Neural SDF for Shadow-aware Unsupervised Structured Light

Kazuto Ichimaru Diego Thomas Takafumi Iwaguchi Hiroshi Kawasaki
Kyushu University, Japan
https://www.cvg.ait.kyushu-u.ac.jp/index.html

Abstract

Among various active 3D measurement techniques, Structured Light (SL) is one of the most popular methods for its robustness and high accuracy. The ordinary SL system consists of a camera and a projector, and by projecting a pre-defined pattern, we can obtain pixel-to-pixel correspondences between the camera and the projector for triangulation. However, if we lack knowledge of the projected pattern for some reason, e.g.*, the projected pattern is not as expected due to lens distortion, inaccurate calibration, undesired optical phenomena like inter-reflection, and so on, the accuracy of conventional SL is severely degraded. As a remedy, we propose* ***unsupervised structured light (USSL)****, which does not explicitly use prior knowledge of the pattern. Inspired by the fact that humans can recognize the scene structure illuminated by an unknown light source (*e.g. *rotating mirror ball), and some prior works have succeeded in novel-view-synthesis under unknown illumination conditions, we implement USSL on Neural Signed Distance Fields (Neural SDF) pipeline with implicit reflection module powered by a neural network. Additionally, since every SL method causes occlusion (shadow) by pattern projection, we must consider it for accurate shape reconstruction. To this end, we integrate shadow volume rendering into the proposed pipeline. Experiments with synthetic and real datasets are conducted to confirm the feasibility of the proposed method.*

1. Introduction

For decades, active 3D measurement has been extensively utilized in various fields, including autonomous vehicle control, human body analysis, and industrial inspection. Among these techniques, active stereo methods, particularly structured light (SL), are favored for their simple configuration and high accuracy, especially where ordinary depth sensor does not work due to extreme environmental conditions (*e.g.* underwater). A typical SL system comprises a camera and a pattern projector. The projector emits predefined patterns, which are captured by the camera to establish pixel-to-pixel correspondences between the camera and the projector. These correspondences are then employed in a stereo triangulation method known as "light-sectioning" to determine scene depth. Utilizing an active light source as a virtual viewpoint, the SL system excels in challenging conditions such as dark environments or sparse-view sequences.

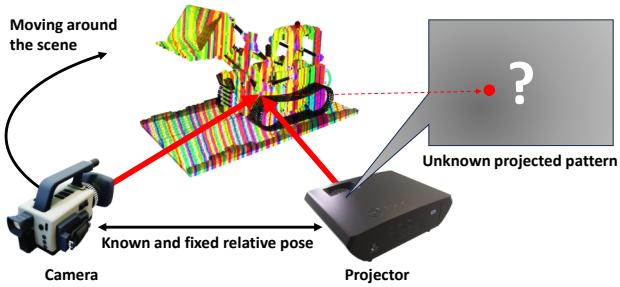

Figure 1. Problem statement and assumed system configuration of the proposed method. The camera and the projector move together, meaning a simple stereo-matching approach will not work. Note that the relative pose between the projector and camera has some flexibility, up to a homography transformation.

Certain SL methods can reconstruct shapes from a single image, known as "one-shot" techniques, making them ideal for dynamic applications like drones or endoscopes. However, these methods often necessitate complex and specialized algorithms or require pattern calibration with a known target for decoding. Additionally, some SL methods adopt a pattern-agnostic correspondence search strategy similar to stereo matching, which can struggle with matching ambiguity or distributional shifts. If it is possible to reconstruct the shape of a scene without prior pattern information, SL algorithms will apply to a wide range of optical systems regardless of projector specifications, including special projection devices like diffractive optical elements, laser, etc.

Recently, Neural Radiance Fields (NeRF) [21] and Neural Signed Distance Fields (Neural SDF) [34, 39] have drawn significant attention for their robustness and versatility across a wide range of applications. These methods have achieved state-of-the-art quality in novel view synthesis and 3D shape reconstruction by addressing the inverse

rendering problem. Additionally, some NeRF / Neural SDF methods are also capable of estimating unknown parameters such as the camera's intrinsic and extrinsic parameters [12, 17, 26, 36, 40], or scene illuminations [38, 41]. Inspired by this progress, we believe it is feasible to acquire an accurate scene shape *without prior knowledge of the projected pattern*. Through this process, the information of projected patterns is implicitly learned. Therefore, as long as the light source position is fixed, some pose parameters can be compensated up to a homography transformation. This flexible setup greatly aids the actual scanning process, and its efficacy has been confirmed in real experiments. The problem statement and assumed system configuration are illustrated in Figure 1.

To achieve this goal, we propose a novel algorithm for unsupervised structured light (SL) using Neural SDF. Specifically, we model reflection implicitly using a multi-layer perceptron (MLP) that processes a feature vector sampled from multi-resolution hash encoding [23] based on pixel coordinates on the projector screen. Furthermore, to mitigate the undesired effects of crisp shadows often created by active light sources, we integrate a shadow-aware projection module employing shadow volume rendering.

Contributions of the work are as follows:

- We propose a pipeline for neural 3D reconstruction using a structured light without prior knowledge of the projected pattern.
- The proposed pipeline incorporates shadow volume rendering module, which efficiently handles occlusion caused by pattern projection.
- Experimental results show the proposed pipeline improves reconstruction quality in both synthetic and real datasets.

2. Related work

2.1. Structured Light

Structured light (SL) is a methodology used to estimate scene shape through light-sectioning using active light sources. It employs projected patterns to establish correspondences for triangulation. To enhance correspondence accuracy, multiple patterns have been traditionally employed in SL, such as Graycode [10] and Phase-shifting [32]. Recent research has explored optimization techniques to reduce the number of required patterns [3, 22]. For applications involving dynamic scenes, single-pattern approaches, known as one-shot scans, have been favored, including grid patterns [6], random dot patterns [9], hamming patterns [37], colorful line patterns [5], and cross laser patterns [24, 25]. Additionally, some studies have proposed calibration-free SL systems [7], which assume that the relative pose between the camera and projector is unknown, while the pattern itself is known. In general, higher-precision SL systems with more unknown variables tend to be more complex and specialized, making them difficult to apply across a wide range of applications.

