Practical Acquisition of Shape and Plausible Appearance of Reflective and Translucent Objects

A. Lin[1,2] (ID)   Y. Lin [2] (ID)   A. Ghosh[1,2] (ID)

[1]Imperial College London, UK
[2]Lumirithmic Ltd.

(a) *Diffuse albedo*   **(b)** *Specular albedo*   **(c)** *Specular normal*   **(d)** *Transmit vector*   **(e)** *Transmit albedo*   **(f)** *3D geometry*   **(g)** *3D rendering*

Figure 1: *Acquired reflectance and transmission maps and surface geometry (a − f) of a liquor bottle, enabling realistic 3D rendering (g) under the Pisa lighting environment.*

Abstract

We present a practical method for acquisition of shape and plausible appearance of reflective and translucent objects for realistic rendering and relighting applications. Such objects are extremely challenging to scan with existing capture setups, and have previously required complex lightstage hardware emitting continuous illumination. We instead employ a practical capture setup consisting of a set of desktop LCD screens to illuminate such objects with piece-wise continuous illumination for acquisition. We employ phase-shifted sinusoidal illumination for novel estimation of high quality photometric normals and transmission vector along with diffuse-specular separated reflectance/transmission maps for realistic relighting. We further employ neural in-painting to fill gaps in our measurements caused by gaps in screen illumination, and a novel NeuS-based neural rendering that combines these shape and reflectance maps acquired from multiple viewpoints for high-quality 3D surface geometry reconstruction along with plausible realistic rendering of complex light transport in such objects.

CCS Concepts

• Computing methodologies → Computational photography; Reflectance modeling; 3D imaging;

1. Introduction

Appearance and shape capture of everyday objects has received a lot of attention in computer graphics and vision for reconstruction and realistic rendering applications. Most efforts here have been dedicated for acquisition of opaque objects, with different acquisition setups designed for objects exhibiting diffuse or glossy reflectance, vs. those exhibiting sharp specular reflectance. Some efforts have also been dedicated to acquisition of transparent glass objects which are also quite common and visually interesting. However, there is very little prior work on acquisition of a special class

of objects that are both reflective and translucent, e.g., a transparent/translucent bottle/container containing a translucent liquid along with opaque reflective sections such as label or bottle cap (see Figure. 1). Such objects are very hard to capture due to the requirements of acquiring both surface shape and reflection as well as transmission/refraction of light through parts of the object. Often such objects exhibit dominantly specular surface reflection (glass or plastic material) with very little diffuse surface texture, making even the acquisition of surface shape challenging.

In this work, we present a practical method and capture setup to

acquire the surface shape and plausible appearance of such reflective and translucent objects for realistic rendering applications. Our capture setup consists of a set of five desktop LCD monitors that are arranged to sequentially illuminate the object with frontal and back illumination respectively in order to separately acquire surface shape and reflectance, as well as transmission/refraction of light through the object. The setup also exploits multiview imaging using a set of cameras placed between the the LCD monitors to acquire view-dependent reflectance and transmission of the object. Direct capture with our setup has some limitations due to gaps in illumination which we overcome using neural in-painting of the acquired maps. Finally, the estimated surface shape and view-dependent appearance maps are employed in a novel NeuS-based neural rendering procedure to create plausibly realistic renderings of the object from novel viewpoints and under novel lighting environments. Note that our approach strives for *plausible* rather than accurate modeling of object appearance. This is because we employ a relatively simple rendering model, driven by our estimated view-dependent appearance maps, which results in a first order approximation of the actual light transport through the object volume. Despite this plausible appearance approximation, our method achieves very compelling rendering results for a visually interesting class of everyday objects that have been extremely hard to digitize until now.

Our main contribution can be summarized as:

- A practical setup comprising commodity screens for scanning target reflective and translucent objects without dedicated illumination hardware.
- A novel photometric method to estimate photometric normals and transmission vectors using phase-shifted sinusoidal illumination.
- A weighted loss function that performs neural inpainting for piece-wise continuous measurements and helps with high quality 3D geometry reconstruction of our target objects.
- Plausibly realistic rendering results for our target class of objects from *novel viewpoints* under *novel lighting* without explicit modeling of internal geometry and medium.

2. Related Work

Reflectance capture for opaque objects have been well researched in computer graphics. Various capture setups with controlled illumination have been proposed for measuring the shape and reflectance of an opaque object [MHP*07; GFT*11; KXH*19; KGX*23]. Most lightstage setups utilize discrete lighting to approximate the illumination. While this works well for rough, opaque surfaces, these setups cannot properly illuminate reflective and translucent objects. Because of the complexity involved, continuous illumination is difficult to produce and setups that provide continuous illumination are usually limited in size and focus on purely reflective objects [LPGD09; TFG*13; KXH*19]. More recently, the monitor-based illumination setup proposed by [LLK*22] provides a convenient way to illuminate a subject with piece-wise continuous illumination, a property that we exploit in this work.

Capturing the appearance of reflective and translucent objects is challenging because of the complex light transport involved in these objects. A BRDF model is insufficient to model the appearance of these objects, and a scattering model would require knowledge of the inner geometry as well as the optical properties of the participating media. While several methods have been proposed to measure the scattering properties of participating media [NGD*06; GZB*13; IRN*22; YX16], simultaneously acquiring the optical properties as well as the geometry of common reflective and translucent object remains challenging. Relighting these objects have been previously done in the context of environment matting methods that, similar to our setup, employ a screen illumination for back-lighting [ZWCS99; PD03; CZH*00]. Image-based relighting of such objects under environmental illumination has been done using a very specialised dual lighstage setup [HED05], as well as by using neural networks for modelling the scene's light transport [RDL*15]. However, these methods are restricted to single-view, and do not provide any 3D geometry information.

A number of other works have focused on 3D geometry reconstruction without appearance modelling. While geometry reconstruction can be done under passive uncontrolled illumination [GHLB15; LYC20], incorporating active illumination methods enables higher-quality capture of meso-structures [FCM*08] and overall geometry [KRG17; LWL*20; MLS*14; WZQ*18]. These active illumination methods involve dedicated illumination and capture setup design that either measures light reflection or transmission/refraction. Most related to our work, [MPZ*02] combined environment matting methods with sparse frontal lighting, and utilized multiview capture to reconstruct the 3D geometry and reflectance for reflective and refractive objects. Their method, however, requires a complicated lighting setup and only provides a rough geometry, as well as sparse specular reflectance measurements.

