Differentiable Adaptive 4D Structured Illumination for Joint Capture of Shape and Reflectance

Huakeng Ding[1] Yaowen Chen[1] Kun Zhou[1,2†] Hongzhi Wu[1†]
[1]State Key Lab of CAD&CG, Zhejiang University
[2]Hangzhou Research Institute of Holographic and AI Technology

Abstract

We present a differentiable framework to adaptively compute 4D illumination conditions with respect to an object, for efficient, high-quality simultaneous acquisition of its shape and reflectance, with a unified spatial-angular structured light and a single camera. Using a simple histogram-based pixel-level probability model for depth and reflectance, we differentiably link the next illumination condition(s) with a loss that encourages the reduction in depth uncertainty. As new structured illumination is cast, corresponding image measurements are used to update the uncertainty at each pixel. Finally, a fine-tuning-based approach reconstructs the depth map and reflectance parameter maps, by minimizing the differences between all physical measurements and their simulated counterparts. The effectiveness of our framework is demonstrated on physical objects with wide variations in shape and appearance. Our depth results compare favorably with state-of-the-art techniques, while our reflectance results are comparable when validated against photographs.

1. Introduction

Capturing the geometry and appearance of physical objects is a key problem in computer vision and graphics. Often represented as a 3D mesh and a 6D spatially-varying bidirectional reflectance distribution function (SVBRDF), the reconstruction result enables photo-realistic image synthesis under any view and lighting conditions. This benefits a wide range of applications, including cultural heritage, e-commerce, special effects and video games.

Active, structured illumination is commonly employed in high-quality acquisition of geometry *or* reflectance. It probes the physical domain to obtain measurements strongly correlated with target information, resulting in a high signal-to-noise ratio (SNR). For shape, spatially-varying light pattern(s) are projected into the 3D space to distinguish rays for triangulation [16, 37]. For reflectance, angular-domain illumination patterns are cast onto the object surface, physically convolving with BRDF slices at different locations. Accurate appearance can be deduced from corresponding measurements [14, 40].

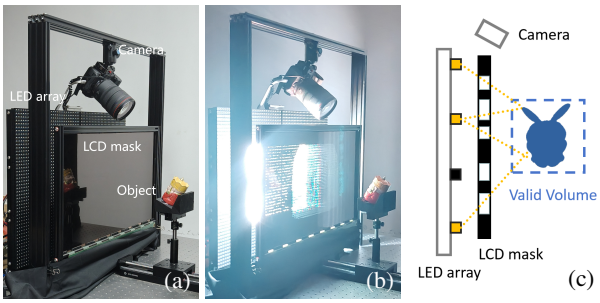

Figure 1. Our acquisition setup. It consists of a camera, an LED array and an LCD mask (a). The setup is working with optimized light/mask pattern (b). A side view is illustrated in (c).

To facilitate *active* capture of both shape *and* reflectance, the first 4D structured light is recently proposed in [43], unifying the aforementioned two types of illumination in a compact form factor. Despite its physical sampling capability in the spatial-angular domain, the acquisition efficiency is not satisfactory: it takes 24 minutes to scan a single view, the majority of which is spent on geometry capture, with only one LED on at a time. Moreover, the illumination conditions in [43] are pre-optimized, which could be *suboptimal* for acquiring a specific object.

To actively capture shape and reflectance at the same time with *high efficiency and quality*, we propose a differentiable framework to automatically optimize the spatial-angular lighting conditions during acquisition, which are tailored to the object being digitized. This is achieved by differentiably connecting the next lighting condition(s) to a loss that minimizes the overall depth uncertainty. By using multiple LEDs instead of one at a time as in previous work, we reduce the exposure time by up to 100× and the total acquisition time by 2×. After acquisition, we further

†Corresponding authors: {kunzhou,hwu}@acm.org.

fine-tune the results by minimizing the differences between physical measurements and simulated ones. The final output is a depth map and multiple texture maps that store GGX BRDF parameters.

The main contributions of this paper are as follows:

- We propose a learned multiplexing scheme for 4D spatial-angular structured illumination, to simultaneously capture shape and reflectance with high efficiency.
- We propose a differentiable framework to on-the-fly optimize complex lighting conditions, which are *adaptive* to an object of interest.

2. Related Work

This section reviews representative work on structured illumination for geometry and/or reflectance, as well as adaptive acquisition, which are two topics most related to our paper. Passive approaches are not discussed here. Interested readers are directed to excellent surveys [7, 15, 18, 36, 42] for a broader view of the topic.

2.1. Structured Illumination

Spatial Lighting. Laser-stripe triangulation [24] and spatial structured lighting [16, 32, 37] are widely used for capturing highly accurate geometry. These techniques project one or more spatially distinctive, discrete or continuous patterns onto the object surface, encoding light rays to facilitate 3D triangulation. Different pattern designs have been explored to enhance robustness [16, 33], improve computational efficiency [9, 10, 34], and increase acquisition speed [21]. While the spatial structures in illumination are historically hand-crafted, recent work demonstrates enhanced quality and efficiency with automatically designed ones [4, 45].

Angular Lighting. Despite its quality, exhaustively sampling the physical appearance domain with a gantry is prohibitively time-consuming [6, 22]. On the other hand, illumination multiplexing is a highly successful class of approaches, which modulates the intensities of lights at multiple angles, to reconstruct reflectance based on measurements taken under different lighting conditions. The light-stage recovers appearance using a pre-computed inverse lookup table [14]. Planar SVBRDF is estimated by analyzing appearance changes relative to a moving linear light source [3, 13]. Isotropic reflectance can be captured via a frequency domain analysis, utilizing an LCD panel as the light source [1]. Similar to spatial illumination, recent trend in angular lighting moves from manual to automatic design of light patterns, substantially increasing the acquisition efficiency [5, 17, 27, 28].

The closest work to ours is [43], which proposes a unified structured light in both spatial and angular domain, with an LED array and an LCD mask. Despite hardware novelty, the prototype *separately* captures geometry (acting like a conventional projector) and then appearance (like a light-stage), with pre-optimized 2D lighting conditions. In comparison, we propose the first learned multiplexing approach for 4D spatial-angular domain to *simultaneously* capture shape and reflectance, with an improved efficiency in both the acquisition time and the number of input photographs. We also establish a differentiable adaptive framework to optimize lighting conditions that are tailored to an object.

