NeLU3D: Neural Inverse Structured Light without Explicitly Modeling the Projector

Giancarlo Pereira[1], David Fouhey[1], Claudio T. Silva[1], and Daniele Panozzo[1]

New York University, New York NY, USA
{giancarlo.pereira,fouhey,csilva,panozzo}@nyu.edu

Abstract. Structured Light (SL) is a popular technique that delivers precise 3D shape acquisition across diverse materials and object sizes. SL requires, however, careful modeling of the set-up, including geometric and radiometric calibration of the camera-projector pair; without it, reconstruction quality degrades. We propose *NeLU3D*, a neural inverse SL method without explicitly modeling the projector. We calibrate the camera–projector system using a neural network that maps 3D positions to a set of projected patterns. Then, with as few as four monochromatic images (or two RGB images), our approach uses differentiable volume rendering to fit a surface to match SL captures. We scan over twenty-five objects of different shapes and reflectances to demonstrate the feasibility and quality of our method in a handful of projector-camera set-ups, including an extremely low-cost projector and an analog projector with fixed RGB pattern. We also showcase sub-millimeter accuracy with sub-optimal patterns, where previous methods recover noisy 3D surfaces. We release an open-source implementation at `https://github.com/geometryprocessing/neural-lookup`.

Keywords: Structured Light · 3D Reconstruction · Neural Method · Inverse Rendering · Differentiable Rendering · Volume Rendering

1 Introduction

3D scanning is concerned with acquiring digital representations of real geometry, be it for cultural preservation, robotic manipulation, *etc*. Structured light (SL) stands out as a well-established non-contact technique that delivers consistent, repeatable results across diverse materials and object sizes. SL systems typically reconstruct 3D geometry using a calibrated projector-camera pair, making it accessible with off-the-shelf components. The camera and projector sit in a stereo configuration; the projector emits time-varying patterns and the camera captures the objects "painted" by said patterns. Through epipolar constraints and triangulation, one can extract a point cloud with sub-millimeter precision.

SL, unfortunately, is limited by the need to emit multiple patterns for accurate pixel correspondences for triangulation. SL also requires modeling each piece

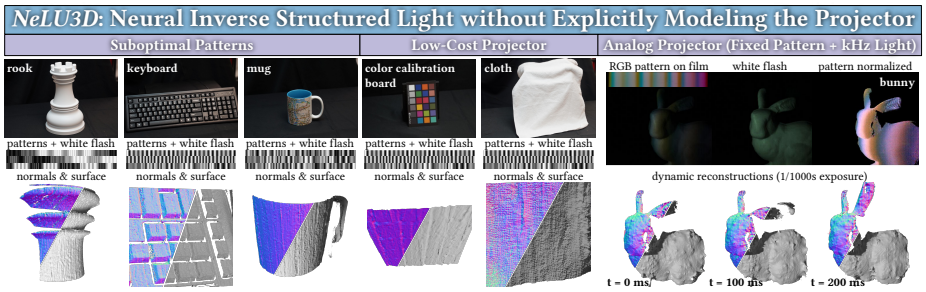

Fig. 1: Neural LookUp3D for Structured Light (SL) Scanning Without Explicitly Modeling the Projector. SL requires careful geometric and radiometric of the camera-projector pair and careful choice of projector patterns to reconstruct object surfaces. Our proposed method, on the other hand, can output accurate normals and geometry from as few as three monochromatic patterns and a white flash. We validate our method by scanning over 25 objects with a variety of camera-projector pairs.

of the set-up very carefully. Since the 3D shape is recovered from triangulating light rays between camera and projector, the final surface quality depends on the proper modeling of these components, *e.g.* pose registration, lens distortions, radiometric calibration, interpixel gaps in a digital projector, and blurring. Errors in modeling any such component result in errors in the reconstructed geometry.

In this work, we propose an end-to-end neural inverse SL scanning approach that sidesteps the explicit modeling of the projector, freeing the surface reconstructions from projector limitations. Our proposed approach *NeLU3D* is inspired by two recent advances in SL scanning: TurboSL [34], a neural inverse structured light scanning method, and LookUp3D [38], a data-driven scanning method that foregoes explicit projector modeling. By carefully crafting a one-time neural network calibration that maps 3D positions to projected colors, our method simplifies the image formation pipeline in inverse rendering, reducing memory requirements and speeding up the time taken by each iteration. *NeLU3D* allows the use of, *e.g.*, an extremely low-cost projector or an analog projector without sacrificing surface reconstruction quality (Fig. 1, Fig. 6).

To demonstrate the quality of our method in realistic, challenging scenarios, we test it using suboptimal patterns (*i.e.* not carefully designed for a particular set-up), a low-cost hard-to-calibrate projector, and an analog projector with a fixed pattern for slow-motion capture. We scan over twenty five objects, ranging from day-to-day items to 3D printed objects to precisely milled ones (see our supplemental for all results). We show results for both static and moving objects with as few as 3 monochromatic patterns and 1 white flash (or 2 trichromatic images) in a handful of projector-camera configurations without ever explicitly modeling the projector. We additionally verify sub-millimeter accuracy against traditional dense SL done with 44 Gray codes (22 horizontal and 22 vertical).

2 Related Work

In this project, we are concerned with accurate 3D reconstruction of objects using visible light. We focus on structured light (SL) scanning, a reliable, accessible, and robust method that reconstructs sub-mm accurate geometry. The minimum requirements for SL scanning are off-the-shelf components: one camera and a controllable, spatially modulated light source (typically a digital projector). Our method foregoes explicitly modeling the projector for SL scanning by leveraging a one-time color calibration and neural inverse rendering for reconstruction. Below, then, we review SL methods and newer improvements to surface reconstruction.

2.1 Structured Light Scanning

By controlling the light cues in a scene, one can efficiently extract a lot of information about the geometry and surface properties of an object [9, 14, 15, 19, 20, 23, 28, 30, 36, 41, 48]. Structured Light (SL) scanning remains an accessible and robust method for accurately acquiring 3D geometry [40, 52]. Researchers have spent decades iterating through and developing different SL codes, each with different benefits and trade-offs [2–4, 6, 10, 11, 13, 21, 22, 33, 39, 49]. The quality of reconstruction, however, unfortunately heavily depends on the set-up used and its proper calibration and modeling. For instance, inter-pixel gaps and limited resolution of digital projectors cause Moiré patterns in reconstructed surfaces [8, 26]. Improper radiometric calibration can make decoding impractical for smooth patterns, such as sinusoidals and Hamiltonians. Additionally, SL typically requires multiple patterns emitted onto the object while it remains static for proper decoding. Researches have made hardware alterations to overcome some limitations from digital projectors [8, 47] and to boost the framerate of SL scanning [7,17,18,23,24,27,43,53], but altering hardware is laborious and requires careful engineering. Our proposal, on the other hand, does not require explicitly modeling the projector: this in turn allows us to use off-the-shelf components as is, without hardware modifications and without worrying about optical misalignment, defocusing, vignetting, geometric or radiometric calibration (Fig. 1).

2.2 Neural Inverse Rendering

The last decade has seen tremendous development of techniques for 3D reconstruction entirely passively from uncalibrated multiview 2D images [5, 25, 29, 32, 37, 45, 46, 50, 51]. Shandilya *et al*. [42] in 2022 brought NeRF to a single-pattern SL system, but the method still required multiple views of the same static scene.

