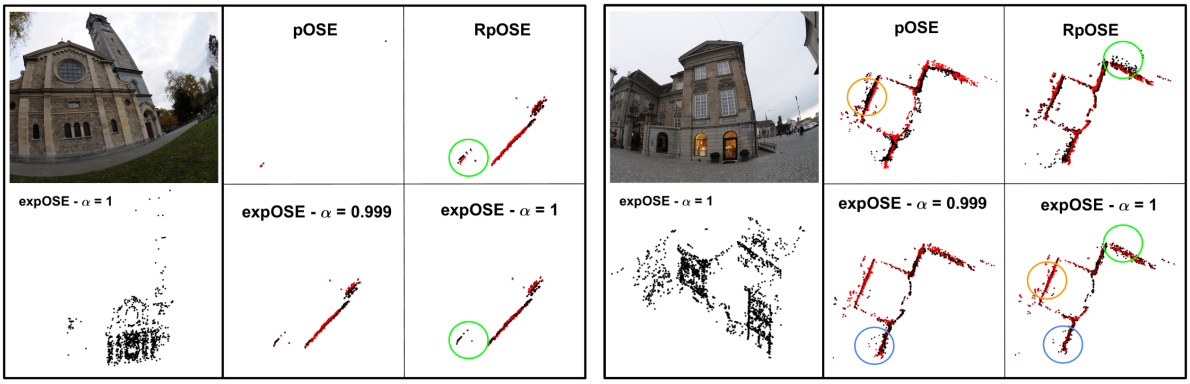


Figure 4. Visualization of reconstructions on the Kirchenge facade and Munsterhof sequence. (Left) An example of one of the images on the sequence. At the bottom, we show a view of the 3D reconstruction of expOSE for $\alpha = 1$. (Right) Comparison between the top view reconstructions (black) obtained with pOSE, RpOSE and expOSE. In red we show the ground-truth 3D point cloud. All reconstructions shown here were not refined with bundle adjustment. We mark with circles of different colors the most relevant visual improvements of expOSE with $\alpha = 1$ compared to the other methods.

presented in Section 5.1 of the paper, in particular for values of $\alpha = 0.5$ and $\alpha = 0.9$.

3.2. TUM sequences

In this section we present additional results on the initial frames of sequences 25 (30cams, 4309pts, 22% missing data), 29 (30cams, 4682pts, 28% missing data), and 31 (15cams, 6669pts, 12% missing data) of the TUM dataset [1]. The pipeline and parameters for expOSE, pOSE, and RpOSE are the same as the ones mentioned in the experiments in the previous section. These sequences do not have ground-truth points-clouds, so we perform Eucliden

Table 2. Results on the Grossmunster, Kirchenge, and Munsterhof datasets (over 10 instances). For each method two rows are presented: the first consists of the results for the output of the factorization method; the second of the output of the Bundle Adjustment (+BA). In green, we show the best results for each metric.