More recently, with the advances of inverse volume rendering, multiple methods have demonstrated high quality novel-view synthesis using just a few input images. Several methods has been proposed to either model specular reflection [VHM*21], refractive light transport [BMR*22], or reconstructing the 4D light field [SESM22]. These methods demonstrated that inverse volume rendering is capable of state-of-the-art novel-view image synthesis. However, relighting reflective and translucent objects remains a challenging problem. Most current works that attempt to reconstruct translucent objects with inverse rendering restrict their model to subsurface scattering [CLZ*20; DLW*22]. Our method differs from these work in that we combine active illumination and inverse volume rendering to achieve high quality 3D rendering with plausible relighting for reflective and translucent objects.

3. Method

3.1. Overview

Our capture setup utilises five desktop LCD monitors for illuminating the object, of which four are used for frontal lighting to acquire outer shape and reflectance. The fifth monitor is used for backlighting to obtain transmission maps. Figure 2 illustrates the capture setup. Six DSLR cameras are evenly placed between the center vertical gap of the frontal monitors (see Figure 6). Placing the cameras in the center gap allows us to obtain good transmission measurements of the object with the back lighting monitor. This placement
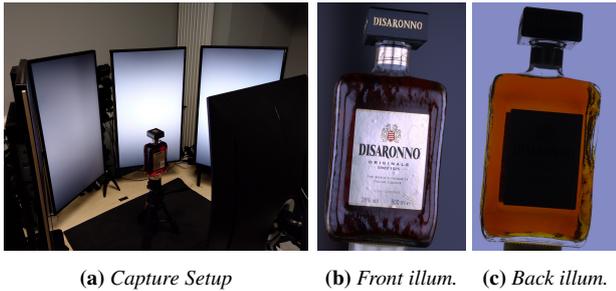
(a) *Capture Setup*   **(b)** *Front illum.*   **(c)** *Back illum.*

Figure 2: *Our capture setup (a) provides piece-wise continuous illumination, allowing practical measurements of mirror-like specular properties of a reflective and translucent object – a liquor bottle. (b), (c) show the object lit with frontal and back illumination separately.*
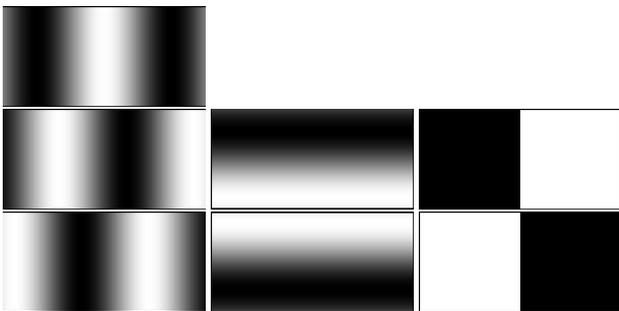

Figure 3: *We use five sinusoidal patterns plus two binary patterns to estimate the diffuse albedo, specular albedo, and specular normal (or the transmission vector for back illumination). This figure shows the patterns we use to illuminate either the front hemisphere or the back hemisphere. Each hemisphere is illuminated sequentially to separately observe the reflective and transmissive properties.*

also allows us to observe the reflection and transmission separately with minimal crosstalk by illuminating the object with only frontal lighting or backlighting. Extending the diffuse-specular separation method of [LPGD09], we illuminate the object being scanned using five sinusoidal and two binary patterns as shown in Figure 3, and further process the captured images to obtain view-dependent reflectance maps. We repeat this process twice, once with frontal illumination only to capture the reflective components and once with back lighting to capture the transmissive components, obtaining diffuse albedo, specular albedo, and specular normal maps for the reflective component, and transmission albedo and transmission vector for the back lighting. The object being scanned is rotated five times about its vertical axis and the above scanning process is repeated per orientation.

The view-dependent appearance maps acquired from multiple viewpoint are thereafter employed to jointly train the recently proposed neural signed distance function (NeuS) representation [WLL*21]. This allows us to obtain a smooth neural interpolation of our acquired view-dependent appearance maps for a novel

viewpoint and perform view-dependent realistic relighting of the scanned object. Inevitably, our acquired specular reflection maps exhibit discontinuity due to gaps in the illumination, limiting the quality of the rendering results. To tackle this problem, we propose a novel loss function for training the NeuS with imperfect reflection normals, and a neural solution to predict an improved specular reflection albedo given the issues with the measured data. Figure 4 illustrates the complete 3D neural reflectance learning pipeline.

3.2. Background

We employ sinusoidal illumination using our capture setup for diffuse-specular separation as proposed by Lamond et al. [LPGD09]. Both reflection and transmission diffuse and specular maps of an object are obtained using the same method, with the only difference being whether the object is front lit or back lit.

Given a sinusoidal illumination pattern:

$$I(\vec{\omega}) = \sin(f\phi + \psi) + 1, \tag{1}$$

where the intensity of the illumination depends on the direction of the irradiance: $\vec{\omega} = (\phi, \theta)$, the reflected radiance for a given surface point is a function of the frequency f and the illumination phase shift ψ:

$$E(f, \psi) = \rho + \cos(\psi)S(f) + \sin(\psi)C(f), \tag{2}$$

ρ, C and S are spatially varying constants that determine the reflected radiance function for each pixel. Since the diffuse response to a sinusoidal illumination decreases to approximately zero when the frequency is high [RH01], we have:

$$\begin{aligned} \rho_s &\approx \sqrt{S^2 + C^2}, \\ \rho_d &\approx \rho - \rho_s, \end{aligned} \tag{3}$$

where ρ_d and ρ_s are the diffuse and specular albedo respectively. Under a fixed frequency f, the reflected radiance for a sinusoidal illumination pattern is a function of the phase shift with three parameters (ρ, C, S). Thus, we measure the reflected radiance under three sinusoidal illumination patterns with different phases $\psi_i, i \in \{0, 1, 2\}$ to estimate the parameters (ρ, C, S) by solving a linear system. While Lamond et al suggested to use $f > 10$ for their method, using such high frequencies would not provide enough information for estimating the specular normal and transmission vector. Instead, we noticed that using $f = 3$ is sufficient for diffuse and specular separation [RH01], and allows us to estimate the specular normal and transmission vector with only two additional binary patterns.

