Supplementary Material

ANONYMOUS AUTHOR(S)

1 EVALUATION

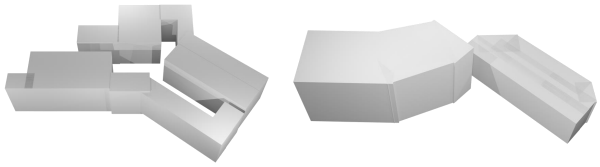

Fig. 1. The proxy we used to crop the scene for evaluation.

Because the NSR results often have many floaters that affect the evaluation, we first use ray casting according to the camera pose to generate a depth map under each image pose. Then, we project the depth map to form a point cloud. Next, we use the proxy to crop the point cloud and compare it with the ground truth.

2 MATCHING COST

We introduce the matching cost which used as a matching quality indicator in Patchmatch operation.

Given N images $\{\boldsymbol{I}_i\}_{i=0}^{N-1}$ with calibrated camera parameters $\{\boldsymbol{P}_i\}_{i=0}^{N-1}$, where $\boldsymbol{P}_i = \boldsymbol{K}_i [\boldsymbol{R}_i \mid \boldsymbol{t}_i]$, we first generate a random 3D plane hypothesis in local coordinate $\boldsymbol{\theta} = \left[\boldsymbol{n}^\top, d\right]^\top$ for each pixel in reference image I_r, where $\boldsymbol{n}$ is the normal of the plane and d is distance from the origin to the plane. The normalization cross correlation (NCC) is defined as:

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{x}', \boldsymbol{H} = \boldsymbol{K}_s \left(\boldsymbol{R}_s \boldsymbol{R}_r^T - \frac{\boldsymbol{R}_s \left(\boldsymbol{R}_s^T \boldsymbol{t}_s - \boldsymbol{R}_r^T \boldsymbol{t}_r \right) \boldsymbol{n}^T}{d} \right) \boldsymbol{K}_r^{-1}, \quad (1)$$

$$NCC\left(I_r\left(q_x\right), I_s\left(q_{x'}\right)\right) = \frac{\operatorname{Cov}\left(I_r\left(q_x\right), I_s\left(q_{x'}\right)\right)}{\sqrt{\operatorname{Var}\left(I_r\left(q_x\right)\right) \operatorname{Var}\left(I_s\left(q_{x'}\right)\right)}}, \quad (2)$$

where x' is the corresponding pixel for x in I_s, $\boldsymbol{H}$ is the plane-induced homography, q_x is the pixels in the 5x5 patch which take x as center, Cov represents the covariance and Var represents the variance. Then the mean of top k largest NCC is set as the final matching cost:

$$\boldsymbol{E}_{NCC} = \frac{1}{k} \left(\sum_k 1 - NCC_k \right). \quad (3)$$

3 GEONEUS

We also tested the GeoNeuS and Geoangelo on the datasets. The results are shown in Figure 7 and Figure 8. We train each method for 300k iterations per block. The Geoangelo adds the NCC module of GeoNeuS to the Bakedangelo. The Geoangelo implemented by us is basically the same as the Geoangelo implemented in SDFStudio regarding reconstruction results. First, the reconstruction results based on the MLP method are too smooth. Using hash tables as the scene representation is promising for large-scale scene reconstruction. In the experiment, we find that the scene often cannot be reconstructed if the NCC loss is provided. The first reason is that the NCC module requires that all sampled rays must come from pixels in one image, instead of randomly sampling from all image pixels like other NSR methods, the reconstruction quality is seriously degraded in large scenes. Another reason is that unlike depth prior, NCC cannot clearly indicate the difference between the current and target states. In a complex outdoor environment, NCC often causes optimization to fall into the local optimization, which fails to reconstruct the scene.

4 MORE EXPERIMENT RESULTS

Here we present more experimental results. Figure 2,3,4,5 show the entire scene mesh of each method on four datasets. Figure 6 shows more close-up comparisons.

