Colored Point Cloud Registration Revisited
Supplementary Material

Jaesik Park   Qian-Yi Zhou   Vladlen Koltun

Intel Labs

Appendix A: RGB-D Image Alignment

Section 3 introduced a joint photometric and geometric objective for RGB-D image alignment. This appendix presents an algorithm that optimizes that objective. This algorithm is used in the reconstruction system presented in Section 5 of the paper.

Objective

The joint photometric and geometric objective for RGB-D image alignment is a nonlinear least-squares objective

$$E(\mathbf{T}) = (1 - \sigma) \sum_{\mathbf{x}} \left(r_I^{\mathbf{x}}(\mathbf{T})\right)^2 + \sigma \sum_{\mathbf{x}} \left(r_D^{\mathbf{x}}(\mathbf{T})\right)^2, \quad \text{(A1)}$$

where $r_I^{\mathbf{x}}$ and $r_D^{\mathbf{x}}$ are the photometric and geometric residuals, respectively:

$$r_I^{\mathbf{x}}(\mathbf{T}) = I_i(\mathbf{g}_{uv}(\mathbf{s}(\mathbf{h}(\mathbf{x}, D_j(\mathbf{x})), \mathbf{T}))) - I_j(\mathbf{x}), \quad \text{(A2)}$$

$$r_D^{\mathbf{x}}(\mathbf{T}) = D_i(\mathbf{g}_{uv}(\mathbf{s}(\mathbf{h}(\mathbf{x}, D_j(\mathbf{x})), \mathbf{T})))$$
$$\quad - \mathbf{g}_d(\mathbf{s}(\mathbf{h}(\mathbf{x}, D_j(\mathbf{x})), \mathbf{T})). \quad \text{(A3)}$$

The definitions of I_i, D_i, $\mathbf{s}$, $\mathbf{h}$, $\mathbf{g}$ are introduced in Section 3.

Optimization

Similar to Section 4.3, this objective is minimized using the Gauss-Newton method. Specifically, we start from an initial transformation $\mathbf{T}^0$ and perform optimization iteratively. In each iteration, we locally parameterize $\mathbf{T}$ with a 6-vector ξ, evaluate the residual $\mathbf{r}$ and Jacobian $\mathbf{J_r}$ at $\mathbf{T}^k$, solve the linear system in (21) to compute ξ, and use ξ to update $\mathbf{T}$. To compute the Jacobian, we need the partial derivatives of the residuals. They are

$$\nabla r_I^{\mathbf{x}}(\mathbf{T}) = \frac{\partial}{\partial \xi_i}(I_i \circ \mathbf{g}_{uv} \circ \mathbf{s}) \quad \text{(A4)}$$

$$= \nabla I_i(\mathbf{g}_{uv})\mathbf{J}_{\mathbf{g}_{uv}}(\mathbf{s})\mathbf{J}_{\mathbf{s}}(\xi), \quad \text{(A5)}$$

$$\nabla r_D^{\mathbf{x}}(\mathbf{T}) = \frac{\partial}{\partial \xi_i}(D_i \circ \mathbf{g}_{uv} \circ \mathbf{s} - \mathbf{g}_d \circ \mathbf{s}) \quad \text{(A6)}$$

$$= \nabla D_i(\mathbf{g}_{uv})\mathbf{J}_{\mathbf{g}_{uv}}(\mathbf{s})\mathbf{J}_{\mathbf{s}}(\xi) - \mathbf{J}_{\mathbf{g}_d}(\mathbf{s})\mathbf{J}_{\mathbf{s}}(\xi). \quad \text{(A7)}$$

