

Article

Data Driven SVBRDF Estimation Using Deep Embedded Clustering

Yong Hwi Kim ¹  and Kwan H. Lee ^{2,*}

¹ IT Convergence Components Research Center, Korea Electronics Technology Institute (KETI), Gwangju 61005, Korea

² School of Mechanical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Korea

* Correspondence: khlee@gist.ac.kr; Tel.: +82-62-715-2386

Abstract: Photo-realistic representation in user-specified view and lighting conditions is a challenging but high-demand technology in the digital transformation of cultural heritages. Despite recent advances in neural renderings, it is still necessary to capture high-quality surface reflectance from photography in a controlled environment for real-time applications such as VR/AR and digital arts. In this paper, we present a deep embedding clustering network for spatially-varying bidirectional reflectance distribution function (SVBRDF) estimation. Our network is designed to simultaneously update the reflectance basis and its linear manifold in the spatial domain of SVBRDF. We show that our dual update scheme excels in optimizing the rendering loss in terms of the convergence speed and visual quality compared to the current iterative SVBRDF update methods.

Keywords: SVBRDF; deep embedding clustering; appearance capture; surface reflectance estimation



Citation: Kim, Y.H.; Lee, K.H. Data Driven SVBRDF Estimation Using Deep Embedded Clustering. *Electronics* **2022**, *11*, 3239. <https://doi.org/10.3390/electronics11193239>

Academic Editor: Andrei Kelarev

Received: 5 September 2022

Accepted: 3 October 2022

Published: 9 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Realistic representation of surface reflectance is a well-known but still challenging problem in computer graphics and vision. For an opaque object, it has been commonly modeled using a Bidirectional reflectance distribution function (BRDF) when the material has homogeneous reflectance properties. However, for an opaque object with inhomogeneous reflectance properties, which occurs when representing a material more realistically, the surface reflectance needs to be represented by a spatially-varying bidirectional reflectance distribution function (SVBRDF). A rich body of studies has estimated various types of SVBRDF using linear decomposition [1–5], independent BRDF reconstruction [6–8], and local linear embedding using BRDF examples [9–12].

The aforementioned methods require a dense sample of surface reflectance at multiple views under given lighting conditions since they are weak for generalization ability for which we can leverage inter-sample variability on SVBRDF estimation. For example, deep learning models further generalize model constraints using self-augmentation [13,14], random mixture of procedural SVBRDFs [15], or rendering with priors on shape and reflectance [16–18]. These methods are able to capture SVBRDF from a few images based on reflectance basis generated by prior information or self-augmentation scheme. Recently, neural radiance fields (NeRFs) have become a mainstream technology for reflectance estimation. They can learn a sparse representation of surface reflectance via view synthesis [19,20]. Beneficial to implicit view synthesis, NeRFs alleviate the complexity of the image capturing process.

However, there is still a high demand for full-sampling of surface reflectance, especially for a non-planar object in cultural heritages and AR/VR [21,22]. Due to the fact that cultural artifacts often consist of a non-planar geometry and a complex surface reflectance including non-diffuse reflectance properties, the current deep learning approaches have several limitations when applied to applications in which a high-quality reproduction of

surface reflectance is required. Conceptually, their methods are capable of approximating SVBRDF with sparse samples. However, practically, their solution space is limited to a flat surface [13–15], or a single view [16–18]. In the same manner, recent advances in NeRFs allow us to handle global illumination changes. However, their solutions are yet bounded for a simple object with a parametric BRDF [23].

In this paper, we present a novel unsupervised deep learning-based method to estimate SVBRDFs from full reflectance samples. In contrast to the recent deep learning networks that infer a parametric BRDF [13,18,23,24], we adopted a general microfacet model [25] using an isotropic Normal Distribution Function (NDF). Our network trains for a non-parametric tabular form of basis BRDFs which allows a better representation of reflectance than the BRDFs using a parametric NDF such as Beckmann [25] and GGX [26]. Furthermore, our network architecture is designed to simultaneously learn a linear reflectance manifold spanned by NDF basis and the NDF basis, itself, by a novel autoencoder structure with two custom layers: the Reprojection layer that updates the NDF basis via rendering loss, and the Distance measure layer that remaps NDF basis in the spatial domain. These properties allow for a fast and more accurate approximation for SVBRDF compared to the current iterative update scheme [1,2,27,28] by the dual gradient search scheme.

The key observation found is that a gradient on a linear manifold of SVBRDF can be represented by a function of both basis BRDFs and the corresponding feature points in the manifold. Based on this observation, we have designed a non-linear autoencoder that simultaneously updates both the basis BRDFs and a spatial weight function that maps the observed reflectance to a linear manifold of the basis BRDFs. Our network design is inspired by the concept of the Deep Embedded Clustering (DEC) [29] which facilitates a joint-optimization of a mapping function and the clusters in an embedded space at the same time. In order to employ their strategy to learn a faithful linear manifold of the SVBRDF, we designed two custom layers that perform to measure similarity between reflectance at a surface point and the basis BRDFs and to render through basis BRDFs.

We believe that it yields a better solution in terms of accuracy and convergence speed compared to the existing methods which have fixed basis functions when forming a linear manifold [3,30], or updating the basis functions and blending weights iteratively [1,2,27]. The benefits of the DEC are also inherited from our network. It is less sensitive to the use of varying hyperparameters and provides state-of-the-art performance for a target application. A variational autoencoder-based deep embedded network is beneficial to high-resolution and non-planar geometry reflectance samples since it does not employ a shape-constrained layer, such as convolutional layers or self-attention modules, on the network.

In summary, the contributions of our research are as follows:

- We proposed Deep embedding clustering-based joint update scheme simultaneously updating non-parametric basis BRDFs and their linear manifold.
- We designed a novel autoencoder that is suited for learning the appearance properties of complex objects using unsupervised learning.
- We demonstrate that our network produces a high-quality rendering result in different illumination conditions.

2. Related Work

2.1. Point-Wise Estimation

The most common way to represent spatially-varying effects over a surface is to treat each surface point as an independent spatial sample. Gardner et al. [6] fitted an isotropic BRDF model at each surface point under a linear light source. Similarly, Aittala et al. [7] used a two specular lobe model to reconstruct the spatially-varying reflectance of a planar object lit by a programmable LCD light source. To simplify the acquisition setup, Dong et al. [31] captured a video sequence in a natural illumination and recovered a data-driven microfacet model at each surface point using a temporal variation in the appearance. Recently, Kang et al. [8] combined a non-linear autoencoder with a near-field acquisition system that emits multiplexing light patterns, in order to efficiently estimate per-point-

based anisotropic BRDFs using fewer samples (16~32 images). All these methods were capable of reconstructing individual BRDFs at each surface point using rich information (e.g., programmed lighting or appearance variation from motion). However, they have not utilized spatial coherency of the surface reflectance which is commonly observed in real materials [11,30,32].

2.2. Estimation Using BRDF Examples

Another class of the SVBRDF estimation method utilizes BRDF examples, called representatives, before manifold mapping in the spatial domain of SVBRDF. These methods form a local linear subspace at every surface point using only the selected representatives that are closed to the observed reflectance. Researchers have proposed various methods to collect representatives: manual acquisition [9,11], NDF synthesis at spatial clusters [32], and the use of a reference chart [10]. Other researchers used data-driven BRDFs (e.g., MERL BRDFs [33]) to jointly optimize the surface normal and reflectance [34,35], or a BRDF dictionary that consists of a set of parametric BRDFs sampled uniformly from a parametric domain. Wu et al. [36] proposed a sparse mixture model of analytic BRDFs to represent Bidirectional Texture Functions (BTFs). Chen et al. [28] utilized a BRDF dictionary with a sparse constraint to reconstruct anisotropic NDFs. Wu et al. [37] used a single parametric BRDF which is best suited for the observed reflectance at each point in order to jointly optimize the geometry, environmental map, and reflectance from an RGBD sequence. Zhou et al. [12] optimized not only blending weights but the number of basis functions blended at a surface point.

