Supplementary material for Radial Distortion Invariant Factorization for Structure from Motion

José Pedro Iglesias[1]
[1]Department of Electrical Engineering
Chalmers University of Technology

Carl Olsson[1,2]
[2]Centre for Mathematical Sciences
Lund University

A. Connection between pOSE models and ML estimates

In this section we show that similar to the the ROSE case the original OSE from [3] is related through linearization to the ML estimate of the regular pinhole camera. Here we let $\boldsymbol{z}_{ij}$ denote the 2-vector containing the first coordinates of a 3-vector representing a 3D point $\boldsymbol{Z}_{ij}$. The number λ_{ij} is the third coordinate of $\boldsymbol{Z}_{ij}$. This point is assumed to be written $\boldsymbol{Z}_{ij} = P_i\boldsymbol{X}_j$, where P_i is a 3×4 camera matrix and $\boldsymbol{X}_j$ is a 4-vector with last coordinate 1 representing a 3D-point in the global coordinate system. The matrix Z containing the vectors $\boldsymbol{Z}_{ij}$ as blocks can then be factorized into

$$Z = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_2 & \dots \end{bmatrix}. \tag{A.1}$$

With this notation the bundle adjustment criteria with a regular pinhole projection can be written

$$\sum_{i,j} \|\frac{1}{\lambda_{ij}}\boldsymbol{z}_{ij} - \boldsymbol{m}_{ij}\|^2. \tag{A.2}$$

where $\boldsymbol{m}_{ij}$ is a 2-vector containing the measured image point.

The original OSE used in [3] is given by

$$\sum_{i,j} \|\boldsymbol{z}_{ij} - \boldsymbol{m}_{ij}\lambda_{ij}\|^2. \tag{A.3}$$

This term penalizes deviations from the viewing ray containing the camera center and the measured point. However in contrast to (A.2), which weights the error by the depth z_{ij}, the term A.3 only measures the perpendicular error to the viewing ray. In addition the OSE term has a shrinking bias as down scaling the matrix X always reduces to objective value. Hence a second term is needed to fix the scale of the reconstruction. For this purpose [3] uses the affine term

$$\sum_{ij} \|\boldsymbol{z}_{ij} - \boldsymbol{m}_{ij}\|^2. \tag{A.4}$$

A convex combination of the OSE and affine terms yeilds the pOSE objective

$$(1-\eta)\sum_{ij} \|\boldsymbol{z}_{ij} - \boldsymbol{m}_{ij}\lambda_{ij}\|^2 + \eta\sum_{ij} \|\boldsymbol{z}_{ij} - \boldsymbol{m}_{ij}\|^2 \tag{A.5}$$

It is clear that the above term in some sense favors solutions with depth 1 due to the affine term. However by selecting η relatively small the idea is that the distance to the viewing lines should dominate the error allowing for deviations from $\lambda_{ij} = 1$.

A straight forward application of Taylor's formula around the point $(\bar{\boldsymbol{z}}, \bar{\lambda})$ yields

$$\frac{1}{\lambda}\boldsymbol{z} \approx \frac{1}{\bar{\lambda}}\bar{\boldsymbol{z}} + \frac{1}{\bar{\lambda}}(\boldsymbol{z} - \bar{\boldsymbol{z}}) - \frac{1}{\bar{\lambda}^2}\bar{\boldsymbol{z}}(\lambda - \bar{\lambda}), \tag{A.6}$$

which gives

$$\frac{1}{\lambda}\boldsymbol{z} - \boldsymbol{m} \approx \frac{1}{\bar{\lambda}}\left(\bar{\boldsymbol{z}} + \boldsymbol{z} - \frac{1}{\bar{\lambda}}\bar{\boldsymbol{z}}\lambda\right) - \boldsymbol{m}. \tag{A.7}$$

If $(\bar{\boldsymbol{z}}, \bar{\lambda}) = (\boldsymbol{m}, 1)$ we get

$$\frac{1}{\lambda}\boldsymbol{z} - \boldsymbol{m} \approx \boldsymbol{z} - \boldsymbol{m}\lambda, \tag{A.8}$$

which is the OSE residual. Hence a Gauss-Newton approach attempting to minimize (A.2) starting from $(\boldsymbol{x}_{ij}, z_{ij}) = (\boldsymbol{m}_{ij}, 1)$ would in its first iteration attempt to minimize the approximation (A.3). We remark however that the starting point may not be feasible since the corresponding matrix X may not be of low rank and hence the factorization (A.1) infeasible. Still this observation opens up the possibility to improve the pOSE approximation in the sense that we can get closer to the bundle objective by updating the linearization. We illustrate in Figure A.1 this by performing two updates of the linearization, for $\eta = 0.5$.

