EUROGRAPHICS Workshop on Graphics and Cultural Heritage (2021)
A. Chalmers and V. Hulusic (Editors)

Exploiting Neighboring Pixels Similarity for Effective SV-BRDF Reconstruction from Sparse MLICs

R. Pintus, M. Ahsan, F. Marton, and E. Gobbetti

CRS4, Italy

Abstract

We present a practical solution to create a relightable model from Multi-light Image Collections (MLICs) acquired using standard acquisition pipelines. The approach targets the difficult but very common situation in which the optical behavior of a flat, but visually and geometrically rich object, such as a painting or a bas relief, is measured using a fixed camera taking few images with a different local illumination. By exploiting information from neighboring pixels through a carefully crafted weighting and regularization scheme, we are able to efficiently infer subtle per-pixel analytical Bidirectional Reflectance Distribution Functions (BRDFs) representations from few per-pixel samples. The method is qualitatively and quantitatively evaluated on both synthetic data and real paintings in the scope of image-based relighting applications.

CCS Concepts

• ***Computing methodologies*** → ***Appearance and texture representations; Reflectance modeling;*** *Scene understanding;*

1. Introduction

A Multi-Light Image Collection (MLIC) is a series of photographs of an object taken from a fixed point of view while changing the lighting condition. They are a powerful source of information on the state of an object, that has found a variety of application in many domains, ranging from Cultural Heritage, natural science, industry, underwater investigation, medical imaging and many more [PDC*19].

The most common use case is the measuring and inspection of objects that have a preferential viewing direction from which the overall depth variation is very small, such as a painting or a bas-relief. Such a globally planar shape is combined with a locally complex geometry at various scales, e.g., variations in roughness or curvature, and a rich optical behavior, with many subtle local variations due to the combination of original material (e.g., brush strokes for a painting) with aging effects. Many practical and affordable acquisition protocols and solutions [CHI19, Mac15, PCS18, GCD*18] have been targeting this use case. In these approaches, objects are measured using a fixed camera position, taking a limited number of high-resolution images with different local illumination from point lights, using a variety of setups targeting both professional and casual users [PDC*19]. The resulting data is then fit to a compact model, which is exploited by interactive visual inspection tools [CHI19, KUL19, JAP*21] to support virtual relighting. The widespread application of this single-view workflow is not only due to the large diffusion of appropriate objects and to the simplicity of the acquisition protocol, but also to the fact that relighting viewers naturally support the analysis of fine surface details with methods resembling the classical physical inspection raking light sources. Moreover, the restriction of camera motion to panning and zooming removes one of the main difficulties of 3D exploration applications, reducing learning curves [JH13].

While classic virtual inspection solutions were restricted to exploiting low-frequency analytical relighting representations, such as PTM or HSH [PDC*19], recent work started targeting physically based rendering from decoupled geometry and appearance representations in the form of spatially varying normal and Bidirectional Reflectance Distribution Functions (BRDF) maps that contain the parameters of an analytical model [JAP*21]. Producing a compact per-pixel normal and analytical BRDF representation from sparsely sampled data is very appealing, since it can be easily distributed, produces a physically reasonable result, and allows for natural integration with standard high-quality and real-time rendering solutions. However, while normal estimation is a well-studied subject [SWM*16], the per-pixel extraction of Spatially Varying BRDF (SV-BRDF) parameters from the small number of samples typically available in sparsely sampled single-view MLICs leads to an under-determined problem [GGG*16, LBFS21]. For this reason, the application of standard per-pixel fitting produces very noisy maps, and the available alternative solutions try to solve the problem either by analyzing the entire object, looking for large normal variation and presence of similar materials, or by deriving extra knowledge from available training sets tuned for the target object kind (Sec. 2).

In this work, we present a practical solution that integrates well with standard local fitting pipelines (Sec. 3). Similarly to previous work (see Sec. 2), we solve for BRDF parameters by minimizing

DOI: 10.2312/gch.20211412

a weighted sum of squared errors between the measurements and an analytical model. However, instead of computing each pixel independently, we exploit information coming from a small neighborhood around the currently computed pixel, and we use weights not only to encode our confidence in the individual measure, but also to include the information on the spatial distance from the central pixel, and the similarity in optical behavior with respect to the computed center. This allows us to expand the angular sampling, as local curvature and roughness variations will modify the pixel's normal, increasing the ability to recover high-frequency information, e.g., in specular areas. Moreover, our bilateral weighting, based on pixel distances and measured values, achieves a non-linear, edge-preserving, and noise-reducing smoothing filter. The assignment of a higher weight to close-by pixels of the same material (e.g., neighboring pixels from the same brush stroke), and a lower weight for far pixels of different optical behavior, is obtained by computing distances of simple pixel descriptors, a much simpler problem than the material classification required by methods dealing with material databases. A regularization term is, in addition, included, to drive the solution to lower-frequency behaviors in case of severely missing data. As a result, our solver can directly replace the fitting module inside all standard per-pixel BRDF fitting pipelines that estimate the parameters in parallel for each pixels. It generates, by construction, relightable models which recover specular information where sufficient data is locally available and fall back to smooth regularized solutions without unwanted high-frequency artifacts in other situations. Finally, the method was qualitatively and quantitatively evaluated on both synthetic data and real objects (paintings) in the scope of image-based relighting for the difficult case of multi-material painted surfaces with or without reflective coating (Sec. 4).

2. Related Work

MLIC acquisition and processing, BRDF fitting, and relightable image modeling and visualization are vast and very active areas, and we refer the reader to established surveys for a general coverage [PDC*19, GGG*16, LBFS21]. In the following, we briefly cover only the approaches most closely related to ours.

Relightable images. This class of methods directly approximates the reflectance signal with an analytical formulation that provides the mapping from lighting parameters to final renderable values, without explicitly separating shape and material information. The seminal approach [MGW01], called Polynomial Texture Mapping (PTM), stores per-pixel coefficients of the second-order bi-quadratic polynomial that best fits the color variations of the pixel as a function of the incident light direction. Different methods try to increase the quality of the final result by changing the polynomial formulation [ZD14], or by improving the fitting algorithm with robust metrics [DHOMH12, PGP*17]. Rather than using simple polynomials, other methods propose a Hemi-Spherical Harmonics (HSH) based models, which are known to work well to represent functions on the surface of a sphere [BJK07, ERF11], or a Discrete Modal Decomposition (DMD) [PLGM*17]. Their compactness and low complexity makes these techniques suitable for fast interactive relighting in local and remote visualization. For this reason, PTM and HSH are the de-facto standard format for relighting applications from MLIC data. Without extra information, however, these methods are limited to model only low-frequency behavior [DHOMH12]. A fundamental limitation of basic relightable image models is the lack of decoupling between shape and material components, which limits shading manipulation and makes it difficult to integrate them in full-fledged rendering frameworks [PDC*19]

SV-BRDF fitting. A number of methods extract a geometric model from the MLIC, e.g., through photometric stereo, and associate it to a material model, in particular in the form of a SV-BRDF [GGG*16]. The nature of common MLIC data, i.e., fixed viewpoint and changing light directions, makes modeling SV-BRDF fields very hard, since the measured per-pixel appearance profile is a very sparse sampling of the high dimensional BRDF. For this reason, normal and BRDF estimation is most commonly applied in hybrid setups, using, e.g., more viewpoints and additional instrument to measure coarse shape for geometry bootstrapping [XR20]. Pure MLIC-based methods try to improve the SV-BRDF reconstruction by defining some constraints, or by augmenting the material data at each pixel location. Several solutions assume that the acquired object have a single BRDF or, conversely, have multiple BRDFs placed on a perfectly planar surface (constant normal map) [AWL13]. These methods are not applicable in the general case of multi-material objects with geometric features. A common strategy is to define the per-pixel BRDF as a weighted sum of few, unknown reference BRDFs [LBAD*06], to build a known BRDF dictionary and to model the material at each location as a position in the non-negative span of the dictionary [HS16], or to extract base materials through a global segmentation and subsequent clustering of appearance profiles [TGVG12]. These techniques require solving a non-trivial global material classification problem, which is especially difficult in the presence of large appearance variations, e.g., due to mixtures. The non-local nature of these methods makes an efficient out-of-core and parallel implementation difficult.