3.3. Estimating Specular Normal and Transmission Vector

While Lamond et al. [LPGD09] used the parameters (ρ, C, S) to separate the diffuse and specular response, the relation between C and S is not utilized. Here, we exploit the fact that the ratio between C and S encodes the phase of the reflected radiance function, and use this to estimate the specular normal and transmission vectors. We can calculate the illumination phase shift that corresponds to the maximum reflected radiance ψ_{max} by rewriting equation 2 as
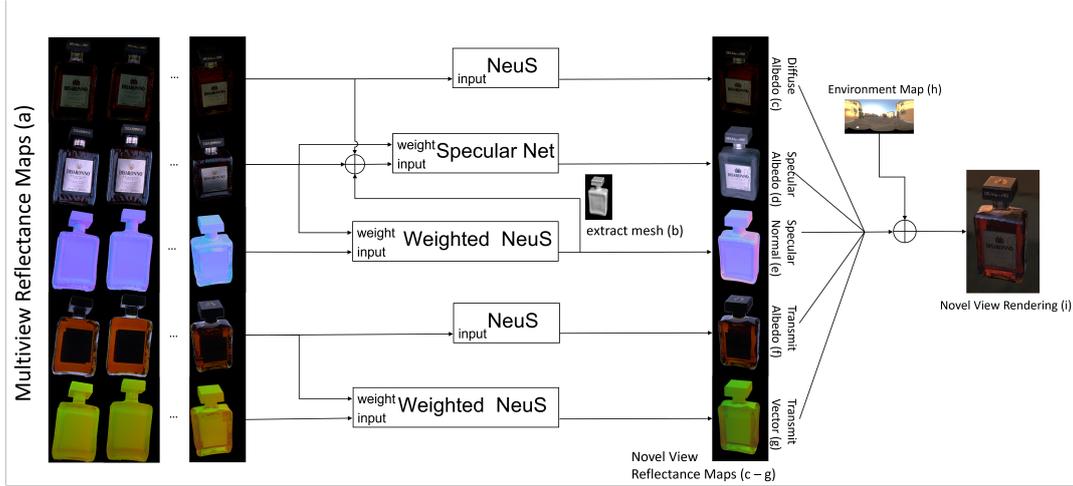
Figure 4: *Overview of the 3D Appearance Pipeline. Once the reflectance maps are acquired (a), we train a NeuS for each reflectance map to acquire a view dependent 3D representation. To fix the discontinuity in the measured specular maps, we propose a weighted loss function while training the specular normal and transmission vector. The geometry acquired from the specular normal is then used as input (with the diffuse map) to train a specular network to predict view dependent specular albedo. Implmentation details are described in section 3.5*

follows:

$$E(f,\psi) = \rho(f) + A(f)\cos(\psi - \psi_{max}(f)),$$
$$A(f) = \sqrt{C(f)^2 + S(f)^2}, \qquad (4)$$
$$\psi_{max}(f) = \arctan(C/S).$$

The observed radiance reaches its maximum when the phase shift of the illumination is $\psi_{max}(f)$. Now consider a reflective, mirror-like surface with a reflection angle ϕ_r, the specular response depends mostly on the illumination from the reflection angle ϕ_r. Thus the maximum specular reflection is observed when the illumination at direction ϕ_r is maximal, referring back to equation 1, we have

$$1 + \sin(f\phi_r + \psi_{max}) = \max_{\phi}(I(\vec{\omega})) = 2, \qquad (5)$$

then the reflection angle can be represented as:

$$\phi_r = \frac{\pi/2 - \psi_{max}}{f} + \frac{2n\pi}{f}, \quad n \in \{0,...,f-1\}. \qquad (6)$$

For a set of sinusoidal patterns with frequency f, we obtain f possible candidates for ϕ_r. Since we chose $f = 3$, and only illuminates half of the sphere during capture, there are only two possible candidates for ϕ_r. ϕ_r can be uniquely identified by acquiring two additional measurements under a horizontal binary pattern and its complement as shown in Figure 3. In the pure reflection case, ϕ_r corresponds to the reflection vector, whereas when only transmission is observed, it is equivalent to the transmission vector.

Now that we know how to estimate the maximum specular response angle along the direction of phase-shift, we can repeat the process along the longitudinal and latitudinal directions respectively, which gives us a pair of angles (θ_r, ϕ_r) that represents the maximum specular response vector in the polar coordinate. Since

we use the same frequency for sinusoidal patterns along the longitudinal and latitudinal directions, we can assume they share the same diffuse component. The resulting linear system will have five unknowns: $\{\rho_d, S_\theta, C_\theta, S_\phi, C_\phi\}$, which requires five observations under different sinusoidal patterns (e.g., three phase-shifts along θ_r, and and two phase-shift along ϕ_r) to solve for the unknowns.

For a reflective and translucent object, both reflection and transmission are usually present under a full sphere of illumination. To separately measure reflection and transmission, we observe the reflective behavior by illuminating only the front hemisphere with the sinusoidal and binary patterns, then repeat the same procedure with only the back hemisphere illuminated. With the back illumination, we can obtain the transmission vector as well as the transmission equivalent of diffuse and specular albedo. In our experiments the observed transmission diffuse albedo is always almost zero, thus in the following sections we only refer to the transmission specular albedo as the transmission albedo.

Figure 5 shows the captured reflectance maps for both reflection and transmission. While these properties do not fully capture the scattering properties, they provide an efficient way to render the translucent behavior of an object without having knowledge of the participating medium and the entire geometry.

3.4. Resolving Issues with Limited Coverage

While the LCD screens provide a convenient way for generating continuous sinusoidal illumination, the gaps between screens and the limited latitudinal coverage introduces specular gaps in the measurements, as seen in Figure 5

To resolve this issue, we use multiview capture to measure the specular reflections from different viewpoints. Specifically, to resolve the limited latitudonal coverage, we place six cameras in a
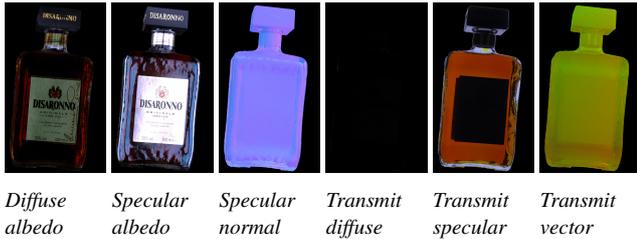
| *Diffuse albedo* | *Specular albedo* | *Specular normal* | *Transmit diffuse* | *Transmit specular* | *Transmit vector* |

