

面向挑战性重建场景的相机参数 与3D高斯联合优化

贾贤志

重庆三峡学院计算机科学与工程学院, 重庆

收稿日期: 2026年2月9日; 录用日期: 2026年3月2日; 发布日期: 2026年3月10日

摘要

3D高斯泼溅(3D Gaussian Splatting, 3DGS)凭借其高效的实时渲染能力与快速收敛特性, 已成为新视角合成领域的重要技术路线。然而, 其重建质量高度依赖于精确的相机内参与外参标定, 这些参数传统上依赖运动恢复结构(SfM)流程(如COLMAP)预先估计。值得注意的是, 在低纹理、重复纹理或视角受限等挑战性场景中, SfM估计往往产生误差, 限制了重建精度。为此, 本文提出一种在训练阶段同步精细化相机参数与3D高斯表示的联合优化框架。具体而言, 我们通过偏移参数化策略增强优化稳定性, 采用对数尺度焦距表示改善数值条件, 并引入余弦衰减自适应正则化项来优化相机参数, 提升建模质量。

关键词

3D高斯泼溅, 相机优化, 新视角合成

Joint Optimization of Camera Parameters and 3D Gaussians for Challenging Reconstruction Scenarios

Xianzhi Jia

College of Computer Science and Engineering, Chongqing Three Gorges University, Chongqing

Received: February 9, 2026; accepted: March 2, 2026; published: March 10, 2026

Abstract

3D Gaussian Splatting (3DGS) has emerged as a prominent technique in novel view synthesis, distinguished by its efficient real-time rendering capabilities and rapid convergence properties. However,

its reconstruction quality heavily relies on accurate estimation of camera intrinsics and extrinsics, which are traditionally obtained through Structure-from-Motion (SfM) pipelines such as COLMAP. Notably, in challenging scenarios characterized by low-texture regions, repetitive patterns, or limited viewpoint distributions, SfM estimations often suffer from inaccuracies, thereby compromising reconstruction fidelity. To address this limitation, we propose a joint optimization framework that simultaneously refines camera parameters and 3D Gaussian representations during the training stage. Specifically, we employ offset parameterization to stabilize optimization, adopt logarithmic focal length representation for improved numerical conditioning, and introduce a cosine annealing adaptive regularization strategy to refine camera parameters and enhance modeling quality.

Keywords

3D Gaussian Splatting, Camera Optimization, Novel View Synthesis

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 背景

随着神经辐射场(NeRF)的引入,新视角合成领域经历了范式转变,NeRF 将场景编码为连续神经函数[1]。最近,3D 高斯泼溅(3DGS)通过将场景表示为显式各向异性 3D 高斯,彻底改变了这一领域,实现了高帧率下的实时真实感渲染。与隐式表示不同,3DGS 利用可微分光栅化进行高效优化,已成为需要交互式新视角合成的实际应用的首选方法[2]。

尽管具有令人印象深刻的能力,但 3DGS 关键依赖于精确的相机内参和外参,这些参数通常通过运动恢复结构(SfM)管道(如 COLMAP)获得。在具有挑战性的真实世界场景中,传统 SfM 方法经常失败[3],原因是:

- (1) 低纹理区域: 缺乏足够的独特特征进行可靠的特征匹配和跟踪。
- (2) 重复模式: 导致模糊对应关系,从而产生错误的姿态估计。
- (3) 受限视角分布: 缺乏足够的基线进行稳定的三角化和束调整。

这种校准误差会在渲染流程中传播,降低学习到的 3D 高斯质量,并最终损害新视角合成性能。基于 NeRF 的场景表示的最新研究表明,场景几何和相机参数的联合优化可以缓解这些问题,然而在 3DGS 背景下,类似的方法仍未得到充分探索。

针对这些问题,本文提出了一个联合优化相机参数和 3D 高斯表示的框架。主要技术包括:

- (1) 偏移参数化,将相机参数表示为从初始 SfM 估计中学习的偏移,确保优化稳定性。
- (2) 对数尺度焦距表示,改善数值条件。
- (3) 余弦衰减自适应正则化,动态调整正则化强度。