Most of the above methods share the same limitation that they rely on various types of given BRDFs. In the case that a prior distribution does not match the appearance of a target object, these methods could fail to produce a proper solution.

2.3. Estimation Using Linear Subspace

Studies in this category seek to use a low-dimensional subspace to represent the spatially-varying reflectance. The linear decomposition methods represent the reflectance at a surface point as a linear combination of basis functions. The basis functions have been constructed by matrix factorization [1,27], or spatial clustering based on the similarity of diffuse albedo [2,4,30,38] and the distance to a cluster center [3,5].

Most methods have fixed the basis functions while forming a linear manifold [3–5,30,38] or iteratively updated basis functions and their blending weights [1,2,27]. In comparison to these methods, our method can make further improvements by performing a simultaneous update of both basis NDFs and blending weights.

2.4. Estimation Using Deep Learning

Deep learning-based methods have also been developed to estimate the reflectance properties and they can be broadly categorized into two types: the supervised and the unsupervised. They differ from one another based on the existence of a separate dataset to train the network. Early work in the supervised method reconstructs the intrinsics of Lambertian surface from a collection of real photographs [39–42]. It is extended to a non-Lambertian surface [17] but their material representation is still based on a simple dichromatic reflection model. For an accurate representation of surface reflectance, many supervised methods have utilized more complex SVBRDFs for inverse rendering. Aittala et al. [43] captured the appearance of a flat surface under a flash light source using Convolutional Neural Network (CNN)-based texture synthesis. Li et al. [13] also recovered the SVBRDF of a flat material that had a homogeneous specular reflection over the surface using CNN-based networks trained by a self-augmentation scheme. Their subsequent work [14] extended it for inexact supervision based on empirical experiments on self-augmentation. Li et al. [24] further refined SVBRDF parameters using learnable non-local smoothness conditions. Deschaintre et al. [15] introduced a rendering loss that compared the appearance of an estimated SVBRDF with the ground truth photographs rendered by in-network rendering

layers. Their work had strength in terms of producing a globally consistent solution. Last but not least, Li et al. [18] utilized U-net-based CNNs for estimating the shape and reflectance under multi-bounce reflections. Most supervised methods for SVBRDF estimation rely on strong supervision over the appearance of materials so that they can produce a reliable solution even under an ill-posed condition. However, they limit the range of target application to a near-planar surface [13,15,24,43] or a single-view case [17,18].

Unsupervised learning methods have received less attention in resolving the inverse rendering problem. Kulkarni et al. [44] proposed a variational autoencoder that captured a distribution over the shape and appearance of face images. Tewari et al. [45] employed CNN-based autoencoders for facial reconstruction from a monocular image. Tanini et al. [46] resolved the photometric stereo problem using two CNNs that captured per-point surface normal and reflectance, respectively. Compared to these methods, however, our method is oriented toward an object-specific SVBRDF estimation which can handle multi-view inputs of a 3D object. A general microfacet model that our network trains on yields a more realistic solution of relighting for virtual environment applications.

3. Preliminaries

In this chapter, we describe the terminology and assumptions used in this paper. We assume that a target object is lit by a dominant light source (e.g., photometric stereo [46], mobile flash photograph [2]) and its surface reflectance is isotropic. We do not consider multi-bounce reflections. The irradiance function $L(\mathbf{x}, \mathbf{i})$ and geometry of the surface are presumed to be given in advance. A BRDF sample $\tilde{\rho}$ is expressed in a local frame where the upward direction is aligned with the surface normal \mathbf{n} at each \mathbf{x} . From the radiance equation introduced under a point light source in [47], $\tilde{\rho}$ can be formulated as

$$\tilde{\rho}(\mathbf{x}, \mathbf{i}, \mathbf{o}) = \frac{I(\mathbf{x}')}{L(\mathbf{x}, \mathbf{i})(\mathbf{i} \cdot \mathbf{n})}, \quad (1)$$

where x' is the pixel location corresponding to \mathbf{x} , and \mathbf{i} and \mathbf{o} are an incident and outgoing vector, respectively.

The surface reflectance is modeled by the general microfacet BRDF [25] as the following:

$$\rho(\mathbf{x}, \mathbf{i}, \mathbf{o}) = \rho_d(\mathbf{x}, \mathbf{i}, \mathbf{o}) + \frac{D(\mathbf{x}, \mathbf{h})G(\mathbf{x}, \mathbf{i}, \mathbf{o})F(\mathbf{x}, \mathbf{i}, \mathbf{o})}{4(\mathbf{i} \cdot \mathbf{n})(\mathbf{o} \cdot \mathbf{n})}, \quad (2)$$

where the half vector \mathbf{h} is calculated by the normalized sum of \mathbf{i} and \mathbf{o} as $\mathbf{h} = (\mathbf{i} + \mathbf{o}) / \|\mathbf{i} + \mathbf{o}\|$.

$\rho_d(\mathbf{x}, \mathbf{i}, \mathbf{o})$ is the diffuse term which is a blackbox expression of non-local effects such as sub-surface scattering and inter-reflections. It is simply modeled by the Lambertian surface model: $\rho_d(\mathbf{x}, \mathbf{i}, \mathbf{o}) = k_d(\mathbf{x}) / \pi$, where $k_d(\mathbf{x})$ is the diffuse coefficient. For isotropic reflectance, the NDF $D(\mathbf{x}, \mathbf{h})$ can be expressed as a univariate function of the angle θ_h between \mathbf{h} and \mathbf{n} : $\theta_h = \cos^{-1}(\mathbf{h} \cdot \mathbf{n})$. We impose the non-negativity constraint: $D(\mathbf{x}, \mathbf{h}) \geq 0$ and normalize it so that $\int_{\Omega_+(n)} (\mathbf{h} \cdot \mathbf{n}) D(\mathbf{x}, \mathbf{h}) d\omega_h = 1$ where $\Omega_+(n)$ is the upper hemispherical domain which satisfies $(\mathbf{h} \cdot \mathbf{n}) > 0$. For this, we set the specular coefficient $k_s(\mathbf{x})$ as

$$k_s(\mathbf{x}) = \int_{\Omega_+(n)} (\mathbf{h} \cdot \mathbf{n}) D(\mathbf{x}, \mathbf{h}) d\omega_h. \quad (3)$$

$G(\mathbf{x}, \mathbf{i}, \mathbf{o})$ is a geometric term for the shadowing and masking effects by microfacets. We use the uncorrelated geometric term expressed as the product of two independent shadowing functions derived by Ashikmin et al. [48]:

$$G(\mathbf{x}, \mathbf{i}, \mathbf{o}) = S(\mathbf{x}, \mathbf{i})S(\mathbf{x}, \mathbf{o}), \quad (4)$$

where

$$S(\mathbf{x}, \mathbf{k}) = \frac{(\mathbf{k} \cdot \mathbf{h})}{\int_{\Omega_+(k) \cap \Omega_+(n)} (\mathbf{k} \cdot \mathbf{h}) D(\mathbf{x}, \mathbf{h}) d\omega_h}, \quad (5)$$

