Estimating Uncertainty in Appearance Acquisition

Z. Zhou[1] , C. Zhang[2] , Z. Dong[2], C. Marshall[2], S. Zhao[1]

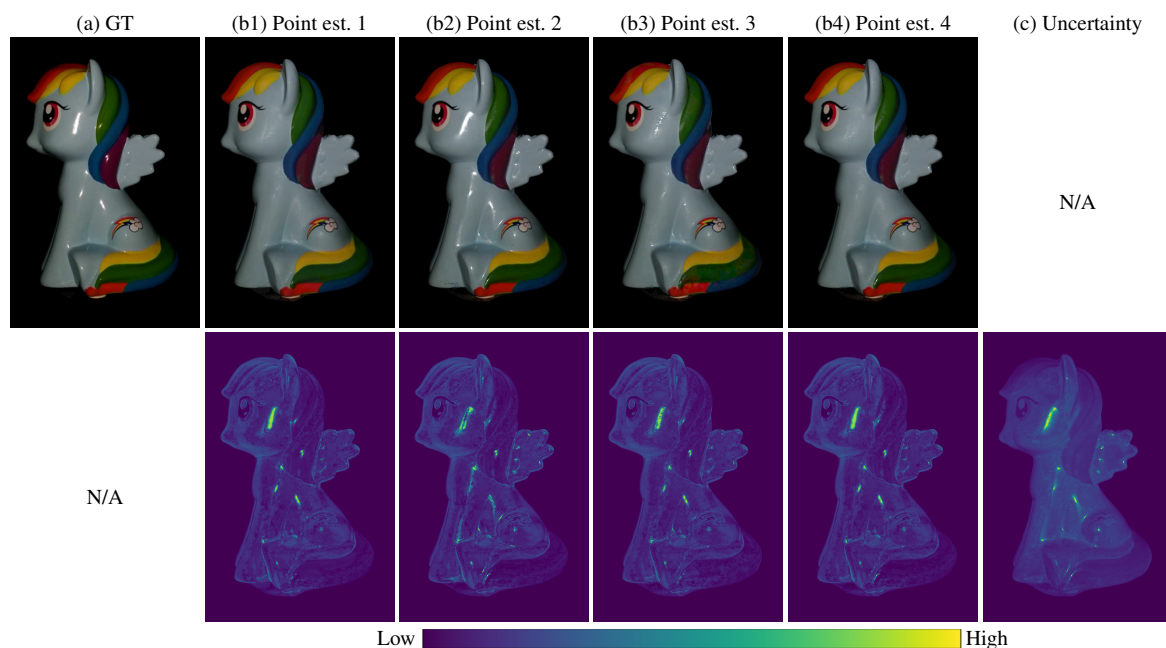[1]University of California, Irvine
[2]Meta Reality Labs

Figure 1: *The inference of a material's appearance parameters based on observations of the material is often under-constrained, causing the point estimates to suffer from ambiguities and prone to baking artifacts. In this paper, we introduce a technique to quantify such ambiguities. Given a set of observations and priors, our technique uses a Bayesian formulation to model uncertainties. This example uses real measured data (under collocated configurations) from [LZBD21]. Using the same input images, we obtain four point estimates of the reflectance of the toy horse (expressed as SVBRDF maps) using different gradient-descent methods (i.e., Adam [KB14] and SGD) with varying initializations. Despite of such differences, renderings of the point estimates under novel conditions (b1–b4) share similar error patterns (as shown on the bottom of these columns) due to uncertainties. Without the need of the groundtruth image (a), our predicted uncertainty (c) correlates strongly with those errors. Note that the color map is used throughout most results.*

Abstract
The inference of material reflectance from physical observations (e.g., photographs) is usually under-constrained, causing point estimates to suffer from ambiguity and, thus, generalize poorly to novel configurations. Conventional methods address this problem by using dense observations or introducing priors.
In this paper, we tackle this problem from a different angle by introducing a method to quantify uncertainties. Based on a Bayesian formulation, our method can quantitatively analyze how under-constrained a material inference problem is (given the observations and priors), by sampling the entire posterior distribution of material parameters rather than optimizing a single point estimate as given by most inverse rendering methods. Further, we present a method to guide acquisition processes by recommending viewing/lighting configurations for making additional observations. We demonstrate the usefulness of our technique using several synthetic and one real example.

delivered by
EUROGRAPHICS DIGITAL LIBRARY
www.eg.org   diglib.eg.org

1. Introduction

Acquiring the spatially varying reflectance of real-world materials is an important topic in computer graphics and vision. Unfortunately, despite decades of research, this remains a challenging problem for multiple reasons.

One main reason, for example, is the limitation of material appearance models. Many real-world materials exhibit complex light transport effects—including translucency, micro-scale interreflection, and wave effects—that cannot be fully captured by widely adopted material models like the microfacet BRDF.

On the other hand, even for materials whose appearances can be accurately modeled, the acquisition can still be challenging due to *ambiguities*. This is because, when recovering a material's appearance parameters based observations (e.g., photographs) of the material, the inference is usually *under-constrained*. In other words, there normally exist multiple feasible solutions that can reproduce the observed appearance of the material—but may suffer from "baking" artifacts and, thus, generalize poorly to non-observed configurations.

At a high level, there are two ways to mitigate the problem of ambiguity. The first approach is to use dense observations to minimize ambiguity. Techniques in this category are usually robust but can require lengthy acquisitions and large quantities of measured data. The second way is to introduce *priors*—which can be hand-crafted or data-driven (e.g., neural)—that models the (unconditional) distribution of "good" material parameters (e.g., those with little artifacts or baking). Methods in this category allow high-quality reconstructions using sparse observations but depend heavily on the choice of priors.

In this paper, we tackle this problem from a new angle by introducing a technique to *quantify ambiguities*. Specifically, provided a set of observations and the prior, our method utilizes a Bayesian formulation and offers the capability of quantitatively analyzing how under-constrained the material acquisition is (using these observations). Based on this method, we also introduce a technique that guides the acquisition process by recommending viewing/lighting configurations for making additional observations.

Concretely, our contributions include:

- Introducing stochastic particle-optimization sampling (SPOS) [ZZCC20] to material acquisition to efficiently sample posterior distributions of material appearance parameters (§3.2);
- Devising a statistical formulation for quantifying uncertainty in both parameter and image spaces (§3.3);
- Developing a method that guides appearance acquisition processes by recommending observation configurations that are likely capable of greatly reducing uncertainty (§3.4).

We demonstrate the effectiveness of our method using several synthetic and real examples in §4.

2. Related Work

Material appearance capture Real-world objects exhibit richly diverse appearance that can be described with models like spatially-varying bidirectional reflectance distribution functions (SVBRDFs).

To minimize ambiguity, many conventional appearance acquisition techniques rely on dense observations (e.g, [Mat03, LKG*03, HLZ10, DWT*10, CDP*14, DWMG15, KCW*18, LZBD21]). Alternatively, to democratize the acquisition, some recent works rely on hand-crafted priors (e.g., sparsity of SVBRDF parameter spaces) to allow reconstructions using fewer observations (e.g., [YDMH99, DCP*14, WWZ15, ZCD*16, KGT*17, PNS18]). In addition, several techniques leverage data-driven neural priors to produce plausible reconstructions of material appearance using small numbers of observations (e.g., [AAL16, HSL*17, GLD*19, DAD*19, GSH*20]). Despite of their convenience, these techniques are usually limited to objects with simple (e.g., planar) geometries.

Most material appearance capture techniques make *point estimates*—That is, return a single estimation of material parameters. However, our work is not the first to consider ambiguity in such capture. [LTH*23] addresses the ambiguity in joint reconstruction of material and environment lighting, by using a diffusion prior. [GHYZ20] performs Bayesian inference of procedural appearance models. Our technique also leverages a Bayesian formulation but with significantly different scope and objective.

