Neural 3D Reconstruction of Planetary Surfaces from Descent-Phase Wide-Angle Imagery

Anonymous CVPR submission

Paper ID 48

Abstract

Digital elevation modeling of planetary surfaces is essential for studying past and ongoing geological processes. Wide-angle imagery acquired during spacecraft descent promises to offer a low-cost option for high-resolution terrain reconstruction. However, accurate 3D reconstruction from such imagery is challenging due to strong radial distortion, and limited parallax from vertically descending, predominantly nadir-facing cameras. Conventional multi-view stereo exhibits limited depth range and reduced fidelity under these conditions, and also lacks domain-specific priors. We present the first study of modern, neural reconstruction methods for planetary descent imaging. We also develop a novel approach that incorporates an explicit neural height field representation, which provides a strong prior since planetary surfaces are generally continuous, smooth, solid, and free from floating objects. This study demonstrates that neural approaches offer a strong and competitive alternative to traditional multi-view stereo (MVS) methods. Experiments on simulated descent sequences over high-fidelity lunar and Mars terrains demonstrate that the proposed approach achieves increased spatial coverage while maintaining satisfactory estimation accuracy.

1. Introduction

Planetary exploration seeks to characterize the planets and moons of the solar system to understand their histories, current conditions, and the prospect of past or present extraterrestrial life. A critical step in this effort is the detailed analysis of planetary surfaces, as surface features can reveal past or present geological and hydrological activity [29]. Surface reconstruction or Digital Elevation Models (DEMs) play a central role in this process [13], enabling scientists to construct accurate 3D models of surface elevation from remote sensing data acquired by orbiters, surface robots, or landers. The primary approach for reconstructing planetary surfaces relies on stereo pairs captured from orbit [19, 25]. Using this method, both the Moon and Mars have been mapped globally in 3D. However, this requires very costly cameras and orbital platforms that can prohibit its use on smaller missions or distant bodies, or new locations.

Descent imagery captured by landers can be used as a low-cost alternative to construct DEMs [6, 29], particularly as new exploration modes such as drones and small descent craft are being developed. Additionally, this method bridges the resolution gap between the orbiter/probe and the lander's local imaging—the orbiter is far and only resolves at lower resolution, while the lander gives fine detail but only of its immediate surroundings. These findings are scientifically valuable because they give a regional context for the local science that the lander will perform, such as imaging, collecting samples, and geochemical experiments. Furthermore, this reconstruction technique enables using archival data from pre-landing imaging campaigns [8, 20] to reconstruct DEMs for further analysis.

A compelling example comes from data of Titan by the Cassini-Huygens mission [29]. River valleys have been detected at nearly all latitudes on Titan, revealing a surprisingly Earth-like landscape shaped not by water, but by liquid methane. To better understand the geomorphological processes responsible for carving these landscapes, researchers have used MicMac [23] structure-from-motion (SfM) [35] and multi-view stereo (MVS) [28] to reconstruct DEMs from images.

Classical stereo-based pipelines often yield sparse or noisy reconstructions and cannot capture view-dependent effects such as reflections (indeed, such effects typically reduce their accuracy). They are also sensitive to parameters and image coverage. While modern, integrated implementations such as COLMAP [26] and Agisoft Metashape [1] have made these pipelines easier to use on high-quality imagery with large overlaps. Extreme imaging conditions, such as wide-angle images of vertical descents, can result in incomplete 3D reconstructions.

Modern NeRF-based [17] approaches instead learn a continuous volumetric representation, producing denser, smoother, and more accurate reconstructions that also cap-

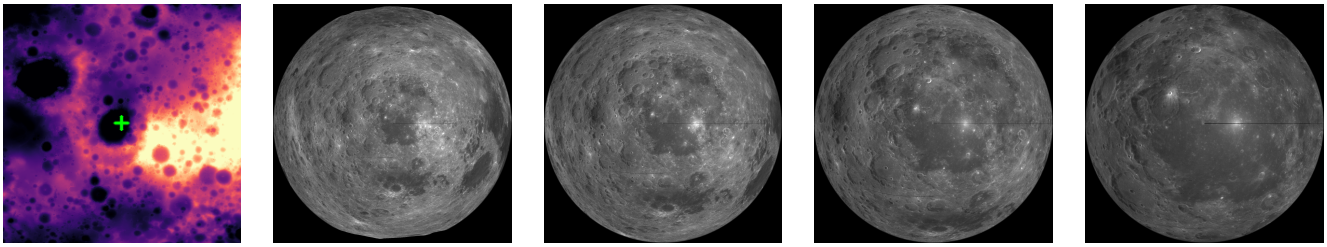


Figure 1. **First:** Height map and x-y camera center (green cross) for the lunar scene. **Remaining:** Example simulated descent views toward the center of the lunar scene, including the top-most and bottom-most views.

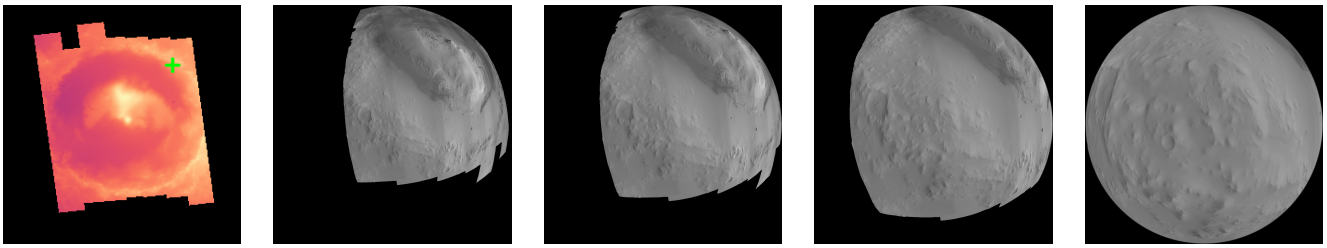


Figure 2. **First:**Height map and x-y camera center (green cross) for the Mars scene. **Remaining:** Example simulated descent views toward the top-right corner, including the top-most and bottom-most views.

ture viewpoint-dependent effects. NeRF replaces traditional MVS by optimizing a neural radiance field that maps 3D coordinates and viewing directions to color and density, taking posed images as input to synthesize novel views and reconstruct detailed scene geometry. There are several extensions to NeRF to support unbounded scenes [3, 4].

