This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

HumanOLAT: A Large-Scale Dataset for Full-Body Human Relighting and Novel-View Synthesis

Timo Teufel[1,*] Pulkit Gera[1,*] Xilong Zhou[1] Umar Iqbal[2]
Pramod Rao[1] Jan Kautz[2] Vladislav Golyanik[1] Christian Theobalt[1]
[1]Max Planck Institute for Informatics, SIC [2]NVIDIA

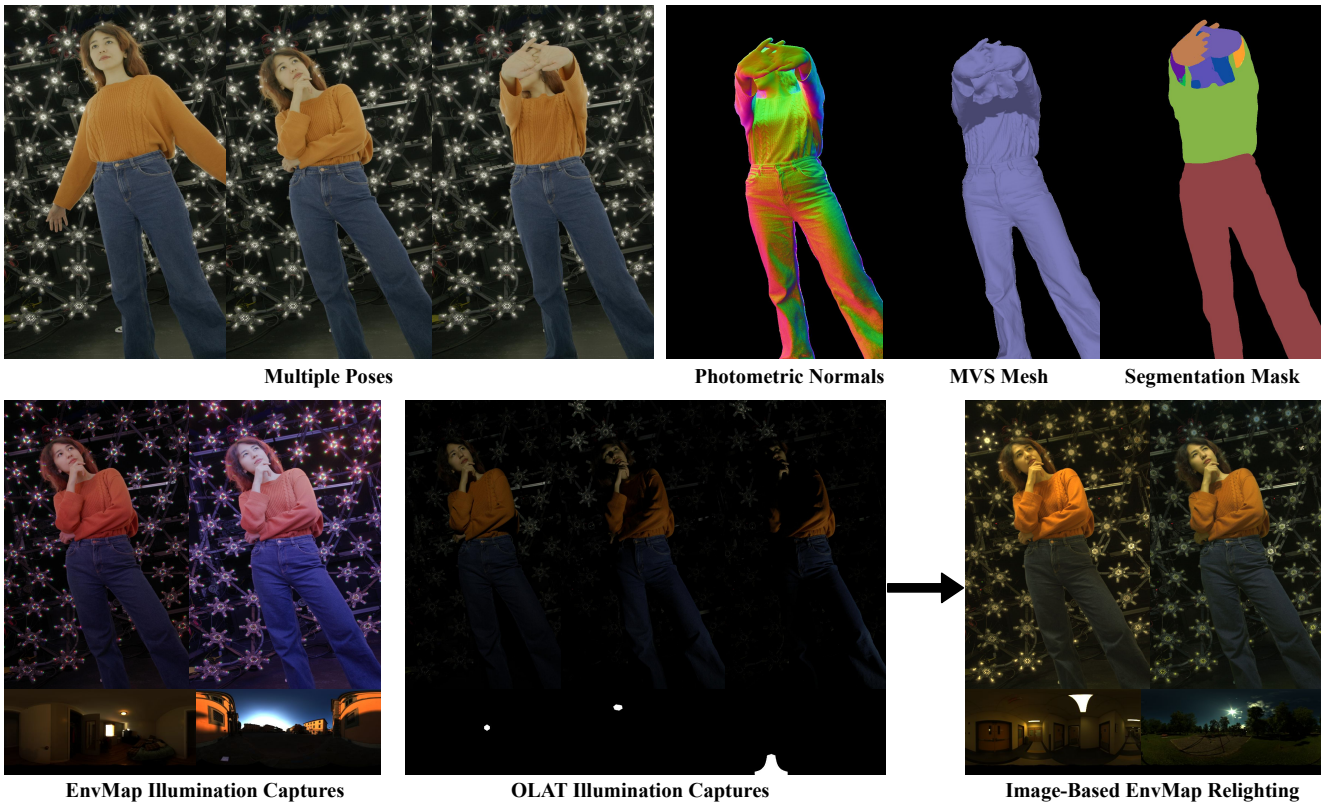


Figure 1. **Overview of the proposed *HumanOLAT* dataset.** We record 21 subjects in three static poses each (top left) and provide pixel-wise photometric normals, multi-view stereo (MVS) meshes and segmentation masks (top right). To enable broad-scale application in various relighting tasks (e.g. illumination harmonization), we capture every pose both under ten environment map illuminations loaded directly into our lightstage (bottom left) as well as 331 OLAT illuminations suitable for image-based relighting [11] (bottom right).

Abstract

Simultaneous relighting and novel-view rendering of digital human representations is an important yet challenging task with numerous applications. Progress in this area has been significantly limited due to the lack of publicly available, high-quality datasets, especially for full-body human captures. To address this critical gap, we introduce the HumanOLAT dataset, the first publicly accessible large-scale dataset of multi-view One-Light-at-a-Time (OLAT) captures of full-body humans. The dataset includes HDR RGB frames under various illuminations, such as white light, environment maps, color gradients and fine-grained OLAT illuminations. Our evaluations of state-of-the-art relighting and novel-view synthesis methods underscore both the dataset's value and the significant challenges still present in modeling complex human-centric appearance and lighting interactions. We believe HumanOLAT will significantly facilitate future research, enabling rigorous benchmarking and advancements in both general and human-specific relighting and rendering techniques.

* both authors contributed equally to this work

1. Introduction

Simultaneous relighting and novel-view rendering of digital human representations is a fundamental and long-standing challenge in computer vision and graphics, with applications spanning game and movie production, virtual telepresence, augmented reality (AR), virtual reality (VR), and mixed reality. Despite significant progress, the problem remains challenging due to inherent ambiguities in human appearance, arising from complex factors such as diverse clothing materials, intricate shadowing effects, and subsurface scattering in human skin.

Historically, the de facto standard for relighting research has been image-based relighting from One-Light-at-a-Time (OLAT) captures [11], which effectively disambiguates appearance components, enabling accurate novel illumination synthesis. However, prior research has primarily focused on upper-body and facial relighting due to the complexities and practical challenges associated with full-body captures. Hence, there are only a handful of works that address this challenge. A critical obstacle in this domain has been the lack of publicly accessible and comprehensive full-body OLAT datasets.