2.2. Adaptive Acquisition

Geometry Capture. Real-time generation of a parametric [20] or code-based [31] single-shot spatial pattern can be performed on-the-fly with respect to scene content. In [25], structured patterns with adaptive color is computed using principal component analysis of the scene image with a projector and two cameras. Adaptation strategy for focus and exposure is proposed in [46] to generate corresponding pattern for robust depth sensing. In [35], the next pattern is selected from an pre-defined set to maximize information gain. Recently, another class of approaches optimize *subject-independent* structured patterns are optimized to adapt to a specific hardware configuration [4, 32].

Appearance Capture. Lensch et al. [23] propose a function to measure the reduction in uncertainty added by one view, to guide view planning for SVBRDF acquisition. In [12], a sampling algorithm is developed to incrementally choose light directions adapted to the properties of a reflectance field. Exploiting a database prior, an adaptive method of accurate interpolation of sparsely measured BRDF is introduced [11]. A sampling scheme is derived from a new BRDF parameterization that automatically adapts to the behavior of a material [8]. Liu et al. [26] apply meta-learning to optimize the physical sampling pattern. Stochastic particle-optimization sampling is adopted in [47] to sample uncertain material parameters to guide acquisition process. Note that a recent approach learns a divide-and-conquer strategy for reflectance reconstruction, with fixed illumination patterns [29].

Unlike the majority existing work in this category, we propose a *differentiable* framework to automatically compute high-dimensional structured illumination for *simultaneous* acquisition of both shape and reflectance.

3. Hardware

We use a single-camera acquisition prototype, similar to the one proposed in [43]. The spatial-angular structured light includes a rectangular RGB LED array (with 64×48 = 3,072 LEDs) and an LCD mask (with a spatial resolution of $1{,}920\times1{,}080$). A 45MP Canon EOS R5 camera is mounted above the mask. We use a focal length of 24mm and an aperture of f/22. A valid volume is defined as a cube of $15\text{cm}\times15\text{cm}\times15\text{cm}$, whose center is 15cm in front of the center of the mask. All intrinsic and extrinsic param-

eters of cameras and lights are differentiably calibrated in an end-to-end fashion (detailed in the supplemental material). Following existing terminology, we refer to the intensity distribution of the LED array as a **light pattern**, and the image displayed on the LCD as a **mask pattern**. Please refer to [43] for more details.

4. Preliminaries

Below we describe the relationship between a pixel measurement, the depth/reflectance corresponding to that pixel and a light/mask pattern, which will be used to drive differentiable optimization described in the next section.

$$\begin{aligned} I_{j,k} &= \sum_l \int_A L_j(\mathbf{x}_l, -\omega_k^i) M_j(\mathbf{x}_l \leftrightarrow \mathbf{x}_k) f_{k,l} F dA, \\ &\approx \sum_l f_{k,l} F \int_A L_j(\mathbf{x}_l, -\omega_k^i) M_j(\mathbf{x}_l \leftrightarrow \mathbf{x}_k) dA, \\ &\approx \sum_l f_{k,l} F L_j(l) \Psi(-\omega_k^i) \int_A L(\mathbf{x}_l) M_j(\mathbf{x}_l \leftrightarrow \mathbf{x}_k^p) dA. \end{aligned} \tag{1}$$

Here $I_{j,k}$ is an image measurement at a pixel k under j-th light/mask pattern. $\mathbf{x}_l$ is a point on an LED with an index of l, modeled as an area light of 2mm $\times$ 2mm according to calibration. $\mathbf{x}_k$ is the 3D position corresponding to the current pixel. The lighting direction ω_k^i is a unit vector pointing from $\mathbf{x}_k$ to $\mathbf{x}_l$, and the view direction ω_k^o from $\mathbf{x}_k$ to the center of the camera. Moreover, $L_j(\mathbf{x}_l, -\omega_k^i)$ is the radiance emitted from the LED l along the direction $-\omega_k^i$. $M_j(\mathbf{x}_l \leftrightarrow \mathbf{x}_k)$ is the mask value, where the ray from x_l to x_k intersects our liquid crystal panel. In addition, $f_{k,l}$ is a BRDF value for (ω_k^i, ω_k^o). We use parametric GGX model [41] in this paper, while other models can also be plugged in here. $F = \frac{(\omega_k^i \cdot \mathbf{n}_k)^+ (-\omega_k^i \cdot \mathbf{n}_l)^+}{||\mathbf{x}_l - \mathbf{x}_k||^2}$ is the form factor, where $\mathbf{n}_k/\mathbf{n}_l$ is the surface normal of $\mathbf{x}_k/\mathbf{x}_l$, respectively. The above integral is calculated over the surface area A of the LED. Due to the small solid angle subtended by A with respect to $\mathbf{x}_k$, we assume constant $f_r/F/\omega_k^i$ in the integral, and factor L as:

$$L_j(\mathbf{x}_l, -\omega_k^i) \approx L_j(l) \Psi(-\omega_k^i) L(\mathbf{x}_l), \tag{2}$$

where $L_j(l)$ is the relative intensity of the LED l in the j-th light pattern, in the range of [0,1]; Ψ is a pre-calibrated angular distribution function for an LED; and $L(\mathbf{x}_l)$ is implemented as a 5×5 kernel, satisfying $\int_A L(\mathbf{x}_l) dA = 1$, whose values are also pre-calibrated, similar to [43].

5. Overview

Our pipeline consists of two stages: differentiable adaptive acquisition and fine-tuning. First, for a physical object, we compute the next light/mask pattern(s) by minimizing depth uncertainty, take photograph(s) with these patterns, update the uncertainty measure with new measurements, and repeat this process until a termination condition is met (Sec. 6). Next, we use the depth/reflectance estimate in the previous stage as initial values, and fine-tune the results by minimizing the differences between physical measurements and corresponding simulated ones (Sec. 7). The final output is a depth map and several texture maps that store parameters of the GGX BRDF model. Similar to the majority of related work, we focus on high-quality reconstruction from a single view, and leave the scanning of a complete 3D object to orthogonal techniques. Please refer to Fig. 2 for a graphical illustration of the pipeline, and Fig. 4 for an example.

