

Geometry-as-context: Modulating Explicit 3D in Scene-consistent Video Generation to Geometry Context

Supplementary Material

7. More results

7.1. Indoor scenes

Please see Figure 5.

7.2. Outdoor scenes

Please see Figure 6.

7.3. Interleaved depth estimation

Please see Figure 7 and Figure 8.

7.4. Failure cases

We present two failure cases of GaC in Figure 9. As can be seen, GaC has not yet generalized well to humans and complex subjects. In addition, for some outdoor scenes, novel view synthesis may produce slightly darkened textures around the boundaries.

8. Improved depth estimation strategy

Specifically, we make two modifications to the deep prediction pipeline of Hunyuan-Voyager:

- We use MoGe2 for metric-scale depth prediction and incorporate the VGGT depth confidence map to refine the depth map.
- We extend most of the operators in the pipeline from numpy to pytorch to support parallelized depth prediction on the GPU.

The relevant code will be open-sourced upon publication.



Reference



View-16



View-32



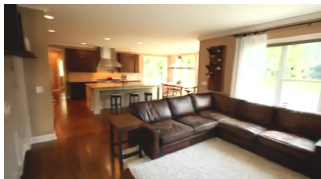
View-48



View-64



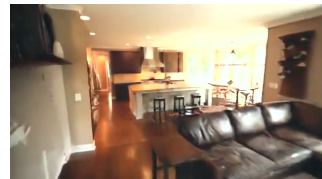
Reference



View-16



View-32



View-48



View-64



Reference



View-16



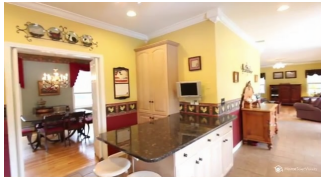
View-32



View-48



View-64



Reference



View-16



View-32



View-48



View-64

Figure 5. Gac's results on indoor scenes.



Reference



View-6



View-12



View-18



View-24



Reference



View-6



View-12



View-18



View-24



Reference



View-8



View-16



View-24



View-32

Figure 6. Gac's results on outdoor scenes.



Reference



View-3



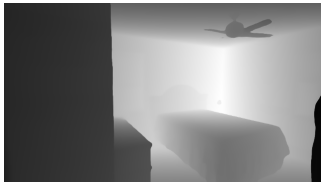
View-6



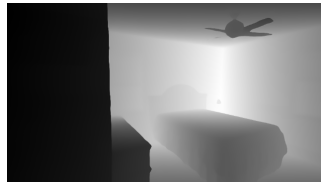
View-9



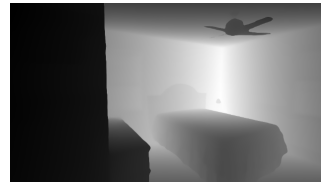
View-12



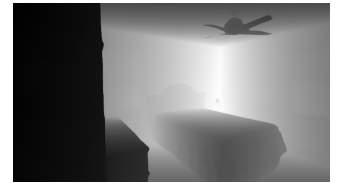
Depth-3



Depth-6



Depth-9



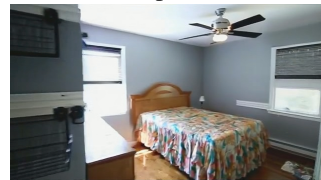
Depth-12



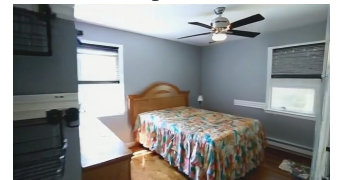
View-15



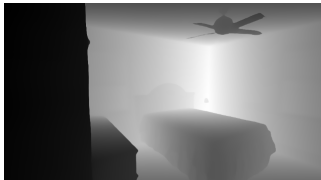
View-18



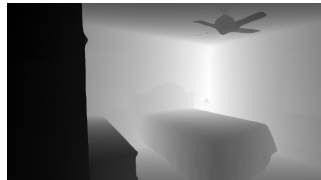
View-21



View-24



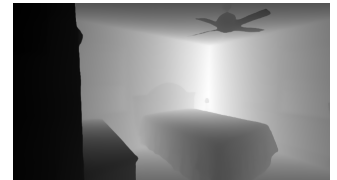
Depth-15



Depth-18



Depth-21



Depth-24

Figure 7. RGB-Depth interleaved output format in indoor scenes.



Reference



View-2



View-4



View-6



View-8



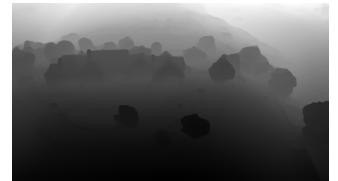
Depth-2



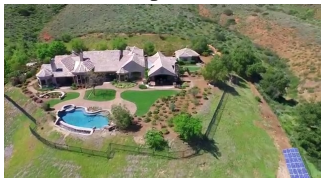
Depth-4



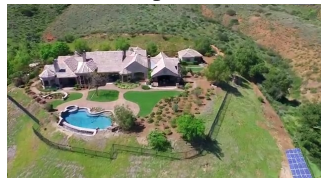
Depth-6



Depth-8



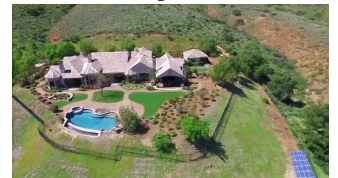
View-10



View-12



View-14



View-16



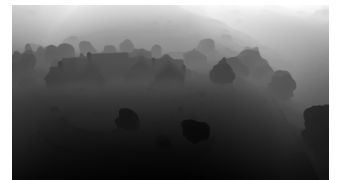
Depth-10



Depth-12



Depth-14



Depth-16

Figure 8. RGB-Depth interleaved output format in outdoor scenes.



Reference



View-4



View-8



View-12



View-16



Reference



View-8



View-16



View-24



View-32

Figure 9. Failure cases.