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 $\alpha = 0.5$	60%	105.29	0.626	18.29
ExpOSE $\alpha = 0.5$ + BA	30%	76.82	0.782	1.51
ExpOSE $\alpha = 0.9$	100%	77.02	1.182	22.99
ExpOSE $\alpha = 0.9$ + BA	40%	0.44	0.008	1.48
ExpOSE α=0.999	100%	44.74	0.227	41.51
ExpOSE α=0.999+BA	100%	0.43	0.007	1.48
ExpOSE α=1	100%	0.18	0.004	1.86
ExpOSE α=1+BA	100%	0.42	0.006	1.48
Kirchenge				
pOSE	100%	160.38	6.844	14.95
+ BA	100%	0.72	0.024	1.22
RpOSE	90%	0.98	0.062	1.94
+ BA	90%	1.06	0.031	1.22
ExpOSE $\alpha = 0.5$	70%	43.20	1.775	14.45
ExpOSE $\alpha = 0.5$ + BA	90%	0.78	0.015	1.22
ExpOSE $\alpha = 0.9$	20%	40.21	1.814	17.42
ExpOSE $\alpha = 0.9$ + BA	70%	0.80	0.015	1.22
ExpOSE α=0.999	60%	24.71	0.022	45.28
ExpOSE α=0.999+BA	80%	1.19	0.021	1.22
ExpOSE α=1	80%	0.51	0.026	1.57
ExpOSE α=1+BA	80%	2.92	0.050	1.22
Munsterhof				
pOSE	100%	14.01	0.230	12.08
+ BA	100%	0.44	0.027	1.70
RpOSE	60%	1.00	0.071	11.96
+ BA	60%	0.44	0.027	1.70
ExpOSE $\alpha = 0.5$	100%	12.43	0.244	11.65
ExpOSE $\alpha = 0.5$ + BA	100%	0.47	0.029	1.70
ExpOSE $\alpha = 0.9$	90%	14.14	0.198	14.35
ExpOSE $\alpha = 0.9$ + BA	100%	0.47	0.029	1.70
ExpOSE α=0.999	100%	20.13	0.021	47.71
ExpOSE α=0.999+BA	100%	0.47	0.029	1.70
ExpOSE α=1	80%	0.12	0.013	3.43
ExpOSE α=1+BA	90%	0.45	0.030	1.70

Table 3. Results on the sequences 25, 29 and 31 of the TUM dataset [1] (over 10 instances). For each method two rows are presented: the first consists of the results for the output of the factorization method; the second for the output of the Bundle Adjustment (+BA). In green we show the best results for each metric.

Sequence 25	Conv. Rate	Rot. [deg]	Trans. [%]	2D [pix]
pOSE	80%	173.01	3.76%	2.65
+ BA	80%	1.97	0.24%	0.72
RpOSE	80%	9.06	1.82%	5.24
+ BA	90%	1.93	0.25%	0.72
ExpOSE $\alpha = 0.5$	60%	175.58	3.59%	2.53
ExpOSE $\alpha = 0.5$ + BA	90%	3.30	0.17%	0.72
ExpOSE $\alpha = 0.9$	60%	175.27	3.23%	2.95
ExpOSE $\alpha = 0.9$ + BA	60%	2.79	0.19%	0.72
ExpOSE $\alpha = 0.999$	60%	174.11	3.16%	4.34
ExpOSE $\alpha = 0.999$ + BA	100%	2.19	0.23%	0.72
ExpOSE $\alpha = 1$	90%	7.58	1.55%	4.28
ExpOSE $\alpha = 1$ + BA	90%	2.77	0.20%	0.72
Sequence 29				
pOSE	100%	39.19	4.18%	2.12
+ BA	100%	0.32	0.12%	0.28
RpOSE	100%	6.91	1.76%	6.43
+ BA	100%	0.32	0.12%	0.28
ExpOSE $\alpha = 0.5$	100%	44.69	4.86%	2.08
ExpOSE $\alpha = 0.5$ + BA	100%	0.30	0.11%	0.28
ExpOSE $\alpha = 0.9$	100%	48.86	4.10%	2.44
ExpOSE $\alpha = 0.9$ + BA	100%	0.32	0.11%	0.28
ExpOSE $\alpha = 0.999$	90%	48.85	3.55%	5.60
ExpOSE $\alpha = 0.999$ + BA	100%	0.34	0.11%	0.28
ExpOSE $\alpha = 1$	50%	0.54	0.33%	1.86
ExpOSE $\alpha = 1$ + BA	70%	0.32	0.12%	0.28
Sequence 31				
pOSE	90%	88.79	6.84%	0.72
+ BA	90%	1.14	0.26%	0.23
RpOSE	50%	19.19	2.69%	2.17
+ BA	50%	1.14	0.26%	0.23
ExpOSE $\alpha = 0.5$	90%	87.25	6.92%	0.70
ExpOSE $\alpha = 0.5$ + BA	90%	1.13	0.27%	0.23
ExpOSE $\alpha = 0.9$	80%	100.77	7.08%	0.79
ExpOSE $\alpha = 0.9$ + BA	80%	1.14	0.26%	0.23
ExpOSE $\alpha = 0.999$	40%	79.95	9.42%	4.93
ExpOSE $\alpha = 0.999$ + BA	50%	1.13	0.27%	0.23
ExpOSE $\alpha = 1$	90%	1.02	0.39%	1.38
ExpOSE $\alpha = 1$ + BA	90%	1.25	0.25%	0.23