6. Differentiable Adaptive Acquisition

To quantitatively measure uncertainty, we first build simple histogram-based probability models for depth and BRDF parameters at a valid pixel (Sec. 6.1). During adaptive acquisition, we differentiably optimize the next light/mask pattern(s), by minimizing the uncertainty computed from our models (Sec. 6.2). We then update the models with the new measurements under optimized patterns (Sec. 6.1). This process is repeated until 72 light/mask patterns are projected in our experiments, while other termination conditions can also be employed here.

6.1. Probability Model

Definition. For a *single* pixel, we assume that the joint distribution of its depth and reflectance can be modeled as independent probability models for each individual parameter (i.e., depth/BRDF parameters) based on probability mass functions defined with histograms. A **candidate** has its depth and BRDF parameters independently sampled from these probability models.

Specifically, to build a probability model for depth, we first determine its minimal and maximal value by intersecting the valid volume (Sec. 3) with the camera ray corresponding to the current pixel. Next, we uniformly split the range into n_{bin} bins (n_{bin} = 100 in our experiments). For each bin, it stores the highest Zero-Normalized Cross-Correlation (ZNCC) score among all randomly sampled candidates, whose depth falls within the current bin: for each candidate, the score is computed between the physical measurements at the current pixel under the projected light/mask pattern(s) so far, and the corresponding simulated measurements (Eq. (1)). Finally, the scores in each bin are normalized to generate a probability distribution. We use ZNCC as it is a common choice of metric in depth acquisition [30, 32]. Please refer to Fig. 3 for an illustration.

For a BRDF parameter, its probability model is similar to that of depth, except that its range is determined by the statistics reported in OpenSVBRDF [28]. Moreover, we calculate the inverse of $\mathcal{L}_1$ distance instead of ZNCC, as

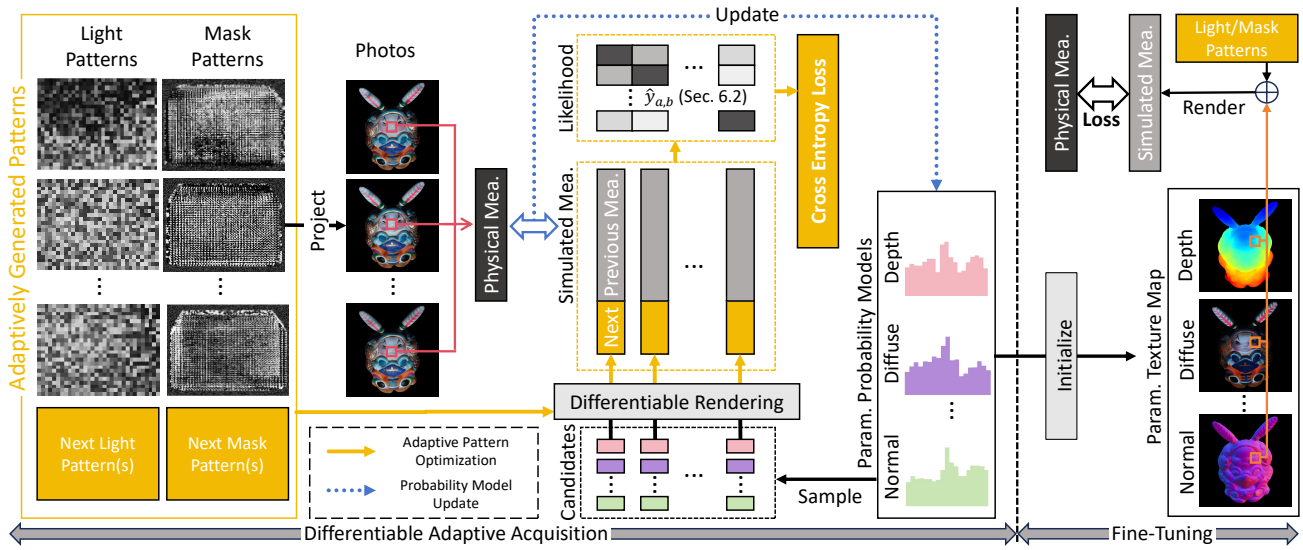

Figure 2. Our pipeline consists of two stages. First, for a physical object, we compute the next light/mask pattern(s) by minimizing the cross entropy among possible candidates sampled from histogram-based probability models. We then take photograph(s) with these patterns, and update probability distributions based on new measurements. This process is repeated until a termination condition is met. Next, we use the depth/reflectance estimate from previous stage as initial values, and fine-tune the results by minimizing the differences between physical measurements and corresponding simulated ones. The final output is a depth map and several texture maps that store parameters of the GGX BRDF model. Params. = parameters, mea. = measurements.

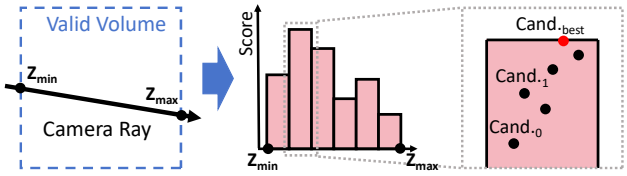

Figure 3. Graphical illustration of our probability model for depth. To build this model, we first determine its range $z_{\min}/z_{\max}$ by intersecting the valid volume with a corresponding camera ray. This range is then split into n_{bin} bins. Each bin stores the highest ZNCC score, computed between physical measurements and simulated ones from each candidate. Cand. = candidate.

$\mathcal{L}_1$ norm is commonly used in state-of-the-art appearance reconstruction [2, 44].

Sample. To sample the depth/reflectance parameter of a candidate from our model, we randomly select a bin based on the scores after normalization (i.e., the probability mass function). Next, we continuously sample a value in the range of the current bin uniformly at random, as the final result (Fig. 3).

Update. All histograms are initialized as a uniform probability distribution. We adopt a Monte-Carlo approach to update them. First, we randomly sample the depth/BRDF parameters for n_{sample} candidates ($n_{\text{sample}} = 600$ in our experiments), according to the current probability models. For each candidate, we compute its ZNCC score and store the highest one among all candidates in the corresponding bin in the probability model for depth. Similar operations are carried out by updating with the inverse of $\mathcal{L}_1$ distance of the current candidate, for the probability models for BRDF parameters. We hope that as more measurements under adaptive illumination are received, the probability distribution on either depth or each reflectance parameter at a valid pixel becomes more concentrated (i.e., we become more certain about the depth and appearance). The bottom row of Fig. 4 shows an example of the histogram of depth at a single pixel that gradually concentrates around the ground truth.

