

Supplementary Material for Exploiting Object Similarity in 3D Reconstruction

Chen Zhou^{1,2} Fatma Güney³ Yizhou Wang^{1,2} Andreas Geiger³

¹Nat'l Engineering Laboratory for Video Technology

²Cooperative Medianet Innovation Center, Peking University, China

³MPI for Intelligent Systems Tübingen

{zhouch, yizhou.wang}@pku.edu.cn

{fatma.guney, andreas.geiger}@tue.mpg.de

Abstract

In this supplementary material we first provide details of our experimental setup and the inference process with respect to the individual optimization blocks. Furthermore, we provide additional quantitative and qualitative results: First, we show the full plots corresponding to Figure 5 in the main paper. Second, we visualize the accuracy and completeness of the individual methods. And finally, we qualitatively compare the results of the proposed approach to the initial reconstruction baseline on novel sequences.

1. Experimental Setup

For capturing our dataset and the 3D ground truth, we used a Velodyne HDL-64 laser scanner, running at 10 fps. For synchronization, we attached a hardware trigger to the scanner (reed contact) which triggers all cameras at the moment the scanner is facing forward. Each scan returns the timestamp of each point which is used to unwarp the point cloud according to the vehicle motion obtained using a high-precision OXTS 3003 RTK GPS/IMU unit (2cm accuracy). The cameras have a global shutter and are in perfect sync. The average driving velocity was 40 km/h.

2. Inference

This section provides details on optimizing the individual blocks in our energy function

$$\operatorname{argmin}_{\pi, \phi, \mathbf{X}, \mathbf{V}, \mathbf{k}} \sum_{i=1}^N \Psi_i(\pi, \phi, \mathbf{X}, \mathbf{V}, \mathbf{k}) \quad (1)$$

where

$$\begin{aligned} \Psi_i(\cdot) &= \psi_{shp}(\pi_i, \phi_{k_i}, \mathbf{x}_i) + \lambda_{scale} \psi_{scale}(\mathbf{s}_i, \mathbf{v}_{k_i}) \\ &\quad + \lambda_{reg} \psi_{reg}^{(i)}(\pi_i) \end{aligned} \quad (2)$$

The reader is referred to section 3.3 of the paper for the definition of each term.

2.1. Block $\{\pi, \mathbf{V}\}$

We optimize Eq. 1 with respect to $\{\pi, \mathbf{V}\}$ via gradient descent. For clarity of exposition we consider a single term Ψ_i , dropping the observation index i . First, we note that the derivatives of $\psi_{scale}(\mathbf{s}, \mathbf{v}_k)$ and $\psi_{reg}(\pi)$ with respect to $\{\pi, \mathbf{V}\}$ are readily given due to their simple quadratic form. The derivative of $\psi_{shp}(\pi, \phi, \mathbf{x})$ with respect to π is given by

$$\frac{\partial \psi_{shp}(\pi, \phi, \mathbf{x})}{\partial \pi} = \sum_{\mathbf{p} \in \Omega} (\phi(\mathbf{p}, \mathbf{x}) - d(\pi(\mathbf{p})))^2 \frac{\partial w}{\partial \pi} - 2w(\pi(\mathbf{p}))(\phi(\mathbf{p}, \mathbf{x}) - d(\pi(\mathbf{p}))) \frac{\partial d}{\partial \pi} \quad (3)$$

Using \mathbf{p}_t to denote $\pi(\mathbf{p}) \in \mathbb{R}^3$ (the transformed point) and the chain rule we have:

$$\frac{\partial w}{\partial \pi} = \frac{\partial w}{\partial \mathbf{p}_t} \frac{\partial \mathbf{p}_t}{\partial \pi} \quad (4)$$

$$\frac{\partial d}{\partial \pi} = \frac{\partial d}{\partial \mathbf{p}_t} \frac{\partial \mathbf{p}_t}{\partial \pi} \quad (5)$$

Where $\frac{\partial w}{\partial \mathbf{p}_t}$ and $\frac{\partial d}{\partial \mathbf{p}_t}$ are gradients of the weight field and the signed distance function field which we obtain using bilinear interpolation. Finally, we compute $\frac{\partial \mathbf{p}_t}{\partial \pi}$. According to our parametrization, the transformed point is given by $\mathbf{p}_t = \mathbf{R}\mathbf{S}\mathbf{p} + [\mathbf{t} \ 0]^T$, where \mathbf{R} is the rotation matrix formed by 2D rotation angle r (i.e., $\begin{bmatrix} \cos r & -\sin r & 0 \\ \sin r & \cos r & 0 \\ 0 & 0 & 1 \end{bmatrix}$), \mathbf{S} is the diagonal scaling matrix (i.e. $\text{diag}(\mathbf{s})$) and \mathbf{t} is the 2D translation vector in the x-y plane (i.e., ground plane).

Thus, we obtain:

$$\frac{\partial \mathbf{p}_t}{\partial r} = \frac{\partial(\mathbf{R}\mathbf{S}\mathbf{p} + \mathbf{t})}{\partial r} \quad (6)$$

$$= \begin{bmatrix} -\sin(r)s_x p_x - \cos(r)s_y p_y \\ \cos(r)s_x p_x - \sin(r)s_y p_y \\ 0 \end{bmatrix} \quad (7)$$

$$\frac{\partial \mathbf{p}_t}{\partial \mathbf{s}} = \frac{\partial(\mathbf{R}\mathbf{S}\mathbf{p} + \mathbf{t})}{\partial \mathbf{s}} \quad (8)$$

$$= \begin{bmatrix} \cos(r)p_x & -\sin(r)p_y & 0 \\ \sin(r)p_x & \cos(r)p_y & 0 \\ 0 & 0 & p_z \end{bmatrix} \quad (9)$$

$$\frac{\partial \mathbf{p}_t}{\partial \mathbf{t}} = \begin{bmatrix} \mathbf{I}_{2 \times 2} \\ \mathbf{0}_{1 \times 2} \end{bmatrix} \quad (10)$$

2.2. Block $\{\phi, \mathbf{X}\}$

We solve the M independent weighted PCA problems using the robust approach of Torre et al. [1].