Several factors contribute to the scarcity of these datasets. Firstly, acquiring OLAT data requires specialized, costly and not widely accessible hardware setups known as lightstages, enabling precise control over illumination and multi-view camera setups. Secondly, capturing full-body OLAT data is considerably challenging due to the necessity for larger capturing space and extended capture durations, during which even minor involuntary movements of subjects result in noticeable artifacts. Consequently, a large number of existing methods rely on synthetic data [19], learned priors or regularizers [9, 36]. These methods, however, often exhibit diminished rendering quality with visible appearance artifacts.

To address the notable absence of a publicly available dataset for this task, this work introduces *HumanOLAT*. Our full-body human OLAT dataset consists of 21 diverse subjects, each captured in three distinct poses, providing a comprehensive resource for the community. The dataset includes precise camera and illumination calibrations, multi-view RGB frames under color-gradient illumination for photometric normal estimation, fine-grained OLAT captures, and images under predefined environment maps. *HumanOLAT* facilitates multiple evaluation scenarios, including static full-body novel-view synthesis, static full-body relighting, and joint full-body novel-view synthesis and lighting. An illustration of the captures and annotations from *HumanOLAT* is depicted in Fig. 1.

We perform extensive experiments using several state-of-the-art relighting and novel-view synthesis methods [4, 15, 58, 62] to demonstrate the utility and challenges posed by our dataset. Our evaluations highlight key remaining challenges in the field, pinpointing opportunities for further methodological improvements.

To summarise, the main contribution of this paper is *HumanOLAT*, a new OLAT multi-illumination dataset providing physically-based ground truth under any lighting condition for addressing different problems in the area of human relighting, novel view synthesis, and potentially other tasks in future. *HumanOLAT* is publicly available at https://vcai.mpi-inf.mpg.de/projects/HumanOLAT/.

2. Related Work

2.1. Lightstages

A light stage is an active illumination system allowing fine-grained control over lighting. First introduced by Debevec et al. [11] to acquire the reflectance field of human faces, these capture systems enable the collection of detailed multi-view image data with known illumination and as such, have become foundational in for developing and benchmarking realistic relighting methods [10, 17, 27, 30, 39, 42, 63]. Crucially, compared to other setups collecting real-world multi-illumination data [36], lightstage's ability to precisely control lighting allows for the capture of detailed reflectance data in the form of one-light-at-a-time (OLAT) images. As light transport is linear [11], these illuminations can be additively combined to enable physically correct image-based relighting under arbitrary target light.

While a variety of works provide lightstage OLAT data for static objects [35, 47], faces [42, 45, 57] and hands[10], to our knowledge, no publicly available OLAT data for full-body humans exists. With our proposed *HumanOLAT* dataset, we aim to close this gap by providing varied multi-illumination data, including OLATs, for a large set of diverse humans.

2.2. Human Relighting

General Object-centric Relighting Recent works exploring the relighting of static objects have largely focused on neural rendering approaches. Early works [5, 44, 60] extend Neural Radiance Fields (NeRF) [38] to enable relighting by decomposing the scene into intrinsic components such as surface normals, albedo, and roughness. While these methods achieve impressive results for simple objects, they struggle with complex materials and lighting effects commonly found in human subjects. Follow-up works [56] improve upon this by incorporating physically-based rendering principles, enabling more accurate modeling of specular reflections and global illumination effects.

Following the success of 3D Gaussian Splatting (3DGS) [25], numerous works have explored its application to relighting static, object-centric scenes [4, 13, 14, 16, 22, 29, 32, 33, 43, 51, 55, 58, 62, 64]. These approaches extend the original splatting technique by parametrizing each 3D

Gaussian primitive with both BRDF reflectance properties (albedo, roughness, specular) and standard Gaussian attributes (opacity, mean, scale) to enable lighting-aware rendering. The field has seen advances in both settings with unknown illumination [7, 14, 16, 22, 32, 43, 51, 64] and known lighting conditions [4, 13, 29, 55, 58, 62].

For these methods, the human body presents significant relighting challenges due to its intricate geometry, diverse material properties (including skin, hair, and clothing), and complex lighting effects, such as subsurface scattering and self-shadowing. Our proposed dataset, therefore, serves as a valuable benchmark for evaluating and stress-testing these approaches under realistic conditions.

Illumination Harmonization Given an arbitrary foreground input and a target lighting condition, such as a text prompt or a background image, illumination harmonization methods aim to generate photo-realistic depictions of the foreground under the new lighting. Recent state-of-the-art methods [18, 21, 23, 27, 39] utilize encoder-decoder architectures [21, 23, 27, 39] or diffusion models [18, 59] to directly learn image-based relighting from a set of ground truth data. The latest advancement, IC-Light by Zhang et al. [59], leverages multiple publicly available datasets [12, 35] and introduces an explicit light transport consistency loss to achieve plausible illumination harmonization.

Recent works [18, 27, 39, 46] have shown that OLAT data is particularly effective for training these methods, as it enables realistic relighting under arbitrary environment maps[11]. With *HumanOLAT*, we aim to provide such data for full-body captures and make it publicly accessible to the research community.

Drivable and Relightable Human Avatars The scarcity of publicly available full-body multi-illumination data has driven numerous studies [8, 19, 31, 34, 48, 53, 61] to focus on creating human avatars using single-illumination datasets like ActorsHQ [20] and MVHumanNet [52]. These methods typically employ a mix of inverse rendering to deduce underlying lighting conditions [8, 31, 53] or use synthetically generated multi-illumination datasets [19, 34, 48, 61] to learn the intrinsic properties essential for physically-based relighting. In contrast, Luvizon et al. [36] capture dynamic subjects under four distinct lighting conditions using light probes to obtain HDRI environment maps. Although this approach provides real-world relightable captures, it lacks fine-grained control over lighting conditions, resulting in a noticeable quality gap compared to methods utilizing light stage-based datasets [17, 49, 63]. Guo et al. [17] and Zhou et al. [63] utilize images captured with white light and color gradients to directly estimate intrinsic properties, such as albedo and geometry, thereby enabling high-quality relighting of both animated and static human models. Notably, the concurrent work by Wang et al. [49] introduces a method for creating drivable and relightable full-body avatars based on 3DGS [25], achieving exceptional quality by learning light transfer from dynamic full-body OLAT data. Similar studies focusing on relightable hand [10, 20] and head models [3, 30, 42, 54] using OLAT data also demonstrate similarly high-quality results.

With the *HumanOLAT* dataset, we aim to further accelerate the development of high-quality relightable human avatars by offering fine-grained OLAT data of static humans for training and benchmarking.

