

Practical Measurements of Translucent Materials with Inter-Pixel Translucency Prior

Zhenyu Chen¹, Jie Guo^{1*}, Shuichang Lai¹, Ruoyu Fu¹, Mengxun Kong¹, Chen Wang¹
Hongyu Sun², Zhebin Zhang², Chen Li², Yanwen Guo^{1†}

¹Nanjing University, Nanjing, China

²OPPO, Seattle, USA

{chenzy, sclai, 502022330022, chenwang}@smail.nju.edu.cn

{hongyu.sun, zhebin.zhang, chen.li}@oppo.com

{guojie, fry, ywguo}@nju.edu.cn



Figure 1. Different types of translucent materials recovered from ordinary photographs (insets) by our practical method.

Abstract

Material appearance is a key component of photorealism, with a pronounced impact on human perception. Although there are many prior works targeting at measuring opaque materials using light-weight setups (e.g., consumer-level cameras), little attention is paid on acquiring the optical properties of translucent materials which are also quite common in nature. In this paper, we present a practical method for acquiring scattering properties of translucent materials, based solely on ordinary images captured with unknown lighting and camera parameters. The key to our method is an inter-pixel translucency prior which states that image pixels of a given homogeneous translucent material typically form curves (dubbed translucent curves) in the RGB space, of which the shapes are determined by the parameters of the material. We leverage this prior in a specially-designed convolutional neural network comprising multiple encoders, a translucency-aware feature fusion module and a cascaded decoder. We demonstrate, through both visual comparisons and quantitative evaluations, that high accuracy can be achieved on a wide range of real-world translucent materials.

1. Introduction

A core demand in 3D content generation is to recover physically-plausible material properties from photographs. Knowledge of the recovered material properties is valuable in a broad range of applications including virtual/augmented reality [22], visual perception [16], 3D printing [31], robot control [54] and aesthetic medicine [41]. Over the past years, much effort has been devoted to acquiring materials from opaque objects that only contain surface reflection [2, 3, 14, 23–25, 30, 39, 40, 45, 48], excluding many real-world translucent materials (e.g., wax, soaps, plastics and beverages shown in Fig. 1) in which sub-surface scattering dominates [6, 34]. For these translucent objects, light will not only reflect at the boundaries, but also undergo a considerable number of scattering or absorption events inside media, resulting in a very complex light field. Often, each pixel of a photograph capturing a translucent object receives light from all points along the line of sight inside the volume [21].

To alleviate the problem, most early works analyze only single scattering in optically thin media [21, 27, 46], covering a very limited range of translucent material in nature. To deal with optically thick media, dedicated hardware systems are necessary, including structured lighting devices,

lasers and expensive high-end cameras [18, 31, 44]. Generally, repetitive calibrations are carried out every few minutes during the capture [31], making the whole process tedious and time-consuming. An alternative choice, which is very appealing recently, is resorting to powerful differentiable renderers [33, 58, 64–66]. Although these methods avoid cumbersome acquisition facilities, they still need camera and lighting parameters which are not always available for low-cost setups. Perhaps more importantly, current differentiable rendering methods are usually hard to converge in inverse rendering optimizations, when initial values are not selected properly [63].

Our goal in this paper is to significantly lower the acquisition cost and to further facilitate the capturing process. To this end, we propose a deep learning-based method that only needs photographs taken by off-the-shelf commodity cameras or mobile devices as the input, without knowing the parameters of lighting, camera and even geometries. Material priors learned from large-scale datasets can lower the ill-posedness of the problem [5], but they tend to generate large biases, manifested as color drifting and incorrect opacity, when training and testing on different datasets. Note that this is always the case since datasets of labeled real-world translucent materials are absent currently. We address this issue by introducing an *inter-pixel translucency prior* based on two observations: 1) image pixels of a homogeneous translucent material typically form curves (dubbed *translucent curves*) that pass through the origin in the RGB space, and 2) their shapes are determined by the material’s properties. We use these observations to derive a *pseudo-albedo map* and a *bias map* for the specific translucent material based on a regression algorithm, from multiple images captured under different lighting conditions.

Based on the inter-pixel translucency prior, we train a neural network *IPTNet* to predict translucent material properties from input images, without knowing the lighting and camera parameters. *IPTNet* is a multi-branch convolutional neural network (CNN) that takes multiple images of the same translucent materials and the derived pseudo-albedo/bias maps as the input. Specifically, the pseudo-albedo map guarantees the correctness of the predicted single-scattering albedo, while the bias map lets *IPTNet* focus more on translucent edges that are crucial to the prediction of the extinction coefficient [19]. A *translucency-aware feature fusion module* and a *cascaded decoder* are developed to further improve the accuracy of prediction.

In summary, we make the following contributions:

- We introduce an inter-pixel translucency prior to describe the distribution of image pixels for a given translucent object, making translucent material prediction accurate and reliable for real-world measurements.
- We design *IPTNet* to predict translucent material properties from multiple images taken under unknown lighting

and camera settings.

- We propose a translucency-aware feature fusion module and a cascaded decoder to further improve the accuracy of prediction.

2. Related Work

Prediction models for translucent materials. Simulating light propagation in translucent materials has been widely studied in graphics [47], building upon the radiative transfer equation (RTE) [4]. These materials usually produce very complex radiance fields, since they allow light to penetrate their surfaces and scatter light in a more complex fashion than opaque materials [11]. Over the past decades, quite a few prediction models for simulating light behavior in translucent materials have been proposed. Using these models inversely is the core of many measurement methods [13, 31]. For optically thick media in which high-order scattering dominates, radiative transport can be simplified by the diffusion approximation [8, 10, 34, 35, 59]. For optically thin media, the phase function (*e.g.*, Henyey-Greenstein (HG) [28], Rayleigh [37], and Lorenz-Mie [42, 43]) can impact appearance in a perceptually important way near thin geometric structures [12, 17].

Measurements of translucent materials. Measuring the optical properties of translucent materials is a notoriously difficult task, since a sensor almost always observes the combined effects of many scattering and absorption events. Therefore, matching the appearance of real translucent objects by manual adjustment of optical properties is nearly impossible [13]. Common measurement methods usually rely on dedicated hardware systems. For instance, integrating spheres are often used in the field of optics to find the the bulk optical properties of turbid media [51–53, 55]. When single scattering dominates, scattering parameters can be measured directly, using static or dynamic light scattering [36]. Narasimhan *et al.* [46] proposed a method based on dilution for measuring the scattering and absorption coefficients as well as the phase function. Gu *et al.* [21] recovered inhomogeneous participating media by projecting structured light. Unfortunately, these methods can only handle optically thin situations in which single scattering dominates. For moderately-thick samples, Mukaigawa *et al.* [44] projected high-frequency stripe patterns to separate single and multiple scattering. Gkioulekas *et al.* [18] proposed a material dictionary to facilitate the optimization of bulk scattering properties of homogeneous materials. To handle arbitrary translucent materials, differentiable rendering based methods [5, 33, 58, 64–66] prevail, but they are usually unstable [63].