In 2024, Mirdehghan *et al*. [34] demonstrated the ability to use neural inverse rendering with a single camera view and single projector for SL scanning. Most of their scenes, however, were scanned using a la carte patterns [33], which are tailored to the specific SL set-up. This is an additional burden, as the user needs to carefully craft these patterns and redesign is needed whenever the camera-projector arrangement changes. As a direct follow-up to [34], Urakawa and Watanabe [44] added a neural network to learn a displacement

field to compensate for motion during multiple pattern projections and used phase-shifted sinusoids, a family of simpler patterns widely used that do not require redesign for different camera-projector arrangements. But their solution ultimately involved two cameras to overcome phase unwrapping ambiguities.

One major constraint of both [34, 44] is the projector. Their systems require careful pose registration, calibration of focal length and distortion parameters, and radiometric calibration to properly reconstruct 3D surfaces. Any such error in those parameters (or additional complications, such as defocusing or interpixel gaps in the projector) gets propagated into the reconstructed surface. As part of their optimization pipeline, for instance, these systems require learning a 11×11 blur kernel to attempt modeling the imperfect optics of the projector. This ultimately limits the types of light sources that can be used for SL, as low-cost projectors tend to be hard to properly calibrate and ultimately produce low-quality reconstructions. It also places a burden on the user to carefully calibrate the projector. Our approach, on the other hand, does not require modeling the projector explicitly. Therefore, typical projector issues, *e.g.* poor optics, low-dynamic range, or interpixel gaps, do not cause final reconstruction quality to degrade. Our method can acquire accurate geometry even with a low-cost projector and an analog projector built for slow-motion capture (Fig. 1).

2.3 Foregoing Explicit Modeling of Projector

In 2003, Scharstein and Szeliski [41] proposed a stereo method using structured light for depth acquisition without calibrating the light source; it required a pair of cameras and at least one light projector. More recently, LookUp3D by Pereira *et al.* [38] introduced a data-driven SL scanning method that bypassed explicitly modeling the projector using a single camera. The method relied on a mapping of observed color values at a single camera pixel to a single scene depth. Since there is no ray triangulation between camera and projector, there is no need to know the extrinsic or intrinsic parameters of the projector. Defocusing, vignetting, interpixel gaps, and other effects get baked into the table, removing the need to model projector imperfections. LookUp3D, however, performs pixelwise decoding (sensitive to noise) and is memory intensive (each pixel stores hundreds of RGB and depth values). Our proposal instead uses an inverse rendering optimization with an implicit representation of the scene geometry, reducing memory requirements (Sec. 3.1) and effectively performing outlier rejection (Fig. 6).

3 *NeLU3D*: Neural LookUp3D Method

The goal of our method is to recover a surface from captured structured light images of a scene without explicitly modeling the light source. We cast this problem as inverse rendering by fitting scene geometry, represented by a multilayer perceptron (MLP) $f : \mathbb{R}^3 \to \mathbb{R}$ mapping 3D points to a signed distance function (SDF), and scene appearance, represented by MLP $a : \mathbb{R}^2 \to [0, 1]$ mapping 2D image coordinates to global illumination contributions. The two MLPs

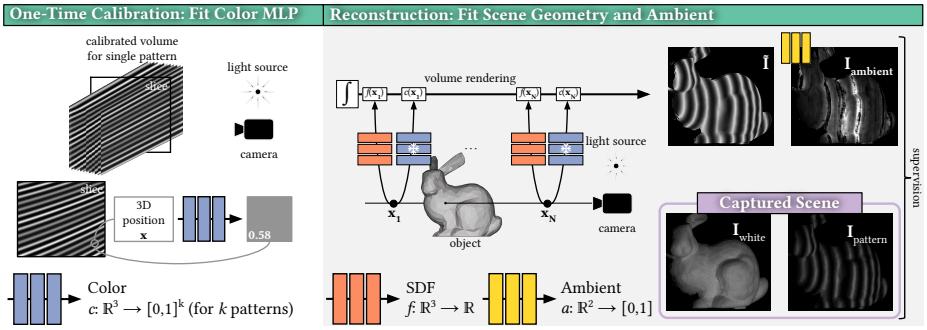

Fig. 2: *NeLU3D* Pipeline. Left: *NeLU3D* works by first mapping a 3D position $\mathbf{x}$ to its color when illuminated by k patterns. This mapping is compressed into an MLP c. **Right**: Given a new capture of $\mathbf{I}_{\text{pattern}}$ and $\mathbf{I}_{\text{white}}$ of an unknown object, *NeLU3D* uses the frozen color MLP c to optimize a scene SDF f and an ambient light image $\mathbf{I}_{\text{ambient}}$ (outputted by 2D MLP a) so that volume rendering matches $\mathbf{I}_{\text{pattern}}$ and $\mathbf{I}_{\text{white}}$.

are optimized to explain the captured images via volumetric rendering, as done in TurboSL by Mirdehghan *et al.* [34] for structured light and NeuS [45] for multiview images. Our key innovation is the careful integration of ideas from LookUp3D by Pereira *et al.* [38], who create a per-pixel lookup table mapping the perceived color of a point to its 3D position when illuminated by projected patterns (after normalizing it by a white flash to remove albedo and other nuisances). After a one-time calibration, we bake this point-color mapping into a color MLP $c : \mathbb{R}^3 \rightarrow [0,1]^k$ (where k is the number of patterns). The functions c, f, and a can be used to render an image via the standard volumetric rendering equations. By holding c fixed at reconstruction time, we can recover an SDF for the scene and use its zero level set to extract a surface. The color to point calibration bypasses geometric and radiometric calibration of the projector, permitting the use of low-cost and analog projectors, and suboptimal patterns not uniquely designed to a specific set-up, without sacrificing quality. We refer to the method, shown in Fig. 2, as Neural LookUp3D (abbreviated to *NeLU3D*).

We begin by introducing the ideas of LookUp3D in a self-contained manner; we then explain how we integrate these ideas with neural rendering (Sec. 3.1). Throughout, we assume that we project k monochromatic patterns onto the scene. During calibration, two additional captures are necessary: one with a white flash (for an 8-bit projector, all pixels set to 255) to skip radiometric calibration and remove albedo and one with the light source off to remove global illumination contributions. During object scanning, the white flash is still required to remove albedo; the ambient term instead is fitted during optimization. Then, we discuss losses optimized during training (Sec. 3.2) and implementation details (Sec. 3.3).

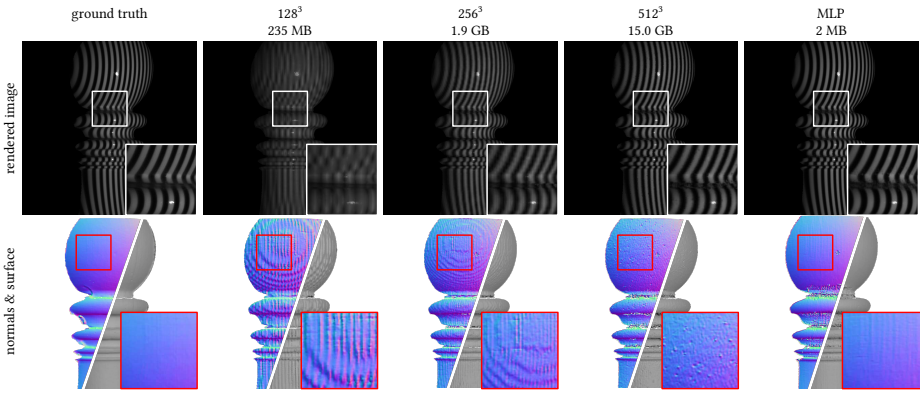

Fig. 3: Making Lookup Table Differentiable. The lookup table L, as presented in LookUp3D [38], maps billions of 3D positions to their respective k projected pattern intensities. The explicit table, however, requires a fine grid to properly resolve details and memory scales cubically with grid size. With a positional encoding (*e.g.* 20 frequency levels as shown here) and a shallow MLP, we can simultaneously compress L and make it differentiable for inverse rendering. With the MLP, we acquire accurate geometry, faithfully match the SL capture, and maintain a smaller memory footprint.