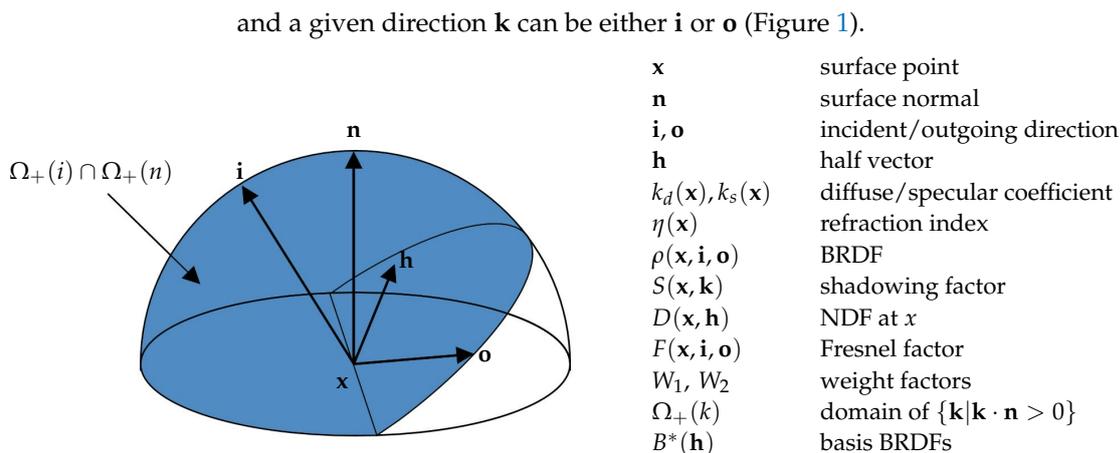


Figure 1. Integration domain of $S(\mathbf{x}, \mathbf{i})$ denoted as the blue area, and notations used in BRDF definition. Integration domain of $S(\mathbf{x}, \mathbf{o})$ can also be computed in the same manner.

$F(\mathbf{x}, \mathbf{i}, \mathbf{o})$ accounts for the Fresnel effect which is dominant at near-grazing angles. We use the Schlick’s approximation [49] for computational efficiency. It can be represented as a linear function of the refraction index $\eta(\mathbf{x})$ at \mathbf{x} as

$$F(\mathbf{x}, \mathbf{i}, \mathbf{o}) = \eta(\mathbf{x}) + (1 - \eta(\mathbf{x}))(1 - (\mathbf{i} \cdot \mathbf{h}))^5. \tag{6}$$

As a result, $\rho(\mathbf{x}, \mathbf{i}, \mathbf{o})$ can be rewritten as

$$\rho(\mathbf{x}, \mathbf{i}, \mathbf{o}) = \frac{k_d(\mathbf{x})}{\pi} + k_s(\mathbf{x}) \frac{D(\mathbf{x}, \mathbf{h})S(\mathbf{x}, \mathbf{i})S(\mathbf{x}, \mathbf{o})F(\mathbf{x}, \mathbf{i}, \mathbf{o})}{4(\mathbf{i} \cdot \mathbf{n})(\mathbf{o} \cdot \mathbf{n})}. \tag{7}$$

Our method takes two input tensors I and P constructed from the pairs of an image, the corresponding light source and camera poses. The observation tensor $I \in \mathbb{R}^{n_p \times m \times 3}$ contains per-point pixels extracted from multi-view images, where n_p is the number of surface points, m is the number of images. The cosine-weight tensor $P \in \mathbb{R}^{n_p \times m \times 4}$ contains four cosine weights $P_{ij} = (\mathbf{i} \cdot \mathbf{n}, \mathbf{o} \cdot \mathbf{n}, \mathbf{h} \cdot \mathbf{n}, \mathbf{i} \cdot \mathbf{h})$ for each I_{ij} where $i = 1, 2, \dots, n_p$ and $j = 1, 2, \dots, m$.

4. Deep Embedding Network

The objective of our network is to reconstruct an observed input image I as close as possible. Our network, the so-called Deep Embedding Network (DEN), does not rely on a prior distribution of the reflectance provided by training data. It instead attempts to estimate the basis BRDFs B^* represented as a general microfacet model [25], and their manifold from I . These two attributes are simultaneously optimized while training the network through the back-propagation of gradients of the rendering loss.

The design of the DEN is inspired by the deep embedded clustering method [29] which jointly updates a mapping function for a low-dimensional manifold and cluster centroids. The deep embedded clustering is specialized for an image classification problem where its kernel function measures the similarity between low-dimensional points and cluster centroids by an auxiliary target distribution based on the Student’s t-distribution. Since the observed reflectance is not well described by a mixture of Gaussian-like distributions due to the fact that the BRDF is represented as a non-linear function of several BRDF terms described in Equation (7), the DEN employs a different kernel function which measures the weighted reflectance difference [2,32] as the similarity between reflectance at \mathbf{x} and B^* . In order to apply this kernel function to the encoder, we have introduced a new layer, called the Distance Measure (DM) layer as the upfront module of the encoder (Section 4.1). We have also customized our decoder using a single layer, called the Re-Projection (RP) layer, so that it can compute the rendering loss of B^* and low-dimensional features z embedded

in the manifold of B^* (Section 4.2). In our network design, the weights used for two custom layers DM and RP, which are represented as B^* , are identical. They share the weights while training the network in order to avoid overlapping updates.

Our network architecture is illustrated in Figure 2. The structure of the DEN has single encoder and single decoder analogs with the structure of a variational autoencoder [29,44,45]. Our non-linear encoder as a non-linear mapping function f_{Θ} consists of the DM layer and a series of Fully-Connected (FC) layers, where Θ indicates learnable parameters of the mapping function f . f_{Θ} maps the observation features I_i into the latent space $Z \in R^{n_c}$ where $I = \{I_i \in R^{m \times 3}\}$, $Z = \{z_i \in R^{n_c}\}$, and n_c is the number of basis BRDFs. The single-layer decoder reconstructs I_i from z_i using the rendering equation [47]. In contrast to [44,45], we do not utilize convolutional layers since they impose a strong constraint on the image resolution and are inefficient in handling masked inputs which are common in multi-view images.

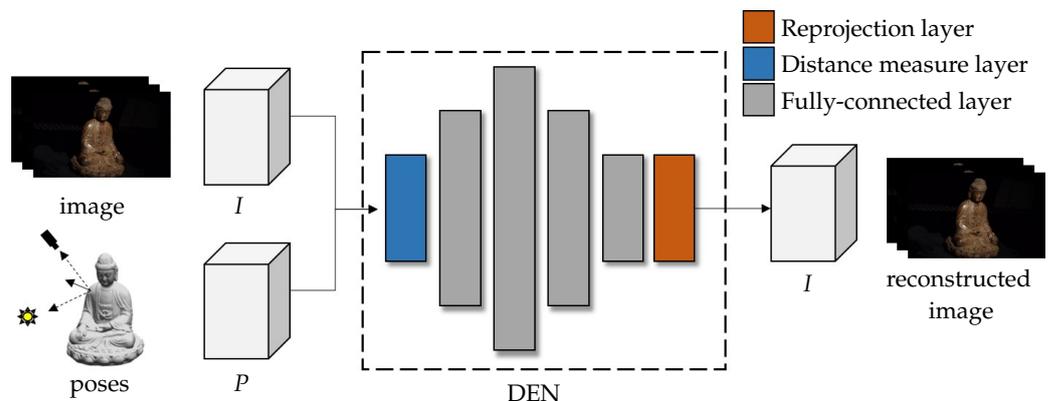


Figure 2. Overview of our method. From two input tensors (I, P), the DEN reconstructs the input I by updating the parameters of the network through the rendering loss for the basis BRDFs and the latent features (Section 4.2). The encoder which maps the input I into the latent space consists of the distance-measure layer and the fully-connected layers. The single-layer decoder (Reprojection layer) reproduces the input I from the encoded features.

