DarkGS: Learning Neural Illumination and 3D Gaussians Relighting for Robotic Exploration in the Dark

Tianyi Zhang[1], Kaining Huang[1], Weiming Zhi[1], Matthew Johnson-Roberson[1]

***Abstract*— Humans have the remarkable ability to construct consistent mental models of an environment, even under limited or varying levels of illumination. We wish to endow robots with this same capability. In this paper, we tackle the challenge of constructing a photorealistic scene representation under poorly illuminated conditions and with a moving light source. We approach the task of modeling illumination as a learning problem, and utilize the developed illumination model to aid in scene reconstruction. We introduce an innovative framework that uses a data-driven approach, *Neural Light Simulators* (NeLiS), to model and calibrate the camera-light system. Furthermore, we present DarkGS, a method that applies NeLiS to create a relightable 3D Gaussian scene model capable of real-time, photorealistic rendering from novel viewpoints. Code released at https://github.com/tyz1030/neuralight.git**

I. INTRODUCTION

Robots and autonomous vehicles have been routinely deployed in poorly illuminated environments for critical missions and tasks such as exploration, inspection, transportation, search and rescue, etc. (see Fig. 1a). Imaging systems consisting of one or multiple RGB cameras and light sources are often equipped on the robot to illuminate and sense the surrounding environment. The streamed image sequence can be further used in downstream tasks, e.g. navigation, mapping, and visualization, to boost the robot autonomy and human understanding of the environment.

Scene reconstruction, or the capability to create accurate internal representations of the environment, is vital for robots operating in unknown environments. Previous vision-only approaches largely rely on identifying common feature points over a set of multiple-view images, and then minimizing a reprojection error [4]. Such procedures like Structure-from-Motion (SfM) or Simultaneous Localization and Mapping (SLAM) also estimate the camera poses of the images. Based on these camera poses, Neural Radiance Field (NeRF) [5] is capable of achieving photorealistic scene reconstruction by optimizing a photometric loss between the representation and the images. However, while achieving huge success in the graphics community, the transition of objective function from reprojection error to photometric loss has raised a new challenge to the robotics community: Can we still build consistent scene representations with a moving light source on the robot platform?

Concretely, the problem tackled in this paper is as such: Given a sequence of images taken in poorly illuminated environments, with one major light source moving with one camera as a rigid body, reconstruct the scene by minimizing photometric loss and achieve photorealistic novel view image synthesis (Fig. 1b). Our contributions are as follows:

(a) Robots in the dark with onboard light sources

(b) Our work build 3D Gaussians and relight the scene in dark

Fig. 1: Robotic imaging systems working in the dark consist of cameras and light sources: Carla Simulator [1], Team CoStar in SubT Challenge [2] and HoloOcean underwater robot simulator [3]. We propose a pipeline that calibrates the camera-light system which helps photorealistic scene reconstruction and relighting from images collected in the dark.

1) A pipeline that consists of light source modeling, camera-light calibration, building 3D Gaussians and scene relighting from illumination-inconsistent images.
2) Neural Light Simulators (NeLiS), a data-driven and physically interpretable illumination model and software for light source modeling and calibration.
3) Dark Gaussian Splatting (DarkGS), a variant of the 3D Gaussian Splatting (3DGS) model that builds photorealistic scene representations under poorly illuminated conditions and relights the scene with global illumination, based on COLMAP [6] and NeLiS results.

II. RELATED WORK

A. Extroceptive Sensor Calibration on Robots

Autonomous robots are usually equipped with perceptual sensors such as cameras and LiDARs. Taking camera calibration as an example, a calibration target with AprilTag [7] or checkerboard pattern is often used. With observations of the target from different perspectives, the focal length, center of projection, distortion coefficient, and pose of the camera can be estimated [8]. Similar target-based approaches have

[1]Authors are with the Robotics Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA {tianyiz4,kaining2,wzhi,mkj}@andrew.cmu.edu