B. More Visualizations and Datasets

B.1. Door, Fountain, Kirchenge, and Grossmunster datasets

In this section we show more visualizations regarding the experiments in Section 4.2, where the complete radial

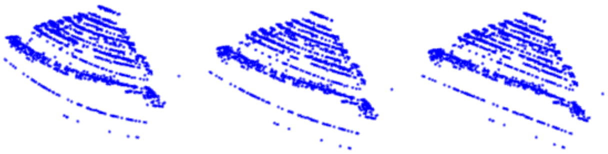

Figure A.1: Resulting 3D reconstruction when no update (left), 1 update (middle), and 2 updates (right) of the linearization with pOSE are performed, for $\eta = 0.5$ and without reducing η in each update. By performing two updates, the reprojection error decreased from 3.57 to 1.48 pixels.

distortion invariant SfM pipeline is started with the solution of RpOSE with 2 updates. In these experiments it was used $\eta = 0.01$, and η is decreased by a factor of 100 in each update. The 3D reconstructions are shown in Figure B.2.

B.2. TUM Dataset

Additionally, the proposed pipeline is evaluated using the TUM Monocular Visual Odometry Dataset [2]. We select 3 of the sequences that have more texture in the initial images. The 2D points are tracked along the first images. Relevant key points are selected in each image frame and tracked using optical flow correspondences [6] with backtracking for matching verification. The principal point and center of distortion are assumed to be the center of the image.

The benchmark provides ground-truth focal length and camera poses. The estimated camera poses are compared to the ground-truth after a similarity transformation. To evaluate our results we define the metrics relative focal length error $\frac{1}{2F}\sum_i^F \left(\frac{|f_x^i - f_x^{\text{GT}}|}{f_x^{\text{GT}}} + \frac{|f_y^i - f_y^{\text{GT}}|}{f_y^{\text{GT}}}\right)$, position error[1] $\frac{1}{FL}\sum_i \|c_i - c_i^{\text{GT}}\|$, rotation error $\frac{1}{F}\sum_i \text{acos}\left(\frac{\text{tr}(R_i^{\text{GT}} R_i^T) - 1}{2}\right)$, reprojection error, and runtime. In these experiments we used $\eta = 0.01$ and decreased η by a factor of 100 in each update. We also compare the performance of the pipeline initialized with RpOSE with the cases where ALM and pOSE ($\eta = 0.05$) are used instead. For the case of pOSE, no local optimization is performed after the factorization method, and the matrix completion step is replaced by an estimation of the distortion parameters only. Note that, contrarily to the experiments in Section 3.2 of the paper, in this experiment there is a compensation for distortion is done for the pOSE case (distortion parameters estimation + bundle adjustment). The setup of the experiments is similar to the experments in Section 4.1 of the paper. For fair comparison, local optimization and bundle adjustment were run until convergence, and the displayed runtimes correspond to the runtime of the factorization methods only. These methods are implemented in MATLAB, thus the relatively slow convergence times.

The results and 3D structure obtained are shown in Table B.1 and Figure B.3, respectively, and show that a pipeline initialized with our method outperforms instances of the same pipeline initialized with similar factorization methods in terms of error metrics and/or runtime.

C. Extension to Non-Rigid Structure from Motion

C.1. NRSfM factorization

The proposed framework can be extended to deal with Non-Rigid Structure from Motion problems by parameterizing X in each frame i as a linear combination of K shape basis $B_k \in \mathrm{R}^{3\times P}, k = 1, \dots, K$, resulting in

$$Z_i = P_i \begin{bmatrix} \sum_k c_{i,k} B_k \\ 1 \end{bmatrix} = \sum_k c_{i,k} P_i^{(1:3)} B_k + P_i^{(4)} = \\ = \begin{bmatrix} c_{i,1} P_i^{(1:3)} & \dots & c_{i,K} P_i^{(1:3)} & P_i^{(4)} \end{bmatrix} \begin{bmatrix} B \\ 1 \end{bmatrix} \tag{C.9}$$