However, scenarios involving wide-angle (fisheye) images and cameras descending vertically toward flat surfaces also present challenges for NeRF-based methods. Wide-angle lenses introduce strong radial distortion that must be properly modeled for accurate ray generation. Predominantly downward-facing cameras imaging a largely planar surface also result in limited viewpoint diversity, reduced parallax, and depth ambiguity [17]. This makes it difficult for NeRF to infer depth and fine geometric details reliably.

In this study, we apply NeRF-based methods to construct digital elevation models (DEMs) of planetary surfaces using wide-angle descent imagery for the first time. Planetary surfaces are generally smooth and continuous, devoid of floating debris, and illuminated primarily by distant sunlight. To avoid limitations of existing NeRF models and to use domain-specific geometric and photometric priors, we propose a novel NeRF-based framework. It integrates an intermediate HeightField representation that maps 2D coordinates to elevation, enabling high-resolution planetary surface height-map generation directly from fisheye imagery. Instead of using a direction-encoded color MLP, we adopt the physically-based Hapke reflectance model [24] to achieve photometrically-consistent and realistic appearance under extraterrestrial lighting conditions. In addition, we incorporate regularization losses to mitigate geometric ambiguity and stabilize training under limited viewpoint diversity.

Our key contributions are:

- Quantitatively evaluating NeRF-based models for Digital Elevation Modeling of planetary surfaces using wide-angle descent imagery for the first time, and showing they achieve higher coverage than Metashape, while preserving satisfactory accuracy.
- A novel reconstruction framework that couples NeRF and heightmap representations to achieve more physically-plausible reconstructions than NeRF
- An angle-aware distortion loss to ensure smooth surface reconstruction while being flexible about the edges
- Analyzing the effectiveness of incorporating a physically-based shading model specific to planetary surfaces, instead of implicitly learning direction-dependent color.

We perform a comprehensive evaluation with simulated Fisheye imagery on existing high-resolution moon [25] and Mars [19] 3D reconstructions. Our proposed approach achieves a balanced trade-off between spatial coverage and elevation accuracy in the generated DEMs, outperforming both Metashape [1] and Nerfacto [33] in coverage at 0.1 relative error across both datasets. Moreover, it demonstrates a clear improvement over Nerfacto [33] for this task, achieving lower relative and absolute elevation differences across both datasets.

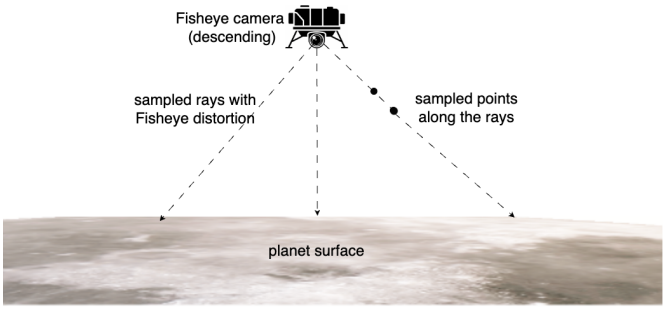


Figure 3. **Fisheye Ray Casting:** maps each pixel through a wide-angle lens to a 3D ray using the lens's radial distortion model.

2. Related Work

2.1. Classical 3D Reconstruction

3D reconstruction methods can be divided into active and passive techniques [14, 38]. Active methods emit controlled energy, such as light or laser, and analyze its interaction with surfaces; examples include structured light [22], LiDAR [27], laser scanning [11, 16], CT scanning [10], and photometric stereo [37]. Although often precise and robust, they require specialized hardware and controlled conditions. Within passive reconstruction, traditional pipelines typically combine Structure-from-Motion (SfM) [9, 21, 26, 30, 31, 35] and Multi-View Stereo (MVS) [28, 32]. Structure-from-Motion (SfM) estimates camera poses and intrinsic parameters while simultaneously reconstructing a sparse 3D point cloud from multiple overlapping images by detecting and matching visual features across views and enforcing geometric constraints such as epipolar consistency [9, 21, 30, 31, 35]. The result is a globally optimized sparse representation of the scene and calibrated camera parameters. Building on this, Multi-View Stereo (MVS) performs dense reconstruction by estimating per-pixel depth maps or dense point clouds using photometric consistency across multiple calibrated images [28, 32]. MVS leverages the camera parameters obtained from SfM to recover detailed surface geometry, producing a dense 3D model suitable for surface reconstruction and DTM generation. A practical implementation of this pipeline is Agisoft Metashape, which integrates image alignment, bundle adjustment, dense reconstruction, and mesh generation into an end-to-end framework.

2.2. 3D Reconstruction with Machine Learning

Recent methods increasingly use deep learning to improve robustness and completeness. These approaches can be broadly divided into explicit representations, such as point clouds, voxels, and meshes, and implicit representations, particularly neural implicit representations [38]. Among the latter, methods based on Neural Radiance Fields (NeRF) [17] have demonstrated high-fidelity novel view synthesis and accurate geometry reconstruction by representing scenes as continuous volumetric radiance fields, which are learned directly from multi-view images and optimized to model both color and density at every point in 3D space. Mip-NeRF [3] reduces aliasing by modeling conical frustums, while Mip-NeRF 360 [4] extends this to 360° unbounded scenes using a non-linear scene parameterization and a distortion-based regularizer that concentrates density along rays, reducing floaters for more realistic reconstructions. However, these methods still face challenges in reconstructing thin surfaces. NeRF-based methods have been applied on Earth for surface elevation modeling using images captured while moving across the surface, rather than vertically descending towards the surface [2, 7]. Alternatively, NeUS [36] learns an implicit surface representation, instead of a density field, from multiview 2D images. However, this method cannot handle fisheye images. More recently, explicit point-based neural representations such as 3D Gaussian Splatting [15, 39] introduce anisotropic Gaussian primitives for scene representation, enabling real-time rendering while maintaining competitive reconstruction accuracy. However, NeRF's volumetric representation allows it to better capture thin structures because the density field is continuous [17].

