

Appendix

A. Linear Form of the Error Term in Eqn. 6

Let the given set of correspondences be:

$$\{(\mathbf{p}^s, \mathbf{q}^s) \mid 1 \leq s \leq \mathcal{S}; \mathbf{p}^s = [(\mathbf{p}^s)' \quad 1]^\top, \mathbf{q}^s = [(\mathbf{q}^s)' \quad 1]^\top \text{ for } \{(\mathbf{p}^s)', (\mathbf{q}^s)'\} \in \mathbb{R}^3\}.$$

Then we can write the error term in Eqn. 6 as

$$e^s(\mathbf{M}(k)) = \left\| \begin{bmatrix} (\mathbf{p}^s)' \\ 1 \end{bmatrix} - (\mathbf{I}_4 + \Delta \mathbf{m}(k))\mathbf{M}(k-1) \begin{bmatrix} (\mathbf{q}^s)' \\ 1 \end{bmatrix} \right\| \quad (\text{A.1})$$

$$\Rightarrow e^s(\mathbf{M}(k)) = \left\| \begin{bmatrix} (\mathbf{p}^s)' - (\mathbf{q}^s)' \\ 0 \end{bmatrix} - \begin{bmatrix} [\boldsymbol{\omega}]_\times & | & \mathbf{u} \\ \mathbf{0} & | & 0 \end{bmatrix} \mathbf{M}(k-1) \begin{bmatrix} (\mathbf{q}^s)' \\ 1 \end{bmatrix} \right\| \quad (\text{A.2})$$

where we get $\Delta \mathbf{m}(k) = \begin{bmatrix} [\boldsymbol{\omega}]_\times & | & \mathbf{u} \\ \mathbf{0} & | & 0 \end{bmatrix}$ from Eqn. 2. Rewriting Eqn. A.2 with $\mathbf{v} = [\boldsymbol{\omega} \quad \mathbf{u}]^\top$, $\mathbf{M}(k-1) = \begin{bmatrix} \mathbf{R}(k-1) & | & \mathbf{t}(k-1) \\ \mathbf{0} & | & 1 \end{bmatrix}$ (from Eqn. 1), and by dropping the trailing 0 for ease of representation, we get,

$$e^s(\mathbf{M}(k)) = \left\| \begin{bmatrix} -[\mathbf{R}(k-1)(\mathbf{q}^s)' + \mathbf{t}(k-1)]_\times & | & \mathbf{I}_3 \\ \mathbf{0} & | & 0 \end{bmatrix} \mathbf{v} - \begin{bmatrix} (\mathbf{p}^s)' - (\mathbf{q}^s)' \\ 0 \end{bmatrix} \right\|. \quad (\text{A.3})$$

Thus, we have in Eqn. 6,

$$\mathbf{A}^s = \begin{bmatrix} -[\mathbf{R}(k-1)(\mathbf{q}^s)' + \mathbf{t}(k-1)]_\times & | & \mathbf{I}_3 \end{bmatrix}, \quad (\text{A.4})$$

$$\mathbf{b}^s = (\mathbf{p}^s)' - (\mathbf{q}^s)'. \quad (\text{A.5})$$

B. Algorithm for Joint Multiview Registration

Similar to our algorithm for robust motion estimation between a pair of 3D scans, we state our solution for the robust motion estimation of a set of $N (\geq 2)$ 3D scans as given in Algorithm B.1.

Algorithm B.1 IRLS estimation of joint multiview 3D registration

Input: $\left\{ \left\{ (\mathbf{p}_i^1, \mathbf{p}_j^1) \cdots (\mathbf{p}_i^{S_{ij}}, \mathbf{p}_j^{S_{ij}}) \right\} \mid (i, j) \in \mathcal{E} \right\}$

(according to the viewgraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$)

Output: $\mathbb{M} = \{\mathbf{M}_i \mid \mathbf{M}_i \in \mathbb{SE}(3) \forall i \in [1 \cdots N]\}$ (Robust estimate of motion of the set of $N = |\mathcal{V}|$ scans)