3.1 Learning the Color Mapping

The basic idea of LookUp3D [38] is to build a lookup table L that maps 3D points to their "color" when illuminated by k projected patterns. L is created by sweeping a white flat target (equipped with fiduciary markers for relative camera pose registration) across a calibration volume with a linear stage. A white flash and ambient capture are needed to remove albedo and indirect light contributions. After calibration, then, L stores the direct light contribution from k patterns at each camera pixel and corresponding depth.

Once the table L is built, depth recovery of an unseen object is a direct search per pixel: one checks all points along a camera ray in L and compares it to the captured color. Mathematically, if $\mathbf{v}_i$ is the ray departing ith pixel and $\mathbf{I}_i$ is the ith pixel of the image, then one solves $\arg\min_{d_i \geqslant 0} ||L[d_i \mathbf{v}_i] - \mathbf{I}_i||$ to find depth d_i. This simple approach is effective and skips explicit projector calibration.

The approach, however, has several downsides that stem from inference being a brute-force pixelwise search. Although parallelizable, the lookup search is costly and memory intensive: each pixel needs to search through all calibrated positions to find the best color match. Additionally, pixelwise decoding is sensitive to noise: a reconstruction can be significantly covered by outliers, which is hard to post-process for downstream tasks [1]. Finally, the arg min search is non-differentiable.

We propose instead to bake the lookup table into a learned MLP $c : \mathbb{R}^3 \to \mathbb{R}^k$, such that, given a 3D position, the MLP predicts the k normalized intensities at that point. We choose an MLP as it is easy to implement and efficient in PyTorch: a multiresolution hash encoding (MHE) [35] and a compact MLP compress the lookup table (*e.g.*, from 50GB to 49MB for $k = 3$ and from 121GB to 49MB

for $k = 11$ patterns) and permit its use in a differentiable rendering framework. During the geometry optimization, the weights for the MLP c remain frozen.

There are other ways make the lookup table differentiable. The table (which contains half a billion points or more) is not in a evenly spaced grid, so a nearest neighbor search grows intractable with trivial implementations. A simpler solution is to force the table into an even grid: with this option we can use PyTorch's *grid_sample* which is fast and handles gradients, but memory grows cubically with the grid resolution (Fig. 3). At grid resolution 1024^3, we run out of memory in an RTX A6000 GPU. The MLP retains far greater detail at a much smaller size; the trade-off is that the MLP needs to be trained (which takes 3 minutes to train an MHE + MLP in our RTX A6000 GPU for 5,000 iterations). For more details and experiments, refer to Supplementary Sec. A.3.

This choice simplifies the neural inverse rendering pipeline from Mirdehghan *et al.* [34]: each iteration of the training loop is faster (Tab. 2) and allows for the use of suboptimal patterns and arbitrary light sources, such as low-cost projectors and analog projectors, without sacrificing quality (Fig. 6, Tab. 1).

3.2 Optimization

NeLU3D trains, per scene, by minimizing the differences between the rendered and captured camera images with respect to neural networks representing the SDF and the ambient term. Our work here builds on the knowledge from [34,44], but ours diverges as we do not need to model any image formation for the projector, nor predict a projector blur kernel or the reflectance of the captured scene. Our rendering pipeline is simplified by the calibrated color MLP c.

Given a 3D point $\mathbf{x}(i, d) = \mathbf{o} + d\mathbf{v}_i$ (where $\mathbf{o} \in \mathbb{R}^3$ is the camera's origin and $\mathbf{v}_i$ is the direction of the ray departing from ith pixel), we can retrieve the color intensity $c(\mathbf{x}) \in \mathbb{R}^k$ for k patterns from the frozen lookup MLP (Sec. 3.1) and $f(\mathbf{x})$ from the SDF MLP. The SDF values are converted into density following NeuS [45] (see supplementary Sec. A.1). We sample the ray departing from $\mathbf{v}_i$ at several depth positions and use standard volumetric rendering equations to render the ith pixel of the image $\tilde{\mathbf{I}}$. This rendered pixel $\tilde{\mathbf{I}}_i$ corresponds solely to the diffuse direct light coming from the light source, without accounting for specularities, indirect light, ambient light, and/or noise. To account for the remaining light components, we also fit the 2D MLP a, which outputs the ambient term $\mathbf{I}_{\text{ambient}}$ (operations are pixelwise and loss is summed over pixels i):

$$\tilde{\mathbf{I}}_{\text{pattern}} = \tilde{\mathbf{I}}(\mathbf{I}_{\text{white}} - \mathbf{I}_{\text{ambient}}) + \mathbf{I}_{\text{ambient}} \tag{1}$$

$$\mathcal{L}_{\text{pattern}} = \sum_i \left\| \mathbf{I}_{\text{pattern}}(i) - \tilde{\mathbf{I}}_{\text{pattern}}(i) \right\| \tag{2}$$

We can also follow the original intuition from LookUp3D and compare the normalized images (operations are pixelwise and loss is summed over pixels i):

$$\mathcal{L}_{\text{LookUp}} = \sum_i \left\| \left(\frac{\mathbf{I}_{\text{pattern}}}{\mathbf{I}_{\text{white}}} \right)(i) - \left(\frac{\tilde{\mathbf{I}}_{\text{pattern}}}{\mathbf{I}_{\text{white}}} \right)(i) \right\| \tag{3}$$

For well-lit scenes (i.e., most of the pixels in $\mathbf{I}_{\text{white}}$ are well exposed) the differences between two loss terms is negligible. We would like to note, however, that the loss equations differ the most on low-light regions, where dividing by a smaller pixel intensity from $\mathbf{I}_{\text{white}}$ tends to change the normalized pixel intensity by many orders of magnitude. For well-lit scenes, this means that $\mathcal{L}_{\text{LookUp}}$ attributes a lot of weight to noisier pixels (see ablation study of pawn in Fig. 7). $\mathcal{L}_{\text{LookUp}}$, however, is necessary for images with low exposure times and low albedo (*e.g.* the monkey statue and black remote controller in Fig. 4). Overall, we opt for a balance between $\mathcal{L}_{\text{pattern}}$ and $\mathcal{L}_{\text{LookUp}}$ with the α parameter $\in [0, 1]$:

$$\mathcal{L}_{\text{image}} = (1 - \alpha)\mathcal{L}_{\text{pattern}} + \alpha\mathcal{L}_{\text{LookUp}} \tag{4}$$

To avoid exploding gradients during the optimization loop (Fig. 7), we constrain the ambient output such that $\mathbf{I}_{\text{ambient}} \leqslant \mathbf{I}_{\text{white}}$. This is theoretically sound, as the white flash capture is done with all projector pixels turned to their maximum value (255 at 8-bit projection). The residual contribution at a single camera pixel, then, cannot exceed the total value integrated during the white flash capture. To enforce this constraint, we simply take the output of the ambient MLP a (which goes through a sigmoid activation and therefore $\mathbf{I}_{\text{ambient}} \in [0, 1]$) and multiply it pixelwise by $\mathbf{I}_{\text{white}}$. This is a trade-off of our method: although we require capturing the white flash, we can remove the reflectance network from the imaging pipeline and can upper bound the ambient contribution from $\mathbf{I}_{\text{ambient}}$.

