Supplementary Materials of SceneShine: Illumination-aware Human Scene Gaussian Re-Splatting from Mobile Device Video

Xuqian Ren[1] Wenjia Wang[2] Mai Ngoc Nguyen[1]
Juho Kannala[3,4] Esa Rahtu[1]
[1] Tampere University, [2] The University of Hong Kong, [3] Aalto University, [4] University of Oulu

1. Implementation Details

2. SceneShinePeople Dataset

To illustrate our methods, we construct two synthetic sequences using Blender. We use the textured, animated people from meshcapade [6], and room assets from turbosquid [7]. We process the sequences following HUGS [4, 9]. The datasets include two subsets: ”walk” with 86 frames and ”stroll” with 171 frames. We divide the datasets into training, validation, and testing sets for both main and relighted sequences. Training is conducted on the training set of the main sequences, while testing is performed on the test set of the relighted sequences. The trajectory is visualized using the tools provided by Neuman [3], as shown in Fig. 1.

3. Method Design

During training, we regulate normal, depth, and PBR properties. We predict normal $\boldsymbol{n_i}$ as learnable parameters for each Gaussian in the scene and splat $\boldsymbol{n_i}$ to generate the normal image $\mathcal{N}$. We then regularize $\mathcal{N}$ with a pseudo normal $\tilde{\mathcal{N}}$, estimated from omnidata [1], which has been shown to produce more robust geometry rendering [8]:

$$\mathcal{L}_n = \|\mathcal{N} - \tilde{\mathcal{N}}\|_2 \tag{1}$$

We further calculate monocular depth with Depthcrafter [2], which is designed for monocular depth estimation from video. We supervise the rendered depth with Pearson correlation loss proposed by FSGS [10]

$$\mathcal{L}_d = \frac{\text{Cov}\left(\mathcal{D}, \mathcal{D}_{mono}\right)}{\sqrt{\text{Var}\left(\mathcal{D}\right)\text{Var}\left(\mathcal{D}_{mono}\right)}} \tag{2}$$

The smoothness constraints on roughness and albedo $\mathcal{L}_{sr,sa}$, was designed as:

$$\mathcal{L}_{sr,sa} = \|\nabla I_{r,a}\|\text{exp}(-\|\nabla I_{gt}\|) \tag{3}$$

Light loss $\mathcal{L}_{\text{light}}$ was applied as following:

$$\mathcal{L}_{\text{light}} = \sum_c \left(L_c - \frac{1}{3}\sum_c L_c\right), c \in \{R, G, B\} \tag{4}$$

4. Experiments

4.1. More comparisons on Real-world Dataset

We present additional qualitative comparisons of our method against HUGS and IC-Light on composite videos, as shown in Fig. 2. Further video demonstrations are included in the supplementary video.

4.2. More comparisons on Single Human Relighting task

We provide more qualitative demos on the single human relighting task in Fig. 3. Our method provides a more natural lighting with a clear face appearance.

References

[1] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *ICCV*, pages 10786–10796, 2021. 1

[2] Wenbo Hu, Xiangjun Gao, Xiaoyu Li, Sijie Zhao, Xiaodong Cun, Yong Zhang, Long Quan, and Ying Shan. Depthcrafter: Generating consistent long depth sequences for open-world videos. *arXiv preprint arXiv:2409.02095*, 2024. 1

[3] Wei Jiang, Kwang Moo Yi, Golnoosh Samei, Oncel Tuzel, and Anurag Ranjan. Neuman: Neural human radiance field from a single video. In *ECCV*, pages 402–418. Springer, 2022. 1