Learned priors. Recent methods try to bypass the formulation of an analytical model for implicit relightable models computation or explicit SV-BRDF fitting by building neural networks that learn to perform the modeling by observing large amount of relighting training examples [XSHR18, RJGW19, RDL*15]. The great advantage of those techniques is that they can model effects such as interreflections or cast shadows, and complex isotropic materials, with relatively small number of images in a MLIC. However, neural networks depend a lot on the training set, and they tend to produce artifacts and hallucinations when used within a general context.

3. Method

Our solution is integrated in standard MLIC processing pipelines that compute per pixel surface characteristics from a the set of images. Starting from the set of input images, the per pixel view- and lighting parameters are computed using standard techniques, resulting in the knowledge of per-pixel view direction, light direction, light color, and light intensity. Any calibration pipeline can be applied (for this paper, we used the recently introduced method by Pintus et al. [PJZ*21]). The initial data in MLIC is also trimmed, to remove under-exposed, over-exposed, and shadow areas, resulting in a variable-size per-pixel appearance profile. From this initial calibration and pruning pipeline, the normal map of the imaged

object is retrieved, using any of the available photometric stereo techniques [PGP*17]. Starting from this information, our method produces, as output, the BRDF parameters of each pixel. Although the method relies on the information gathered from the appearance profiles of the pixels in a local neighborhood, the computation of the BRDF of a pixel is completely independent from the computation of others, resulting in a highly parallelizable problem that requires to maintain in-core only the data from the local neighborhood of the currently computed pixels. For these reasons, in the next sections, we will explain the proposed technique by focusing only on the calculation performed on a single pixel. Finally, our method belongs to the class of BRDF reconstruction techniques that consider a local illumination model, and does not take into account other more complex (and global) optical effects, e.g., inter-reflections, sub-surface scattering, etc.

3.1. Problem definition

A pixel's BRDF is computed locally by exploiting not only its appearance profile, but also the appearance profiles collected from its neighborhood. The main idea is that there is a high chance that a pixel in the neighborhood has the same material of the central pixel, but different normal; so it might add new sampling values in the sparsely measured BRDF space, thus increasing the robustness (and quality) of the BRDF computation. The extent of the contribution of each measurement in the neighborhood is then evaluated on the basis of a weighting strategy that reflects the probability that a neighbor has the same material as the central pixel. The input data and the weights are the known values of the following objective function that needs to be minimized:

$$\underset{\Pi}{argmin} \sum_{\Omega} \sum_{\Delta} \left(w^2_{\omega,\delta} \left\| m_{\omega,\delta} - f_r\left(\Pi, N_\omega, V_\omega, L_{\omega,\delta}\right) \right\|^2 \right) + \\ + \lambda R^2_\Omega \tag{1}$$

The function $f_r(\dots)$ is the BRDF of the analyzed pixel. The typical analytic representation of the BRDF is in terms of the incoming light direction L, the outgoing direction V, and the spatial position x. Here we remove the spatial term since we are considering a single location (i.e., the single pixel), and we want to highlight the difference between the known and unknown terms. For this reason we distinguish the known term of the BRDF, i.e. N_ω, V_ω, and $L_{\omega,\delta}$, with the unknown Π, which represents one or more unknown parameters. The number of parameters depends on the analytic representation that we choose for the BRDF. For now the eq. 1 is in its general form, and says only that the BRDF is analytic (through the definition of Π), but does not define which type of representation we use. The neighborhood is represented by Ω, while the set of different lighting conditions in the MLIC are expressed with Δ. Hence, the normal N and the view V depends only on the spatial position of different pixels in the neighborhood Ω, and they remain fixed for the same pixel while the light changes (fixed camera condition). Conversely, the light depends both on Ω and Δ, as well as the measurements $m \in M$; M is the set of all the measurements across Ω and Δ. Due to occlusions, cast shadows, and other filtering preprocessing steps, the dimension of each appearance profile may vary across the MLIC, so it is worth to note that the Δ set is a function of ω; we omit this dependence for simplicity. As we will see in later sections, the weights w not only give a confidence value about material similarity between central pixels and the neighbors, but they also assist the fitting by compensating/equalizing the contribution of different orders of magnitude in the input measurements (i.e., measurement compression). Finally, the term λR^2_Ω is a regularization term that controls the fitting behavior in under-determined cases to avoid disturbing visual artifacts. For instance, if a BRDF is sampled only in the diffuse zone, the system without this regularization term would be free to create any BRDF with an arbitrary and possibly high specular signal. To take this situation under control, we introduce a single virtual BRDF measurement in a highly specular direction (i.e., for V_ω and $L_{\omega,\delta}$ collinear to N_ω) with a value set to the largest measured value for that pixel among all sampled light directions. As the weight λ is very low, this underestimated value only affects the solutions when no measured data is available for those angles.

3.2. Weighting Strategy

The extracted neighborhood Ω, which is the domain of our computation, contains a lot of information that we want to exploit as much as possible in order to increase the confidence in the computation of the BRDF of the central pixel.

Radial weight. The first consideration is that a pixel closer to the center has the highest probability of being of the same material as the central pixel than a pixel far away from it. Hence, we define a radial weighting function as:

$$w^{rad}_\omega = 1 - \frac{r^2_\omega}{R^2}\ , \tag{2}$$

where r_ω is the distance between the pixel ω in the neighborhood and the central pixel, while R is the radius of the neighborhood. This function is equal to 1 at the central pixel and decreases to 0 at the boundary of the neighborhood. The weight plays the same role as the spatial closeness factors in bilateral filtering, and is the same for all the measures in the appearance profile of pixel ω.

Similarity weight. The similarity weight must assign a value to all the measurements from a neighboring pixels proportional to the likelihood that they represent the same material as the central pixel. While the general problem of material identification from sparsely sampled values is very complex, we can use here very simplified solutions, as we are concentrating only in a small neighborhood, where we can expect that close-by pixels that roughly exhibit the same behavior across the various angles are very likely to come from the same material (e.g., because they are part of the same brush stroke). Despite this simplification, the problem is still not trivial, since we need to infer this similarity from a small number of samples. Moreover, since we are looking to enhance the representation by fusing surface areas with different normals, these sparse samples will be taken at different angles. In order to compute the similarity, we transform each pixel's sparse sample into a feature vector F that provides a compact regularized representation that simplifies distance computation. This representation consists in a fixed amount of B bins that contain the maximum sampled color for each region of the Θ_h angle in the Rusinkiewicz parameterization [Rus98]. Θ_h is a function of the angle between the half vector H and the normal N, θ_h. Since smaller variations are expected for high values of Θ_h, corresponding to the diffuse areas of the BRDF, the bins are not uniformly distributed, but have a size proportional to $\sqrt[3]{\Theta_h}$. Note,

moreover, that, due to the sparsity of sampling some of the bins in the feature vector might remain empty after accumulation.

