Seeing A 3D World in A Grain of Sand

Supplementary Material

We provide the following in this supplementary document: 1) detailed geometric derivations of angles and conditions used in Section 3.2; 2) more details and intermediate results of the pre-processing steps shown in Section 4.1; 3) comparison between different mirror configurations; and 4) additional experimental results on synthetic and real data.

A. Geometric Derivations

A.1. Derivation of Angles in Section 3.2

Here we show how to derive the angles we used in Section 3.2, when formulating the effective viewing volume. Our goal is derive the half apex angle of the viewing volume (θ), given the tilting angles of the two mirrors (α_1 and α_2). For ease of reference, we introduce auxiliary angles labeled in numbers. All the angles that we have referred to are annotated in Fig. 1.

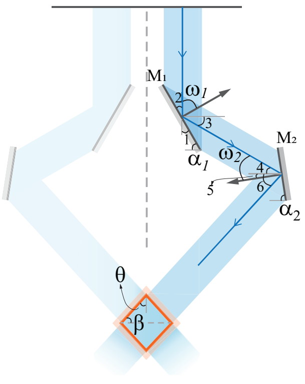

Figure 1. Angle annotations.

Since $\angle 1$ and $\angle 2$ are vertical angles, we have $\angle 2 = \angle 1 = 90^\circ - \alpha_1$. Since $\angle 2$ and ω_1 are complementary, we can calculate the incident/exit angle of reflection on M_1 as:

$$\omega_1 = 90^\circ - \angle 2 = \alpha_1. \quad (1)$$

Since $\angle 3$ and $\angle 4$ are alternate angles, we have $\angle 4 = \angle 3 = 2\omega_1 - 90^\circ$. By substituting ω_1 with Eq. 1, we have $\angle 4 = 2\alpha_1 - 90^\circ$. Since $\angle 5$ and α_2 are complementary, we have $\angle 5 = 90^\circ - \alpha_2$. Therefore, the incident/exit angle of reflection on M_2 can be calculated as:

$$\omega_2 = \angle 4 + \angle 5 = 2\alpha_1 - \alpha_2. \quad (2)$$

Since $\angle 6$ and β are congruent, we have $\beta = \angle 6 = \angle 5 + \omega_2$. By substituting ω_2 and $\angle 5$, we have $\beta = 90^\circ - 2\Delta\alpha$, where $\Delta\alpha = \alpha_2 - \alpha_1$. The half apex angle of the effective viewing volume, being complenentary to β, is thus:

$$\theta = 90^\circ - \beta = 2\Delta\alpha. \quad (3)$$

A.2. Derivation of Conditions in Section 3.2

Derivation of condition (i). This condition is introduced to allow light to travel through the lens from one end to the other, after being reflected by the two mirrors in a pair. In addition, the multi-view images formed by the eight mirror pairs should have overlaps, in order to be practical for scene reconstruction.

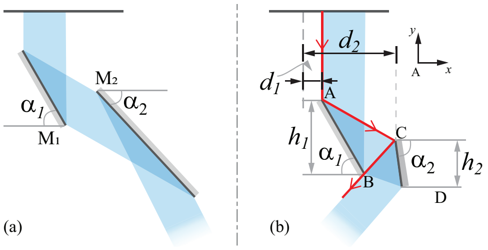

Figure 2. (a) The diverging situation when $\alpha_2 < \alpha_1$; (b) The extreme situation without inter-reflection, in which the reflected ray from M_2 intersects with the bottom edge of M_1.

With $\alpha_1 > 45^\circ$ and $\alpha_2 < 90^\circ$, we guarantee that light from the scene could travel through our mirror lens and reach the camera on the other end (*i.e.*, light path wouldn't turn around inside of the lens). If $\alpha_2 < \alpha_1$, the light exiting the lens (after reflected by M_2) would be diverging (see Fig. 2 (a)), resulting none-overlapping multi-view images. So we have $45^\circ < \alpha_1 < \alpha_2 < 90^\circ$.

Derivation of condition (ii). Here we derive the minimum vertically projected height of M_2 (denoted as h_2), such that it can cover the entire light beam reflected from M_1.