2.3. Previous Work in Planetary Elevation Modeling using Fisheye Descent Images

Although 3D reconstruction has recently seen significant advances with the advent of NeRF-based methods, space science continues to rely primarily on classical methods to reconstruct DEMs. Commercial software such as SOCET SET has been used to generate DEMs from descent images [29] taken by the Huygens probe on Titan, enabling photogrammetric adjustment of overlapping image pairs to recover camera orientations and probe positions. More recently, IPGP-DMT leverages the open-source photogrammetry software MicMac [23] with Shape-from-Motion [34] to produce higher-quality DEMs. MicMac's reconstruction quality is highly sensitive to parameter settings and image coverage, often yielding noisy or incomplete DEMs. Furthermore, surface points were generated by triangulating stereo matches between two images of different altitudes [6] and evaluated by measuring errors at varying radii from the image centre. This approach is limited by its reliance on rectified images and inherently sparse coverage, often producing holes or artifacts in the reconstructed surface—issues that are not captured by the evaluation metrics. Recently, Metashape [1], which automates the entire SfM–MVS workflow, supporting end-to-end 3D reconstruction from image alignment through textured mesh generation, and is widely used in planetary science. However, Metashape depends on high-quality, well-overlapping images, which can lead to insufficient coverage in DEMs. Al-

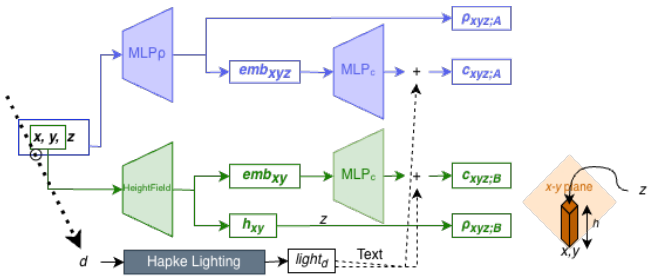


Figure 4. Our method uses two scene representations that each define the density and color of sampled points $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ along the ray. In the **NeRF representation (purple)**, an $\mathrm{MLP_d}$ maps $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ into density $\rho_{\mathbf{xyz;A}}$ and density embedding $\mathbf{emb_{xyz}}$. $\mathrm{MLP_{rgb}}$ maps $\mathbf{emb_{xyz}}$ into colour, which is then multiplied by ray direction dependent lighting ($\mathbf{light_d}$) to get final colour: $\mathbf{rgb_{xyz;A}}$. In the **heightmap representation (green)**, given our world frame is defined so the surface is perpendicular to the z-axis, the $(\mathbf{x}, \mathbf{y})$ is passed onto HeightField to predict heights $\mathbf{h_{xy}}$ and the density embedding of the $(\mathbf{x}, \mathbf{y})$ column: $\mathbf{emb_{xy}}$. Then, $\rho_{\mathbf{xy}} = \rho_{\mathbf{xy}}^{\text{const}}$ below the height, smoothly decaying to zero via a sigmoid above it. $\mathrm{MLP_{rgb}}$ maps $\mathbf{emb_{xy}}$ into colour of the column along z; which is then multiplied by $\mathbf{light_d}$ to get $\mathbf{rgb_{xyz;B}}$ for each z along $(\mathbf{x}, \mathbf{y})$ column.

though NerF has been applied to surface reconstruction of the lunar surface using stereo pairs [12], fisheye descent image -based surface reconstruction is unexplored.

3. Methodology

In this study, we evaluate the effectiveness of NeRF-based approaches to reconstruct DEMs from wide-angle descent imagery. We propose an extended NeRF approach specifically for this task to reconstruct smooth planetary surfaces while reducing the limitations of Fisheye distortion and less viewpoint diversity.

Our method reconstructs a high-resolution heightmap of the planetary surface from a fisheye image sequence, $\mathbf{I_n}$; $\mathbf{n} = \mathbf{1}, \ldots, \mathbf{N}_{\text{img}}$, acquired during the descent of the spacecraft for camera extrinsics: $[\mathbf{R}_n \mid \mathbf{t}_n]$ and intrinsics including distortion parameters: $\{\mathbf{K}_n, \boldsymbol{\xi}_n\}$. A binary mask $\mathbf{m_n}$ is provided as additional input to exclude the camera housing and any visible spacecraft structures.

3.1. Background: Neural Radiance Fields (NeRF)

NeRF [17, 33] represents a scene as a continuous volumetric field using an MLP. To render a pixel, a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ is cast from the camera center $\mathbf{o}$ through the pixel along direction $\mathbf{d}$. Distances $\mathbf{t}$ (sorted between near plane t_{near} and far plane t_{far}) are sampled along the ray. For each point $\mathbf{x} = \mathbf{r}(t_k)$, a positional encoding is applied. The encoded points are fed into an MLP (which also takes view direction, omitted for brevity) to predict density τ_k and color $\mathbf{c}_k$:

$$[\tau_\mathbf{k}, \mathbf{c}_k] = \mathrm{MLP}(\gamma(\mathbf{r}(\mathbf{t_k})); \Theta). \tag{1}$$

The pixel color is computed via numerical quadrature of the volume rendering integral:

$$\mathbf{C}(\mathbf{r}; \Theta, \mathbf{t}) = \sum_k T_k(1 - \exp(-\tau_\mathbf{k}\Delta_k))\mathbf{c}_k, \tag{2}$$

$$T_k = \exp\left(-\sum_{k'<k} \tau_{\mathbf{k}'}\Delta_{k'}\right), \tag{3}$$

with $\Delta_k = \mathbf{t_{k+1}} - \mathbf{t_k}$. Training minimizes the error between rendered and ground-truth pixel colors.