2.3. Block $\{\mathbf{X}, \mathbf{k}\}$

Recall that $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ are the coefficient vectors for all observations and $\mathbf{k} = \{k_1, \dots, k_N\}$ denote the model associations with $k_i \in \{1, \dots, M\}$. With slight change in notation, let the linear model associated with observation i be described by the mean $\boldsymbol{\mu}_{k_i} \in \mathbb{R}^{|\Omega|}$ and the D -dimensional orthonormal basis $\boldsymbol{\xi}_{k_i}^d \in \mathbb{R}^{|\Omega|}$, $d \in \{1, \dots, D\}$. The coefficient vector \mathbf{x}_i for observation i is obtained by solving the following weighted least squares estimation problem (in closed form):

$$\mathbf{B}_{k_i}^T \mathbf{W}_i \mathbf{B}_{k_i} \mathbf{x}_i = \mathbf{B}_{k_i} \mathbf{W}_i (\mathbf{d}_i - \boldsymbol{\mu}_{k_i}) \quad (11)$$

Here, $\mathbf{B} = \{\boldsymbol{\xi}_{k_i}^1, \dots, \boldsymbol{\xi}_{k_i}^D\}$ denotes the $|\Omega| \times D$ basis matrix, $\mathbf{W}_i = \text{diag}(w(\pi_i(\mathbf{p}))) \in \mathbb{R}^{|\Omega| \times |\Omega|}$, and $\mathbf{d}_i = d(\pi_i(\mathbf{p})) \in \mathbb{R}^{|\Omega|}$ is the vector of signed distance function values for observation i . We estimate the solution to Eq. 11 for each $k_i \in \{1, \dots, M\}$ in order to find the minimizer of Eq. 9 in the main paper.

3. Additional Results

In this section, we first show the full plots for all methods corresponding to Figure 5 in the main paper (Fig. 1). Next, we visualize the accuracy (Fig. 2) and the completeness (Fig. 3) of the different methods for the sequence presented in the main paper. We also include zoomed in comparisons of our fused result (PC 0) to our initial reconstruction in terms of both accuracy (Fig. 4, Fig. 6) and completeness (Fig. 5, Fig. 7). Finally, we show additional qualitative results on other sequences from our dataset in Fig. 8 – Fig. 12.

References

[1] F. D. la Torre and M. J. Black. Robust principal component analysis for computer vision. In *ICCV*, pages 362–369, 2001.

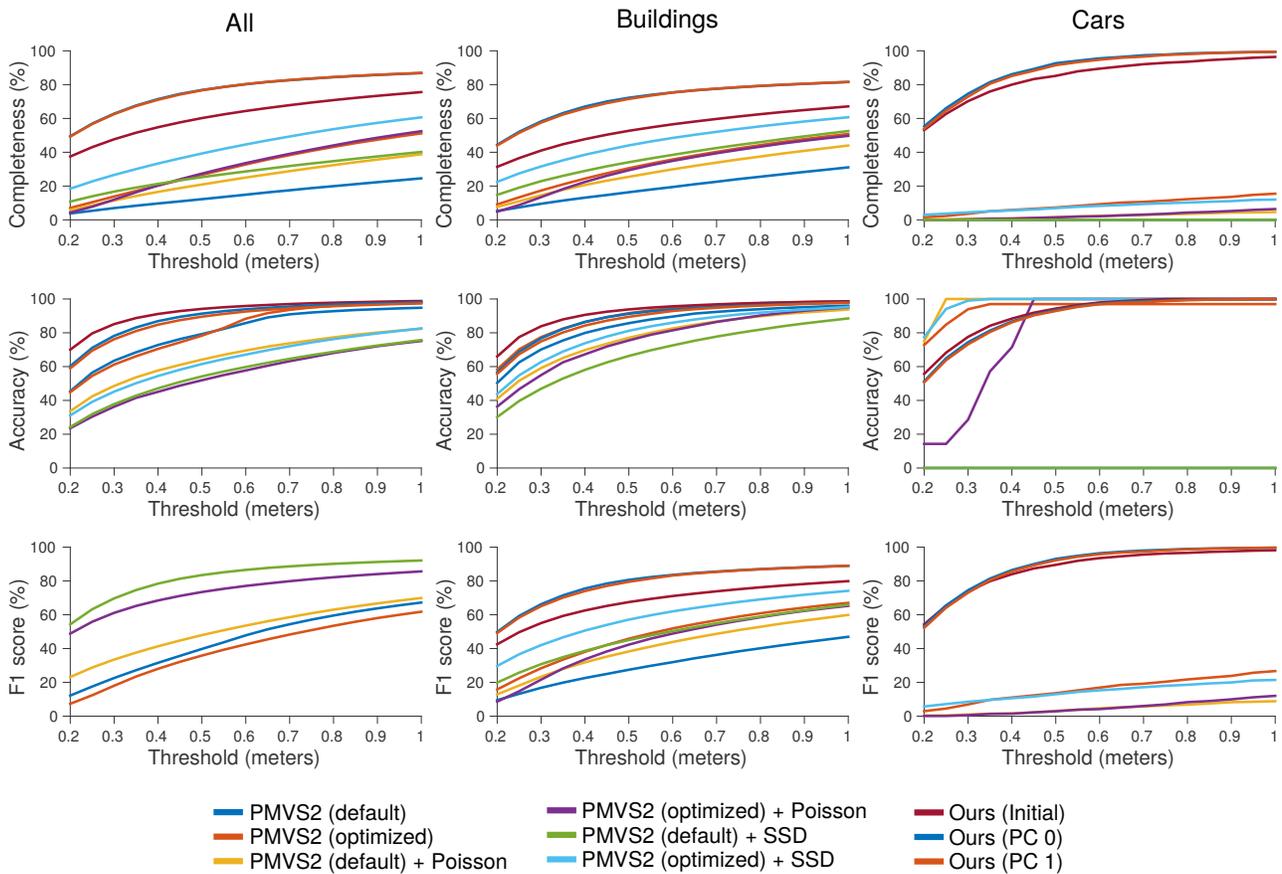


Figure 1: **Varying the Evaluation Distance.** This figure shows quantitative results in terms of completeness, accuracy and F1 score when varying the evaluation distance τ between 0.2 and 1.0 m.