In addition to the image losses, we also calculate a mask loss, an eikonal loss, and a sparsity loss, all common to other methods that train SDFs on 2D image supervision [34,44,45,51]. To remove density from shadowed/dark regions of the scene, we calculate a binary mask $\mathcal{M}$ and calculate a mask loss $\mathcal{L}_{\text{mask}}$:

$$\mathcal{L}_{\text{mask}} = \underset{i}{\text{BCE}}\left(\mathcal{M}(i), \int_{d_{min}}^{d_{max}} T(i,\delta)\sigma(i,\delta)d\delta\right) \tag{5}$$

where T is the transmittance and σ is the density function obtained from the SDF values (refer to NeuS [45] and Sec. A.1 of our supplemental document for details on volume rendering). BCE is the binary-cross entropy loss and the integral computes the total attenuation along the ray for ith pixel.

The eikonal loss $\mathcal{L}_{\text{eik}}$ pushes the gradient of the SDF to equal 1 and the sparsity loss $\mathcal{L}_{\text{sp}}$ regularizes the 3D points where the SDF equals zero:

$$\mathcal{L}_{\text{eik}} = \sum_{i,d}(\|\nabla f(\mathbf{x}(i,d))\|_2 - 1)^2, \quad \mathcal{L}_{\text{sp}} = \sum_{i,d}\exp\left(-|f(\mathbf{x}(i,d))|\right) \tag{6}$$

3.3 Implementation Details

We train every scene on an RTX A6000 GPU for 10,000 iterations. The loss functions are weighed as follows: $L1$ and $L2$ norms for the lookup loss $\mathcal{L}_{\text{LookUp}}$ and the pattern loss $\mathcal{L}_{\text{pattern}}$ are set to 1.0 and 10.0, respectively. The weights for the mask loss $\mathcal{L}_{\text{mask}}$, eikonal loss $\mathcal{L}_{\text{eik}}$, and sparsity loss $\mathcal{L}_{\text{sp}}$ are set to 0.1, 0.1, and 0.01, respectively. We experimentally test (see supplementary Sec. B.4) and set $\alpha = 0.1$ for all results in this paper and in the supplemental document.

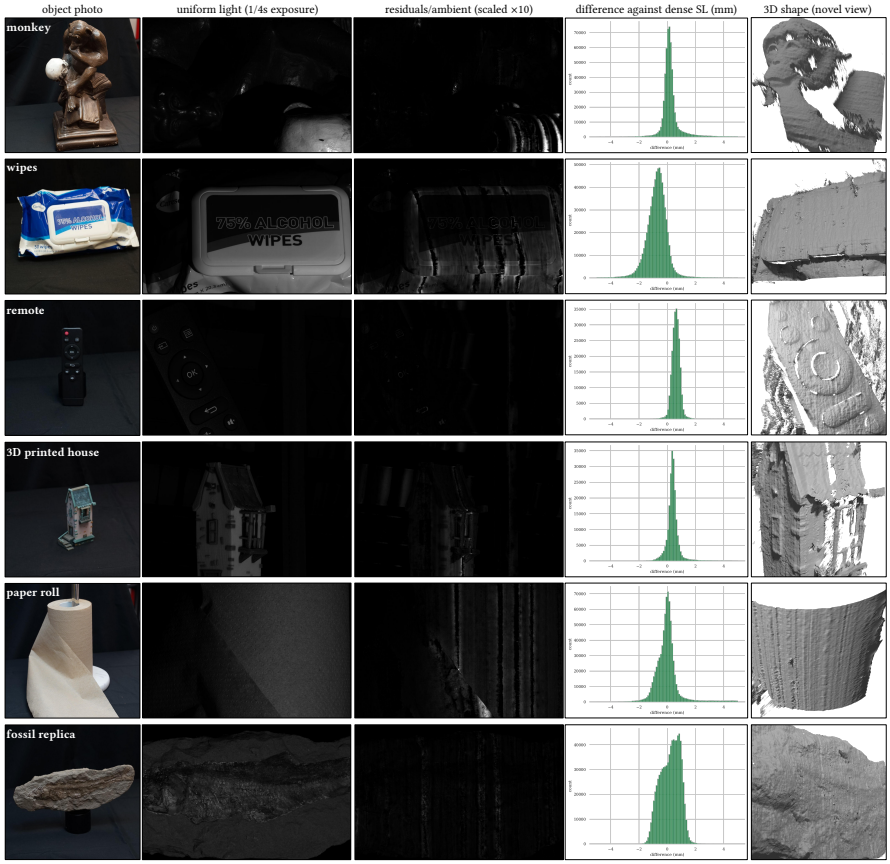

Fig. 4: Static Results with Atlas camera and DLP Projector. We highlight *NeLU3D* results with $k = 3$ patterns (1 sinusoidal, 1 triangular, and 1 random – the same from Fig. 6's third column) and the white flash. *NeLU3D* reconstructs fine geometric details, even with interreflections, low albedo, and subsurface scattering.

4 Experimental Results

In this section, we show results from four different structured light set-ups: two with digital projectors (Fig. 4, Fig. 6, and Fig. 7) and two with an analog projector (Fig. 5). The DLP projector is the Texas Instruments DLP4710EVM-LC, with a resolution of 1920×1080 and a DMD chip. The LCD projector is the VOPLLS Mini Beaner N3, a 50USD projector with cheap optics. With these two digital projectors, we use the monochromatic Lucid Vision Labs Atlas 31.4MP with an Edmund Optics APS-C 50mm lens, with a fixed exposure of 1/4 second.

The other two set-ups rely on the open-source analog projector from Pereira *et al*. [38]. This is a projector with two perpendicular 100W LEDs (with a beam splitter in between): one for the white flash capture, the other paired with a

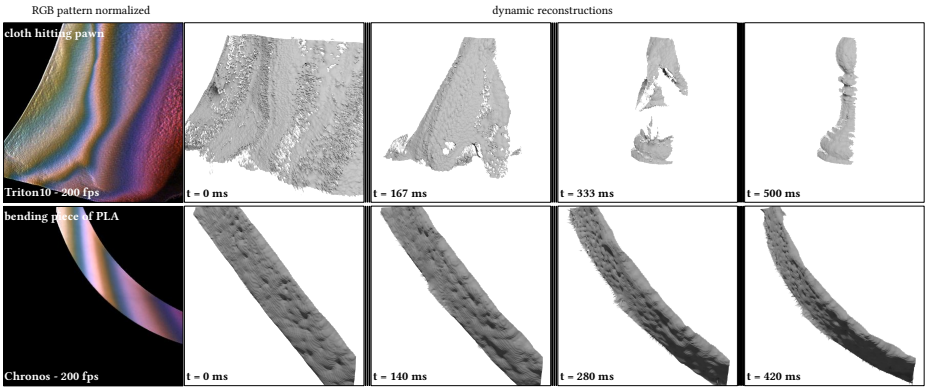

Fig. 5: Dynamic Results with Trichromatic Cameras, Analog Projector, and RGB Pattern on Film. We highlight two dynamic scenes using an analog projector with a helical RGB pattern developed on 35mm film. **Top row (Triton10 at 200fps)**: cloth hitting the pawn. **Bottom row (Chronos at 200fps)**: piece of PLA bending.

RGB pattern developed on 35mm film. Since this projector can flicker the LEDs at kilohertz frequency, we use it with two high-speed cameras equipped with Bayer filters for RGB captures: (1) Chronos HD camera equipped, which achieves 1k fps at 2 Megapixel, paired with a Sigma 35mm f1.4 lens; (2) the Triton10 5MP camera from Lucid Vision Labs (model TRX051S), which achieves 200 fps and 5 Megapixel resolution, paired with an Edmund Optics 12mm C Series Fixed Focal Length lens (stock number 58-001). We test these two cameras at different exposure times and framerates, but all below 1/200 seconds.