In order to make the model capable of having unbounded scenes, a scene contraction function transforms the Euclidean space into a bounded domain before mapping each point $(\mathbf{x})$ into density(τ_k) as shown below [4, 33]:

$$\text{contract}(\mathbf{x}) = \begin{cases} \mathbf{x} & \|\mathbf{x}\| \leq 1 \\ \left(2 - \frac{1}{\|\mathbf{x}\|}\right)\frac{\mathbf{x}}{\|\mathbf{x}\|} & \|\mathbf{x}\| > 1 \end{cases} \tag{4}$$

Second, a distortion loss regularizes the ray weight distribution to be compact, suppressing floaters and discouraging background collapse [4, 33]:

$$\mathcal{L}_{\text{dist}} = \iint w(\mathbf{u})w(\mathbf{v})|\mathbf{u} - \mathbf{v}|\,\mathbf{du}\,\mathbf{dv}; \tag{5}$$

where each integral should be interpreted as a path integral along a ray.

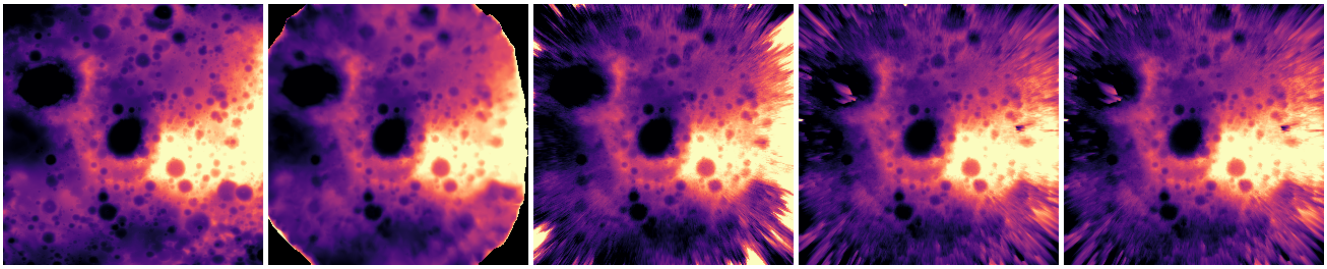
Moreover, we use a multi-resolution hash encoding for positional encoding, enabling fast training and high-fidelity 3D reconstruction with minimal parameter count [18].

3.2. Our Heightmap Reconstruction Method

Our study builds on NeRF and develops a novel approach specialized for the task of reconstructing DEMs from wide-angle descent images. Our objective is to learn a continuous function HeightField that maps planar surface coordinates $(\mathbf{x}, \mathbf{y})$ to their corresponding elevation values. The input data are fisheye RGB descent images, along with their camera intrinsics and extrinsics.

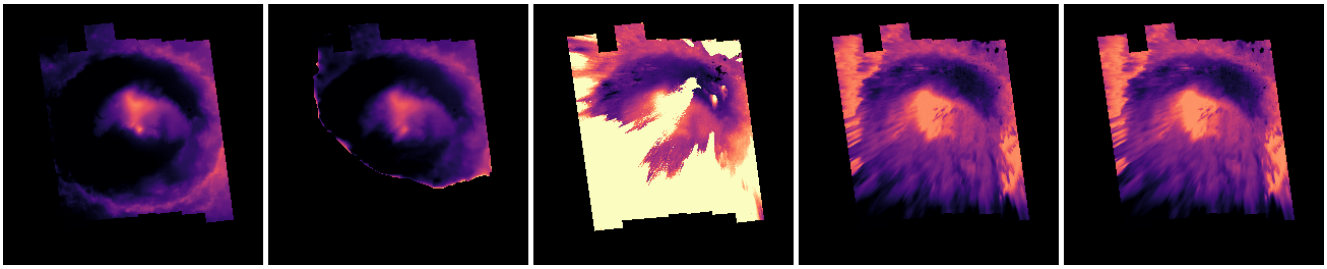
Scene representation. Our approach simultaneously optimizes two representations of the planetary surface. The first is a standard NeRF representation (Sec. 3.1), i.e. a neural network maps 3D coordinates to density. The second is the function HeightField that maps 2D coordinates in a ground plane to elevations. These two representations have complementary properties. NeRF is less prone to local optima during optimization, while the heightmap encodes the prior knowledge that planetary surfaces are typically solid and lacking significant overhangs, avoiding pathological solutions with floating geometry. Importantly, we can also directly read off elevations from HeightField to define a DEM.

In the NeRF representation (purple in Figure 4), the coordinates of a given position $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ are passed through a



GT **Metashape** **Nerfacto** **ours - NeRF representation** **ours-heightmap representation**

Figure 5. Qualitative comparison of DEM reconstructions from simulated lunar fisheye descent imagery against a high-resolution mesh ground truth. The height is calculated from the top-most camera. Our method achieves improved surface coverage while maintaining satisfactory elevation accuracy.