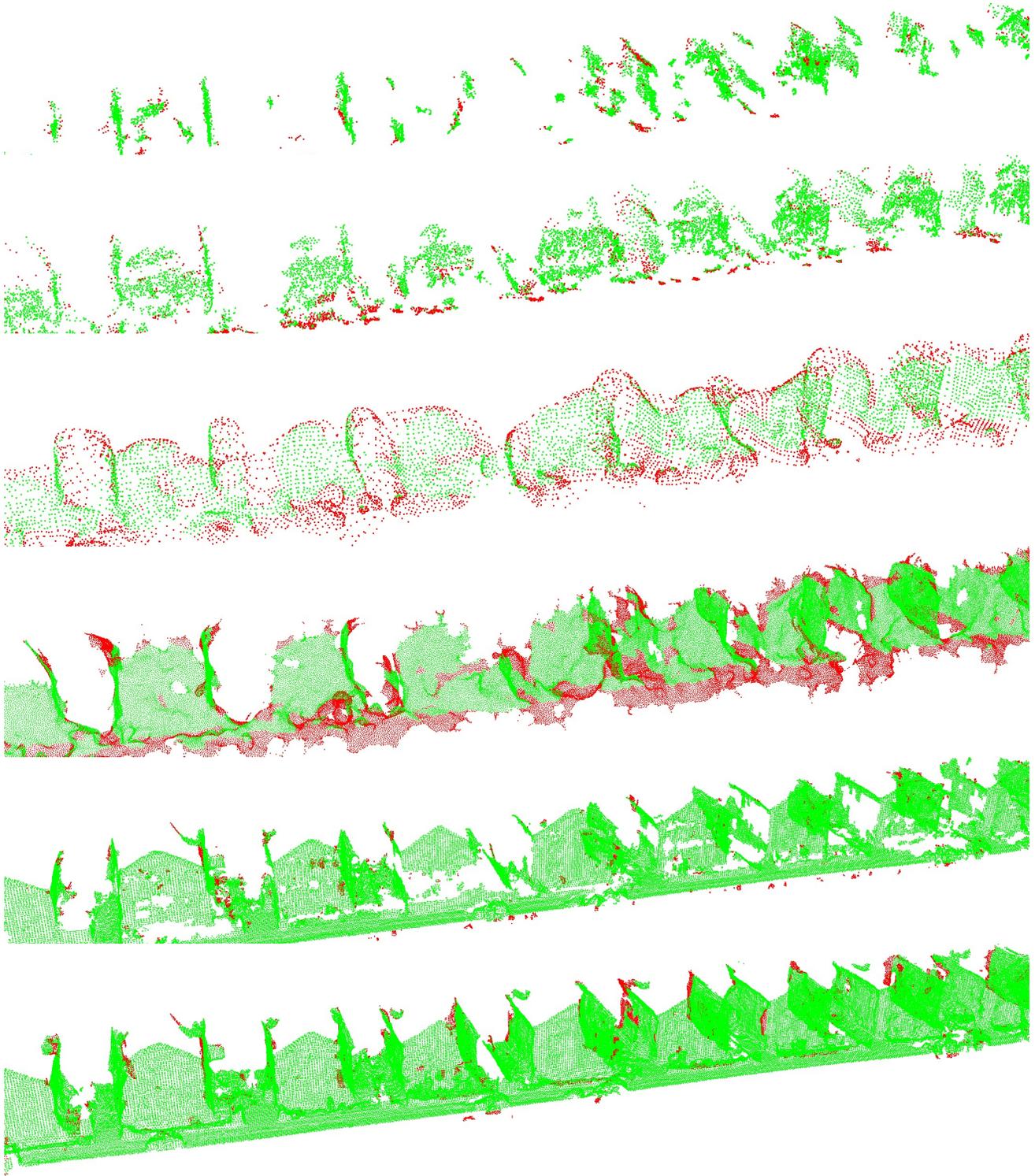


Figure 2: **Accuracy Visualization.** From top to bottom: Point cloud from PMVS2 using the default parameter setting, point cloud from PMVS2 with optimized parameters, PMVS2 (optimized) + Poisson, PMVS2 (optimized) + SSD, our initial reconstruction, and our fused result (PC 0). Green-coloured points are reconstructed 3D points for which at least one ground truth 3D point is within a distance of $\tau = 0.5$ m and red-coloured points are other (falsely) reconstructed points.

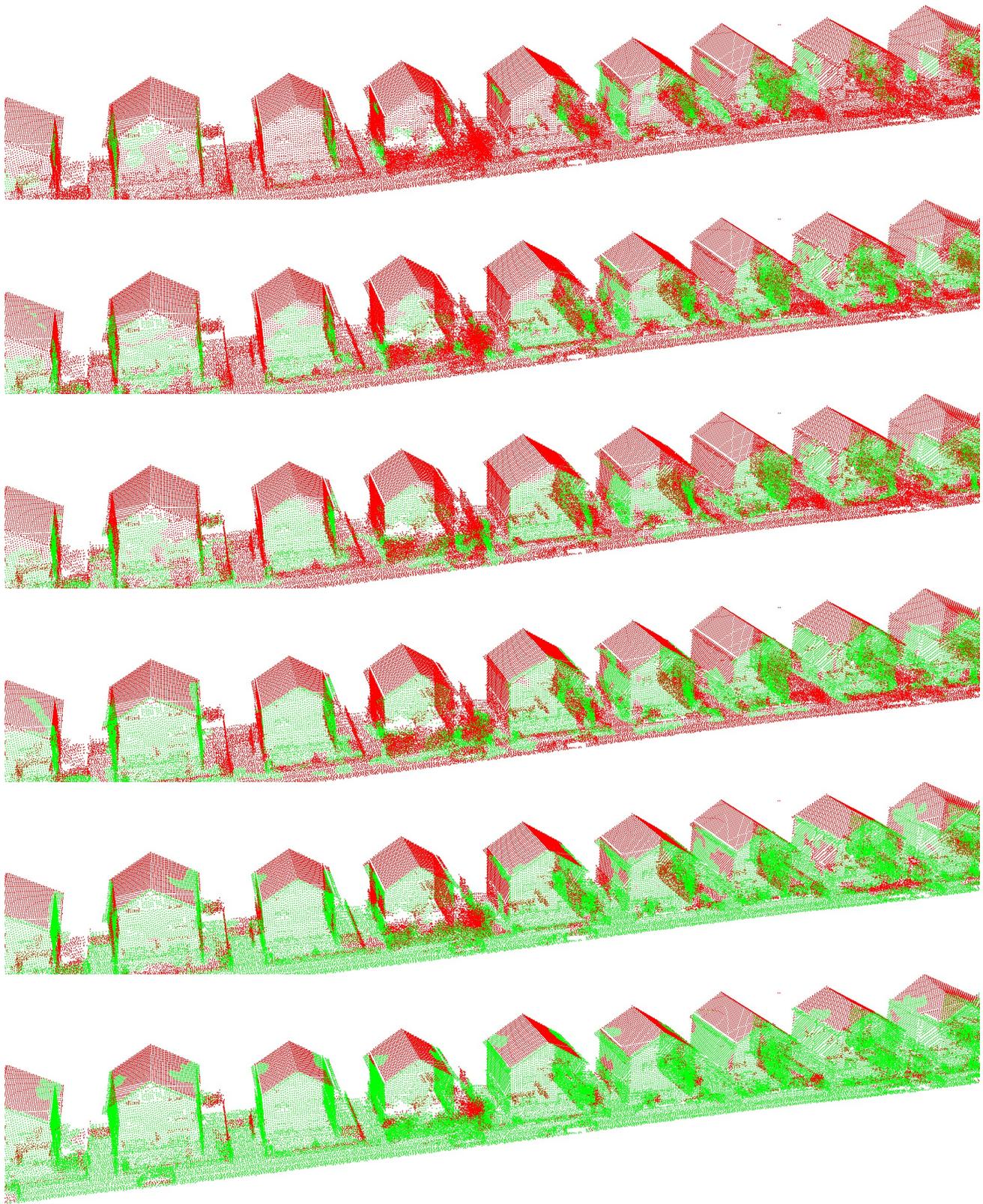


Figure 3: **Completeness Visualization.** From top to bottom: Point cloud from PMVS2 using the default parameter setting, point cloud from PMVS2 with optimized parameters, PMVS2 (optimized) + Poisson, PMVS2 (optimized) + SSD, our initial reconstruction, and our fused result (PC 0). Green-coloured points are ground truth 3D points for which at least one reconstructed 3D point is within a distance of $\tau = 0.5$ m and red-coloured points are other (missed) ground truth points.