Figure 5: *Captured Reflectance Maps. We observed that the transmit diffuse albedo is always almost zero, even with objects that are highly scattering. The transmit diffuse albedo will not be shown in the following discussion.*
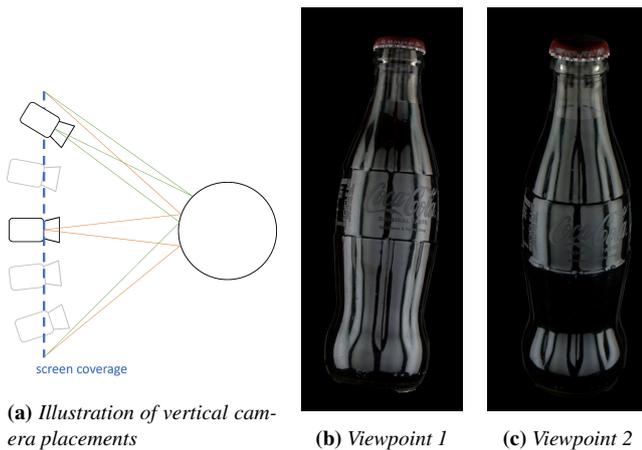

(a) *Illustration of vertical camera placements*

(b) *Viewpoint 1*

(c) *Viewpoint 2*

Figure 6: *Multiple cameras placed vertically provides wider range for valid specular measurements*

vertical line, extending the measurable specular reflections (Figure 6). To resolve the issues due to the gaps between screens, we slightly rotate the object horizontally 5 times during each capture. As a result, for each object we would measure the reflectance maps from 30 different viewpoints.

3.5. Novel View Synthesis

After acquiring the view-dependent reflection and transmission maps, synthesising novel view maps can be achieved by training multiple NeuS [WLL*21] independently for each reflectance maps across different views. However, as we have gaps between the frontal monitors, the specular response from the object exhibits discontinuity in reflection, resulting in noticeable imperfections in the acquired specular albedo and normals. To tackle this problem, we make two assumptions:

1. Low specular albedo measurement is caused by the gaps and limited coverage of the illumination.
2. For each surface point, there are one or more good specular observations, i.e. the surface reflects the screen illumination to at least one of the camera views.

With such assumptions, we can directly relate the acquired specular normal quality to its acquired specular albedo, and there are
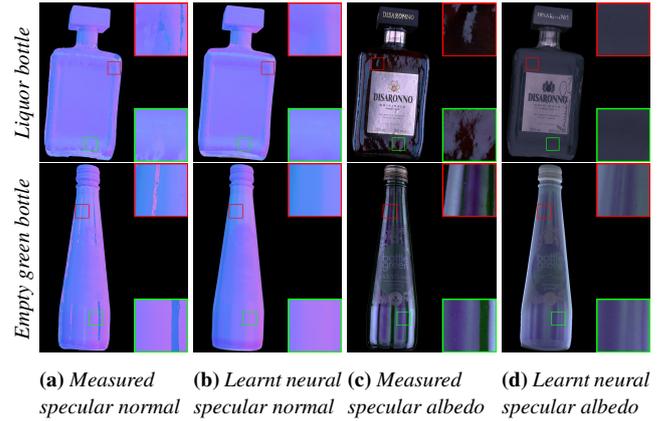

(a) *Measured specular normal* **(b)** *Learnt neural specular normal* **(c)** *Measured specular albedo* **(d)** *Learnt neural specular albedo*

Figure 7: *Our neural volumetric NeuS-based specular normals **(b)** learns to fill in the gaps in measurement **(a)** by taking information from other viewpoints. The reconstructed mesh is then used with the diffuse albedo to guide an MLP that predicts an in-painted specular albedo map **(d)** from the original measurements **(c)**.*

one or more reasonably accurate normal acquired for a surface point across the 30 views we acquire of the object's frontal surface. Therefore, we employ NeuS to perform neural in-painting of the gaps in the specular measurement. For this purpose, we first train a NeuS for the specular normal minimizing the weighted objective function:

$$\mathcal{L}_{normal} = \sum_{i=0}^{N} (\hat{\vec{n}} - \vec{n})^2 \cdot \rho_s, \tag{7}$$

where $\hat{\vec{n}}$ is the estimated specular normal from NeuS, $\vec{n}$ is the acquired normal, and ρ_s is the acquired specular albedo.

Figure 7(a, b) demonstrates comparisons between the acquired specular normal and the hole-filled specular normal by training a NeuS [WLL*21] with our weighted loss function. Our refined specular normal maps are globally consistent and smooth, unlike the directly measured specular normal maps which have noise caused by the piece-wise continuous illumination. As a result, the 3D geometry extracted from the trained NeuS for the normal is of high quality and accuracy, and is much better to the geometry obtained from directly training NeuS with images captured under a fixed lighting (see Figure 9).

A similar process could be applied to refine the specular albedo. However, as NeuS implicitly optimises for surfaces, highly view-dependent specular albedo maps tend to introduce unpleasant surface variations to compensate for the view dependency. Instead, we borrow the mesh extracted from the NeuS trained using normal maps, and we directly predict the specular albedo using a trained MLP with the fixed mesh. We assume the specular albedo is strongly correlated to the diffuse albedo, and define the training loss as:

$$\mathcal{L}_{specular} = \sum_{i=0}^{N} (f(\rho_d, c, (\vec{v} \cdot \vec{n})) - \rho_s)^2 \cdot \rho_s, \tag{8}$$

where ρ_s is the acquired specular albedo, c is the world positional

coordinate of the surface point. To model the view-dependent Fresnel effect, we provide the function f with the dot product of the view vector $\vec{v}$ and the surface normal $\vec{n}$. We train an MLP to explicitly learn the function f, and estimate the specular albedo for a surface point c viewed at direction $\vec{v}$, given the diffuse albedo ρ_d. Here, the diffuse albedo itself is obtained from a separate NeuS trained using the acquired multi-view diffuse albedo maps for novel view synthesis. Figure 7(c, d) shows a comparison between the acquired specular albedo which exhibits gaps in the reflection and the neural prediction of the specular albedo which performs smooth in-painting of the gaps.

Our acquired transmission vector and albedo maps inherently encode complicated transmission and scattering events. As we only target relighting plausibility rather than absolute physical accuracy of light transport, we propose to train separate NeuS for both the transmission vector and transmission albedo for novel view synthesis of these acquired parameters without any additional refinement step. While the transmit vector is highly view-dependent, it varies smoothly across viewpoints and we find that a neural representation such as NeuS is capable of learning to interpolate the transmission vector for novel-view rendering.