Capture guidance has also been extensively used in computer vision. For brevity we only discuss its use in appearance capture. BRDF sampling techniques such as [NJR15] [XNY*16] are able to obtain a measured BRDF from an optimized set of limited samples. [FBLS07] adaptively samples a reflectance field, which can then be used for image-space relighting. These methods are specialized for their specific measurement applications. [LLSS03] is most related to our work. It analyzes the ambiguity in a captured SVBRDF and plans a best view for capture. However, its analysis is limited to a linear appearance model, utilizing only the local property of the problem, which isn't general enough for most parametric appearance models where non-linearity and multi-modality are common. In contrast, our method generalizes to non-linear appearance models and arbitrary objective functions, with capability of analyzing multi-modal distributions.

Posterior sampling A key ingredient for Bayesian inference is drawing samples from (potentially unnormalized) posterior distributions. Conventionally, this is achieved using Markov-Chain Monte Carlo methods such as the Metropolis-Hastings (MH) algorithm [MRR*53]. Unfortunately, these methods are known to have difficulties scaling to large sets of observations. To address this problem, stochastic gradient (SG)-MCMC methods that leverage Langevin Monte Carlo (LMC) [WT11] or Hamiltonian Monte Carlo (HMC) [CFG14] have been proposed.

Another relatively new category of Bayesian inference algorithm is particle-optimization methods. Unlike MCMC methods that generate (and propose) one sample at a time, the particle-optimization methods operate on multiple samples, or particles, simultaneously. Widely adopted particle-optimization methods include Stein variational gradient descent (SVGD) [LW16] as well as its many variants including annealed SVGD [DF21], MP-SVGD [ZLS*18], SVGD w/ matrix valued kernels [WTBL19],

and stochastic particle-optimization sampling (SPOS) [ZZCC20] (among others [DCM*18, CWC*19, CG20]).

In this paper, we utilize SPOS to draw samples of material appearance parameters from posterior probabilities.

3. Our Method

In what follows, we describe our method in details. In §3.1, we present the main ingredient of our method—which is based on Bayesian inference of material appearance. Then, in §3.3, we demonstrate how this formulation can be used to quantify the uncertainty of estimated material parameters. Lastly, in §3.4, we discuss another application of our technique: Selecting viewing/lighting configurations to minimize uncertainty.

3.1. Bayesian Formulation

Our technique utilizes a Bayesian formulation that infers material appearance parameters x (e.g., SVBRDF maps) given a set of observations $\mathcal{Y} := \{y_1, y_2, \ldots\}$ (e.g., photographs) as the estimation of the **posterior probability** $p_{\text{post}}(x \mid \mathcal{Y})$ of x conditioned on $\mathcal{Y}$.

According to Bayes' theorem, the posterior probability $p_{\text{post}}(x \mid \mathcal{Y})$ satisfies that

$$p_{\text{post}}(x \mid \mathcal{Y}) \propto L(\mathcal{Y} \mid x)\, p_{\text{pri}}(x), \tag{1}$$

where $L(\mathcal{Y} \mid x)$ is the **likelihood** of $\mathcal{Y}$ given x, and $p_{\text{pri}}(x)$ is the **prior probability** of x.

In practice, the likelihood $L(\mathcal{Y} \mid x)$ can be modeled based on the difference between each observation y_i (under lighting and viewing conditioned expressed by θ_i) and the corresponding simulated result $\mathcal{R}(x; \theta_i)$. For instance, we use the following variant motivated by L1 loss which is commonly used in inverse rendering

$$L(\mathcal{Y} \mid x) = \prod_i \exp\left(-\frac{\|y_i - \mathcal{R}(x; \theta_i)\|_1}{\sigma_{\text{err}}}\right), \tag{2}$$

where $\|\cdot\|_1$ denotes 1-norm, and $\sigma_{\text{err}} > 0$ is a hyper-parameter usually set to a small value, e.g. 0.01.

Lastly, the prior probability $p_{\text{pri}}(x)$ is considered a user input that can be, for example, a uniform distribution over all valid material appearance parameters x.

3.2. Sampling Posterior Distribution

With the posterior distribution defined in Eq. (1), we draw a discrete set of samples $\mathcal{X} := \{x_1, x_2, \ldots, x_n\}$ from this distribution. As we will explain in §3.3, these samples can be used to quantify properties like uncertainly of the posterior distribution.

To draw the samples, we utilize a technique called *stochastic particle-optimization sampling* (SPOS) [ZZCC20] originated in Stein variational inference [LW16]. To our knowledge, this is the first time SPOS is used for material appearance acquisition in computer graphics. In the following, we describe how our sampling works in details.

SPOS considers each sample x_i as the position of a "particle". By

iteratively refining these positions, SPOS ensures that, after convergence, the particles are distributed proportionally to a nonnegative **target function** $f(x)$—which we set based on the posterior probability $p_{\text{post}}(x \mid \mathcal{Y})$ to

$$f(x) = L(\mathcal{Y} \mid x)\, p_{\text{pri}}(x). \tag{3}$$

Let $x_i^{(t)}$ denote the position of the i-th particle at iteration t (for $i = 1, 2, \ldots, n$). Then, SPOS computes the position $x_i^{(t+1)}$ at the next iteration via

$$x_i^{(t+1)} = x_i^{(t)} - \frac{\lambda_t}{n} \sum_{j=1}^{n} k\left(x_j^{(t)}, x_i^{(t)}\right) \nabla U_t\left(x_j^{(t)}\right) \tag{4}$$

$$+ \frac{\lambda_t}{n} \sum_{j=1}^{n} \nabla_{x_j^{(t)}} k\left(x_j^{(t)}, x_i^{(t)}\right) - \frac{\lambda_t}{\beta} \nabla U_t\left(x_i^{(t)}\right) + \sqrt{\frac{2\lambda_t}{\beta}}\, \xi_i^{(t)},$$

where:

- λ_t and λ_t/β are step sizes that can be further preconditioned using methods like Adam [KB14];

- k is the commonly adopted *RBF kernel* [LW16, ZZCC20] controlling the repulsive force between particles:

$$k(x_j, x_i) := \exp\left(-\frac{\log n}{M^2} \|x_j - x_i\|^2\right), \tag{5}$$

where M denotes the median of distances between all pairs of particles in $\mathcal{X}$.

- U_t is given by the logarithm of the target function f modulated by an *annealing* [DF21] term γ:

$$U_t(x) := -\gamma(t) \log f(x), \tag{6}$$

where $\gamma(t) := \tanh[(1.4t/T)^8]$ with T being the total number of iterations.

- $\xi_i^{(t)} \sim \mathcal{N}(0, I)$ is a Gaussian noise.

Intuitively, the algorithm works by letting these particles repel against each other to achieve divergence and collectively conform to the shape of the target distribution. The added Gaussian noise helps particles escape local modes, improving its robustness.

3.2.1. Comparison with simpler methods

While our core technique is orthogonal to the choice of posterior sampling algorithm and comparing the performance of various posterior sampling algorithms is not the focus of this paper, we nonetheless provide a comparison under a simple setting to illustrate the difference between SPOS and other simpler alternatives.

One naive way of drawing multiple appearance samples is to run an inverse rendering algorithm multiple times with different initial values. This indeed gives some variety in the results, which may be interpreted in the same way we treat samples from the posterior distribution. However, this is not a sampling algorithm and may produce biased results.

A popular inference algorithm is Langevin Monte Carlo (or SGLD) [WT11]. In practice, we find that SPOS offers improved accuracy and robustness compared to SGLD.