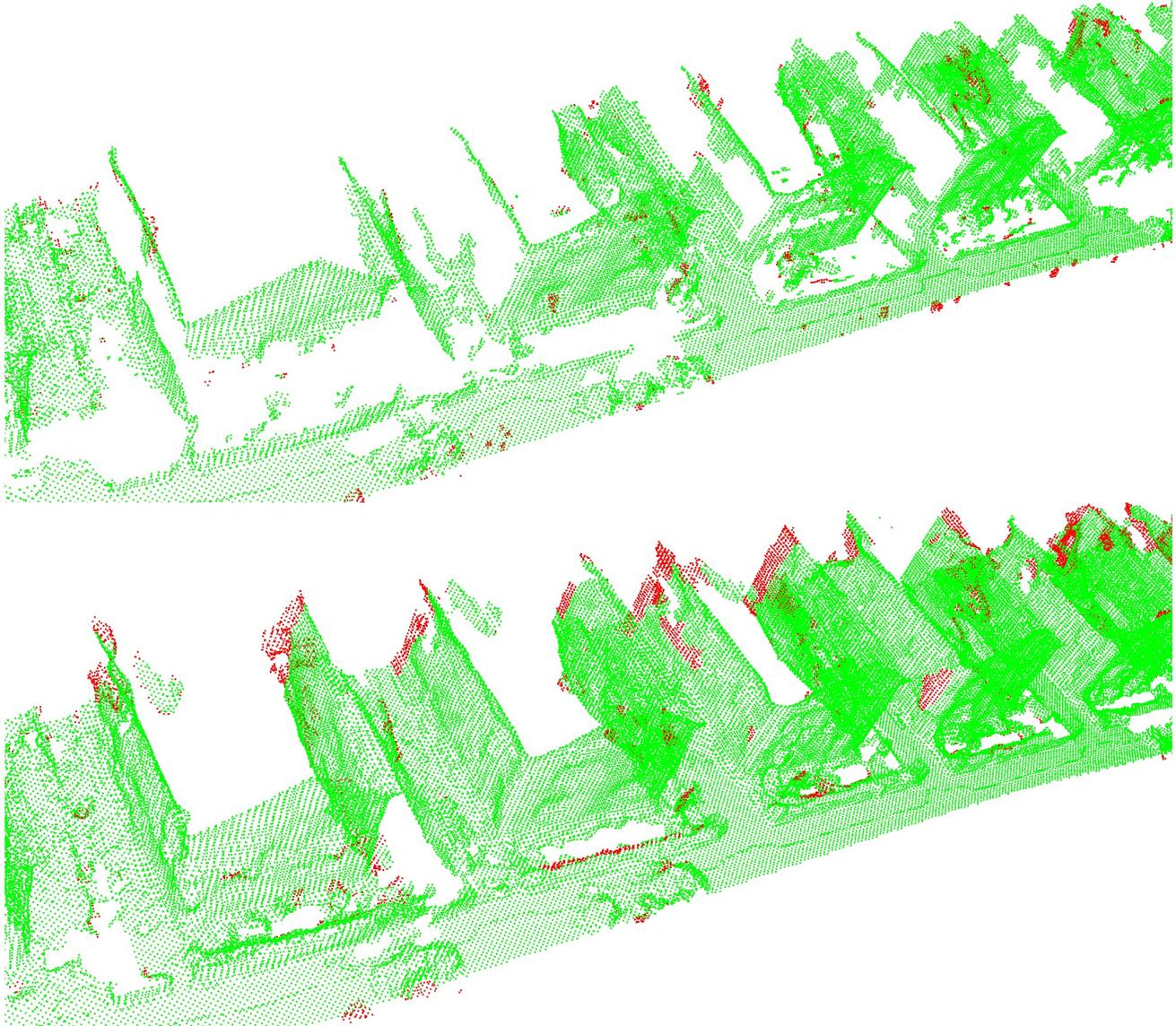


Figure 4: **Accuracy Visualization.** Our initial reconstruction and our fused result (PC 0). Green-coloured points are reconstructed 3D points for which at least one ground truth 3D point is within a distance of $\tau = 0.5$ m and red-coloured points are other (falsely) reconstructed points.