1.2. 相关工作

1.2.1. 3D 高斯泼溅及其扩展

3D 高斯泼溅将场景表示为由位置、不透明度、缩放、旋转和球谐系数参数化的各向异性 3D 高斯集合[4]。通过可微分光栅化,3DGS 实现了具有实时渲染能力的真实感新视角合成,在推理速度方面显著

优于隐式神经表示[5]。最近的扩展改进了原始公式，包括压缩技术、抗锯齿和动态场景建模[6]。然而，这些方法主要假设精确的相机校准，在优化过程中将相机参数视为固定。

1.2.2. 神经渲染中的相机姿态优化

基于 NeRF 的方法，认识到不准确的相机姿态会显著降低新视角合成质量，这激发了关于联合姿态 - 场景优化的广泛研究。光束法平差神经辐射场(BARF)在 NeRF 优化的同时引入了渐进姿态校正，证明了精确姿态对于高保真重建至关重要[7]。后续工作探索了稀疏控制点、空间自适应姿态细化和鲁棒优化策略来处理具有挑战性的初始化场景。虽然这些方法取得了令人印象深刻的结果，但它们主要关注隐式神经表示，与显式 3D 表示相比，隐式表示在训练和推理速度上较慢[8]。传统多视图立体(MVS)和运动恢复结构方法长期以来一直结合束调整进行联合姿态和几何细化。在可微分渲染的背景下，最近的工作探索了神经辐射场的相机优化和基于网格的表示。然而，3D 高斯泼溅的相机参数优化仍然相对未被探索，尽管它具有重要的实用意义和基于点渲染带来的独特挑战[9]。有效的参数化对于稳定的神经网络优化至关重要[10]。偏移参数化(可学习参数表示与初始估计的偏差)已在姿态图优化和神经网络训练中得到广泛采用。对于相机内参，对数空间参数化防止了不切实际的缩放并改善了数值条件。自适应正则化(正则化强度在优化期间变化)在计算机视觉任务中已被证明是有效的。基于调度的正则化通过动态调整约束强度防止过拟合并实现稳定收敛[11]。偏移参数化(将可学习参数表示为与初始估计的偏差)在姿态图优化和神经网络训练中已得到广泛应用。

2. 本文的方法

本文基于 3D 高斯泼溅提出了独特特性的联合相机-3D 优化统一框架，目的是在训练 3D 高斯的同时联合优化相机参数，以获得更精确的场景表示

2.1. 联合优化

- (1) 3D 高斯参数：位置、不透明度、缩放、旋转、球谐系数。
- (2) 相机参数：外参(旋转四元数 $\mathbf{q} \in \mathbb{R}^4$ 、平移 $\mathbf{t} \in \mathbb{R}^3$)和内参(焦距 f_x, f_y 、主点 c_x, c_y)。

将相机参数表示为 11 维向量：

$$\theta = [\mathbf{q}, \mathbf{t}, \delta_{f_x}, \delta_{f_y}, \delta_{c_x}, \delta_{c_y}] \in \mathbb{R}^{11}$$

为了实现稳定有效的联合优化，采用以下关键技术：

- (1) 偏移参数化：将相机参数表示为 COLMAP 初始估计的偏移，确保数值稳定性并增强可解释性。
- (2) 对数尺度焦距：在对数空间中参数化焦距以获得更好的数值条件。
- (3) 自适应正则化。

2.2. 偏移参数化

本文不直接优化原始相机参数，而是将它们参数化为初始 COLMAP 估计的偏移：

外参(直接优化，无偏移)：

- (1) 四元数旋转： $\mathbf{q} \in \mathbb{R}^4$ (单位四元数)。
- (2) 平移： $\mathbf{t} \in \mathbb{R}^3$ 。

内参(带偏移)：

对于焦距，我们采用对数尺度参数化：

$$f_x = f_x^{\text{init}} \cdot \exp(\delta_{f_x}), f_y = f_y^{\text{init}} \cdot \exp(\delta_{f_y})$$