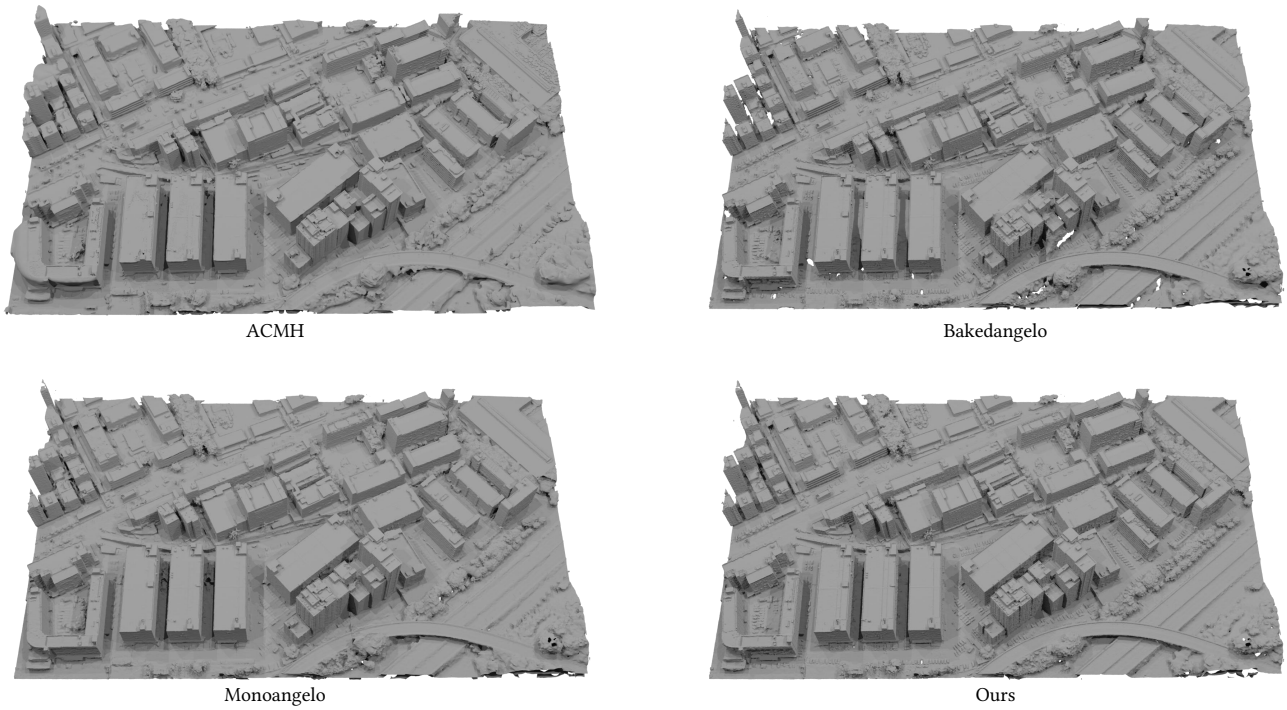

Fig. 2. The reconstruction results of the entire Residence.