6.2. Optimizing the Next Pattern(s)

Loss function. It is defined as the sum of the depth uncertainty at each valid pixel. For a single pixel, we compute its depth uncertainty as the cross entropy:

$$-\sum_{a,b} y_{a,b} \log(\hat{y}_{a,b}), \tag{3}$$

where we cast the problem as standard multi-class classification [43]. Specifically, we randomly sample candidates according to the current probability distributions (Sec. 6.1), and treat each candidate as a class of its own. $y_{a,b}$ is the ideal likelihood for classifying candidate a to class b. It is 1 if $a = b$, and 0 otherwise. $\hat{y}_{a,b}$ is the computed likelihood for classifying candidate a to class b from the simulated measurements under existing and the next pattern(s),

as:

$$\hat{y}_{a,b} = \frac{e^{\mathrm{ZNCC}(\{I_{j,a}\}_j, \{I_{j,b}\}_j)}}{\sum_b e^{\mathrm{ZNCC}(\{I_{j,a}\}_j, \{I_{j,b}\}_j)}}. \quad (4)$$

Here $\{I_{j,a}\}_j$/$\{I_{j,b}\}_j$ is a vector of measurements, each of whose elements is calculated with Eq. (1), corresponding to a particular light/mask pattern.

Intuitively, our loss function in Eq. (3) encourages the distinctiveness from one candidate to another in terms of the corresponding measurements under the patterns being optimized. Note that the loss is defined with respect to depth uncertainty only, without explicit consideration of the uncertainty in reflectance. The reason is that per-pixel reflectance can be faithfully recovered by using even a small number of photographs under varying illumination, as in state-of-the-art techniques [28, 43]. So we would like to devote our effort to the more difficult problem of depth reconstruction, as reflectance can be easily recovered as a by-product: our experiments confirm that photographs taken with adaptive illumination that minimizes depth uncertainty are sufficient to produce high-quality reflectance results (see Sec. 9 and additional results in supplementary material).

Optimization. According to the relationship in Eq. (1), we can differentiably optimize the next light and mask pattern(s) adaptive to the current situation, by minimizing the loss function in Eq. (3). Please refer to Fig. 2 for a graphical illustration.

For physical plausibility and ease of calibration, the value of each light/mask pattern pixel goes through a sigmoid function to fit in the range of [0,1]. In particular, to generate a pixel in a mask pattern, we multiply a free variable with a large scalar (10^8 in our experiments) before going to the sigmoid, to encourage the final result to be either 0 or 1. Note that in our experiments, we optimize n_{batch} light/mask patterns at a time, to amortize the training cost (n_{batch} = 3 in our experiments). Please refer to Fig. 4 for a visualization.

We find that each candidate is not equally important in pattern optimization. For candidates whose ZNCC score is considerably smaller than the current best, it is unlikely that any of them will be the final solution. Moreover, for candidates whose depth is very close to the one with the highest ZNCC score, it is not feasible to obtain disambiguating pattern(s), as it is beyond the resolution of our LCD panel. Therefore, in practice, we find it effective to use only the candidates with top n_{peak} peak ZNCC scores in computing the cross entropy (n_{peak} = 3 in our experiments). Local maximum filtering with adaptive thresholding is applied for peak detection.

7. Fine-Tuning

This is a two-step process. The first step is to initialize depth and reflectance for each valid pixel, from the corresponding probability models after differentiable adaptive acquisition. Specifically, for each histogram, we subdivide each of its bins into 5 smaller ones, pick the bin with the highest score, and uniformly at random sample a value within its range as the initial value for the corresponding parameter (i.e., depth or a GGX parameter).

The second step is simultaneous fine-tuning of both depth and reflectance parameters, by minimizing the differences between physical and synthetic measurements, which are simulated according to Eq. (1). As directly optimizing native GGX parameters is difficult [43], we reparameterize the BRDF model with a 16D neural latent vector, along with 5 MLPs. Each MLP converts the latent to respective GGX parameters, to facilitate differentiable optimization. Before fine-tuning, for each valid pixel, we precompute a latent that closely predicts the initial values of GGX parameters from the first step. Please refer to [43] for details.

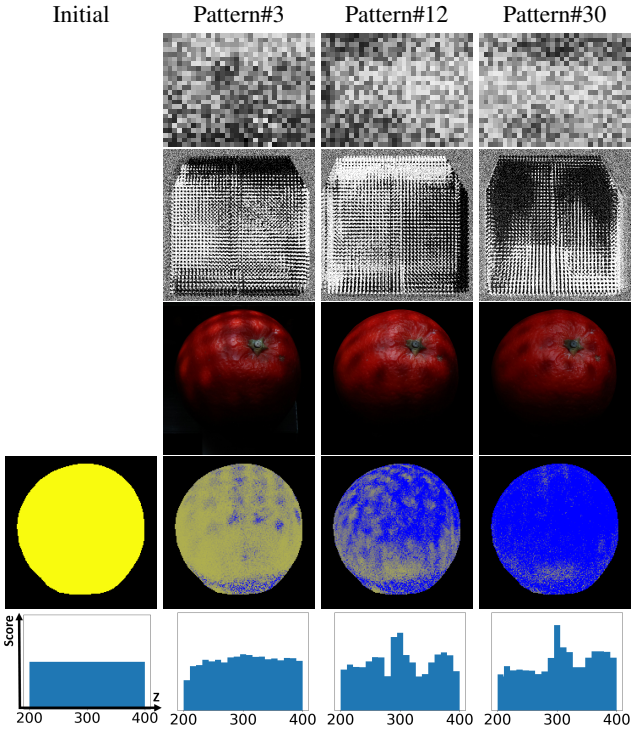

Figure 4. Visualization of various parts in adaptive acquisition. From the left column to right, after the initialization, after pattern#3, #12 and #30 is projected. From the top row to bottom, light pattern, mask pattern, corresponding photograph, depth uncertainty visualization (yellow = uncertain, blue = certain), and the visualization of the probability model at a single pixel.