GT **Metashape** **Nerfacto** **ours-NeRF representation** **ours-heightmap representation**

Figure 6. Qualitative comparison of reconstructed DEMs from simulated Mars fisheye descent imagery, using a high-resolution mesh as ground truth. Our method achieves superior surface coverage while maintaining satisfactory elevation accuracy. The height is calculated from the top-most camera. Notably, Nerfacto suffers from floating artifacts, whereas our approach effectively suppresses these floaters, resulting in a smoother and more coherent surface reconstruction.

hash embedding then into MLP_ρ to predict density $\rho_{\mathbf{xyz};\mathrm{A}}$ and density embedding of the point $\mathbf{emb}_{\mathbf{xyz}}$. Then, density embedding is passed into MLP_c to predict viewing direction independent color of the point, which is then multiplied by a coefficient $\mathrm{light}_\mathrm{d}$ given by the Hapke shading model [24] to get $\mathbf{c}_{\mathbf{xyz};\mathbf{A}}$.

In the heightmap representation (green in Figure 4), assuming that surface is roughly perpendicular to the z-axis, the $(\mathbf{x}, \mathbf{y})$ coordinates of the point are passed onto our $\mathrm{HeightField}$ to predict the height of the surface $\mathbf{h}_{\mathbf{xy}}$ and the density embedding of the x,y column:$\mathbf{emb}_{\mathbf{xy}}$. The raw height output by the MLP is scaled and bounded by applying a scaled hyperbolic tangent. Then density $\rho_{\mathbf{xyz};\mathbf{B}}$ is calculated as:

$$\rho_{\mathbf{xyz};\mathbf{B}} = k_2 \,\mathrm{sigmoid}\left(k_1\left(\mathbf{h}_{\mathbf{xy}} - \mathbf{z}\right)\right). \tag{6}$$

Then $\mathbf{emb}_{\mathbf{xy}}$ is passed onto MLP_c to predict the constant color of the column; which is then multiplied by Hapke lighting $\mathrm{light}_\mathrm{d}$ to get colour of the point: $\mathbf{c}_{\mathbf{xyz};\mathbf{B}}$.

Optimization. Both representations define a density and color at each point in 3D space, and we can render them to pixels following Eqs. 2–3. Similar to NeRF, we minimize the difference between the rendered reconstruction and the original descent imagery, on minibatches of randomly sampled pixels. We also encourage the NeRF and heightmap to express similar geometry, allowing the heightmap to regularize the NeRF and distill geometric information from it.

For each input wide-angle image $\mathbf{I}_\mathbf{n}$, $\mathbf{n} = \mathbf{1}, \ldots, \mathbf{N}_{\text{img}}$, we cast rays $\mathbf{r}_n$ into the scene, accounting for the *fisheye distortion* as shown in Figure 3, considering only pixels marked as valid by $\mathbf{m}_\mathbf{n}$. The ray direction (d) is computed by mapping the pixel to the normalized camera frame using fisheye intrinsics and distortion: $\tilde{\mathbf{d}}_n = f_{\text{fisheye}}(u, v; \mathbf{K}_n, \boldsymbol{\xi}_n)$, and then rotating to world coordinates using the extrinsic rotation $\mathbf{R}_n$: $\mathbf{d}_n = \mathbf{R}_n \tilde{\mathbf{d}}_n$. We can then sample points x along the rays, denoting each sampled point as $(\mathbf{x}, \mathbf{y}, \mathbf{z})$. Sampled points are spatially distorted as described in Equation 5 to allow capturing distant surface points, and $\mathbf{z}$ is bounded by approximate z_{min}, z_{max} values depending on the planetary curvature and topography.

We first define separate RGB loss terms for the rendering

of each scene representation, as follows:

$$\mathcal{L}_{c;A/B} = \lambda_{c,A/B}\,\mathrm{MSE}(\mathbf{c}_{\mathbf{u},\mathbf{v}}, \mathbf{c}_{\mathbf{u},\mathbf{v};\mathbf{A}/\mathbf{B}}) \tag{7}$$

Secondly, we incorporate a distortion loss to regularize spatial inconsistencies in the reconstructed points (floaters) in the NeRF pipeline, as shown in Equation 5 [4]. When applying this to fisheye images, the central pixels have a small angular spread, leading to small 3D sampling intervals and therefore a naturally lower distortion loss contribution. Edge pixels in a fisheye image are spatially compressed but correspond to a large angular spread, producing elongated sampling intervals along rays that can disproportionately amplify the distortion loss and introduce reconstruction errors. Replacing the standard distance-based distortion loss (Eq. 5) with an angle-based formulation mitigates this issue, leading to more stable and accurate results. Our refined distortion loss is as follows:

$$\mathcal{L}_{dist;A} = \lambda_{dist}\cos\theta_{\mathbf{d}} \iint \mathbf{w}(\mathbf{u})\mathbf{w}(\mathbf{v})|\mathbf{u}-\mathbf{v}|\,\mathbf{du}\,\mathbf{dv} \tag{8}$$

In order to transfer geometric knowledge from the NeRF to the heightmap, we define a grid at the height ($\mathbf{h_k}$) of the highest camera, covering the required spatial region. Let $\mathbf{q}_{ij}$ denote the sampled point in the $(\mathbf{i},\mathbf{j})$-th grid cell. We perform stratified sampling within each cell to obtain a ray origin $\mathbf{o}_{ij}$, and cast a vertical ray along the z-axis with direction $\mathbf{d} = (0, 0, 1)$.

Along each ray, we sample points and compute the expected depth $\delta_{\mathbf{ij}}$ using the NeRF MLP_ρ, and the corresponding height using the HeightField representation:

$$\delta_{\mathbf{ij}} = \mathbb{E}_{z\sim\mathrm{Ray}(\mathbf{o}_{ij},\mathbf{d})}\left[\mathrm{MLP}_\rho(\mathbf{o}_{ij} + z\mathbf{d})\right] \tag{9}$$

$$h_{ij} = \mathrm{HeightField}(\mathbf{o}_{ij}). \tag{10}$$

We define a loss term as follows:

$$\mathcal{L}_{height;AB} = \lambda_{height}\,\mathrm{L1}(\mathbf{h_k} + \delta_{\mathbf{ij}}, \mathbf{h_{ij}}) \tag{11}$$

We use L1 loss here to reduce the effect of outliers.