where $c_{i,k}, k = 1, \dots, K$ are the shape coefficients for the i:th view, $B \in \mathrm{R}^{3K\times P}$ is a matrix consisting of a vertical concatenation of all B_k, and $P_i^{(1:3)}$ and $P_i^{(4)}$ are the first three columns and the last column of P_i, respectively. Expanding to all views, we get

$$Z = \tilde{P} \begin{bmatrix} B \\ 1 \end{bmatrix} = \begin{bmatrix} c_{1,1} P_1^{(1:3)} & \dots & c_{1,K} P_1^{(1:3)} & P_1^{(4)} \\ \vdots & & & \\ c_{F,1} P_F^{(1:3)} & \dots & c_{F,K} P_F^{(1:3)} & P_F^{(4)} \end{bmatrix} \begin{bmatrix} B \\ 1 \end{bmatrix}. \tag{C.10}$$

Note that Z is now a rank $3K+1$, and $K = 1$ consists of the rigid case considered in the main paper. The factorization $\tilde{P} \begin{bmatrix} B \\ 1 \end{bmatrix}$ can then be used in

$$\underset{\tilde{P}, B}{\text{minimize}} \quad \left\| \mathcal{A}\left(\tilde{P} \begin{bmatrix} B \\ 1 \end{bmatrix}\right) - b \right\|^2 \tag{C.11}$$

where $\mathcal{A}$ and b are defined in the main paper.

C.2. Weighted Nuclear Norm Regularization

It is possible to weight differently the contribution of each of the K shape basis in the reconstruction [5, 4] by adding to (C.11) a regularization term on the singular values of Z, leading to the optimization problem

$$\underset{Z}{\text{minimize}} \quad \sum_i w_i \sigma_i(Z) + \|\mathcal{A}(Z) - b\|^2. \tag{C.12}$$

---

[1] The position error is normalized by the length of the path along the selected images.