3. The *HumanOLAT* Dataset

In the following, we describe the contents and the capture setup of the proposed dataset in detail. Specifically, we describe our capture setup in Sec. 3.1 and provide an overview of the recorded data in Sec. 3.2, followed by an explanation of our data processing pipeline in Sec. 3.3. Finally, we compare *HumanOLAT* to existing publicly available lightstage datasets in Sec. 3.4.

3.1. Lightstage Capture Setup

Fig. 2 illustrates our capture setup, which features a spherical dome equipped with 40 RED Komodo 6K cameras and 331 individually controllable LEDs capable of emitting red, green, blue, amber, and white light (RGBAW). The cameras and lights are arranged 360° around the subject, enabling the capture of synchronized multi-illumination sequences at 30 FPS with the 5K image resolution.

3.2. Dataset Description

HumanOLAT is a comprehensive multi-view, multi-illumination dataset designed for full-body relighting. It includes 21 subjects, each captured in three distinct poses: an A-pose and two creative poses. For each recording, we capture the subject's appearance from 40 different views under the following diverse lighting conditions:

- **One white light illumination** utilized for camera calibration, mesh reconstruction, segmentation, and providing ground truth for albedo [63];
- **Two color gradient illuminations** [17, 63] employed to estimate pixel-wise photometric normals;
- **Ten environment map illuminations** [17, 63] used directly by methods that assume a known environment map [36, 62];
- **331 OLAT Illuminations** applied in methods that rely on single light images [4, 58, 62] and enable image-based relighting under arbitrary target lighting [11].

Samples of captured illuminations and subjects are shown in Fig. 1 and Fig. 3. For each of the 21 subjects, the dataset includes approximately: 3 Poses $\times$ 40 Views $\times$ 344 Illuminations $\approx$ 40K Frames resulting in a total of around 850K frames.

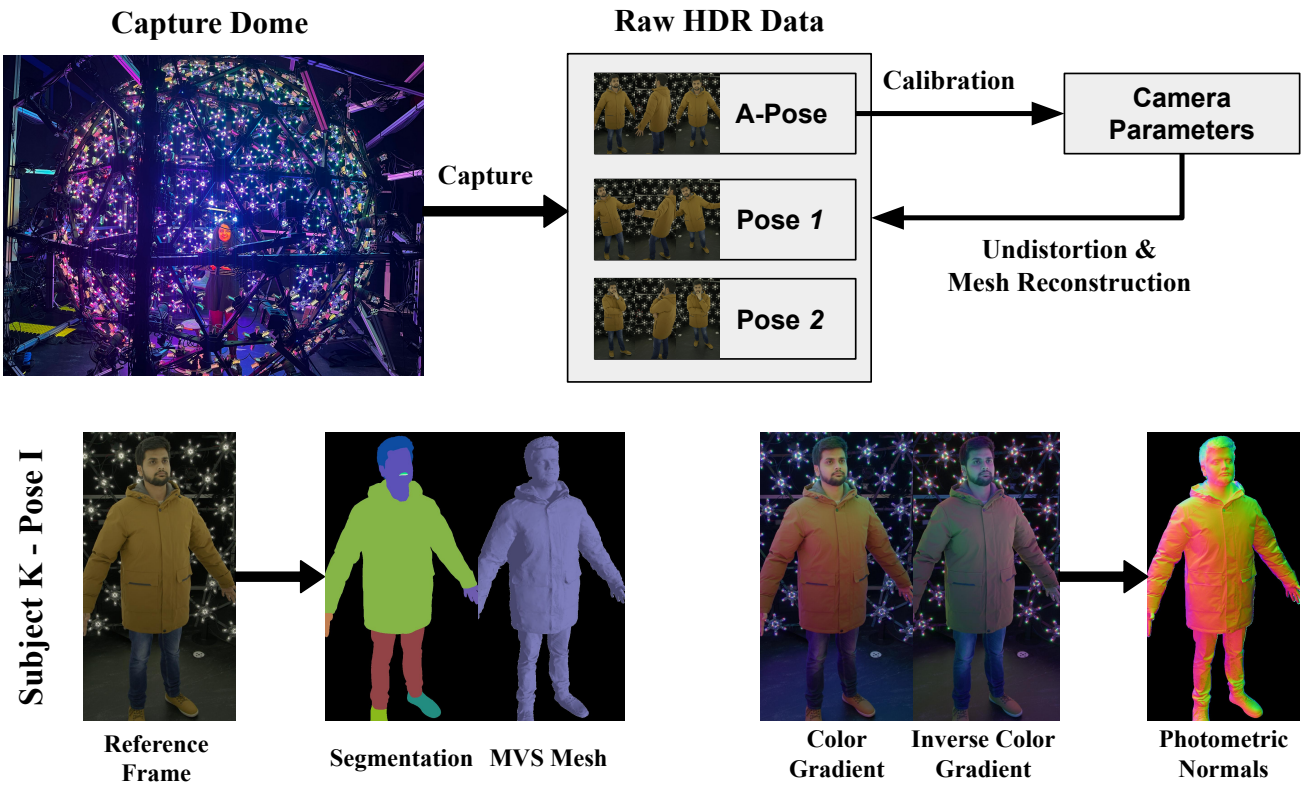


Figure 2. **Scheme of our data processing pipeline**. For each subject, we capture multi-view multi-illumination data under an A-pose as well as two creative poses. Afterwards, we use Metashape [2] to establish per-subject camera parameters from the A-pose. Taking a frame illuminated under white light as reference, we construct a multi-view stereo (MVS) mesh for each captured pose using the same software and employ Sapiens [26] for generating segmentation masks. Finally, we deduce pixel-wise photometric normals from color gradient illuminations [17, 63] recorded during the capture.

As shown in Fig. 3, the subjects include both male and female individuals wearing a variety of clothing, ranging from tight-fitting shirts and jumpers to loose, volumetric clothing such as hoodies and jackets. Specifically, among the recorded subjects, there are 13 males and 8 females. Of these, 5 subjects are wearing tight-fitting shirts and jumpers, while 16 are wearing loose clothing, which includes 7 jumpers, 5 hoodies, and 3 jackets. Please refer to Fig. 11 in the supplement for additional data samples.

We also provide OpenPose [6] annotations and SMPL-X [40] parameters. For more details, please refer to App. D.