Measurements of opaque materials. Our work is also closely related to surface appearance acquisition which has received continuous interest in computer vision and graphics [9, 61]. Traditional methods usually require significant

equipment and expense [7, 15, 20, 29, 38, 56]. To reduce the acquisition cost, it is attractive to use off-the-shelf commodity mobile devices [1, 45]. Inspired by the success of deep learning for a variety of vision and graphics tasks, recent work has considered CNN-based material acquisition [2, 3, 14, 23, 25, 39, 40, 45]. Notably, these opaque materials behave locally and do not consider particle scattering inside objects, making them easier to model than translucent materials.

3. Inter-Pixel Translucency Prior

3.1. Preliminaries

A translucent material containing particles that scatter and absorb light is represented by its macroscopic bulk optical parameters, typically consisting of the extinction coefficient σ_t , the single-scattering albedo Λ , and the phase function $p(\theta)$. For the phase function, we adopt the HG distribution which is parameterized by a single parameter: g . Alternatively, we can use the scattering coefficient σ_s and the absorption coefficient σ_a in place of σ_t and Λ , according to the relationship: $\sigma_t = \sigma_s + \sigma_a$ and $\Lambda = \sigma_s/\sigma_t$.

Generally, measuring the optical properties of a translucent material is difficult since the above optical properties cannot be easily separated. Often, two completely different sets of material parameters can present almost identical visual effects in certain circumstances, which is known as the principle of similarity [62, 67] for translucency. The first-order similarity offers another set of optical parameters:

$$\sigma'_s = (1 - g)\sigma_s, \quad \sigma'_t = \sigma'_s + \sigma_a, \quad \Lambda' = \sigma'_s/\sigma'_t \quad (1)$$

which are less correlated than other sets [5], and will be the target of our prediction. Moreover, the physical parameter space for translucent materials is extremely large and perceptually non-uniform [17, 57]. All these issues make translucent material acquisition challenging, especially under a light-weight setting: input images are captured by a consumer-level camera with unknown parameters.

3.2. Observations

To alleviate the problem, we introduce the inter-pixel translucency prior, which is derived from a variable number of images taken of the same homogeneous translucent object under uncalibrated lighting directions. We assume that the object is positioned in front of a black background and lit by a white illuminant (e.g., a build-in flash light of a mobile phone). Under this setting, we observe that image pixels from the same position across multiple images ($\mathbf{I}_i, i = 1, 2, \dots, n$) typically form a curve (or line) in the RGB space. We term these curves *translucent curves* to emphasize this characteristic. We validate and quantify this observation on both synthetic and real-world examples in Fig. 2. Here, we collect pixels across multiple images (shown

in the insets) of the same translucent material captured under varying lighting conditions. Two different patches¹ are shown for each example. We clearly see from Fig. 2 that pixels from the same position on the object are distributed roughly along curves. These curves pass through the origin and stem from shading variations within the object.

Let's analyze how these curves form and their relationship with the bulk optical properties of the underlying translucent material. For optically thick parts in which multiple scattering dominates, Pat and Wolfgang [26] have found that there is a non-linear relationship between the single-scattering albedo Λ' and the surface reflectance \mathbf{R} , which could be written as $\mathbf{R} = f(\Lambda')$. Since these parts reflect light in a similar way as opaque surfaces, the pixels in these areas take values proportional to \mathbf{R} , i.e., $\mathbf{I} \propto f(\Lambda')$. Therefore, when pixel values are plotted under different lighting conditions, they are roughly distributed along a line of which the gradient is determined by Λ' . For optically thin parts, e.g., the translucent edges [19], single scattering plays an important role. Pixel values in these parts are closely related to the contribution of the attenuated light paths inside the medium. The attenuation is determined by the extinction coefficient and usually follows an exponential law. It varies across channels, indicating that pixel values of different channels will have different rates of attenuation.

3.3. The pseudo-albedo map and bias map

Based on the observations, we derive a pseudo-albedo map \mathbf{A} and a bias map \mathbf{B} from multiple input images of the same translucent object, using an efficient regression method.

Assuming that the collected pixels at the same position (x, y) across multiple images form a potential line $\mathbf{I}^r(x, y)/k_{x,y}^r = \mathbf{I}^g(x, y)/k_{x,y}^g = \mathbf{I}^b(x, y)/k_{x,y}^b$ passing through the origin, where superscripts $\{r, g, b\}$ indicate channels, our method produces optimized per-pixel and per-channel slopes by:

$$\begin{aligned} k_{x,y}^1 &= \arg \min_{k_{x,y}^*} \sum_{i=1}^n (k_{x,y}^* \mathbf{I}_i^b(x, y) - \mathbf{I}_i^r(x, y))^2 \\ k_{x,y}^2 &= \arg \min_{k_{x,y}^*} \sum_{i=1}^n (k_{x,y}^* \mathbf{I}_i^b(x, y) - \mathbf{I}_i^g(x, y))^2 \end{aligned} \quad (2)$$

where $k_{x,y}^1 = k_{x,y}^r/k_{x,y}^b$ and $k_{x,y}^2 = k_{x,y}^g/k_{x,y}^b$. The optimization is conducted on a set of n images. As we could scale $\{k_{x,y}^r, k_{x,y}^g, k_{x,y}^b\}$ without alternating the direction of the line, it is safe to assume $k_{x,y}^b = 1$ and $k_{x,y}^1 = k_{x,y}^r$, $k_{x,y}^2 = k_{x,y}^g$. Then, the pixel value of \mathbf{A} is given by:

$$\mathbf{A}(x, y) = \mathcal{N}(k_{x,y}^r, k_{x,y}^g, 1) \quad (3)$$

¹To make the visualization in Fig. 2 clear, we analyze a small patch instead of a single pixel for each case.