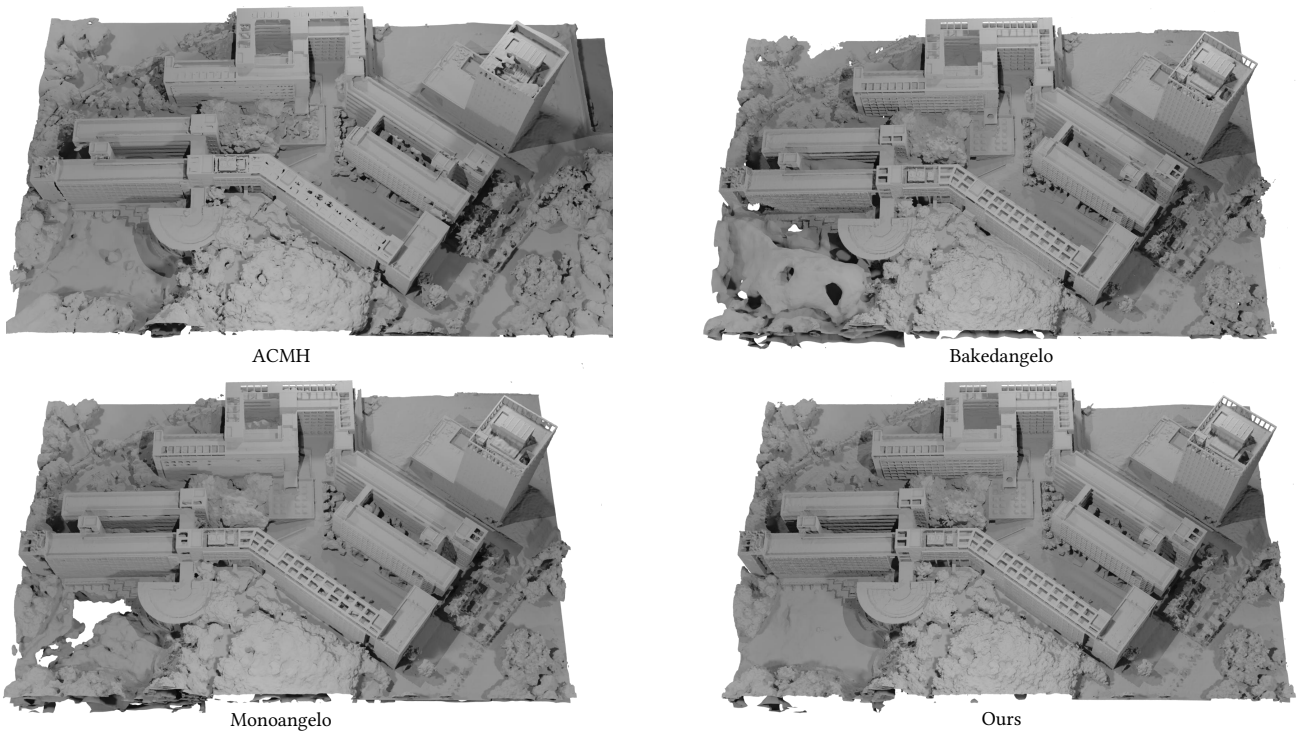

Fig. 3. The reconstruction results of the entire SciArt.

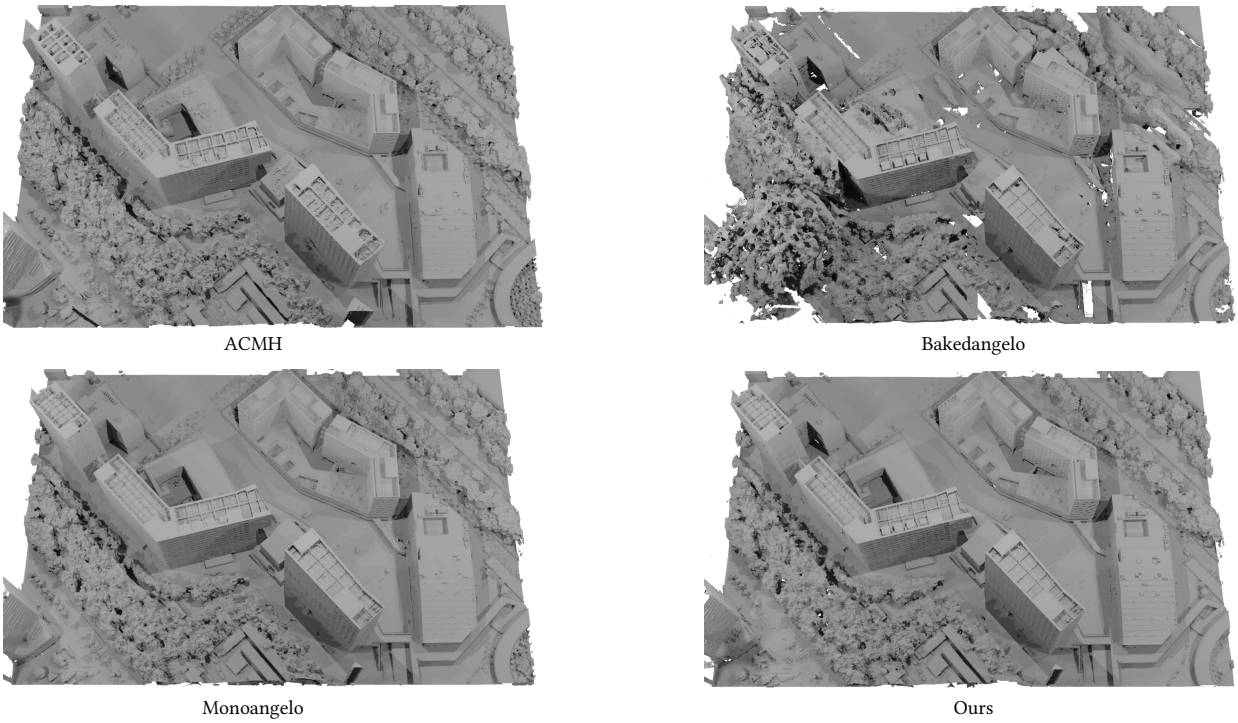

Fig. 4. The reconstruction results of the entire Polytech.