3.3. Data Processing

3.3.1. Calibration and Mesh Reconstruction

For estimating camera parameters and suitable initial geometry, we rely on feature-based algorithms implemented in the proprietary software Metashape [2]. To ensure consistent camera intrinsics and extrinsics across different poses, we calculate a single camera calibration for each subject using the A-pose capture as a reference. Since the quality of feature-based calibration and reconstruction is directly dependent on the quantity and quality of features, we use the frame illuminated under flat white light to ensure optimal conditions for feature detection. Finally, we use a marker-based setup to transform estimated camera poses and geometry into a predefined real-world canonical coordinate system. The positions of individual lights are also provided within the same canonical system by measuring their physical location. However, we note that while most of the LEDs are static, a subset of around $\sim$20 lights are attached to a hatch used to enter the lightstage. Since the hatch moves, fully accurate 3D light positions cannot be guaranteed for this subset, and we recommend ignoring them during training and evaluation. Note, however, that methods that do not rely on accurate 3D light positions, such as ones performing relighting from environment maps, are unaffected.

To quantitatively validate our calibrations, we calculate the average re-projection error of triangulated keypoints across a representative subset of subjects in the dataset. In sum, we arrive at an average error of 0.819 pixels. Additionally, we qualitatively assess the quality of the extracted mesh in Fig. 4 by projecting the extracted textured mesh onto the camera image plane using the calibrated camera parameters. It can be seen that the mesh aligns well with the reference frame, demonstrating the accuracy of camera calibration and the mesh extraction.



Figure 3. Samples of captures of *HumanOLAT*. The subjects are wearing a variety of clothing and take three different poses each.



Figure 4. Illustration of the MVS meshes provided within *HumanOLAT*. From left to right: (1) one of the 40 frames used for reconstruction, (2) the mesh rendered under flat color, (3) the mesh rendered with texture and (4) border of the mesh rendered onto the reference frame.

3.3.2. Mask Generation

Using estimated camera distortion parameters, we undistort captured images in accordance with the pin-hole camera model. Afterwards, to generate masks, we rely on the Sapiens models introduced by Khirodkar et al. [26]. We find that Sapiens, being designed for human-centric vision tasks, generally produces more accurate masks in our case compared to other popular methods, such as SAM [28, 41].

3.3.3. Motion Compensation

We note that as a single recording takes around 11 seconds, the subjects are unable to remain fully static over the course of a single capture and sway slightly. As illustrated in Fig. 5, for image-based relighting according to Debevec et al. [11], this leads to blurry results unsuitable for training. To combat this, we follow the method by Wenger et al. [50] to perform motion compensation: injecting a white light frame every 21th OLAT frame, we calculate optical flow towards a target frame—in our case the white illumination frame also used for mesh reconstruction and segmentation—using tracking-any-point (TAP) foundation model Co-Tracker3 [24] by tracking points sampled on the target frame over the concatenated tracking frames. Note that to keep computational cost reasonable, we track only $\sim$12k sparse grid points and linearly interpolate between them to arrive at the final dense flow. Finally, to motion-correct a given OLAT image, we linearly interpolate the flow for the two adjacent tracking frames and warp towards the target frame. See Fig. 5 for the qualitative results comparing summed raw and motion-compensated OLATs.



Figure 5. Comparison of summing OLAT illumination without (left) and with (right) motion compensation applied.

3.3.4. Normals and Image-Based Relighting

Following Guo et al. [17] and Zhou et al. [63], we provide pixel-wise photometric normals $\mathbf{n}$ computed from color and inverse color gradient illuminations g^+ and g^- as

$$\mathbf{n} = \frac{\mathbf{d}}{|\mathbf{d}|}, \quad \text{with} \quad \mathbf{d} = \frac{g^+ - g^-}{g^+ + g^-}. \tag{1}$$

We leverage the linearity of lighting to achieve accurate

image-based relighting under arbitrary environment maps. Specifically, given a target environment map E_{target}, we obtain the weighting color $\mathbf{c}_i$ for each OLAT image by masking E_{target} with each OLAT environment mask E_i and subsequent per-channel averaging. According to Debevec et al. [11], we then perform a per-channel linear combination of N_{OLAT} of motion-corrected OLAT frames I_i to acquire the final relit image I_{target}. This process is formulated as

$$I_{\text{target}} = \sum_{i=0}^{N_{\text{OLAT}}} \mathbf{c}_i I_i. \tag{2}$$

We visualize the estimated photometric normals from Eq. (1) and show relighting results obtained with Eq. (2) in Figs. 1 and 2.

3.4. Comparison with Existing Lightstage Datasets

A comparison of *HumanOLAT* with publicly available lightstage multi-illumination datasets is shown in Tab. 1. For object-centric scenes, ReNé [47] and OpenIllumination [35] both capture multi-view OLAT images under a variety of light positions, with the latter also providing illuminations under white light and pre-defined lighting patterns. Compared to their work, our dataset not only includes OLAT and white light illuminations but also provides color gradients and images captured under environment lights emulated by our lightstage. Moreover, our OLAT data is more fine-grained, with 331 illuminations compared to 40 in ReNé and 142 in OpenIllumination.

In the realm of human-centric lightstage captures, a majority of publicly available datasets focus only on the face, with ICT-3DRFE [37, 45], Dynamic OLAT [57] and Goliath-4 [10, 42] belonging to this list. Note that the latter also provides relightable hand captures. ICT-3DRFE aims to enable experimentation on relightable facial expressions and only provides a 3D face model with the intrinsics necessary for relighting. Dynamic OLAT and Goliath-4 focus on capturing white light and grouped OLAT illuminations of moving subjects, limiting their use outside the dynamic setting. The dataset most similar to ours is Ultrastage [63], which contains static relightable captures of 100 subjects performing different actions under a comparable number of views as in *HumanOLAT*. However, Ultrastage only provides three illuminations—consisting of white light and color gradients—for each capture, limiting their application in methods relying on detailed illumination data. In comparison to these works, we focus on recording multi-view images of the full human body under a multitude of illuminations, including white light, color gradients, environment maps and fine-grained OLATs, for broad-scale applications.