对于主点，采用归一化偏移：

$$c_x = c_x^{\text{init}} + \delta_{c_x} \cdot \frac{W}{2}, \quad c_y = c_y^{\text{init}} + \delta_{c_y} \cdot \frac{H}{2}$$

其中 $f_x^{\text{init}}, f_y^{\text{init}}, c_x^{\text{init}}, c_y^{\text{init}}$ 是 COLMAP 的初始估计值， $\delta_{f_x}, \delta_{f_y}, \delta_{c_x}, \delta_{c_y}$ 是可学习的偏移参数， W 和 H 分别是图像的宽度和高度。

该公式确保：

- (1) 焦距在对数空间的数值稳定性。
- (2) 主点偏移的归一化尺度。
- (3) 可解释性：接近零的偏移表示对 COLMAP 的置信度。

偏移参数化的核心思想是将相机参数的优化问题转化为对初始估计的修正问题。这种参数化方式具有以下优势：(1) 减小优化空间的搜索范围，使优化过程更加稳定；(2) 保持对初始估计的可解释性，偏移量的大小直接反映了优化的程度；(3) 避免参数漂移，确保优化结果不会偏离合理的物理约束。

2.3. 自适应正则化

我们应用正则化以防止相机参数偏离初始估计太远。正则化损失定义为：

$$\mathcal{L}_{\text{reg}} = (\|\mathbf{q}\| - 1)^2 + \delta_{f_x}^2 + \delta_{f_y}^2 + \delta_{c_x}^2 + \delta_{c_y}^2$$

其中第一项 $(\|\mathbf{q}\| - 1)^2$ 是单位四元数约束，确保旋转四元数保持单位长度；其余项约束焦距和主点的偏移参数接近零。

总损失函数为：

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \lambda(t) \cdot \mathcal{L}_{\text{reg}}$$

其中 $\mathcal{L}_{\text{recon}}$ 是重建损失(L1 损失和 SSIM 损失的组合)， $\lambda(t)$ 是随时间变化的自适应正则化权重，采用余弦衰减策略：

$$\lambda(t) = \lambda_{\text{final}} + (\lambda_{\text{initial}} - \lambda_{\text{final}}) \cdot \frac{1 + \cos(\pi t/T)}{2}$$

其中 t 是当前迭代次数， T 是总迭代次数。在我们的实现中， $\lambda_{\text{initial}} = 0.01$ ， $\lambda_{\text{final}} = 0.001$ 。正则化权重初始较强以防止早期发散，然后逐渐衰减，允许训练进展时更大的灵活性。余弦衰减策略的设计基于以下观察：在训练初期，3D 高斯参数尚未稳定，此时需要较强的正则化约束来防止相机参数的剧烈变化；随着训练的进行，3D 高斯逐渐收敛，可以适当放宽正则化强度，允许相机参数进行更精细的调整。这种自适应策略平衡了优化稳定性和最终精度。

2.4. 实现细节

我们的实现使用 Adam 优化器，初始学习率为 $1e-4$ ，通过余弦退火调度降至 $1e-5$ 。总训练迭代次数为 15000 次，分别在 7000 次和 15000 次迭代时进行评估。相机参数采用 11 维向量表示，包含外参(四元数旋转和平移)和内参(焦距和主点的偏移)。正则化权重从初始值 0.01 按余弦衰减至 0.001。

3. 实验设置

3.1. 数据集

主要数据集：杏果实数据集(自制) - 使用手机拍摄的真实世界数据集，包含低纹理和复杂光照等具有

挑战性的场景。

模拟真实场景误差：**Garden** 数据集(扰动实验) - 选自 Mip-NeRF 360 数据集的户外场景，包含丰富的植被和复杂几何结构。为了系统评估相机优化框架对初始 COLMAP 估计误差的校正能力，我们对相机参数进行了扰动处理。具体而言，考虑到移动设备拍摄场景中的典型误差来源(如运动模糊、手持抖动、低纹理区域等)，我们向 COLMAP 输出的相机参数添加了温和级别的高斯噪声：