We train the parameters of HeightField,$\mathrm{MLP_d}$, and $\mathrm{MLP_c}$ to minimize the loss function:

$$\mathcal{L} = \mathcal{L}_{c;A} + \mathcal{L}_{c;B} + \mathcal{L}_{dist;B} + \mathcal{L}_{height;AB} \tag{12}$$

In summary, $\mathcal{L}_{\mathrm{rgb;A}}$ and $\mathcal{L}_{\mathrm{rgb;B}}$ guide -learning from image data, $\mathcal{L}_{\mathrm{dist;B}}$ enforces surface smoothness, and $\mathcal{L}_{\mathrm{height;AB}}$ helps the HeightField capture geometric information through MLP_ρ.

4. Experiments

We evaluate classical MVS, standard NeRF, and our specialized pipeline on two realistic datasets. Our hyperparameters are provided in the supplementary material.

Dataset	Model	AED (m) ↓	RED (m) ↓	Coverage@0.1 (%) ↑
Lunar	Metashape	196.49	23.05	92.30
	Nerfacto	198.16	60.30	99.16
	Ours	188.98	46.63	**99.52**
Mars	Metashape	1388.09	72.64	70.32
	Nerfacto	9549.76	2022.66	49.88
	Ours	1560.59	92.53	**95.58**

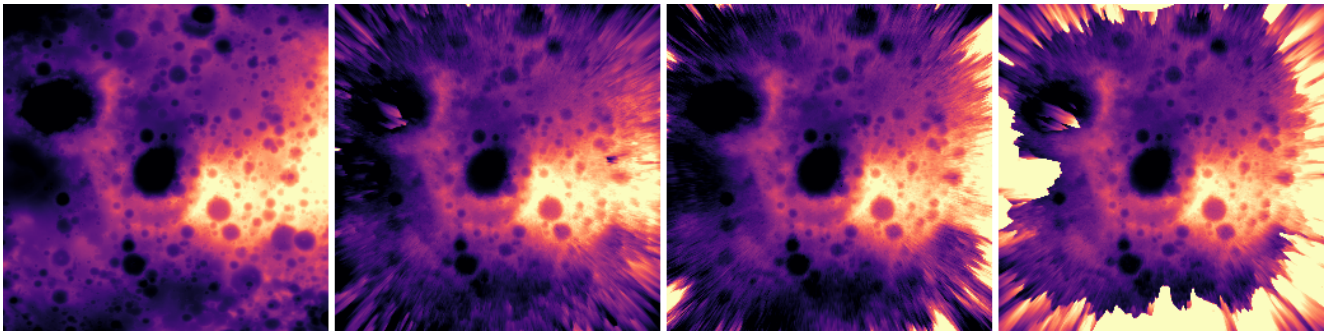
Table 1. Quantitative comparison with Metashape and Nerfacto on lunar and Mars datasets. AED and RED were computed after filling NaNs with the spatially nearest available point, while Coverage@0.1 counts a point as correct only if the relative elevation difference is below 0.1, treating NaNs as incorrect. Ours has shown better Coverage@0.1 compared to baselines.

Datasets. For evaluation, we use two simulated descent sequences, one on the Moon and one on Mars. These model the descent of a 150° fisheye camera in lunar and Mars scenes using SIMply [5]. The lunar scene was constructed from an existing high-resolution DEM mosaic [25] combined with resolved maps of the surface Hapke parameters [24]. In our evaluation, we use a region that approximately covers 100 × 100km and 30 descent images of 1000 × 1000 resolution that descend towards the middle of the scene (Figure 1). The Mars scene was constructed from a Mars Context Camera (CTX) DEM mosaic and a CTX normal albedo mosaic, covering a 150 × 220km region surrounding Gale Crater [19]. We simulated 40 descent images of 1024 × 1024 resolution (Figure 2). We generated masks to make the model ignore the black borders in all images when casting rays.

Metrics. We evaluate predicted heightmaps of size 256 × 256 pixels using three complementary metrics: *Absolute Elevation Difference* (AED), *Relative Elevation Difference* (RED), and *Coverage@0.1*. The *Absolute Elevation Difference* measures the mean absolute vertical error between the predicted height-map $\hat{H}$ and the ground-truth height-map H, defined as $\frac{1}{N}\sum_p |\hat{H}(p) - H(p)|$, where p indexes valid pixels in the ground-truth height-map and N is the total number of evaluated pixels. This metric captures global elevation accuracy and penalizes systematic height offsets. The *Relative Elevation Difference* evaluates local shape fidelity by replacing $\hat{H}(p)$ in the above equation by the mean of a 1km square window around that pixel. When calculating these two metrics, we replaced invalid pixels with the closest valid pixel's $\hat{H}$. *coverage@0.1* measures the proportion of valid points in $\hat{H}$ where the Elevation Difference relative to the ground is less than 0.1. It is defined as:

$$\text{Coverage@0.1} = \frac{\sum_p \mathbf{1}\left(\frac{|\hat{H}-H|}{|H|} \le 0.1\right)}{N} \tag{13}$$

In general, high coverage is desirable, since it maximizes the benefit of capturing wide-angle images during the descent.



GT | **ours-heightmap representation** | **with MiP-NeRF 360's distortion loss** | **with NeRF-like direction dependent colour**

Figure 7. Qualitative comparisons of DEMs using simulated Moon fisheye descent images, with a high-resolution mesh as ground truth. Our full model demonstrates superior coverage while maintaining high-quality reconstruction compared to two ablations.

4.1. Baselines

Metashape. A widely used practical implementation of the pipeline that combines Structure-from-Motion (SfM) [35] and Multi-View Stereo (MVS) [28] is Agisoft Metashape [1]. It integrates feature detection and matching, bundle adjustment, dense depth-map estimation, point cloud densification, mesh reconstruction, and texture mapping into a unified photogrammetry framework. It automates the complete SfM–MVS workflow, enabling end-to-end 3D reconstruction from image alignment to textured mesh generation, and is extensively adopted in applications such as robotics mapping, cultural heritage documentation, surveying, and computer vision research.