Initialization: \mathbb{M} is set to initial guess $\mathbb{M}_{initial}$

```

while  $\|\mathbf{v}\| > N\epsilon$  do
    1.  $\mathbf{v} = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_N]^\top$  Compute
     $\{(\mathbf{A}_{ij}^s, \mathbf{b}_{ij}^s) \mid \forall s \in [1 \cdots S_{ij}]; (i, j) \in \mathcal{E}\}$  using
    Eqn. 10
    2. Compute weights  $w_{ij}^s = \frac{\rho'(e_{ij}^s)}{e_{ij}^s}$  as used in
    Eqn. 12
    3. Estimate  $\mathbf{v}$  as IRLS solution for Eqn. 12
    4. Update  $\mathbf{M}_i \leftarrow \exp(\hat{\mathbf{v}}_i)\mathbf{M}_i \forall i \in [1 \cdots N]$ 
end while

```

Table C.1: Median rotation angle error (RAE) (in degrees), median translation norm error (TNE) (in units of the mean scene diameter) and mean running time (in milliseconds) of the motion step of each method for each sequence in the UWA dataset.

Method	Median RAE	Median TNE	Mean Time
FGR [3]	1.276	0.152	32.7
Our $L_{\frac{1}{2}}$	0.319	0.034	10.1
Our L_1	0.824	0.108	7.0
Our GM	1.276	0.152	14.5

C. More Results

UWA dataset: Table C.1 reports the performance of the motion step of the Fast Global Registration (FGR) method of [3] as well as all the 3 versions of our method on the UWA dataset [2]. This dataset consists of 5 objects and 50 scenes, with each scene consisting of a combination of these objects. The task is to align individual objects to the scenes, given that each scene contains substantial clutter and occlusion. A total of 188 such alignment tasks are provided in the dataset. As we can see from Table C.1, the $L_{\frac{1}{2}}$ version of our method produces the lowest median rotation angle and median translation norm errors. It is also significantly faster than FGR [3].

Relative improvement of joint multiview approach over two-stage motion averaging ap-

proach: As described in Section 5.2, we show in Table C.2 that the reconstruction error of the scenes in the Augmented ICL-NUIM dataset [1] decreases when we use our joint multiview estimation procedure on top of the two-stage motion averaged approach. We show the improvement achieved using the $L_{\frac{1}{2}}$ loss, which is our best-performing version.

Augmented ICL-NUIM dataset: We also show full reconstructions of the livingroom 1, livingroom 2, office 1 and office 2 sequences from the Augmented ICL-NUIM dataset [1] in Figures C.1, C.3 and C.4 respectively, as provided by the $L_{\frac{1}{2}}$ version of our method.

D. An Illustration of the Limitation of FPFH Feature-Matching Based Registration

As discussed in Section 6.2, we have presented a scenario where the FPFH feature-matching based registration technique breaks down due to unreliability of the feature matches themselves. In this particular scenario, we have 23 scans of a life-size statue of Mahatma Gandhi collected with a standard commercial depth camera. Figure D.2 shows the plan view of a schematic of the cameras (represented as balls) around the statue, as recovered by our ICP-based multiview approach. Recall that these cameras are the nodes of the viewgraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. We also display a schematic of the edges in the viewgraph \mathcal{G} (using sticks). The thickness of each edge is proportional to the number of FPFH feature matches found between the corresponding camera (or equivalently scan) pair.

We can observe from Figure D.2 that the thinnest edges are found between pairs of cameras at different depths, implying that there are extremely few FPFH feature matches between these cameras. Compounding this observation with the fact that FPFH features are noisy to begin with, the resultant motions between these cameras, even with our robust cost function, are grossly incorrect. In contrast, our ICP-based mul-

tiview approach can, albeit at a higher computational cost, align these cameras correctly and produce the desired reconstruction.

References

- [1] Sungjoon Choi, Qian-Yi Zhou, and Vladlen Koltun. Robust Reconstruction of Indoor Scenes. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5556–5565. IEEE, 2015. 2
- [2] Ajmal S. Mian, Mohammed Bennamoun, and Robyn Owens. Three-Dimensional Model-Based Object Recognition and Segmentation in Cluttered Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1584–1601, 2006. 1
- [3] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast Global Registration. In *European Conference on Computer Vision*, pages 766–782. Springer, 2016. 1

Table C.2: Mean registration error (in meters) before and after applying the joint multiview (MV) procedure with $L_{\frac{1}{2}}$ loss on the motion averaged estimate, for full reconstruction from the fragments of each sequence in the Augmented ICL-NUIM dataset.

	livingroom 1	livingroom 2	office 1	office 2
Before MV	0.07	0.07	0.06	0.07
After MV	0.04	0.05	0.03	0.04