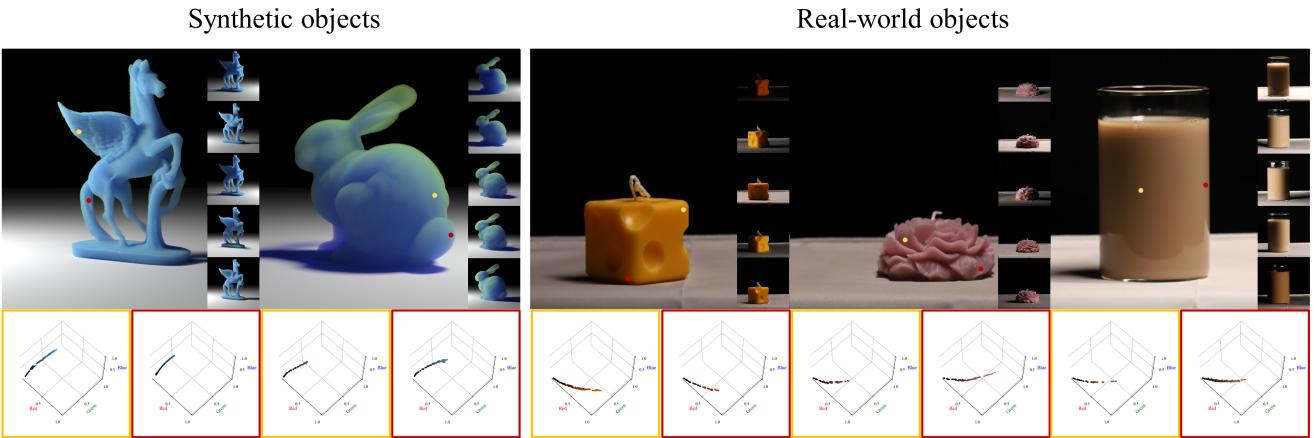


Figure 2. Illustration of the inter-pixel translucency prior on two synthetic (left) and three real-world (right) translucent objects. The last row shows that RGB space pixel values from the same position of a translucent material are distributed roughly along curves.

where \mathcal{N} denotes the vector normalization operation. The bias map \mathbf{B} evaluates the mean Euclidean distance between normalized image pixels from \mathbf{I}_i and pixels from \mathbf{A} :

$$\mathbf{B}(x, y) = \frac{1}{n} \sum_{i=1}^n [\mathcal{N}(\mathbf{I}_i(x, y)) - \mathbf{A}(x, y)]^2. \quad (4)$$

Generally, the pseudo-albedo map \mathbf{A} , encoding per-pixel pseudo-albedo $f(\Lambda')$, is not the real albedo map, but a proxy very close to it. The closeness depends on the optical thickness of the medium. The bias map \mathbf{B} marks translucent edges that do not generally follow a linear distribution of pixels, and serves as a weight map to reflect the spatially-varying optical thickness of a give translucent object. Intuitively, a pixel from the optically thick region reaches almost zero in the bias map \mathbf{B} , while a large pixel value in the bias map \mathbf{B} indicates that the corresponding pixel lies on a translucent edge. This information serves as an important cue for translucent material acquisition. Several examples of \mathbf{A} and \mathbf{B} are shown in Fig. 4.

4. IPTNet

The objective of IPTNet (Inter-Pixel Translucent Network) is to predict the bulk optical parameters of a homogeneous translucent material, with the help of the inter-pixel translucency prior.

4.1. Network architecture

Our IPTNet is a multi-image CNN that comprises several encoders and two specially-designed modules: a translucency-aware feature fusion module and a cascaded decoder, as illustrated in Fig. 3. To extract different material features \mathcal{F}_I in various image regions, a weight-sharing

encoder E_I is trained on multiple images to extract features from the optical characteristics of the material. The number of encoders is dynamically determined to match the number of inputs.

Additionally, two extra encoders E_A and E_B are adopted to extract features from the generated pseudo-albedo map and bias map. The translucency-aware feature fusion module tries to combine features from different parts. Leveraging the combined feature maps, the cascaded decoder predicts the scattering parameters step-by-step, thereby preserving both the global and local information of the material for the $\{\sigma'_t, g\}$ decoder branch.

4.2. Translucency-aware feature fusion

To better exploit translucent cues from input images, we introduce a translucency-aware feature fusion module. This module enables efficient extraction and summarization of features from multiple images while highlighting translucent edges that allow light to pass through. Specifically, the intermediate feature maps $\{\mathcal{F}_I^1, \dots, \mathcal{F}_I^n\}$ produced by the weight-sharing encoders E_I are fused into a single joint feature map \mathcal{F}_J by selecting the maximum value reported by any intermediate feature map at each pixel and channel. This process of max-pooling guarantees that each intermediate feature map has equal means to contribute to the content of the joint feature map, without smoothing out high-frequency features. To highlight the per-channel translucency, the bias map's feature map \mathcal{F}_B is multiplied with \mathcal{F}_J to produce a new weighted feature map \mathcal{F}_T :

$$\mathcal{F}_T = \mathcal{F}_J \otimes (E_B(\mathbf{B})) \quad (5)$$

where \otimes means element-wise multiplication. The motivation is that we let the network pay more attention to features

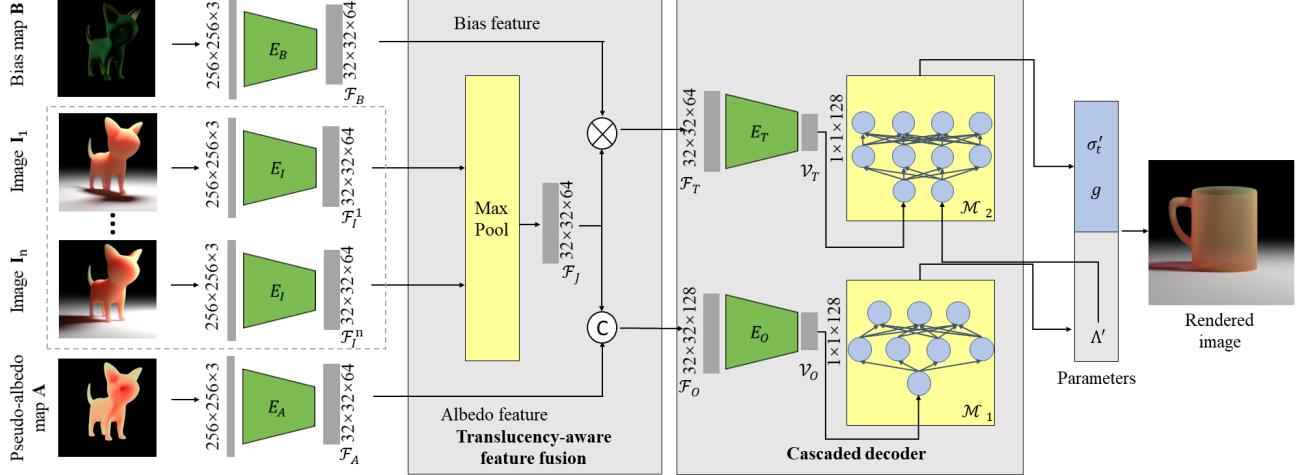


Figure 3. Network architecture of IPTNet. A set of n images $\{\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_n\}$ for a translucent material captured under different lighting conditions are resized to 256×256 pixels. From these images, we obtain a pseudo-albedo map **A** and a bias map **B**. These images pass through multiple encoders, a translucent-aware feature fusion module and a cascaded decoder, predicting a set of material parameters.

extracted from raw input images (rather than the pseudo-albedo map **A**) for optically thin parts. To achieve a similar goal, previous work [5] utilized image edge detection to produce meaningful weight maps. However, edge detection is incapable of detecting translucency features on different channels, and it may fail to detect edges on the dark regions.