8. Implementation Details

We apply [19] to perform foreground segmentation to determine the pixels of interest. For efficient computation during adaptive acquisition, we downsample each photo-

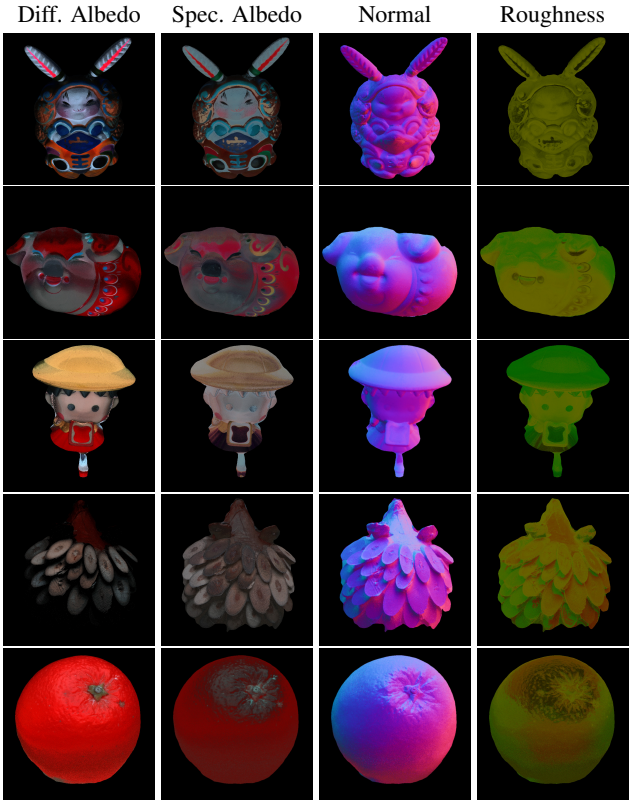

Figure 5. Reflectance results represented as GGX BRDF parameters map. Tangent maps are not shown here due to limited space.

graph from its original resolution to 127×64. We find that this resolution is suitable for our prototype, achieving decent quality without incurring an excessive computational burden (Sec. 9.2). A higher resolution can be used when more computational power is available. After acquisition, we start with the low resolution of 127×64, and gradually upsample it to the original resolution during fine-tuning (Sec. 7). Our differentiable pipeline is implemented with PyTorch. Adam optimizer is used in all experiments, with a learning rate of 10^{-3} and a weight decay of 10^{-6}.

9. Results and Discussions

We capture the shape and reflectance of 10 physical objects from a single view. The maximum dimension of each object ranges from 9 to 15cm. The appearance ranges from diffuse-dominant clay, wood, plastic, metallic paint, to highly specular natural wax coating. An exposure time of 0.2s is set for taking one photograph, and only LDR input images are used in our pipeline. All computation is conducted on a workstation with dual Intel Xeon 4210 CPUs, 256GB DDR4 memory and a GeForce RTX 3090 GPU. It takes about 10 minutes for adaptive acquisition (the majority of which is spent on pattern optimization, with only 15 seconds of total exposure time) and 2 hours for joint fine-tuning of about 1024×1024 depth map and multiple textures map for GGX parameters. Note that we focus on the former, as the latter is relatively easier to scale up via means like parallelization.

9.1. Comparisons

Geometry. We compare our shape reconstruction with related work in Fig. 6. According to [39], we measure geometric quality in the following metrics: RMSE in estimated depth, percentage of inliers (absolute depth error $<$ 3mm), and RMSE among the inliers. The ground-truth shape is captured with a commercial 3D scanner [38].

In the 1st,3rd and 4th column of the figure, we compare with [43], the closest state-of-the-art technique to ours, and [16], a representative method on traditional structured illumination. Our approach employs 3×24 = 72 adaptive light/mask patterns. For [43], the same number of patterns are pre-trained for a fair comparison. Note that each of their lighting conditions requires 20s of exposure time, due to the limited power of a single LED, which is two orders of magnitude longer than our approach with multiple LEDs on simultaneously. For [16], based on the effective spatial resolution of our LCD panel, a frequency band consisting of 16 pixels and 15 frequencies are used to generate 34 patterns. Our geometric results quantitatively outperform the two related techniques. For ORANGE, we are unable to scan a ground-truth shape as it is non-rigid. Yet one can qualitatively inspect the results: ours exhibit less banding artifacts.

In Fig. 7, we further demonstrate the benefit of making full use of the LED array during acquisition, compared with [43] using a single light source, as is common in the majority of work on structured illumination. Our result is more complete, as we are able to cast light from different sources to reduce the shadow region, where no reliable depth estimation can be made.

Reflectance. Fig. 5 shows texture maps that store the GGX BRDF parameters. Next, in the last 3 columns of Fig. 6, we compare our relighting results with [43], while validating against corresponding photographs taken under novel lighting conditions. In all cases, our results are comparable to [43] and closely resemble the photographs. Please refer to the supplementary material for more details.

9.2. Ablations

In first two columns of Fig. 6, we demonstrate the effectiveness of adaptive acquisition, by comparing to a variant that pre-trains 72 fixed light/mask patterns from synthetic training data, sampled according to the statistics in [28]. In Fig. 8, the depth accuracy improves with the increasing number of patterns, as more information is acquired to disambiguate current uncertainty.

We ablate various parameters of our approach as follows.