Given two feature vectors F^0 for the central pixel and F^ω for a pixel in the neighborhood, we compute the similarity based on a distance metric $d(...)$ between the two features:

$$w_\omega^{sim} = 1 - \min\left(1, d\left(F^\omega, F^0\right)\right) \tag{3}$$

Clearly, this metric must compare these two features taking into account the overlap of bins. If there are no common bins, which can happen if the normals are wildly different, we have no information on how to compute the distance, and, thus, we return the maximum distance. Similarly, the maximum distance is returned if some overlapping bins have large chromaticity or intensity differences, as the materials likely behave in a very different way for similar view and light configurations. In all other cases, we return a distance proportional to the relative difference of spectral values, using the log2 metric of Sun et al. [SJR18]. To sum up the distance will be:

$$d\left(F^0, F^\omega\right) = \begin{cases} 1, & \text{if } \sum_{i=0}^{B-1} b_i = 0 \qquad \text{Overlap} \\ 1, & \text{if } \min_i \frac{F_i^0}{\|F_i^0\|} \cdot \frac{F_i^\omega}{\|F_i^\omega\|} \leq \cos(\chi) \qquad \text{Chroma} \\ 1, & \text{if } \max_i \log^2\left(\frac{\|F_i^0\|+\varepsilon}{\|F_i^\omega\|+\varepsilon}\right) \geq \rho \qquad \text{Magnitude} \\ \frac{1}{\rho \sum_{i=0}^{B-1} b_i} \sum_{i=0}^{B-1} \log^2\left(\frac{\|F_i^0\|+\varepsilon}{\|F_i^\omega\|+\varepsilon}\right), & \text{otherwise} \end{cases} \tag{4}$$

The two thresholds that determine whether individual bins are very dissimilar are selected so as to let only very similar neighbors to contribute to the BRDF estimation of the central pixel; the angle χ is 5 degrees, while $\rho = \log^2(1.1)$, which means maximum allowed relative magnitude deviation of 10%. Note that, since we compare only the overlapping regions, smooth variations of normals make it possible to include more and more information in the BRDF coming from neighboring pixels, allowing for the recovery of non-diffuse behaviors due to local roughness or shape variations.

Compression weight. Finally, it has been demonstrated [LBFS21] that BRDF fitting benefits from two other data manipulations, i.e., data clamping and measurement compression. A simple data clamping strategy is to discard measurements that are associated with raking angles (typically bigger than $\alpha = 80$ degrees [NDM05]). On the other hand, compression applies a non-linear transformation to the measurement m and the evaluated BRDF $f(...)$ before computing the fitting error, to attenuate peak values. Instead of applying the standard cube-root transformation [LBFS21], we apply a weight equal to $w_{\omega,\delta}^{comp} = m_{\omega,\delta}^{-2/3}$, which applies the same compression as cube-root to the measured value. As we will see below, avoiding the introduction of a nonlinearity allows us to exploit optimized solutions for weighted least squares fitting.

Taking into account all these aspects, the final weight for a measurement becomes:

$$w_{\omega,\delta} = \begin{cases} 0 & \text{if } N_\omega \cdot L_{\omega,\delta} < \cos(\alpha) \\ w_{\omega,\delta}^{comp} w_\omega^{rad} w_\omega^{sim} & \text{otherwise} \end{cases} \tag{5}$$

This weighting technique implements an edge-preserving bilateral filtering strategy that maximizes the amount of information useful for estimating the BRDF with a denser angular sampling than the single pixel approach.

3.3. BRDF representation and optimization approach

While our method is in principle applicable to any parametric BRDF, for efficiency of representation we specialize our solver for a BRDF that consists in a sum of terms (or BRDF components), each of which is a multiplication between a multi-dimensional spectral value and a scalar function, i.e.:

$$f_r(\Pi, N, V, L) = \sum_{\gamma=0}^{\Gamma-1} K_\gamma^S f_\gamma(\Pi_\gamma, N, V, L) \ . \tag{6}$$

The term K^S represents the spectral value of the BRDF component, and it is a multi-spectral color vector of dimentionality S; for an RGB signal $K^3 = \left(k^R, k^G, k^B\right)$. The function $f(\dots)$ is the component BRDF; for instance, for a Lambertian component of the material $f(\dots) = const = 1/\pi$. Π_γ is the set of parameters of a single BRDF component (e.g., it is an empty set for a Lambertian BRDF, while it is a single parameter set for the classic Ward model), while Π is the union of all the Π_γ and K_γ^S sets for $\gamma = \{0, 1, \dots, \Gamma-1\}$. The number of parameters in this formulation is:

$$\#Params = 6 + S + |\Pi| = 6 + S + S \cdot \Gamma + \sum_{\gamma=0}^{\Gamma-1} |\Pi_\gamma| \ , \tag{7}$$

where the operator $|\cdot|$ stands for the cardinality of the set. The number $6+S$ is the sum of the two values for the normal, view, and light directions, and the S spectral values of the light intensity. Given this analytic formulation, the equation 1 of the objective function becomes:

$$\underset{\Pi}{argmin} \sum_\Omega \sum_\Delta \left(w_{\omega,\delta}^2 \left\| m_{\omega,\delta} - \sum_{\gamma=0}^{\Gamma-1} K_\gamma^S f_\gamma\left(\Pi_\gamma, N_\omega, V_\omega, L_{\omega,\delta}\right) \right\|^2 \right) + \\ + \lambda R_\Omega^2 \tag{8}$$

Using this formulation is convenient for two main reasons. First, it is so general that is capable of handling a large range of phenomenological and physically-based BRDF models [GGG*16]. Moreover, in combination with our linear weighting solutions, it makes it possible to express the complex non-linear problem into two sub-problems, i.e., one simple linear problem for finding the spectral values embedded within a non-linear problem with a decreased dimensionality for finding the shaping parameters. In this paper, in particular, we test the proposed framework with the analytic BRDF f set to Duer's variant of the original isotropic Ward model [Dü06]:

$$f_r(\alpha, N, V, L) = K_d^S f_d + K_s^S f_s(\alpha, N, V, L) = \tag{9}$$

$$= K_d^S \frac{1}{\pi} + K_s^S \frac{1}{4\pi\alpha^2 \sqrt{(N \cdot L)(N \cdot V)}} e^{-\frac{(H \cdot X)^2 + (H \cdot Y)^2}{\alpha^2 (H \cdot N)^2}} \ , \tag{10}$$

where K_d^S and K_s^S are respectively the diffuse and specular color, f_d is the constant Lambertian BRDF, while f_s is Ward's specular term. The parameter α drives the material roughness. In this case the non-linear search is one-dimensional.