Since the pseudo-albedo map **A** can offer supplementary details pertaining to the surface reflectivity of translucent materials [26], we directly concatenate its feature map \mathcal{F}_A with the joint feature map to form the feature map \mathcal{F}_O :

$$\mathcal{F}_O = \mathcal{F}_J \odot (E_A(A)) \quad (6)$$

where \odot denotes concatenation operation.

4.3. Cascaded decoder

As aforementioned, visual appearance of optically thick parts is primarily dominated by Λ' [26, 50], while optically thin parts which are highlighted by the bias maps reveal the effect of blending and correlation between all parameters [5, 19]. To enhance the accuracy of estimation, we propose a cascaded decoder that fully exploits the inherent optical property decomposition effect.

The cascaded decoder starts with two feature encoders, E_O and E_T , which convert \mathcal{F}_O and \mathcal{F}_T into latent vectors v_O and v_T , respectively. Following this, an MLP \mathcal{M}_1 is employed to predict Λ' . It is important to note that the global information of the material is concealed by the bias map, which is necessary as the translucent edges display the correlated effect of the scattering parameters. To preserve both the global information and the details of the translucent edges, we reintroduce the estimated global information of the material regarding Λ' into the decoding MLP \mathcal{M}_2 of

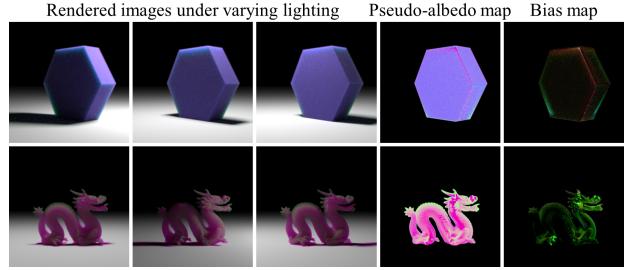


Figure 4. Two training examples from our dataset.

$\{\sigma'_t, g\}$. The whole process of the cascaded decoder can be summarized as

$$\begin{cases} \Lambda' = \mathcal{M}_1(E_O(\mathcal{F}_O)) \\ \{\sigma'_t, g\} = \mathcal{M}_2(E_T(\mathcal{F}_T), \Lambda'). \end{cases} \quad (7)$$

4.4. Dataset preparation

Generating a large number of training images and labels in the real world from previously acquired real scattering parameters [18, 46] is a tedious and time-consuming process that requires accurate and repetitive photo captures. Currently, only a few types of scattering parameters can be precisely acquired, which is insufficient to cover the parameter space of translucent materials. Alternatively, training a neural network using synthetic datasets has been proven to be a reasonable choice [1, 23]. In our work, we generate a synthetic dataset that is rendered using the physically based renderer, Mitsuba [60]. This renderer is capable of generating photorealistic images that capture a wide range of translucency effects.

Our synthetic dataset contains images of translucent objects with diverse geometries, illuminations, viewpoints and

scattering parameters. The dataset comprises 22 distinct shapes with a variety of thin and thick geometrical characteristics. Each scene is illuminated by a small area light source with arbitrary brightness and position and is observed from 5 different viewpoints. Inspired by the previous work [5], we choose the parameter space of translucent materials as $\sigma'_t \in [25\text{mm}^{-1}, 300\text{mm}^{-1}]$, $\Lambda' \in [0.3, 0.95]$, and HG phase functions with $g \in [0, 0.9]$. For each combination of shape and viewpoint, we render 10 images under different lighting conditions, resulting in a total of 16,800 groups of images in our dataset.

To assess the effectiveness and generalization ability of the trained model, we collect three additional test sets, each of which comprises 300 groups of images:

- the *TrView* test set contains 2 additional viewpoints which do not exist in the training set,
- the *TrLight* test set contains significantly different lighting configurations from those in the training set, and
- the *TrShape* test set contains 3 new shapes which are never seen in the training set, accompanied with novel viewpoints and lighting configurations.

5. Experiments

We implement our pipeline using PyTorch[49]. The learning rate is initialized to 0.00005, adjusted with the powered of 0.92 every five epochs. All other hyperparameters are set by default. The loss function computed on the predicted parameters is $l1$ norm. We train the network with a batch size of 4 for 80 epochs, taking about 18 hours on an NVIDIA Tesla V100 graphics card.

5.1. Experiments on synthetic data

We compare our method with two state-of-the-art approaches that can handle translucent material acquisition. Specifically, Inverse Transport Network (ITN) [5] is a neural network-based translucent material acquisition solution that combines a CNN and a differentiable rendering based loss. We also choose Mitsuba 3 [32] along with Dr.jit [33] as an optimization-based baseline. To verify the effectiveness of different methods on the three test sets, we adopt three metrics: 1) MSE over (normalized) parameters, 2) PSNR over rendered images, and 3) SSIM over rendered images. For each method, the predicted parameters of all three test sets will be applied on three different shapes : 1) a hexahedron model (see the first row of Fig. 4) labeled as H, 2) a duck model (see the second row of Fig. 6) labeled as D, and 3) a dragon model (see the second row of Fig. 4) labeled as R.

We first compare our method with three variants of ITN:

- The ITN(S) method uses a single image as its input, accompanying with a differentiable renderer.
- The ITN(M) method runs ITN(S) repeatedly on all 10 images in a test set group and obtains averaged results, en-

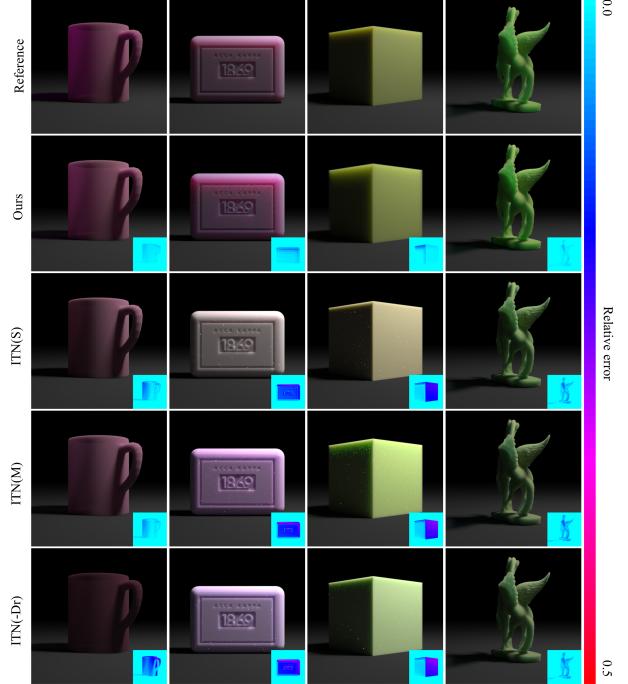


Figure 5. Visual comparison with different variants of ITN [5]. Each column shows rendered images with predicted optical parameters from different methods. The insets show error maps as compared with the reference.

abling fair comparisons with ours.