Algorithm 1 RGB-D image alignment

Input: A pair of RGB-D images (I_i, D_i) and (I_j, D_j), initial transformation $\mathbf{T}^0$
Output: Transformation $\mathbf{T}$ that aligns (I_j, D_j) to (I_i, D_i)
1: Build image pyramids $\{(I_i^l, D_i^l)\}$ and $\{(I_j^l, D_j^l)\}$
2: Precompute image gradients $\{(\nabla I_i^l, \nabla D_i^l)\}$
3: $\mathbf{T} \leftarrow \mathbf{T}^0, L \leftarrow \text{max\_pyramid\_level}$
4: **for** $l \in \{L, L-1, \cdots, 0\}$ **do** ▷ From coarsest to finest
5:     **while** not converged **do**
6:         $\mathbf{r} \leftarrow \mathbf{0}, \mathbf{J_r} \leftarrow \mathbf{0}$
7:         Compute a mask M that prunes correspondences
8:         **for** $\mathbf{x} \in M$ **do**
9:             Compute $r_I^{\mathbf{x}}, r_D^{\mathbf{x}}$ at $\mathbf{T}$ (Eq. A2,A3)
10:        Compute $\nabla r_I^{\mathbf{x}}, \nabla r_D^{\mathbf{x}}$ at $\mathbf{T}$ (Eq. A5,A7)
11:        Update $\mathbf{r}$ and $\mathbf{J_r}$ accordingly
12:        Solve linear system 21 to get ξ
13:        Update $\mathbf{T}$ using Equation 20 then map to $SE(3)$
14: Validate if $\mathbf{T}$ aligns (I_j, D_j) to (I_i, D_i)

Steps A5 and A7 apply the chain rule. ∇I_i and ∇D_i are the gradient of I_i and D_i respectively. They are computed by applying a normalized Scharr kernel over I_i and D_i. $\mathbf{J}_{\mathbf{g}_{uv}}$ and $\mathbf{J}_{\mathbf{g}_d}$ are the Jacobian matrices of $\mathbf{g}_{uv}$ and $\mathbf{g}_d$, derived from (5). $\mathbf{J_s}$ is the Jacobian of $\mathbf{s}$ with respect to ξ, derived from (4) and (20).

Correspondence pruning

Equation (2) constructs a correspondence from pixel $\mathbf{x}$ in image (I_j, D_j) to pixel $\mathbf{x}'$ in image (I_i, D_i). Since two images are viewed from different perspectives, $\mathbf{x}'$ can be occluded in image (I_i, D_i). In this case the correspondence is invalid and can corrupt the optimization. We compare $D_i(\mathbf{x}')$ and $\mathbf{g}_d(\mathbf{s}(\mathbf{h}(\mathbf{x}, D_j(\mathbf{x})), \mathbf{T}))$. If $\mathbf{x}'$ is occluded, the two depth values are apart. We use this criterion to create an image mask M that prunes invalid correspondences

$$M = \left\{ \mathbf{x} \,\middle|\, \mathbf{x} \in (I_j, D_j) \text{ and } \left|r_D^{\mathbf{x}}(\mathbf{T}^k)\right| < \delta \right\}. \quad \text{(A8)}$$

$r_D^{\mathbf{x}}$ is defined in (A3). δ is an empirical threshold: 7 centimeters. In each iteration, we recompute M and optimize objective (A1) over correspondences that fall within M.

Coarse-to-fine processing

Similar to Section 4.4, we apply the optimization in a coarse-to-fine manner: an RGB-D image pyramid is built and the optimization is performed from the coarsest pyramid level to the finest. This makes the algorithm more robust to bad initialization. Similar ideas have been exploited in [6, 4].

Algorithm 1 summarizes our RGB-D image alignment algorithm.

Appendix B: Parameter σ

The joint photometric and geometric optimization objectives (7) and (12) have a parameter σ that balances the photometric term and the geometric term. We find its optimal value by grid search.
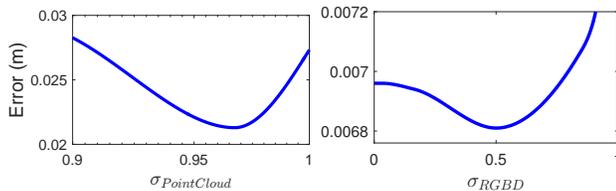

Figure 1. Grid search of σ for colored point cloud registration (left) and for RGB-D image alignment (right).

To find the optimal σ for colored point cloud registration, we took the experiment setup in Section 7.1 and perturb the true pose in the rotational components by $10°$. The average RMSE as a function of σ is shown in Figure 1. This function has a U-shape, indicating that optimizing a joint objective achieves lower error than optimizing a photometric objective or geometric objective alone. The optimal σ for colored point cloud registration is 0.968. Similarly, we can setup an experiment for RGB-D image alignment and find its optimal σ. It is 0.5.