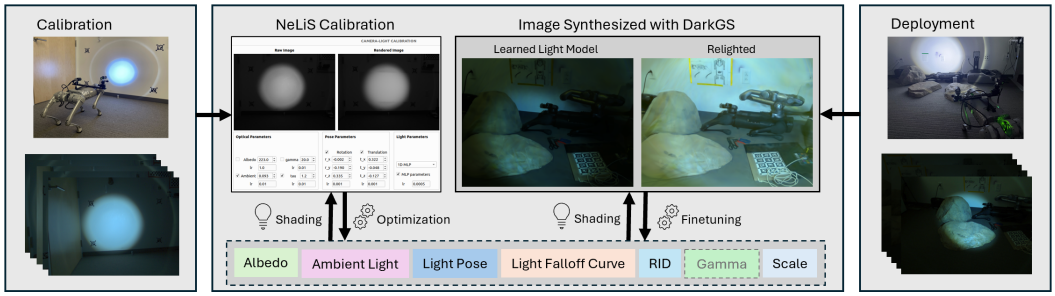


Fig. 2: Our proposed workflow: Images for camera-light calibration are first collected at a calibration target. With NeLiS, we manually initialize the parameters and then optimize the light model. The model can then be used to build DarkGS, present the scene with learned or relighted illumination.

been used to calibrate LiDARs [9], radars [10] and acoustic sensors [11] for downstream sensor fusion tasks [12]. Analogous to sensor calibrations mentioned above, in this paper, we propose to use a target consisting of AprilTags [7] and blank space for calibrating the light in a camera-light system, including estimating the transformation between camera's and light's coordinate system, radiant intensity distribution (RID) and light fall-off curve of the light source.

B. Light Calibration

Existing methods use various kinds of customized calibration targets: [13][14] propose to use a target of AprilTags and pins to infer the position of point light source from the shadows of the pins. Alternatively, [15] [16] propose to use a Lambertian sphere and estimate the light source parameters by learning to reconstruct the sphere. [17] shows the most relevance to our work, which calibrates the pose of a light source and the metric scale given the RID curve.

C. NeRF and 3D Gaussian Splatting

Based on the success of SfM [4], NeRF learns to represent scenes using a continuous function, e.g. a multilayer perceptron (MLP), and can achieve photorealistic novel-view image synthesis. Gaussianshader [18] and Relightable 3DG [19] introduce physical properties into the 3DGS framework but model illumination as a constant that does not change frame-by-frame. According to our experiments, none of the above mentioned methods handles the illumination-inconsistancy issue on a real-robot imaging system. In our proposed DarkGS, by modeling the physical property of the scene and taking advantage of NeLiS, we can not only build a consistent 3DGS from poorly illuminated images, but also relight the scene with global illumination.

III. METHODOLOGY

Our proposed workflow for building NeLiS and DarkGS is shown in Fig. 2. We build NeLiS by calibrating the camera-light system at a planar calibration target. After obtaining a calibrated NeLiS, the robot can be deployed in a dark environment, using the collected images to build a DarkGS, finetune the model, and relight the scene.

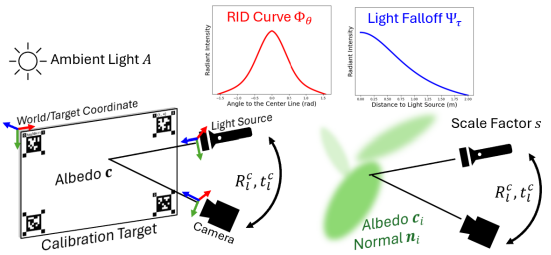


Fig. 3: Our shading model: (Left) In NeLiS, camera poses are localized by AprilTags on the calibration target.

A. Shading Model for NeLiS

The calibration data is taken by capturing photos at a calibration target from different views while the light source moves with the camera as a rigid body. The calibration target is a white plane with four AprilTags positioned as four corners of a rectangle (as shown in Fig. 3). We attach the origin of the world coordinate to the top-left corner of the calibration target. We assume that the camera is precalibrated with distortions removed. One of the key problems for NeLiS to solve is estimating the relative pose of light to the camera $R_l^c \in SO(3)$ and $t_l^c \in \mathbb{R}^3$.