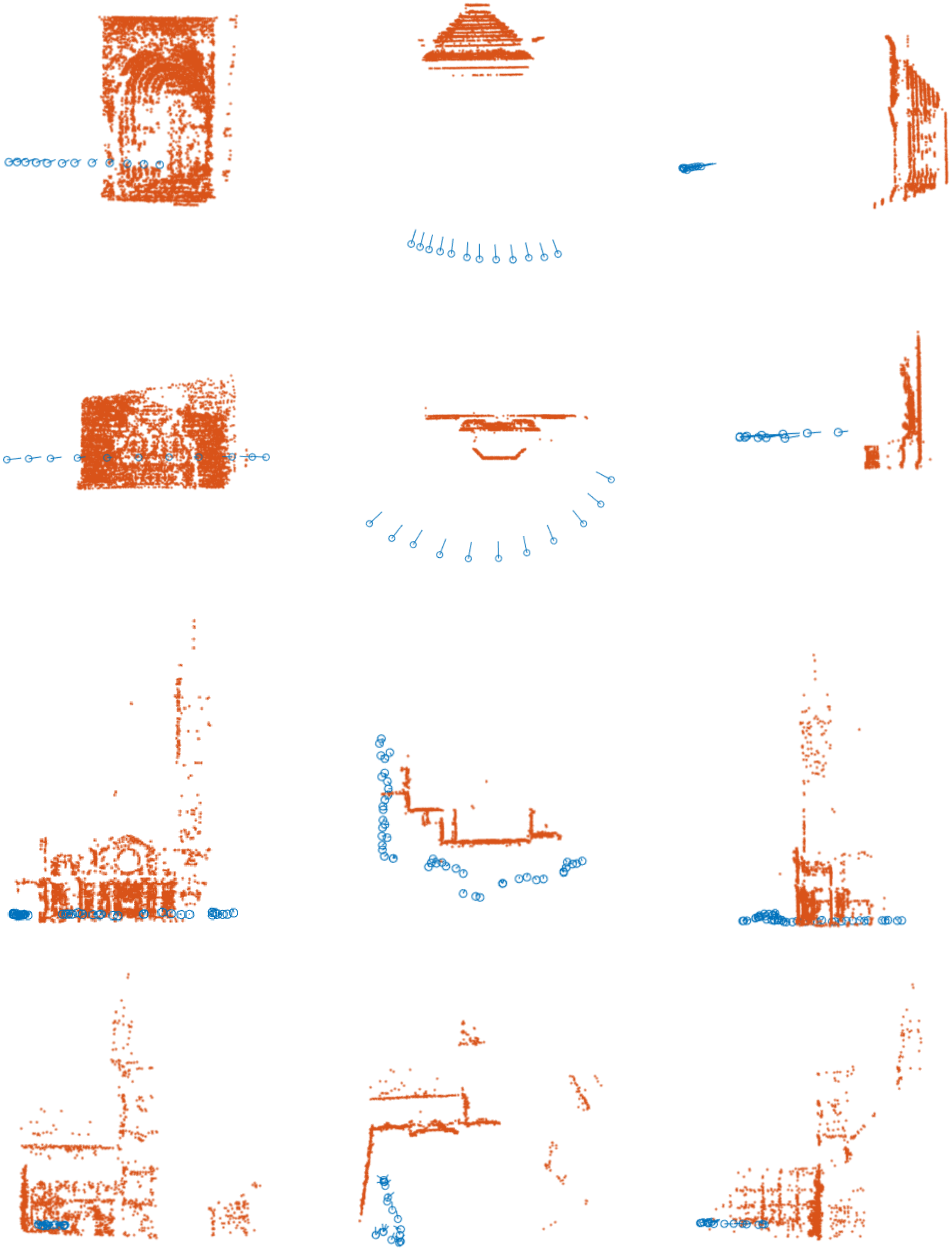

Figure B.2: 3D reconstruction of the Door, Fountain, Kirchenge and Grossmunster datasets (top to bottom).

Table B.1: Evaluation metrics for the experiments with the TUM datasets. Each sequence has [Fc, Pp, W%], where F is the number of viewpoints, P the number of tracked points and W the percentage of available data entries.

		sequence_25 [30c,4822p,75.5%]	sequence_29 [30c,5377p,60.3%]	sequence_31 [15c,7606p,88.0%]
Rotation [deg]	RpOSE	**1.235**	**0.265**	1.065
	ALM	3.099	129.987	13.932
	pOSE	7.507	0.395	**0.802**
Position [%]	RpOSE	**0.44**	1.34	**0.59**
	ALM	0.49	26.07	5.87
	pOSE	1.06	**0.94**	0.69
2D [pix]	RpOSE	**0.394**	**0.146**	**0.149**
	ALM	0.420	219.19	0.644
	pOSE	**0.394**	**0.146**	**0.149**
Focal [%]	RpOSE	**6.00**	**5.63**	13.41
	ALM	12.34	98.66	**7.37**
	pOSE	9.18	7.31	10.06
Runtime [s]	RpOSE	**152**	**145**	**107.88**
	ALM	1073	1050	2036
	pOSE	420	798	1713