Nerfacto. Nerfacto [33] is a neural rendering method built on the NeRF [17] framework for efficiently reconstructing 3D scenes from images. It extends the standard NeRF formulation with techniques from Mip-NeRF 360 [4] to handle unbounded, large-scale scenes, and from Instant-NGP [18] to improve computational efficiency. This combination allows Nerfacto to produce detailed geometry and high-quality novel views, making it well-suited for planetary surface reconstruction from wide-angle descent imagery. As our baseline, we further adapted the model to accept fisheye images and per-image masks as inputs.

4.2. Comparison with Baselines

As seen in Figure 5 and Figure 6, our approach delivers broader coverage on both lunar and Mars data with a significant precision of fine-scale reconstruction. Additionally, we can see in Figure 6 that ours has predicted a smoother surface without floaters compared to Nerfacto. Table 1 further demonstrates that our method outperforms the baselines in terms of coverage, for relative elevation differences below 0.1. Notably, ours achieves superior performance compared to Nerfacto across all metrics on both datasets: AED, RED, and Coverage@0.1, demonstrating that our enhancements are effective. Ours also shows a performance increase of 7.22% on lunar data with respect to Coverage@0.1, and a very substantial increase of 25.26% on Mars data. Note that the empty regions of their reconstruction were filled with the spatially nearest value; This is disadvantageous for NeRF-based methods, including ours, as they may predict outliers rather than leaving invalid regions as NaN.

4.3. Ablations

Angle-based distortion loss. We introduce an angle-based distortion loss that adapts the regularization strength according to ray direction (Eq. 8). This formulation increases regularization for central rays, while relaxing it for edge rays, resulting in more accurate and stable surface reconstruction. To assess its impact, we optimized an additional model using the standard distortion loss with the same weighting factor on lunar data. As shown in Figure 7, the model without angle-based distortion loss produces lower-quality elevation estimates near the edges. It achieves only 69.63% Coverage@0.1, demonstrating reduced spatial completeness compared to our model.

Hapke lighting. Typically, NeRF passes direction encoding to the MLP that generates the color of each sampled point to implicitly model direction-dependent color. We instead use a physically-realistic shading model designed specifically to model planetary surfaces [24], that calculates reflected color based on incident light, viewing, and approximate surface normal directions, considering that we know the direction of sunlight at the time of the descent. To evaluate the effectiveness of this additional physical prior, we optimized another model on lunar data with implicit handling of lighting by passing direction encoding onto the model. We can see in Figure 7 that a more accurate lighting model is important for high-quality DEMs. Quantitatively, this model shows only 88.55% Coverage@0.1 compared to

99.52% coverage of our full model.

5. Conclusion

We have presented the first study of modern neural 3D reconstruction techniques for building digital elevation models using wide-angle descent imagery. As well as evaluating existing methods, we proposed a novel technique specialized for this domain, incorporating a heightmap representation and physical priors. Quantitative and qualitative evaluations of Metashape, Nerfacto, and our proposed approach on lunar and Mars datasets show that these methods have promise for this domain. Our method achieves greater coverage while maintaining high accuracy in DEM reconstruction.

References

[1] Agisoft metashape: Intelligent photogrammetry software. https://www.agisoft.com/. Accessed: 2026-02-27. 1, 2, 3, 7

[2] Dante Abate, Kyriakos Themistocleous, and D Hadjimitsis. The application of neural radiance fields (nerf) in generating digital surface models from uav imagery. In *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, pages 10228–10231. IEEE, 2024. 3

[3] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5855–5864, 2021. 2, 3

[4] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5470–5479, 2022. 2, 3, 4, 6, 7

[5] George Brydon. Image simulation for camera development—python image simulator for planetary exploration (simply). *Space Science and Technology*, 5: 0319, 2025. 6

[6] G. Brydon, D.M. Persaud, and G.H. Jones. Planetary topography measurement by descent stereophotogrammetry. *Mullard Space Science Laboratory, University College London*, 2023. The Centre for Planetary Sciences at UCL/Birkbeck, London, UK. 1, 3

[7] Adam Dai, Shubh Gupta, and Grace Gao. Neural elevation models for terrain mapping and path planning. *arXiv preprint arXiv:2405.15227*, 2024. 3

[8] I. J. Daubar, A. G. Hayes, G. C. Collins, and et al. Planned geological investigations of the europa clipper mission. *Space Science Reviews*, 220:18, 2024. 1

[9] Frank Dellaert, Steven M. Seitz, Charles E. Thorpe, and et al. Em, mcmc, and chain flipping for structure from motion with unknown correspondence. *Machine Learning*, 50:45–71, 2003. 3

[10] Alexander Flisch, Joachim Wirth, Robert Zanini, Michael Breitenstein, Adrian Rudin, Florian Wendt, Franz Mnich, and Roland Golz. Industrial computed tomography in reverse engineering applications. *DGZ-fP-Proceedings BB*, 4 (7):45–53, 1999. 3

[11] Werner Göbel, Björn M Kampa, and Fritjof Helmchen. Imaging cellular network dynamics in three dimensions using fast 3d laser scanning. *Nature methods*, 4(1):73–79, 2007. 3

[12] Margaret Hansen, Caleb Adams, Terrence Fong, and David Wettergreen. Analyzing the effectiveness of neural radiance fields for geometric modeling of lunar terrain. In *2024 IEEE Aerospace Conference*, pages 1–12. IEEE, 2024. 4

[13] B. M. Hynek. Extraterrestrial digital elevation models: Constraints on planetary evolution, with focus on mars. *International Journal of Remote Sensing*, 31(23):6259–6274, 2010. 1