We also compare *NeLU3D* with three other SL methods (Sec. 4.1) and perform an ablation study of our system (Sec. 4.2). For more *NeLU3D* results, check supplementary Sec. B.6 (static), Sec. B.7 (dynamic), and our video. All results use a single camera view - the camera remains fixed after color calibration.

4.1 Comparisons

We compare *NeLU3D* to TurboSL [34], LookUp3D with low-rank approximation [38], and ZNCC [31, 33] in Tab. 1, Tab. 2, and Fig. 6. LookUp3D and ZNCC are pixelwise decoders; TurboSL is a neural inverse structured light method. Pixelwise decoders are paralellizable and can output point clouds in fractions of a second (Tab. 2); they are, however, sensitive to noise and often output point clouds covered by outliers (**first** and **second rows** of Fig. 6). In the neural inverse SL set-up, the inductive bias of the SDF acts as an effective mechanism for outlier rejection, outputting smoother surfaces with good normal estimation. TurboSL [34], however, cannot properly fit the scene geometry and appearance when we use $k = 3$ suboptimal patterns or when we use a very cheap projector without careful calibration (**third row** of Fig. 6). The cheap projector lens has strong vignetting and distortion and the projected image does not focus on a

Table 1: Quantitative results of ZNCC [31, 33], LookUp3D [38], TurboSL [34], and our proposed method *NeLU3D*. We report mean depth (in mm) and disparity (in pixels) errors against dense SL results. The dense SL results are obtained with 44 Gray codes (11 horizontal and their inverse, 11 vertical and their inverse) using ZNCC. For a fair comparison, all methods also ingest the white flash and ambient captures. See supplemental Sec. B.10 for plots and more results.

		mean absolute error against dense SL (mm / px)			
objects	patterns	ZNCC	LookUp3D	TurboSL	*NeLU3D*
wipes	3 Hilbert	196.70 / 832.95	8.81 / 153.23	29.91 / 181.68	**3.87** / **8.13**
first aid	3 Random	298.44 / 514.57	13.27 / 38.93	122.00 / 415.71	**0.53** / **1.39**
monkey	3 Random	265.24 / 525.16	41.11 / 298.44	40.04 / 409.94	**0.96** / **2.90**
teapot	3 A La Carte	92.44 / 333.71	5.70 / 16.82	**2.90** / **6.41**	9.17 / 25.43
pawn	4 MPS	44.98 / 232.86	3.17 / 9.70	2.76 / 7.54	**1.02** / **2.78**
keyboard	11 Gray	0.39 / 1.14	33.09 / 137.16	0.97 / 2.90	**0.33** / **0.84**

single plane (for more details on projector hardware we used, check supplemental Sec. A.4); these issues typically make SL struggle, but *NeLU3D* maintains sub-mm accuracy in these scenarios. In Tab. 1, we also report the mean absolute error for depth (in millimeters) and for disparity (in pixels) for a variety of objects and pattern combinations. For suboptimal patterns, such as random and Hilbert, *NeLU3D* maintains good accuracy while previous methods struggle without any additional, careful calibration and modeling. With better pattern design for SL scanning, *e.g.* 11 Gray code, 4 micro phase shifting patterns [12], and 3 a la carte patterns [33], the previous methods (TurboSL in particular) perform a lot better. For additional comparisons and visualizations, see supplemental Sec. B.10.

We would like to note that *NeLU3D* **requires** the white flash capture for reconstruction. That means a reconstruction with $k = 3$ patterns has three distinct patterns plus the white flash (*i.e.* 255 intensity for 8-bit projector pixels), totaling four captures. LookUp3D requires both the white flash and ambient captures to reconstruct a scene, totaling 5 captures for $k = 3$ patterns. In Fig. 6, Tab. 1, and Tab. 2, we feed the ambient and flash captures to all methods for a fair comparison. One advantage of ZNCC and TurboSL is that they rely only on traditional calibration steps for SL scanning, while LookUp3D and *NeLU3D* require the lookup table calibration as a trade-off from geometric and radiometric calibrations of the projector. *NeLU3D* then requires training the MLP c of these color mappings (please refer to Sec. 3.1 and supplementary A.3 for more details). We believe this is a reasonable trade-off, as it allows SL scanning to break free from many typical limitations of projector and the projected patterns. We can use patterns not uniquely optimized for particular set-ups and cheap, hard-to-calibration projectors to do SL scanning without sacrificing accuracy (Fig. 6).

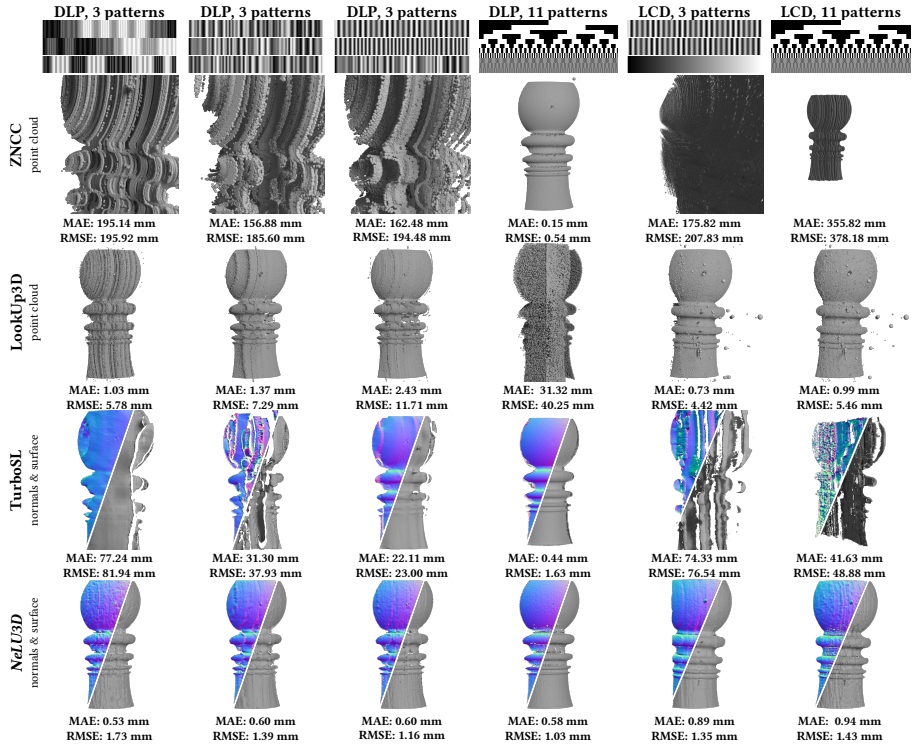

Fig. 6: Comparisons. We compare *NeLU3D* to ZNCC [31,33] (**first row**), LookUp3D [38] (**second row**), and TurboSL [34] (**third row**). *NeLU3D* (**last row**) accurately reconstructs the pawn with suboptimal patterns and with a low-cost projector without any need to explicitly model the light source, while previous methods struggle with many noisy outliers or severe distortions. **Columns**: **(1)** 3 Hilbert patterns [16]; **(2)** 3 random patterns; **(3)** 1 sinusoidal, 1 triangular, 1 random pattern; **(4)** 11 Gray codes; **(5)** 2 phase-shifted sinusoidals and 1 ramp with low-cost projector; and **(6)** 11 Gray codes with low-cost projector. We display mean absolute error (MAE) and root mean squared error (RMSE) against the CAD model of the pawn for LCD projector and against dense SL for DLP projector.