Table 1 shows the mean square error in estimating the mean and

variance of a posterior distribution arising in appearance acquisition, using samples drawn by different algorithms. The distribution represents a common scenario where diffuse reflectance and roughness are ambiguously inferred from a single reflectance observation. The Mean square errors of estimated expectance/variance of both diffuse reflectance and roughness are added together.

Algorithm	MSE (Expectance)	MSE (Variance)
SPOS	0.0023	0.0002
SGLD	0.0078	0.0010
multiple SGD	0.0045	0.0039

Table 1: *Mean square error of estimated expectance and variance from 16 samples drawn using different methods, averaged over 20 runs.*

3.3. Point Estimates and Uncertainties

With the material samples $\mathcal{X} := \{x_i : i = 1, 2, \ldots, n\}$ drawn from the posterior distribution (§3.2), we use these samples to (i) obtain a point estimate of the material appearance; and (ii) estimate the uncertainty associated with the observations $\mathcal{Y}$ and prior p_{pri}.

3.3.1. Obtaining Point Estimates

Material acquisition applications usually take *point estimates* (instead of posterior probabilities) of material parameters as output. Although obtaining point estimates is **not** our main focus, our technique is capable of providing those with little computational overhead.

Specifically, we provide maximum-a-posteriori (MAP) estimates x_{MAP} that maximize the posterior probability $p_{\text{post}}(x_{\text{MAP}} | \mathcal{Y})$ and hence the target function defined in Eq. (3):

$$x_{\text{MAP}} = \arg\max_x f(x). \tag{7}$$

To obtain x_{MAP}, we first set it to the position of the sampled particle where the target function is maximized: $x_{\text{MAP}} = \arg\max_{x \in \mathcal{X}} f(x)$. Then, as an optional step, we apply gradient-based optimization (e.g., Adam [KB14]) to further refine x_{MAP}.

3.3.2. Estimating Uncertainties

A main benefit of using a Bayesian formulation is the capability of estimating uncertainties. Specifically, our technique offers the capability of estimating uncertainties defined over both the *parameter space* and the *image space*—which we explain in the following.

Parameter-space uncertainty We first define the **parameter-space uncertainty** σ_x as the component-wise standard deviation of the posterior probability:

$$\sigma_x[j] := \text{std}_{x \sim p_{\text{post}}(x | \mathcal{Y})} x[j], \tag{8}$$

where $x[j]$ denotes the j-th component of x, and std indicates the standard deviation.

In practice, we approximate this uncertainty using the standard deviation of the samples $\mathcal{X} = \{x_i : i = 1, 2, \ldots, n\}$:

$$\sigma_x[j] \approx \text{std}\{x_i[j] : i = 1, 2, \ldots, n\}. \tag{9}$$

The parameter-space uncertainty captures how "tight" the posterior distribution is around any potential point estimate (such as x_{MAP}), and a greater standard deviation implies being more uncertain.

Image-space uncertainty Estimated material appearance parameters are commonly used for renderings under some novel illumination and/or viewing conditions. Given a novel condition specified with θ, we define the **image-space uncertainty** $\sigma_{\mathcal{R}}$ as an image given by

$$\sigma_{\mathcal{R}}(\theta)[j] := \text{std}_{x \sim p_{\text{post}}(x | \mathcal{Y})} \mathcal{R}(x; \theta)[j], \tag{10}$$

where $\sigma_{\mathcal{R}}(\theta)[j]$ and $\mathcal{R}(x; \theta)[j]$ denote the j-th pixels of the uncertainty image $\sigma_{\mathcal{R}}(\theta)$ and the rendering $\mathcal{R}(x; \theta)$ using the material parameter x and viewing/lighting configuration θ, respectively.

In practice, similar to the parameter-space uncertainty in Eq. (9), we approximate $\sigma_{\mathcal{R}}(\theta)$ using

$$\sigma_{\mathcal{R}}(\theta)[j] \approx \text{std}\{\mathcal{R}(x_i; \theta) : i = 1, 2, \ldots, n\}. \tag{11}$$

Discussion We note that higher parameter-space uncertainty does not always lead to higher image-space uncertainty because, for example, under smooth illumination, variations of surface roughness may only lead to small changes in object appearance.

Additionally, we emphasize that the uncertainties estimated using Eqs. (9) and (11) are **not** specific to any point-estimation methods—including our MAP-based one presented in §3.3.1—but statistical properties determined by the posterior probability. We demonstrate this in Figures 1 and 4.

3.4. Guiding Appearance Acquisitions

Besides quantifying uncertainties (§3.3), another application of our SPOS-based sampling (§3.2) is to guide appearance acquisition processes. Specifically, given a set of observations $\mathcal{Y}$, we introduce a technique to recommend viewing/lighting configurations θ so that additional observations acquired under these configurations are likely to be "informative" (and greatly reduce uncertainties).

In what follows, we introduce a formulation that quantifies the amount of information a *not-yet-acquired* observation with given viewing/lighting conditions is likely to offer in §3.4.1. Based on this formulation, we develop an algorithm for recommending new viewing/lighting configurations in §3.4.2.

3.4.1. Mutual Information

To this end, we consider the unknown observation $Y(\theta)$ taken under some fixed viewing/lighting condition θ. Motivated by the likelihood $L(\mathcal{Y} | x)$ defined in Eq. (2), we formulate $Y(\theta)$ as a random variable

$$Y(\theta) = R(\theta) + W, \tag{12}$$

where $R(\theta) := \mathcal{R}(X; \theta)$ is a random rendering generated using the same viewing/lighting conditions θ and random material parameters X subject to the posterior probability $p_{\text{post}}(x | \mathcal{Y})$, and the term W is an independent noise image with zero mean. In other

words, we treat the random rendering $R(\theta)$—which we can examine statistically—as an approximation of the unknown observation $Y(\theta)$.

Inspired by the inequality in information theory $h(A|B) \leq h(A)$, which states that having additional observation will decrease uncertainty, we formulate the selection of an informative configuration θ as maximizing such uncertainty reduction, i.e. the **mutual information** between material parameters X and the random rendering $R(\theta)$, which models the amount of information gained about one variable by observing the other variable [Sha48]:

$$h(X) - h(X\,|\,R(\theta)) = I(X \leftrightarrow R(\theta)) \\ = h(R(\theta)) - h(R(\theta)\,|\,X), \tag{13}$$

where h denotes the **differential entropy** and is defined as

$$h(Z) := -\int_{\Omega} p(z) \log p(z)\, \mathrm{d}\Omega(z), \tag{14}$$

for any random variable Z over some domain Ω with the probability density $p(z)$. At a high level, high differential entropy implies high uncertainty, and vice versa.

With fixed material parameters x and viewing/lighting conditions θ, we consider the corresponding rendering $\mathcal{R}(x;\theta)$ a mostly constant image except for having a small amount of pixel-wise independent rendering noise which can result from, for example, Monte Carlo integration and is not to be confused with the noise W. It follows that, on the right-hand side of Eq. (13), the conditional differential entropy $h(R(\theta)\,|\,X)$ emerges solely from the rendering noise, which we model as a small Gaussian noise when using sufficient render samples. Then, we approximate the maximization of the mutual information $I(X \leftrightarrow R(\theta))$ with maximizing the differential entropy $h(R(\theta))$:

$$\arg\max_{\theta} I(X \leftrightarrow R(\theta)) \approx \arg\max_{\theta} h(R(\theta)). \tag{15}$$

According to Eq. (14), estimating $h(R(\theta))$ requires evaluating the corresponding probability $p(r)$ which, unfortunately, has no analytical expression. To this end, we utilize kernel density estimation (KDE) [Sil86] as follows. Let $\mathcal{X} := \{x_i : i = 1, 2, \ldots, n\}$ be n samples of material parameters drawn using SPOS (§3.2). Then, for any image r, the probability density can be approximated via:

$$p^{\mathrm{KDE}}(r; \mathcal{X}, \theta) = \frac{1}{n\,\bar{\sigma}^k\,(2\pi)^{k/2}} \sum_{i=1}^{n} \exp\left(-\frac{\|r - \mathcal{R}(x_i;\theta)\|^2}{2\bar{\sigma}^2} \right), \tag{16}$$

where $\bar{\sigma} \in \mathbb{R}_+$ is the **kernel bandwidth**, k denotes the number of pixels in the image r, and $\mathcal{R}(x_i;\theta)$ is the image rendered using material parameters x_i and viewing/lighting conditions θ.