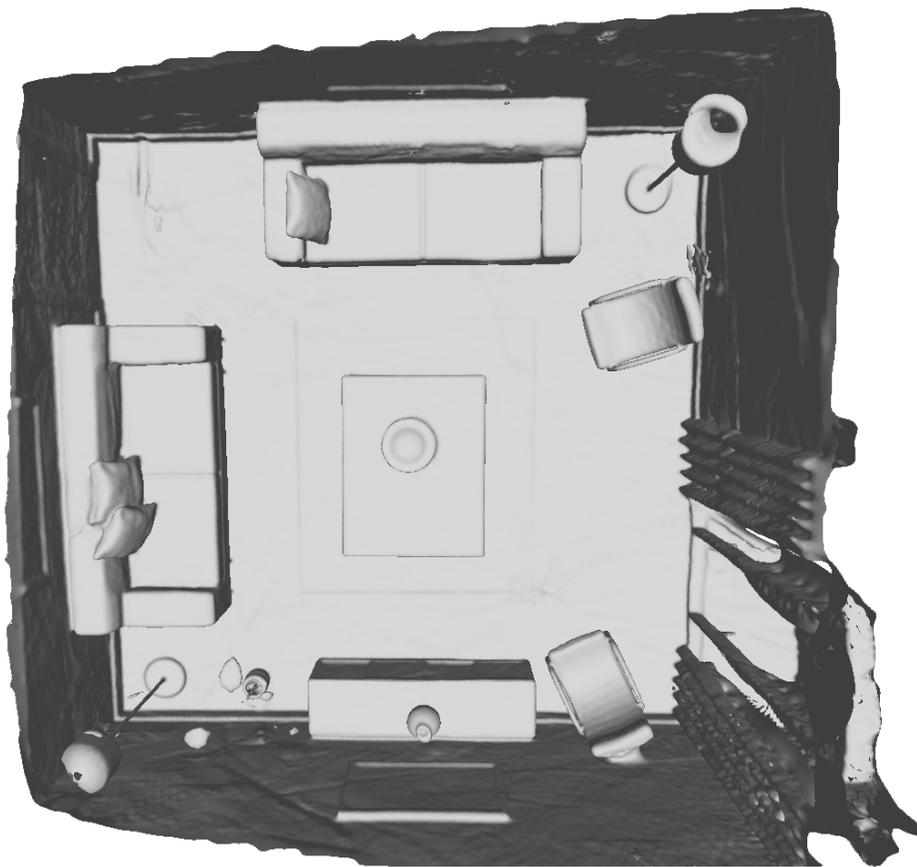


Figure C.1: Reconstruction of the livingroom 1 sequence from the Augmented ICL-NUIM dataset, as given by our method with the $L_{\frac{1}{2}}$ loss function