- The ITN(-Dr) method uses a single image as its input without differentiable rendering, avoiding potential errors caused by differentiable rendering.

Table 1 summarizes the quantitative results, showing that IPTNet achieves significant performance gains over all ITN variations on all three test sets in both the parameter domain and the image domain. It is noteworthy that IPTNet provides much more accurate prediction results on Λ' than other methods with the assistance of pseudo-albedo map and translucency-aware feature fusion module. This emphasizes the accurate restoration of chromaticity of opaque regions in Fig. 5, where ITN-based approaches offer scattering parameters with varying degrees of aberration on both translucent and opaque object parts.

We next compare our method with a differentiable rendering based inverse rendering method in Mitsuba 3 [32]. We only compare our method with it in the image domain due to the similarity relation [62]. We compare our approach with two variants of it using a stochastic starting point: 1) $\text{Dr}(\sigma'_t)$ optimizes only one parameter σ'_t and 2) $\text{Dr}(\sigma'_t, \Lambda', g)$ optimizes all parameters. Without precise knowledge of scene lighting and geometries, these differentiable rendering based methods are hard to converge. Even providing this information to these methods, they still

Table 1. Quantitative comparison with different variants of ITN [5] on three test sets: TrView, TrShape and TrLight. The best results of each group are highlighted in **bold**.

	TrView				TrShape				TrLight			
	Ours	ITN(S)	ITN(M)	ITN(-Dr)	Ours	ITN(S)	ITN(M)	ITN(-Dr)	Ours	ITN(S)	ITN(M)	ITN(-Dr)
σ'_t	0.014	0.021	0.015	0.023	0.026	0.035	0.027	0.036	0.015	0.026	0.038	0.047
	0.001	0.026	0.045	0.052	0.002	0.011	0.013	0.087	0.010	0.167	0.087	0.102
	0.025	0.200	0.079	0.505	0.033	0.068	0.055	0.053	0.027	0.196	0.065	0.408
H	38.68	34.88	35.94	34.02	37.41	33.91	34.58	33.52	36.34	28.41	29.62	29.01
	37.37	33.08	33.86	32.02	35.96	31.77	31.94	31.31	33.45	24.22	25.73	24.96
	38.87	34.76	35.77	33.89	37.55	33.70	34.07	33.29	35.74	27.31	28.82	28.06
PSNR ↓	0.993	0.972	0.975	0.964	0.993	0.986	0.983	0.984	0.988	0.950	0.954	0.950
	0.993	0.966	0.969	0.957	0.994	0.985	0.982	0.984	0.986	0.932	0.937	0.932
	0.996	0.977	0.979	0.971	0.996	0.988	0.987	0.988	0.990	0.950	0.954	0.951
MSE ↓												

achieve sub-optimal results and may even converge to incorrect results (see the last row of Fig. 6). Our IPTNet, without any additional information about lighting and geometries, beats these differentiable rendering based methods on all cases in Fig. 6. Moreover, the inverse rendering technique takes about 20 minutes to converge while our IPTNet only needs roughly 30 milliseconds to extract information from a scene on the same device.

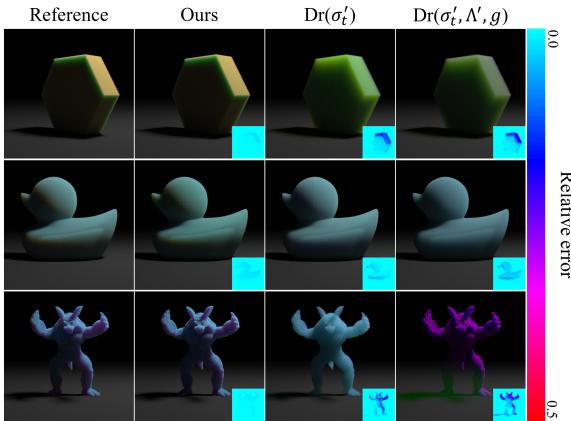


Figure 6. Visual comparison with a differentiable rendering (Dr) based optimization method from Mitsuba 3 [32].

5.2. Experiments on real-world examples

We conduct multiple experiments on a wide range of real-world translucent objects. Fig. 7 shows reconstruction results of three waxy cubes from our method and ITN(M). The input images were captured using a camera equipped with a standard LED area light source. The acquired LDR images can be utilized as input for our method following a straightforward post-processing step, inverse gamma correction, as our algorithm for generating the pseudo-albedo map and bias map disregards over-exposed or under-exposed pixels, which are infrequent and easy to identify.

Our method provides superior prediction results that benefit from the inter-pixel translucency prior, particularly on edges, in comparison to ITN(M), whose predicted scattering parameters exhibit reduced saturation and translucency on rendered images. As our method narrows the gap between real and virtual images, it demonstrates that the results supplied by our method almost match captured photos visually and statically.

Fig. 8 compares our method with a conventional 3D reconstruction method based on multi-view stereo (MVS). Without special designs, existing MVS methods fail to recover translucent appearance from ordinary images. They tend to produce incorrect diffuse reflectance from translucent objects. As expected, our method faithfully reproduces the translucent appearance which is highlighted by the back lit renderings (Back) in the fourth column of Fig. 8.

Our method is robust to a wide range of translucency. To show this, we test our method on waxy blocks with different thicknesses in Fig. 9 and orange juice at various concentrations in Fig. 10. Our recovered results closely match the reference images in all cases, even for objects with great translucency (the thin waxy block in the first column of Fig. 9 and the 16x diluted orange juice in the third column of Fig. 10). More results of real-world translucent materials accompanied with complex geometries are shown in Fig. 1.