4.2 Ablation Study

We conduct an ablation study of different components of *NeLU3D* using a milled pawn (approx. 10cm $\times$ 16cm $\times$ 10cm) coated with a glossy paint (Fig. 7). We capture $k = 3$ patterns (1 sinusoidal, 1 triangular, and 1 random – the same ones from the third column of Fig. 6) and the white flash with the DLP projector and Atlas camera. In all reconstructions, the mask, eikonal, and sparsity losses remain enabled. First, we disable the ambient MLP a and reconstruct the scene using only $\mathcal{L}_{\text{LookUp}}$. The approach fails to accurately model the captured images, and darker pixels and global illumination corrupt the reconstructed surface. With only $\mathcal{L}_{\text{pattern}}$ instead, we mitigate noise in darker regions but the model still

Table 2: Comparison of memory (in GB of VRAM of an RTX A6000 GPU) and time (in seconds) for outputting a 3D reconstruction for three previous methods and ours for $k = 3, 6, 11$ patterns in two different scenes. Our method requires less time and less GPU memory than the state-of-the-art neural inverse structured light TurboSL [34].

		pawn			monkey		
		3	6	11	3	6	11
ZNCC [33]	VRAM (GB)	0.1	0.2	0.2	0.1	0.2	0.3
	Time (s)	0.0022	0.0022	0.0041	0.0997	0.1000	0.1008
LookUp3D [38]	VRAM (GB)	10.2	17.0	29.2	10.2	17.0	29.2
	Time (s)	0.0016	0.0016	0.0022	0.0022	0.0018	0.0025
TurboSL [34]	VRAM (GB)	4.3	4.6	5.2	4.3	4.6	5.2
	Time (s)	1020	1560	2364	1160	1650	2468
NeLU3D	VRAM (GB)	3.7	4.0	4.4	3.7	4.0	4.4
	Time (s)	420	423	431	552	562	534

fails to overcome global illumination effects. However, adding the MLP a but not enforcing $\mathbf{I}_{\text{ambient}} \leqslant \mathbf{I}_{\text{white}}$ causes the optimization to break. Because we calculate $\mathbf{I}_{\text{white}} - \mathbf{I}_{\text{ambient}}$, if this value becomes negative, it causes the gradients to explode during backpropagation. To prevent that from happen, we use a sigmoid activation to enforce ambient network outputs between 0 and 1 and multiply it, pixelwise, by $\mathbf{I}_{\text{white}}$. This way the network always obeys the constrain $\mathbf{I}_{\text{ambient}} \leqslant \mathbf{I}_{\text{white}}$. This leads to our full model, with the pawn reconstructed surface close to the CAD model and the lowest MAE. For further experiments on $\alpha \in [0, 1]$, which balances $\mathcal{L}_{\text{LookUp}}$ and $\mathcal{L}_{\text{pattern}}$, check supplemental Sec. B.4. We also study the impact of more patterns (Sec. B.2), different patterns (Sec. B.3), and exposure time (Sec. B.5) on reconstruction quality in our supplemental document.

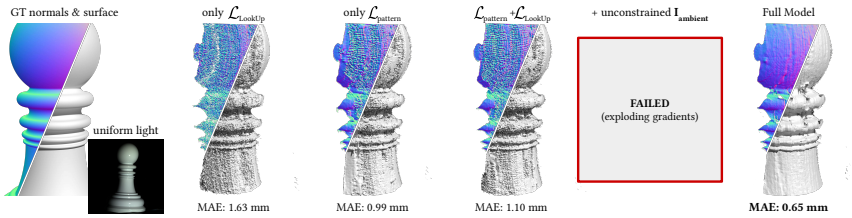

Fig. 7: Ablation Study. We assess the impact on reconstruction error (the mean absolute error is measured against the CAD model of the pawn) of different components of our proposed model. Without the ambient term $\mathbf{I}_{\text{ambient}}$ and with only the LookUp loss $\mathcal{L}_{\text{LookUp}}$, the final surface is extremely noisy. With the pattern loss $\mathcal{L}_{\text{pattern}}$, instead, the final surface of the pawn is smoother and closer to the ground truth, but far from ideal. By using both $\mathcal{L}_{\text{LookUp}}$ and $\mathcal{L}_{\text{pattern}}$, while enforcing $\mathbf{I}_{\text{ambient}} \leqslant \mathbf{I}_{\text{white}}$, we get our full model and the best reconstruction results of the milled pawn object. The ambient term $\mathbf{I}_{\text{ambient}}$ is necessary to compensate for global illumination.

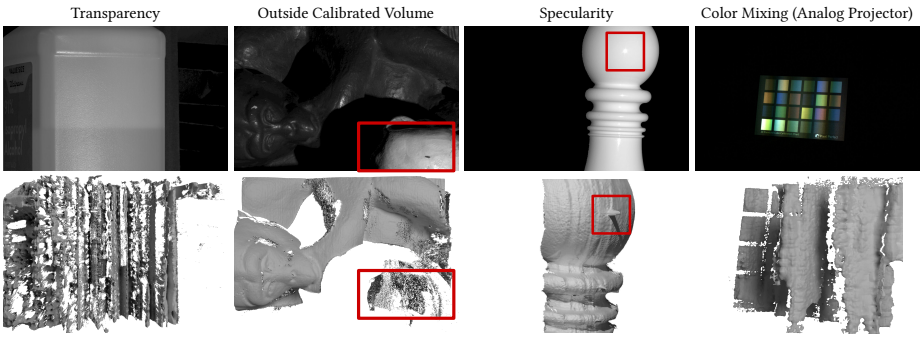

Fig. 8: Failures of *NeLU3D*. Our proposed method has limitations that are typical to other SL methods. *NeLU3D* also gets restricted to the calibrated volume, just like LookUp3D [38]. We have also observed that, with the analog projector and a trichromatic camera, there is cross-talk between different color channels in the camera sensor and the projected RGB pattern developed on film, making the reconstruction fail.

4.3 Limitations

In this section and in Fig. 8, we discuss some of the limitations of our proposed method. As this project expands upon Pereira *et al.* [38], scanning is restricted to the volume of calibration (*i.e.* it cannot extrapolate what the lookup tables are outside of the calibrated region). For example, the skull the monkey holds was outside the volume of calibration; its surface, therefore, was not accurately reconstructed. Similar to many other SL approaches, we cannot scan transparent, translucent, and mirror surfaces. Additionally, the ambient term is not able to fully learn specularities, as seen in the pawn reconstruction - likely due to sensor overexposure and loss of information in those pixels.

Additionally, each new set of patterns requires calibrating the color mapping again. If one would like to switch, for example, the film pattern in the analog projector, one has to calibrate the lookup table again and retrain the color MLP. Finally, our implementation, similar to [34, 44], requires around 10 minutes of training the SDF and ambient MLPs to output a single frame – these systems are therefore not yet equipped for real-time acquisition.

5 Conclusion

We proposed a neural inverse rendering method for structured light without the need to explicitly model the projector, building upon two recent ideas in 3D scanning with SL (TurboSL [34] and LookUp3D [38]). This simplifies the scanning pipeline since there is no need to calibrate the intrinsics and extrinsic parameters of the projector or modeling/learning any effect associated with imperfect optics, such as blurring, defocusing, and vignetting. We showed our system is capable of capturing accurate surface normals and geometry for both

static and dynamic objects with a variety of camera-projector configurations, including a low-cost, hard-to-calibrate projector and analog projector built with a simple, bright LED flashing at kilohertz frequency for slow-motion scanning.

We thank the authors of TurboSL and LookUp3D for providing their source code and hardware specifications, as this accelerated our research for this project. In the same spirit and to continue fostering reproducibility and adoption of these SL techniques, we released an open-source implementation of our approach[1].