1. ****外参扰动****: 向平移向量 $\mathbf{t} \in \mathbb{R}^3$ 添加噪声 $\epsilon \sim \mathcal{N}(0, (0.01m)^2 \mathbf{I}_3)$ 。
2. ****内参扰动****: 焦距参数按 $f' = f \cdot (1 + \epsilon_f)$ 缩放，其中 $\epsilon_f \sim \mathcal{N}(0, 0.005^2)$ 。

所有实验训练 15,000 次迭代，在迭代 7000 和 15,000 时在训练集上评估指标，在单个 GPU 上使用默认 3DGS 超参数进行训练。

3.2. 评估指标

使用三个标准指标来评估重建质量：

PSNR(峰值信噪比)：衡量重建图像与真实图像之间的像素级差异，定义为：

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right)$$

其中 MAX_I 是图像的最大可能像素值， MSE 是均方误差。

SSIM(结构相似性指数)：评估结构信息的保留程度，定义为：

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

其中 μ_x, μ_y 是均值， σ_x^2, σ_y^2 是方差， σ_{xy} 是协方差， c_1, c_2 是稳定常数。

LPIPS (学习感知图像块相似性)：使用预训练深度网络提取特征并计算感知距离，定义为：

$$\text{LPIPS}(x, y) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot (\hat{x}_{hw}^l - \hat{y}_{hw}^l) \right\|_2^2$$

其中 l 表示网络层， \hat{x}_{hw}^l 是第 l 层的特征图， w_l 是学习的通道权重。

4. 实验结果

4.1. 杏果实数据集

如图 1 所示，基线 3DGS (原始 3DGS) 与本文方法的训练 L1 损失曲线对比。蓝色实线表示基线 3DGS，红色实线表示本文方法，本文的相机优化方法的训练 L1 损失降低 47.47%，说明重建质量有所提升。

如表 1 所示，本文的相机优化方法在所有指标上均优于基线 3DGS，PSNR 提升至 37.22，SSIM 提升至 0.970，LPIPS 降低至 0.157。

图 2 展示了在杏果实数据集上基线 3DGS 与我们的相机优化方法的渲染结果对比。左边为基线 3DGS 渲染结果，右边为本文的相机优化方法渲染结果，可以看出本文方法在细节保留和整体视觉质量上均有改善。

4.2. Garden 数据集

如表 2 所示，Garden 数据集上，本文的相机优化方法在所有指标上略优于基线 3DGS，PSNR 提升至 24.19，SSIM 提升至 0.631，LPIPS 降低至 0.357。这说明在存在相机参数扰动(1.0 cm 平移误差 + 0.5% 焦距误差)的情况下，本文方法能够有效校正初始估计误差，实现高质量的 3D 重建。这证明了联合优化相