Some work utilized projected pattern as virtual texture [14], or edited images to add clues based on sparse LiDAR depth [1], to alleviate correspondence ambiguity on stereo matching. Note that these approaches assume that multiple cameras simultaneously observe patterns projected from the same projector pose, while we assume a projector is moving with a single camera.

2.2. NeRF and Neural SDF

Neural Radiance Fields (NeRF) represent scenes using a volumetric function that outputs density and color for 3D points from multiple viewpoints [21]. NeRF leverages deep neural networks (DNNs) for accurate interpolation and extrapolation, enabling high-quality novel view synthesis and scene shape estimation. Some variants utilize NeRF as a tool for acquiring albedo or re-lighting scenes [2]. However, NeRF's performance in 3D shape reconstruction is limited due to challenges in representing solid surfaces with volumetric density functions.

In contrast, VolSDF and Neural SDF replace volumetric density functions with signed distance functions (SDF), which provide the signed distance to the nearest surface [16, 34, 35, 39]. This approach significantly improves 3D shape reconstruction accuracy by incorporating specific structural biases. While density function-based methods achieve precise reconstruction in some scenarios [27], SDF is well-suited for various downstream geometric tasks, including normal regularization via Eikonal loss [8], shape integration via TSDF [11], and shape editing [33].

2.3. NeRF / Neural SDF for Structured Light

Several studies have employed NeRF or Neural SDF to enhance the accuracy of structured light (SL) based shape reconstruction [13, 15, 30]. For instance, Li *et al*. combined Neural SDF with Graycode for handling inter-reflections, employing a conventional decoding strategy for image-to-pattern correspondence [15]. Shandilya *et al*. integrated pattern projection into a volume rendering pipeline, separately modeling direct and indirect illumination, albeit requiring reference images without pattern projection, which may limit practical applicability [30]. Ichimaru *et al*. proposed a generalized Neural SDF approach accommodating various patterns and system configurations, assuming Lambertian surfaces exclusively [13]. Notably, none of these approaches explicitly addressed occlusion caused by pattern projection within the volume rendering process.

In this work, inspired by [13], we extend these concepts by implicitly modeling complex reflections using a multi-layer perceptron (MLP). Additionally, we explicitly model occlusion (shadows) induced by pattern projection

by computing a differentiable occlusion mask with volume rendering from the projector's viewpoint. Our occlusion model draws inspiration from recent shadow supervision techniques proposed in NeRF-based methods [18]. Although some previous work has employed an MLP to represent visibility fields [31,42], the approximation capability of the MLP is generally insufficient to accurately capture the shadows cast by small studs.

3. Shadow-aware Unsupervised SL

3.1. Method overview

We consider a structured light (SL) system composed of a camera and an arbitrary number of projectors. The relative transformations between the camera and the projectors are known during capture, and the projected patterns remain constant throughout the process. Given that humans can infer the shape of a scene by observing the movement of patterns on surfaces, even without prior knowledge of the pattern (*e.g.* a room illuminated by a rotating mirror ball), it stands to reason that it may also be possible to reconstruct a scene's shape from unknown patterns.

To this end, we propose a Neural SDF pipeline, as depicted in Figure 2. The pipeline consists of four primary modules: the **geometry module**, **pattern feature module**, **shadow volume module**, and **rendering module**. The overall process is as follows:

1. Sample 3D points $\boldsymbol{P}_i$ on the ray cast from the camera's pixel.
2. Compute SDF values $f(\boldsymbol{P}_i)$, features $\boldsymbol{z}_i$, normals $\boldsymbol{n}_i$, and surface color $\boldsymbol{c}_s(\boldsymbol{P}_i, \boldsymbol{v}_i, \boldsymbol{z}_i, \boldsymbol{n}_i)$ in geometry module ($\boldsymbol{v}_i$ is viewing direction).
3. Back-project each $\boldsymbol{P}_i$ onto the projector screen coordinate system to get $\boldsymbol{p}_i^{proj}$, and sample feature vector $\boldsymbol{w}_i$ in pattern feature module.
4. Sample 3D points $\boldsymbol{P}'_{ij}$ on the ray cast from the projector's optical center to each $\boldsymbol{P}_i$.
5. Compute visibility from the projector to each $\boldsymbol{P}'_{ij}$ through volume rendering to get shadow mask s_i in shadow volume module.
6. Compute reflected color $\boldsymbol{c}_r$ using $\boldsymbol{P}_i, \boldsymbol{z}_i, \boldsymbol{n}_i, \boldsymbol{v}_i, \boldsymbol{w}_i, s_i$ and light incoming direction $\boldsymbol{l}_i$, then blend it with $\boldsymbol{c}_s$ to get observed color $\boldsymbol{c}_o(\boldsymbol{P}_i, \boldsymbol{v}_i)$ in rendering module.
7. Render the final color $\boldsymbol{c}$ for the pixel via volume rendering.
8. Optimize parameters in each modules to minimize $\mathcal{L}_{render} = |\boldsymbol{c} - \boldsymbol{c}_{gt}|$, where $\boldsymbol{c}_{gt}$ is the ground-truth color of the pixel.

We elaborate on the processes 4, 5, and 6, which are inherent to the proposed method, in the following sections. For other processes, they are identical to [13].

3.2. Unsupervised pattern projection

A major challenge lies in computing the reflected color $\boldsymbol{c}_r$ from unknown incoming light. When the incoming light is known, the reflected color can be computed using a physical model, such as the micro-facet BRDF, as demonstrated in [20]. However, since we assume the pattern is unknown, we model the reflection implicitly using an MLP that processes a feature vector sampled according to the pixel coordinates on the projector screen.

To address this, we incorporate the pattern feature module and the rendering module into our pipeline. For the pattern feature module, we first back-project each 3D point $\boldsymbol{P}_i$ onto the projector screen as follows:

$$\boldsymbol{p}_i^{proj} = K^{proj}(R^{proj}\boldsymbol{P}_i + \boldsymbol{t}^{proj}), \tag{1}$$

where K^{proj} is the intrinsic matrix and $R^{proj}, \boldsymbol{t}^{proj}$ represent the relative transformation of the projector, converting 3D points in the world coordinate system into the projector screen coordinate system. Subsequently, we employ 2D multi-resolution hash encoding (hashgrid encoding) [23] to sample a feature vector $\boldsymbol{w}_i$. Note that K^{proj} is calibrated so that $\boldsymbol{p}_i^{proj}$ ranges from 0 to 1. Hashgrid encoding is effective in preserving very high-resolution information on a pixel-by-pixel basis, which is why we selected this method.