Equation 1 depicts a general non-linear optimization with a search

space dimension equal to $|\Pi|$. Conversely, the formulation in equation 8 is expressed in terms of a weighted linear least squares problem of size $S \cdot \Gamma$ for finding the spectral values once the shaping parameters are known, embedded within a non-linear problem with $\sum_{i=0}^{\Gamma-1} |\Pi_\gamma|$ unknowns for finding the shaping parameters; the latter is typically $\ll |\Pi|$, especially when dealing with multi-spectral acquisitions. A similar approach was taken by Ngan et al. [NDM05]. In our current implementation, we use the locally-biased DIRECT global optimization algorithm for the nonlinear search [Jon01] and a SVD solver for the linear least squares subproblem.

4. Results

We validate the proposed solution by analyzing its performance in the reconstruction of SV-BRDFs from sparsely sampled MLICs with a variable number of images. The comparison is made with respect to a standard state-of-the-art single pixel approach that employs the same SV-BRDF model and fitting strategy, but relies on samples coming only from the single pixel without considering its neighborhood. For all our tests, we use RGB images and the isotropic Ward analytic BRDF; so we have to solve a seven dimensional optimization problem, i.e., three unknown parameters for the diffuse and specular colors, and one for the gloss. We use a regularization weight $\lambda = 10^{-4}$ and a neighborhood radius of 10.

The main goal of our evaluation is to show that starting from a small number of images (a sparse BRDF sampling) it is possible to increase the quality of the final reconstruction and relighting of challenging flat, visually/geometrically rich objects. This aspect has immediate practical importance. First of all, an extremely dense MLIC capture is costly and very rarely employed in daily work scenarios. Moreover, even with lots of images, the a single-view MLIC acquisition intrinsically provides an undersampled set of BRDF measurements, due to the fixed view point.

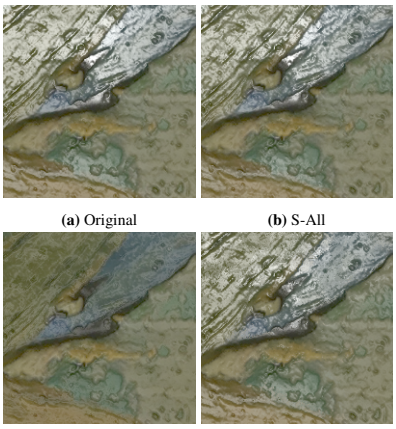
In the following, we first present a quantitative analysis of the quality achievable when reconstructing BRDFs (Sec. 4.1) and then report the results of a user test, which provides a perceptual human feedback for the visualization quality of the proposed method for relighting applications (Sec. 4.2).

4.1. Quantitative evaluation of reconstructions from sparse MLICs

We performed an evaluation of reconstruction quality on both synthetic and real-world datasets. For each MLIC dataset (synthetic or real) we apply the same testing procedure. We first take the entire MLIC with the total number of images (52 for this paper) and we compute the SV-BRDF with the single pixel algorithm; we will consider this as the reference result (we call it *S-All*). Then, we consider different subsets of the MLIC, by removing more and more images (from one to twelve images), and for each subset we compute the SV-BRDF with the single pixel algorithm and with our proposed method; we respectively call *S-X* and *K-X* the single pixel or the k-neighbor based computation when the subset is obtained by removing X images from the entire MLIC. Then, we use each computed SV-BRDF to simulate a virtual relighting of all the images in the MLIC, by using the same lighting condition, and we compare each pair real vs virtual image with a perceptual metric that quantifies image quality reproduction; we use the Structural Similarity Index (SSIM) [WBSS04] for that purpose. As a final value for the quality of the computed SV-BRDF, we take the worst SSIM value among all the real vs virtual image comparison in the MLIC. In addition, we also present for each test a visual comparison between the relightings obtained by the *S-All*, the *S-X*, and our *K-X* approach.

Synthetic tests. Synthetic tests on rendered models make it possible to evaluate the behavior of the methods in a fully controlled case where ground truth is available. We selected two synthetic models (*Paint-Texture-16* and the *Paint-Texture-14*) from the *EveryTexture* database [Tex21], since they exhibit a detailed shape and appearance similar to the type of real objects we are interested in. The data is provided through *Diffuse*, *Bump*, and *Normal* maps. These maps were used to create a synthetic MLICs using a fixedd camera and 52 directional lights. Fig. 1 shows one original image from the synthetic MLIC of the *Paint-Texture-16* (Fig. 1a), together with the same image virtually relighted from the SV-BRDF computed with the entire MLIC and the single pixel algorithm *S-All* (Fig. 1b). In the bottom row of the same figure, we compare the virtual relighting after computing the SV-BRDF by removing the image in Fig. 1a from the MLIC, and by applying both the single pixel (*S-01*) and our (*K-01*) solution. Even with this slightly sparser input (one less image), it is clear how the information from the neighbor pixels helps to retrieve a better surface optical response, so that the virtual relighted image exhibits a similar level of specular reflection as the original one (Fig. 1d). Moreover, our edge-preserving strategy properly keeps the sharpness of the original image. Conversely, the single pixel algorithm produces an almost diffuse image (Fig. 1c), without any gloss component. A similar behaviour can be seen in the second synthetic dataset (Fig. 2). The highlight signal in the blue part is completely lost with the standard procedure (*S-01*, Fig. 2c), while it is largely recovered by our approach (*K-01*, Fig. 2d). In order to quantitatively measure the improvement of our algorithm compared to the standard single pixel technique, Fig. 3 shows the SSIM statistics when we remove one to twelve images from the original 52 image MLIC. It can be seen that removing images deteriorates the quality of the optical characterization, but we can clearly see how our solution can provide a better SV-BRDF reconstruction than the single pixel approach, especially for the sparser models.

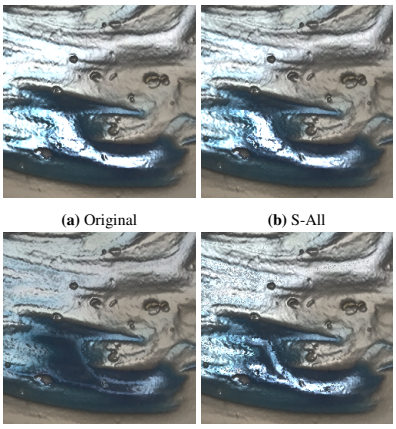
Real-world scenes. In order to test our solution in a real-world scenario, we consider six painting mockups with heterogeneous spatially-varying material distribution over the surface. Mockups were realized on painting paper, with standard acrylic colors. Our aim was to have different kind of color mixtures and geometries. Regarding color mixtures we ranged from fully fresh mixed colors to fully separated color layers. Regarding the geometry, we tested different configurations, ranging from a thin layer of flat color to a typical brush texture created with a brush filled with a good quantity of color, up to even stronger geometry features obtained depositing the color directly from the tube; in general, the maximum depth range of the mockups is in the order of millimeters. Finally, when all the color layers were completely dry, we covered half of each mockup with a thin coating of gloss varnish, in order to have both a quite diffusive surface and a very shiny one. MLICs for these mockups have been acquired by a custom light dome with a radius