[14] Francesco Isgro, Francesca Odone, and Alessandro Verri. An open system for 3d data acquisition from multiple sensor. In *Seventh international workshop on computer architecture for machine perception (CAMP'05)*, pages 52–57. IEEE, 2005. 3

[15] 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. 3

[16] Karl Kraus and Norbert Pfeifer. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and remote Sensing*, 53(4): 193–203, 1998. 3

[17] 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. *Communications of the ACM*, 65(1):99–106, 2021. 1, 2, 3, 4, 7

[18] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 4, 7

[19] Divya M. Persaud. On multi-resolution 3d orbital imagery and visualisation for mars geological analysis. UCL Discovery, 2022. University College London repository. 1, 2, 6

[20] Cynthia B. Phillips, Jennifer E. C. Scully, Marissa E. Cameron, Kathleen L. Craft, Cyril Grima, Divya M. Persaud, and the Europa Clipper Reconnaissance Focus Group. A reconnaissance strategy for landing on europa, based on europa clipper data. In *55th Lunar and Planetary Science Conference (LPI Contrib. 3040)*, page 1672, 2024. 2024LPICo3040.1672P. 1

[21] Srikumar Ramalingam, Suresh K Lodha, and Peter Sturm. A generic structure-from-motion framework. *Computer Vision and Image Understanding*, 103(3):218–228, 2006. 3

[22] CMPPC Rocchini, Paulo Cignoni, Claudio Montani, Paolo Pingi, and Roberto Scopigno. A low cost 3d scanner based on structured light. In *computer graphics forum*, pages 299–308. Wiley Online Library, 2001. 3

[23] Ewelina Rupnik, Mehdi Daakir, and Marc Pierrot Deseilligny. Micmac–a free, open-source solution for photogrammetry. *Open geospatial data, software and standards*, 2(1): 14, 2017. 1, 3

[24] H. Sato, M. S. Robinson, B. Hapke, B. W. Denevi, and A. K. Boyd. Resolved hapke parameter maps of the moon. *Journal of Geophysical Research: Planets*, 119(8):1775–1805, 2014. 2, 5, 6, 7

[25] F. Scholten, J. Oberst, K.-D. Matz, T. Roatsch, M. Wählisch, E. J. Speyerer, and M. S. Robinson. Gld100: The near-global lunar 100 m raster dtm from lroc wac stereo image data. *Journal of Geophysical Research: Planets*, 117(E12): E00H17, 2012. 1, 2, 6

[26] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4104–4113, 2016. 1, 3

[27] Brent Schwarz. Mapping the world in 3d. *Nature Photonics*, 4(7):429–430, 2010. 3

[28] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, pages 519–528. IEEE, 2006. 1, 3, 7

[29] Laurence A Soderblom, Martin G Tomasko, Brent A Archinal, Tammy L Becker, Michael W Bushroe, Debbie A Cook, Lyn R Doose, Donna M Galuszka, Trent M Hare, Elpitha Howington-Kraus, et al. Topography and geomorphology of the huygens landing site on titan. *Planetary and Space Science*, 55(13):2015–2024, 2007. 1, 3

[30] Peter Sturm. Multi-view geometry for general camera models. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pages 206–212. IEEE, 2005. 3

[31] Peter Sturm, Srikumar Ramalingam, and Suresh Lodha. On calibration, structure from motion and multi-view geometry for generic camera models. In *Imaging Beyond the Pinhole Camera*, pages 87–105. Springer, 2006. 3

[32] R. Szeliski. A multi-view approach to motion and stereo. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, pages 157–163 Vol. 1, 1999. 3

[33] Matthew Tancik, Ethan Weber, Evonne Ng, Ruilong Li, Brent Yi, Justin Kerr, Terrance Wang, Alexander Kristoffersen, Jake Austin, Kamyar Salahi, Abhik Ahuja, David McAllister, and Angjoo Kanazawa. Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings*, 2023. 2, 4, 7, 1

[34] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams: A factorization method. *Proceedings of the National Academy of Sciences of the United States of America*, 90(21):9795–9802, 1993. 3

[35] Shimon Ullman. The interpretation of structure from motion. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 203(1153):405–426, 1979. 1, 3, 7

[36] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 3

[37] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):139–144, 1980. 3

[38] Linglong Zhou, Guoxin Wu, Yunbo Zuo, Xuanyu Chen, and Hongle Hu. A comprehensive review of vision-based 3d reconstruction methods. *Sensors*, 24(7):2314, 2024. 3

[39] Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. Ewa volume splatting. In *Proceedings Visualization, 2001. VIS'01.*, pages 29–538. IEEE, 2001. 3

Neural 3D Reconstruction of Planetary Surfaces from Descent-Phase Wide-Angle Imagery

Supplementary Material

6. Implementation and Resource details

We developed our methodology on top of Nerfacto by Nerfstudio [33]. We optimized our dual-pipeline on each dataset for 100K iterations and optimized Nerfacto baseline for 30K iterations on a local cluster running Ubuntu 22.04.1 with CUDA version 12.4. The cluster nodes are equipped with four NVIDIA A5000 GPUs (24GB VRAM each) and a single 16-core CPU with 128GB RAM. Each training run utilized a single GPU.

Tuned hyperparameters for each dataset are listed below:

Table 2. Hyperparameters for each dataset.

Hyperparameter	**Lunar**	**Mars**
t_{near}	0.15	0.10
t_{far}	500.0	10000.0
$\lambda_{c;A}$	1.0	1.0
$\lambda_{c;B}$	0.001	0.00001
λ_{height} $(it < 2000)$	0.0	0.0
λ_{height} $(2000 \leq it < 4000)$	1.0	1.0
λ_{height} $(it \geq 4000)$	0.001	0.001
λ_{dist}	0.01	0.5
k_1	10000.0	10000.0
k_2	1000.0	1000.0