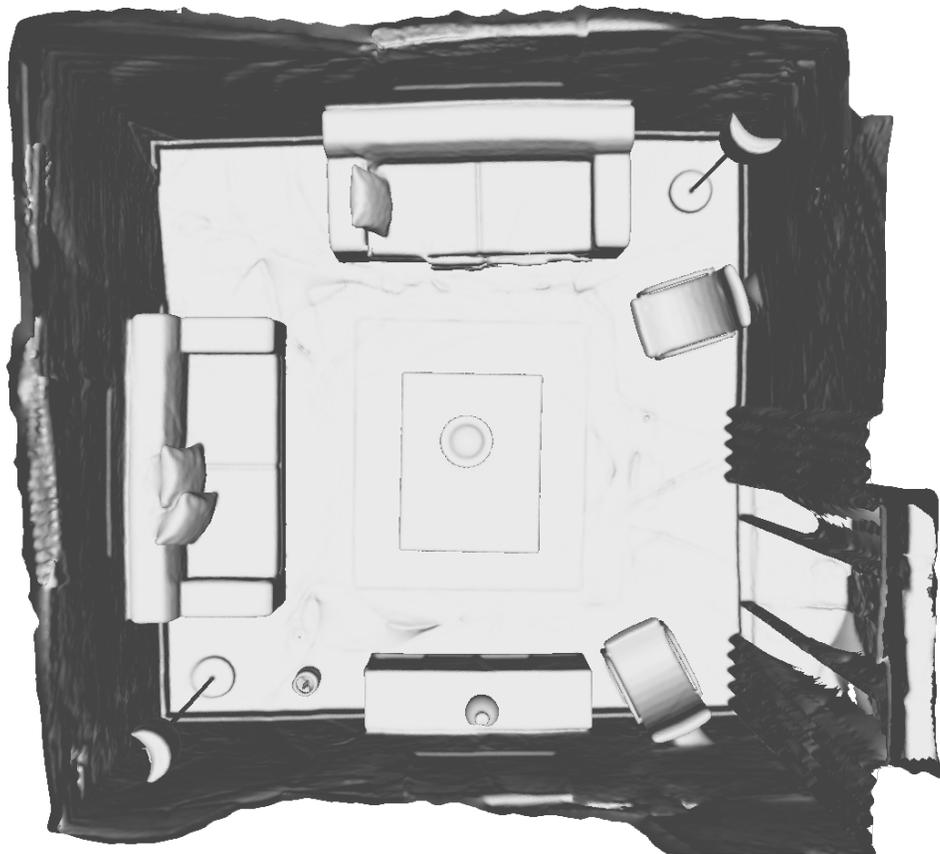


Figure C.2: Reconstruction of the `livingroom 2` sequence from the Augmented ICL-NUIM dataset, as given by our method with the $L_{\frac{1}{2}}$ loss function

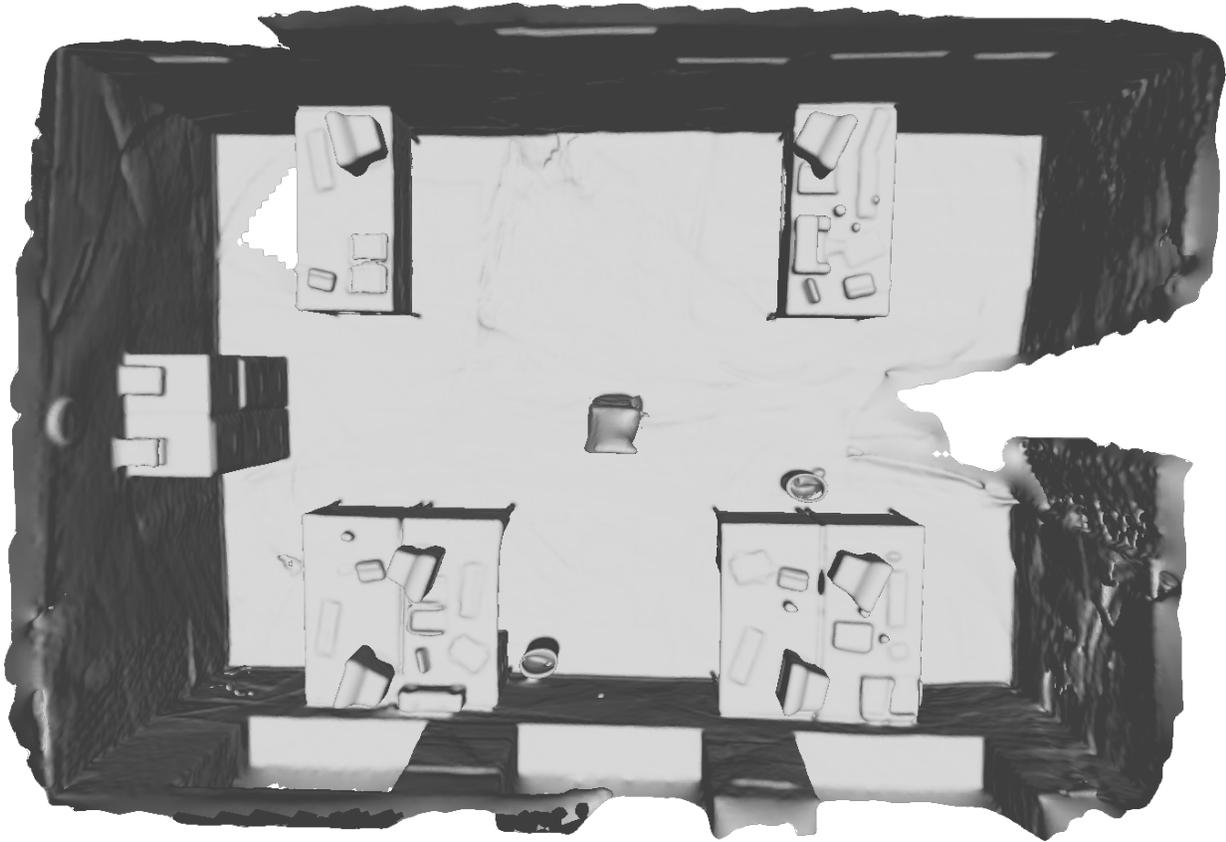


Figure C.3: Reconstruction of the `office_1` sequence from the Augmented ICL-NUIM dataset, as given by our method with the $L_{\frac{1}{2}}$ loss function