Unfortunately, since r is usually high-dimensional, for $p^{\mathrm{KDE}}(r)$ defined in Eq. (16) to accurately approximate the true probability $p(r)$, an impractically large number of samples (i.e., large n) may be needed. To mitigate this obstacle, we focus on the case when each dimension (i.e., pixel) $r[j]$ of the image r is approximately independent to the others, leading to

$$p(r) \approx \prod_{j=1}^{k} p(r[j]). \tag{17}$$

Then, the differential entropy $h(R(\theta))$ of the random image $R(\theta)$ can be computed by summing the entropy $h(R(\theta)[j])$ of the image's individual pixel $R(\theta)[j]$. That is,

$$I(X \leftrightarrow R(\theta)) = h(R(\theta)) - h(R(\theta)\,|\,X) \\ \approx \sum_{j=1}^{k} h(R(\theta)[j]) - h_{\mathrm{noise}}[j] \\ = -\sum_{j=1}^{k} \int p_j(r;\theta) \log p_j(r;\theta)\, \mathrm{d}r - h_{\mathrm{noise}}[j], \tag{18}$$

where $p_j(r;\theta)$ is the probability density function of the rendered j-th pixel; $h_{\mathrm{noise}}[j]$ quantifies the uncertainty of the j-th pixel due to render noise. The probability density p_j for the j-th pixel can be estimated using one-dimensional KDE via

$$p_j^{\mathrm{KDE}}(r; \theta, \mathcal{X}) = \frac{1}{n\,\bar{\sigma}_j\,\sqrt{2\pi}} \sum_{i=1}^{n} \exp\left(-\frac{(r - \mathcal{R}(x_i;\theta)[j])^2}{2\bar{\sigma}_j^2} \right). \tag{19}$$

In this equation, $\mathcal{R}(x_i;\theta)[j]$ denotes the j-th pixel of the rendering $\mathcal{R}(x_i;\theta)$, and $\bar{\sigma}_j$ is the kernel bandwidth for the j-th pixel.

In practice, we use Silverman's rule of thumb [Sil86]

$$\bar{\sigma}_j = 0.9\,n^{-1/5} \min(\sigma,\ \mathrm{IQR}/1.34), \tag{20}$$

where $\sigma_j := \mathrm{std}\{\mathcal{R}(x_i;\theta)[j] : i = 1, 2, \ldots, n\}$ is the sample standard deviation, and IQR is the interquartile range defined as the difference between 75-th and 25-th percentile of the samples.

We note that although image-space uncertainty (standard deviation) can also be used in place of the mutual information, i.e. maximizing image-space uncertainty instead of the mutual information of the chosen configuration, it's both less theoretically grounded and offering less uncertainty reduction in practice.

Numerical estimation In practice, given a set $\mathcal{X}$ of sampled material parameters and viewing/lighting configurations θ, we estimate Eq. (18) using Monte Carlo estimation. Specifically, for each $j = 1, 2, \ldots, k$, we draw r_j with the probability density $p_j^{\mathrm{KDE}}(r_j; \theta, \mathcal{X})$—which can be achieved by (i) sampling $r_{j,i} \in \mathbb{R}$ from the normal distribution $\mathcal{N}(\mathcal{R}(x_i;\theta), \bar{\sigma}_j^2)$ for $i = 1, 2, \ldots, n$; and (ii) choosing r_j from $\{r_{j,1}, r_{j,2}, \ldots, r_{j,n}\}$ uniformly at random. With all r_j sampled, we obtain the following single-sample Monte Carlo estimator for the differential entropy $h(R(\theta))$:

$$\langle h(R(\theta)) \rangle = -\sum_{j=1}^{k} \log p^{\mathrm{KDE}}(r_j; \theta, \mathcal{X}). \tag{21}$$

Handling redundant pixels When inferring SVBRDF parameters using high-resolution renderings/observations, multiple pixels may correspond to a single texel, violating our assumption of pixel-wise independency.

We address this problem by analyzing pixel footprint in the texture space. Precisely, if one texel contains the footprints of multiple pixels (under the viewing configuration encoded in θ), we treat all these pixels identical. When formulating $h(R(\theta))$ using Eq. (18), we exclude all but one of these pixels from the summation on the right-hand side.

ALGORITHM 1: Material acquisition guided with our technique

1 `GuidedAcquisition`$(\mathcal{Y}, \bar{\Theta}, m; n)$
   Input: Preexisting observations $\mathcal{Y}_0$; candidate viewing/lighting
       configurations $\bar{\Theta}$; number m of additional observations
   Output: Selected configurations Θ with $|\Theta| = m$
2 **begin**
3    $\mathcal{Y} \leftarrow \mathcal{Y}_0$;
4    $\Theta \leftarrow \emptyset$;
5   **while** $|\Theta| < m$ **do**
6   Draw n samples $\mathcal{X}$ from $p_{\text{post}}(x | \mathcal{Y})$ using SPOS ;   // §3.2
7       $\theta^* \leftarrow \arg\max\{I(X \leftrightarrow R(\theta)) : \theta \in \bar{\Theta}\}$;   // Eq. (18)
8   Acquire the observation y^* using the configuration θ^*;
9       $\mathcal{Y} \leftarrow \mathcal{Y} \cup \{y^*\}$;
10      $\Theta \leftarrow \Theta \cup \{\theta^*\}$;
11   **end**
12   **return** Θ;
13 **end**

3.4.2. The Guiding Algorithm

With the mutual information formulated using Eq. (18), we now present an algorithm that guides material acquisition processes.

Specifically, let $\bar{\Theta} := \{\theta_1, \theta_2, \ldots\}$ be a (discrete) set of candidate viewing/lighting configurations, and $\mathcal{Y}_0 := \{y_1, y_2, \ldots\}$ be a set of (already-conducted) observations. Given the number m of additional observations needed, our method selects the m-element subset $\Theta \subseteq \bar{\Theta}$ so that, under these configurations, the observations are likely to maximize information (and minimize uncertainty).

To this end, as outlined in Algorithm 1, we conduct a greedy process that iteratively selects the configuration θ^* that maximizes the mutual information $I(X \leftrightarrow R(\theta))$ given by Eq. (18).

Batched recommendations In Algorithm 1, the SPOS sampling (Line 6) and mutual information computation (Line 7) steps can both be computationally intensive. To improve the performance of Algorithm 1, we introduce an approximation that allows a batch of observations to be added in each iteration. Specifically, after drawing the samples $\mathcal{X}$ (Line 6), our goal is to select K configurations $\theta_1^*, \theta_2^*, \ldots, \theta_K^* \in \bar{\Theta}$ (where K is a hyper-parameter).

We note that adding one observation with configuration θ^* would cause all the other configurations similar to θ^* to be non-informative afterward. Based on this observation, to select the K configurations, we repeatedly select the top candidate $\theta^* \in \bar{\Theta}$ (that maximizes mutual information) before removing all candidates that reside within a hyper-sphere centered at θ^* from the set $\bar{\Theta}$.