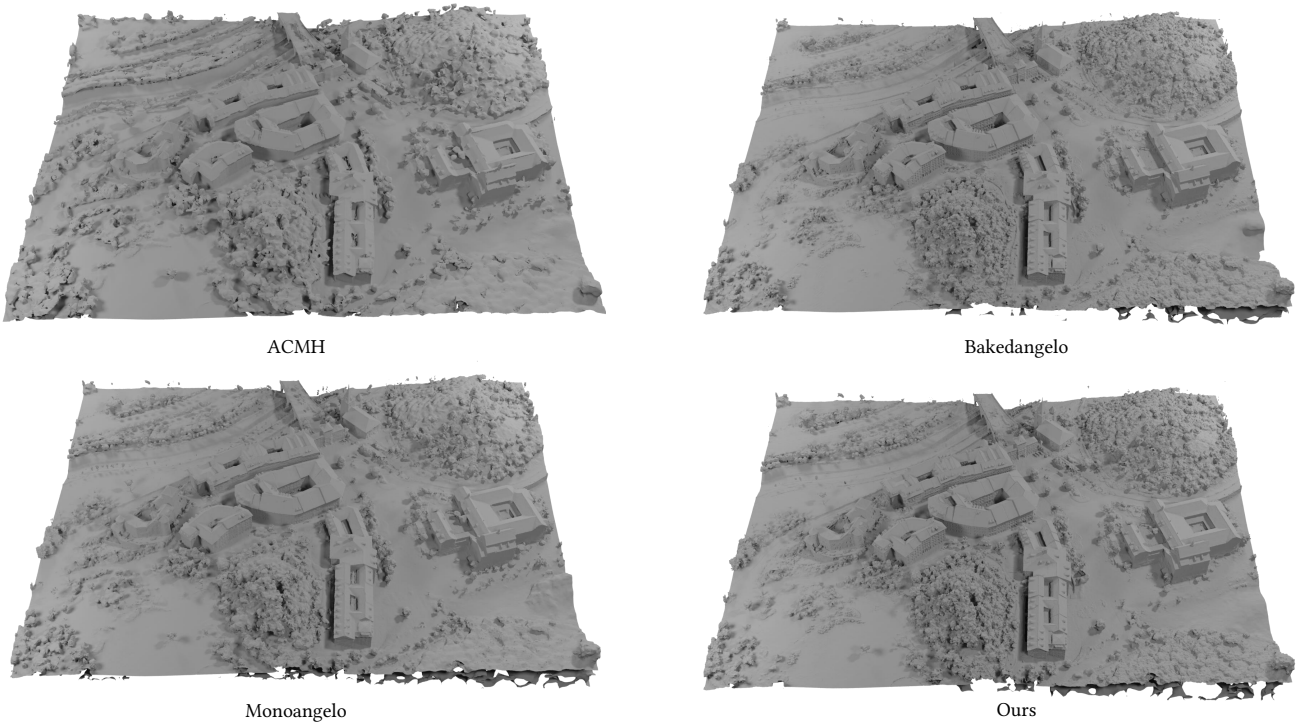

Fig. 5. The reconstruction results of the entire Songshanhu.