Appendix C: Evaluation on the Cathedral Scene

The setup of this evaluation is detailed in Section 7.1. The results are shown in Figure 2.

Appendix D: Qualitative Evaluation of Scene Reconstruction

Section 7.2 summarized the quantitative performance of different scene reconstruction systems on the presented dataset. (Table 3 in the paper.) Accuracy was measured by the F-score (harmonic mean of precision and recall) for a fixed threshold $\tau = 20$ millimeters. In this appendix we report the F-score for each reconstruction system on each scene for varying thresholds τ. These results are shown in

Figure 3. Our system outperforms the baselines across distance thresholds.

Appendix E: Visualization of the Dataset

Figure 4 shows the ground-truth models of the five scenes in the dataset presented in Section 6. For this visualization, the ground-truth point clouds were meshed using Poisson surface reconstruction [3]. The renderings thus exhibit meshing artifacts that are not present in the ground-truth point clouds themselves.

Appendix F: SceneNN Scenes

Figure 3 in the paper showed results on two randomly sampled scenes from the SceneNN dataset [2]. In Figure 5 in this supplement, we show five more scenes from that dataset.

References

[1] S. Choi, Q.-Y. Zhou, and V. Koltun. Robust reconstruction of indoor scenes. In *CVPR*, 2015. 5

[2] B. Hua, Q. Pham, D. T. Nguyen, M. Tran, L. Yu, and S. Yeung. SceneNN: A scene meshes dataset with annotations. In *3DV*, 2016. 2, 5

[3] M. M. Kazhdan and H. Hoppe. Screened Poisson surface reconstruction. *ACM Transactions on Graphics*, 32(3), 2013. 2

[4] C. Kerl, J. Sturm, and D. Cremers. Robust odometry estimation for RGB-D cameras. In *ICRA*, 2013. 2

[5] H. Kim and A. Hilton. Influence of colour and feature geometry on multi-modal 3D point clouds data registration. In *3DV*, 2014. 3

[6] F. Steinbrücker, J. Sturm, and D. Cremers. Real-time visual odometry from dense RGB-D images. In *ICCV Workshops*, 2011. 2