Given the true size of the calibration target, we can apply Perspective-n-Point algorithm (PnP) [20] to extract $R_c^w \in SO(3)$ and $t_c^w \in \mathbb{R}^3$ which transform points from the camera coordinate to the world coordinate. The position of the light source is then given by $\mathbf{o}_l = R_c^w t_l^c + t_c^w \in \mathbb{R}^3$. Since we align the z axis of the light coordinate with the centerline of the light, the direction of the centerline can be denoted by $\boldsymbol{\omega}_l = R_c^w R_l^c [0,0,1]^\top \in \mathbb{R}^3$. Both $\mathbf{o}_l$ and $\boldsymbol{\omega}_l$ are in the world coordinate frame.

We only use the area bounded by 4 AprilTags as the Region of Interest (ROI) to do calibrations. We assume that this area is a Lambertian plane and has the same normal $\mathbf{n} \in \mathbb{R}^3$ and diffusive albedo $\mathbf{c} \in \mathbb{R}^\lambda$ anywhere on the plane ($\lambda = 1$ for grayscale images and $\lambda = 3$ for RGB). For each pixel in the ROI, we find the intersection of the corresponding camera ray and the target plane in the world coordinate system, denoted by $\mathbf{x} \in \mathbb{R}^3$. To infer the incident radiance at $\mathbf{x}$, one needs to model the RID, light falloff function, and ambient light.

1) RID: RID is commonly modeled as a function of the angle between the centerline $\boldsymbol{\omega}_l$ of the light and light ray $\boldsymbol{\omega}_\mathbf{x} = \mathbf{x} - \mathbf{o}_l$. Previous work assume that RID is given [16], [17], [21]. However, this assumption may not hold for in-the-wild robot deployment. Instead, we remove this dependency by learning a neural RID from calibration data:

$$\Phi_\theta(\mathbf{x}) = \text{MLP}_\theta(\cos^{-1}(\frac{\boldsymbol{\omega}_\mathbf{x}}{\|\boldsymbol{\omega}_\mathbf{x}\|_2} \cdot \boldsymbol{\omega}_l)) \tag{1}$$

here θ denotes the parameters of the MLP.

2) Light Falloff Curve: Inverse square law is widely used to model light falloff, based on the assumption of a point light source. However, when objects are closer to the light source, the inverse square law starts to fail. We choose to model the light falloff in the form of a Lorentzian function of the distance $\|\boldsymbol{\omega}_\mathbf{x}\|_2$, as suggested by [22]:

$$\Psi_\tau(\mathbf{x}) = \frac{1}{\tau + \|\boldsymbol{\omega}_\mathbf{x}\|_2^2} \tag{2}$$

Instead of estimating τ from hand measurement of the light surface [22], we designate it to be a learnable parameter.

3) Ambient Light: We model ambient light as a learnable parameter A. The incident radiance at the point $\mathbf{x}$ can thereby be modeled as:

$$I_\mathbf{x} = \Psi_\tau(\mathbf{x})\Phi_\theta(\mathbf{x}) + A \tag{3}$$

4) BRDF: We opt for the Lambertian reflection model, eliminating the need to optimize any parameters in our bidirectional reflectance distribution function (BRDF). We use $f_r(\boldsymbol{\omega}_\mathbf{x}, \mathbf{n}) = \max(\boldsymbol{\omega}_\mathbf{x} \cdot \mathbf{n}, 0)$ as our BRDF, giving the rendering equation:

$$\hat{L}_\mathbf{x} = I_\mathbf{x} f_r(\boldsymbol{\omega}_\mathbf{x}, \mathbf{n})\mathbf{c} \tag{4}$$

With captured pixel intensity $L_\mathbf{x}$, we use L1 loss and formulate the NeLiS optimization problem as:

$$\min_{\theta, A, \tau, R_l^c, t_l^c, \mathbf{c}} \sum_{\mathbf{x} \in \text{ROI}} \|L_\mathbf{x} - \hat{L}_\mathbf{x}\|_1 \tag{5}$$

B. Building DarkGS

Within the framework of 3DGS, we model the scene with a point cloud of Gaussians $\mathbf{G}$ (as shown in Fig. 3 right). Each Gaussian g_i in the point cloud encompasses attributes including position $\mathbf{p}_i$, covariance Σ_i, opacity α_i, albedo $\mathbf{c}_i$ and normal $\mathbf{n}_i$, that $g_i = \{\mathbf{p}_i, \Sigma_i, \alpha_i, \mathbf{c}_i, \mathbf{n}_i\} \in \mathbf{G}$.