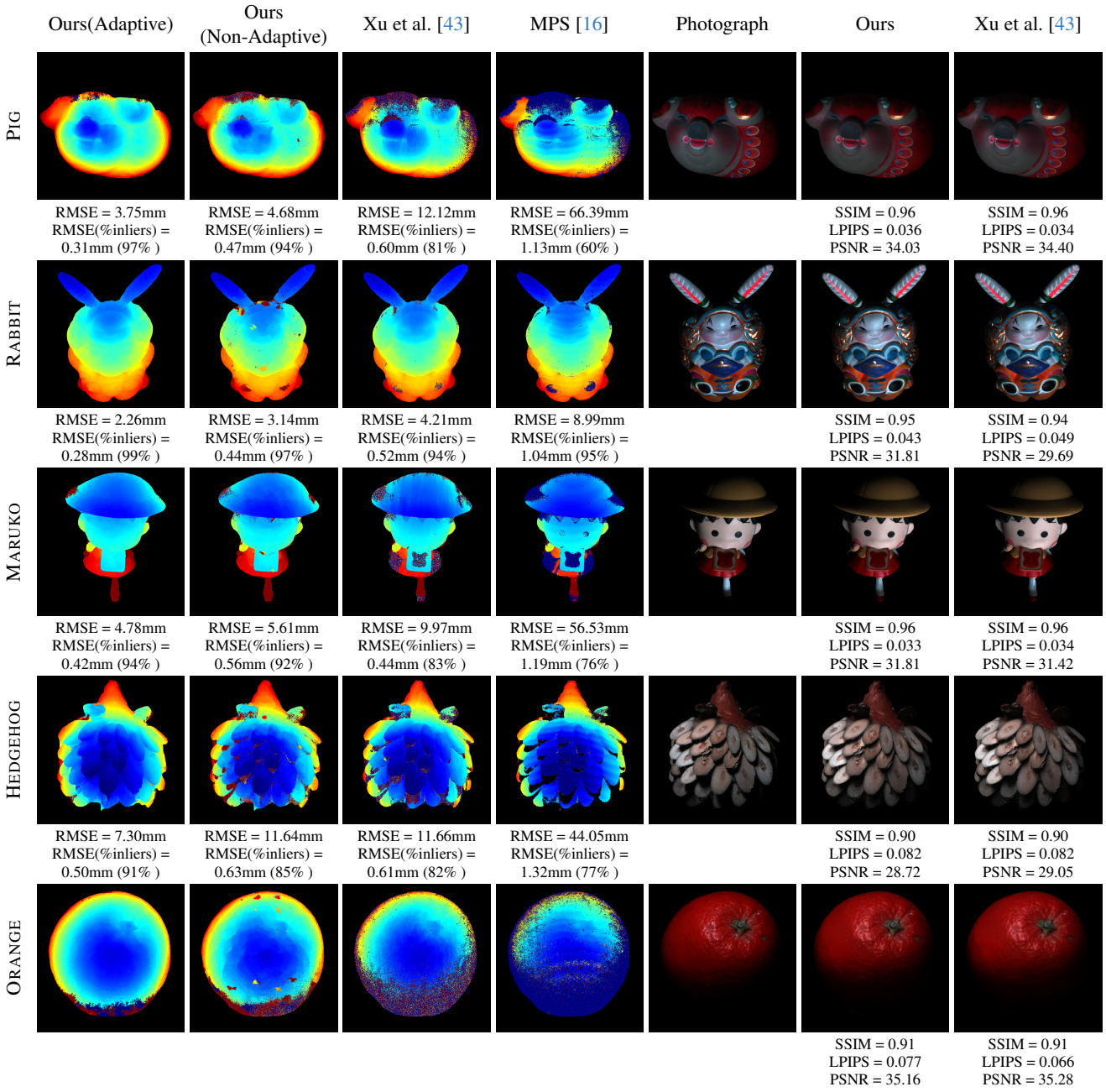

Figure 6. Comparisons with state-of-the-art techniques on shape and reflectance capture. From the left column to right: depth reconstruction with our approach (adaptive/non-adaptive patterns), [43] and MPS [16]; photograph under a lighting condition not used in optimization, rendering with the reflectance results of our approach and [43]. Quantitative errors are listed below each related image.

In Fig. 9, higher quality results are obtained with a larger number of Monte Carlo samples, at the expense of higher computational burden. Fig. 10 evaluates the impact of the batch size of simutaneously optimized next patterns in the same amount of total acquisition time. Our current choice of n_{batch} is empirically determined from this experiment. Fig. 11 ablates n_{peak}, which needs to strike a balance between missing good solutions and the speed to find the optimal value. In Fig. 12, finer bins improve the quality with more computations.

Finally, we validate the effectiveness of using a low image resolution for computing adaptive patterns in Fig. 13. The current choice is made after balancing accuracy and performance.

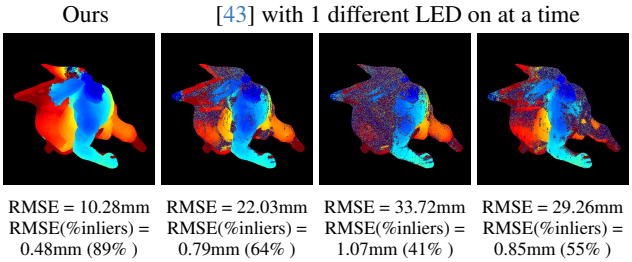

Figure 7. Comparison with a single-source structured light [43]. From the left to right: our result, the result of [43] when the LED at the center, left, or right corner of the LED array is on.

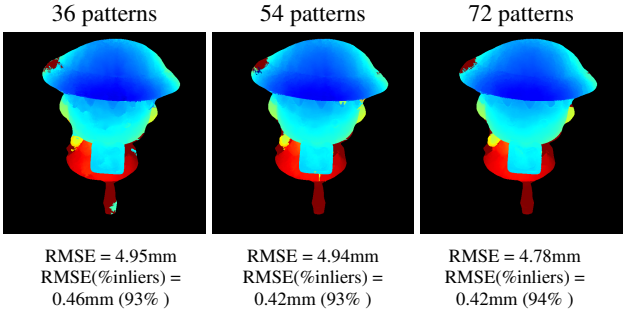

Figure 8. Impact of the total number of adaptive light/mask patterns over the depth quality.

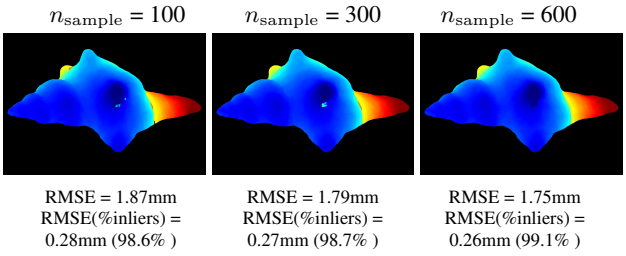

Figure 9. Impact of n_{sample} over the depth quality.

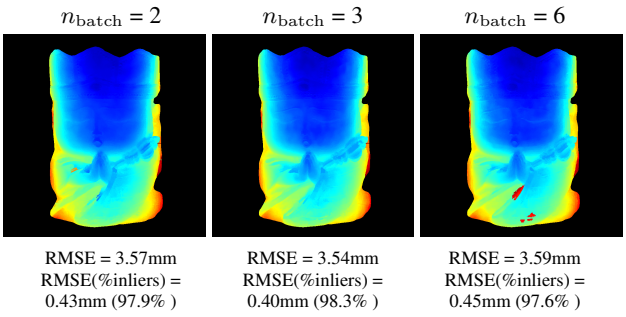

Figure 10. Impact of the number of simultaneously optimized next patterns (n_{batch}) over the depth quality, with the same total acquisition time.