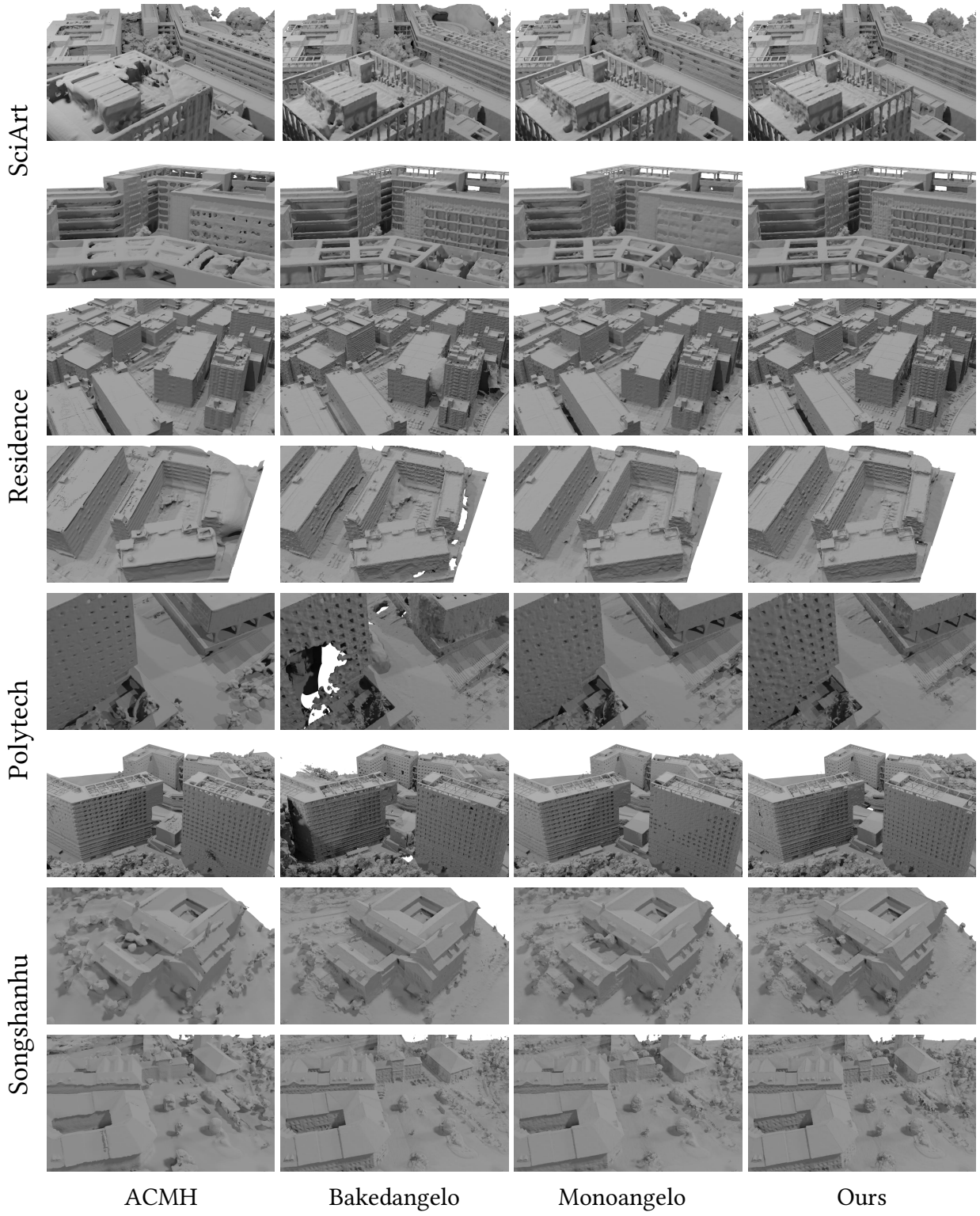

Fig. 6. More qualitative results.