4.1. Manifold Mapping

Our encoder learns the linear manifold of the basis BRDFs $B^*(\mathbf{i}, \mathbf{o})$ which are defined as a non-parametric tabulated function that considers three color bands: $B^*(\mathbf{i}, \mathbf{o}) \in R^{(n_s+1) \times n_c \times 3}$ where n_s is the number of samples in the NDF domain. The diffuse albedo of each of the basis BRDFs is also listed in the tensor so that there is one additional entry in the first dimension of B^* [2]. In this paper, we manually set n_s as 90 according to Matusik et al. [33]. We further transform the NDF domain θ_h by $\theta'_h = \sqrt{\theta_h}$ in order to densely sample θ_h near 0 [33].

The reflectance at each \mathbf{x} is represented as the sum of $B^*(\mathbf{i}, \mathbf{o})$ by

$$\rho(\mathbf{x}, \mathbf{i}, \mathbf{o}) = \sum_{c=1}^{n_c} z_{ic} B_c^*(\mathbf{i}, \mathbf{o}), \tag{8}$$

where $B_c^*(\theta_i, \theta_o)$ is a slice of B^* which represents a single basis BRDF $B_c^* \in R^{91 \times 3}$ and z_{ic} is the corresponding feature in an embedded feature vector z_i .

A direct mapping from I into Z [29,44] is undesirable for this case as we mentioned in the above section. Therefore, the DM layer in front of f_{Θ} firstly transforms I into a low-dimensional space by measuring the weighted mean squared distance $d_i \in R^{n_c}$ between a partial observation of reflectance at each \mathbf{x} and each B_c^* :

$$d_{ic} = \frac{\sum (B_c^*(\mathbf{i}, \mathbf{o}) - \tilde{\rho}(\mathbf{x}_i, \mathbf{i}, \mathbf{o}))(\mathbf{h} \cdot \mathbf{n})}{\sum (\mathbf{h} \cdot \mathbf{n})}. \tag{9}$$

The following eight FC layers map d_i into Z . We flexibly set the number of nodes in FC layers according to n_c so that each layer has $\{2n_c, 4n_c, 8n_c, 16n_c, 8n_c, 4n_c, 2n_c, n_c\}$ nodes,

respectively. In this paper, the learnable parameter Θ of the mapping function corresponds to the weights of FC layers.

For all FC layers, we apply the ‘Tanh’ activation function except for the last one. In the last layer, the ‘softmax’ activation function is applied so that z_i has non-negative weights for B^* where $\sum_c z_{ic} = 1$. During the training stage, we set the weight of the DM layer to be non-trainable and copy B^* from the RP layer to the DM layer after each batch.

4.2. Rendering Loss

The RP layer is a single-layer decoder that computes the rendering loss under a single light source condition similar to [15,46]. We have further customized it to allow for the use of the non-parametric BRDFs B^* .

The rendering loss is defined as

$$L = \|I(\mathbf{x}') - \rho(\mathbf{x}, \mathbf{i}, \mathbf{o})L(\mathbf{x}, \mathbf{i})(\mathbf{i} \cdot \mathbf{n})\| \quad (10)$$

where $\|\cdot\|$ indicates the mean squared difference. Gradients for the two major parameters D^* and z_i are defined as

$$\begin{aligned} \frac{\partial L}{\partial z_{ic}} &= -\frac{2}{m} \sum W_1 W_2 B_c^*, \\ \frac{\partial L}{\partial B_c^*} &= -\frac{2}{m} \sum W_1 W_2 z_{ic}. \end{aligned} \quad (11)$$

where $W_1 \in R^{m \times 3}$ and $W_2 \in R^{m \times 3}$ are weight tensors which can be formulated as

$$\begin{aligned} W_1(\mathbf{x}, \mathbf{i}, \mathbf{o}) &= \frac{k_s(\mathbf{x})S(\mathbf{x}, \mathbf{i})S(\mathbf{x}, \mathbf{o})F(\mathbf{x}, \mathbf{i}, \mathbf{o})}{4(\mathbf{i} \cdot \mathbf{n})(\mathbf{o} \cdot \mathbf{n})} \\ W_2(\mathbf{x}, \mathbf{i}, \mathbf{o}) &= \{I(\mathbf{x}') - \rho(\mathbf{x}, \mathbf{i}, \mathbf{o})L(\mathbf{x}, \mathbf{i})(\mathbf{i} \cdot \mathbf{n})\}L(\mathbf{x}, \mathbf{i})(\mathbf{i} \cdot \mathbf{n}), \end{aligned} \quad (12)$$

4.3. Network Initialization

For the initialization of the DEN, the weights of FC layers are initialized by the ‘glorot’ uniform distribution [50]. To initialize the weights of the DM and RP layers, we firstly performed a manual separation of the diffuse and the specular components for each \mathbf{x} by using a minimum filter [10,32]. We grouped I_i to n_c clusters using a k-means algorithm based on the L2 distance between cluster centroids and $k_d(\mathbf{x}_i)$ in the RGB color space. An initial guess of B^* was calculated for collected reflectance samples in each cluster. For this, we excluded the samples of near-grazing angles where $\cos(\theta_i) < 0.2$ or $\cos(\theta_o) < 0.2$ and assumed $F(\mathbf{x}, \mathbf{i}, \mathbf{o}) = 1$ [32]. $S(\mathbf{x}, \mathbf{i})$ and $S(\mathbf{x}, \mathbf{o})$ are also assumed to be equal to 1 at the starting point. After the calculation of B^* has been performed, the weights of the DM and RP layers are initialized to the value of B^* .

5. Results

5.1. Comparison for the Reconstruction Quality

We validated our network on 3D objects (Buddha, Minotaur, and Terracotta) provided by Schwartz et al. [51]. This 3D dataset aims to estimate a dense BTF of objects which have complex shapes and surface reflectance so that they captured $151 \times 151 = 22,801$ sample images. For 3D objects, we parameterized the surface of a triangle mesh into a 2D texture map using the ABF++ mesh parameterization method [52]. Each 2D texture map is rasterized in the resolution of 1024×1024 . To construct the input tensors I and P , we selected the co-located images, where the camera and light source positions are identical [34].

We compared the DEN with the Discrete Spatial Clustering (DSC) [4] and the Log-Mapping PCA (LMPCA) [53] method. The DSC fits a parametric BRDF to reflectance data collected in each cluster. In the comparison, we used the Cook–Torrance [25] BRDF with a parametric NDF for DSC and LMPCA.

Table 1 shows the PSNR of reconstruction results with respect to the ground truth. Overall, the DEN outperforms the competitive methods. This improvement is derived from the mechanism of the DEN that can learn non-parametric NDFs with its blending weights in the spatial domain. Since the non-parametric NDF is capable of representing a comprehensive surface reflection by giving more degrees of freedom in the shape of specular reflection, the reconstruction result of the DEN is improved in comparison with the method using a parametric BRDF [4] or a pre-defined BRDF [53].

Figure 3 shows the visual comparison of 3D objects. In contrast to the DSC which fails to capture underlying surface reflectance including an irregular mixture of materials, the DEN robustly reproduced the surface reflectance. Meanwhile, the result of the LMPCA seems to be over-saturated in specular regions.



Figure 3. Visual comparison on Buddha, Minotaur, and Terracotta dataset: (a) Ground truth photograph; (b) DEN; (c) DSC; (d) LMPCA.

Table 1. PSNR comparison with DEN, DSC, and LMPCA. We evaluated the PSNR of each reconstruction result with respect to the ground truth input tensor I .