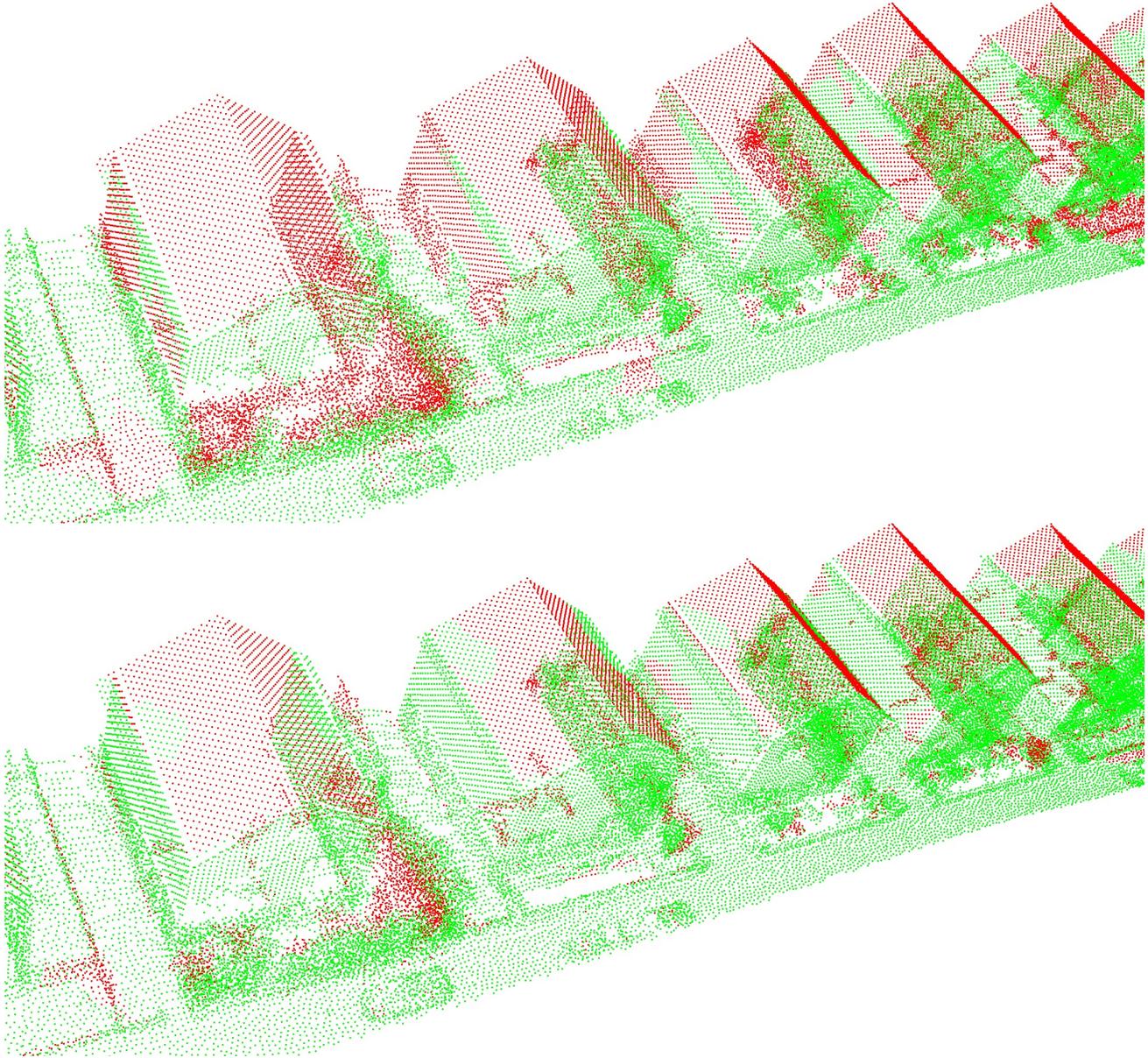


Figure 5: **Completeness Visualization.** Our initial reconstruction and our fused result (PC 0). Green-coloured points are ground truth 3D points for which at least one reconstructed 3D point is within a distance of $\tau = 0.5$ m and red-coloured points are other (missed) ground truth points.

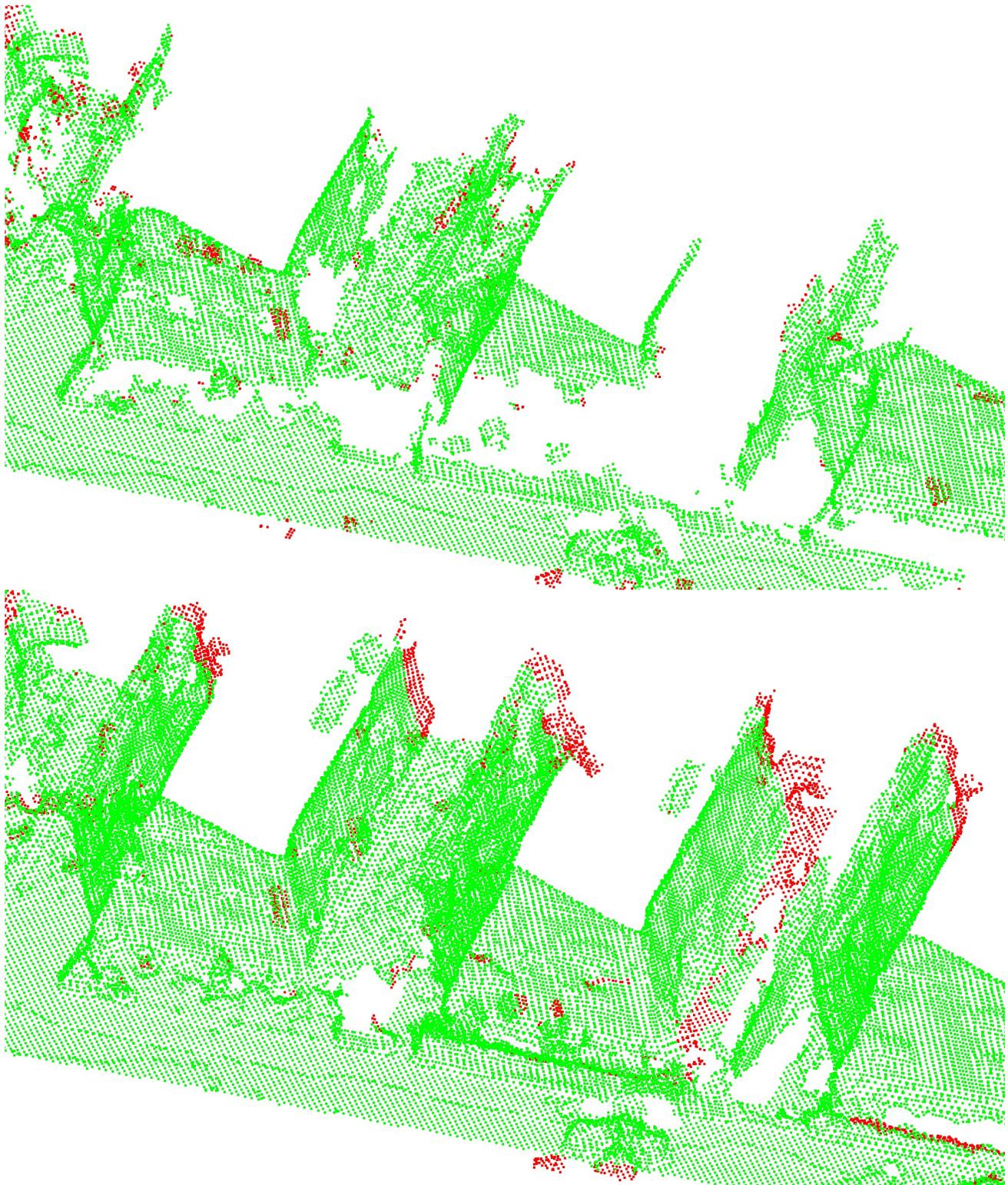


Figure 6: **Accuracy Visualization.** Our initial reconstruction and our fused result (PC 0). Green-coloured points are reconstructed 3D points for which at least one ground truth 3D point is within a distance of $\tau = 0.5$ m and red-coloured points are other (falsely) reconstructed points.

