Supplementary Material for Learning to Reconstruct 3D Non-Cuboid Room Layout from a Single RGB Image

Cheng Yang[1,2*†] Jia Zheng[3*] Xili Dai[1,4] Rui Tang[3] Yi Ma[4] Xiaojun Yuan[1‡]
[1]University of Electronic Science and Technology of China [2]Hikvision Research Institute
[3]Manycore Tech Inc. (Kujiale) [4]University of California, Berkeley

https://github.com/CYang0515/NonCuboidRoom

In this supplementary material, we provide the details of the potential intersection line region and additional reconstruction results.

Potential intersection line region. To classify the 3D geometric relationship of two adjacent walls, we define a potential intersection line region between these two adjacent walls. If these two walls are physically connected in 3D space, their projected intersection line calculated by their 3D parameters should lie in their potential intersection line region. If not, these two walls are physically disconnected in 3D space and the occlusion occurs. We select at most one detected line in each potential intersection line region as the corresponding detected intersection line or occlusion line. We use the detected line to adjust the 3D parameters of walls or handle the occlusion, depending on their geometric relationship in 3D space.

The potential intersection line region is defined by the bounding boxes of walls and is a wall size-adaptive range. Let $p_i = (x_i, y_i, w_i, h_i)$ and $p_j = (x_j, y_j, w_j, h_j)$ be the 2D bounding boxes of two adjacent walls, where plane p_i appears to the left of the plane p_j, *i.e.*, $x_i < x_j$. We first get the right-most boundary x_i^{right} of the plane p_i and left-most boundary x_j^{left} of the p_j:

$$x_i^{\text{right}} = x_i + w_i/2, \tag{1}$$

$$x_j^{\text{left}} = x_j - w_j/2. \tag{2}$$

The potential intersection line region $R_{i,j} = (x_R^{\text{left}}, x_R^{\text{right}})$ is defined as the overlapping region or gap between the two walls:

$$x_R^{\text{left}} = \min(x_i^{\text{right}}, x_j^{\text{left}}) \tag{3}$$

$$x_R^{\text{right}} = \max(x_i^{\text{right}}, x_j^{\text{left}}). \tag{4}$$

To tolerate the errors of plane and line detections, we then enlarge the region $R_{i,j}$ according to the sizes of the bounding boxes as follows:

$$x_R^{\text{left}} = \begin{cases} \max(x_R^{\text{left}} - \min(\epsilon \cdot w_i, \delta), x_i), & x_j^{\text{left}} > x_i^{\text{right}}, \\ \max(x_R^{\text{left}} - \min(\epsilon \cdot w_j, \delta), x_i), & \text{otherwise}, \end{cases} \tag{5}$$

$$x_R^{\text{right}} = \begin{cases} \min(x_R^{\text{right}} + \min(\epsilon \cdot w_j, \delta), x_j), & x_j^{\text{left}} > x_i^{\text{right}}, \\ \min(x_R^{\text{right}} + \min(\epsilon \cdot w_i, \delta), x_j), & \text{otherwise}, \end{cases} \tag{6}$$

where ϵ and δ are hyper-parameters to control the range of the potential intersection line region. In our implementation, we set $\epsilon = 0.25$ and $\delta = 10$ with respect to the image width of 640.

More reconstruction results. Figure 1 shows more 3D layout reconstruction results on Structured3D dataset [1].

References

[1] Jia Zheng, Junfei Zhang, Jing Li, Rui Tang, Shenghua Gao, and Zihan Zhou. Structured3d: A large photo-realistic dataset for structured 3d modeling. In *ECCV*, pages 519–535, 2020.

*Equal contributions.
†This work was done when Yang Cheng was a student at UESTC.
‡Corresponding author.