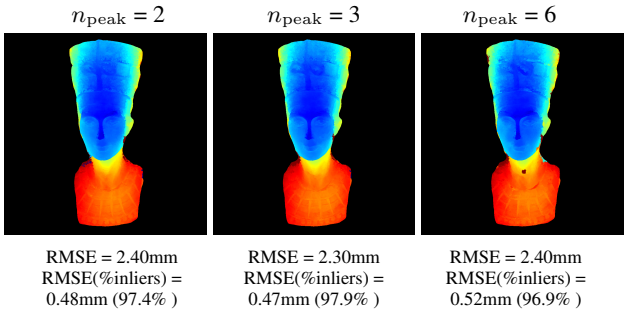

Figure 11. Impact of n_{peak} over the depth quality.

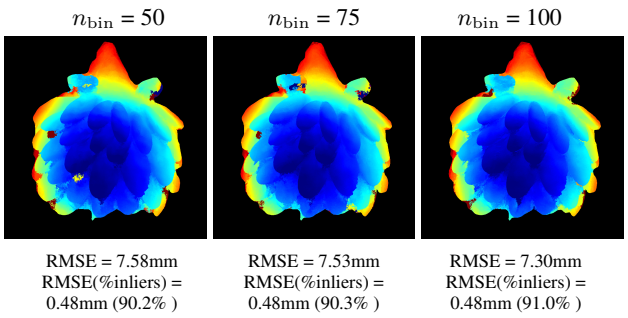

Figure 12. Impact of n_{bin} over the depth quality.

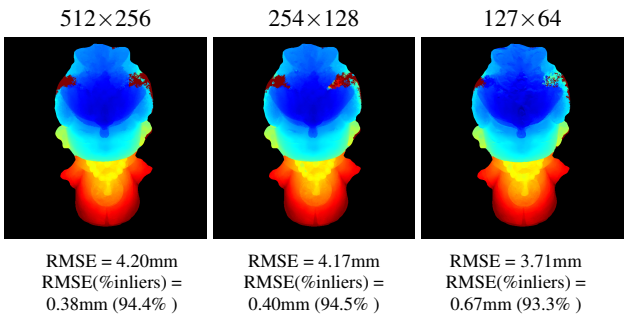

Figure 13. Impact of the image resolution for computing adaptive patterns (listed on top) over the depth quality. Note that the same reconstruction resolution is used in all cases.

10. Limitations & Future Work

Our work is subject to a number of limitations, which could inspire exciting future research possibilities. First, our depth uncertainty does not consider indirect illumination during adaptive pattern optimization. Second, the expressive power of depth maps and parametric BRDFs are limited. It will be intriguing to combine our approach with more advanced representations, such as Gaussian Splatting representation that supports high-quality relighting [2]. Finally, it seems promising to apply our idea to free-form scanning with a hand-held device [27], equipped with a miniature version of the spatial-angular structured light.

Acknowledgements. We would like to thank Fan Pei, Kaizhang Kang, and Xianmin Xu for their help. This work is partially supported by NSF China (62227806, 62332015, & 62421003), the XPLORER PRIZE, Information Technology Center, State Key Lab of CAD&CG, Zhejiang University and a gift from Adobe.

References

[1] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. Practical svbrdf capture in the frequency domain. *ACM Trans. Graph.*, 32(4), 2013.

[2] Zoubin Bi, Yixin Zeng, Chong Zeng, Fan Pei, Xiang Feng, Kun Zhou, and Hongzhi Wu. Gs^3: Efficient relighting with triple gaussian splatting. In *SIGGRAPH Asia 2024 Conference Papers*, 2024.

[3] Guojun Chen, Yue Dong, Pieter Peers, Jiawan Zhang, and Xin Tong. Reflectance scanning: estimating shading frame and brdf with generalized linear light sources. *ACM Trans. Graph.*, 33(4), 2014.

[4] Wenzheng Chen, Parsa Mirdehghan, Sanja Fidler, and Kiriakos N. Kutulakos. Auto-tuning structured light by optical stochastic gradient descent. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5969–5979, 2020.

[5] Seokjun Choi, Seungwoo Yoon, Giljoo Nam, Seungyong Lee, and Seung-Hwan Baek. Differentiable display photometric stereo. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11831–11840, 2024.

[6] Kristin J Dana, Bram Van Ginneken, Shree K Nayar, and Jan J Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions On Graphics (TOG)*, 18(1):1–34, 1999.

[7] Yue Dong. Deep appearance modeling: A survey. *Visual Informatics*, 3(2):59–68, 2019.

[8] Jonathan Dupuy and Wenzel Jakob. An adaptive parameterization for efficient material acquisition and rendering. *ACM Trans. Graph.*, 37(6), 2018.

[9] Sean Ryan Fanello, Christoph Rhemann, Vladimir Tankovich, Adarsh Kowdle, Sergio Orts Escolano, David Kim, and Shahram Izadi. Hyperdepth: Learning depth from structured light without matching. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5441–5450, 2016.

[10] Sean Ryan Fanello, Julien Valentin, Christoph Rhemann, Adarsh Kowdle, Vladimir Tankovich, Philip Davidson, and Shahram Izadi. Ultrastereo: Efficient learning-based matching for active stereo systems. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6535–6544, 2017.

[11] Jirí Filip, Radomír Vávra, Michal Haindl, Pavel id, Mikulá Krupika, and Vlastimil Havran. Brdf slices: Accurate adaptive anisotropic appearance acquisition. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1468–1473, 2013.

[12] Martin Fuchs, Volker Blanz, Hendrik P. A. Lensch, and Hans-Peter Seidel. Adaptive sampling of reflectance fields. *ACM Trans. Graph.*, 26:10, 2007.

[13] Andrew Gardner, Chris Tchou, Tim Hawkins, and Paul Debevec. Linear light source reflectometry. *ACM Trans. Graph.*, 22(3):749–758, 2003.

[14] Abhijeet Ghosh, Tongbo Chen, Pieter Peers, Cyrus A. Wilson, and Paul Debevec. Estimating specular roughness and anisotropy from second order spherical gradient illumination. In *Proceedings of the Twentieth Eurographics Conference on Rendering*, page 1161–1170, Goslar, DEU, 2009. Eurographics Association.

[15] Dar'ya Guarnera, Giuseppe Claudio Guarnera, Abhijeet Ghosh, Cornelia Denk, and Mashhuda Glencross. Brdf representation and acquisition. *Computer Graphics Forum*, 35, 2016.