registration to the ground-truth camera centers instead. The image points tracks are built using optical flow with bidirectional coherence filtering, starting from SIFT-detected features in the initial frame. The 3D error metric is replaced by relative translation error, computed as $e_{\text{translation}} = \sum_i \|c'_i - c_i^{\text{GT}}\|/L$, where c_i is the estimated camera center of the ith view after registration to GT, and L is the length of the GT camera path. In Figure 5 we present some visualizations of the reconstructed sequences for pOSE, RpOSE, and expOSE ($\alpha = 0.999$ and $\alpha = 1$).

The results validate our approach, with expOSE ($\alpha = 1$) outperforming both pOSE and RpOSE before refinement

while keeping a similar convergence rate. In Figure 5 (left) it is also possible to see the quality of the obtained 3D reconstruction before the refinement.

3.3. Other benchmark datasets

Additionally, we also present reconstruction results for datasets/sequences [2, 3] without radial distortion. Without radial distortion, step 2 in the pipeline described in Section 5 can be skipped for $\alpha = 0.5$, and only the third rows of the camera matrices are estimated for $\alpha = 1$. The sequences used are house_martenstorget (12cams, 7500pts, 59% missing data), lund_cath_small (17cams, 7500pts, 74% missing data), herz-jesu-p8 (8cams, 6552pts, 50% missing data), castle-p19 (19cams, 7500pts, 78% missing data), fountain-p11 (11cams, 7500pts, 57% missing data), and Alcatraz Countyard (133cams, 7500, 11% missing data). Some of these sequences have more than 7500 pts detected over all views, but we capped it to 7500 for faster inference.

For these sequences, we only present convergence rate and 2D reprojection errors since some of the datasets do not have ground-truth reconstructions. The results are presented in Table 4 and Figure 6. The conclusions are similar to the previous experiments, with the exception that in this case, without radial distortion, expOSE with $\alpha = 0.5$ had the best performance with reprojection errors extremely close to the BA refined solution. It is also possible to notice that radial models resulted in higher errors than the pinhole models as they use fewer data for the estimation as explained in Section 2. The results also show that expOSE for $\alpha = 0.5$ and $\alpha = 1$ consistently outperform pOSE and RpOSE, respectively, hence validating the hypothesis that exponential regularization results in more accurate factorization in the Structure-from-Motion context.

The sequence castle-p19 was the only with a low convergence rate for all methods, including expOSE. This sequence has two almost disjoint sub-sequences with the intersecting views between them having very few detections. This makes the problem close to degenerate and consequently harder to solve, resulting in lower convergence rates when initializing from random solutions.