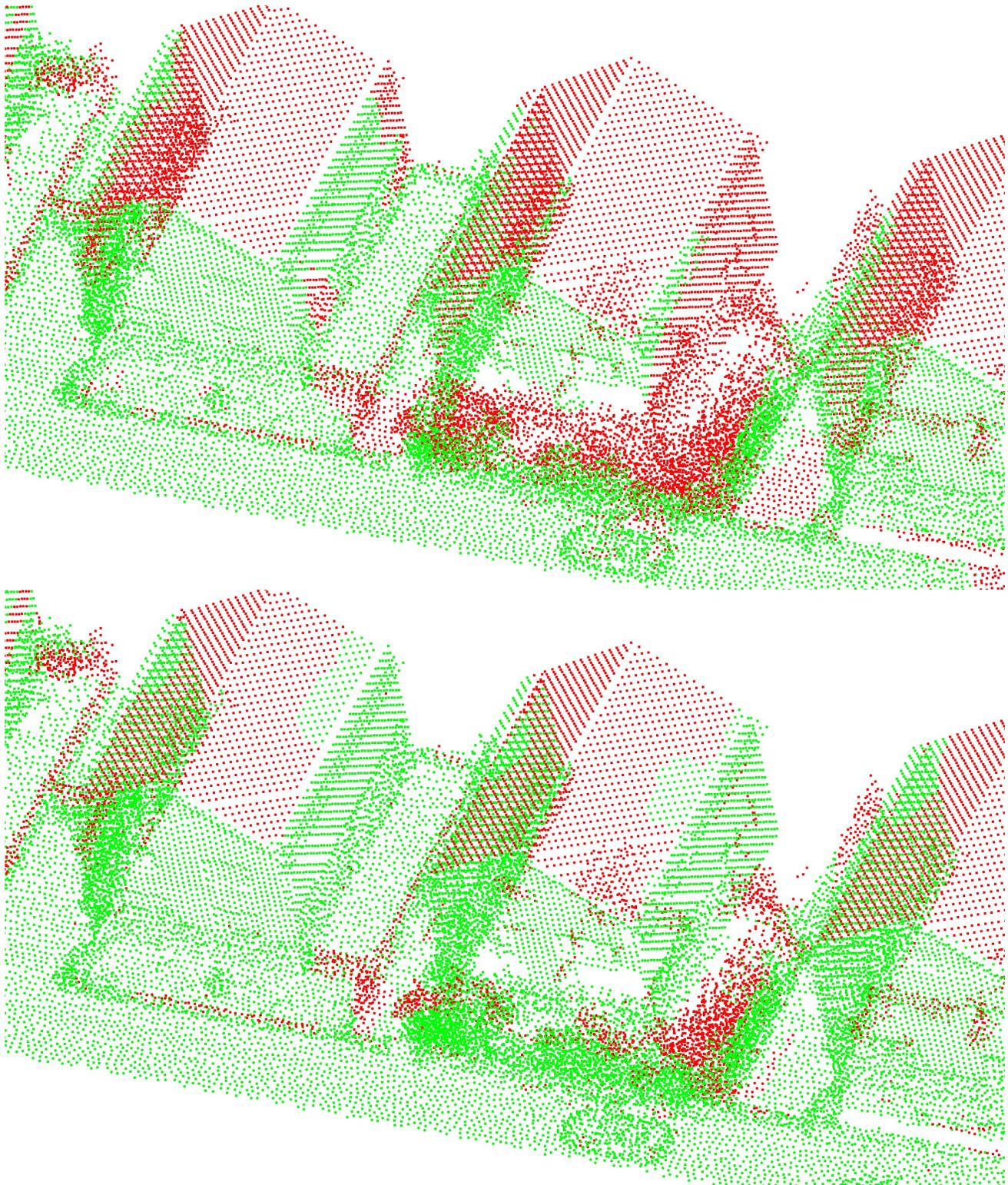


Figure 7: **Completeness Visualization.** Our initial reconstruction and our fused result (PC 0). Green-coloured points are ground truth 3D points for which at least one reconstructed 3D point is within a distance of $\tau = 0.5$ m and red-coloured points are other (missed) ground truth points.

4. Additional Qualitative Result

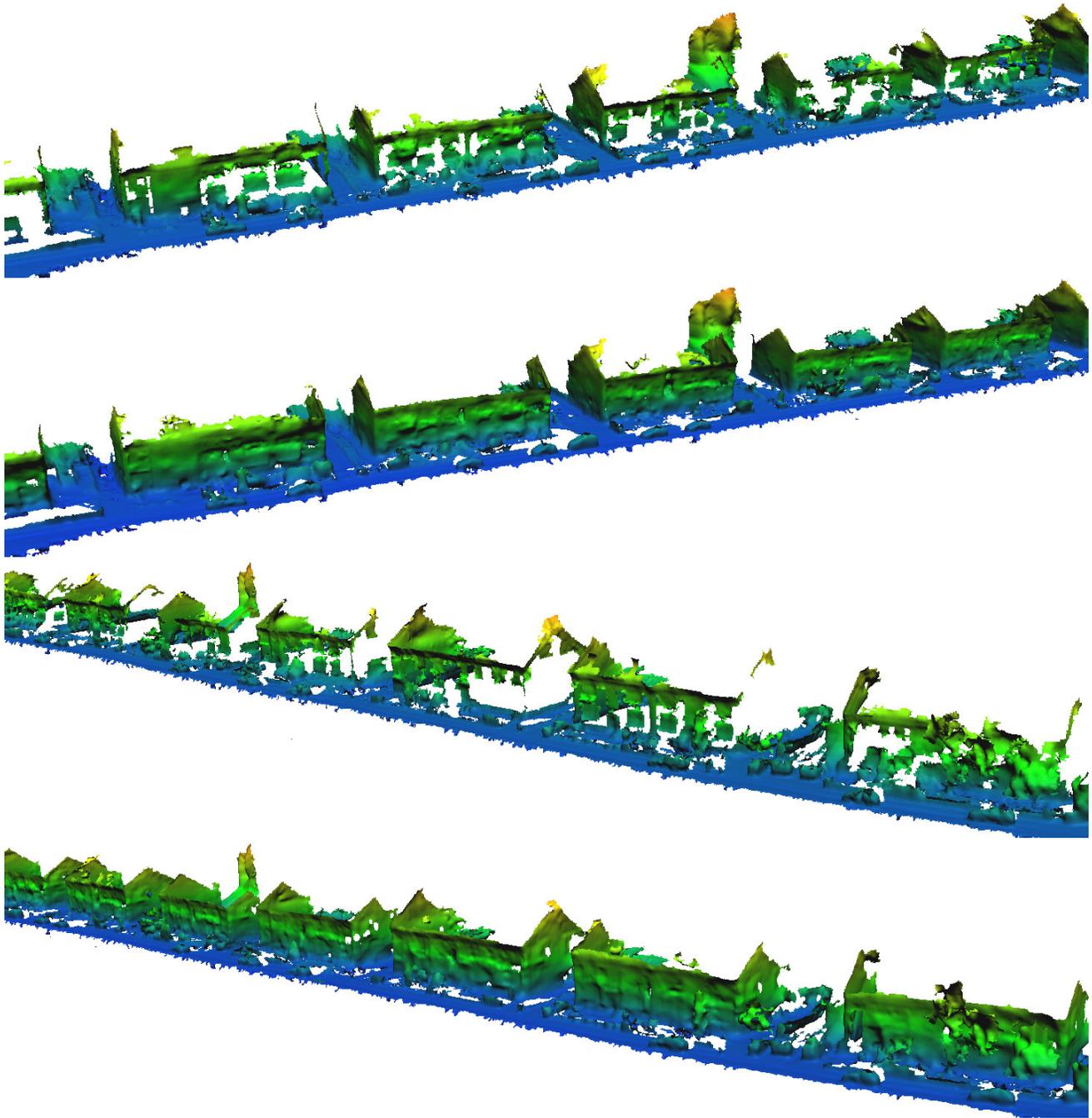


Figure 8: **Reconstruction of a sequence with elongated buildings from two different viewpoints.** The first and third rows show the original (initial) reconstruction, the second and fourth row show our final result. Note how our model is able to complete missing walls by reasoning jointly about buildings of similar shape.