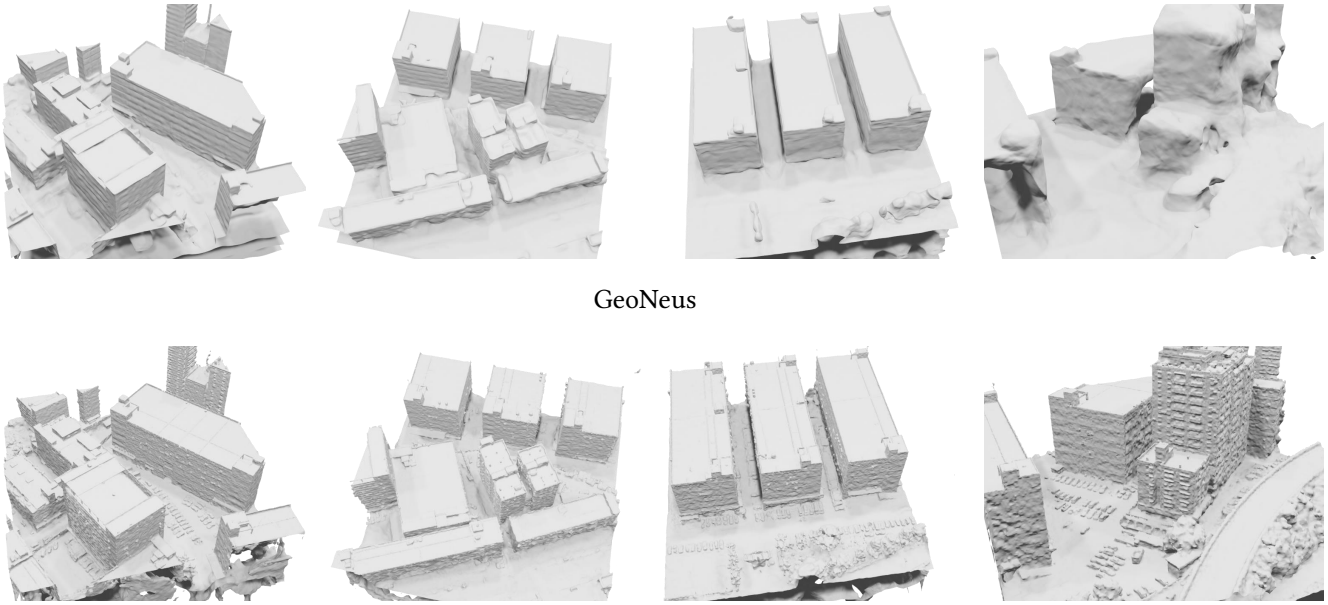

Fig. 7. Reconstruction results of GeoNeuS.

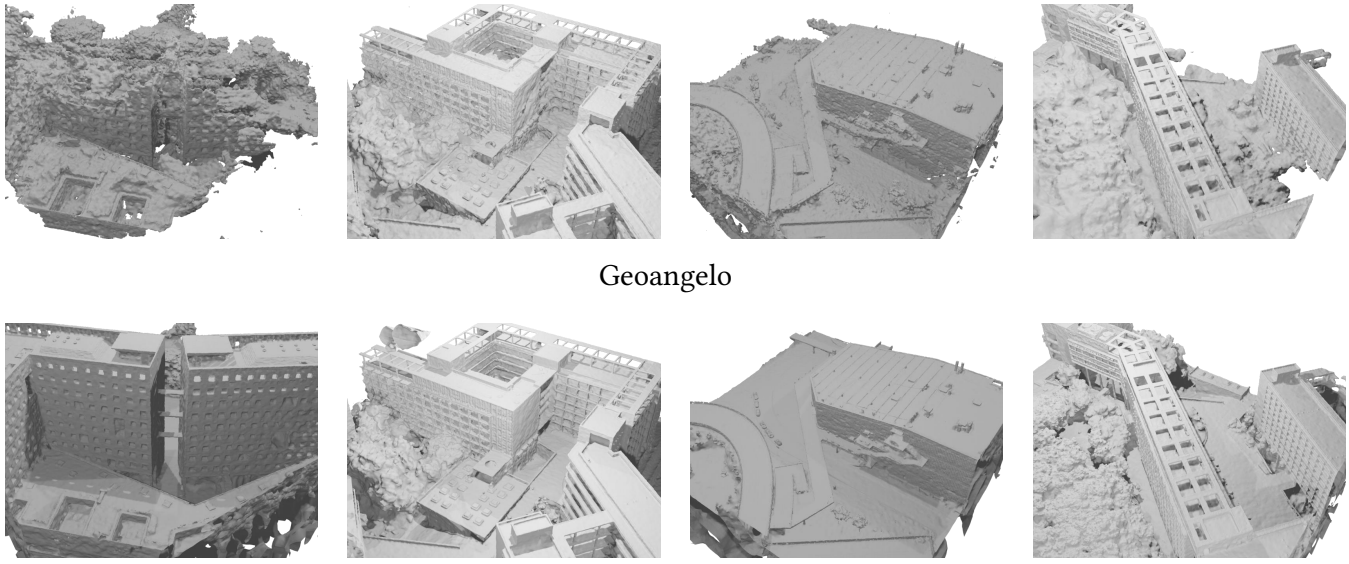

Fig. 8. Reconstruction results of Geoangelo.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009