Figure C.4: Reconstruction of the `office_2` sequence from the Augmented ICL-NUIM dataset, as given by our method with the $L_{\frac{1}{2}}$ loss function

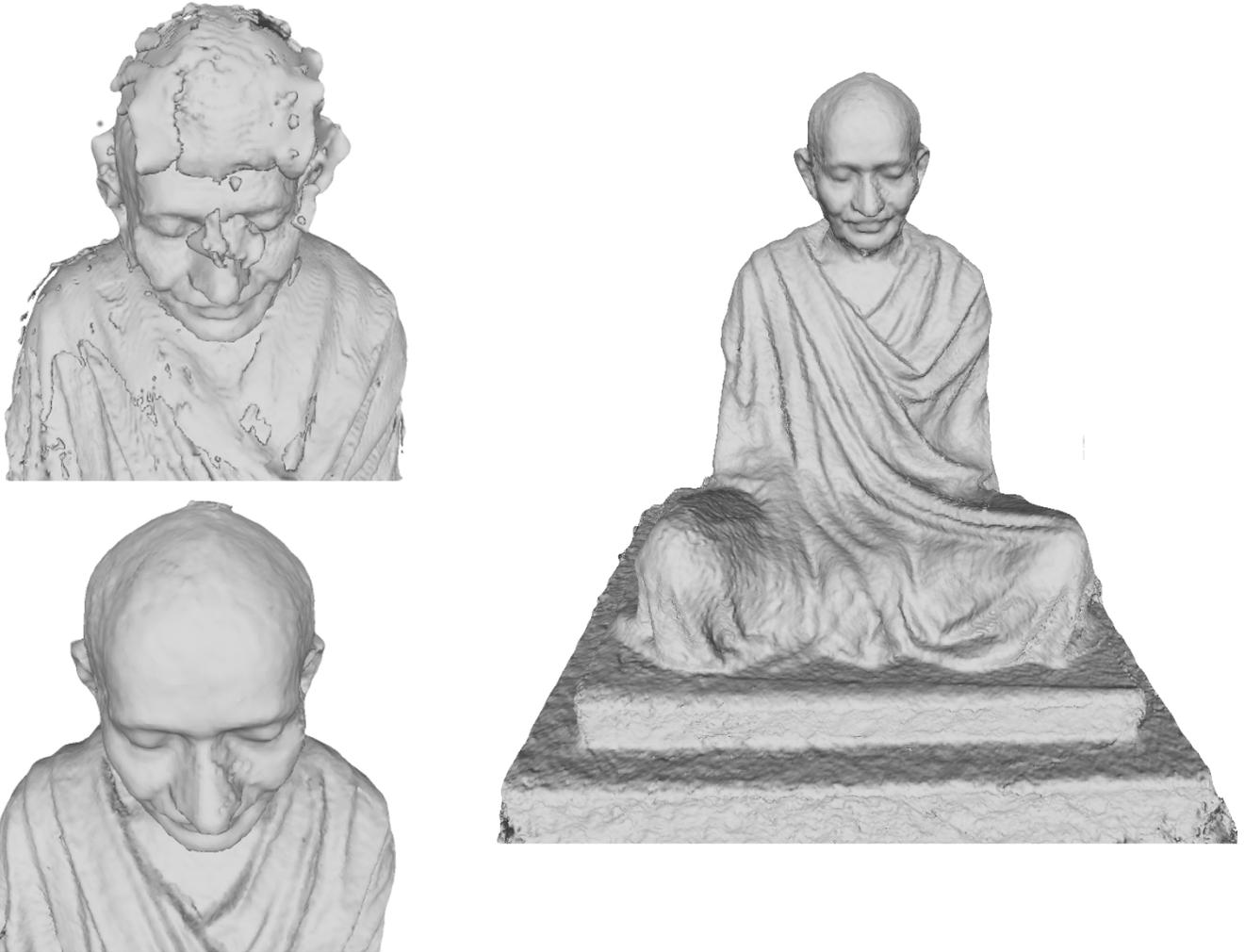


Figure D.1: Full scale version of Figure 3 of the paper showing a 3D reconstruction of a statue of Mahatma Gandhi. The close-up on the top left shows that joint multiview registration using FPFH features fails whereas the close-up on the bottom left shows successful registration using our robust pairwise motion estimation within a multiview ICP routine. The full reconstruction is shown on the right.