The width of parallel light beam reflected from M_1 is $w = h_1 / \tan\alpha_1$, where h_1 is the vertically projected height of M_1. In order to cover the entire beam, the length of M_2 (denoted as l_2) should satisfy:

$$l_2 \geq \frac{w}{\sin(\alpha_2 - \angle 4)} = \frac{w}{\cos(2\alpha_1 - \alpha_2)}. \quad (4)$$

Substituting $l_2 = h_2 / \sin\alpha_2$ and $w = h_1 / \tan\alpha_1$, we can rewrite Eq. 4 as:

$$h_2 \geq \frac{\sin\alpha_2}{\tan\alpha_1 \cdot \cos(2\alpha_1 - \alpha_2)} \cdot h_1. \quad (5)$$

Derviation of condition (iii). Here we derive the minimum separation between the two mirrors in order to avoid interreflection. We quantify this distance as $d_2 - d_1$ (where d_1 and d_2 are the distances from M_1 and M_2's upper edges to the central ray), when given their vertically projected heights h_1, h_2 and tilting angles α_1, α_2. We consider the extreme situation when the leftmost ray of the light beam intersects with the bottom edge of M_1 after reflecting from M_2 (see Fig. 2 (b)).

We denote the end points of M_1 and M_2 in the 2D cross-section plot as A, B, C, and D. We setup a coordinate system with A as the origin as shown in Figure 2(b). The line equation for M_1 (line AB) can be written as:

$$y = -\tan\alpha_1 \cdot x. \quad (6)$$

The line equation for the leftmost ray incident to M_2 (line AC) can be written as:

$$y = \cot 2\alpha_1 \cdot x. \quad (7)$$

Since $x_C = d_2 - d_1$, we plug it into Eq. 7 and calculate the coordinate of C as $(d_2 - d_1,\ \cot 2\alpha_1 \cdot (d_2 - d_1))$. The line equation for leftmost ray reflected from M_2 (line BC) can thus be calculated as:

$$y = \cot\Delta\alpha \cdot (x - (d_2 - d_1)) + \cot 2\alpha_1 \cdot (d_2 - d_1), \quad (8)$$

where $\Delta\alpha = \alpha_2 - \alpha_1$. By combining Eq. 6 and Eq. 8, we can calculate the x coordinate of B as:

$$x_B = \frac{\cot 2\Delta\alpha - \cot 2\alpha_1}{\tan\alpha_1 + \cot 2\Delta\alpha} \cdot (d_2 - d_1). \quad (9)$$

To avoid inter-reflection, x_B should be satisfy: $x_B \geq h_1/\tan\alpha_1$. Subsituting x_B with Eq. 9, we obtain the third condition regarding the mirror distances:

$$d_2 \geq \frac{\tan\alpha_1 + \cot 2\Delta\alpha}{\tan\alpha_1 \cdot (\cot 2\Delta\alpha - \cot 2\alpha_1)} \cdot h_1 + d_1. \quad (10)$$

B. More Details on Pre-processing Steps

Fig. 3 shows how our captured raw image is processed into multi-view input to 3DGS. A raw image captured by our portable lens prototype is shown in Fig 3 (a). Its resolution is 2448 × 2048. We first apply a multi-view mask to extract the effective regions formed through mirror reflection. The filtered image is shown in Fig. 3(b). Then, for each sub-image, we re-project it to allow smooth view transition (we update camera poses after re-projection). We also mask out the background and only reconstruct the foreground objects. The processed image for one sub-view (highlighted in red) is shown in Fig. 3 (c). This image is with resolution 800 × 800. The eight sub-view images processed in this way are used as input to 3DGS.

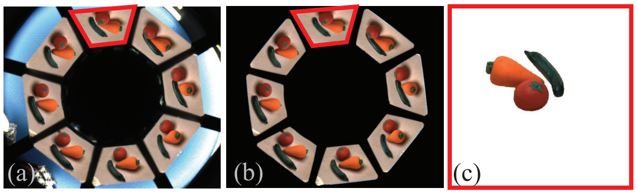

Figure 3. (a) Our captured raw image; (b) Image filtered by the multi-view mask; (c) Re-projected image of the highlighted view.