4. Baseline Experiments

This section evaluates several relighting baseline methods using our *HumanOLAT* dataset. We first assess 3DGS-based relighting methods designed for OLAT illumination (Sec. 4.1). Then, we test our dataset on the illumination harmonization task (Sec. 4.2). We also provide an evaluation of Wang et al. [48] in App. B.3. These evaluations aim to demonstrate the effectiveness and versatility of our dataset in various relighting scenarios. We present some results in the main paper and refer to App. B for more results.

4.1. Inverse Rendering from OLAT Illuminations

We evaluate the performance of four recent Gaussian-based inverse rendering methods that utilize OLAT illumination from our dataset: PRT-Gaussians [58], GS^3 [4], RNG [15], and BiGS [62]. To facilitate efficient training, we downsample the images to 1K resolution and evenly select 100 lights from 32 cameras, resulting in ≈3000 training frames. The remaining cameras and lights are reserved for validation. Since these methods assume static objects, we apply motion compensation as detailed in Section 3.3.3 to all frames and empirically brighten them by a factor of 10 to strengthen the training signal. We conduct baseline comparisons using the source code provided by the authors in all cases.

We evaluate the quantitative results using average PSNR, LPIPS, and SSIM metrics for six representative captures from our dataset, as shown in Tab. 2. These six captures were chosen to provide a balanced representation of the dataset's diversity. Out of the tested methods, GS^3 performs the best overall both qualitatively and numerically.

We present qualitative results of the tested methods in Fig. 6. While PRT-Gaussian captures some aspects of light transfer, it struggles to accurately reconstruct geometry from the OLAT frames, leading to poor quality and ghosting artifacts. In contrast, GS^3, RNG [15], and BiGS [62] show superior performance, achieving better geometry and more plausible relighting results. However, as illustrated in Fig. 6, the renderings remain somewhat blurry and display noticeable artifacts in the hand and face regions. These issues might stem from an insufficient number of Gaussians. Although we initialize the method with 300K sampled points, many are culled during training, resulting in final reconstructions with only approximately 50K–150K Gaussians. While GS^3 achieves the best overall performance, even here complex lighting effects like specular highlights and sharp shadows are not effectively captured.

In summary, while the evaluated methods perform well on simple object-centric scenes, they face challenges with the complex human-centric scenarios in our dataset. Therefore, *HumanOLAT* serves as a crucial benchmark for evaluating and stress testing object-centric relighting methods.

Dataset	Type	#Frames	#Subj.	#View	Res.	Illuminations			
						White	ColorGrad.	EnvMap	OLAT
ReNe [47]	objects	40k	20	50	1K	✗	✗	✗	✔
OpenIllumination [35]$^{+}$	objects	108k	64	72	4K	✔	✗	✗	✔
ICT-3DRFE [37, 45]†	heads	14k	23	2	1K	✗	✔	✗	✗
Dynamic OLAT [57]*	heads	603k	4*	1	4K	✔	✗	✗	✔
Goliath-4 [10, 42]*	hands/heads	>1M	4	150/110	4K	✔	✗	✗	✔
Ultrastage [63]	full bodies	192k	100	32	8K	✔	✔	✗	✗
HumanOLAT	full bodies	850k	21	40	5K	✔	✔	✔	✔

Table 1. Comparison of *HumanOLAT* with existing publicly available multi-illumination lightstage datasets regarding the number of frames (#Frames), captured subjects (#Subj.), views (#View), resolution (Res.) and provided types of illuminations. For the latter, "✔" and "✗"' denote whether a specific illumination is or is not contained within the dataset, while "✔" implies the illuminations are present in limited capacity. ("$^{+}$": OpenIllumination provides 13 pattern illuminations in addition to OLAT data; "†": ICT-3DRFE does not provide raw captures, but diffuse and specular normal and reflectance maps estimated from color gradients; "*": Dynamic OLAT and Goliath-4 provide *grouped* OLAT illuminations—with a subset of the lights on—for dynamically moving subjects).

4.2. Illumination Harmonization

To demonstrate potential applications of our dataset in illumination harmonization, we perform a qualitative survey for the state-of-the-art work IC-Light [59] on our data. To evaluate their method, we use the officially released model with our masked foreground images captured under white light and relight using both our background captures and the sample backgrounds provided by IC-Light for six subjects. Qualitative representative results are depicted in Fig. 7. For detailed quantitative results, please refer to App. B.1.

We observe that while the method can produce generally plausible relighting results, it struggles with photorealistic complex light transport effects, such as subsurface scattering, and consistently fails to preserve crucial facial details like eye color and mouth shape. Notably, IC-Light tends to focus on relighting based on normal direction and ambient occlusion, with self-shadows from elements like the head, arms, or clothing rarely observed.

We attribute these limitations primarily to the training data used by IC-Light, which relies on a mix of publicly available data, including augmented in-the-wild images, synthetically generated multi-illumination data, and lightstage captures for faces and objects, totaling approximately 10^7 frames. Additionally, they include an unspecified internal OLAT dataset with $2 \cdot 10^4$ appearances. To our knowledge, the method does not utilize significant full-body OLAT data for training, which limits its effectiveness on datasets like ours.

5. Discussion and Conclusion

Open Challenges Due to complex light transport caused by intricate geometry and diverse materials, photo-realistic depictions of full-body humans under novel illumination remains a challenging problem for many human-centric [8, 19, 31, 34, 36, 48, 53, 61] and general [4, 7, 13, 14, 16, 18, 21–23, 29, 32, 33, 43, 51, 55, 58, 62, 64] relighting approaches. While historically many works rely on data with limited lighting information, various recent works show that captures recorded in a lightstage—in particular, OLAT illuminations—provide suitable ground truth for developing photo-realistic relighting techniques [10, 27, 30, 39, 42, 49]. *HumanOLAT* is the first dataset to make such data publicly available, opening the associated avenues of research to the broader community.

Method	PSNR ↑	LPIPS ↓	SSIM ↑
PRT-Gaussian [58]	24.06	0.212	0.810
***GS*^{3}** [4]	**30.04**	**0.152**	0.892
RNG [15]	27.38	0.139	0.905
BiGS [62]	26.72	0.201	**0.936**

Table 2. Quantitative comparison between the performance of PRT-Gaussian, GS^3, RNG and BiGS on OLAT data from the *HumanOLAT* dataset. Results measure the quality of generated images under both novel view and novel illumination, averaged across six subjects with a randomly chosen pose.