Object	PSNR		
	DEN	DSC [4]	LMPCA [53]
Buddha	31.5	29.7	30.0
Minotaur	36.9	31.8	32.4
Terracotta	40.8	34.5	35.4

To observe whether the DEN is over-fitted to training data, we evaluate the RMSE of test data that has a novel view and illumination conditions. The result in Table 2 indicates that the SVBRDF estimated from the DEN is not over-fitted and well applied to different views and illumination conditions from that of images in training data. This characteristic is very important to guarantee visual consistency in real-time rendering (Section 5.7).

Table 2. RMS errors of the DEN for the training data and the test data in the BTF dataset. The number of basis BRDFs n_c are also listed in the table.

Object	n_c	RMSE of Training Data	RMSE of Test Data
Buddha	9	0.0264	0.0274
Minotaur	10	0.0143	0.0141
Terracotta	9	0.0091	0.0090

5.2. Comparison with the Iterative Optimization Method

Iterative optimization methods which separately optimize the basis functions of the SVBRDF and the blending weights have been widely used [1,2,27,28]. We compared the convergence speed with that of the iterative optimization method. For comparison, we implemented a quadratic programming-based optimizer that updates B^* and z_i by iteratively solving two quadratic equations in Equation (10). Refer to Nam et al. [2] for the details of constructing these quadratic equations. While training the iterative method, we used the same constraints imposed on the DEN. Figure 4 shows RMS errors in terms of the computation time. Our method converged faster than the iterative optimization due to the simultaneous update of B^* and z_i . Nevertheless, our method shows a good visual quality as shown in Figure 4.

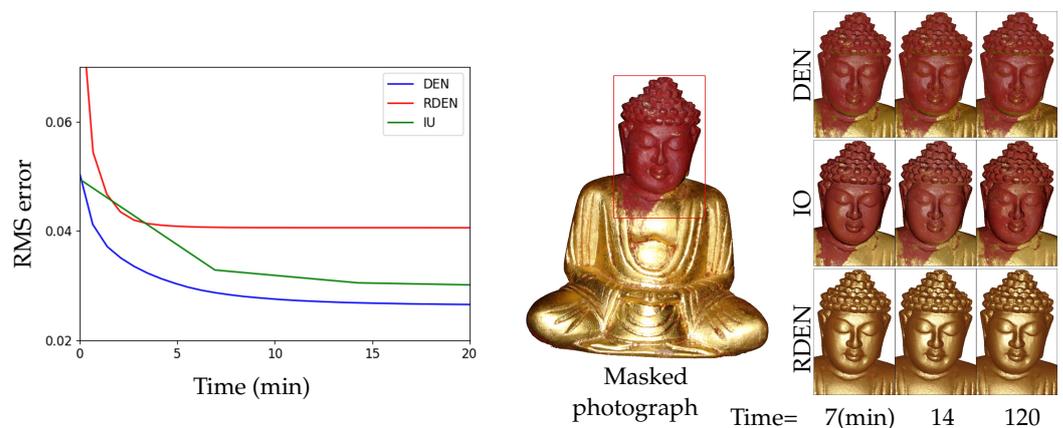


Figure 4. (left) RMS errors between reconstructions and photographs according to the computation time. (right) DEN: DEN with initialization of B^* , RDEN: Randomly-initialized DEN by gloriot uniform distribution, IO: Iterative Optimization. The middle figure is a masked photograph. The highlighted figures are the reconstruction result of each method at $t = 7, 14$, and 120 min.

5.3. Gradient Observation

Figure 5 shows the histogram of the magnitude of the gradients $\left| \frac{\partial L}{\partial B_c^*} \right|$ according to $|z_{i,c}|$ at the first epoch. As noted in Equation (11), $\left| \frac{\partial L}{\partial B_c^*} \right|$ is linearly proportional to $|z_{i,c}|$. It indicates that B_c^* which is closer to a partial observation $\tilde{B}(x, \theta_i)$ has been updated more rapidly. This characteristic plays an important role in the convergence of the DEN. We observed that a random initialization of B^* leads to local minima in which the DEN gave a poor result (Figure 4(left)). However, when an initial value of B^* computed by a discrete spatial clustering [2,3,5,38] is given to the DEN, it outperforms the iterative optimization method for B^* and z_j .

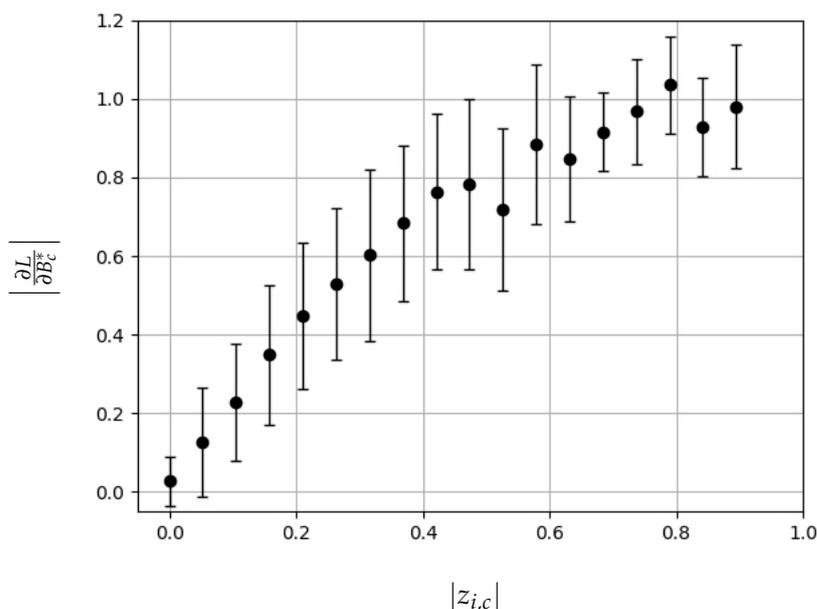


Figure 5. Histogram of the first gradients for the basis NDF $\left| \frac{\partial L}{\partial B_c^*} \right|$ according to the magnitude of blending weight $|z_{i,c}|$. We uniformly spaced 20 bins in the domain of $|z_{i,c}|$. The dots and the error bars indicate the mean and the standard deviation of samples in each bin. Note that there shows no sample in the last two bins.

5.4. Comparison of Photometric Stereo Data

The DiLiGenT benchmark dataset (Shi et al. [54]) provides 96 photometric stereo images and the corresponding light source positions for 10 objects, as well as their ground truth surface normal. They also provide the BRDF fitting results of Alldrin et al. [27] under the given surface normal and the given number of basis functions as three.

Table 3 summarizes the comparison result in terms of the PSNR. Our method shows a better performance than the ACLS [27]. Although the PSNR in several objects is marginally improved by the proposed method, we have observed that the DEN significantly improves the visual quality of the reconstruction. Figure 6 shows highlighted results for three objects: Harvest, Reading, and Ball, to compare the visual quality of the two methods. The DEN produces more realistic solutions than the competitive method in both specular and diffuse regions.