Given $\mathbf{p}_i$, the incident radiance I_i can be calculated by Eq. 3. With $\mathcal{N}$ ordered points for pixel (u, v), the rendering equation then becomes:

$$\hat{L}_{u,v} = \sum_{i \in \mathcal{N}} I_i f_r(\boldsymbol{\omega}_i, \mathbf{n}_i)\mathbf{c}_i\alpha_i \prod_j^{i-1}(1 - \alpha_j) \tag{6}$$

1) Scale Recovery: The framework of 3DGS and its variants are often based on monocular SfM solutions such as COLMAP [4]. Monocular SfM only gives up-to-scale poses for building 3D Gaussians. Here, we introduce a scaling factor $s > 0$ as a learnable parameter so that we can obtain the positions with scale $\mathbf{p}_i' = s\mathbf{p}_i$. With captured pixel intensity $L_{u,v}$, the 3D Gaussian can be built by solving the follow optimization problem:

$$\min_{A, \mathbf{G}, s} \sum \|L_{u,v} - \hat{L}_{u,v}\|_1 \tag{7}$$

2) Training Warm-Up: In some cases, a large discrepancy between the initial scale and the true scale often leads to divergence and local minimum at the beginning of optimization. We define a warm-up factor that grows with iterations, so that in the m^{th} iteration, this factor is $\frac{m}{k}$. Here, we denote Lie exponential map and Lie logarithm map as follows:

$$\exp(\cdot) : \mathbb{R}^3 \to SO(3), \quad \log(\cdot) : SO(3) \to \mathbb{R}^3 \tag{8}$$

so that in warm-up stage we replace R_l^c and t_l^c with:

$$\hat{R}_l^c = \exp(\frac{m}{k}\log(R_l^c)), \quad \hat{t}_l^c = \frac{m}{k} t_l^c \tag{9}$$

3) Relighting: Once the DarkGS model is built, we can relight the scene by replacing the components in Eq. 3, i.e. Ψ_τ, Φ_θ and A, by carefully designed values and functions. For example, replacing the MLP function in Φ_θ with a constant can create Lambertian illumination without a pattern.

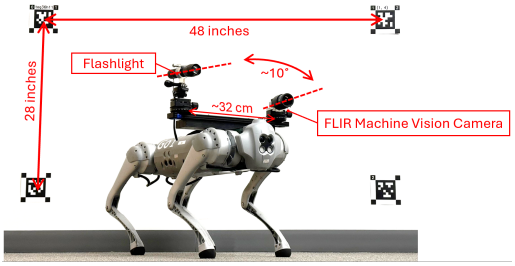


Fig. 4: Our experiment setup: The imaging system is installed on a legged robot platform (Unitree GO1). We use a FLIR machine vision camera to stream the images in RAW format.

IV. EXPERIMENTS

A. Experiments Setup

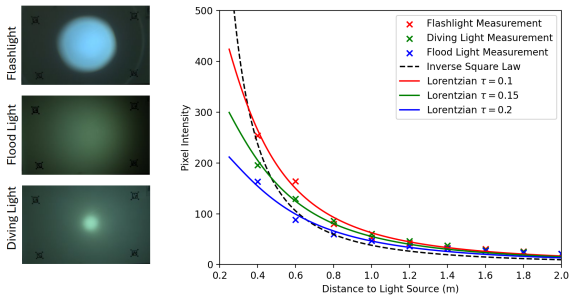
Our imaging system consists of a FLIR machine vision camera (Model Firefly S) and a light source, together mounted on a rigid structure on top of our legged robot, as shown in Fig. 4. The baseline between the camera and the light source is approximately 32 cm, measured by hand. We experimented with three different light sources: a flashlight, a diving light, and a flood light.

B. Can existing 3DGS methods do the job?

We first investigated whether existing 3DGS methods can reconstruct the scene using the images we collected with our robotic setup in the dark. We experimented with: 1) vanilla 3DGS [23] which models the scene with constant radiance 2) RawNeRF [24] which is developed to reconstruct scene from RAW HDR images 3) Relightable 3DGS [19] which models the shading and environmental light map.