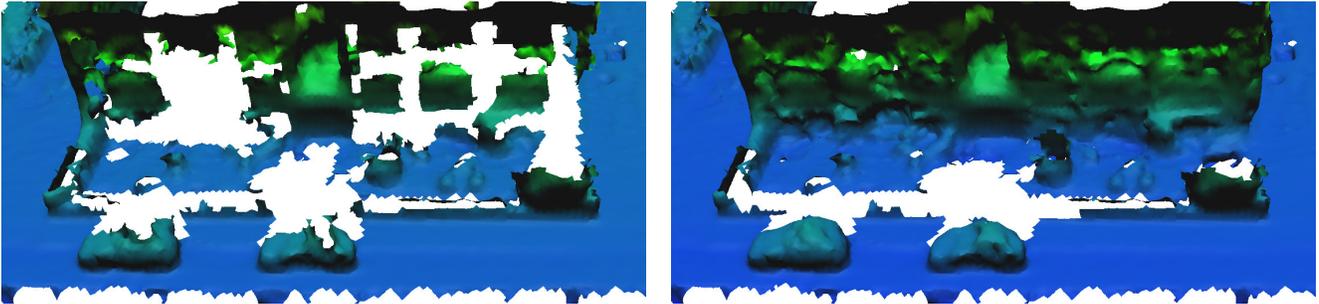


Figure 9: **Reconstruction of cars.** The left figure shows the original reconstruction, the right figure shows our final result.

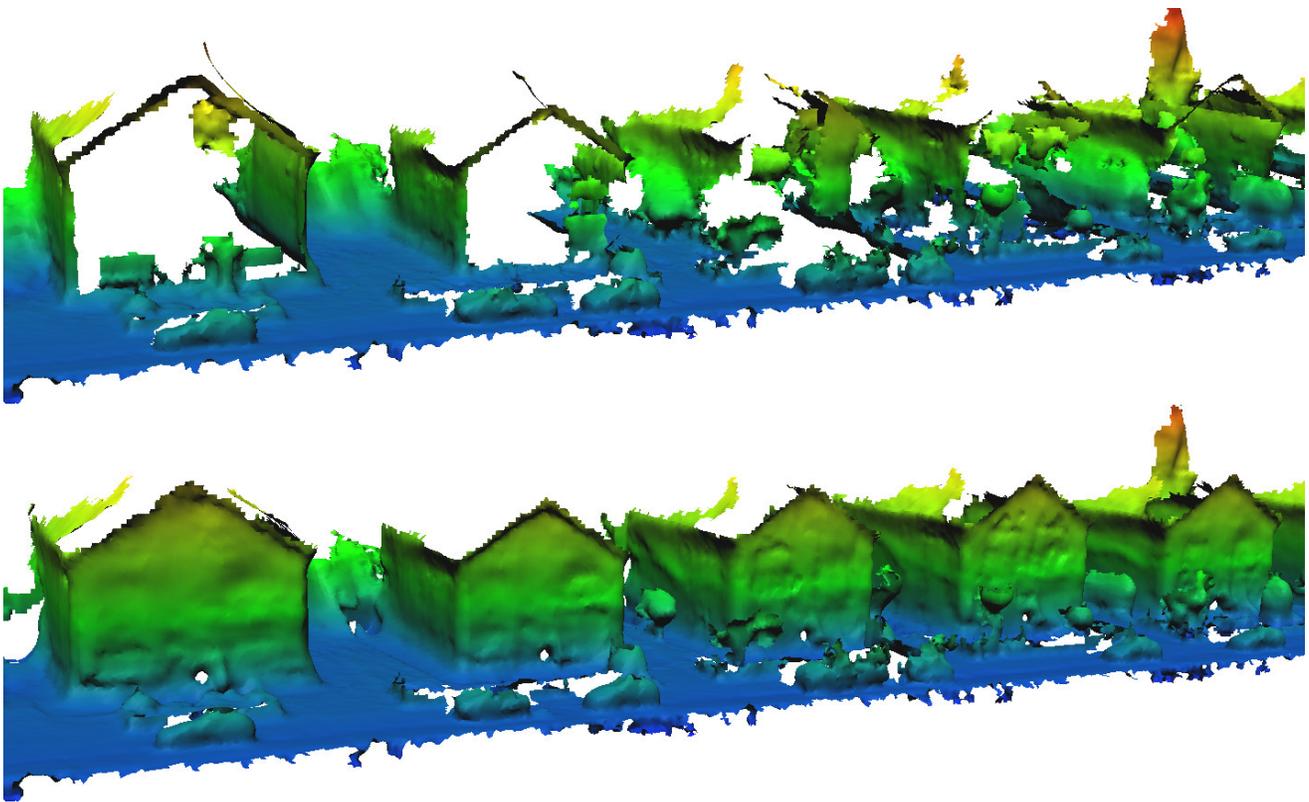


Figure 10: **Reconstruction of a third sequence.** The first row shows the original reconstruction, the second row shows our final result. Note how the proposed approach is to complete buildings even in the presence of very little surface information in the initial reconstruction.