5.1 Future Work

The calibration using the planar board with fiduciary markers allows skipping the explicit modeling of the projector. Nonetheless, we are still required to calibrate the camera intrinsics. For future work, we are interested in also bypassing the camera calibration, accumulating all the different calibrations into a single linear stage sweep of the calibration board. We expect that this would also enable high-accuracy 3D reconstructions while preventing limitations on the camera side.

Since we rely on two captures (one RGB pattern and one white flash) for dynamic scenes with an analog projector, fast motion can cause errors (see, *e.g.*, the right corner of the bunny's ear in Fig. 1). Similar to what Urakawa and Watanabe [44] achieved in their work, we can explore a neural network to model a displacement field between the pattern capture and the flash capture. It is an interesting avenue to potentially mitigate errors caused by extreme motion.

Additionally, captures at extremely low exposure (*e.g.* 1/1000s) can suffer from severe noise (dark noise, photon noise, read noise, *etc.*), as we show in our supplementary material (Sec. B.5). Another future research direction is to incorporate noise into the image formation model to better reconstruct surfaces.

Acknowledgments

This work was partially supported by NSF grants OAC-2411349 and OAC-2411221 and the NYU IT High Performance Computing resources, services, and staff expertise.

[1] `https://github.com/geometryprocessing/neural-lookup`

References

1. Berger, M., Tagliasacchi, A., Seversky, L.M., Alliez, P., Levine, J.A., Sharf, A., Silva, C.T.: State of the art in surface reconstruction from point clouds. In: Lefebvre, S., Spagnuolo, M. (eds.) 35th Annual Conference of the European Association for Computer Graphics, Eurographics 2014 - State of the Art Reports, Strasbourg, France, April 7-11, 2014. pp. 161–185. Eurographics Association (2014). https://doi.org/10.2312/EGST.20141040, https://doi.org/10.2312/egst.20141040
2. Boyer, K.L., Kak, A.C.: Color-Encoded Structured Light for Rapid Active Ranging. IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-9**(1), 14–28 (Jan 1987). https://doi.org/10.1109/TPAMI.1987.4767869, https://ieeexplore.ieee.org/document/4767869
3. Carrihill, B., Hummel, R.: Experiments with the intensity ratio depth sensor. Computer Vision, Graphics, and Image Processing **32**(3), 337–358 (Dec 1985). https://doi.org/10.1016/0734-189X(85)90056-8, https://www.sciencedirect.com/science/article/pii/0734189X85900568
4. Caspi, D., Kiryati, N., Shamir, J.: Range imaging with adaptive color structured light. IEEE Transactions on Pattern Analysis and Machine Intelligence **20**(5), 470–480 (May 1998). https://doi.org/10.1109/34.682177, http://ieeexplore.ieee.org/document/682177/
5. Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H.: Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14124–14133 (2021)
6. Chen, W., Mirdehghan, P., Fidler, S., Kutulakos, K.N.: Auto-tuning structured light by optical stochastic gradient descent. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
7. Cossairt, O., Matsuda, N., Gupta, M.: Robust 3d acquisition using motion contrast 3d scanning. In: Imaging and Applied Optics 2015. p. CT2E.1. COSI, OSA (2015). https://doi.org/10.1364/cosi.2015.ct2e.1, http://dx.doi.org/10.1364/COSI.2015.CT2E.1
8. Drouin, M.A., Blais, F., Godin, G.: High resolution projector for 3d imaging. In: 2014 2nd International Conference on 3D Vision. vol. 1, pp. 337–344 (2014). https://doi.org/10.1109/3DV.2014.86
9. Guo, K., Lincoln, P., Davidson, P., Busch, J., Yu, X., Whalen, M., Harvey, G., Orts-Escolano, S., Pandey, R., Dourgarian, J., Tang, D., Tkach, A., Kowdle, A., Cooper, E., Dou, M., Fanello, S., Fyffe, G., Rhemann, C., Taylor, J., Debevec, P., Izadi, S.: The relightables: volumetric performance capture of humans with realistic relighting. ACM Trans. Graph. **38**(6) (Nov 2019). https://doi.org/10.1145/3355089.3356571, https://doi.org/10.1145/3355089.3356571
10. Gupta, M., Nayar, S.K.: Micro Phase Shifting. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 813–820. IEEE, Providence, RI (Jun 2012). https://doi.org/10.1109/CVPR.2012.6247753, http://ieeexplore.ieee.org/document/6247753/
11. Gupta, M., Nakhate, N.: A geometric perspective on structured light coding. In: Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XVI. p. 90–107. Springer-Verlag, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-030-01270-0_6, https://doi.org/10.1007/978-3-030-01270-0_6

12. Gupta, M., Nayar, S.K.: Micro phase shifting. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 813–820 (2012). https://doi.org/10.1109/CVPR.2012.6247753
13. Hall-Holt, O., Rusinkiewicz, S.: Stripe boundary codes for real-time structured-light range scanning of moving objects. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 2, pp. 359–366. IEEE Comput. Soc, Vancouver, BC, Canada (2001). https://doi.org/10.1109/ICCV.2001.937648, http://ieeexplore.ieee.org/document/937648/
14. Hertzmann, A., Seitz, S.: Example-based photometric stereo: shape reconstruction with general, varying brdfs. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(8), 1254–1264 (2005). https://doi.org/10.1109/TPAMI.2005.158
15. Horn, B.K.P.: Height and gradient from shading. International Journal of Computer Vision **5**(1), 37–75 (Aug 1990). https://doi.org/10.1007/bf00056771, http://dx.doi.org/10.1007/BF00056771
16. Horn, E., Kiryati, N.: Toward optimal structured light patterns1based on "toward optimal structured light patterns" by e. horn and n. kiryati which appeared in proceedings of 3-d digital imaging and modelling, ottawa, may 97; (pp. 28—35). © 1999 ieee.1. Image and Vision Computing **17**(2), 87–97 (1999). https://doi.org/https://doi.org/10.1016/S0262-8856(98)00113-9, https://www.sciencedirect.com/science/article/pii/S0262885698001139
17. Hu, P., Yang, S., Zhang, Guofeng fand Deng, H.: High-speed and accurate 3d shape measurement using dic-assisted phase matching and triple-scanning. Optics and Lasers in Engineering **147**, 106725 (Dec 2021). https://doi.org/10.1016/j.optlaseng.2021.106725, http://dx.doi.org/10.1016/j.optlaseng.2021.106725
18. Huang, X., Zhang, Y., Xiong, Z.: High-speed structured light based 3d scanning using an event camera. Optics Express **29**(22), 35864 (Oct 2021). https://doi.org/10.1364/oe.437944, http://dx.doi.org/10.1364/OE.437944
19. Ikeuchi, K., Horn, B.K.: Numerical shape from shading and occluding boundaries. Artificial Intelligence **17**(1), 141–184 (1981). https://doi.org/https://doi.org/10.1016/0004-3702(81)90023-0, https://www.sciencedirect.com/science/article/pii/0004370281900230
20. Ikeuchi, K., Oishi, T., Takamatsu, J., Sagawa, R., Nakazawa, A., Kurazume, R., Nishino, K., Kamakura, M., Okamoto, Y.: The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. Int. J. Comput. Vision **75**(1), 189–208 (Oct 2007). https://doi.org/10.1007/s11263-007-0039-y, https://doi.org/10.1007/s11263-007-0039-y
21. Inokuchi, S., Sata, K., Matsuda, F.: Range imaging system for 3-d object recognition. Proceedings of the International Conference on Pattern Recognition 1984 pp. 806–808 (1984)
22. Je, C., Lee, S.W., Park, R.H.: Colour-stripe permutation pattern for rapid structured-light range imaging. Optics Communications **285**(9), 2320–2331 (May 2012). https://doi.org/10.1016/j.optcom.2012.01.025, https://linkinghub.elsevier.com/retrieve/pii/S003040181200048X
23. Jones, A., Gardner, A., Bolas, M., Mcdowall, I., Debevec, P.: Simulating spatially varying lighting on a live performance. In: The 3rd European Conference on Visual Media Production (CVMP 2006) - Part of the 2nd Multimedia Conference 2006. pp. 127–133 (2006). https://doi.org/10.1049/cp:20061934
24. Jones, A., Lang, M., Fyffe, G., Yu, X., Busch, J., McDowall, I., Bolas, M., Debevec, P.: Achieving eye contact in a one-to-many 3d video teleconferencing system. ACM