机参数与 3D 高斯表示的框架在实际标定不精确场景中的可行性。

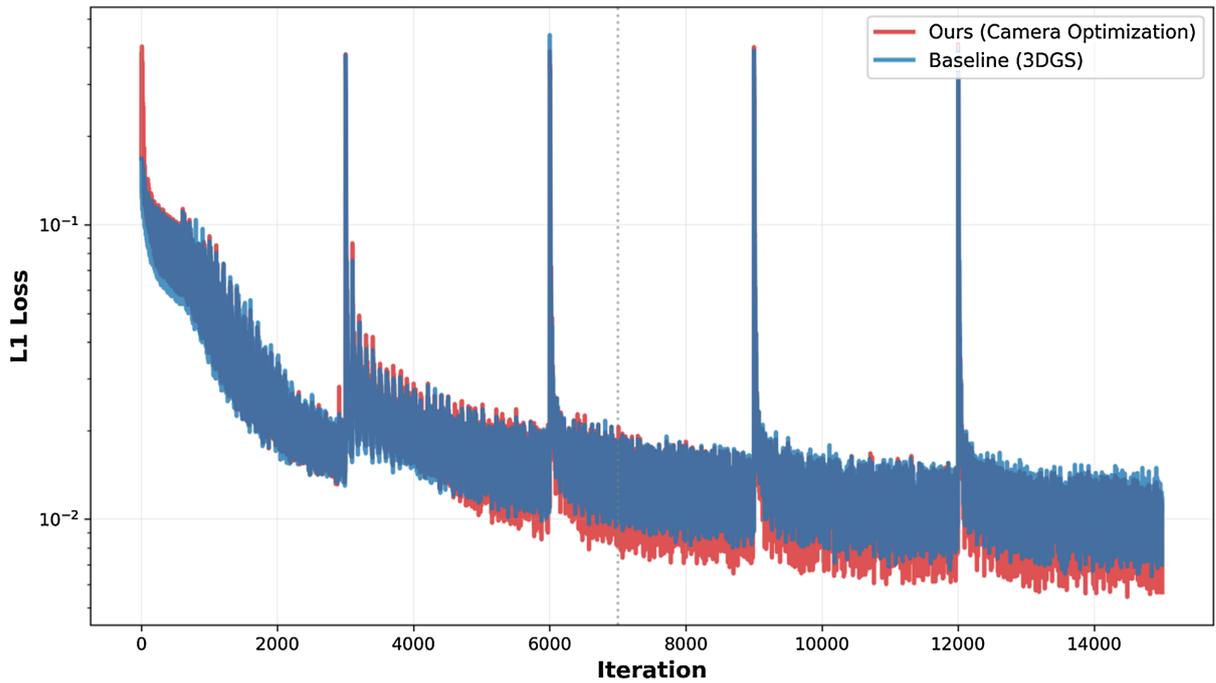


Figure 1. L1 loss comparison of baseline 3DGS vs. our method on apricot fruit dataset

图 1. 杏果实数据集基线 3DGS 与本文方法的 L1Loss 的对比

Table 1. Comparative analysis of training set results on apricot fruit dataset

表 1. 杏果实数据集训练集结果对比分析

方法	PSNR↑	SSIM↑	LPIPS↓
3DGS(基线)	36.97	0.969	0.161
本文(相机优化)	37.22	0.970	0.157



Figure 2. Comparison between baseline 3DGS and the proposed method

图 2. 基线 3DGS 与本文方法对比

Table 2. Comparative analysis of training set results on the garden dataset
表 2. Garden 数据集训练集结果对比分析

方法	PSNR↑	SSIM↑	LPIPS↓
3DGS (基线)	23.84	0.588	0.373
本文(相机优化)	24.19	0.631	0.357

4.3. 讨论

真实世界场景中的有效性：在自制杏果实数据集以及带有扰动的 garden 数据集上的实验表明，相机参数优化能够显著改善重建质量，所有评估指标均有提升。这些改善验证了联合优化框架在真实世界应用中的有效性。在手机拍摄的数据集中，COLMAP 的估计可能因运动模糊、低纹理或复杂光照而不准确。本文方法通过偏移参数化和自适应正则化，能够在优化 3D 高斯的同时微调相机参数，获得更精确的场景表示。

适用场景：本方法特别适用于以下具有挑战性的真实世界场景：移动设备拍摄(存在运动模糊和相机抖动)、低纹理环境、重复模式场景，以及视角受限的数据集。在这些情况下，COLMAP 的初始估计往往不够精确，而相机优化能够有效校正误差。