C. Lens Design Comparison

Here we show comparison between two lens designs with different mirror configuration. Prototypes of the two designs are shown in Fig. 4. The two lenses have the same base lengths for the inner and outer pyramids, with different tilting angles for the mirrors. The parameters we use are shown in Table 1. Images taken with the two lenses are shown in Fig. 4.

We can see that design (b), which has larger $\Delta\alpha$, has better coverage of the side views (*e.g.*, the figurine's face becomes visible in (b)). This is equivalent to having virtual cameras with more oblique angles. Such configuration is preferred since it provides fuller coverage of the scene. This observation is consistent with our guidelines on optimizing the mirror configuration.

Table 1. Mirror parameters of the two different designs.

	α_1	α_2	$\Delta\alpha$
Design (a)	75°	85°	10°
Design (b)	60°	85°	25°

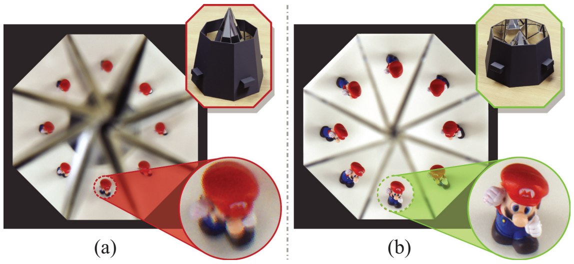

Figure 4. Comparison between two lens designs. Here we show the lens prototypes and their captured images with zoom-in views.

D. Additional Experimental Results

D.1. Ablation on Depth Loss

Fig. 5 compares depth maps obtained by different methods for a real scene (*i.e.*, the "frog" scene). Specifically, the MiDaS [3] depth is used by FSGS [6]; Depth Anything V2 [5] is used by Hierarchical 3DGS [1]; and the visual hull depth is used by our approach. We can see that Depth Anything

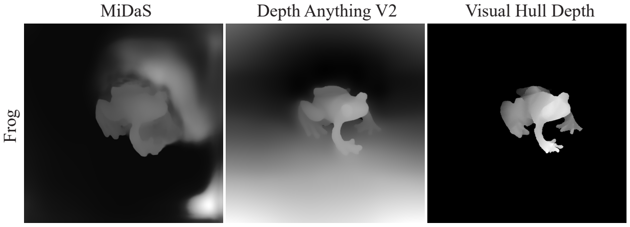

Figure 5. Comparison of depth map obtained by different methods.