For the rendering module, we utilize a projection MLP that takes the 3D point $\boldsymbol{P}_i$, normal $\boldsymbol{n}_i$, SDF feature $\boldsymbol{z}_i$, viewing direction $\boldsymbol{v}_i$, light incoming direction $\boldsymbol{l}_i$, and the sampled pattern feature $\boldsymbol{w}_i$ to output the reflected color $\boldsymbol{c}_r(\boldsymbol{P}_i, \boldsymbol{v}_i, \boldsymbol{z}_i, \boldsymbol{n}_i, \boldsymbol{l}_i)$. We anticipate that the MLP will effectively learn the reflection properties of the scene, as these inputs provide sufficient information to represent common BRDFs. Once $\boldsymbol{c}_r$ is obtained, following [13], it is blended with the surface color $\boldsymbol{c}_s$ using learnable coefficients i_s and i_p to derive the observed color $\boldsymbol{c}_o(\boldsymbol{P}_i, \boldsymbol{v}_i)$ as follows:

$$\boldsymbol{c}_o(\boldsymbol{P}_i, \boldsymbol{v}_i) = \boldsymbol{c}_s + (i_s\boldsymbol{c}_s + i_p)\boldsymbol{c}_r. \tag{2}$$

Although the projection MLP itself receives sufficient information to directly output $\boldsymbol{c}_o(\boldsymbol{P}_i, \boldsymbol{v}_i)$, we deliberately adopt a blending strategy based on [4]. This approach, which might appear redundant, introduces an inductive bias that separates the surface color from the projected color, and as a result, reconstruction quality is improved as demonstrated in the ablation study.

3.3. Shadow volume rendering with ray pruning

Another significant challenge is how to represent shadows in Neural SDF. To treat shadows as additional supervision rather than a disturbance, we compute shadows in a fully differentiable manner using volume rendering, similar to the rendering of camera rays. Specifically, we cast rays from the projector's optical center $\boldsymbol{O}_{proj}$ to each 3D point

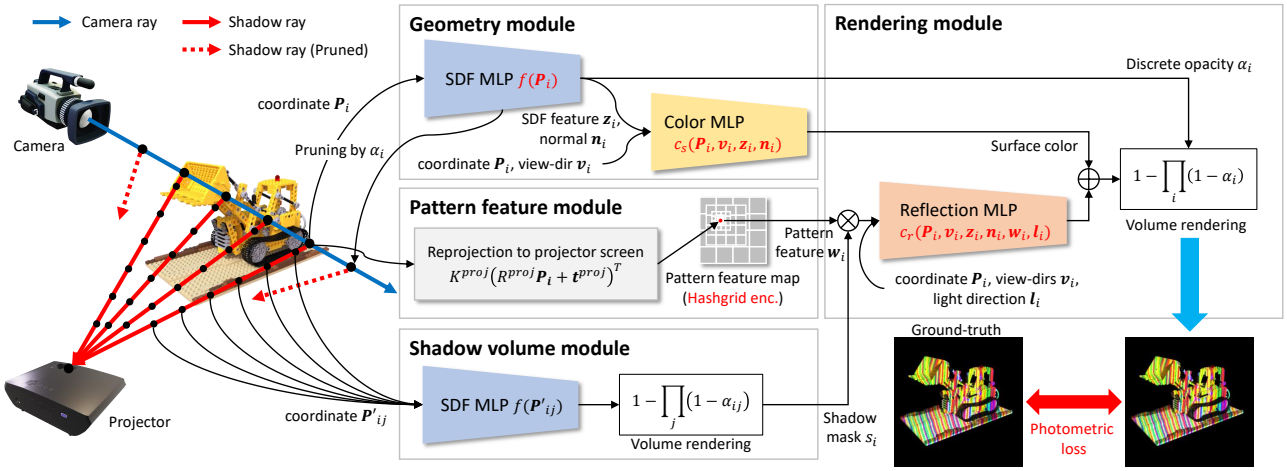

Figure 2. Pipeline of the proposed method. Red texts indicate network parameters are updated during training.

$\boldsymbol{P}_i$ and sample 3D points along the ray as follows:

$$\boldsymbol{P}'_{ij} = \boldsymbol{O}_{proj} + t\frac{\boldsymbol{P}_i - \boldsymbol{O}_{proj}}{|\boldsymbol{P}_i - \boldsymbol{O}_{proj}|}, \tag{3}$$

where $u \leq t \leq 1 - 2s$, and u represents the distance to the intersection of the ray on the unit sphere. We intentionally add a bias of $2s$ to shorten the far clip of the ray, thereby avoiding "Z-fighting" caused by Neural SDF, where s is the standard deviation of the logistic density function. In practice, we employ importance sampling [21] instead of uniform sampling to maximize precision on small structures.

Next, we compute SDF values using the SDF MLP f to obtain $f(\boldsymbol{P}'_{ij})$. Finally, these values are integrated along the ray to compute the accumulated density (shadow mask) s_i, representing the visibility from the projector to each $\boldsymbol{P}_i$, as follows:

$$s(\boldsymbol{P}_i) = 1 - \prod_j (1 - \alpha_{ij}), \tag{4}$$

where α_{ij} is a discrete opacity computed from $f(\boldsymbol{P}'_{ij})$.

However, directly computing shadow volume rendering for each 3D point is computationally expensive. Therefore, we introduce two ray pruning strategies to reduce the total number of $\boldsymbol{P}'_{ij}$. The first pruning strategy omits rays to points with very low density, as the pattern is not observable on such points regardless. Specifically, we set a threshold on the discrete opacity calculated from SDF values, excluding points with discrete opacity smaller than $1e-3$ from shadow computation. The second pruning strategy omits rays if $min_j(f(\boldsymbol{P}'ij)) > 0$, indicating that none of the points $\boldsymbol{P}'_{ij}$ are inside the scene shape (i.e., $\boldsymbol{P}_i$ is completely visible from the projector).