3.6. Implementation Details

For our neural fields we follow the implementation of [WLL*21]. We additionally restrict the output to have length 1 for specular normal and transmission vector. We input the view vector without positional encoding for training the specular normal and diffuse albedo nets. For transmission albedo and transmission vector, the view vector is encoded with six frequencies following [WLL*21]'s implementation. The spatial position is encoded with nine frequencies when training the specular normal as this gives fine geometry details without introducing high-frequency artefacts. The rest of the reflectance maps are trained by encoding the spatial position with twelve frequencies. Our specular albedo network forms a four-layer MLP, and the world position is encoded with six frequencies. All reflectance maps are trained for 300k iterations, using an Adam optimizer with a learning rate of 10^{-5}.

4. Relighting with Acquired Maps

To relight an object under a given target lighting environment map, we employ a very simple normal-based rendering model. We render both the reflective and transmissive components using their respective diffuse (reflection only) and specular albedo and respective normals. We do not estimate the specular roughness for the reflective or transmissive components and hence this parameter is emperically chosen. In practice, we perform Gaussian blur on the environment map, then use the surface normal or the transmission vector to directly index the blurred envrionment map to simulate specular roughness as proposed in [KVHS00]. Since the specular roughness is usually different between the label and the reflective and translucent segments, we employ the diffuse intensity to segment the label, and apply a different level of gaussian blur on the environment for the label and reflective and translucent segments.

Note that transmissive normals acquired by our capture process are somewhat affected by the near-field nature of the back-lighting from the back monitor. However, we employ them as-is to index into environment maps during rendering of the backlighting. This is again an approximation since environment maps assume directional lighting. However, as has been shown in the work of [KRFB06], humans have a poor perception of any inaccuracies in transmission/refraction through convex transparent objects. Hence, we rely on our approximation of backlighting using the transmissive normal to still produce very compelling rendering results.

5. Results

Our method produces high quality 3D shape of the outer surface along with highly plausible appearance maps of common reflective and translucent objects for realistic rendering. Examples of these can be seen in Figure. 8. In this section we perform qualitative and quantitative validations of our 3D geometry, as well as comparison of rendering results with actual photographs. We also make comparisons between our 3D novel-view renderings and other state-of-the-art methods.

Comparison of Geometry Reconstruction

Figure 9 shows the geometry reconstruction results using our acquired specular normal compared to reconstruction using uniform lighting. To further evaluate the effectiveness of our method of using specular normal and weighted loss function, we also compare the geometry obtained from the diffuse and specular albedo obtained by Lamond et al's method [LPGD09]. All reconstructions are obtained by training a NeuS [WLL*21] and extracting a mesh from the SDF. Our specular normal provides enough view-independent textures on the reflective outer surface, yielding reconstruction result with more details and higher quality. Note that without the specular normal, textureless surfaces with highly view-dependent specular/transmission highlights cannot be correctly reconstructed, even when using the separated diffuse and specular albedo.

Relighting Based on Plausible Reflectance Maps

In Figure 10 we compare the rendered results with real photographed images. Since we do not measure the roughness of the objects, as mentioned previously, we empirically set that parameter for rendering with pre-filtered environment mapping [KVHS00]. We show the rendering results with image-space rendering using the directly acquired view-dependent appearance maps (the left most column of each illumination setting in Figure 10), as well as novel-view 3D rendering with neural NeuS-based interpolation of the aquired view-dependent appearance maps (the second to the right column of each illumination setting in Figure 10). The 3D neural appearance maps are trained with 30 views and in this case without the view used for validation in Figure 10. Thus, the rendered result represents how well the 3D interpolation of appearance maps matches the actual appearance for a novel view of such a object. Table 1 shows the quantitative results on the peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned image patch similarity (LPIPS). Although our plausible appearance maps do not provide pixel-accurate renderings, we note that the rendered results match the photograph quite well for the same viewpoint.

| Diffuse albedo | Specular albedo | Specular normal | Transmit albedo | Transmit vector | Geometry | Rendering (pisa) | Rendering (eucalyptus) |

Figure 8: *Results of acquired shape and appearance maps of reflective and translucent objects and 3D rendering in novel lighting environments.*

(a) *Liquor bottle*   **(b)** *Empty green bottle*   **(c)** *Coke bottle*   **(d)** *Rice syrup bottle*   **(e)** *Orange juice bottle*
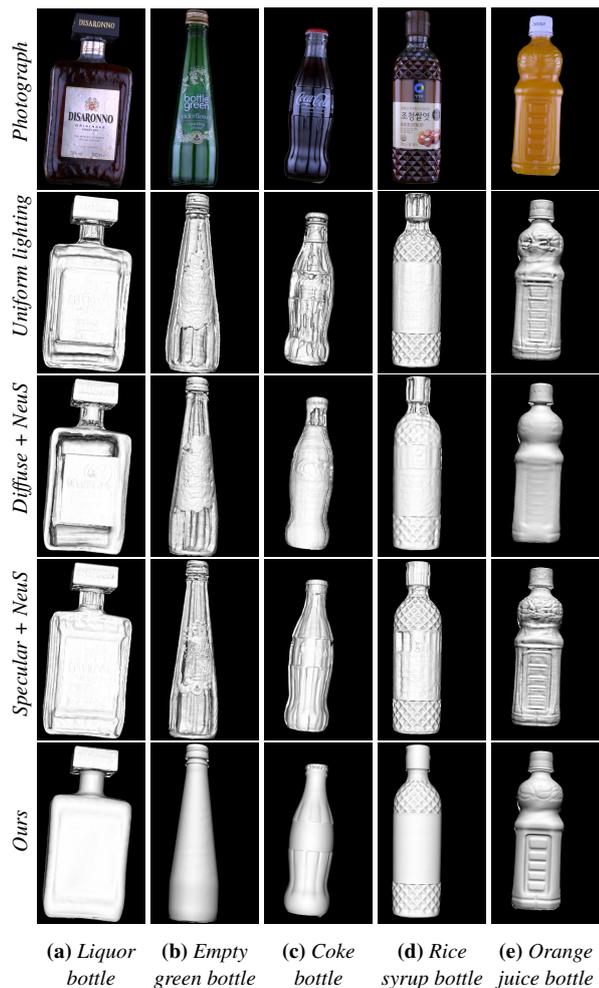
Figure 9: *Geometry reconstruction with our method is significantly better than NeuS geometry under uniform lighting (first row), with diffuse albedo (second row), and with specular albedo (third row). Only our method with the specular normal plus weighted loss provides consistent geometry reconstruction without showing artifacts from the specular/transmission highlights.*

And while the error increases for a missing novel viewpoint, the results are highly believable and produce convincing reflection and transmission properties.