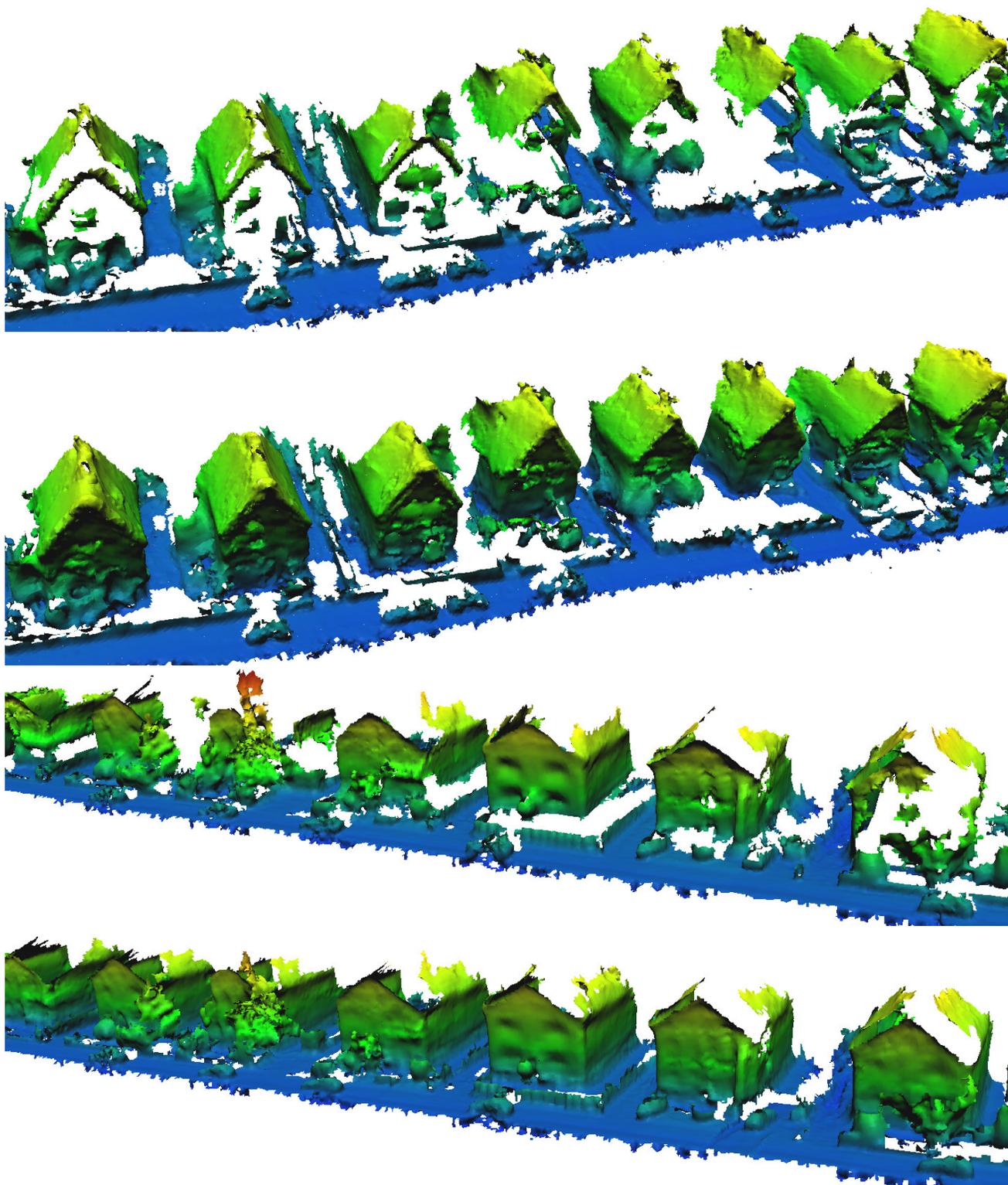


Figure 11: Comparison of reconstruction for two more sequences.

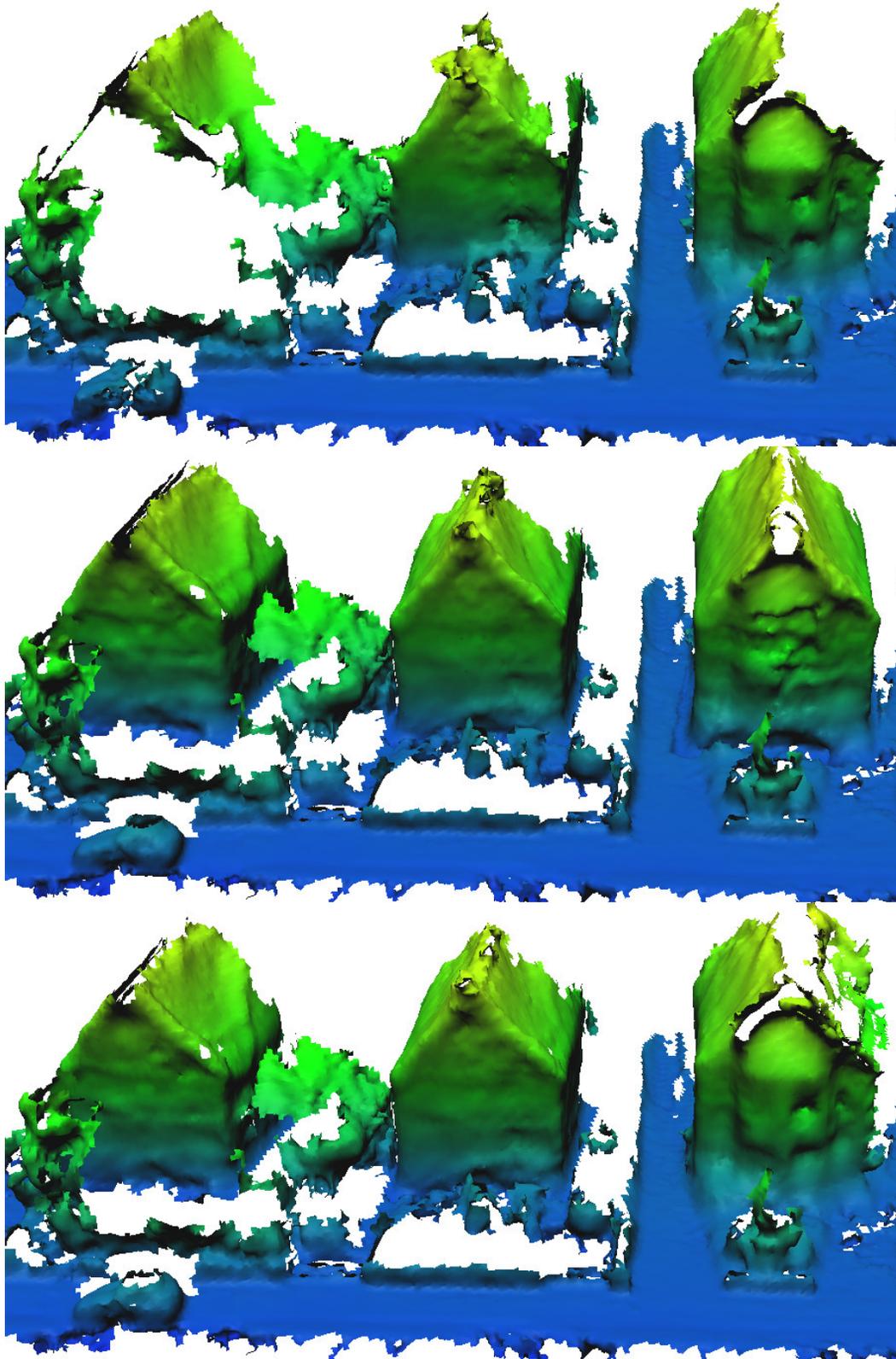


Figure 12: **Comparison of reconstruction result using PC0 and PC1.** From top to bottom: initial reconstruction, PC0 and PC1. Note for the righthmost building, PC1 correctly captures the variation in the front wall, however noise are introduced on the roof because of weaker regularization.