Empirically, these pruning strategies resulted in the omission of approximately 90% of the rays, while the rendered images with the pruning strategies remained indistinguishable from those rendered without any pruning.

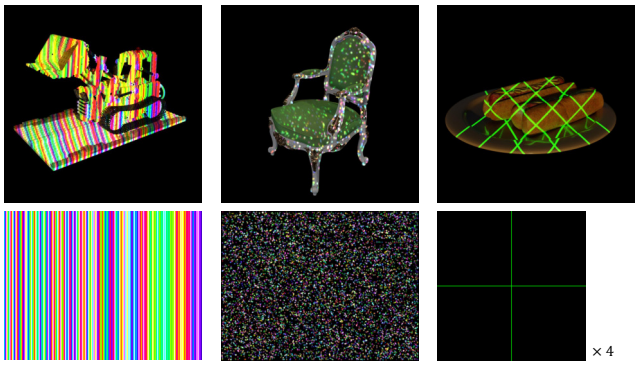

Figure 3. Examples of the dataset images and pattern images. **Left to Right**: Hamming, random dot, cross-lasers. We can see shadows and inter-reflections caused by pattern projection.

3.4. Implementation detail

As for the loss function, we used L1 error $\mathcal{L}_{render}$ for the photometric loss on the rendered pixels. Additionally, we used mask loss $\mathcal{L}_{mask}$ and eikonal loss $\mathcal{L}_{eikonal}$ for better convergence [8]. Finally, the full optimization objective $\mathcal{L}$ of the pipeline is,

$$\mathcal{L} = \mathcal{L}_{render} + \beta\mathcal{L}_{mask} + \gamma\mathcal{L}_{eikonal}, \tag{5}$$

where $\beta = \gamma = 0.1$ in our implementation.

We implemented the proposed method on NeuS [34], and trained the networks for 300k steps in every experiment, which took approximately 12 hours on a single RTX A6000.

Table 1. Comparison results on a synthetic dataset. **Bold**: Ours superior to active stereo-based methods. Underline: Ours superior to passive stereo-based methods. N/A indicates no shape reconstructed.

Pattern	Method	40 views					10 views				
		Lego	Chair	Hotdog	Mic	Average	Lego	Chair	Hotdog	Mic	Average
No	COLMAP	8.39	7.98	7.91	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	NeuS	6.91	6.55	7.60	6.56	6.91	7.81	6.53	7.52	6.87	7.18
Random dot	Light-sectioning	8.58	6.39	7.00	10.08	8.01	9.78	8.01	9.04	10.02	9.21
	ActiveNeuS	7.89	7.87	8.64	6.78	7.80	7.59	6.89	7.29	6.42	7.04
	USSL (Ours)	**6.47**	**5.97**	**6.67**	**6.20**	**6.33**	**7.51**	**6.36**	7.58	6.61	**7.02**
Hamming	Light-sectioning	10.24	8.33	9.63	11.33	9.88	11.61	9.56	11.40	10.34	10.73
	ActiveNeuS	7.17	6.94	7.73	6.74	7.15	7.40	6.72	7.29	6.70	7.03
	USSL (Ours)	**6.84**	**6.30**	**6.71**	**6.39**	**6.56**	7.43	**6.37**	**6.59**	**6.36**	**6.69**
Cross-lasers	Light-sectioning	10.07	9.06	9.45	9.45	9.51	11.18	10.16	10.91	10.13	10.60
	ActiveNeuS	7.15	7.03	8.14	6.50	7.20	7.45	6.77	7.49	6.49	7.05
	USSL (Ours)	**6.95**	**6.31**	**6.75**	**6.19**	**6.55**	7.50	**6.41**	**6.74**	6.56	**6.80**