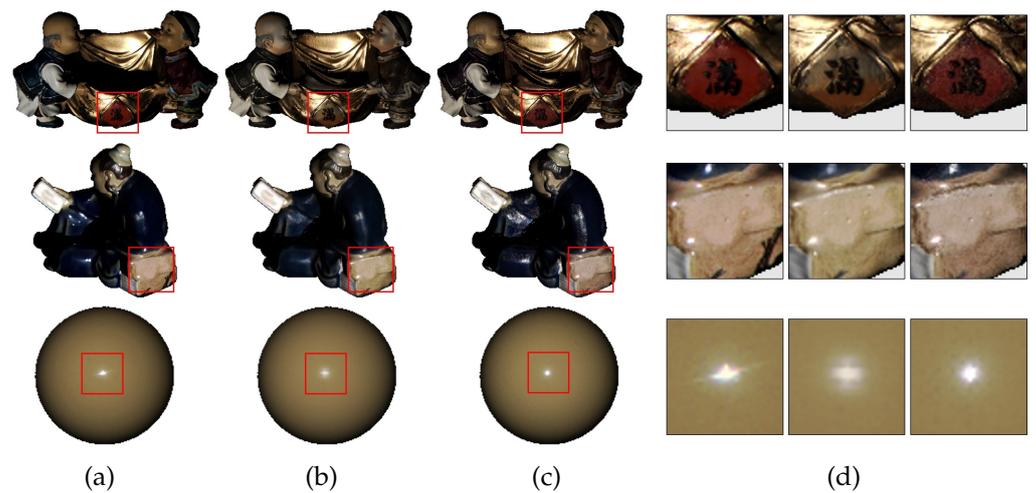


Figure 6. Visual comparison on the Harvest, Reading, and Ball dataset: (a) Ground truth photograph; (b) Alldrin et al. [27]; (c) Ours; (d) Highlighted regions in (a–c).

Table 3. Comparison of PSNR with ACLS for the “DiLigenT” dataset. We generated the reconstruction results using the ground truth surface normal for both methods.

Object	PSNR	
	ACLS [27]	DEC
Ball	38.9	43.3
Bear	43.1	41.4
Buddha	38.2	39.2
Cat	43.1	43.3
Cow	40.3	39.8
Goblet	36.7	35.7
Harvest	30.4	30.8
Pot1	45.8	48.9
Pot2	45.6	49.2
Reading	27.5	29.0

5.5. Sensitivity for Normal Error

The reconstruction quality of the DEN can be affected by the noise in the surface normal since it relies on the given surface geometry and camera configurations. We measured the sensitivity of the reconstruction results in terms of the normal error of the 10 objects for the DiLiGenT dataset. For each object, we constructed 11 cosine-weight tensors P corresponding to a single observation tensor I by adding uniform random noise to angles of the surface normal in the polar coordinates. We gradually increased the amplitude of noise μ as $\mu = \{0, 1, 2, \dots, 9, 10\}^\circ$. As shown in Figure 7, the RMS error of reconstruction results is increased as μ increases. Shiny objects tend to be more sensitive to normal errors as reported by Zhou et al. [12].

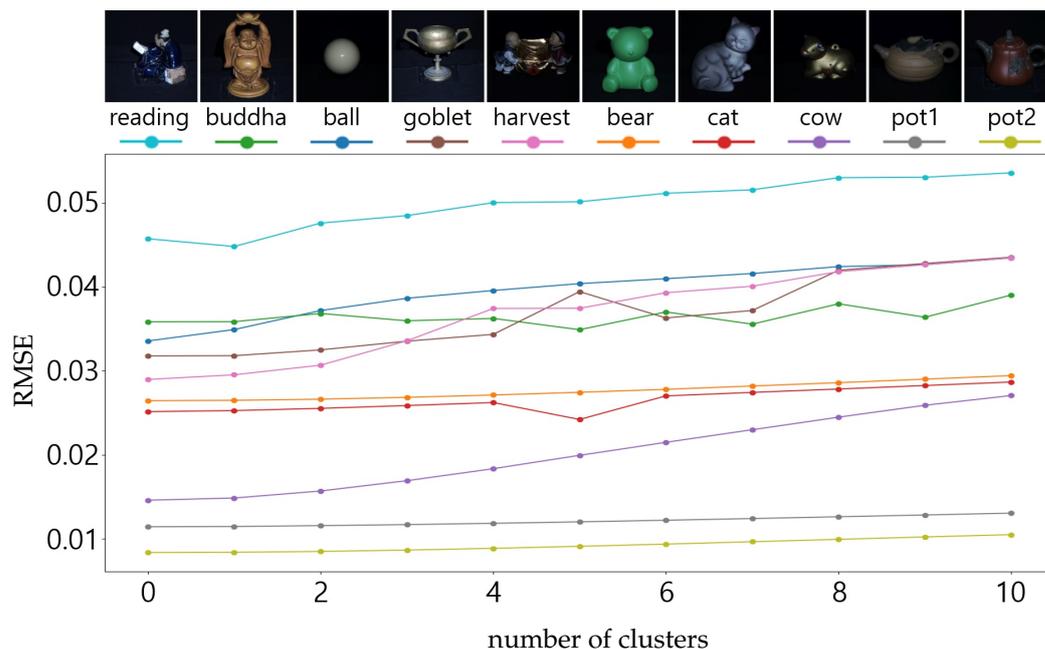


Figure 7. RMS errors in terms of the degree of uniform random noise in the surface normal. Shiny objects (reading, ball, goblet, harvest, and cow) tend to be more affected by normal perturbation. This figure is best viewed in color.

5.6. Selection of n_c

The number of basis functions n_c is the most dominant hyper-parameter of the DEN. Similar to a grid hyperparameter tuning scheme for linear manifold search [55], we empirically select it by a brute-force search [1,2,10,27]. Figure 8 shows the tendency of the RMS errors according to n_c . In our experiments, a small number of basis functions is enough to represent the reflectance of many surface points (300~500 k). The reconstruction quality is insignificantly improved when $n_c > 5$.

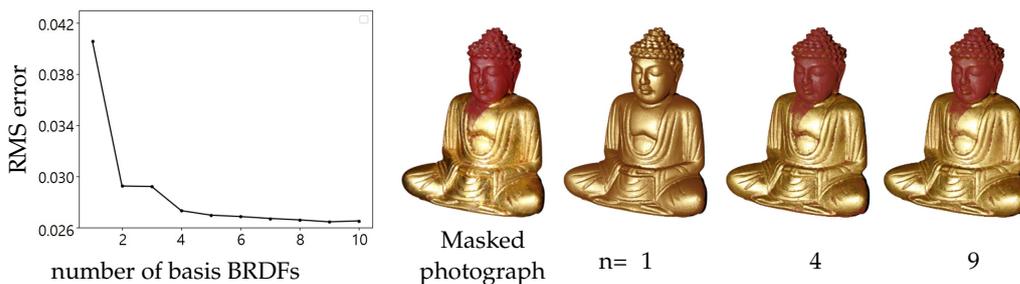


Figure 8. (left) RMS error between reconstructions and photographs according to the number of basis BRDFs. (right) corresponding reconstruction result for the ground photograph.

5.7. Relighting Result

Figure 9 shows the relighting result under various environmental maps. We applied the final weight of the RP layer as the basis NDFs and the final output of the encoder as the weight of the basis NDFs. As shown in Figure 9, our network can generate a high-fidelity reconstruction in given view and illumination conditions.

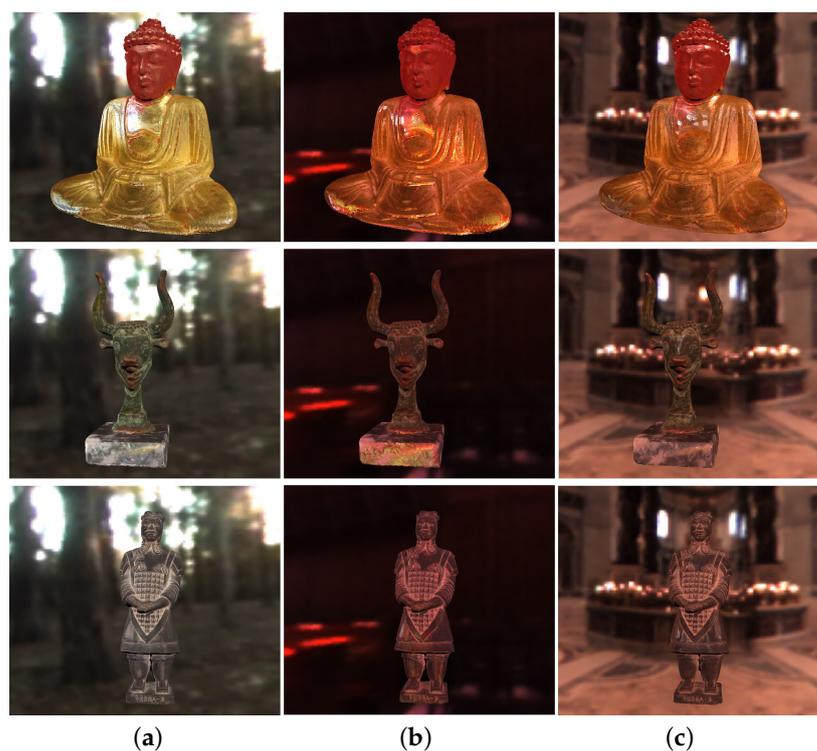


Figure 9. Relighting results of three objects for different environmental maps provided by P. Debevec: (a) Eucalyptus Grove; (b) Grace Cathedral; (c) St. Peter's Basilica.