[7] T. Whelan, R. F. Salas-Moreno, B. Glocker, A. J. Davison, and S. Leutenegger. ElasticFusion: Real-time dense SLAM and light source estimation. *International Journal of Robotics Research*, 35(14), 2016. 5
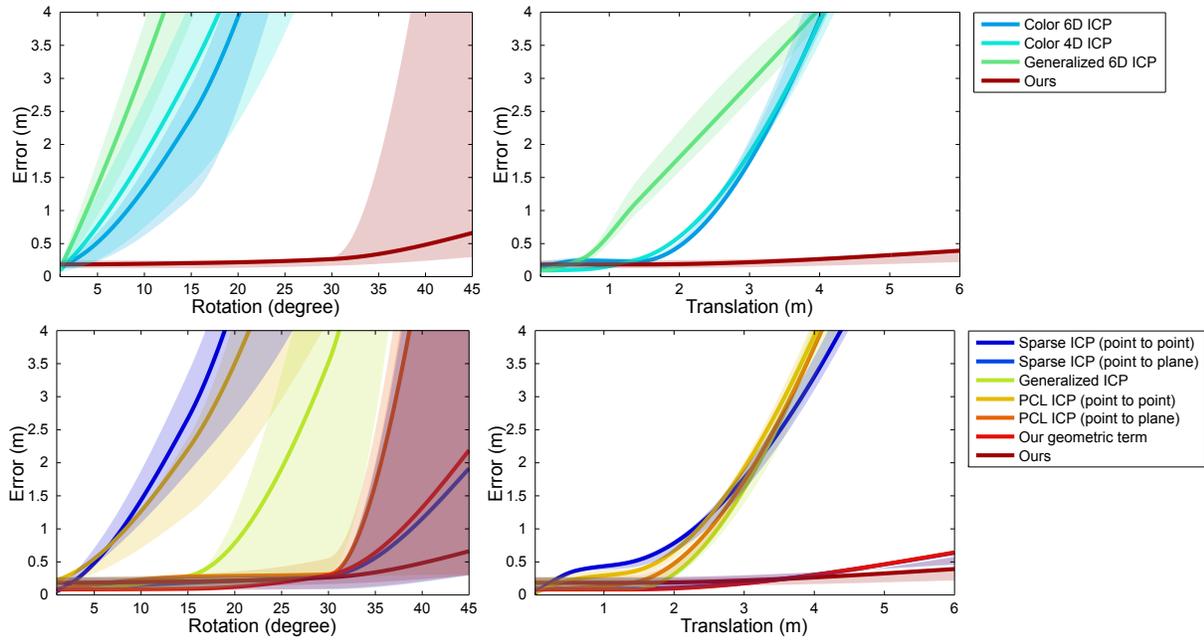
Figure 2. Evaluation of LiDAR point cloud alignment on the Cathedral Scene [5]. The presented algorithm is compared to prior algorithms that use color (top) and to algorithms that do not (bottom). The algorithms are initialized with transformations that are perturbed away from the true pose in the rotational component (left) and the translational component (right). The plot shows the median RMSE at convergence (bold curve) and the 40%-60% range of RMSE across trials (shaded region). Lower is better.
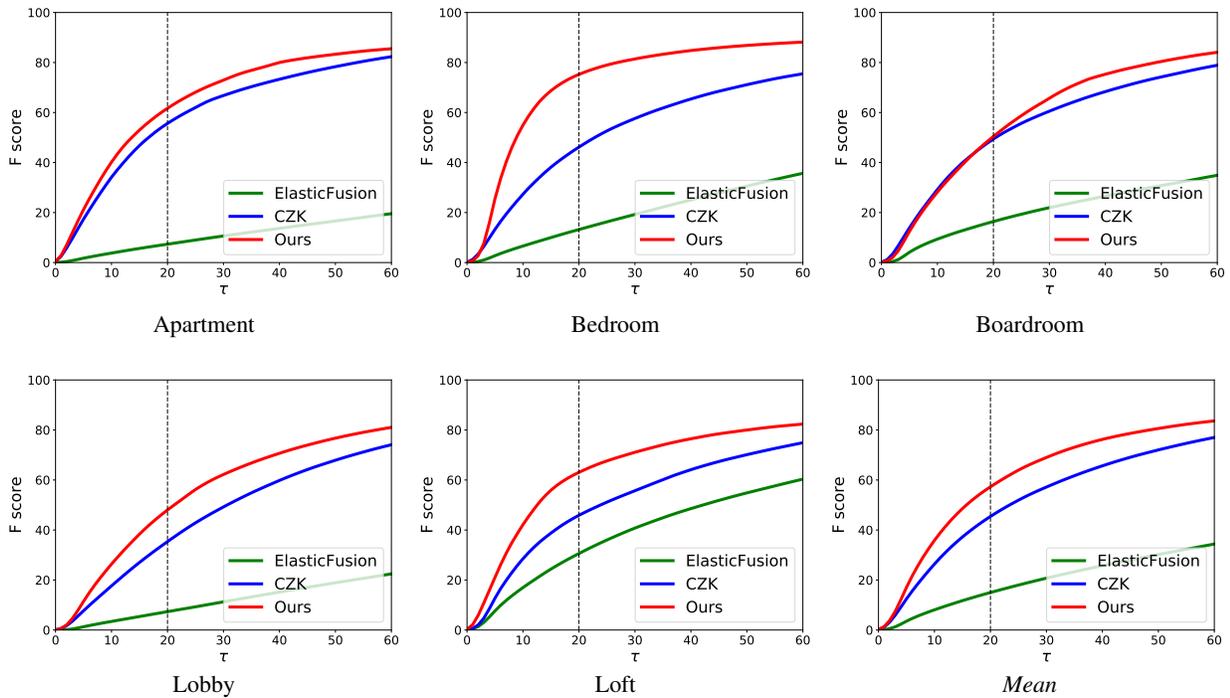