Comparison with Novel View Synthesis Methods

In Figure 11 we provide comparisons of our novel view 3D renderings with the novel view synthesis results using NeRF [MST*21] and the recently proposed eikonal fiels for refractive novel-view synthesis [BMR*22]. For the comparison, these two works are trained under sinusoidal illumination emitted by the desktop monitors in our setup, then compared with our method's result rendered under the same sinusoidal environment map. Note that for objects with strong absorption photographed under low frequency background, the method of [BMR*22] cannot effectively learn the eikonal field, thus failing to reach an optimal solution for novel view synthesis. Table 2 show the quantitative comparison metrics.

And while our rendering result achieves numerically comparable performance when compared to NeRF, qualitatively it is a better match to the comparison photograph with sharper reconstruction of high-frequency details while not being limited to rendering under a fixed lighting condition.

Limitations

While our method works well on common reflective and translucent object such as beverage bottles, our reflectance estimation suffers from inter-reflections within the object. Inter-reflection is not dominant in translucent objects, but is much more dominant in purely transparent objects. For transparent objects or bottles with opaque internal objects (e.g., a straw), our method acquires noisy specular normal affected by the internal object, as well as imperfect diffuse and specular separation. These objects objects are much better handled by the method of [BMR*22] and cannot be captured with our method.

For translucent objects with observable inter-reflections, the inter-reflections may interfere with the diffuse-specular separation, causing artefacts in the diffuse albedo map (Figure 13a). While we can use a method similar to how we treat the specular albedo to learn an MLP that removes the artefacts, we found it loses high-frequency details in the resulting diffuse albedo map, as seen in Figure 13b. Our learned specular albedo retains the details since it is conditioned on the diffuse albedo as input to the MLP. We found the rendered result using the improved diffuse albedo with eliminated inter-reflections (Figure 13d) do not show visual improvement over the rendered result with the original diffuse albedo. Hence, we decide to employ the original diffuse albedo to retain the high-frequency details.

6. Conclusion

In this work, we proposed a practical method that acquires high quality 3D shape (outer surface) and plausible appearance of common reflective and translucent objects. Our measurements employ a set of desktop monitors and just seven lighting conditions for acquisition including novel usage of sinusoidal illumination patterns for estimation of specular (reflective and transmissive) normals. We further showed that a multiview inverse volume rendering approach improves any artifacts in the appearance maps caused by gaps in the monitor illumination and provides a neural representation for 3D rendering and relighting. Qualitative studies show that our method creates high-quality and believable rendering results under different lighting environments. For the target class of objects, the method also holds up against recent fully volumetric methods for novel view synthesis. However, the method has limitations and cannot handle completely transparent objects where inter-reflections and internal objects can cause errors in the estimated appearance maps and a fully volumetric solution may be preferable in this case. The relighting achieved is also plausible rather than accurate and we incur some approximation errors due to the employed rendering model as well as near-field effects of illumination.
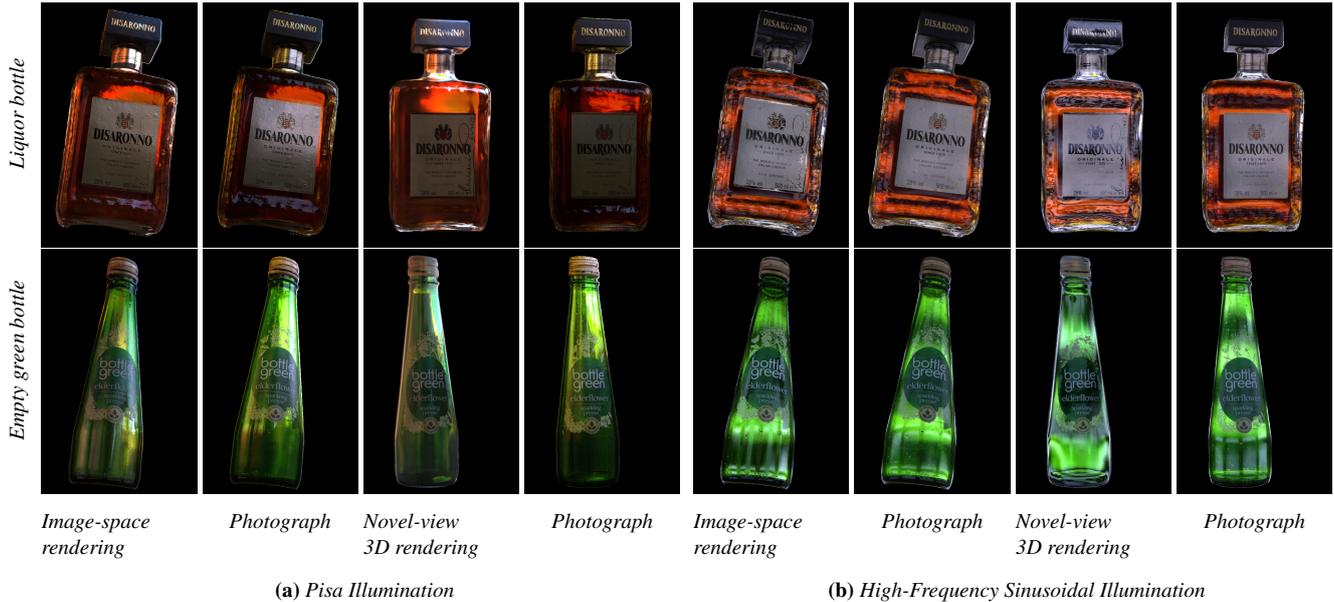
| | Image-space rendering | Photograph | Novel-view 3D rendering | Photograph | Image-space rendering | Photograph | Novel-view 3D rendering | Photograph |

(a) *Pisa Illumination*   **(b)** *High-Frequency Sinusoidal Illumination*

Figure 10: *Rendering validation with actual photographs. The image-space renderings are created by directly using the captured reflectance images (as shown in Figure 5). For image-space rendering, we validate how our captured reflectance maps allow us to create a realistic rendering for that specific view. For novel-view 3D rendering, we test how well the learned 3D reflectance maps interpolate to an unseen novel viewpoint.*

Table 1: *Quantitative comparison of image-space view-dependent rendering and 3D rendering for a novel view to actual photograph. The reflectance maps used for neural 3D rendering are trained using 30 different views (without training on maps acquired from the validation viewpoint). Note that the reference image is lit with surrounding LCD screens, thus exhibiting near-field effects that aren't directly accounted for in the employed rendering model.*