5.3. Ablation studies

We perform some ablation studies to verify the validity of our design of the network architecture. We test three ablated models of IPTNet: 1) the *mean* model replaces the max pooling layer in IPTNet with mean pooling layer, 2) the *-pr* model predicts scattering parameters solely using images as its input without additional bias maps and pseudo-albedo maps, and 3) the *single* model uses a single decoder to predict all parameters, instead of our cascaded decoder. Table 2 demonstrates that our complete model consistently beats these ablated models on most metrics. In particular, the absence of the proposed prior (*i.e.*, the *-pr* model) degrades

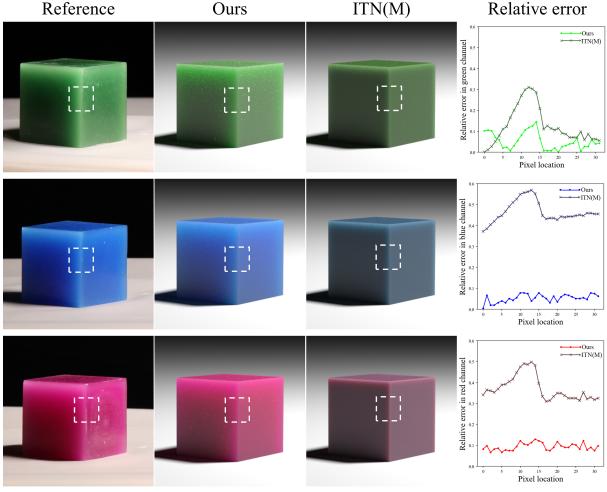


Figure 7. Quantitative experiment on real-world examples. We compare images rendered according to predicted parameters of our method and ITN(M) on waxy cubes. The fourth column shows their relative errors to real captures within dotted boxes.

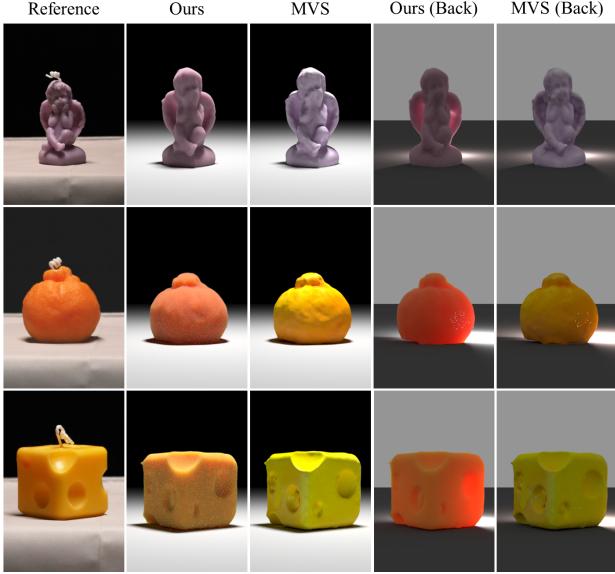


Figure 8. Visual comparison with a conventional 3D reconstruction method based on multi-view stereo (MVS) which can only reproduce purely diffuse appearance.

the performance, leading to less accurate prediction.

5.4. Limitations

Our method has several limitations. First, the regression algorithm for the pseudo-albedo/bias maps requires at least three images. Otherwise, it will not work properly. Second, for near-transparent objects that will mix its color with the scene background, the proposed inter-pixel translucency prior is no longer valid, easily resulting in inaccurate pre-

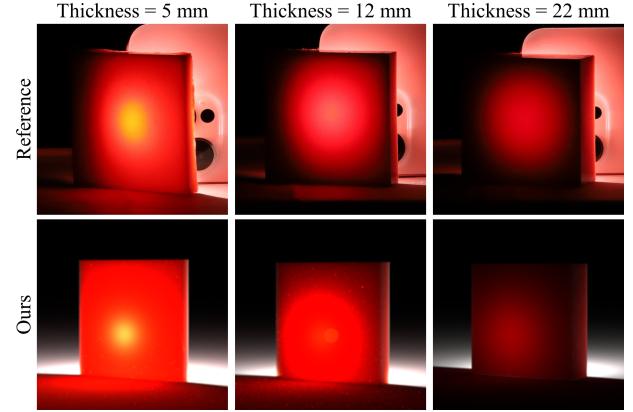


Figure 9. Test on red waxy blocks with different thicknesses.

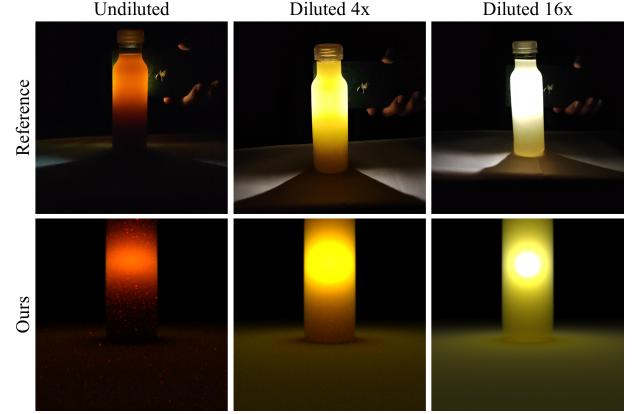


Figure 10. Test on orange juice at various concentrations.

Table 2. Quantitative comparison with different variants of ITN

		TrShape				TrLight			
		Ours	mean	-pr	single	Ours	mean	-pr	single
MSE↓	σ'_t	0.026	0.040	0.038	0.033	0.015	0.011	0.033	0.018
	Λ'	0.002	0.003	0.003	0.003	0.010	0.019	0.034	0.026
	g	0.033	0.075	0.098	0.040	0.027	0.057	0.092	0.076
PSNR↑	H	37.41	35.53	35.53	36.35	36.34	34.63	31.74	32.92
	D	35.96	33.30	33.92	34.27	33.45	30.43	28.61	28.22
	R	37.55	35.33	33.80	36.21	35.74	33.75	31.37	31.67
SSIM↑	H	0.993	0.991	0.990	0.992	0.988	0.985	0.976	0.982
	D	0.994	0.991	0.992	0.992	0.986	0.980	0.967	0.975
	R	0.996	0.994	0.994	0.995	0.990	0.986	0.978	0.984

diction. Third, our method may fail for input images with large portions of over-saturated pixels due to specular materials. More discussions on limitations are provided in the supplemental document.

6. Conclusion

In this paper, we have proposed a practical method for translucent material acquisition, which relies only on a low-cost setup and ordinary images. The success of this method owes to an inter-pixel translucency prior, which demonstrates that pixels collected from multiple images of a translucent material form translucent curves in the RGB space, and a specially-designed neural network that takes a pseudo-albedo map and a bias map stemming from this prior into consideration. We have shown from experiments on both synthetic data and real-world examples that the proposed method achieves state-of-the-art performance and supports a wide range of translucent materials in our daily life.