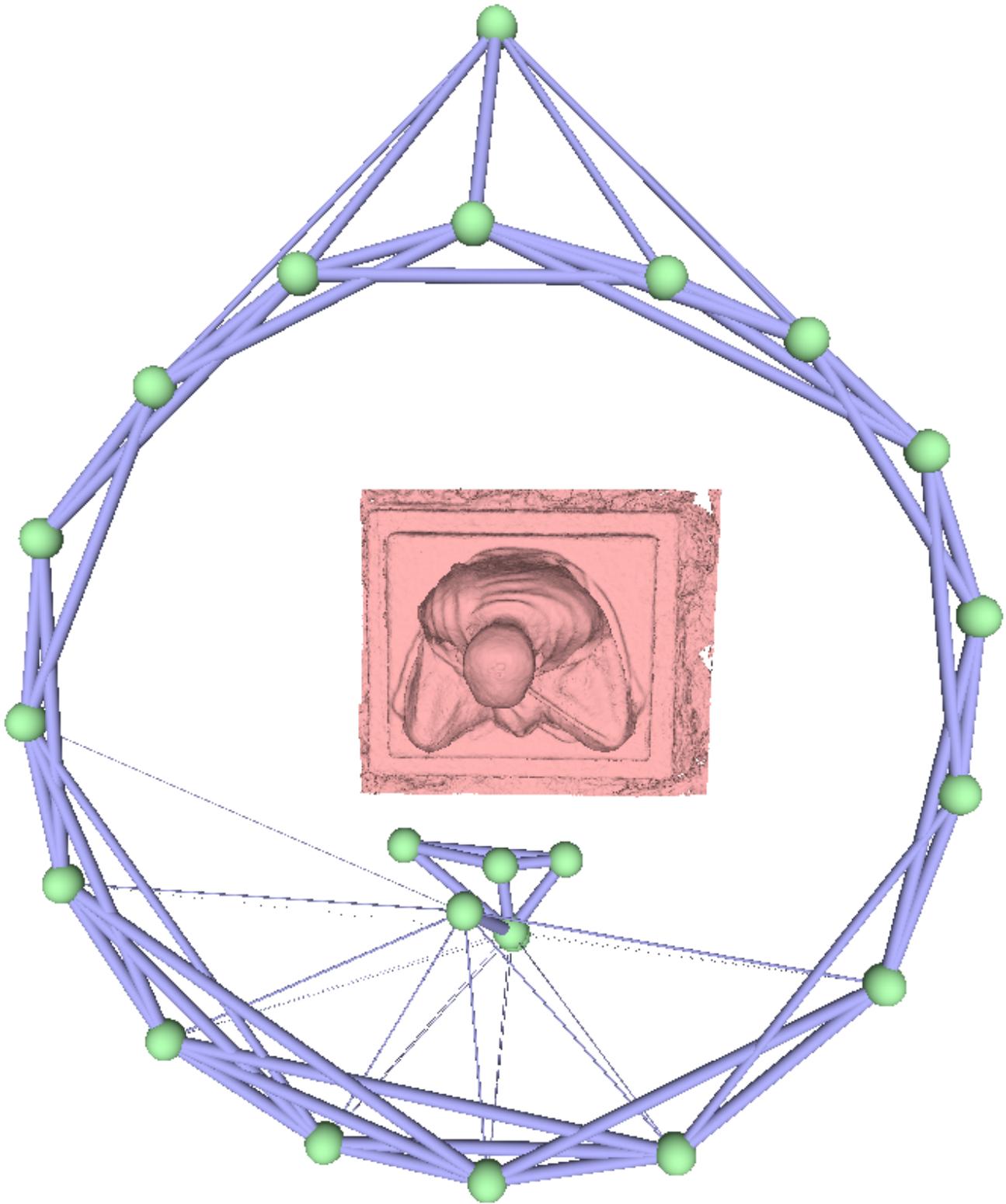


Figure D.2: Plan view of a schematic representation of the viewgraph used for reconstruction of the statue of Mahatma Gandhi. See text for details.