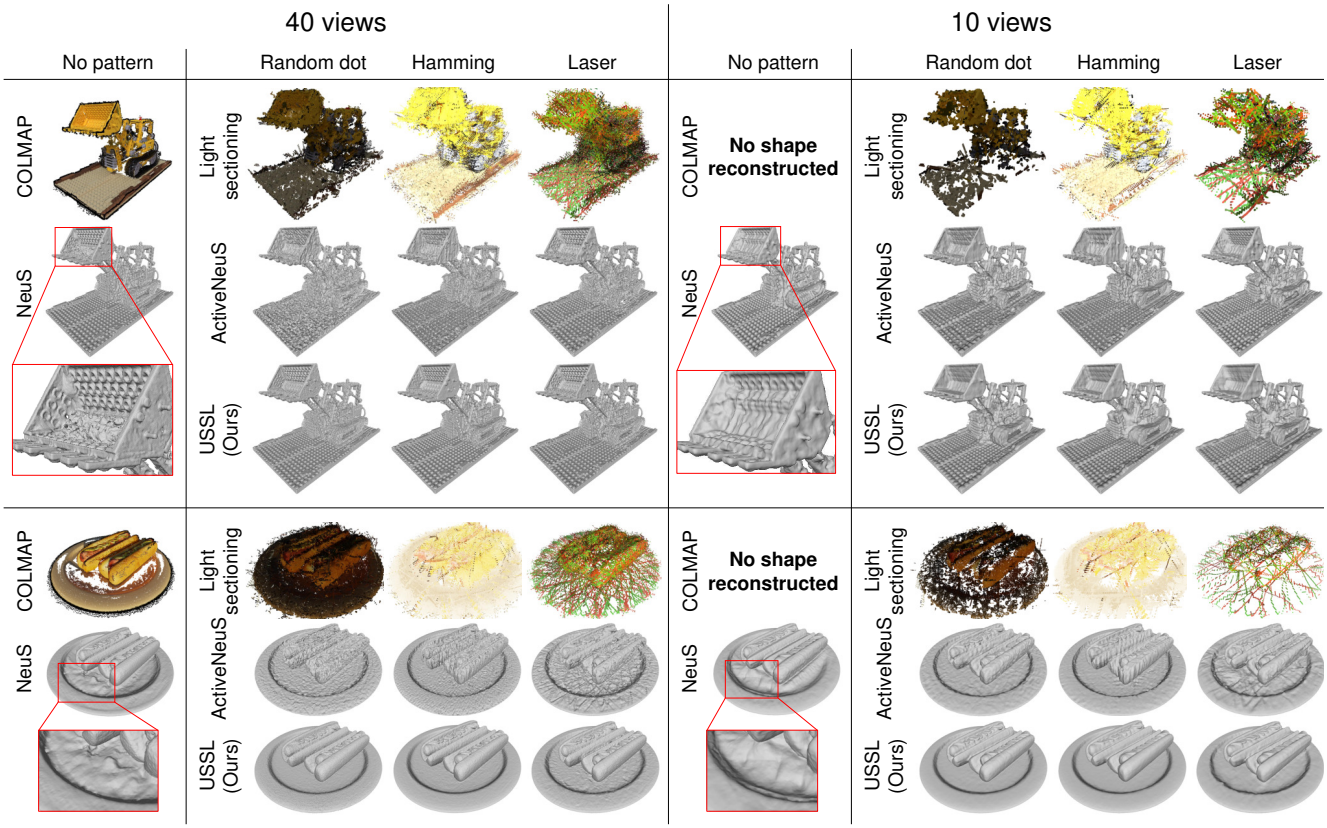

Figure 4. Qualitative comparison results on a synthetic dataset.

4. Experiments

4.1. Evaluation with synthetic dataset

To confirm the feasibility of the proposed method, we conducted several evaluations using a synthetic dataset. We employed the NeRF-Synthetic dataset [21], which includes the Lego, Chair, Hotdog, and Mic scenes from the original Blender files. Images were rendered with Cycles from various viewpoints with pattern projection. For the projected patterns, we used random dot [9], hamming [37], and cross-

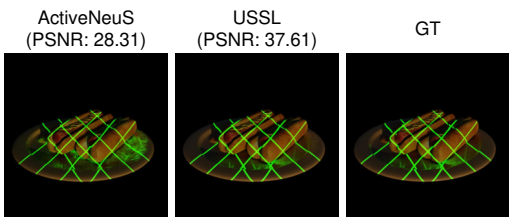

Figure 5. Example rendering images with severe inter-reflection.

lasers [24], as shown in Figure 3.

First, we compared the shape reconstruction accuracy with several conventional methods. For passive stereo based methods, we used COLMAP [28,29] (MVS) and NeuS [34], and for active stereo based methods, we used conventional light-sectioning with the projected patterns [9, 24, 37], and ActiveNeuS [13] (We used RAFT-Stereo [19] as a stereo matching engine). We used 40 images (*i.e.* 40 viewpoints) for each method, but to evaluate the robustness of the methods under a smaller number of viewpoints, we also ran the same evaluations with 10 images. For 40 images, we utilized hashgrid encoding within the SDF MLP to achieve higher accuracy and fidelity. This approach was also applied in NeuS to ensure a fair comparison, which is substantially equivalent to Neuralangelo without curvature loss [16]. As for the evaluation metrics, we used Chamfer distance. Note, since COLMAP naturally has scale ambiguity, we applied Iterative Closest Point (ICP) with the scale on COLMAP results to align with the ground-truth shapes.

Table 1 shows the quantitative comparison results. We can see the proposed method (USSL) outperforms in most scenes in both 40 views and 10 views. Importantly, USSL achieved higher accuracy compared to NeuS, which means USSL does not only learn the information of the projected pattern and reflection model but also utilizes spatial consistency of the pattern feature to improve shape reconstruction accuracy. In other words, projecting some pattern is valuable for accurate shape reconstruction, even if the pattern is unknown. Additionally, we show qualitative comparison results on Figure 4. We can observe bumpy reconstructions in NeuS and ActiveNeuS, while they are successfully compensated in USSL.

We also observed the presence of line-shaped artifacts reconstructed by ActiveNeuS in the hotdog scene (laser). We believe these artifacts are caused by inter-reflection, as our method implicitly handles pattern projection, whereas ActiveNeuS does not. To investigate further, we rendered images using the trained models (Figure 5). As shown, ActiveNeuS suffers from severe artifacts, likely due to inter-reflection, while USSL achieves accurate rendering, even capturing the reflected lasers on the dish. This demonstrates that USSL is more robust against inter-reflection, thanks to its implicit representation.

Second, to confirm each module in the proposed pipeline

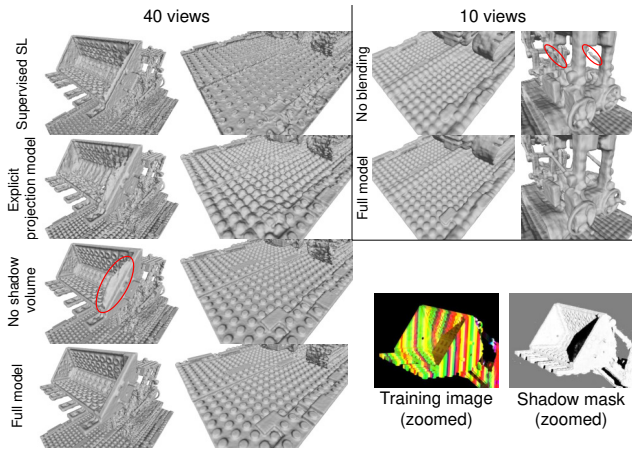

Figure 6. Qualitative comparison of reconstructed shapes on the ablation study. For no shadow volume result, we can see an extra wall reconstructed on the bucket side. It is due to shadow caused by pattern projection, as shown in the training image.

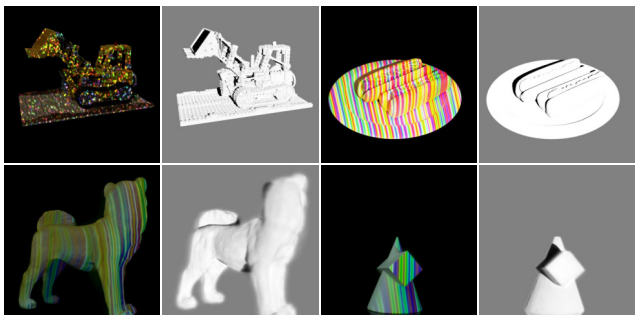

Figure 7. Example images of computed shadow masks with the proposed method. **Top**: Synthetic dataset. **Bottom**: Real dataset.

Table 2. Quantitative comparison of reconstructed shapes on the ablation study. **Bold**: Best. Our full model achieved the best accuracy in almost all cases.

