Exploiting Local Shape and Material Similarity for Effective SV-BRDF Reconstruction from Sparse Multi-Light Image Collections

RUGGERO PINTUS, MOONISA AHSAN, ANTONIO ZORCOLO, FABIO BETTIO, FABIO MARTON, and ENRICO GOBBETTI, Visual Computing Group, CRS4, Italy

We present a practical solution to create a relightable model from small 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 and visually pleasing per-pixel analytical Bidirectional Reflectance Distribution Functions (BRDFs) representations from few per-pixel samples. The method has a low memory footprint and is easily parallelizabile. We qualitatively and quantitatively evaluated it on both synthetic and real data in the scope of image-based relighting applications.

CCS Concepts: • **Computing methodologies** → **Appearance and texture representations**; **Reflectance modeling**; *Scene understanding*.

Additional Key Words and Phrases: MLIC, Reflectance computation, BRDF fitting, Virtual Relighting, Paintings, Bas-reliefs

ACM Reference Format:
Ruggero Pintus, Moonisa Ahsan, Antonio Zorcolo, Fabio Bettio, Fabio Marton, and Enrico Gobbetti. 2022. Exploiting Local Shape and Material Similarity for Effective SV-BRDF Reconstruction from Sparse Multi-Light Image Collections. *ACM J. Comput. Cult. Herit.* 1, 1 (September 2022), 31 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

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 [23].

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 [5, 9, 18, 27] 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 [23]. The resulting data is then

Authors' address: Ruggero Pintus, ruggero.pintus@crs4.it; Moonisa Ahsan, moonisa@crs4.it; Antonio Zorcolo, zarco@crs4.it; Fabio Bettio, fabio@crs4.it; Fabio Marton, marton@crs4.it; Enrico Gobbetti, gobbetti@crs4.it, Visual Computing Group, CRS4, Ex-Distilleria, Via Ampère 2, Cagliari, Italy, 09134.

XXXX-XXXX/2022/9-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Fig. 1. Capture and relighting of a painting (panel of the polyptych retable of Saint Bernardino (1455), Cagliari, Italy). The optical response of the painting surface to variable illumination is measured by taking a few tens of photos using a fixed reflex camera and a hand-held LED (left). The Multi-Light Image Collection is then transformed to a shape and material representation, which is used for interactive relighting (right)

fit to a compact model, which is exploited by interactive visual inspection tools [4, 5, 13, 15] to support virtual relighting (Fig. 1). 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 with 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 [12].

While classic virtual inspection solutions were restricted to exploiting low-frequency analytical relighting representations, such as PTM or HSH [23], 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 [13]. Having a compact per-pixel normal and an analytical BRDF representation from sparsely sampled data is very appealing, since it can be easily distributed using low-bitrate representations, it produces a physically reasonable result, and it allows for natural integration with standard high-quality and real-time rendering solutions. However, while normal estimation is a well-studied subject [31], the per-pixel extraction of Spatially Varying BRDF (SV-BRDF) parameters of a multi-material object from the small number of samples typically available in sparsely sampled single-view MLICs leads to an under-determined problem [10, 16]. For this reason, the application of standard per-pixel fitting to MLICs with a small number of images may produce very noisy maps, especially in presence of glossy materials. Heavily increasing the number of captured images increases capture complexity and effort for the common dome-based and free-form setups used in CH, making the method much less appealing for the practitioners. The available alternative solutions to produce compelling models from a small number of samples try to solve the problem either by deriving extra knowledge from available training sets tuned for a particular target object kind and/or by analyzing the entire object to find similar materials in areas with significant normal variation, so as to cover the largest possible angular sampling space. Effectively combining data from different locations requires, however, the solution of complex material classification problems (Sec. 2).

Since we primarily target planar objects that are visually and geometrically rich, such as paintings or bas-reliefs, for each considered surface point, the probability of finding a different surface orientation is mostly independent

from the distance to that point, while the probability of finding the same exact optical behavior is much larger in a small neighborhood than far from it.