4. Results

We implement the algorithms described in §3 in PyTorch with the rendering function $\mathcal{R}$ handled with a GPU-based differentiable renderer based on [ZZC*22]. On a workstation equipped with an NVIDIA RTX 4090, our uncertainty estimation takes a couple to 20 minutes depending on the required number of renderings—That is, how many times $\mathcal{R}$ needs to be evaluated in Eq. (19).

Ablation of particle count As presented in §3.2, the SPOS method we utilize for drawing sampling from posterior probabilities maintains and updates the positions of some n particles with n being a hyper-parameter. In Figure 2, we conduct an ablation on the value of n by estimating parameter-space uncertainty using Eq. (9) for a synthetic SVBRDF inference problem using 20 observations. As shown in this figure, the estimated uncertainty approximately converges after $n \geq 16$. Hence, we use $n = 20$ for all the following experiments.

4.1. Uncertainty Estimation

A main component of our technique is the estimation of parameter-space and image-space uncertainties (§3.3)—which we evaluate using several SVBRDF inference problems.

Evaluation setup Given a fixed set of observations of an object (with known geometry and lighting), we estimate uncertainties emerging from inferring the SVBRDF parameters expressed as per-texel diffuse albedo and roughness (based on the microfacet model). To evaluate our estimated uncertainties, we obtain (one or multiple) point estimates and examine the correlation between the uncertainties and (parameter- or image-space) errors.

As discussed at the end of §3.3, our estimated uncertainties are not specific to any point estimate but rather quantify the intrinsic ambiguity emerging from the posterior (determined by the observations and the priors).

Parameter-space uncertainty We evaluate our estimation of parameter-space uncertainty in Figure 3, where the SVBRDF of a planar wooden object is inferred using different numbers of observations (under both collocated and environment light configurations). As shown in this figure, our estimated parameter-space uncertainties do reduce when the number of observations increases. On the other hand, these uncertainties do not correlate closely with point estimate error maps. This is somewhat expected because, when there is high uncertainty, one point estimate can still get "lucky" and return an answer with low error. In other words, we expect to observe low error in regions with low uncertainty and, on average, higher error in regions with high uncertainty—which is the case in this result.
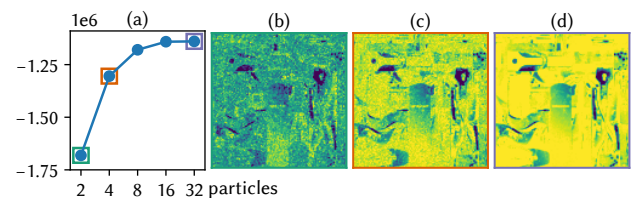

Figure 2: *We conduct an **ablation** to study how our uncertainty estimation is affected by the number n of particles used by the SPOS process. To this end, we obtain multiple estimates of the parameter-space uncertainty using Eq. (9) for a synthetic SVBRDF inference problem using $n = 2, 4, 8, 16, 32$. Column (b–d) visualize estimated uncertainties in surface roughness using $n = 2, 4, 32$, respectively.*
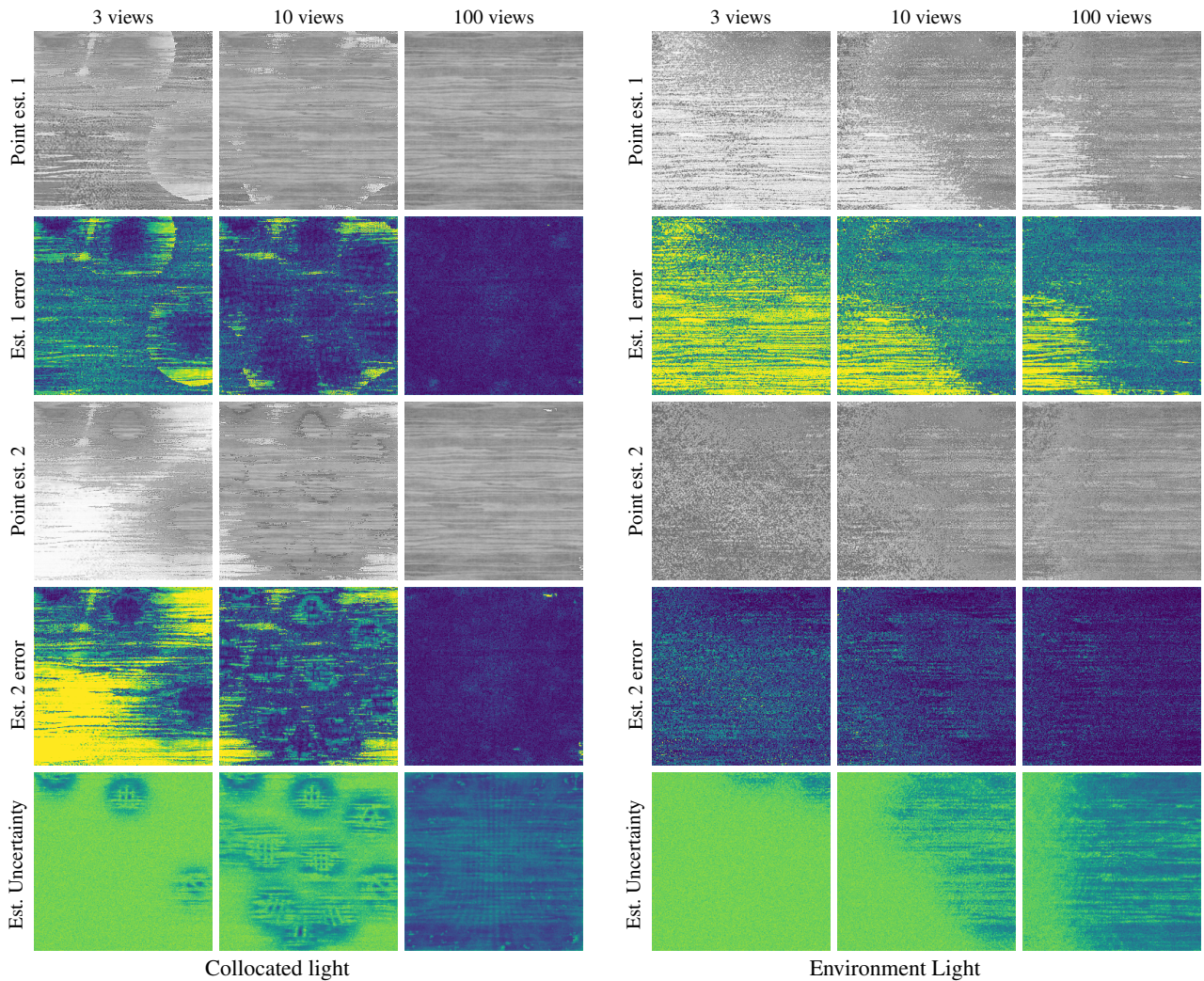
Figure 3: *Parameter-space uncertainty: We estimate parameter-space uncertainties using a synthetic SVBRDF inference problem where a wooden plane is observed from different directions under two different lighting configurations. In the 1st and 3nd row, we show 2 different point estimates of surface roughness using 3, 10, and 100 observations. In the 2nd and 4th row, we show L1 error of these point estimates. In the last row, we show our estimated parameter-space uncertainties of surface roughness. The error and uncertainty visualizations use the same color map as Figure 1.*

Image-space uncertainty We now evaluate our estimations of image-space uncertainty.

Figure 4 shows results obtained using two synthetic SVBRDF inference problems. For each problem, we compare our estimated uncertainties with the errors of novel renderings of four point estimates obtained with: (b1) Adam optimizer with high initial roughness; (b2) Adam optimizer with low initial roughness; (b3) SGD optimizer with random initial roughness; (b4) our point estimate described in §3.3.1. As shown in column (c), our estimated uncertainties correlate with the errors closely.