As shown in Fig. 5, we found that none of the existing methods is able to build valid scene reconstructions. We see an excessive amount of artifacts in the center area of the synthesized image. The key reason is that the capability to model varying illumination is missing from existing methods.

Fig. 5: None of the existing methods can solve the problem: Results of Vanilla 3DGS [23], RawNeRF [24] Relightable 3DGS [19] show heavy artifacts and fail to converge.



(a) RID (b) Light Fall-off Curve

Fig. 6: (a) Light sources on real robots have various RID patterns; (b) Real world measurements of light falloff show that the inverse square law is insufficient to model any of our light sources, but Lorentzian functions [22] with learnable parameter τ fit them well.

C. Why do we need to learn RID?

As shown in Fig. 6a, different light has different RID patterns. Existing methods often model RID as known [17], or general functions such as the power of cosine functions [16] or a Gaussian distribution. However, for most light sources, and all light sources we used in this research, RID is not given. Modeling them with general functions such as a Gaussian distribution will not be adaptive and expressive enough to reflect the true RID of different light patterns for building a photorealistic renderer. The ablation study in Fig. 7 (columns 2 and 5) shows that while a Gaussian distribution fits the RID of flood light well, it performs much worse on the other two types of light. In comparison, our MLP-based model fits all 3 kinds of light with the best performance, showing good adaptivity to different light patterns.

D. Why do we need to learn light falloff?

For simplicity, previous studies [16], [17], [21] use an inverse square law to model the light falloff. It is not a good fit for real-world light systems. We measure the light falloff of the center point of 3 different light sources as shown in Fig. 6b. The plot show that when the range is close, light fall-off does not follow the inverse square law as the point light source assumption starts to fail. However, with a Lorentzian model suggested by [22], the light falloff can be better fitted with our introduced learnable parameter τ. We further show numerical results in Fig. 7 (columns 2 and 3) that by learning τ, the rendering performance on testing set of all 3 kinds of lights gets improved, which implies that our light fall-off model better approximates the true light falloff.

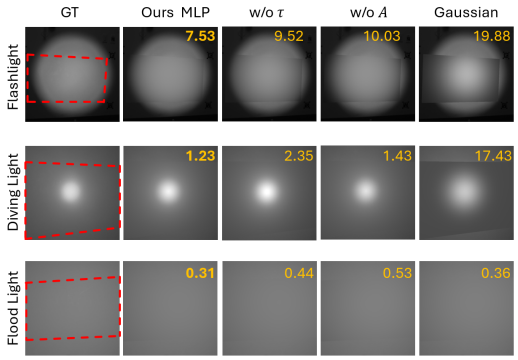


Fig. 7: Ablation study: We show that our model with an MLP-based RID, τ and A can effectively improve the rendering performance. MSE of the testing set are highlighted in yellow. The red dashed box bounds the regions to be rendered.

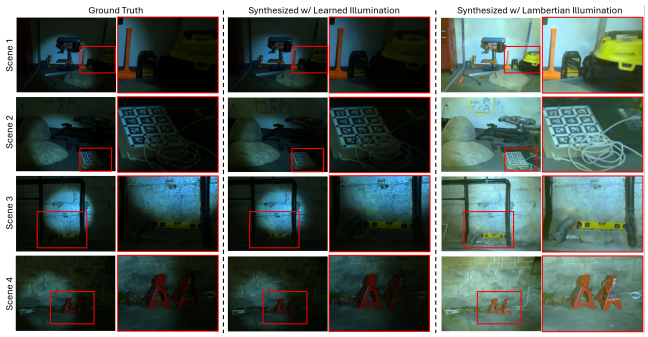


Fig. 8: Visulization of results from multiple scenes: We show that with DarkGS, we can reconstruct the scene with RAW images from robotic deployments in dark environments and relight the scene.

We deployed our system in various real-world environments and the results are shown in Fig. 8. Compared with ground truth, our model is able to reconstruct the image with photorealistic quality with learned illumination. Then we replace the light source in our model with a Lambertian light to create a global illumination that illuminates the entire field of view (FOV) of the camera with photorealistic rendering quality.