References

- [1] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. Two-shot svbrdf capture for stationary materials. *ACM Trans. Graph.*, 34(4), 2015. [3](#), [5](#)
- [2] Louis-Philippe Asselin, Denis Laurendeau, and Jean-François Lalonde. Deep SVBRDF Estimation on Real Materials. In *International Conference on 3D Vision (3DV)*, 2020. [1](#), [3](#)
- [3] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik P.A. Lensch, and Jan Kautz. Two-shot spatially-varying brdf and shape estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. [1](#), [3](#)
- [4] Subrahmanyam. Chandrasekhar. *Radiative transfer*. Dover, 1960. [2](#)
- [5] Chengqian Che, Fujun Luan, Shuang Zhao, Kavita Bala, and Ioannis Gkioulekas. Towards learning-based inverse subsurface scattering. In *2020 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12, 2020. [2](#), [3](#), [5](#), [6](#), [7](#)
- [6] Per H. Christensen. An approximate reflectance profile for efficient subsurface scattering. In *ACM SIGGRAPH 2015 Talks*, New York, NY, USA, 2015. Association for Computing Machinery. [1](#)
- [7] Kristin J. Dana, Bram van Ginneken, Shree K. Nayar, and Jan J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Trans. Graph.*, 18(1):1–34, 1999. [3](#)
- [8] Eugene D’Eon and Geoffrey Irving. A quantized-diffusion model for rendering translucent materials. *ACM Trans. Graph.*, 30(4), 2011. [2](#)
- [9] Yue Dong. Deep appearance modeling: A survey. *Visual Informatics*, 3(2):59 – 68, 2019. [2](#)
- [10] Craig Donner and Henrik Wann Jensen. Light diffusion in multi-layered translucent materials. *ACM Trans. Graph.*, 24 (3):1032–1039, 2005. [2](#)
- [11] Julie Dorsey, Holly Rushmeier, and Franois Sillion. *Digital Modeling of Material Appearance*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2007. [2](#)
- [12] Jeppe Revall Frisvad, Niels Jørgen Christensen, and Henrik Wann Jensen. Computing the scattering properties of participating media using lorenz-mie theory. *ACM Trans. Graph.*, 26(3):60–es, 2007. [2](#)
- [13] J. R. Frisvad, S. A. Jensen, J. S. Madsen, A. Correia, L. Yang, S. K. S. GregerSEN, Y. Meuret, and P.-E. Hansen. Survey of models for acquiring the optical properties of translucent materials. *Computer Graphics Forum*, 39(2):729–755, 2020. [2](#)
- [14] Duan Gao, Xiao Li, Yue Dong, Pieter Peers, Kun Xu, and Xin Tong. Deep inverse rendering for high-resolution svbrdf estimation from an arbitrary number of images. *ACM Trans. Graph.*, 38(4), 2019. [1](#), [3](#)
- [15] Abhijeet Ghosh, Tim Hawkins, Pieter Peers, Sune Frederiksen, and Paul Debevec. Practical modeling and acquisition of layered facial reflectance. *ACM Trans. Graph.*, 27(5), 2008. [3](#)
- [16] Davit Gigilashvili, Jean-Baptiste Thomas, Jon Yngve Harderberg, and Marius Pedersen. Translucency perception: A review. *Journal of Vision*, 21(8):4–4, 2021. [1](#)
- [17] Ioannis Gkioulekas, Bei Xiao, Shuang Zhao, Edward H. Adelson, Todd Zickler, and Kavita Bala. Understanding the role of phase function in translucent appearance. *ACM Trans. Graph.*, 32(5), 2013. [2](#), [3](#)
- [18] Ioannis Gkioulekas, Shuang Zhao, Kavita Bala, Todd Zickler, and Anat Levin. Inverse volume rendering with material dictionaries. *international conference on computer graphics and interactive techniques*, 2013. [2](#), [5](#)
- [19] Ioannis Gkioulekas, Bruce Walter, Edward H. Adelson, Kavita Bala, and Todd Zickler. On the appearance of translucent edges. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5528–5536, 2015. [2](#), [3](#), [5](#)
- [20] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, 2010. [3](#)
- [21] Jinwei Gu, Shree K. Nayar, Eitan Grinspun, Peter N. Belhumeur, and Ravi Ramamoorthi. Compressive structured light for recovering inhomogeneous participating media. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):1–1, 2013. [1](#), [2](#)
- [22] Giuseppe Claudio Guarnera, Abhijeet Ghosh, Ian Hall, Mashuda Glencross, and Dar’ya Guarnera. Material capture and representation with applications in virtual reality. In *ACM SIGGRAPH 2017 Courses*, New York, NY, USA, 2017. Association for Computing Machinery. [1](#)
- [23] Jie Guo, Shuichang Lai, Chengzhi Tao, Yuelong Cai, Lei Wang, Yanwen Guo, and Ling-Qi Yan. Highlight-aware two-stream network for single-image svbrdf acquisition. *ACM Trans. Graph.*, 40(4), 2021. [1](#), [3](#), [5](#)
- [24] Jie Guo, Shuichang Lai, Qinghao Tu, Chengzhi Tao, Changqing Zou, and Yanwen Guo. Ultra-high resolution svbrdf recovery from a single image. *ACM Trans. Graph.*, 42(3), 2023.
- [25] Yu Guo, Cameron Smith, Miloš Hašan, Kalyan Sunkavalli, and Shuang Zhao. Materialgan: Reflectance capture using a generative svbrdf model. *ACM Trans. Graph.*, 39(6), 2020. [1](#), [3](#)