This result demonstrates that image-space uncertainties—which take into consideration not only posterior probabilities but also rendering configurations—tend to be more informative than parameter-space uncertainties.

In addition, Figure 1 shows an example obtained using an SVBRDF inference problem with real measurements from the work by Luan et al. [LZBD21]. Similar to the synthetic experiment (Figure 4), our predicted image-space uncertainty correlates well with rendering errors under novel configurations.

Prior awareness As discussed in §3.3, our estimation of parameter-space and image-space uncertainties are prior-aware due to the fact that the prior probability is a factor of the posterior probability. We demonstrate such prior awareness in Figure 5 that uses a simple setting where the SVBRDF of a homogeneous object is inferred using observations under collocated configurations. Without

a strong prior on roughness (shown on the top), the point estimate suffers from severe baking. With a prior modeled as a Gaussian around the groundtruth roughness, the point estimate matches the groundtruth much better with little baking. Our estimated uncertainties successfully capture these effects, as shown in (c).

Using Neural priors Lastly, we show a proof-of-concept example that goes beyond per-texel SVBRDF capture. Fig. 6 uses an example where the material reflectance is modeled using a neural prior—the MaterialGAN [GSH*20]. In other words, instead of directly specifying diffuse albedo and roughness per texel, this models specifies material appearance uses a latent representation which is then decoded (using a pre-trained GAN network) into the SVBRDF maps.

We apply our algorithm on the latent space of MaterialGAN, inferring the posterior distribution of $\theta = \{\mathbf{w}, \mathbf{z}\}$ where $\mathbf{w}, \mathbf{z}$ are the latent vector and noise respectively, in the same practice as [GSH*20]. We do not apply any additional prior on the latent space, hence a uniform prior $p_{\text{pri}}(\theta) = 1$ is used.

Our estimation of image-space uncertainties generalizes well to this setup. As shown in the figure, our estimated uncertainties correlate well with errors in novel renderings.

4.2. Guiding Captures

Lastly, we evaluate our capture guidance method (§4.2) using two synthetic settings in Figure 7.

The top and bottom sections of this figure use, respectively, SVBRDF inference problems where the handplane model is being observed under collocated and light-stage-like configurations, respectively. For the collocated light configuration, we have 1000 candidate camera positions. For the light stage configuration, we have 5000 candidate camera positions and 30 available lights.

We compare rendering errors (under novel configurations) and estimated image-space uncertainty using observations determined by two methods:

- **Baseline method:** As a baseline, we use a low-discrepancy sampler to evenly sample viewing directions. For the light stage configuration, we pair each camera with a randomly selected light.

- **Our method:** Using our guidance scheme, for the collocated light configuration we recommend 5 views in addition to 5 uniformly drawn viewing configurations (using the same low-discrepancy sampler), and in each subsequent round add 10 more recommended cameras. For the light stage configuration, we start with 3 uniformly drawn viewing configurations and recommend 3 views in each round.

As shown in Figure 7, observations recommended by our method consistently outperforms those given by the baseline method. Specifically, as shown in the uncertainty (c) and rendering error plots (b), using the same number of total observations, our technique offers consistently lower image-space uncertainty (averaged over all pixels) as well as rendering error (using one point estimate under a novel configuration).

Below the plots, we also show novel renderings, rendering errors, and estimated uncertainties obtained with 10 observations under collocated light setup and 9 observations under light stage setup using both methods.

5. Discussion and Conclusion

Limitations and future work As discussed in §1, errors in reflectance acquisition can emerge from not only ambiguities but also limitations of appearance models. Our method does not model the latter and, thus, can fail when that type of error is severe.

When the computational budget for rendering is limited, the render noise will be a significant factor that affects estimating uncertainties. Formulating and practically handling render noise will be a challenging topic that we leave for future work.

In addition, our technique—especially the guiding method discussed in §3.4—remains computationally expensive due to the need of generating large numbers of renderings (when evaluating Eq. (19) for each particle j and candidate configuration θ). Accelerating this process will drastically improve its practical usefulness in, for example, an interactive appearance reconstruction pipeline.

Conclusion In this paper, we introduced a Bayesian formulation to quantify ambiguities in material appearance acquisition. Specifically, given a set of observations and priors, our technique uses stochastic particle-optimization sampling (SPOS) to efficiently draw material parameter samples from posterior probabilities determined by the input observations and priors. Based on this process, we developed a method to estimate uncertainties over the parameter-space and the image-space. In addition, we proposed a pipeline that guides acquisition processes by iteratively recommending viewing/lighting configurations for making future observations. We evaluated our methods using several synthetic and one real examples.

References

[AAL16] AITTALA M., AILA T., LEHTINEN J.: Reflectance modeling by neural texture synthesis. *ACM Trans. Graph. 35*, 4 (2016), 1–13. 2

[CDP*14] CHEN G., DONG Y., PEERS P., ZHANG J., TONG X.: Reflectance scanning: estimating shading frame and BRDF with generalized linear light sources. *ACM Trans. Graph. 33*, 4 (2014), 1–11. 2

[CFG14] CHEN T., FOX E., GUESTRIN C.: Stochastic gradient hamiltonian monte carlo. In *International conference on machine learning* (2014), PMLR, pp. 1683–1691. 2

[CG20] CHEN P., GHATTAS O.: Projected stein variational gradient descent. *Advances in Neural Information Processing Systems 33* (2020), 1947–1958. 3

[CWC*19] CHEN P., WU K., CHEN J., O'LEARY-ROSEBERRY T., GHATTAS O.: Projected stein variational newton: A fast and scalable bayesian inference method in high dimensions. *Advances in Neural Information Processing Systems 32* (2019). 3

[DAD*19] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Flexible SVBRDF capture with a multi-image deep network. In *Computer Graphics Forum* (2019), vol. 38, Wiley Online Library, pp. 1–13. 2

[DCM*18] DETOMMASO G., CUI T., MARZOUK Y., SPANTINI A., SCHEICHL R.: A stein variational newton method. *Advances in Neural Information Processing Systems 31* (2018). 3

[DCP*14] DONG Y., CHEN G., PEERS P., ZHANG J., TONG X.: Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Trans. Graph. 33*, 6 (2014), 1–12. 2

[DF21] D'ANGELO F., FORTUIN V.: Annealed stein variational gradient descent, 2021. arXiv:2101.09815. 2, 3

[DWMG15] DONG Z., WALTER B., MARSCHNER S., GREENBERG D. P.: Predicting appearance from measured microgeometry of metal surfaces. *ACM Trans. Graph. 35*, 1 (2015), 1–13. 2

[DWT*10] DONG Y., WANG J., TONG X., SNYDER J., LAN Y., BEN-EZRA M., GUO B.: Manifold bootstrapping for SVBRDF capture. *ACM Trans. Graph. 29*, 4 (2010), 1–10. 2

[FBLS07] FUCHS M., BLANZ V., LENSCH H. P., SEIDEL H.-P.: Adaptive sampling of reflectance fields. *ACM Transactions on Graphics (TOG) 26*, 2 (2007), 10–es. 2

[GHYZ20] GUO Y., HAŠAN M., YAN L., ZHAO S.: A bayesian inference framework for procedural material parameter estimation. *Computer Graphics Forum 39*, 7 (2020), 255–266. 2

[GLD*19] GAO D., LI X., DONG Y., PEERS P., XU K., TONG X.: Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images. *ACM Trans. Graph. 38*, 4 (2019), 134–1. 2

[GSH*20] GUO Y., SMITH C., HAŠAN M., SUNKAVALLI K., ZHAO S.: MaterialGAN: Reflectance capture using a generative SVBRDF model. *ACM Trans. Graph. 39*, 6 (2020), 254:1–254:13. 2, 8, 11