(a) Original **(b)** S-All

(c) S-01 **(d)** K-01

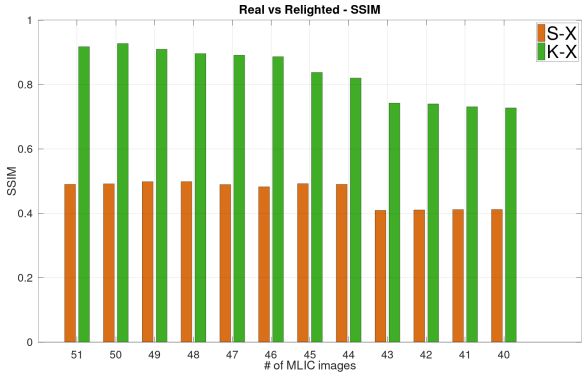
Figure 1: *We compare the original image (a) with its virtual relighting obtained from the SV-BRDF computed with (b)* S-All*, (c)* S-01*, and (d)* K-01*. For the last two we removed from the MLIC the image in (a). It is clear how the proposed solution exploits neighbor pixels to retrieve the glossy signal, which is completely lost in the standard single pixel strategy.*



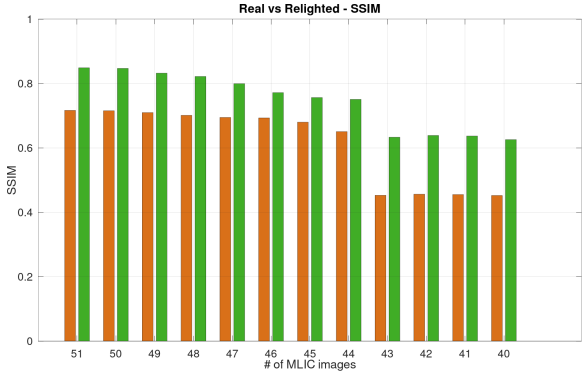
(a) Original **(b)** S-All

(c) S-01 **(d)** K-01

Figure 2: *We compare the original image (a) with its virtual relighting obtained from the SV-BRDF computed with (b)* S-All*, (c)* S-01*, and (d)* K-01*. For the last two we removed from the MLIC the image in (a). The highlight in the blue part is completely lost by using the standard single pixel strategy, while it is largely recovered by our solution.*



(a) Paint-Texture-16



(b) Paint-Texture-14

Figure 3: *We use SSIM metric to compare the performance of the single pixel strategy* S-X *with our k-neighbor based solution* K-X*. The graphs show the worst SSIM value in the relighted MLIC obtained from SV-BRDFs computed after removing from one to twelve images from the original* 52 *image MLIC. The more images are removed, the worse is the relighting quality. However, our solution can always provide better performances than the state-of-the-art single pixel approach.*

of about 30*cm*, and with 52 LED lights. The LEDs are neutral white lights that cover the entire visible spectrum. The capture device is a 36.3 Mpixels DSLR FX Nikon D810 Camera with a 50*mm* AF Nikkor Lens. The acquisition system has been calibrated with four glossy spheres (for light direction), and with a *Spectralon* target by using a flat field light intensity calibration technique. As we did for synthetic datasets, we first visually compare one original image and with the virtually relighted ones, using the *S-All*, *S-01*, and *K-01* approaches. Again, for the last two we removed from the MLIC the shown original image. Fig. 4 shows that the outcomes of the real experiment confirm what we have found in the synthetic case. The removal of even one original images has a high impact on the relighting result in the case of the standard single pixel algorithm. Conversely, our approach is capable of retrieving the information lost by a sparse sampling of the surface appearance by looking at neighbor pixels; *K-01* results contain many of the highlights present in the original photograph, while the *S-01* strategy results in a more diffuse surface. In particular, in the first column of Fig. 5, we show one original photograph from the MLIC (Up) and the same photo virtually relighted from the SV-BRDF computed with the *S-All* method (Bottom). The other columns show the relighting results obtained with the *S-X* (Up) and *K-X* (Bottom) method, by respectively removing (from left to right) one, five, eleven, and twelve images. The first image we remove (second column) is exactly the same original image we show in the first column (Up); this is the worst case, since we are removing exactly the information from the light direction we are using to compute the relighted image. The other images are removed from front to raking light, in order to remove first the data that are more statistically correlated to the BRDF glossy region (raking images typically have more diffuse signal). We can see how, even with twelve discarded images, our approach keeps the surface glossiness, while the single pixel method almost completely loses it after the removal of just one image. As before, for each of those mockups, we report the SSIM statistics in Fig. 6. Now, let's suppose we have a central and a neighbor pixel that are both sampled in the diffuse part of the BRDF domain, and only the neighbor is sampled

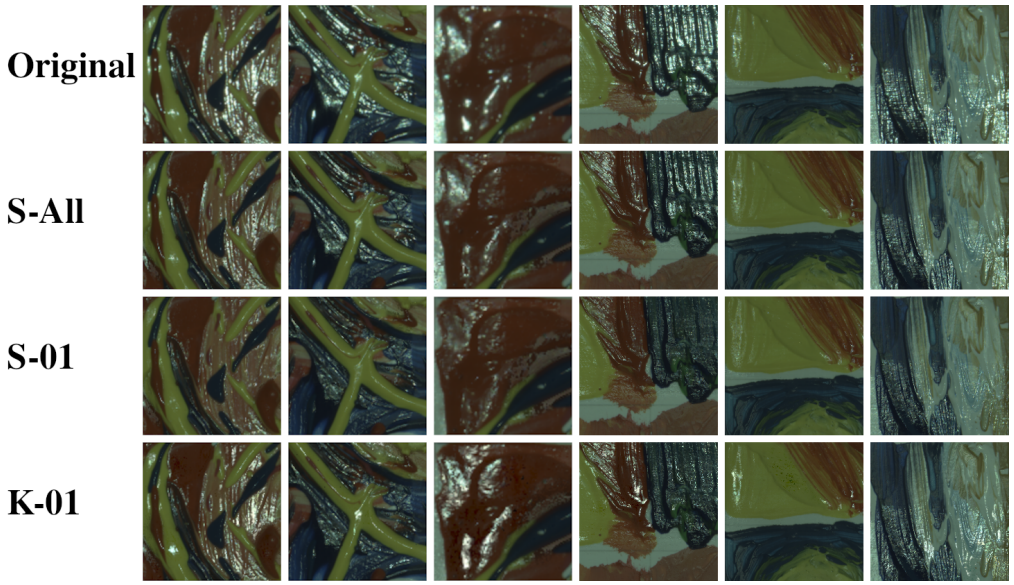


Figure 4: *We present six different datasets (painting mockups), and we visually compare one original image from the MLIC with the results of a virtual relighting done with the same lighting condition and with the SVBRDF computed by different algorithms. For each column: original image of the mockup (first row); single pixel algorithm applied to the entire MLIC (S-All); single pixel algorithm applied to the same MLIC without the image in the first row (S-01); our algorithm applied to the same MLIC subset used for the third row (K-01). We can see how the proposed algorithm exploits the information in the neighborhood region to recover most of the highlights that are lost in the single pixel approach, while preserving the original image sharpness.*