Figure 3. Accuracy of different reconstruction systems on the five scenes in the presented dataset, measured by F-score with varying distance thresholds τ (in millimeters). The last plot shows the mean F-score for each threshold over all five scenes in the dataset.
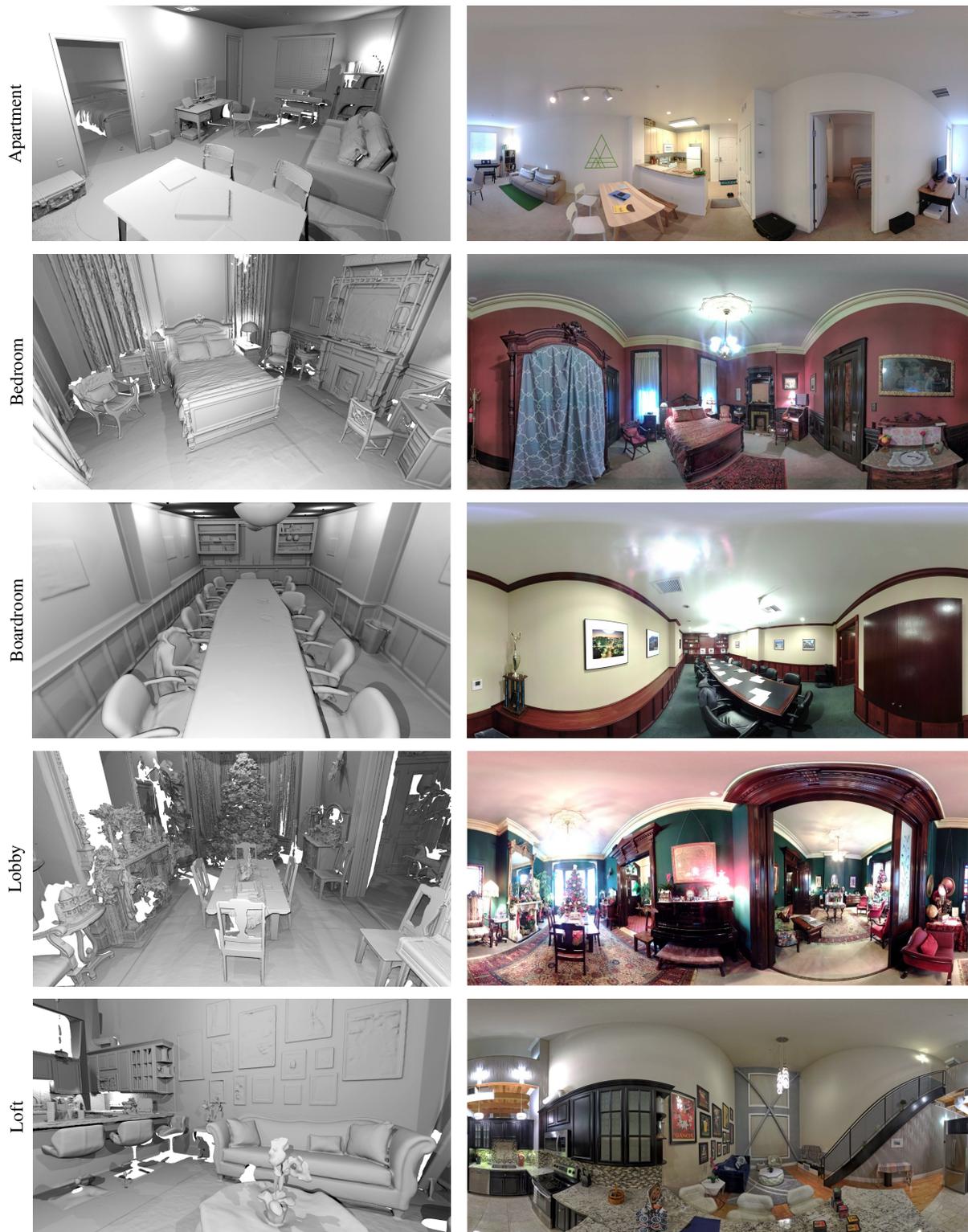
Figure 4. Visualization of the dataset presented in Section 6 of the paper. Left: ground-truth model of each scene, acquired using an industrial laser scanner. For this visualization, the ground-truth point