# of views	Method	Lego	Chair	Hotdog	Mic	Average
40	Supervised SL	6.75	6.54	7.47	6.58	6.84
	Explicit projection model	6.56	6.40	6.75	6.50	6.55
	No shadow volume	6.84	6.69	**6.69**	6.43	6.66
	Full model	**6.38**	**6.29**	6.72	**6.35**	**6.44**
10	No blending	7.56	6.41	7.24	**6.30**	6.88
	Full model	**7.43**	**6.37**	**6.59**	6.36	**6.69**

contributed to the reconstruction accuracy, we conducted an ablation study on Hamming scenes. Specifically, we removed or replaced the following modules:

1. **Supervised SL** Instead of pattern feature, RGB-values of the projected pattern are inputted to the projection MLP.
2. **Explicit projection model** Projection MLP is replaced with a conventional reflection model with diffuse component and ambient component. We also inserted another MLP layer, which takes pattern feature as input to predict

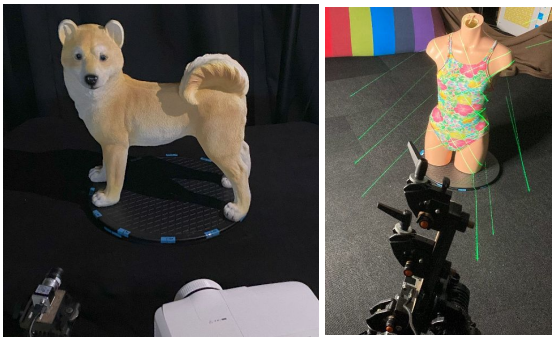

Figure 8. Experimental rigs to capture the real dataset.

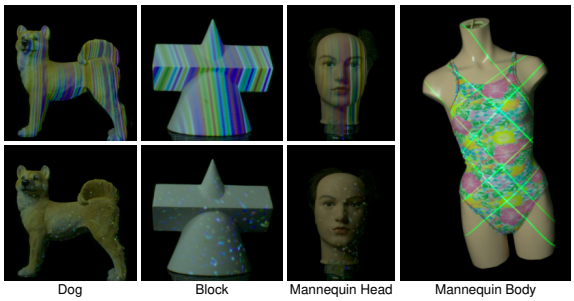

Figure 9. Example images of the real dataset.

RGB-values as incoming light color.

3. **No shadow volume** Shadow volume module is removed, *i.e.*, occlusion caused by pattern projection is simply ignored.
4. **No blending** Blending mechanism introduced in subsection 3.2 is removed, *i.e.*, the projection MLP directly computes observed colors for each 3D point.

Figure 6 and Table 2 present the results of the ablation study. As demonstrated in these figures, both "Supervised SL" and "Explicit projection model" generated numerous artifacts due to their inability to handle unexpected optical phenomena such as inter-reflection. The "No shadow volume" configuration achieved almost identical accuracy to our full model, except in areas where the projected pattern created shadows. Shadows computed by the proposed method are shown in Figure 7, demonstrating that the occlusion by pattern projection is successfully handled. For "No blending" with 10 viewpoints, we observed that its fidelity was inferior to that of our full model.

4.2. Evaluation with real dataset

To evaluate the robustness of the proposed method in real-world scenes, we conducted additional experiments using a real dataset. Specifically, we used two SL systems:

- A camera and a video projector were fixed on a platform, while the measurement target was placed on a rotating table to capture images from various directions. We captured a ceramic dog, a cone-shaped block, and a mannequin head using two patterns: random dot and hamming (Figure 9 left 3 columns).
- A camera and four cross-laser projectors (wavelength: $520nm$) were fixed on a platform. The subsequent steps were the same as above. We captured a mannequin body (Figure 9 rightmost column).

For all scenes, we captured 36 images (36 viewpoints) and ran evaluations with 36 images and 12 images. To calibrate the system poses relative to the objects, we used COLMAP [28] and ICP on the reference point clouds obtained by Graycode technique [10]. We also captured ground-truth shapes of the objects with KinectFusion [11].

Figure 10 and Table 3 present the results of the evaluation. We confirmed that the proposed method achieved superior accuracy compared to NeuS and ActiveNeuS in almost all scenes. However, for the Block scene with random dot projection, our results were slightly degraded. We believe this is due to the projected pattern not being dense enough, as observed in Figure 9 (2nd column, bottom row), which made it difficult to efficiently utilize spatial consistency.

From the qualitative comparison results, it is evident that the proposed method successfully reconstructed accurate shapes compared to NeuS and ActiveNeuS, especially in areas with insufficient textural information. Although we expected ActiveNeuS to be more robust against texture-less surfaces due to its explicit use of projected pattern information, it produced severe artifacts, particularly noticeable in the hamming scenes. We attribute this to ActiveNeuS modeling reflection with flat shading, which lacks the representation power for various optical phenomena, whereas our method uses an implicit reflection model with an MLP. Furthermore, ActiveNeuS suffered from wrong reconstruction due to projector occlusion (Figure 10, red ellipses), while ours is robust against projector occlusion thanks to the computed shadow masks (Figure 7).

We also observed that for block scenes, the reconstructed shapes using 12 views were consistently more accurate than those using 36 views. We attribute this to the presence of some frames with erroneous poses in the 36-view case.

5. Conclusion

In this paper, we propose an unsupervised structured light (USSL) approach that does not explicitly rely on prior knowledge of the projected pattern. We demonstrate that accurate scene reconstruction is achievable without pattern information, even with a limited number of viewpoints, where passive stereo methods typically fail catastrophically. Our approach also incorporates a shadow volume rendering module to account for shadows caused by pattern projection, thereby enhancing reconstruction accuracy, as evidenced by our ablation study. For future work, we aim to

Table 3. Comparison results on a real dataset. **Bold**: Ours superior to active stereo-based methods. Underline: Ours superior to passive stereo-based methods.