V. CONCLUSIONS AND FUTURE WORK

This work aims to solve the problem of building 3D Gaussians and scene relighting from images taken by a moving camera-light system. Our proposed pipeline consists of NeLiS, a camera-light simulation and calibration model, and DarkGS which build's photolistic representation for scenes in the dark. The results show that our proposed pipeline can build relightable Gaussians from images taken by the robot platform deployed in the dark Future work includes modeling shadows and non-Lambertian objects, and bringing white balance and tone mapping into the loop.

ACKNOWLEDGEMENTS

This work is supported by NOAA NA22OAR0110624.

REFERENCES

[1] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[2] A. Agha, K. Otsu, B. Morrell, *et al.*, "Nebula: Quest for robotic autonomy in challenging environments; team costar at the darpa subterranean challenge," *arXiv preprint arXiv:2103.11470*, 2021.

[3] E. Potokar, S. Ashford, M. Kaess, and J. G. Mangelson, "Holoocean: An underwater robotics simulator," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 3040–3046.

[4] S. Agarwal, Y. Furukawa, N. Snavely, *et al.*, "Building rome in a day," *Communications of the ACM*, vol. 54, pp. 105–112, 2011.

[5] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *ECCV*, 2020.

[6] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[7] E. Olson, "Apriltag: A robust and flexible visual fiducial system," in *2011 IEEE International Conference on Robotics and Automation*, 2011, pp. 3400–3407.

[8] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2016, pp. 4304–4311.

[9] J. Beltrán, C. Guindel, A. de la Escalera, and F. García, "Automatic extrinsic calibration method for lidar and camera sensor setups," *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[10] J. Domhof, J. F. P. Kooij, and D. M. Gavrila, "A joint extrinsic calibration tool for radar, camera and lidar," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 3, pp. 571–582, 2021.

[11] D. Yang, B. He, M. Zhu, and J. Liu, "An extrinsic calibration method with closed-form solution for underwater opti-acoustic imaging system," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6828–6842, 2020.

[12] J. Song, L. Zhao, and K. A. Skinner, "Lirafusion: Deep adaptive lidar-radar fusion for 3d object detection," *arXiv preprint arXiv:2402.11735*, 2024.

[13] H. Santo, M. Waechter, M. Samejima, Y. Sugano, and Y. Matsushita, "Light structure from pin motion: Simple and accurate point light calibration for physics-based modeling," in *European Conference on Computer Vision (ECCV)*, 2018.

[14] H. Santo, M. Waechter, W.-Y. Lin, Y. Sugano, and Y. Matsushita, "Light structure from pin motion: Geometric point light source calibration," *International Journal of Computer Vision (IJCV)*, 2020.

[15] L. Ma, Y. Liu, J. Liu, *et al.*, "A fast led calibration method under near field lighting based on photometric stereo," *Optics and Lasers in Engineering*, vol. 147, p. 106 749, 2021.

[16] L. Ma, X. Liu, Y. Liu, X. Pei, and S. Guo, "Robust point light source calibration method for near-field photometric stereo using feature points selection," *Appl. Opt.*, vol. 62, no. 36, pp. 9512–9522, Dec. 2023.

[17] Y. Song, F. Elibol, M. She, D. Nakath, and K. Köser, *Light pose calibration for camera-light vision systems*, 2020.

[18] Y. Jiang, J. Tu, Y. Liu, *et al.*, "Gaussianshader: 3d gaussian splatting with shading functions for reflective surfaces," *arXiv preprint arXiv:2311.17977*, 2023.

[19] J. Gao, C. Gu, Y. Lin, *et al.*, "Relightable 3d gaussian: Real-time point cloud relighting with brdf decomposition and ray tracing," *arXiv:2311.16043*, 2023.

[20] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, 381–395, Jun. 1981.

[21] L. Ma, J. Liu, X. Pei, Y. Hu, and F. Sun, "Calibration of position and orientation for point light source synchronously with single image in photometric stereo," *Opt. Express*, vol. 27, no. 4, pp. 4024–4033, Feb. 2019.

[22] A. Ryer, *Light Measurement Handbook*. International Light, 1997.

[23] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, Jul. 2023.

[24] B. Mildenhall, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, and J. T. Barron, "NeRF in the dark: High dynamic range view synthesis from noisy raw images," *CVPR*, 2022.