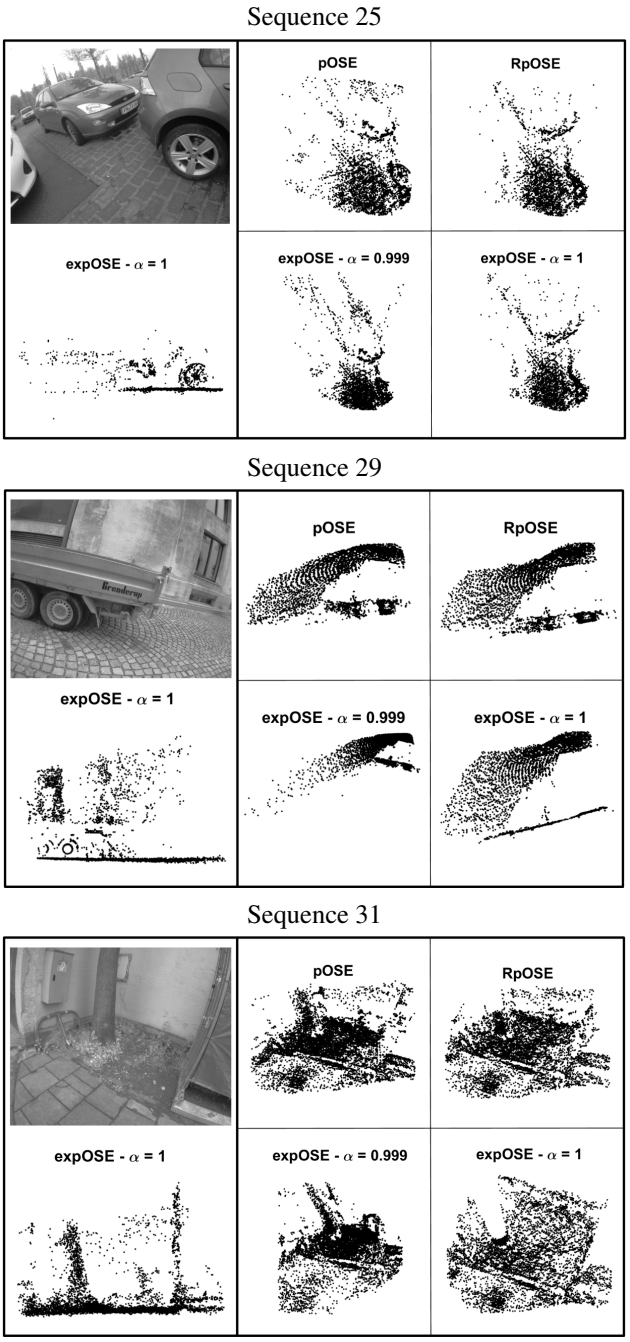


Figure 5. Visualization of reconstructions on sequences 25, 29, and 31 of the TUM dataset. (Left) An example of one of the images on the sequence. At the bottom, we show a view of the 3D reconstruction of expOSE for $\alpha = 1$. (Right) Comparison between the top view reconstructions obtained with pOSE, RpOSE and expOSE. All reconstructions shown here were not refined with bundle adjustment.

References

[1] J. Engel, V. Usenko, and D. Cremers. A photometrically calibrated benchmark for monocular visual odometry. In *arXiv:1607.02555*, July 2016. 3, 4