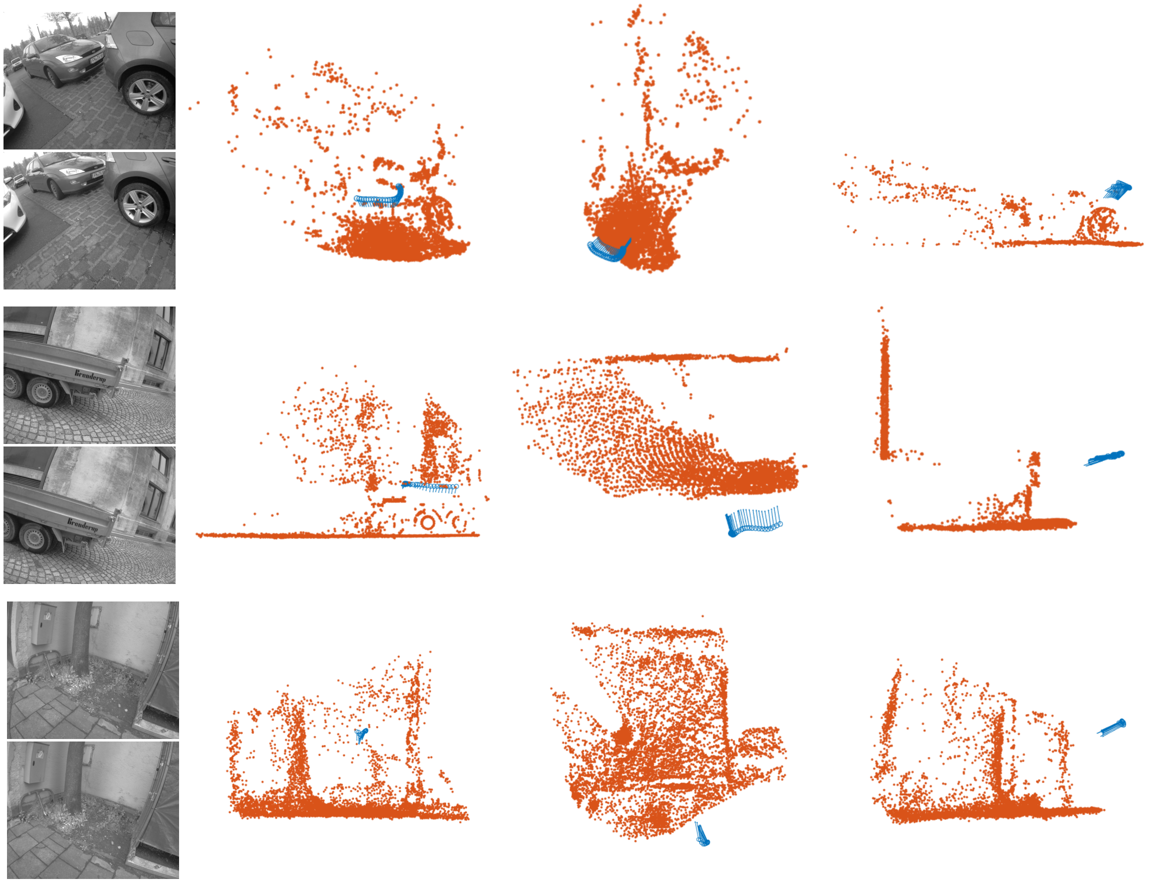

Figure B.3: (Left) Examples of images for the sequences of TUM dataset. (Right) 3D reconstructions using the pipeline initialized with RpOSE for sequence 25, 29, and 31 (top to bottom) of the TUM dataset.

This regularization is commonly known as Weighted Nuclear Norm, and as shown by Iglesias *et al.* [4], it can be applied to factorization formulations by solving instead the equivalent problem

$$\underset{\tilde{P},B}{\text{minimize}} \quad \sum_i w_i \frac{\|\tilde{P}^{(i)}\|^2 + \|B^{T(i)}\|^2}{2} + \left\| \mathcal{A}\left(\tilde{P}\begin{bmatrix} B \\ 1 \end{bmatrix}\right) - b \right\|^2, \tag{C.13}$$

where $\tilde{P}^{(i)}$ and $B^{T(i)}$ are the i:th column of $\tilde{P}$ and B^T, respectively.

C.3. Estimation of correction matrix

Similarly to the rigid case, the solution $\tilde{P}$ obtained from (C.21) does not necessarily have the desired structure shown in (C.10). Dai *et al.* [1] propose an algorithm to estimate a correction matrix G_k such that for each view i

$$\tilde{P}_i G_k \propto f c_{i,k} R_i. \tag{C.14}$$

From the orthogonality property of R_i, we get

$$\tilde{P}_i \underbrace{G_k G_k^T}_{Q} \tilde{P}_i^T \propto f^2 c_{i,k}^2 \mathcal{I}_2, \tag{C.15}$$

from which is possible to obtain a linear system such that

$$A\text{vec}(Q) = 0 \tag{C.16}$$

The solution Q is then obtained from the $2K^2 - K$ nullspace of A in (C.16) and must verify the intersection

$$\{A\text{vec}(Q) = 0 \quad \cap \quad Q \succeq 0 \quad \cap \quad \text{rank}(Q) = 3\}, \tag{C.17}$$

which is solved with SDP. After estimating Q, G_k can be obtained through Cholesky decomposition of Q, and the rotations R_i from (C.14). For further details we refer the reader to [1].