Conclusion This paper introduces *HumanOLAT*, the first publicly available dataset to provide extensive full-body multi-view captures of humans under multiple illuminations, including white light, environment maps, color gradients and fine-grained OLATs. Further, we test our proposed dataset with state-of-the-art baselines for relighting under novel illuminations and novel views. The results demonstrate that while current methods achieve plausible results, they do not capture fully the complex lighting effects characteristic to human bodies. We believe the proposed dataset forms a valuable basis for developing and benchmarking future general and human-centric relighting methods.

Acknowledgement. This research was supported by NVIDIA.



Figure 6. Qualitative results for GS^3 [4], PRT-Gaussian [58], RNG [15] and BiGS [62] under novel views and illuminations.

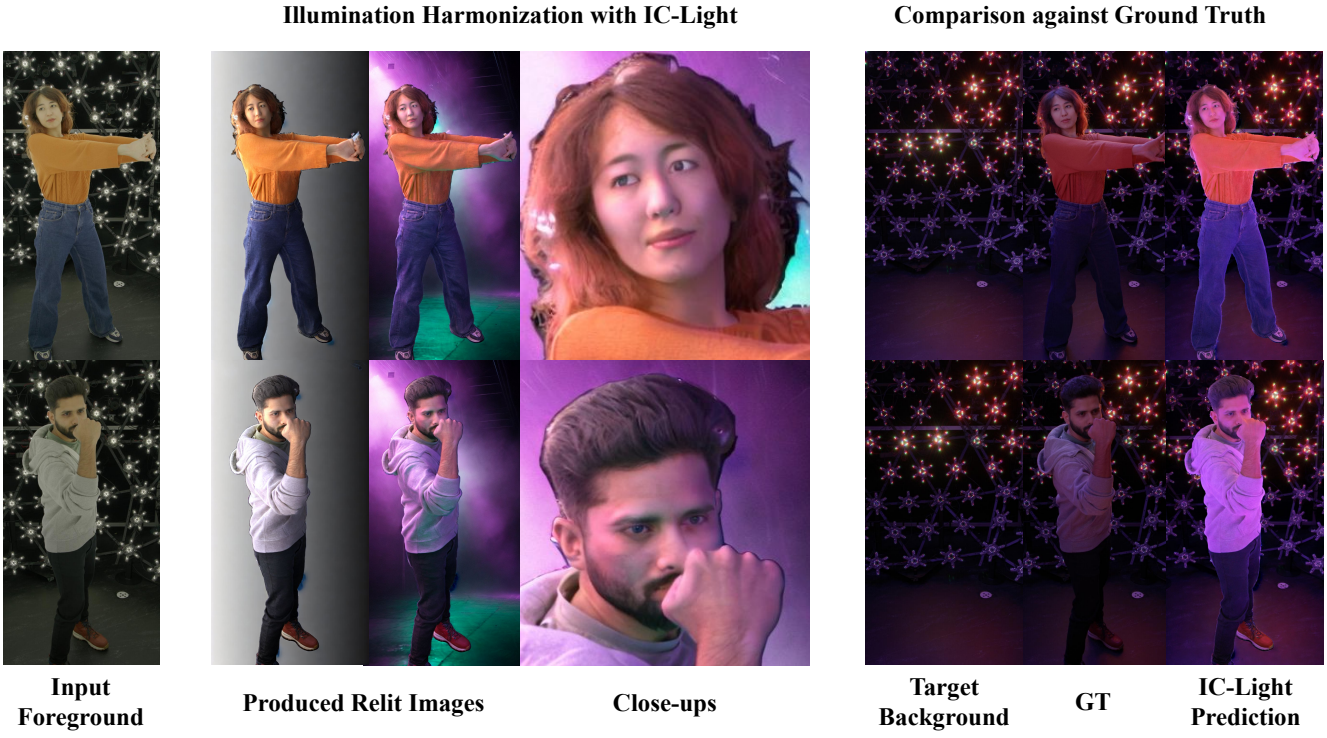


Figure 7. Illumination harmonization with IC-Light [59] on *HumanOLAT*. Given an input image (left), we show both relighting results for background images sampled from the selection provided in IC-Light [59] (middle) and for a background image of the empty lightstage (right). For latter, we also provide the actual expected illumination of the target given the light emitted by the lightstage.

References

[1] Easymocap - make human motion capture easier. Github, 2021. 1, 2

[2] Agisoft LLC. Metashape. https://www.agisoft.com/downloads/installer/, 2025. Version retrieved July 2025. 4

[3] Sai Bi, Stephen Lombardi, Shunsuke Saito, Tomas Simon, Shih-En Wei, Kevyn Mcphail, Ravi Ramamoorthi, Yaser Sheikh, and Jason Saragih. Deep relightable appearance models for animatable faces. *ACM Trans. Graph.*, 40(4), 2021. 3

[4] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. Gs3: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia*, 2024. 2, 3, 6, 7, 8, 1, 4

[5] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. Nerd: Neural reflectance decomposition from image collections. In *International Conference on Computer Vision (ICCV)*, 2021. 2

[6] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 4, 1, 2

[7] Hongze Chen, Zehong Lin, and Jun Zhang. Gi-gs: Global illumination decomposition on gaussian splatting for inverse rendering. In *International Conference on Learning Representations (ICLR)*, 2025. 3, 7

[8] Yushuo Chen, Zerong Zheng, Zhe Li, Chao Xu, and Yebin Liu. Meshavatar: Learning high-quality triangular human avatars from multi-view videos. In *European Conference on Computer Vision (ECCV)*, 2024. 3, 7

[9] Zhaoxi Chen and Ziwei Liu. Relighting4d: Neural relightable human from videos. In *European Conference on Computer Vision (ECCV)*, 2022. 2

[10] Zhaoxi Chen, Gyeongsik Moon, Kaiwen Guo, Chen Cao, Stanislav Pidhorskyi, Tomas Simon, Rohan Joshi, Yuan Dong, Yichen Xu, Bernardo Pires, He Wen, Lucas Evans, Bo Peng, Julia Buffalini, Autumn Trimble, Kevyn McPhail, Melissa Schoeller, Shoou-I Yu, Javier Romero, Michael Zollhöfer, Yaser Sheikh, Ziwei Liu, and Shunsuke Saito. URhand: Universal relightable hands. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3, 6, 7