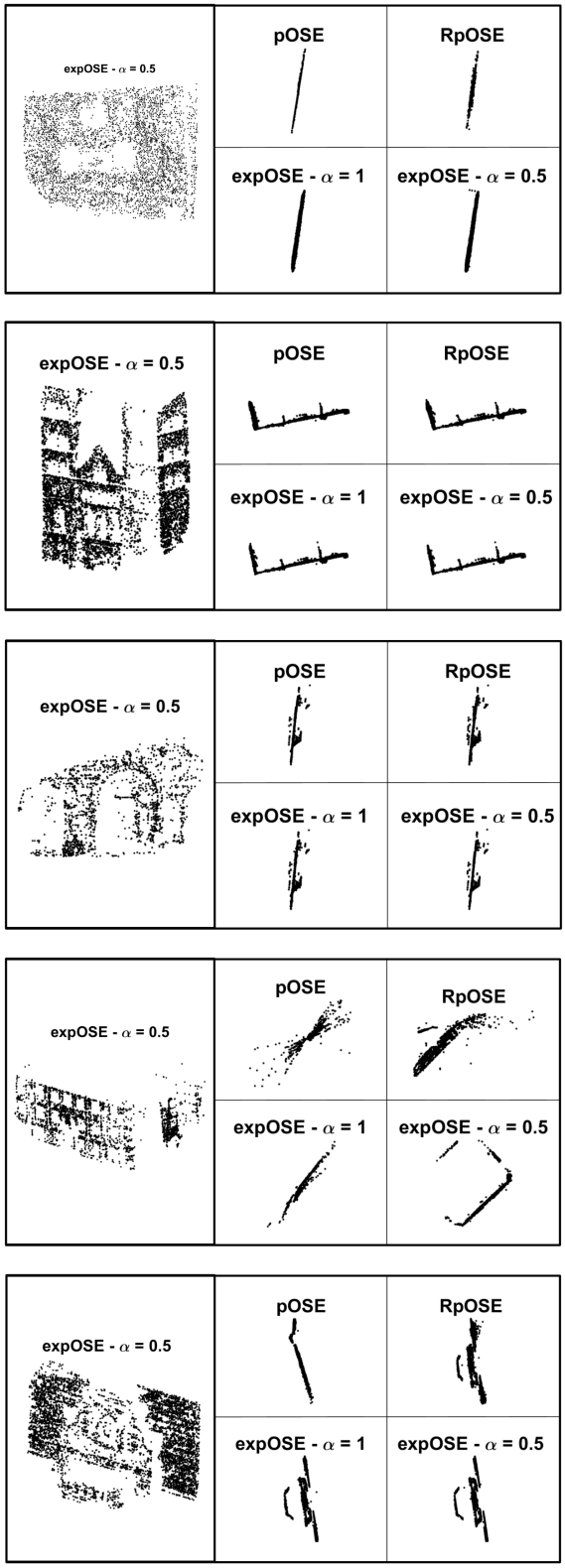
[2] Carl Olsson and Olof Enqvist. Stable structure from motion for unordered image collections. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA)*, pages 524–535, 2011. 5

[3] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008. 5

Table 4. Results on other benchmark datasets (over 10 instances). In green, we show the best results for each metric.

house_martenstorget		Convergence Rate	2D reproj. [pix]
pOSE		100%	1.86
	+ BA	20%	0.82
RpOSE		50%	11.14
	+ BA	90%	0.79
ExpOSE	$\alpha = 0.5$	80%	0.79
	$\alpha = 0.5$ + BA	80%	0.79
	$\alpha = 1$	10%	3.01
	$\alpha = 1$ + BA	10%	0.79
lund_cath_small		Convergence Rate	2D reproj. [pix]
pOSE		90%	1.18
	+ BA	80%	0.67
RpOSE		20%	6.53
	+ BA	50%	0.66
ExpOSE	$\alpha = 0.5$	100%	0.66
	$\alpha = 0.5$ + BA	100%	0.66
	$\alpha = 1$	30%	2.52
	$\alpha = 1$ + BA	30%	0.66
herz-jesu-p8		Convergence Rate	2D reproj. [pix]
pOSE		90%	1.11
	+ BA	90%	0.49
RpOSE		60%	7.61
	+ BA	100%	0.49
ExpOSE	$\alpha = 0.5$	100%	0.49
	$\alpha = 0.5$ + BA	100%	0.49
	$\alpha = 1$	100%	2.49
	$\alpha = 1$ + BA	100%	0.49
castle-p19		Convergence Rate	2D reproj. [pix]
pOSE		50%	2.65
	+ BA	40%	1.37
RpOSE		10%	16.69
	+ BA	10%	7.51
ExpOSE	$\alpha = 0.5$	40%	0.72
	$\alpha = 0.5$ + BA	10%	0.72
	$\alpha = 1$	10%	13.70
	$\alpha = 1$ + BA	10%	4.08
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

Figure 6. Corresponding visualizations of the sequences of the table on the left. Layout similar to Figure 5.



Alcatraz Courtyard		Convergence Rate	2D reproj. [pix]
pOSE		100%	1.02
	+ BA	90%	0.81
RpOSE		70%	14.00
	+ BA	60%	0.81
ExpOSE	$\alpha = 0.5$	100%	0.49
	$\alpha = 0.5$ + BA	60%	0.52
	$\alpha = 1$	100%	2.26
	$\alpha = 1$ + BA	100%	0.81