Pattern	Method	36 views				12 views			
		Dog	Block	Mannequin head	Mannequin body	Dog	Block	Mannequin head	Mannequin body
No	NeuS	7.62	3.58	6.33	2.76	8.19	3.43	6.07	3.34
Random dot	ActiveNeuS	7.73	3.50	6.11	-	8.25	3.27	6.37	-
	USSL	**7.36**	3.51	**5.48**	-	**7.98**	3.36	**5.85**	-
Hamming	ActiveNeuS	8.24	3.98	6.97	-	8.46	4.02	6.25	-
	USSL	**6.70**	**3.41**	**5.38**	-	**7.35**	**3.34**	**5.50**	-
Cross-lasers	ActiveNeuS	-	-	-	2.65	-	-	-	3.16
	USSL	-	-	-	**2.54**	-	-	-	**2.68**

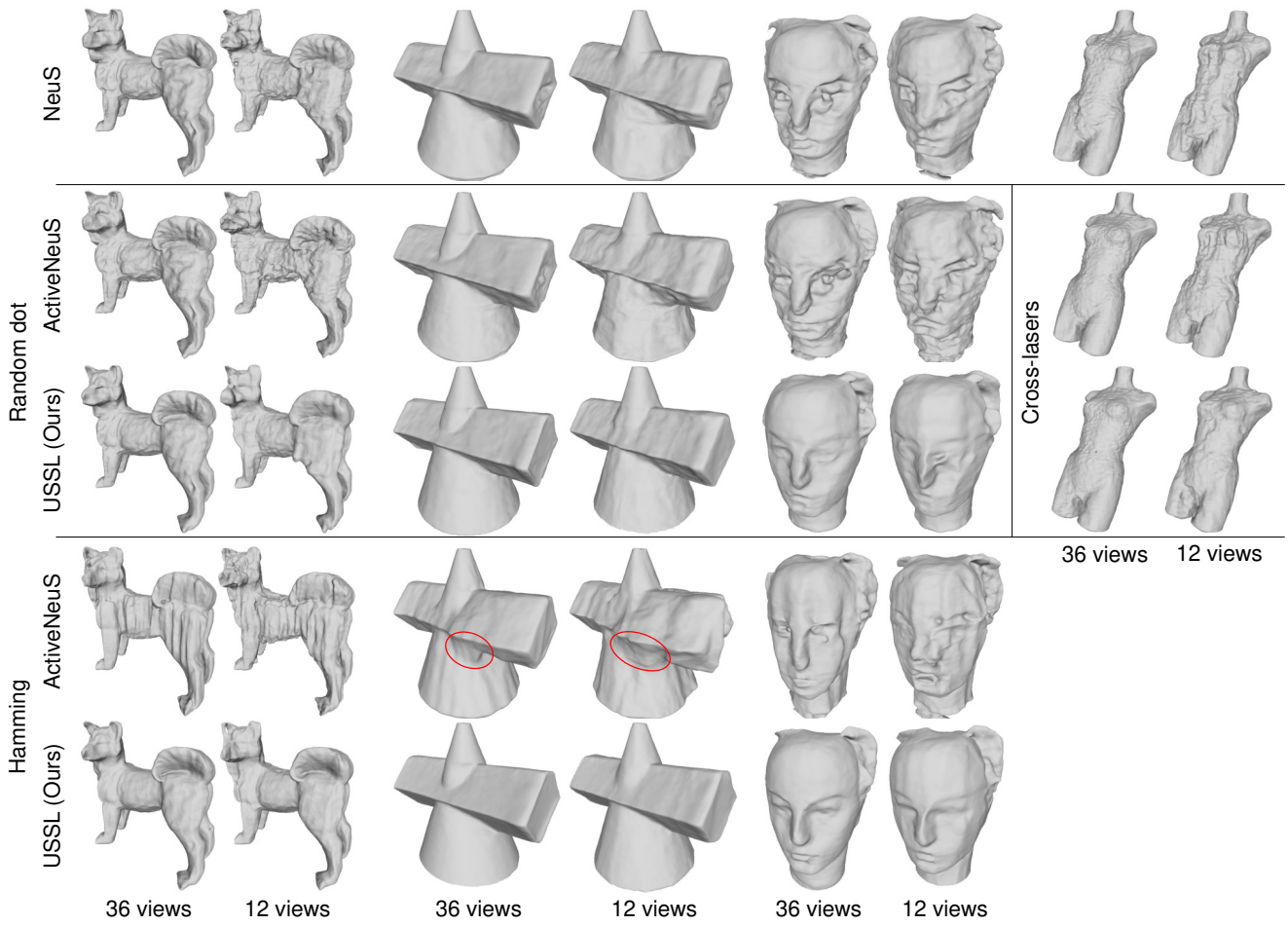

Figure 10. Qualitative comparison results on a real dataset. Red ellipses indicate where the reconstruction was likely to be incorrect due to occlusion.

improve the performance of USSL in few-shot scenes and extend its application to dynamic scenes.

ACKNOWLEDGMENT

This work was supported by JST Startup JPMJSF23DR and JSPS/KAKENHI JP20H00611 and JP23H03439 in Japan.

References

[1] Bartolomei, L., Poggi, M., Tosi, F., Conti, A., Mattoccia, S.: Active stereo without pattern projector. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 18470–18482 (October 2023) 2

[2] Boss, M., Braun, R., Jampani, V., Barron, J.T., Liu, C., Lensch, H.P.: Nerd: Neural reflectance decomposition from image collections. In: IEEE International Conference on Computer Vision (ICCV) (2021) 2

[3] 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) 2

[4] Cho, S.Y., Chow, T.: A neural-learning-based reflectance model for 3-d shape reconstruction. IEEE Transactions on Industrial Electronics (2000) 3

[5] Fernandez, S., Salvi, J.: A novel structured light method for one-shot dense reconstruction. In: IEEE International Conference on Image Processing (2012) 2

[6] Furukawa, R., Mikamo, M., Sagawa, R., Kawasaki, H.: Single-shot dense active stereo with pixel-wise phase estimation based on grid-structure using cnn and correspondence estimation using gcn. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4001–4011 (January 2022) 2

[7] Furukawa, R., Oka, S., Kotachi, T., Okamoto, Y., Tanaka, S., Sagawa, R., Kawasaki, H.: Fully auto-calibrated active-stereo-based 3d endoscopic system using correspondence estimation with graph convolutional network. In: Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (2020) 2