clouds were meshed using Poisson surface reconstruction, and the renderings exhibit meshing artifacts that are not present in the ground-truth point clouds themselves. Right: 360° panoramic images of the physical scenes that were scanned.
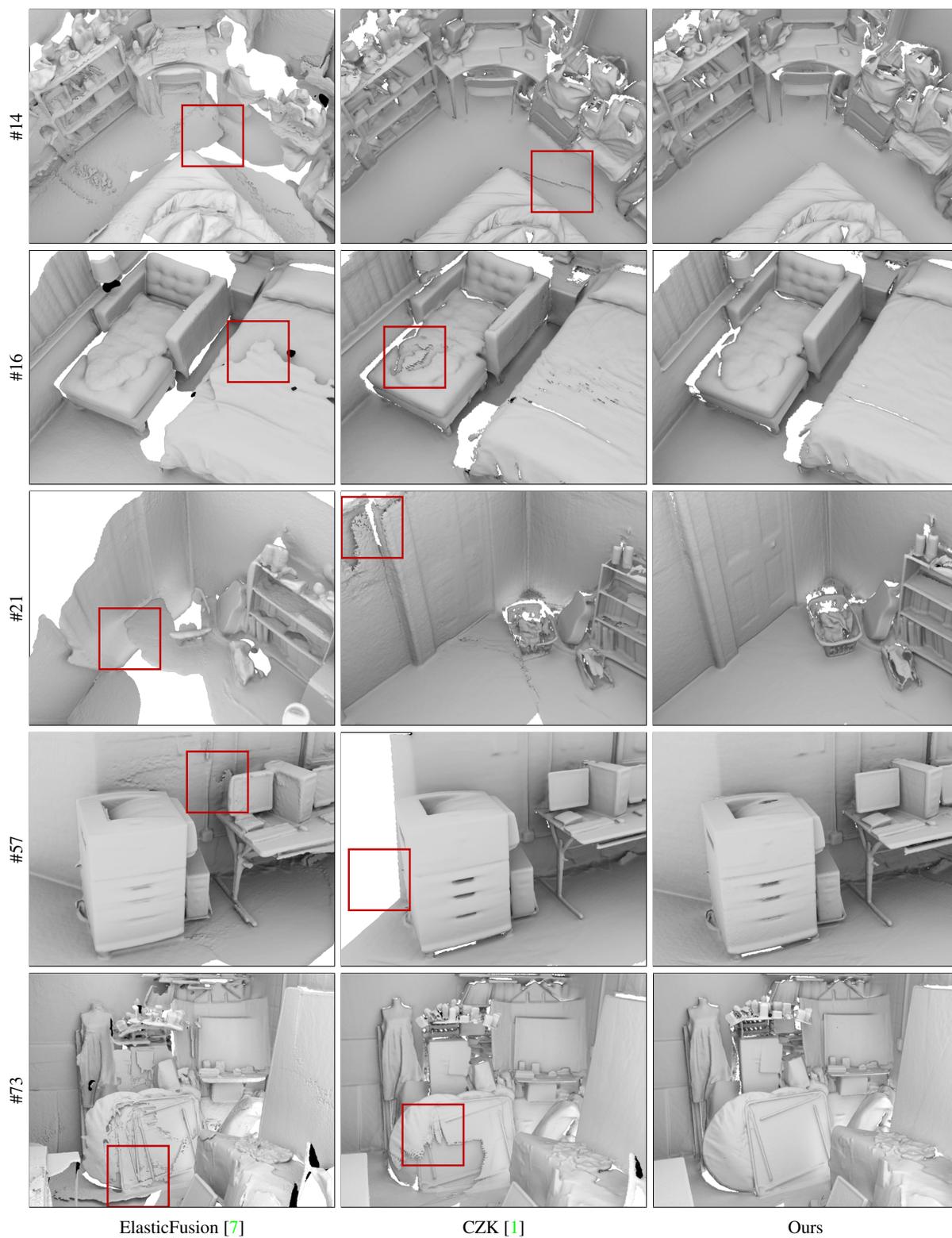
Figure 5. Reconstructions of five more scenes from the SceneNN dataset [2]. This extends Figure 3 in the paper. Prior systems suffer from inaccurate surface alignment and produce broken geometry. Our system produces much cleaner results.