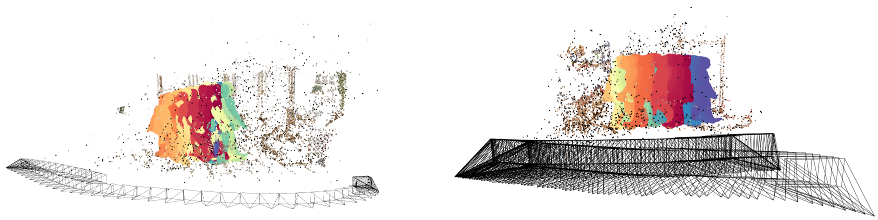
[4] Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, and Anurag Ranjan. Hugs: Human gaussian splats. In *CVPR*, pages 505–515, 2024. 1, 3

[5] Wenbin Lin, Chengwei Zheng, Jun-Hai Yong, and Feng Xu. Relightable and animatable neural avatars from videos. In *AAAI*, pages 3486–3494, 2024. 3

[6] Meshcapade Inc. Meshcapade: The digital human company. `https://meshcapade.com`, 2024. Accessed: 2024-11-17. 1

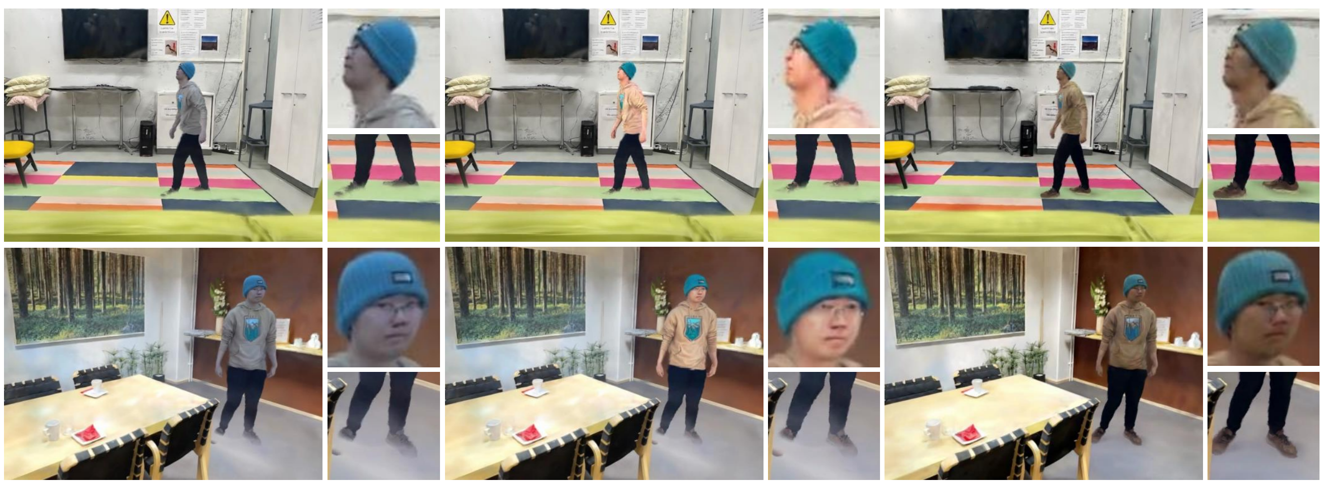
[7] TurboSquid, Inc. Turbosquid, 2024. Accessed: 2024-11-17. 1

[8] Matias Turkulainen, Xuqian Ren, Iaroslav Melekhov, Otto Seiskari, Esa Rahtu, and Juho Kannala. Dn-splatter: Depth and normal priors for gaussian splatting and meshing, 2024. 1



(a) Walk (b) Stroll

Figure 1. Visualization of the two sequences in SceneShinePeople.



HUGS **IC-Light** **Ours**

Figure 2. We show more visualization comparisons between our method with HUGS and IC-Light from novel views, where the human subject is animated with new poses. Our method demonstrates more realistic relighting while maintaining identity consistency.