[16] Mohit Gupta and Shree K. Nayar. Micro phase shifting. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 813–820, 2012.

[17] Kaizhang Kang, Zimin Chen, Jiaping Wang, Kun Zhou, and Hongzhi Wu. Efficient reflectance capture using an autoencoder. *ACM Trans. Graph.*, 37(4), 2018.

[18] Behnaz Kavoosighafi, Saghi Hajisharif, Ehsan Miandji, Gabriel Baravdish, Wen Cao, and Jonas Unger. Deep svbrdf acquisition and modelling: A survey. *Computer Graphics Forum*, 43(6):e15199, 2024.

[19] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, and Ross Girshick. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4015–4026, 2023.

[20] T.P. Koninckx and L. Van Gool. Real-time range acquisition by adaptive structured light. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):432–445, 2006.

[21] Sanjeev J. Koppal, Shuntaro Yamazaki, and Srinivasa G. Narasimhan. Exploiting dlp illumination dithering for reconstruction and photography of high-speed scenes. *International Journal of Computer Vision*, 96:125 – 144, 2011.

[22] Jason Lawrence, Aner Ben-Artzi, Christopher DeCoro, Wojciech Matusik, Hanspeter Pfister, Ravi Ramamoorthi, and Szymon Rusinkiewicz. Inverse shade trees for non-parametric material representation and editing. *ACM Transactions on Graphics (ToG)*, 25(3):735–745, 2006.

[23] Hendrik P. A. Lensch, Jochen Lang, Asla Medeiros Sá, and Hans-Peter Seidel. Planned sampling of spatially varying brdfs. *Computer Graphics Forum*, 22, 2003.

[24] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital michelangelo project: 3d scanning of large statues. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, page 131–144, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[25] Qiang Li, Moyuresh Biswas, Mark R. Pickering, and Michael R. Frater. Dense depth estimation using adaptive structured light and cooperative algorithm. In *CVPR 2011 WORKSHOPS*, pages 21–28, 2011.

[26] Chen Liu, Michael Fischer, and Tobias Ritschel. Learning to learn and sample brdfs. *Computer Graphics Forum*, 42(2): 201–211, 2023.

[27] Xiaohe Ma, Kaizhang Kang, Ruisheng Zhu, Hongzhi Wu, and Kun Zhou. Free-form scanning of non-planar appearance with neural trace photography. *ACM Trans. Graph.*, 40(4), 2021.

[28] Xiaohe Ma, Xianmin Xu, Leyao Zhang, Kun Zhou, and Hongzhi Wu. Opensvbrdf: A database of measured spatially-varying reflectance. *ACM Trans. Graph.*, 42(6), 2023.

[29] Xiaohe Ma, Yaxin Yu, Hongzhi Wu, and Kun Zhou. Efficient reflectance capture with a deep gated mixture-of-experts. *IEEE Transactions on Visualization and Computer Graphics*, 30(7):4246–4256, 2024.

[30] Jérôme Martin and James L. Crowley. Experimental comparison of correlation techniques. 2007.

[31] Xavier Maurice, Pierre Graebling, and Christophe Doignon. Real-time structured light coding for adaptive patterns. *J. Real-Time Image Process.*, 8(2):169–178, 2013.

[32] Parsa Mirdehghan, Wenzheng Chen, and Kiriakos N. Kutulakos. Optimal structured light a la carte. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6248–6257, 2018.

[33] Daniel Moreno, Kilho Son, and Gabriel Taubin. Embedded phase shifting: Robust phase shifting with embedded signals. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2301–2309, 2015.

[34] Rukun Qiao, Hiroshi Kawasaki, and Hongbin Zha. Depth reconstruction with neural signed distance fields in structured light systems. In *2024 International Conference on 3D Vision (3DV)*, pages 770–779, 2024.

[35] Guy Rosman, Daniela Rus, and John W. Fisher. Information-driven adaptive structured-light scanners. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 874–883, 2016.

[36] Joaquim Salvi, Jordi Pagès, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004. Agent Based Computer Vision.

[37] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, pages I–I, 2003.

[38] Shining3D. Einscan pro 2x plus handheld industrial scanner. https://www.einscan.com/handheld-3d-scanner/2x-plus/, 2024.

[39] Varun Sundar, Sizhuo Ma, Aswin C. Sankaranarayanan, and Mohit Gupta. Single-photon structured light. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17844–17854, 2022.

[40] Borom Tunwattanapong, Graham Fyffe, Paul Graham, Jay Busch, Xueming Yu, Abhijeet Ghosh, and Paul Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Trans. Graph.*, 32(4), 2013.

[41] Bruce Walter, Stephen R. Marschner, Hongsong Li, and Kenneth E. Torrance. Microfacet models for refraction through rough surfaces. In *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, page 195–206, Goslar, DEU, 2007. Eurographics Association.

[42] Michael Weinmann, Fabian Langguth, Michael Goesele, and Reinhard Klein. Advances in geometry and reflectance acquisition. In *Proceedings of the 37th Annual Conference of the European Association for Computer Graphics: Tutorials*, Goslar, DEU, 2016. Eurographics Association.

[43] Xianmin Xu, Yuxin Lin, Haoyang Zhou, Chong Zeng, Yaxin Yu, Kun Zhou, and Hongzhi Wu. A unified spatial-angular structured light for single-view acquisition of shape and reflectance. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 206–215, 2023.

[44] Chong Zeng, Guojun Chen, Yue Dong, Pieter Peers, Hongzhi Wu, and Xin Tong. Relighting neural radiance fields with shadow and highlight hints. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings*, page 1–11. ACM, 2023.

[45] Yixin Zeng, Zoubin Bi, Mingrui Yin, Xiang Feng, Kun Zhou, and Hongzhi Wu. Real-time acquisition and reconstruction of dynamic volumes with neural structured illumination. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20186–20195, 2024.

[46] Yueyi Zhang, Zhiwei Xiong, Pengyu Cong, and Feng Wu. Robust depth sensing with adaptive structured light illumination. *Journal of Visual Communication and Image Representation*, 25(4):649–658, 2014. 3D Video Processing.

[47] Zhiqian Zhou, Cheng Zhang, Zhao Dong, Carl Marshall, and Shuang Zhao. Estimating Uncertainty in Appearance Acquisition. In *Eurographics Symposium on Rendering*. The Eurographics Association, 2024.