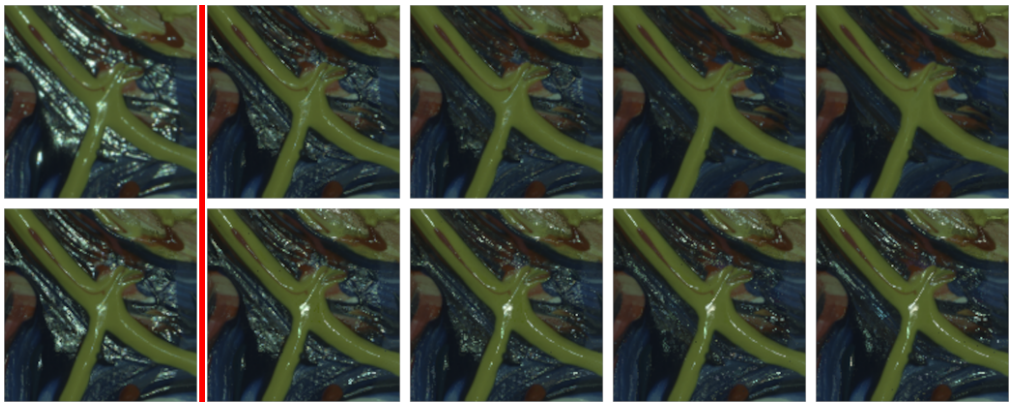


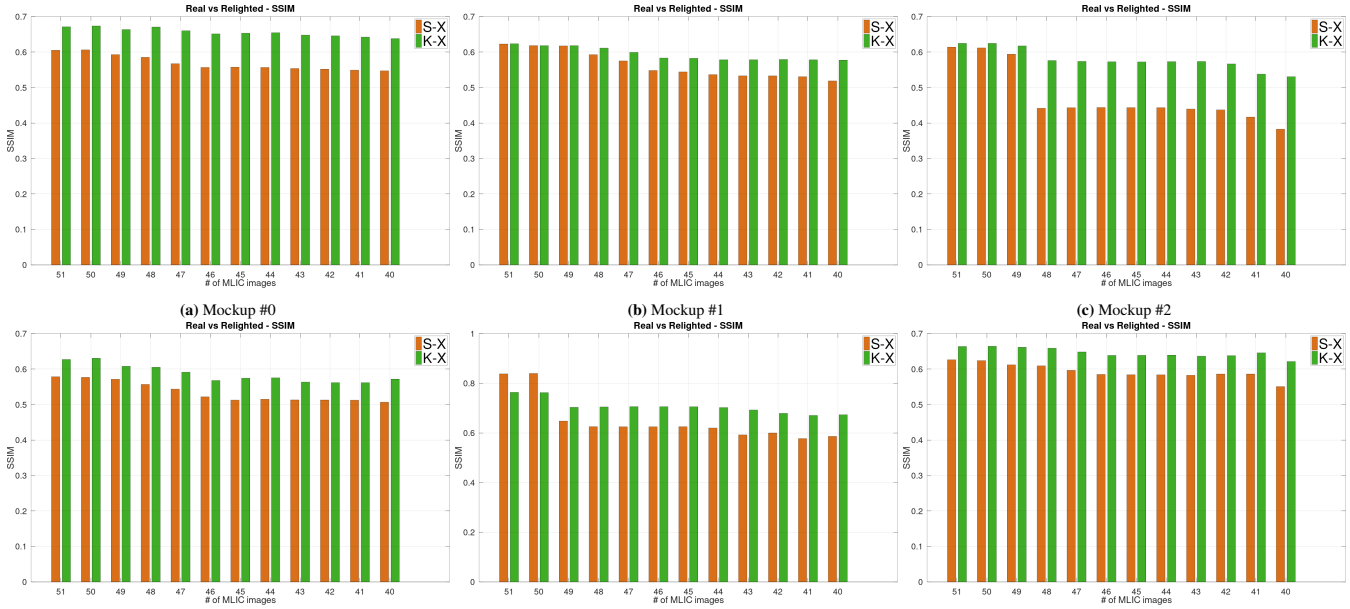
Figure 5: *From column two to five, we visually compare the relighting results obtained with the* S-X *(Up) and* K-X *(Bottom) method, by removing (from left to right) 1, 5, 11, and 12 images. On the first column we show the original image (Top) and the result of* S-All *(Bottom). In all the relightings, our approach keeps the glossy signal, that* S-X *method almost completely loses after the removal of just one image.*

in the glossy region. If the two diffuse regions are highly similar, the algorithm assumes that the unsampled glossy region of the central pixel is similar too, and assigns the proper color and glossiness value to that retrieved material. While this is a strength of our algorithm, since we are capable of recovering some highlights otherwise lost, on the other hand it could be sometimes a limitation as well. Sometimes, the unsampled glossy region is actually different from the well-sampled neighbor pixel, so we should marked a low similarity between them. When this scenario happens, in the relighted images this causes a spatial increase of highlight regions, as in Fig. 4, third or sixth columns. Beside this limitation, which will be the subject of future investigations, in general our solution can provide a better and more stable performance than the standard, state-of-the-art single pixel approach, even when the input MLIC provides a really sparse BRDF sampling.

4.2. User evaluation

We also performed a user test with the aim of assessing the advantages of the presented method for generating data to be used in relighting applications.

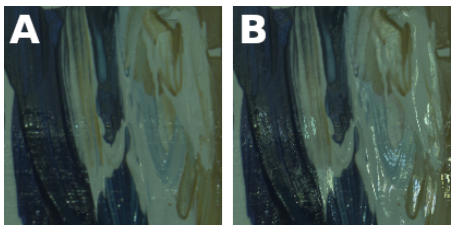
Goal. The main goal of the evaluation is to assess if the new SV-BRDF computation is more adequate for the usage in the typical scenario of object relighting, inspection and daily research activity, where many users with different skills and experiences try to interactively explore virtually relighted artworks. As already done for



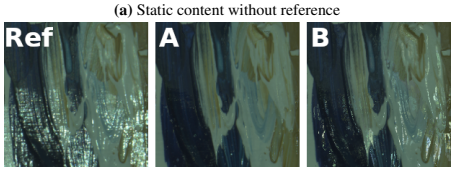
(a) Mockup #0 (b) Mockup #1 (c) Mockup #2

(d) Mockup #3 (e) Mockup #4 (f) Mockup #5

Figure 6: *We use SSIM metric to compare the performance of the single pixel strategy* S-X *with our k-neighbor based solution* K-X. *The graphs show the worst SSIM value in the relighted MLIC obtained from SV-BRDFs computed after removing from one to twelve images from the original MLICs of the six real painting mockups. The more images are removed, the worse is the relighting quality. However, our solution can generally provide better and more stable performances than the state-of-the-art single pixel approach.*



(a) Static content without reference



(b) Static content with reference

Figure 7: *The user was presented with two types of questions, i.e., the choice between two synced parallel videos (or static images) (a) without any reference, or (b) with a reference video/image.*

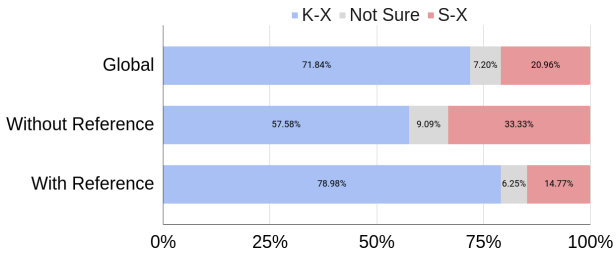


Figure 8: *We present here the global scores across all types of user tests, the scores of only the tests without and with reference. Globally, more than two third of the votes go to the proposed* K-X *solution, while* 20% *to the standard single pixel approach* S-X. *Although the proposed solution get the majority of votes when a reference video/image has been provided (With Reference row),* K-X *is capable of producing more natural and convincing relighting results even when no clue has been shown to the users (Without Reference row). The "Not sure" answer means that the algorithm performances are judged similar.*

the quantitative analysis in the previous section, we compare our neighbor-aware method (*K-X*) and the single pixel approach (*S-X*).