计算权衡：相机优化会增加每次迭代的计算时间，但考虑到重建质量的全面提升，这种额外的计算开销是值得的。对于对重建质量要求高的应用，这种权衡是可接受的。

5. 结论

本文提出了一个统一的框架，用于在训练期间联合优化相机参数和 3D 高斯表示。通过采用偏移参数化、对数尺度焦距表示、自适应余弦衰减正则化，来对相机参数进行优化。本文方法在具有挑战性的真实世界数据上相比基线 3DGS 取得了显著改进。与基线 3DGS 相比，本文方法使 PSNR 提高了 0.25 db (36.97→37.22)，SSIM 提高了 0.001 (0.969→0.970)，LPIPS 降低了 0.004 (0.161→0.157)，训练损失降低了 45.44%。这些结果证明了对于难以获得精确相机校准的场景，联合相机 - 几何优化既实用又有效。本工作强调了在实际 3DGS 应用中考虑相机参数不确定性的重要性，特别是对于移动设备拍摄以及表现出低纹理、重复模式或受限视角分布的环境。所提出的优化策略能够在不修改核心 3DGS 渲染管道的情况下实现鲁棒的场景重建，使本文方法可直接应用于现有实现。

局限性与未来工作。虽然本文方法取得了显著成果，但仍有一些方向值得进一步研究：

- (1) 进一步实验：验证三个模块对性能的影响。
- (2) 动态场景：扩展到时变场景和基于视频的新视角合成。
- (3) 高级相机模型：集成径向畸变、偏斜等其他内参以实现更全面的建模。
- (4) 自适应优化：开发自动检测机制，判断何时相机优化有益以及校准中的不确定性。

参考文献

- [1] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R. and Ng, R. (2020) NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 405-421. https://doi.org/10.1007/978-3-030-58452-8_24
- [2] Kerbl, B., Kopanas, G., Leimkuehler, T. and Drettakis, G. (2023) 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, **42**, 1-14. <https://doi.org/10.1145/3592433>
- [3] Schonberger, J.L. and Frahm, J. (2016) Structure-from-Motion Revisited. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 4104-4113. <https://doi.org/10.1109/cvpr.2016.445>

-
- [4] Yu, Z., Chen, A., Huang, B., Sattler, T. and Geiger, A. (2024) Mip-Splatting: Alias-Free 3D Gaussian Splatting. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 17-21 June 2024, 19447-19456. <https://doi.org/10.1109/CVPR52733.2024.01839>
- [5] Liu, S., Yang, M., Xing, T. and Yang, R. (2025) A Survey of 3D Reconstruction: The Evolution from Multi-View Geometry to NeRF and 3DGS. *Sensors*, **25**, Article 5748. <https://doi.org/10.3390/s25185748>
- [6] Luiten, J., Kopanas, G., Leibe, B. and Ramanan, D. (2024) Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis. 2024 *International Conference on 3D Vision (3DV)*, Davos, 18-21 March 2024, 800-809. <https://doi.org/10.1109/3DV62453.2024.00044>
- [7] Lin, C.H., Ma, W.C., Torralba, A. and Lucey, S. (2021) BARF: Bundle-Adjusting Neural Radiance Fields. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 11-17 October 2021, 5741-5751. <https://doi.org/10.1109/ICCV48922.2021.00569>
- [8] Ku, K., Jubery, T.Z., Rodriguez, E., Balu, A., Sarkar, S., Krishnamurthy, A. and Ganapathysubramanian, B. (2025) SC-NeRF: NeRF-Based Point Cloud Reconstruction using a Stationary Camera for Agricultural Applications. 2025 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 11-15 June 2025, 5472-5481. <https://doi.org/10.1109/CVPRW67362.2025.00544>
- [9] Triggs, B., McLauchlan, P.F., Hartley, R.I. and Fitzgibbon, A.W. (2000) Bundle Adjustment—A Modern Synthesis. In: Triggs, B., Zisserman, A. and Szeliski, R., Eds., *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 298-372. https://doi.org/10.1007/3-540-44480-7_21
- [10] Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J. and Zhou, J. (2021) NerfingMVS: Guided Optimization of Neural Radiance Fields for Indoor Multi-View Stereo. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Montreal, 11-17 October 2021, 5610-5619. <https://doi.org/10.1109/ICCV48922.2021.00556>
- [11] Park, K., Henzler, P., Mildenhall, B., Barron, J.T. and Martin-Brualla, R. (2023) CamP: Camera Preconditioning for Neural Radiance Fields. *ACM Transactions on Graphics (TOG)*, **42**, 1-11. <https://doi.org/10.1145/3618401>