C.4. Structure estimation from known camera rotations

Dai *et al.* [1] also propose a method for structure estimation from known rotation matrices obtained from the method described in Section C.3. In this case, the 3D structure and the camera translations are the unknowns, and the parameterization becomes

$$Z = Rg(X^\sharp) + t\mathbb{1}^T, \tag{C.18}$$

where $R = \text{blkdiag}(R_1, \ldots, R_F) \in \mathrm{R}^{2F\times 3F}$, $t \in \mathrm{R}^{2F}$ is a vector corresponding to the translations (or $P^{(4)}$), $X = g(X^\sharp)$, $X^\sharp \in \mathrm{R}^{F\times 3P}$ is a reshaped version of X such that

$$X^\sharp = \begin{bmatrix} X_x & X_y & X_z \end{bmatrix}, \tag{C.19}$$

and $X_x, X_y, X_z \in \mathrm{R}^{F,P}$ are the x, y, and z-coordinates of the 3D structure over the F views. Using this reshaped matrix, we can factorize $X^\sharp = CB^\sharp$, with

$$C = \begin{bmatrix} c_{1,1} & \ldots & c_{1,K} \\ \vdots & \ddots & \vdots \\ c_{F,1} & \ldots & c_{F,K} \end{bmatrix} \quad \text{and} \quad B^\sharp = \begin{bmatrix} g^{-1}(B_1) \\ \vdots \\ g^{-1}(B_K) \end{bmatrix}. \tag{C.20}$$

Note that since the rank of $X^\sharp$ is K, the matrices C and $B^\sharp$ will have K columns and rows, respectively. This is a significant dimensionality reduction when compared to the $3K$ columns/rows of the factors in Section C.1, and thus leads to a smaller (and more constrained) problem.

The regularization mentioned in Section C.2 can also be applied directly on the singular values of $X^\sharp$, resulting in the following optimization problem

$$\underset{Z=Rg(CB^\sharp)+t\mathbb{1}^T}{\text{minimize}} \quad \sum_i w_i \frac{\|C^{(i)}\|^2 + \|B^{\sharp T(i)}\|^2}{2} + \|\mathcal{A}Z - b\|^2. \tag{C.21}$$

We refer the reader to Iglesias *et al.* [4] for further details on how to optimize (C.21).

C.5. Proposed extension to NRSfM

We use a similar pipeline to the one proposed by Dai *et al.* [1] to extend our method to NRSfM. The pipeline consists of the following steps:

1. Obtain a factorization $\{\tilde{P}, B\}$ using the formulation described in Sections C.1 and C.2;
2. Estimate the camera rotations using the method described in Section C.3;
3. Estimate the 3D structure and translation by solving the matrix factorization problem described in Section C.4.

We qualitatively evaluate our method using the Back [7], Heart [8], Paper [9], with $K = 2, 3$ and 2, respectively, and $\eta = 0.05$. The weights w_i chosen for each dataset are

- Back: $w_i = 10^{-3}\begin{bmatrix}1 & 1 & 1 & 2 & 2 & 2\end{bmatrix}$ for C.1/C.2; $w_i = 10^{-3}\begin{bmatrix}1 & 2\end{bmatrix}$ for C.4;
- Heart: $w_i = 10^{-4}\begin{bmatrix}1 & 1 & 1 & 2 & 2 & 2 & 3 & 3 & 3\end{bmatrix}$ for C.1/C.2; $w_i = 10^{-3}\begin{bmatrix}1 & 2 & 3\end{bmatrix}$ for C.4;
- Paper: $w_i = 10^{-4}\begin{bmatrix}1 & 1 & 1 & 2 & 2 & 2\end{bmatrix}$ for C.1/C.2; $w_i = 1 \times 10^{-4}\begin{bmatrix}1 & 2\end{bmatrix}$ for C.4.

The results, shown in Figures C.4, C.5, and C.6, as well as in the three videos and .fig files submitted along with this document, show that our method can easily be extended to NRSfM problems with promising results. Note that these consist of preliminary results.