Setup. The test is conducted through a web-based questionnaire carried out by a number of volunteers. The questionnaire aims at measuring the quality of our algorithm based on user-perception. The experimental setup consists in two generic and eight specific sections. The first section tries to understand the type of users that submit the questionnaire, and some general opinion/feeling after they have performed the visual test. The specific sections ask users to visually compare relighting results obtained by the *K-X* and *S-X* algorithms, without knowing which algorithm generated which image. As relighting is typically used for interactive inspection, these section include both static and dynamic data. Static data consists in rendered image, while dynamic data consists in short clips of interactive relighting sessions with a fixed view and exactly the same light motion for all the presented choices. Moth static and dynamic tests are performed using two types of comparisons. The first is a comparison between two relightings without any reference image (or video). The second, is the same setup but with a reference relighting. The objects used for the tests are the same as included in our quantitative evaluation (Sec. 4.1).

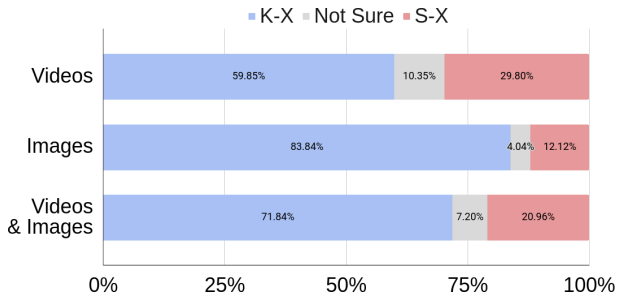


Figure 9: *Comparison between statistics related to only dynamic (videos) or static (images) tests. The percentage of "Not Sure" answer shows how it is more difficult to spot differences in the dynamic test. When the user is more confident about the answer, the votes for the proposed* K-X *method get to more than* 80%.

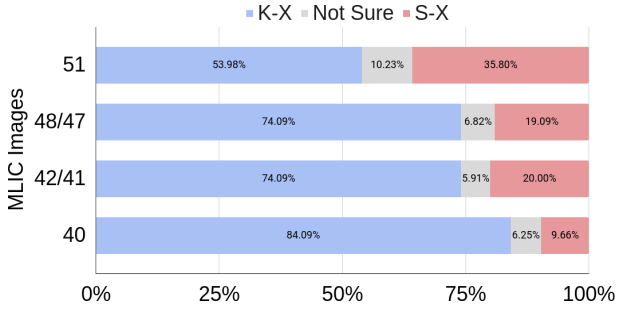


Figure 10: *User visual test performed after removing one* (51)*, four or five* (48/47)*, ten or eleven* (42/41)*, and twelve* (40) *images from the* 52 *image MLIC. In the first row the uncertainty is very high and the two methods get similar results. The more images we remove, the more the user appreciate the results produce by the proposed algorithm.*

Tasks. Each user was first presented with two synced parallel videos (or static images) without any reference (e.g., Fig. 7a). The user was asked to choose the video (or image) that looks more natural/photo-realistic, solely based on his/her visual perception and knowing the fact that it depicts an illuminated painted surface. After these sections, each dynamic and static representation is respectively coupled with a reference video/image (e.g., Fig. 7b), and users are asked to choosing the option that looks more similar to the reference. In all these scenarios, the users are requested to select one option, but, if the outcomes of the two relighting algorithms are very similar, and they cannot decide, there is also the possibility to choose not to pick any one. We set limitless time for the experiment, as we want users to inspect and select their options carefully. We choose to perform the two experiments (both without and with the reference) since there might be cases in which the algorithm reconstructs a SV-BRDF that, for instance, recovers the highlights but adds some artifacts in the final relighted image. In this scenario, we want to test if, with a reference image, the user would prefer the image with the artifacts rather than the image with all the highlights completely lost; on the other hand, without the reference, we want to test if the user would choose the image without the artifacts, since it looks more clean and photo-realistic.

Participants. 22 participants were recruited, spanning different ages (from 15 to 56), backgrounds, levels of education, and computer based skills. We include conservation scientists, researchers in computer science, students, teachers, and others (e.g., bioengineers and administrative staff). This distribution aims at testing the visual outcome of the methods not only from the point of view of expert conservation or visual computing scientists, but also from the perspective of more general cultural dissemination and virtual presentation to the public.

Evaluation of results. Let's start by analysing the scores across different types groups of visual tests. Fig. 8 shows that globally the majority of the votes goes to our solution (*K-X*), while one fifth only to the standard single pixel strategy (*S-X*). Although the choice between *K-X* and *S-X* is more clear when we provide a reference video or image (Fig. 8, third row), even when no clue has been provided, the proposed method is capable of producing more natural and convincing relighting results (Fig. 8, second row). In few cases (about from 6% to 9%) the outcomes of the two methods are very similar, and the user is not capable of making a clear decision. In Fig. 9 we evaluate the scores given by the users in two separate subgroups of the tests, i.e., only dynamic (videos) and only static (images) tests. In static tests, the user has a higher chance to better analyze the subtle differences between the two rendering, and the statistics exhibit a less percentage of uncertainty (about 4% of the cases). Conversely, in the videos the uncertainty arises up to ten percent. When the user is more confident about the answer, the votes for the proposed *K-X* method get to more than 80%. In the last graph (Fig. 10, we subdivide the statistics with respect to the number of images removed from the MLIC when computing the SV-BRDF and the relighting. We saw in the quantitative evaluation that removing images decreases the fidelity of the reconstruction, but our method exhibits a better statistics even in extreme cases. The user test confirms this behavior. We show four groups, i.e., the test performed removing one image (51), four or five images (48/47), ten or eleven images (42/41), and twelve (40) images from the original 52 image MLIC. Here, since we are grouping by removed number of images, *Y* stands for both *K* or *S*. We can see that, by removing only one image, the uncertainty is very high and the two methods are almost similar in terms of performances. The more images we remove, the more the user tends to vote and appreciate the results produce by the *K-X* algorithm, up to about 85% when we remove twelve images from the original MLIC. As a conclusion, we can deduce that, even with a strong undersampling, the virtual relighting, done with the digitally characterized SV-BRDF computed by the proposed approach, still appears photo-realistic, with significant improvements with respect to single-pixel approaches.

5. Conclusions

We have presented a practical solution to create a relightable model from Multi-light Image Collections (MLICs) acquired using standard acquisition pipelines. Our approach targets the difficult problem of creating shape and material models from a limited number of single-view acquisitions of flat but visually and geometrically rich objects. By exploiting information from neighboring pixels through a carefully crafted weighting and regularization scheme, we are able to efficiently infer subtle per-pixel analytical Bidirectional Reflectance Distribution Functions (BRDFs) representations from few per-pixel samples. As each pixel reconstruction is independent, the

proposed method can be easily integrated in common scalable out-of-core pipelines that estimate per-pixel characteristics in parallel.