6. Conclusions

In this paper, we have presented a novel Deep Embedding Network (DEN) for the simultaneous update of the basis NDFs and their linear manifold via the dual gradient (Equation (10)) in order to search for a photo-realistic representation of an opaque object using SVBRDF. We have empirically shown that the basis NDFs and their blending weights are linear-proportionally updated in our autoencoder network. We also demonstrated that our scheme excels in optimizing the rendering loss in terms of the convergence speed and the visual quality compared to the iterative SVBRDF update methods [2,28].

Our method is specialized to model intra-class variations of the surface reflectance. It is advantageous for restoring the high-definition appearance of a 3D object measured in a laboratory environment. However, our method does not take into account a sparse measurement [15,18,43], or unknown illumination conditions [16,37]. In future work, we plan to extend our work to estimate not only the surface reflectance but the surface geometry [56]. This could be achieved by extending our network to jointly estimate the surface normal and the reflectance as similar to Tani et al. [46] or Nam et al. [2]. To achieve the goal, we can directly adopt the CNN-based reflectance estimator [46] and the iterative quadratic solver [2].

Author Contributions: Conceptualization, Y.H.K. and K.H.L.; methodology, Y.H.K.; software, Y.H.K.; validation, K.H.L., and Y.H.K.; formal analysis, K.H.L.; investigation, K.H.L.; resources, Y.H.K.; data curation, Y.H.K.; writing—original draft preparation, Y.H.K.; writing—review and editing, K.H.L.; visualization, Y.H.K.; supervision, K.H.L.; project administration, K.H.L.; funding acquisition, Y.H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2022 (Project Name: The expandable Park Platform technology development and demonstration based on Metaverse-AI convergence, Project Number: R2021040269).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lawrence, J.; Ben-Artzi, A.; DeCoro, C.; Matusik, W.; Pfister, H.; Ramamoorthi, R.; Rusinkiewicz, S. Inverse shade trees for non-parametric material representation and editing. *ACM Trans. Graph.* **2006**, *25*, 735–745. [\[CrossRef\]](#)
2. Nam, G.; Lee, J.H.; Gutierrez, D.; Kim, M.H. Practical SVBRDF acquisition of 3D objects with unstructured flash photography. In Proceedings of the SIGGRAPH Asia 2018 Technical Papers, Hong Kong, China, 19–22 November 2018; p. 267.
3. Lensch, H.P.; Kautz, J.; Goesele, M.; Heidrich, W.; Seidel, H.P. Image-based reconstruction of spatially varying materials. In *Rendering Techniques 2001*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 103–114.
4. Holroyd, M.; Lawrence, J.; Zickler, T. A coaxial optical scanner for synchronous acquisition of 3D geometry and surface reflectance. *ACM Trans. Graph.* **2010**, *29*, 99. [\[CrossRef\]](#)
5. Albert, R.A.; Chan, D.Y.; Goldman, D.B.; O'Brien, J.F. Approximate svBRDF estimation from mobile phone video. In Proceedings of the Eurographics Symposium on Rendering: Experimental Ideas & Implementations, Karlsruhe, Germany, 1–4 July 2018; pp. 11–22.
6. Gardner, A.; Tchou, C.; Hawkins, T.; Debevec, P. Linear light source reflectometry. *ACM Trans. Graph.* **2003**, *22*, 749–758. [\[CrossRef\]](#)
7. Aittala, M.; Weyrich, T.; Lehtinen, J. Practical SVBRDF capture in the frequency domain. *ACM Trans. Graph.* **2013**, *32*, 110–111. [\[CrossRef\]](#)
8. Kang, K.; Chen, Z.; Wang, J.; Zhou, K.; Wu, H. Efficient reflectance capture using an autoencoder. *ACM Trans. Graph.* **2018**, *37*, 127. [\[CrossRef\]](#)
9. Debevec, P.; Tchou, C.; Gardner, A.; Hawkins, T.; Poullis, C.; Stumpf, J.; Jones, A.; Yun, N.; Einarsson, P.; Lundgren, T.; et al. Estimating surface reflectance properties of a complex scene under captured natural illumination. *Cond. Accept. ACM Trans. Graph.* **2004**, *19*, 1–11.
10. Ren, P.; Wang, J.; Snyder, J.; Tong, X.; Guo, B. Pocket reflectometry. *ACM Trans. Graph.* **2011**, *30*, 45. [\[CrossRef\]](#)
11. Dong, Y.; Wang, J.; Tong, X.; Snyder, J.; Lan, Y.; Ben-Ezra, M.; Guo, B. Manifold bootstrapping for SVBRDF capture. *ACM Trans. Graph.* **2010**, *29*, 98. [\[CrossRef\]](#)
12. Zhou, Z.; Chen, G.; Dong, Y.; Wipf, D.; Yu, Y.; Snyder, J.; Tong, X. Sparse-as-possible SVBRDF acquisition. *ACM Trans. Graph.* **2016**, *35*, 189. [\[CrossRef\]](#)
13. Li, X.; Dong, Y.; Peers, P.; Tong, X. Modeling surface appearance from a single photograph using self-augmented convolutional neural networks. *ACM Trans. Graph.* **2017**, *36*, 45. [\[CrossRef\]](#)
14. Ye, W.; Li, X.; Dong, Y.; Peers, P.; Tong, X. Single Image Surface Appearance Modeling with Self-augmented CNNs and Inexact Supervision. *Comput. Graph. Forum* **2018**, *37*, 201–211. [\[CrossRef\]](#)
15. Deschaintre, V.; Aittala, M.; Durand, F.; Drettakis, G.; Bousseau, A. Single-image SVBRDF capture with a rendering-aware deep network. *ACM Trans. Graph.* **2018**, *37*, 128. [\[CrossRef\]](#)
16. Rematas, K.; Ritschel, T.; Fritz, M.; Gavves, E.; Tuytelaars, T. Deep reflectance maps. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 4508–4516.
17. Shi, J.; Dong, Y.; Su, H.; Stella, X.Y. Learning non-lambertian object intrinsics across shapenet categories. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 5844–5853.
18. Li, Z.; Xu, Z.; Ramamoorthi, R.; Sunkavalli, K.; Chandraker, M. Learning to reconstruct shape and spatially-varying reflectance from a single image. In Proceedings of the SIGGRAPH Asia 2018 Technical Papers, Vancouver, BC, Canada, 12–16 August 2018; p. 269.
19. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106. [\[CrossRef\]](#)
20. Wang, Q.; Wang, Z.; Genova, K.; Srinivasan, P.P.; Zhou, H.; Barron, J.T.; Martin-Brualla, R.; Snavely, N.; Funkhouser, T. Ibrnet: Learning multi-view image-based rendering. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4690–4699.
21. Paliokas, I.; Patenidis, A.T.; Mitsopoulou, E.E.; Tsita, C.; Pehlivanides, G.; Karyati, E.; Tsafaras, S.; Stathopoulos, E.A.; Kokkalas, A.; Diplaris, S.; et al. A gamified augmented reality application for digital heritage and tourism. *Appl. Sci.* **2020**, *10*, 7868. [\[CrossRef\]](#)
22. Marto, A.; Gonçalves, A.; Melo, M.; Bessa, M. A survey of multisensory VR and AR applications for cultural heritage. *Comput. Graph.* **2022**, *102*, 426–440. [\[CrossRef\]](#)
23. Boss, M.; Braun, R.; Jampani, V.; Barron, J.T.; Liu, C.; Lensch, H. Nerf: Neural reflectance decomposition from image collections. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 12684–12694.
24. Li, Z.; Sunkavalli, K.; Chandraker, M. Materials for Masses: SVBRDF Acquisition with a Single Mobile Phone Image. *arXiv* **2018**, arXiv:1804.05790.
25. Cook, R.L.; Torrance, K.E. A reflectance model for computer graphics. *ACM Trans. Graph.* **1982**, *1*, 7–24. [\[CrossRef\]](#)
26. Walter, B.; Marschner, S.R.; Li, H.; Torrance, K.E. Microfacet models for refraction through rough surfaces. In Proceedings of the 18th Eurographics conference on Rendering Techniques, Grenoble, France, 25–27 June 2007; pp. 195–206.