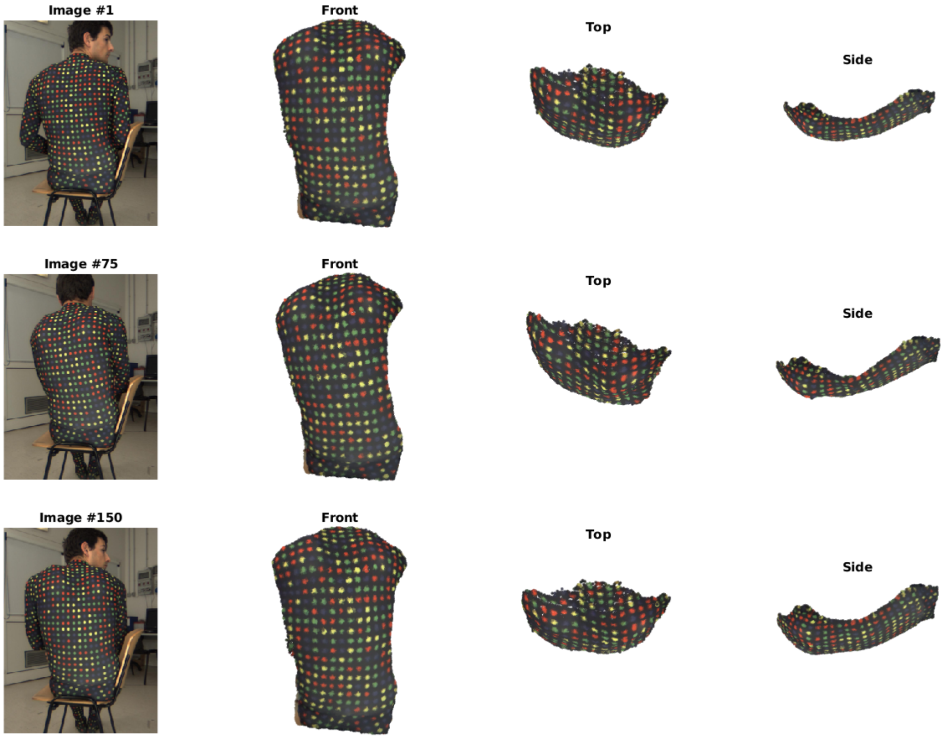

Figure C.4: Back dataset.

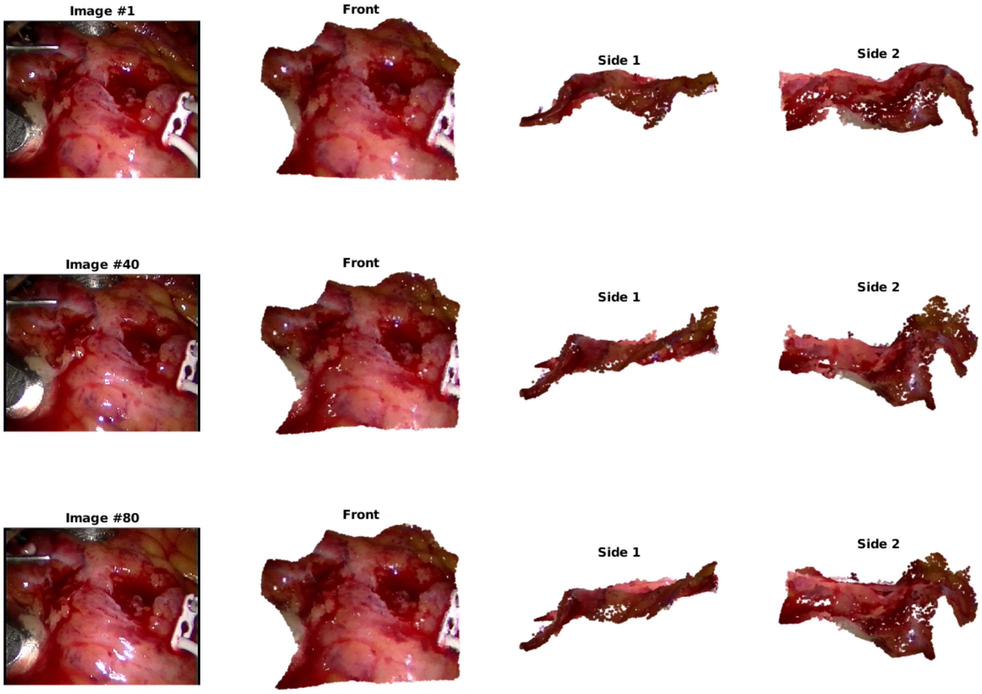

Figure C.5: Heart dataset.