Render Type	Liquor Bottle						Empty Green Bottle					
	Pisa			High Freq.			Pisa			High Freq.		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Image-based	23.876	0.835	0.110	19.519	0.751	0.169	22.870	0.896	0.081	23.416	0.881	0.085
3D Novel-view	17.573	0.752	0.170	17.265	0.657	0.226	19.916	0.834	0.124	16.644	0.788	0.151

Table 2: *Quantitative comparison of our novel-view rendering with NeRF and [BMR*22]. Our rendered image achieves comparable numerical performance with NeRF trained under low frequency sinusoidal illumination while having additional relighting capability under different lighting conditions. Note that [BMR*22]'s method is intended to work under highly varying background, and is more suitable for novel-view synthesis for complex scenes rather than reconstructing a single object under more controlled uniform illumination, as our work targets.*

Render Type	Liquor Bottle			Empty Green Bottle		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
NeRF	21.925	0.936	0.077	22.618	0.943	0.059
Bemana et al.	20.231	0.918	0.102	21.139	0.928	0.082
Ours	21.465	0.922	0.069	22.545	0.937	0.060

References

[BMR*22] BEMANA, MOJTABA, MYSZKOWSKI, KAROL, REVALL FRISVAD, JEPPE, et al. "Eikonal Fields for Refractive Novel-View Synthesis". *Special Interest Group on Computer Graphics and Interactive Techniques Conference Proceedings*. 2022, 1–9 2, 8–10.

[CLZ*20] CHE, CHENGQIAN, LUAN, FUJUN, ZHAO, SHUANG, et al. "Towards learning-based inverse subsurface scattering". *2020 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2020, 1–12 2.

[CZH*00] CHUANG, YUNG-YU, ZONGKER, DOUGLAS E, HINDORFF, JOEL, et al. "Environment matting extensions: Towards higher accuracy and real-time capture". *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. 2000, 121–130 2.

[DLW*22] DENG, XI, LUAN, FUJUN, WALTER, BRUCE, et al. "Reconstructing Translucent Objects using Differentiable Rendering". *ACM SIGGRAPH 2022 Conference Proceedings*. 2022, 1–10 2.

[FCM*08] FRANCKEN, YANNICK, CUYPERS, TOM, MERTENS, TOM, et al. "High quality mesostructure acquisition using specularities". *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. 2008, 1–7 2.
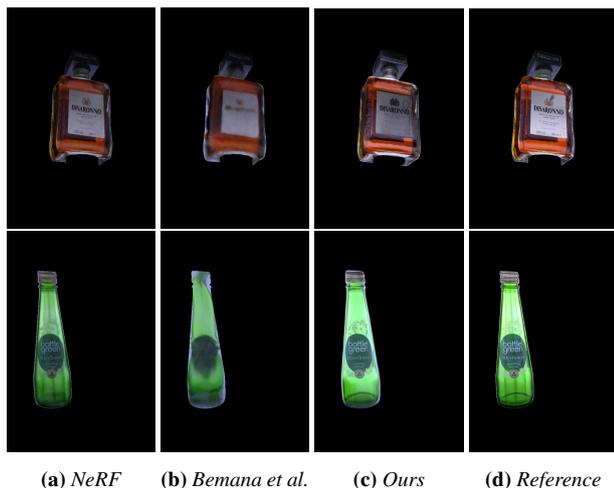
(a) *NeRF*   **(b)** *Bemana et al.*   **(c)** *Ours*   **(d)** *Reference*

Figure 11: *Comparisons of our novel-view renderings with NeRF [MST*21], and eikonal fields [BMR*22]. The background is masked out for better comparisons.*
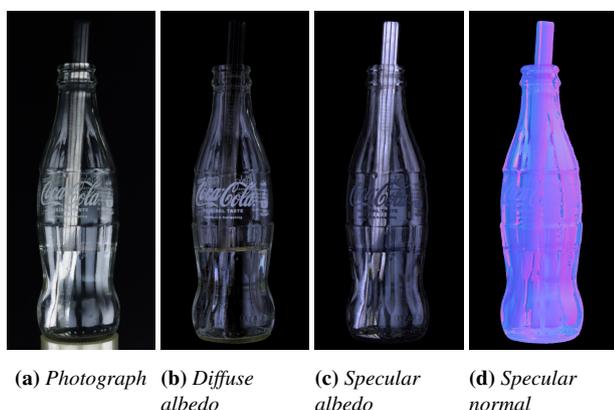

(a) *Photograph*   **(b)** *Diffuse albedo*   **(c)** *Specular albedo*   **(d)** *Specular normal*

Figure 12: *Transparency and internal objects causes internal reflections that hinder our separation result. Note that the diffuse albedo contains a lot of specular reflections caused by internal reflections, the specular normal is thus noisy and fragmented.*
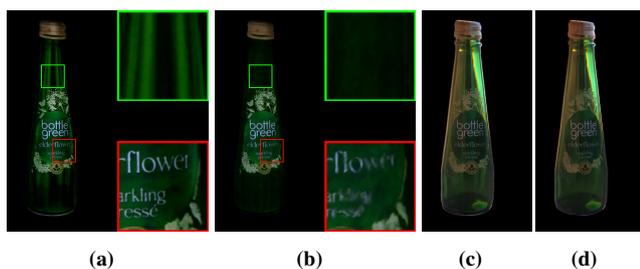

(a)   **(b)**   **(c)**   **(d)**

Figure 13: *Training an MLP for diffuse prediction eliminates interreflection (b) when compared to the orginal diffuse measurement (a). However, the quality of labels are deteriorated, and the rendered image quality with non-inter-reflecting diffuse (d) does not show obvious visual improvements over rendering useing measured diffuse (c).*

[GFT*11] GHOSH, ABHIJEET, FYFFE, GRAHAM, TUNWATTANAPONG, BOROM, et al. "Multiview face capture using polarized spherical gradient illumination". *Proceedings of the 2011 SIGGRAPH Asia Conference*. 2011, 1–10 2.

[GHLB15] GODARD, CLEMENT, HEDMAN, PETER, LI, WENBIN, and BROSTOW, GABRIEL J. "Multi-view reconstruction of highly specular surfaces in uncontrolled environments". *2015 International Conference on 3D Vision*. IEEE. 2015, 19–27 2.

[GZB*13] GKIOULEKAS, IOANNIS, ZHAO, SHUANG, BALA, KAVITA, et al. "Inverse volume rendering with material dictionaries". *ACM Transactions on Graphics (TOG)* 32.6 (2013), 1–13 2.