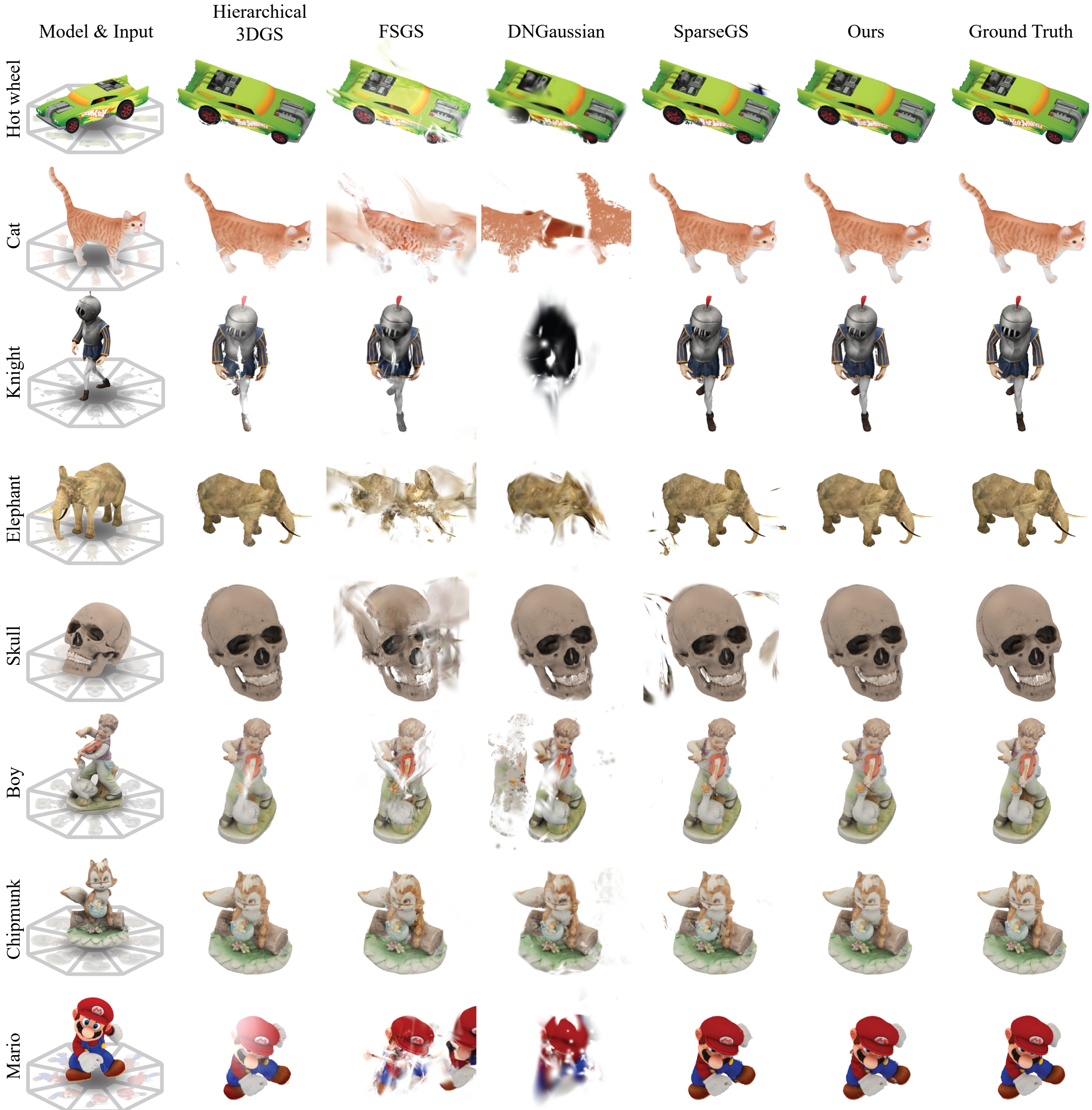

Figure 6. Additional visual comparison results on synthetic data.

provides much better depth prior than MiDaS depth. Our visual hull depth outperforms Depth Anything result in details (*e.g.*, the frog legs have more discernible depth variation in the visual hull depth). Moreover, the visual hull projection provides depth values in absolute scale, whereas the other two learning-based methods estimate relative depths.

We performed an ablation study on depth loss using the "skull" scene (see Fig. 6). We compare the PSNR of synthesized novel views for three variants of our algorithm: without depth loss, with monocular depth (Depth Anything V2 [5] depth), and with visual hull depth (VH depth). The table below shows the ablation study on depth loss.

Table 2. Ablation study on depth loss.