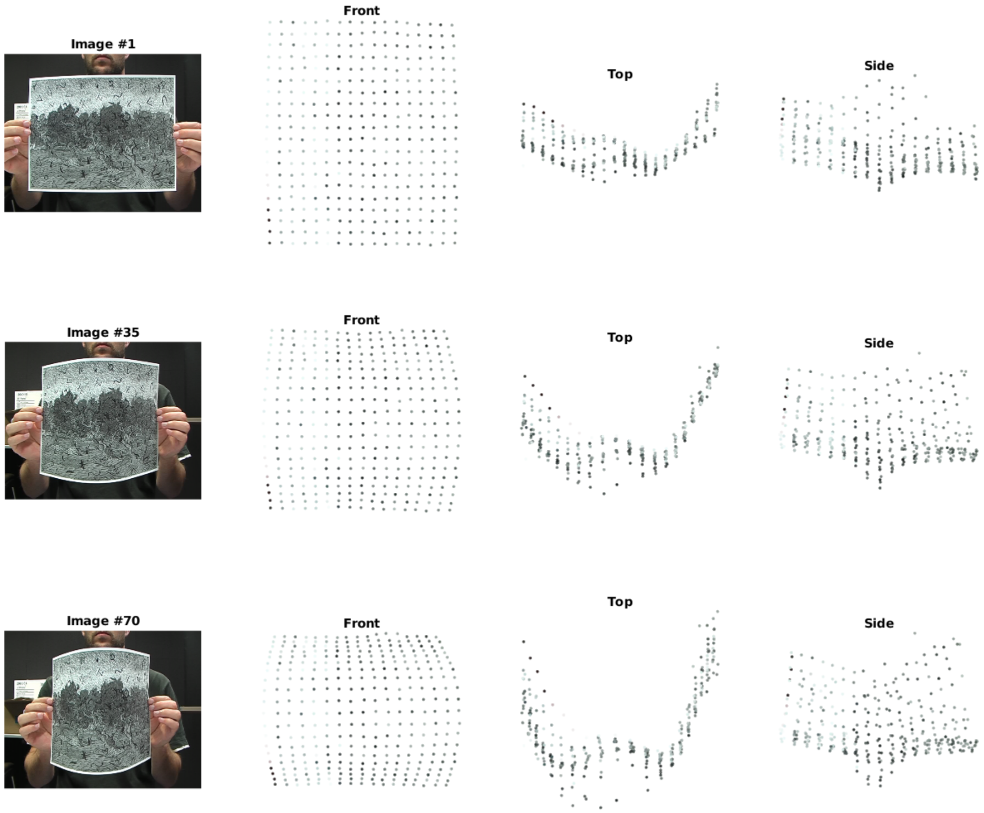

Figure C.6: Paper dataset.

References

[1] Y. Dai, H. Li, and M. He. A simple prior-free method for non-rigid structure-from-motion factorization. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2018–2025, 2012. 5

[2] J. Engel, V. Usenko, and D. Cremers. A photometrically calibrated benchmark for monocular visual odometry. In *arXiv:1607.02555*, July 2016. 2

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

[4] José Pedro Iglesias, Carl Olsson, and Marcus Valtonen Örnhag. Accurate optimization of weighted nuclear norm for non-rigid structure from motion. In *European Conference on Computer Vision (ECCV)*, 2020. 2, 5

[5] Suryansh Kumar. Non-rigid structure from motion: Prior-free factorization method revisited. In *IEEE Winter Conference on Applications of Computer Vision, WACV 2020, Snowmass Village, CO, USA, March 1-5, 2020*, pages 51–60. IEEE, 2020. 2

[6] Bruce Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision (ijcai). volume 81, 04 1981. 2

[7] Chris Russell, Joao Fayad, and Lourdes Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3009 – 3016, 07 2011. 5

[8] Danail Stoyanov, George P. Mylonas, Fani Deligianni, Ara Darzi, and Guang Zhong Yang. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In James S. Duncan and Guido Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, pages 139–146, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg. 5

[9] Aydin Varol, Mathieu Salzmann, Engin Tola, and Pascal Fua. Template-free monocular reconstruction of deformable surfaces. In *International Conference on Computer Vision (ICCV)*, pages 1811 – 1818, 11 2009. 5