- [26] Pat Hanrahan and Wolfgang Krueger. Reflection from layered surfaces due to subsurface scattering. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, page 165–174, New York, NY, USA, 1993. Association for Computing Machinery. 3, 5
- [27] Tim Hawkins, Per Einarsson, and Paul Debevec. Acquisition of time-varying participating media. *ACM Trans. Graph.*, 24(3):812–815, 2005. 1
- [28] L. G. Henyey and J. L. Greenstein. Diffuse radiation in the Galaxy. *The Astrophysical Journal*, 93:70–83, 1941. 2
- [29] Michael Holroyd, Jason Lawrence, and Todd Zickler. A coaxial optical scanner for synchronous acquisition of 3d geometry and surface reflectance. In *ACM SIGGRAPH 2010 Papers*, New York, NY, USA, 2010. Association for Computing Machinery. 3
- [30] Zhuo Hui, Kalyan Sunkavalli, Joon-Young Lee, Sunil Hadap, Jian Wang, and Aswin C. Sankaranarayanan. Reflectance capture using univariate sampling of brdfs. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 1
- [31] Tomáš Iser, Tobias Rittig, Emilie Nogué, Thomas Klaus Nindel, and Alexander Wilkie. Affordable spectral measurements of translucent materials. *ACM Trans. Graph.*, 41(6), 2022. 1, 2
- [32] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. Mitsuba 3 renderer, 2022. <https://mitsuba-renderer.org>. 6, 7
- [33] Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. Dr.jit: A just-in-time compiler for differentiable rendering. *Transactions on Graphics (Proceedings of SIGGRAPH)*, 41(4), 2022. 2, 6
- [34] Henrik Wann Jensen and Juan Buhler. A rapid hierarchical rendering technique for translucent materials. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, page 576–581, New York, NY, USA, 2002. Association for Computing Machinery. 1, 2
- [35] Henrik Wann Jensen, Stephen R. Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 511–518, New York, NY, USA, 2001. Association for Computing Machinery. 2
- [36] C. Johnson and D. Gabriel. *Laser light scattering*. Dover, 1994. 2
- [37] RAYLEIGH J. W. S. L. On the scattering of light by small particles. *Philosophical Magazine*, 64:447–454, 1871. 2
- [38] 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. In *ACM SIGGRAPH 2006 Papers*, pages 735–745, New York, NY, USA, 2006. Association for Computing Machinery. 3
- [39] Xiao Li, Yue Dong, Pieter Peers, and Xin Tong. Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *ACM Trans. Graph.*, 36(4), 2017. 1, 3
- [40] Zhengqin Li, Kalyan Sunkavalli, and Manmohan Chandraker. Materials for masses: Svbrdf acquisition with a single mobile phone image. In *Computer Vision – ECCV 2018*, pages 74–90, Cham, 2018. Springer International Publishing. 1, 3
- [41] Hibernon Lopes Filho, Lúcio E.G. Maia, Marcus Vinicius A. Araújo, and Antônio Carlos O. Ruellas. Influence of optical properties of esthetic brackets (color, translucence, and fluorescence) on visual perception. *American Journal of Orthodontics and Dentofacial Orthopedics*, 141(4):460–467, 2012. 1
- [42] Ludvig Lorenz. Lysbevægelser i og uden for en af plane lysbølger belyst kugle. *Det kongelig danske Videnskabernes Selskabs Skrifter*, pages 2–62, 1890. 2
- [43] Gustav Mie. Beiträge zur optik trüber medien, speziell kolloidalen metallösungen. *Annalen der Physik*, 330(3):377–445, 1908. 2
- [44] Yasuhiro Mukaigawa, Yasushi Yagi, and Ramesh Raskar. Analysis of light transport in scattering media. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 153–160, 2010. 2
- [45] Giljoo Nam, Joo Ho Lee, Diego Gutierrez, and Min H. Kim. Practical svbrdf acquisition of 3d objects with unstructured flash photography. *ACM Trans. Graph.*, 37(6), 2018. 1, 3
- [46] Srinivasa G. Narasimhan, Mohit Gupta, Craig Donner, Ravi Ramamoorthi, Shree K. Nayar, and Henrik Wann Jensen. Acquiring scattering properties of participating media by dilution. *ACM Trans. Graph.*, 25(3):1003–1012, 2006. 1, 2, 5
- [47] Jan Novák, Iliyan Georgiev, Johannes Hanika, and Wojciech Jarosz. Monte carlo methods for volumetric light transport simulation. *Computer Graphics Forum*, 37(2):551–576, 2018. 2
- [48] Geoffrey Oxholm and Ko Nishino. Shape and reflectance estimation in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):376–389, 2016. 1
- [49] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 6
- [50] Matt Pharr and Wenzel Jakob. *Physically Based Rendering, fourth edition: From Theory to Implementation*. MIT Press, London, England, 2023. 5
- [51] John William Pickering, Christian J. M. Moes, Henricus J. C. M. Sterenborg, Scott A. Prahl, and M.J.C. van Gemert. Two integrating spheres with an intervening scattering sample. *Journal of The Optical Society of America A-optics Image Science and Vision*, 9:621–631, 1992. 2
- [52] John W. Pickering, Scott A. Prahl, Niek van Wieringen, Johan F. Beek, Henricus J. C. M. Sterenborg, and Martin J. C. van Gemert. Double-integrating-sphere system for measuring the optical properties of tissue. *Appl. Opt.*, 32(4):399–410, 1993.

- [53] Scott A. Prahl, Martin J. C. van Gemert, and Ashley J. Welch. Determining the optical properties of turbid media by using the adding–doubling method. *Appl. Opt.*, 32(4): 559–568, 1993. [2](#)
- [54] Ashutosh Saxena, Justin Driemeyer, Justin Kearns, and Andrew Ng. Robotic grasping of novel objects. In *Advances in Neural Information Processing Systems*. MIT Press, 2006. [1](#)
- [55] Sven Schröder, Alexander von Finck, and Angela Duparré. Standardization of light scattering measurements. *Advanced Optical Technologies*, 4:361 – 375, 2015. [2](#)
- [56] 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. [3](#)
- [57] Eric Veach. *Robust Monte Carlo Methods for Light Transport Simulation*. PhD thesis, Stanford, CA, USA, 1997. [3](#)
- [58] Delio Vicini, Sébastien Speierer, and Wenzel Jakob. Path replay backpropagation: Differentiating light paths using constant memory and linear time. *ACM Trans. Graph.*, 40(4), 2021. [2](#)
- [59] Jiaping Wang, Shuang Zhao, Xin Tong, Stephen Lin, Zhouchen Lin, Yue Dong, Baining Guo, and Heung-Yeung Shum. Modeling and rendering of heterogeneous translucent materials using the diffusion equation. *ACM Trans. Graph.*, 27(1), 2008. [2](#)
- [60] Wenzel Jakob. Mitsuba renderer. [5](#)
- [61] Tim Weyrich, Jason Lawrence, Hendrik Lensch, Szymon Rusinkiewicz, and Todd Zickler. Principles of appearance acquisition and representation. In *ACM SIGGRAPH 2008 Classes*, New York, NY, USA, 2008. Association for Computing Machinery. [2](#)
- [62] D. R. Wyman, M. S. Patterson, and B. C. Wilson. Similarity relations for anisotropic scattering in monte carlo simulations of deeply penetrating neutral particles. *J. Comput. Phys.*, 81(1):137–150, 1989. [3, 6](#)
- [63] Jiankai Xing, Fujun Luan, Ling-Qi Yan, Xuejun Hu, Houde Qian, and Kun Xu. Differentiable rendering using rgbxy derivatives and optimal transport. *ACM Trans. Graph.*, 41(6), 2022. [2](#)
- [64] Tizian Zeltner, Sébastien Speierer, Iliyan Georgiev, and Wenzel Jakob. Monte carlo estimators for differential light transport. *ACM Trans. Graph.*, 40(4), 2021. [2](#)
- [65] Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. A differential theory of radiative transfer. *ACM Trans. Graph.*, 38(6), 2019.
- [66] Cheng Zhang, Zihan Yu, and Shuang Zhao. Path-space differentiable rendering of participating media. *ACM Trans. Graph.*, 40(4), 2021. [2](#)
- [67] Shuang Zhao, Ravi Ramamoorthi, and Kavita Bala. High-order similarity relations in radiative transfer. *ACM Trans. Graph.*, 33(4), 2014. [3](#)