[11] Paul Debevec, Tim Hawkins, Chris Tchou, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *SIGGRAPH*, 2000. 1, 2, 3, 5, 6

[12] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Computer Vision and Pattern Recognition (CVPR)*, 2022. 3

[13] Jan-Niklas Dihlmann, Arjun Majumdar, Andreas Engelhardt, Raphael Braun, and Hendrik Lensch. Subsurface scattering for gaussian splatting. In *Neural Information Processing Systems (NeurIPS)*, 2024. 2, 3, 7

[14] Kang Du, Zhihao Liang, and Zeyu Wang. Gs-id: Illumination decomposition on gaussian splatting via diffusion prior and parametric light source optimization. In *International Conference on Computer Vision (ICCV)*, 2025. 2, 3, 7

[15] Jiahui Fan, Fujun Luan, Jian Yang, Milos Hasan, and Beibei Wang. Rng: Relightable neural gaussians. In *Computer Vision and Pattern Recognition (CVPR)*, 2025. 2, 6, 7, 8, 1, 4

[16] Jian Gao, Chun Gu, Youtian Lin, Hao Zhu, Xun Cao, Li Zhang, and Yao Yao. Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing. In *European Conference on Computer Vision (ECCV)*, 2024. 2, 3, 7

[17] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, Danhang Tang, Anastasia Tkach, Adarsh Kowdle, Emily Cooper, Mingsong Dou, Sean Fanello, Graham Fyffe, Christoph Rhemann, Jonathan Taylor, Paul Debevec, and Shahram Izadi. The relightables: volumetric performance capture of humans with realistic relighting. *ACM Trans. Graph.*, 38(6), 2019. 2, 3, 4, 5

[18] Mingming He, Pascal Clausen, Ahmet Levent Taşel, Li Ma, Oliver Pilarski, Wenqi Xian, Laszlo Rikker, Xueming Yu, Ryan Burgert, Ning Yu, and Paul Debevec. Diffrelight: Diffusion-based facial performance relighting. In *SIGGRAPH Asia*, 2024. 3, 7

[19] Umar Iqbal, Akin Caliskan, Koki Nagano, Sameh Khamis, Pavlo Molchanov, and Jan Kautz. Rana: Relightable articulated neural avatars. In *International Conference on Computer Vision (ICCV)*, 2023. 2, 3, 7

[20] Shun Iwase, Shunsuke Saito, Tomas Simon, Stephen Lombardi, Timur Bagautdinov, Rohan Joshi, Fabian Prada, Takaaki Shiratori, Yaser Sheikh, and Jason Saragih. Relightablehands: Efficient neural relighting of articulated hand models. In *Computer Vision and Pattern Recognition (CVPR)*, 2023. 3

[21] Chaonan Ji, Tao Yu, Kaiwen Guo, Jingxin Liu, and Yebin Liu. Geometry-aware single-image full-body human relighting. In *European Conference on Computer Vision (ECCV)*, 2022. 3, 7

[22] Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, and Yuexin Ma. Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3

[23] Yoshihiro Kanamori and Yuki Endo. Relighting humans: occlusion-aware inverse rendering for full-body human images. 37(6), 2018. 3, 7

[24] Nikita Karaev, Iurii Makarov, Jianyuan Wang, Natalia Neverova, Andrea Vedaldi, and Christian Rupprecht. Cotracker3: Simpler and better point tracking by pseudo-labelling real videos. In *International Conference on Computer Vision (ICCV)*, 2025. 5

[25] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 3

[26] Rawal Khirodkar, Timur Bagautdinov, Julieta Martinez, Su Zhaoen, Austin James, Peter Selednik, Stuart Anderson, and

Shunsuke Saito. Sapiens: Foundation for human vision models. In *European Conference on Computer Vision (ECCV)*, 2024. 4, 5

[27] Hoon Kim, Minje Jang, Wonjun Yoon, Jisoo Lee, Donghyun Na, and Sanghyun Woo. Switchlight: Co-design of physics-driven architecture and pre-training framework for human portrait relighting. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3, 7

[28] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *International Conference on Computer Vision (ICCV)*, 2023. 5

[29] Zhiyi Kuang, Yanchao Yang, Siyan Dong, Jiayue Ma, Hongbo Fu, and Youyi Zheng. Olat gaussians for generic relightable appearance acquisition. In *SIGGRAPH Asia*, 2024. 2, 3, 7

[30] Junxuan Li, Chen Cao, Gabriel Schwartz, Rawal Khirodkar, Christian Richardt, Tomas Simon, Yaser Sheikh, and Shunsuke Saito. Uravatar: Universal relightable gaussian codec avatars. In *SIGGRAPH*, 2024. 2, 3, 7

[31] Zhe Li, Yipengjing Sun, Zerong Zheng, Lizhen Wang, Shengping Zhang, and Yebin Liu. Animatable and relightable gaussians for high-fidelity human avatar modeling. *arXiv preprint arXiv:2311.16096v4*, 2024. 3, 7

[32] Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, and Kui Jia. Gs-ir: 3d gaussian splatting for inverse rendering. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3, 7

[33] Zhihao Liang, Hongdong Li, Kui Jia, Kailing Guo, and Qi Zhang. Gus-ir: Gaussian splatting with unified shading for inverse rendering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP:1–15, 2025. 2, 7

[34] Wenbin Lin, Chengwei Zheng, Jun-Hai Yong, and Feng Xu. Relightable and animatable neural avatars from videos. In *Association for the Advancement of Artificial Intelligence (AAAI)*, 2024. 3, 7

[35] Isabella Liu, Linghao Chen, Ziyang Fu, Liwen Wu, Haian Jin, Zhong Li, Chin Ming Ryan Wong, Yi Xu, Ravi Ramamoorthi, Zexiang Xu, et al. Openillumination: A multi-illumination dataset for inverse rendering evaluation on real objects. In *Neural Information Processing Systems (NeurIPS)*, 2023. 2, 3, 6, 7

[36] Diogo Luvizon, Vladislav Golyanik, Adam Kortylewski, Marc Habermann, and Christian Theobalt. Relightable neural actor with intrinsic decomposition and pose control. In *European Conference on Computer Vision (ECCV)*, 2024. 2, 3, 7

[37] Wan-Chun Ma, Tim Hawkins, Pieter Peers, Charles-Felix Chabert, Malte Weiss, and Paul Debevec. Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In *Eurographics Symposium on Rendering (EGSR)*, 2007. 6, 7