[HLZ10] HOLROYD M., LAWRENCE J., ZICKLER T.: A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Trans. Graph. 29*, 4 (2010), 1–12. 2

[HSL*17] HUI Z., SUNKAVALLI K., LEE J.-Y., HADAP S., WANG J., SANKARANARAYANAN A. C.: Reflectance capture using univariate sampling of BRDFs. In *ICCV* (2017), IEEE, pp. 5362–5370. 2

[KB14] KINGMA D. P., BA J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014). 1, 3, 4

[KCW*18] KANG K., CHEN Z., WANG J., ZHOU K., WU H.: Efficient reflectance capture using an autoencoder. *ACM Trans. Graph. 37*, 4 (2018), 127–1. 2

[KGT*17] KIM K., GU J., TYREE S., MOLCHANOV P., NIESSNER M., KAUTZ J.: A lightweight approach for on-the-fly reflectance estimation. In *ICCV* (2017), IEEE, pp. 20–28. 2

[LKG*03] LENSCH H. P., KAUTZ J., GOESELE M., HEIDRICH W., SEIDEL H.-P.: Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graph. 22*, 2 (2003), 234–257. 2

[LLSS03] LENSCH H. P., LANG J., SÁ A. M., SEIDEL H.-P.: Planned sampling of spatially varying brdfs. In *Computer graphics forum* (2003), vol. 22, Wiley Online Library, pp. 473–482. 2

[LTH*23] LYU L., TEWARI A., HABERMANN M., SAITO S., ZOLL-HÖFER M., LEIMKÜHLER T., THEOBALT C.: Diffusion posterior illumination for ambiguity-aware inverse rendering. *ACM Trans. Graph. 42*, 6 (dec 2023). URL: https://doi.org/10.1145/3618357, doi:10.1145/3618357. 2

[LW16] LIU Q., WANG D.: Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in neural information processing systems 29* (2016). 2, 3

[LZBD21] LUAN F., ZHAO S., BALA K., DONG Z.: Unified shape and svbrdf recovery using differentiable Monte Carlo rendering. *Computer Graphics Forum 40*, 4 (2021), 101–113. 1, 2, 7

[Mat03] MATUSIK W.: *A data-driven reflectance model*. PhD thesis, Massachusetts Institute of Technology, 2003. 2

[MRR*53] METROPOLIS N., ROSENBLUTH A. W., ROSENBLUTH M. N., TELLER A. H., TELLER E.: Equation of state calculations by fast computing machines. *The journal of chemical physics* (1953). 2

[NJR15] NIELSEN J. B., JENSEN H. W., RAMAMOORTHI R.: On optimal, minimal BRDF sampling for reflectance acquisition. *ACM Trans. Graph. 34*, 6 (2015), 1–11. 2

[PNS18] PARK J. J., NEWCOMBE R., SEITZ S.: Surface light field fusion. In *3DV* (2018), IEEE, pp. 12–21. 2

[Sha48] SHANNON C. E.: A mathematical theory of communication. *The Bell system technical journal 27*, 3 (1948), 379–423. 5

[Sil86] SILVERMAN B. W.: *Density estimation for statistics and data analysis*. Chapman & Hall/CRC monographs on statistics and applied probability. Chapman and Hall, London, 1986. URL: https://cds.cern.ch/record/1070306. 5

[WT11] WELLING M., TEH Y. W.: Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)* (2011), pp. 681–688. 2, 3

[WTBL19] WANG D., TANG Z., BAJAJ C., LIU Q.: Stein variational gradient descent with matrix-valued kernels. *Advances in neural information processing systems 32* (2019). 2

[WWZ15] WU H., WANG Z., ZHOU K.: Simultaneous localization and appearance estimation with a consumer rgb-d camera. *IEEE TVCG 22*, 8 (2015), 2012–2023. 2

[XNY*16] XU Z., NIELSEN J. B., YU J., JENSEN H. W., RAMAMOORTHI R.: Minimal BRDF sampling for two-shot near-field reflectance acquisition. *ACM Trans. Graph. 35*, 6 (2016), 1–12. 2

[YDMH99] YU Y., DEBEVEC P., MALIK J., HAWKINS T.: Inverse global illumination: Recovering reflectance models of real scenes from photographs. In *Proc. Computer graphics and interactive techniques* (1999), pp. 215–224. 2

[ZCD*16] ZHOU Z., CHEN G., DONG Y., WIPF D., YU Y., SNYDER J., TONG X.: Sparse-as-possible SVBRDF acquisition. *ACM Trans. Graph. 35*, 6 (2016), 1–12. 2

[ZLS*18] ZHUO J., LIU C., SHI J., ZHU J., CHEN N., ZHANG B.: Message passing stein variational gradient descent. In *International Conference on Machine Learning* (2018), PMLR, pp. 6018–6027. 2

[ZZC*22] ZHENG S., ZHOU Z., CHEN X., YAN D., ZHANG C., GENG Y., GU Y., XU K.: Luisarender: A high-performance rendering framework with layered and unified interfaces on stream architectures. *ACM Trans. Graph. 41*, 6 (nov 2022). URL: https://doi.org/10.1145/3550454.3555463, doi:10.1145/3550454.3555463. 6