[HED05] HAWKINS, TIM, EINARSSON, PER, and DEBEVEC, PAUL. "A Dual Light Stage". *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*. EGSR '05. Konstanz, Germany: Eurographics Association, 2005, 91–98. ISBN: 3905673231 2.

[IRN*22] ISER, TOMÁŠ, RITTIG, TOBIAS, NOGUÉ, EMILIE, et al. "Affordable Spectral Measurements of Translucent Materials". *ACM Transactions on Graphics (TOG)* 41.6 (2022), 1–13 2.

[KGX*23] KANG, KAIZHANG, GU, MINYI, XIE, CIHUI, et al. "Neural Reflectance Capture in the View-Illumination Domain". *IEEE Transactions on Visualization and Computer Graphics* 29.2 (2023), 1450–1462. DOI: 10.1109/TVCG.2021.3117370 2.

[KRFB06] KHAN, ERUM ARIF, REINHARD, ERIK, FLEMING, ROLAND W., and BÜLTHOFF, HEINRICH H. "Image-Based Material Editing". *ACM SIGGRAPH 2006 Papers*. SIGGRAPH '06. Boston, Massachusetts: Association for Computing Machinery, 2006, 654–663. ISBN: 1595933646. DOI: 10.1145/1179352.1141937. URL: https://doi.org/10.1145/1179352.1141937 6.

[KRG17] KIM, JAEWON, RESHETOUSKI, ILYA, and GHOSH, ABHIJEET. "Acquiring axially-symmetric transparent objects using single-view transmission imaging". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017, 3559–3567 2.

[KVHS00] KAUTZ, JAN, VÁZQUEZ, PERE-PAU, HEIDRICH, WOLFGANG, and SEIDEL, HANS-PETER. "Unified Approach to Prefiltered Environment Maps". *Proceedings of the Eurographics Workshop on Rendering Techniques 2000*. Berlin, Heidelberg: Springer-Verlag, 2000, 185–196. ISBN: 3211835350 6.

[KXH*19] KANG, KAIZHANG, XIE, CIHUI, HE, CHENGAN, et al. "Learning Efficient Illumination Multiplexing for Joint Capture of Reflectance and Shape". *ACM Trans. Graph.* 38.6 (Nov. 2019). ISSN: 0730-0301. DOI: 10.1145/3355089.3356492. URL: https://doi.org/10.1145/3355089.3356492 2.

[LLK*22] LATTAS, ALEXANDROS, LIN, YIMING, KANNAN, JAYANTH, et al. "Practical and scalable desktop-based high-quality facial capture". *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI*. Springer. 2022, 522–537 2.

[LPGD09] LAMOND, BRUCE, PEERS, PIETER, GHOSH, ABHIJEET, and DEBEVEC, PAUL. "Image-based separation of diffuse and specular reflections using environmental structured illumination". *2009 IEEE International Conference on Computational Photography (ICCP)*. IEEE. 2009, 1–8 2, 3, 6.

[LWL*20] LYU, JIAHUI, WU, BOJIAN, LISCHINSKI, DANI, et al. "Differentiable refraction-tracing for mesh reconstruction of transparent objects". *ACM Transactions on Graphics (TOG)* 39.6 (2020), 1–13 2.

[LYC20] LI, ZHENGQIN, YEH, YU-YING, and CHANDRAKER, MANMOHAN. "Through the looking glass: Neural 3d reconstruction of transparent shapes". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, 1262–1271 2.

[MHP*07] MA, WAN-CHUN, HAWKINS, TIM, PEERS, PIETER, et al. "Rapid Acquisition of Specular and Diffuse Normal Maps from Polarized Spherical Gradient Illumination." *Rendering Techniques* 2007.9 (2007), 10 2.

[MLS*14] MA, CHENGUANG, LIN, XING, SUO, JINLI, et al. "Transparent Object Reconstruction via Coded Transport of Intensity". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2014 2.

[MPZ*02] MATUSIK, WOJCIECH, PFISTER, HANSPETER, ZIEGLER, REMO, et al. "Acquisition and rendering of transparent and refractive objects". (2002) 2.

[MST*21] MILDENHALL, BEN, SRINIVASAN, PRATUL P, TANCIK, MATTHEW, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis". *Communications of the ACM* 65.1 (2021), 99–106 8, 10.

[NGD*06] NARASIMHAN, SRINIVASA G, GUPTA, MOHIT, DONNER, CRAIG, et al. "Acquiring scattering properties of participating media by dilution". *ACM SIGGRAPH 2006 Papers*. 2006, 1003–1012 2.

[PD03] PEERS, PIETER and DUTRÉ, PHILIP. "Wavelet environment matting". *Proceedings of the 14th Eurographics workshop on Rendering*. 2003, 157–166 2.

[RDL*15] REN, PEIRAN, DONG, YUE, LIN, STEPHEN, et al. "Image based relighting using neural networks". *ACM Transactions on Graphics (ToG)* 34.4 (2015), 1–12 2.

[RH01] RAMAMOORTHI, RAVI and HANRAHAN, PAT. "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object". *JOSA A* 18.10 (2001), 2448–2459 3.

[SESM22] SUHAIL, MOHAMMED, ESTEVES, CARLOS, SIGAL, LEONID, and MAKADIA, AMEESH. "Light Field Neural Rendering". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, 8269–8279 2.

[TFG*13] TUNWATTANAPONG, BOROM, FYFFE, GRAHAM, GRAHAM, PAUL, et al. "Acquiring reflectance and shape from continuous spherical harmonic illumination". *ACM Transactions on graphics (TOG)* 32.4 (2013), 1–12 2.

[VHM*21] VERBIN, DOR, HEDMAN, PETER, MILDENHALL, BEN, et al. "Ref-nerf: Structured view-dependent appearance for neural radiance fields". *arXiv preprint arXiv:2112.03907* (2021) 2.

[WLL*21] WANG, PENG, LIU, LINGJIE, LIU, YUAN, et al. "NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction". *NeurIPS* (2021) 3, 5, 6.

[WZQ*18] WU, BOJIAN, ZHOU, YANG, QIAN, YIMING, et al. "Full 3D reconstruction of transparent objects". *ACM Transactions on Graphics (TOG)* 37.4 (2018) 2.

[YX16] YANG, JINGJIE and XIAO, SHUANGJIU. "An inverse rendering approach for heterogeneous translucent materials". *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1*. 2016, 79–88 2.

[ZWCS99] ZONGKER, DOUGLAS E, WERNER, DAWN M, CURLESS, BRIAN, and SALESIN, DAVID H. "Environment matting and compositing". *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. 1999, 205–214 2.