[38] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, 2020. 2

[39] Rohit Pandey, Sergio Orts-Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul E Debevec, and Sean Ryan Fanello. Total relighting: learning to relight portraits for background replacement. *ACM Trans. Graph.*, 40(4):43–1, 2021. 2, 3, 7

[40] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3D hands, face, and body from a single image. In *Computer Vision and Pattern Recognition (CVPR)*. 4, 1

[41] Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. In *International Conference on Learning Representations (ICLR)*, 2025. 5

[42] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3, 6, 7

[43] Yahao Shi, Yanmin Wu, Chenming Wu, Xing Liu, Chen Zhao, Haocheng Feng, Jian Zhang, Bin Zhou, Errui Ding, and Jingdong Wang. Gir: 3d gaussian inverse rendering for relightable scene factorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–12, 2025. 2, 3, 7

[44] Pratul P. Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T. Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[45] Giota Stratou, Abhijeet Ghosh, Paul Debevec, and Louis-Philippe Morency. Effect of illumination on automatic expression recognition: A novel 3d relightable facial database. In *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2011. 2, 6, 7

[46] Tiancheng Sun, Jonathan T. Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. Single image portrait relighting. *ACM Trans. Graph.*, 38(4), 2019. 3

[47] Marco Toschi, Riccardo De Matteo, Riccardo Spezialetti, Daniele De Gregorio, Luigi Di Stefano, and Samuele Salti. Relight my nerf: A dataset for novel view synthesis and relighting of real world objects. In *Computer Vision and Pattern Recognition (CVPR)*, 2023. 2, 6, 7

[48] Shaofei Wang, Božidar Antić, Andreas Geiger, and Siyu Tang. Intrinsicavatar: Physically based inverse rendering of dynamic humans from monocular videos via explicit ray tracing. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 3, 6, 7, 1, 2

[49] Shaofei Wang, Tomas Simon, Igor Santesteban, Timur Bagautdinov, Junxuan Li, Vasu Agrawal, Fabian Prada, Shoou-I Yu, Pace Nalbone, Matt Gramlich, Roman Lubachersky, Chenglei Wu, Javier Romero, Jason Saragih, Michael Zollhoefer, Andreas Geiger, Siyu Tang, and Shunsuke Saito.

Relightable full-body gaussian codec avatars. In *SIGGRAPH*, 2025. 3, 7

[50] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Transactions on Graphics (TOG)*, 24:756–764, 2005. 5

[51] Tong Wu, Jia-Mu Sun, Yu-Kun Lai, Yuewen Ma, Leif Kobbelt, and Lin Gao. Deferredgs: Decoupled and relightable gaussian splatting with deferred shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP, 2025. 2, 3, 7

[52] Zhangyang Xiong, Chenghong Li, Kenkun Liu, Hongjie Liao, Jianqiao Hu, Junyi Zhu, Shuliang Ning, Lingteng Qiu, Chongjie Wang, Shijie Wang, Shuguang Cui, and Xiaoguang Han. Mvhumannet: A large-scale dataset of multi-view daily dressing human captures. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[53] Zhen Xu, Sida Peng, Chen Geng, Linzhan Mou, Zihan Yan, Jiaming Sun, Hujun Bao, and Xiaowei Zhou. Relightable and animatable neural avatar from sparse-view video. In *Computer Vision and Pattern Recognition (CVPR)*, 2024. 3, 7

[54] Haotian Yang, Mingwu Zheng, Wanquan Feng, Haibin Huang, Yu-Kun Lai, Pengfei Wan, Zhongyuan Wang, and Chongyang Ma. Towards practical capture of high-fidelity relightable avatars. In *SIGGRAPH Asia*, 2023. 3

[55] Silong Yong, Venkata Nagarjun Pudureddiyur Manivannan, Bernhard Kerbl, Zifu Wan, Simon Stepputtis, Katia Sycara, and Yaqi Xie. Omg: Opacity matters in material modeling with gaussian splatting. In *International Conference on Learning Representations (ICLR)*, 2025. 2, 3, 7

[56] Kai Zhang, Fujun Luan, Qianqian Wang, Kavita Bala, and Noah Snavely. PhySG: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In *Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[57] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Neural video portrait relighting in real-time via consistency modeling. In *International Conference on Computer Vision (ICCV)*, 2021. 2, 6, 7

[58] Libo Zhang, Yuxuan Han, Wenbin Lin, Jingwang Ling, and Feng Xu. Prtgaussian: Efficient relighting using 3d gaussians with precomputed radiance transfer. In *2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2024. 2, 3, 6, 7, 8, 1, 4

[59] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In *International Conference on Learning Representations (ICLR)*, 2025. 3, 7, 8, 1, 4

[60] Xiuming Zhang, Pratul P. Srinivasan, Boyang Deng, Paul Debevec, William T. Freeman, and Jonathan T. Barron. Nerfactor: neural factorization of shape and reflectance under an unknown illumination. *ACM Trans. Graph.*, 40(6), 2021. 2

[61] Yang Zheng, Qingqing Zhao, Guandao Yang, Wang Yifan, Donglai Xiang, Florian Dubost, Dmitry Lagun, Thabo Beeler, Federico Tombari, Leonidas Guibas, and Gordon Wetzstein. Physavatar: Learning the physics of dressed 3d avatars from visual observations. In *European Conference on Computer Vision (ECCV)*, 2024. 3, 7

[62] Liu Zhenyuan, Yu Guo, Xinyuan Li, Bernd Bickel, and Ran Zhang. Bigs: Bidirectional gaussian primitives for relightable 3d gaussian splatting. In *International Conference on 3D Vision (3DV)*, 2025. 2, 3, 6, 7, 8, 1, 4

[63] Taotao Zhou, Kai He, Di Wu, Teng Xu, Qixuan Zhang, Kuixiang Shao, Wenzheng Chen, Lan Xu, and Jingyi Yu. Relightable neural human assets from multi-view gradient illuminations. In *Computer Vision and Pattern Recognition (CVPR)*, 2023. 2, 3, 4, 5, 6, 7

[64] Zuo-Liang Zhu, Beibei Wang, and Jian Yang. Gs-ror^2: Bidirectional-guided 3dgs and sdf for reflective object relighting and reconstruction, 2025. 2, 3, 7