[ZZCC20] ZHANG J., ZHANG R., CARIN L., CHEN C.: Stochastic particle-optimization sampling and the non-asymptotic convergence theory. In *International Conference on Artificial Intelligence and Statistics* (2020), PMLR, pp. 1877–1887. 2, 3
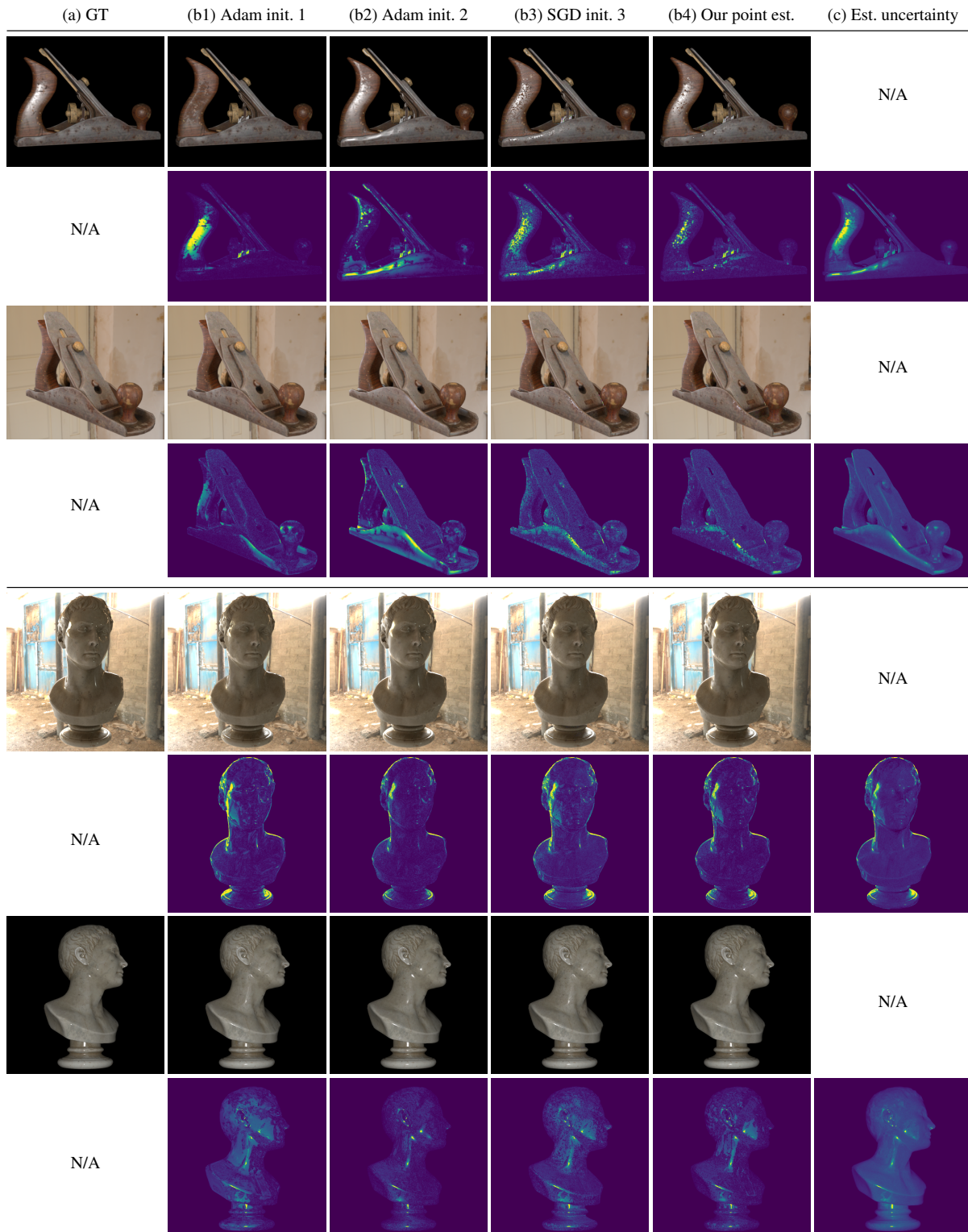
Figure 4: *Image-space uncertainty:* We compare our estimated image-space uncertainties for two objects each under two novel lighting/viewing conditions. The handplane on the top and bust on the bottom have their SVBRDFs inferred using 20 observations under collocated and environmental illuminations, respectively. In this figure, we show renderings of the inferred reflectance using four point estimates (b1–b4) under novel viewing/lighting conditions. Our estimated image-space uncertainties (c) correlate closely with the rendering errors.
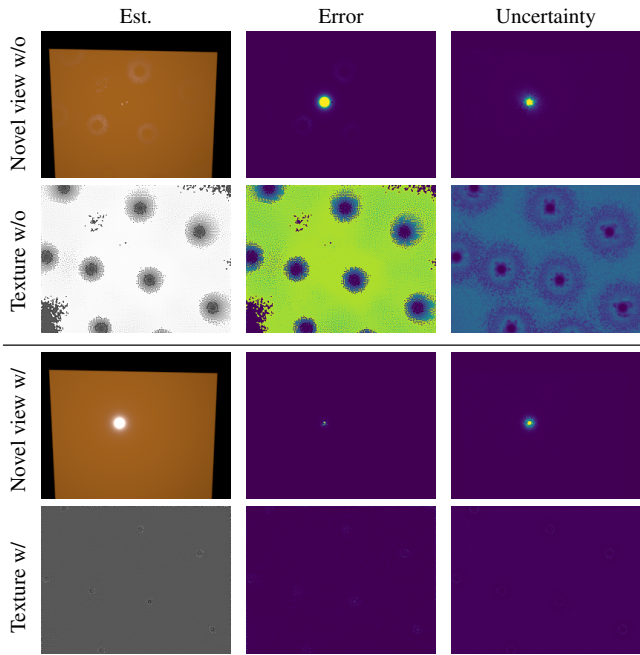
Figure 5: ***Prior awareness:*** *Reconstruction of a smooth plane captured from 10 angles without and with the low-roughness prior.*
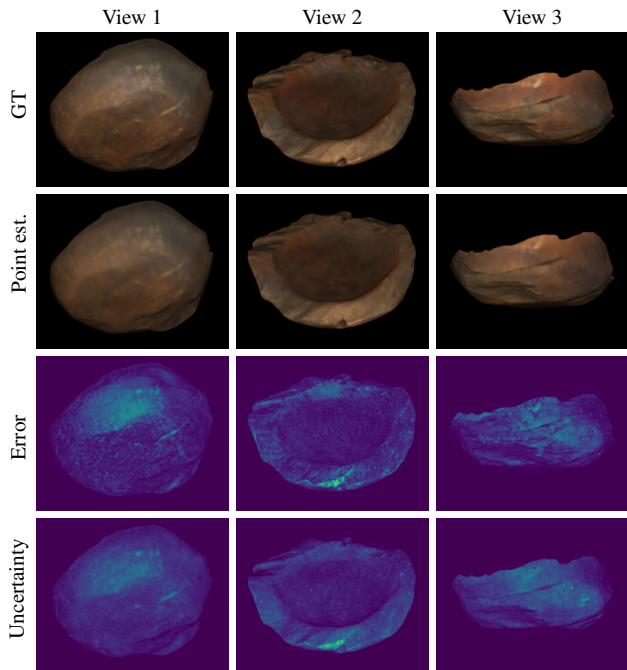

Figure 6: ***Using Neural priors:*** *Appearance of a wooden bowl reconstructed using MaterialGAN [GSH*20], captured from 7 views.*
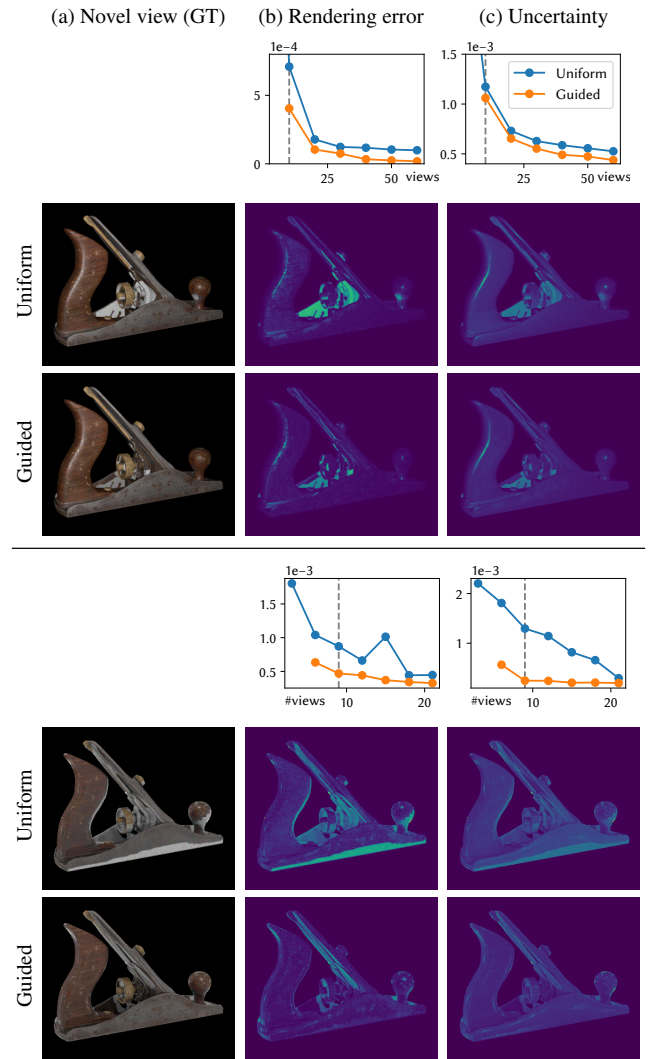

Figure 7: ***Guiding captures:*** *We show two examples with each using an SVBRDF inference problem using varying numbers of observations. For both examples, using observations recommended by our method leads to lower uncertainties (c), allowing point estimates to produce lower error under novel configurations (b). Gray dashed line illustrates number of views used in shown examples.*