27. Alldrin, N.; Zickler, T.; Kriegman, D. Photometric stereo with non-parametric and spatially-varying reflectance. In Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AL, USA, 24–26 June 2008.
28. Chen, G.; Dong, Y.; Peers, P.; Zhang, J.; Tong, X. Reflectance scanning: Estimating shading frame and BRDF with generalized linear light sources. *ACM Trans. Graph.* **2014**, *33*, 117. [[CrossRef](#)]
29. Xie, J.; Girshick, R.; Farhadi, A. Unsupervised deep embedding for clustering analysis. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016; pp. 478–487.
30. Goldman, D.B.; Curless, B.; Hertzmann, A.; Seitz, S.M. Shape and spatially-varying BRDFs from photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1060–1071. [[CrossRef](#)]
31. Dong, Y.; Chen, G.; Peers, P.; Zhang, J.; Tong, X. Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Trans. Graph.* **2014**, *33*, 193. [[CrossRef](#)]
32. Wang, J.; Zhao, S.; Tong, X.; Snyder, J.; Guo, B. Modeling anisotropic surface reflectance with example-based microfacet synthesis. *ACM Trans. Graph.* **2008**, *27*, 41. [[CrossRef](#)]
33. Matusik, W.; Pfister, H.; Brand, M.; McMillan, L. A Data-Driven Reflectance Model. *ACM Trans. Graph.* **2003**, *22*, 759–769. [[CrossRef](#)]
34. Hui, Z.; Sunkavalli, K.; Lee, J.Y.; Hadap, S.; Wang, J.; Sankaranarayanan, A.C. Reflectance capture using univariate sampling of brdfs. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; Volume 2.
35. Hui, Z.; Sankaranarayanan, A. Shape and Spatially-Varying Reflectance Estimation from Virtual Exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2060. [[CrossRef](#)] [[PubMed](#)]
36. Wu, H.; Dorsey, J.; Rushmeier, H. A sparse parametric mixture model for BTF compression, editing and rendering. *Comput. Graph. Forum* **2011**, *30*, 465–473. [[CrossRef](#)]
37. Wu, H.; Wang, Z.; Zhou, K. Simultaneous localization and appearance estimation with a consumer rgb-d camera. *IEEE Trans. Vis. Comput. Graph.* **2016**, *22*, 2012–2023. [[CrossRef](#)]
38. Palma, G.; Callieri, M.; Dellepiane, M.; Scopigno, R. A statistical method for SVBRDF approximation from video sequences in general lighting conditions. *Comput. Graph. Forum* **2012**, *31*, 1491–1500. [[CrossRef](#)]
39. Tang, Y.; Salakhutdinov, R.; Hinton, G. Deep lambertian networks. In Proceedings of the 29th International Conference on Machine Learning, Edinburgh, UK, 26 June–1 July 2012; pp. 1419–1426.
40. Bell, S.; Bala, K.; Snavely, N. Intrinsic images in the wild. *ACM Trans. Graph.* **2014**, *33*, 159. [[CrossRef](#)]
41. Zhou, T.; Krahenbuhl, P.; Efros, A.A. Learning data-driven reflectance priors for intrinsic image decomposition. In Proceedings of the IEEE International Conference on Computer Vision, Washington, DC, USA, 7–13 December 2015; pp. 3469–3477.
42. Narihira, T.; Maire, M.; Yu, S.X. Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In Proceedings of the International Conference on Computer Vision, Boston, MA, USA, 8–10 June 2015; p. 2992.
43. Aittala, M.; Aila, T.; Lehtinen, J. Reflectance modeling by neural texture synthesis. *ACM Trans. Graph.* **2016**, *35*, 65. [[CrossRef](#)]
44. Kulkarni, T.D.; Whitney, W.F.; Kohli, P.; Tenenbaum, J. Deep convolutional inverse graphics network. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 2539–2547.
45. Tewari, A.; Zollhofer, M.; Kim, H.; Garrido, P.; Bernard, F.; Perez, P.; Theobalt, C. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3715–3724.
46. Tani, T.; Maehara, T. Neural inverse rendering for general reflectance photometric stereo. In Proceedings of the International Conference on Machine Learning, Stockholm, Sweden, 10–15 July 2018; pp. 4864–4873.
47. Kautz, J.; McCool, M.D. Interactive rendering with arbitrary BRDFs using separable approximations. In *Rendering Techniques' 99*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 247–260.
48. Ashikmin, M.; Premoze, S.; Shirley, P. A microfacet-based BRDF generator. In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, LA, USA, 23–28 July 2000; pp. 65–74.
49. Schlick, C. An inexpensive BRDF model for physically-based rendering. *Comput. Graph. Forum* **1994**, *13*, 233–246. [[CrossRef](#)]
50. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
51. Schwartz, C.; Weinmann, M.; Ruiters, R.; Klein, R. Integrated High-Quality Acquisition of Geometry and Appearance for Cultural Heritage. *Proc. VAST* **2011**, *2011*, 25–32.
52. Sheffer, A.; Lévy, B.; Mogilnitsky, M.; Bogomyakov, A. ABF++: Fast and robust angle based flattening. *ACM Trans. Graph.* **2005**, *24*, 311–330. [[CrossRef](#)]
53. Nielsen, J.B.; Jensen, H.W.; Ramamoorthi, R. On optimal, minimal BRDF sampling for reflectance acquisition. *ACM Trans. Graph.* **2015**, *34*, 186. [[CrossRef](#)]
54. Shi, B.; Wu, Z.; Mo, Z.; Duan, D.; Yeung, S.K.; Tan, P. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NE, USA, 27–30 June 2016; pp. 3707–3716.

-
55. Ngoc, T.T.; Le Van Dai, C.M.T.; Thuyen, C.M. Support vector regression based on grid search method of hyperparameters for load forecasting. *Acta Polytech. Hung.* **2021**, *18*, 143–158. [[CrossRef](#)]
 56. Molnar, A.; Lovas, I.; Domozi, Z. Practical Application Possibilities for 3D Models Using Low-resolution Thermal Images. *Acta Polytech. Hung.* **2021**, *18*, 199–212. [[CrossRef](#)]