Our qualitative and quantitative results on both synthetic and real data shows that we are able to recover high-frequency specular information where sufficient data is locally available, falling back to regularized solutions without unwanted high-frequency artifacts in other situations. In order to provide an evaluation of the approach with quantitative measures compared to ground truth, we have concentrated on the measurement of limited size samples of various characteristics using a light dome. We are currently working on applying the technique to large scale acquisition on paintings, also using fully free-form setups.

Acknowledgments. We thank the anonymous reviewers for their helpful comments and suggestions. The project received funding from the European Union's H2020 research and innovation programme under grant 813170 (EVOCATION), and from Sardinian Regional Authorities under project VIGECLAB (POR FESR 2014-2020).

References

[AWL13] AITTALA M., WEYRICH T., LEHTINEN J.: Practical SVBRDF capture in the frequency domain. *ACM Transactions on Graphics (TOG) 32*, 4 (2013), 110–1. 2

[BJK07] BASRI R., JACOBS D., KEMELMACHER I.: Photometric stereo with general, unknown lighting. *International Journal of Computer Vision 72*, 3 (2007), 239–257. 2

[CHI19] CHI: Cultural heritage imaging website, 2019. [Online; accessed-July-29-2021]. URL: http://culturalheritageimaging.org. 1

[DHOMH12] DREW M. S., HEL-OR Y., MALZBENDER T., HAJARI N.: Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing 30*, 4-5 (2012), 317–331. 2

[Dü06] DÜR A.: An improved normalization for the ward reflectance model. *Journal of Graphics Tools 11*, 1 (2006), 51–59. 4

[ERF11] ELHABIAN S. Y., RARA H., FARAG A. A.: Towards accurate and efficient representation of image irradiance of convex-Lambertian objects under unknown near lighting. In *Proc. ICCV* (2011), pp. 1732–1737. 2

[GCD*18] GIACHETTI A., CIORTAN I. M., DAFFARA C., MARCHIORO G., PINTUS R., GOBBETTI E.: A novel framework for highlight reflectance transformation imaging. *Computer Vision and Image Understanding 168* (2018), 118–131. 1

[GGG*16] GUARNERA D., GUARNERA G. C., GHOSH A., DENK C., GLENCROSS M.: BRDF representation and acquisition. *Computer Graphics Forum 35*, 2 (2016), 625–650. 1, 2, 4

[HS16] HUI Z., SANKARANARAYANAN A. C.: Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence 39*, 10 (2016), 2060–2073. 2

[JAP*21] JASPE VILLANUEVA A., AHSAN M., PINTUS R., GIACHETTI A., GOBBETTI E.: Web-based exploration of annotated multi-layered relightable image models. *ACM Journal on Computing and Cultural Heritage 14*, 2 (May 2021), 24:1–24:31. 1

[JH13] JANKOWSKI J., HACHET M.: A survey of interaction techniques for interactive 3D environments. In *Eurographics STAR* (2013). 1

[Jon01] JONES D. R.: DIRECT global optimization algorithm. *Encyclopedia of optimization* (2001), 431–440. 5

[KUL19] KUL: PLD software KU-Leuven, 2019. [Online; accessed-July-29-2021]. URL: https://portablelightdome.wordpress.com/software. 1

[LBAD*06] 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. In *ACM Transactions on Graphics (TOG)* (2006), vol. 25-3, pp. 735–745. 2

[LBFS21] LAVOUÉ G., BONNEEL N., FARRUGIA J.-P., SOLER C.: Perceptual quality of brdf approximations: dataset and metrics. *Computer Graphics Forum 40*, 2 (2021), 327–338. 1, 2, 4

[Mac15] MACDONALD L. W.: *Realistic visualisation of cultural heritage objects*. PhD thesis, UCL (University College London), 2015. 1

[MGW01] MALZBENDER T., GELB D., WOLTERS H.: Polynomial texture maps. In *Proc. ACM SIGGRAPH* (2001), pp. 519–528. 2

[NDM05] NGAN A., DURAND F., MATUSIK W.: Experimental analysis of BRDF models. In *Rendering Techniques* (2005), pp. 117–126. 4, 5

[PCS18] PONCHIO F., CORSINI M., SCOPIGNO R.: A compact representation of relightable images for the web. In *Proc. ACM Web3D* (2018), pp. 1–10. 1

[PDC*19] PINTUS R., DULACHE T., CIORTAN I., GOBBETTI E., GIACHETTI A.: State-of-the-art in multi-light image collections for surface visualization and analysis. *Computer Graphics Forum 38*, 3 (2019), 909–934. 1, 2

[PGP*17] PINTUS R., GIACHETTI A., PINTORE G., GOBBETTI E., ET AL.: Guided robust matte-model fitting for accelerating multi-light reflectance processing techniques. In *Proc. BMVC* (2017), BMVA Press. 2, 3

[PJZ*21] PINTUS R., JASPE VILLANUEVA A., ZORCOLO A., HADWIGER M., GOBBETTI E.: A practical and efficient model for intensity calibration of multi-light image collections. *The Visual Computer 37* (2021). 2

[PLGM*17] PITARD G., LE GOÏC G., MANSOURI A., FAVRELIÈRE H., DESAGE S.-F., SAMPER S., PILLET M.: Discrete modal decomposition: a new approach for the reflectance modeling and rendering of real surfaces. *Machine Vision and Applications 28*, 5-6 (2017), 607–621. 2

[RDL*15] REN P., DONG Y., LIN S., TONG X., GUO B.: Image based relighting using neural networks. *ACM Transactions on Graphics (TOG) 34*, 4 (2015), 111:1–111:12. 2

[RJGW19] RAINER G., JAKOB W., GHOSH A., WEYRICH T.: Neural BTF compression and interpolation. *Computer Graphics Forum 38*, 2 (2019), 235–244. 2

[Rus98] RUSINKIEWICZ S. M.: A new change of variables for efficient BRDF representation. In *Rendering techniques* (1998), pp. 11–22. 3

[SJR18] SUN T., JENSEN H. W., RAMAMOORTHI R.: Connecting measured BRDFs to analytic BRDFs by data-driven diffuse-specular separation. *ACM Trans. Graph. 37*, 6 (Dec. 2018). 4

[SWM*16] 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* (2016), pp. 3707–3716. 1

[Tex21] TEXTURE E.: 3D texture database, 2021. [Online; accessed-July-29-2021]. URL: http://https://everytexture.com/. 5

[TGVG12] TINGDAHL D., GODAU C., VAN GOOL L.: Base materials for photometric stereo. In *Proc. ECCV* (2012), pp. 350–359. 2

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing 13*, 4 (2004), 600–612. 5

[XR20] X-RITE: Scanner TAC7 total appearance capture, 2020. [Online; accessed-January-17-2020]. URL: https://www.xrite.com/categories/appearance/tac7. 2

[XSHR18] XU Z., SUNKAVALLI K., HADAP S., RAMAMOORTHI R.: Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG) 37*, 4 (2018), 126. 2

[ZD14] ZHANG M., DREW M. S.: Efficient robust image interpolation and surface properties using polynomial texture mapping. *EURASIP Journal on Image and Video Processing 2014*, 1 (2014), 25. 2