Variant	w/o depth	w. monodepth	w. VH depth
PSNR	29.2052	29.3698	29.3856

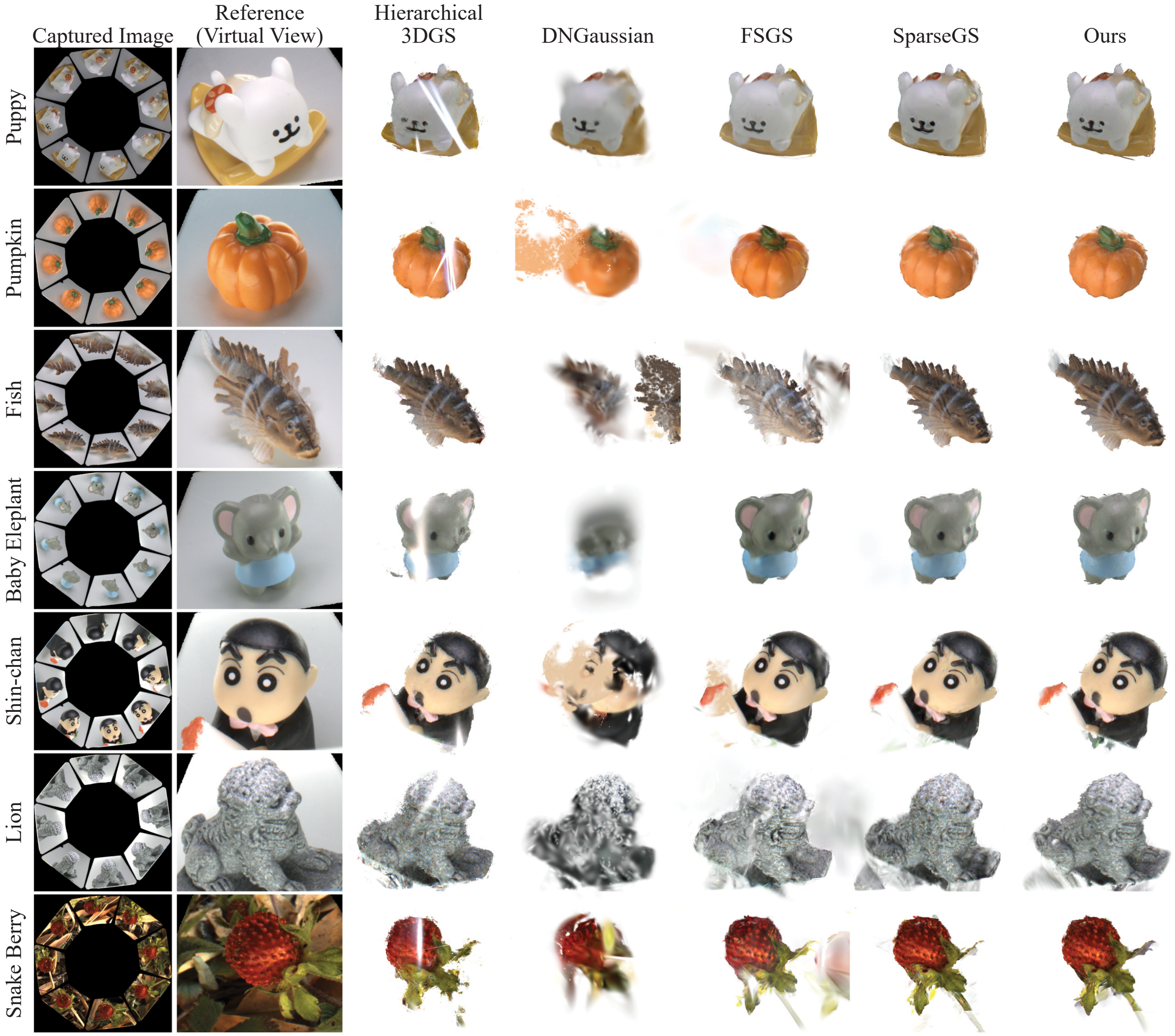

Figure 7. Additional visual comparison results on real data.

D.2. Additional Synthetic Results

We show more visual results on synthetic data in Fig. 6. We compare with recent state-of-the-art 3DGS algorithms: Hierarchical 3DGS [1], FSGS [6], DNGaussian [2], and SparseGS [4]. Most of these methods are optimized for spare view input. We can see that our results outperform the state-of-the-arts and resemble the ground truths.

D.3. Additional Real Results

Fig. 7 shows more visual comparison results on real data in comparison with state-of-the-arts. The “snake berry” scene is captured outdoor with our portable lens.

References

[1] Bernhard Kerbl, Andreas Meuleman, Georgios Kopanas, Michael Wimmer, Alexandre Lanvin, and George Drettakis. A hierarchical 3d gaussian representation for real-time rendering of very large datasets. *ACM TOG*, 43(4), 2024. 2, 4

[2] Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, and Lin Gu. Dngaussian: Optimizing sparse-view 3d gaussian radiance fields with global-local depth normalization. In *CVPR*, 2024. 4

[3] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE TPAMI*, 44(3), 2022. 2

[4] Haolin Xiong, Sairisheek Muttukuru, Rishi Upadhyay,

Pradyumna Chari, and Achuta Kadambi. SparseGS: Real-time 360° sparse view synthesis using gaussian splatting. *Arxiv*, 2023. 4

[5] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *arXiv:2406.09414*, 2024. 2, 3

[6] Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. FSGS: Real-time few-shot view synthesis using gaussian splatting. In *ECCV*, 2024. 2, 4