Trans. Graph. **28**(3) (Jul 2009). https://doi.org/10.1145/1531326.1531370, https://doi.org/10.1145/1531326.1531370
25. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics **42**(4) (July 2023), https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/
26. Koch, S., Piadyk, Y., Worchel, M., Alexa, M., Silva, C., Zorin, D., Panozzo, D.: Hardware design and accurate simulation of structured-light scanning for benchmarking of 3d reconstruction algorithms. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2) (2021), https://openreview.net/forum?id=bNL5VlTfe3p
27. Koppal, S.J., Yamazaki, S., Narasimhan, S.G.: Exploiting dlp illumination dithering for reconstruction and photography of high-speed scenes. International Journal of Computer Vision **96**(1), 125 – 144 (January 2012)
28. Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., Ginzton, M., Anderson, S., Davis, J., Ginsberg, J., Shade, J., Fulk, D.: The digital michelangelo project: 3d scanning of large statues. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. p. 131–144. SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., USA (2000). https://doi.org/10.1145/344779.344849, https://doi.org/10.1145/344779.344849
29. Li, Z., Muller, T., Evans, A., Taylor, R., Unberath, M., Liu, M., Lin, C.: Neuralangelo: High-fidelity neural surface reconstruction. In: Proceedings - 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023. pp. 8456–8465. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society (2023). https://doi.org/10.1109/CVPR52729.2023.00817, publisher Copyright: © 2023 IEEE.; 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023 ; Conference date: 18-06-2023 Through 22-06-2023
30. Ma, W.C., Hawkins, T., Peers, P., Chabert, C.F., Weiss, M., Debevec, P.: Rapid acquisition of specular and diffuse normal maps from polarized spherical gradient illumination. In: Proceedings of the 18th Eurographics Conference on Rendering Techniques. p. 183–194. EGSR'07, Eurographics Association, Goslar, DEU (2007)
31. Martin, J., Crowley, J.L.: Comparison of correlation techniques. In: Intelligent Autonomous Systems. pp. 86–93 (1995)
32. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: ECCV (2020)
33. Mirdehghan, P., Chen, W., Kutulakos, K.N.: Optimal structured light à la carte. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
34. Mirdehghan, P., Wu, M., Chen, W., Lindell, D.B., Kutulakos, K.N.: Turbosl: Dense accurate and fast 3d by neural inverse structured light. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 25067–25076 (June 2024)
35. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. **41**(4) (Jul 2022). https://doi.org/10.1145/3528223.3530127, https://doi.org/10.1145/3528223.3530127
36. Nayar, S.K., Ikeuchi, K., Kanade, T.: Shape from interreflections. International Journal of Computer Vision **6**(3), 173–195 (Aug 1991). https://doi.org/10.1007/bf00115695, http://dx.doi.org/10.1007/BF00115695

37. Oechsle, M., Peng, S., Geiger, A.: Unisurf: Unifying neural implicit surfaces and radiance fields for multi-view reconstruction. In: International Conference on Computer Vision (ICCV) (2021)
38. Pereira, G., Gao, Y., Piadyk, Y., Fouhey, D., Silva, C.T., Panozzo, D.: Lookup3d: Data-driven 3d scanning. In: Proceedings of the SIGGRAPH Asia 2025 Conference Papers. SA Conference Papers '25, Association for Computing Machinery, New York, NY, USA (2025). https://doi.org/10.1145/3757377.3763986, https://doi.org/10.1145/3757377.3763986
39. Posdamer, J., Altschuler, M.: Surface measurement by space-encoded projected beam systems. Computer Graphics and Image Processing **18**(1), 1–17 (Jan 1982). https://doi.org/10.1016/0146-664X(82)90096-X, https://linkinghub.elsevier.com/retrieve/pii/0146664X8290096X
40. Salvi, J., Pagès, J., Batlle, J.: Pattern codification strategies in structured light systems. Pattern Recognition **37**(4), 827–849 (Apr 2004). https://doi.org/10.1016/j.patcog.2003.10.002, https://linkinghub.elsevier.com/retrieve/pii/S0031320303003303
41. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. p. 195–202. CVPR'03, IEEE Computer Society, USA (2003)
42. Shandilya, A., Attal, B., Richardt, C., Tompkin, J., O'Toole, M.: Neural fields for structured lighting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3512–3522 (2023)
43. TABATA, S., NOGUCHI, S., WATANABE, Y., ISHIKAWA, M.: High-speed 3d sensing with three-view geometry using a segmented pattern. Transactions of the Society of Instrument and Control Engineers **52**(3), 141–151 (2016). https://doi.org/10.9746/sicetr.52.141, http://dx.doi.org/10.9746/sicetr.52.141
44. Urakawa, Y., Watanabe, Y.: Neural inverse rendering for high-accuracy 3d measurement of moving objects with fewer phase-shifting patterns. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 27692–27701 (October 2025)
45. Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W.: Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. NeurIPS (2021)
46. Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L.: Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023)
47. Wissmann, P., Forster, F., Schmitt, R.: Fast and low-cost structured light pattern sequence projection. Opt. Express **19**(24), 24657–24671 (Nov 2011). https://doi.org/10.1364/OE.19.024657, https://opg.optica.org/oe/abstract.cfm?URI=oe-19-24-24657
48. Woodham, R.J.: Photometric method for determining surface orientation from multiple images. Optical Engineering **19**(1) (Feb 1980). https://doi.org/10.1117/12.7972479, http://dx.doi.org/10.1117/12.7972479
49. Wust, C., Capson, D.W.: Surface profile measurement using color fringe projection. Machine Vision and Applications **4**(3), 193–203 (Jun 1991). https://doi.org/10.1007/BF01230201, http://link.springer.com/10.1007/BF01230201
50. Yariv, L., Gu, J., Kasten, Y., Lipman, Y.: Volume rendering of neural implicit surfaces. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W.

(eds.) Advances in Neural Information Processing Systems. vol. 34, pp. 4805–4815. Curran Associates, Inc. (2021), https://proceedings.neurips.cc/paper_files/paper/2021/file/25e2a30f44898b9f3e978b1786dcd85c-Paper.pdf

51. Yariv, L., Kasten, Y., Moran, D., Galun, M., Atzmon, M., Ronen, B., Lipman, Y.: Multiview neural surface reconstruction by disentangling geometry and appearance. Advances in Neural Information Processing Systems **33** (2020)
52. Zhang, S.: High-speed 3D shape measurement with structured light methods: A review. Optics and Lasers in Engineering **106**, 119–131 (Jul 2018). https://doi.org/10.1016/j.optlaseng.2018.02.017, https://linkinghub.elsevier.com/retrieve/pii/S0143816617313246
53. Zhang, S., Van Der Weide, D., Oliver, J.: Superfast phase-shifting method for 3-D shape measurement. Optics Express **18**(9), 9684 (Apr 2010). https://doi.org/10.1364/OE.18.009684, https://opg.optica.org/abstract.cfm?URI=oe-18-9-9684