[8] Gropp, A., Yariv, L., Haim, N., Atzmon, M., Lipman, Y.: Implicit geometric regularization for learning shapes. In: Proceedings of Machine Learning and Systems 2020, pp. 3569–3579 (2020) 2, 4

[9] Gu, F., Song, Z., Zhao, Z.: Single-shot structured light sensor for 3d dense and dynamic reconstruction. Sensors **20**(4), 1094 (2020) 2, 5, 6

[10] Inokuchi, S., Sato, K., Matsuda, F.: Range imaging system for 3-d object recognition. In: International Conference on Pattern Recognition (1984) 2, 7

[11] Izadi, S., Kim, D., Hilliges, O., Molyneaux, D., Newcombe, R., Kohli, P., Shotton, J., Hodges, S., Freeman, D., Davison, A., Fitzgibbon, A.: Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera. In: UIST '11 Proceedings of the 24th annual ACM symposium on User interface software and technology. pp. 559–568. ACM (October 2011), https://www.microsoft.com/en-us/research/publication/kinectfusion-real-time-3d-reconstruction-and-interaction-using-a-moving-depth-camera/ 2, 7

[12] Jeong, Y., Ahn, S., Choy, C., Anandkumar, A., Cho, M., Park, J.: Self-calibrating neural radiance fields. In: ICCV (2021) 2

[13] Kazuto, I., Ikeda, T., Thomas, D., Iwaguchi, T., Kawasaki, H.: Activeneus: Neural signed distance fields for active stereo. In: International Conference on 3D Vision (3DV) (2024) 2, 3, 6

[14] Konolige, K.: Projected texture stereo. In: IEEE International Conference on Robotics and Automation (2010) 2

[15] Li, C., Hashimoto, T., Matsumoto, E., Kato, H.: Multi-view neural surface reconstruction with structured light. In: The British Machine Vision Conference (BMVC) (2022) 2

[16] Li, Z., Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H.: Neuralangelo: High-fidelity neural surface reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2023) 2, 6

[17] Lin, C.H., Ma, W.C., Torralba, A., Lucey, S.: Barf: Bundle-adjusting neural radiance fields. In: IEEE International Conference on Computer Vision (ICCV) (2021) 2

[18] Ling, J., Wang, Z., Xu, F.: Shadowneus: Neural sdf reconstruction by shadow ray supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2023) 3

[19] Lipson, L., Teed, Z., Deng, J.: Raft-stereo: Multilevel recurrent field transforms for stereo matching. In: International Conference on 3D Vision (3DV) (2021) 6

[20] Liu, Y., Wang, P., Lin, C., Long, X., Wang, J., Liu, L., Komura, T., Wang, W.: Nero: Neural geometry and brdf reconstruction of reflective objects from multiview images (2023) 3

[21] 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) 1, 2, 4, 5

[22] Mirdehghan, P., Chen, W., Kutulakos, K.N.: Optimal structured light a la carte. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018) 2

[23] Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. **41**(4), 102:1–102:15 (Jul 2022). https://doi.org/10.1145/3528223.3530127, https://doi.org/10.1145/3528223.3530127 2, 3

[24] Nagamatsu, G., Ikeda, T., Iwaguchi, T., Thomas, D., Takamatsu, J., Kawasaki, H.: Self-calibration of multiple-line-lasers based on coplanarity and epipolar constraints for wide area shape scan using moving camera. In: International Conference on Pattern Recognition (ICPR) (2022) 2, 6

[25] Nagamatsu, G., Takamatsu, J., Iwaguchi, T., Thomas, D., Kawasaki, H.: Self-calibrated dense 3d sensor using multiple cross line-lasers based on light sectioning method and visual odometry. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2021) 2

[26] Park, K., Henzler, P., Mildenhall, B., Barron, J.T., Martin-Brualla, R.: Camp: Camera preconditioning for neural radiance fields. ACM Trans. Graph. (2023) 2

[27] Rakotosaona, M.J., Manhardt, F., Arroyo, D.M., Niemeyer, M., Kundu, A., Tombari, F.: Nerfmeshing: Distilling neural radiance fields into geometrically-accurate 3d meshes (2023) 2

[28] Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 6, 7

[29] Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV) (2016) 6

[30] Shandilya, A., Attal, B., Richardt, C., Tompkin, J., O'Toole, M.: Neural fields for structured lighting. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (2023) 2

[31] Srinivasan, P.P., Deng, B., Zhang, X., Tancik, M., Mildenhall, B., Barron, J.T.: Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In: CVPR (2021) 3

[32] Srinivasan, V., Liu, H.C., Halioua, M.: Automated phase-measuring profilometry of 3-d diffuse objects. Applied Optics **23**, 3105–3108 (1984) 2

[33] Tzathas, P., Maragos, P., Roussos, A.: 3d neural sculpting (3dns): Editing neural signed distance functions. In: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). IEEE (January 2023) 2

[34] 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) 1, 2, 4, 6

[35] 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) 2

[36] Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V.A.: NeRF−−: Neural radiance fields without known camera parameters. arXiv preprint arXiv:2102.07064 (2021) 2

[37] Yamazaki, S., Nukada, A., Mochimaru, M.: Hamming color code for dense and robust one-shot 3d scanning. In: The British Machine Vision Conference (BMVC) (2011) 2, 5, 6

[38] Yao, Y., Zhang, J., Liu, J., Qu, Y., Fang, T., McKinnon, D., Tsin, Y., Quan, L.: Neilf: Neural incident light field for physically-based material estimation. In: European Conference on Computer Vision (ECCV) (2022) 2

[39] Yariv, L., Gu, J., Kasten, Y., Lipman, Y.: Volume rendering of neural implicit surfaces. In: Thirty-Fifth Conference on Neural Information Processing Systems (2021) 1, 2

[40] Zhang, J., Zhan, F., Yu, Y., Liu, K., Wu, R., Zhang, X., Shao, L., Lu, S.: Pose-free neural radiance fields via implicit pose regularization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3534–3543 (October 2023) 2

[41] Zhang, J., Yao, Y., Li, S., Liu, J., Fang, T., McKinnon, D., Tsin, Y., Quan, L.: Neilf++: Inter-reflectable light fields for geometry and material estimation. In: International Conference on Computer Vision (ICCV) (2023) 2

[42] Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T.: Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. In: SIGGRAPH Asia (2021) 3