[9] ZCWzy. preprocessforhugs. https://github.com/ZCWzy/preprocessForHugs, 2024. GitHub repository, Accessed: 2024-11-17. 1

[10] Zehao Zhu, Zhiwen Fan, Yifan Jiang, and Zhangyang Wang. Fsgs: Real-time few-shot view synthesis using gaussian splatting. In *ECCV*, pages 145–163. Springer, 2025. 1

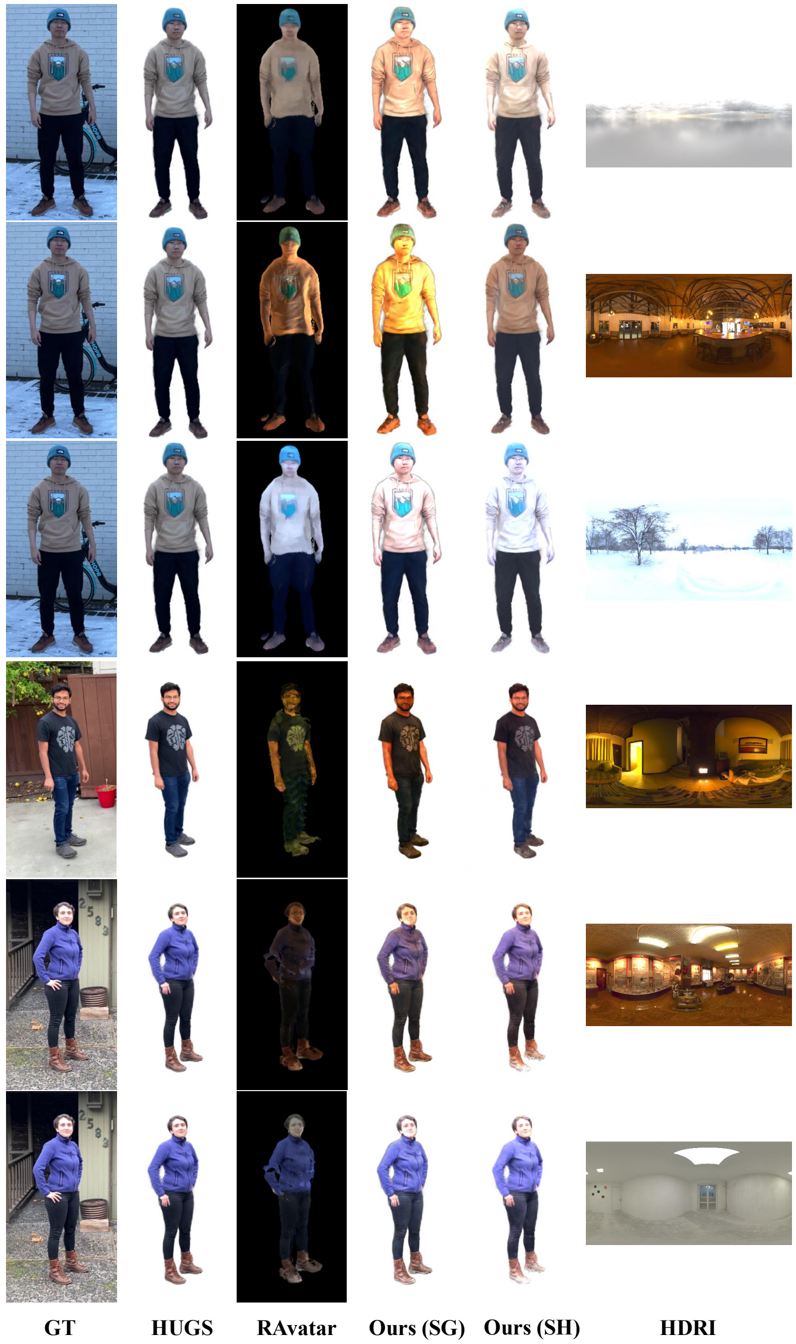


Figure 3. Single-human relighting results under the given HDRI map. Our method produces sharper textures while ensuring a more natural lighting effect compared to HUGS [4] and RAvatar [5]. Please zoom in to see the details.