Based on this consideration, we introduce a practical solution that computes the BRDF parameters of a pixel by carefully combining all the observations available in a small neighborhood (Sec. 3). As for usual fitting pipelines, BRDF fitting is formulated as an optimization for the set of BRDF parameters that minimizes the sum of weighted squared differences between measured surface reflectance samples and the corresponding values from the analytical BRDF model. In our case, however, the weights do not only measure the importance of the particular measure, but determine our confidence that the given measure comes from a surface area with the same material as the computed pixel (Sec. 4). 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, by including bilateral weighting, based on pixel distances and measured values, we also obtain a non-linear, edge-preserving, and noise-reducing smoothing. 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 (Sec. 5). While the method is applicable to improve the fitting of any parametric BRDF, our weighted least squares formulation permits to apply important optimizations for the common case of a multi-lobe BRDF composed of a sum of terms (Sec. 6). As a result of its design, our solver can directly replace the fitting module inside all standard per-pixel BRDF fitting pipelines that estimate the parameters in parallel for each pixel of the image, and can operate with little memory overhead, without requiring global access to all images (Sec. 7). 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.

This article is an invited extended version of our contribution to the 19^{th} *Eurographics Workshop on Graphics and Cultural Heritage (GCH 2021)* [22]. Here, we provide a much more thorough exposition, as well as significant new material. Our main novel contributions with respect to the conference version include an improved similarity metric among pixel descriptors (see Sec. 4.3.2), a novel technique for computing similarity weights inside a region through the diffusion inside the pixel neighborhood (see Sec. 4.3.3), and a novel pruning technique for reducing solution costs for large neighborhoods (see Sec. 7). We also present an extended and improved qualitative and quantitative evaluation on synthetic and captured data, and discuss the application of the method to the free-form capture and subsequent interactive relighting of a painting. Our results demonstrate significant improvements with respect to standard single-pixel solutions and to our previous work both in terms of accuracy of fitting and perceptual quality of results (Sec. 8).

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 [10, 16, 23]. In the following, we briefly cover only the approaches most closely related to ours.

2.1 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 [19], called Polynomial Texture Mapping (PTM), stores per-pixel coefficients of a 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 [39], or by improving the fitting algorithm with robust metrics [6, 24]. 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 [3, 8], or a Discrete Modal Decomposition (DMD) [26]. 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 [6]. 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 [23]

2.2 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 [10]. 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 [37]. 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) [1]. 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 [17], 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 [11], or to extract base materials through a global segmentation and subsequent clustering of appearance profiles [34]. 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.

2.3 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 [28, 29, 38]. 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 OVERVIEW

Multi-light reflectance data is acquired by taking from the same pixel-registered viewpoint a set of photographic images of an object, with each image illuminated from a different direction. As a result, a calibrated processing pipeline receives as input a so-called per-pixel appearance profile, i.e., a list of pairs for each pixel p that couple, for each entry δ, a viewing and lighting configuration, consisting in a view direction V_p, a light direction $L_{p,\delta}$, and a light intensity $I_{p,\delta}$, to the corresponding measured value $m_{p,\delta}$. Any calibration pipeline can be applied to obtain this information (for this article, we used the recently introduced method by Pintus et al. [25]). Moreover,

the set of observations is trimmed, to discard shadow areas, which do not provide data, as well as under-exposed and over-exposed pixels, which provide unreliable information.

Pixel saturation, where the incident light at a pixel causes one of the color channels of the camera sensor to respond at its maximum value, produces an unreliable measure that cannot be used for fitting, as we only know that the measured intensity is higher or equal to the pixel value. For this reasons, we consider those pixels over-exposed, and discard them. Similarly, pixels that are too dark also to not provide a reliable measure, as it is not easy to distinguish if the dark color is due to a cast shadow, and therefore should not be used for computing the response of the source to the illumination coming from the light. Moreover, when the measured value is too low, the incidence of noise is also high. In this work, we set a threshold for the lowest possible intensity value equal to the 0.1% of the maximum possible value, and discard the under-exposed pixels with a measured intensity below that threshold. This trimming results in a variable-size per-pixel appearance profile.

Surface shape and appearance can then be characterized for each pixel p by converting this information to shape and appearance descriptors. In this work, we assume that the shape is determined by the normal N_p at each pixel, computed using any of the available photometric stereo techniques (for our results, we use the method of Pintus et al. [24]), and focus on the extraction of a per-pixel reflectance mode by fitting the set of multi-light measurments to a fixed analytical BRDF f_r parameterized by per-pixel parameters Π_p. Prior to fitting, we further reduce the size of appearance profiles by discarding measurements at grazing $V_p \cdot N_p$ and $L_p \cdot N_p$ angles bigger than 80 degrees, as performed by Ngan et al. [20] and demonstrated beneficial for BRDF processing by Lavoué et al. [16] to remove likely unreliable measurmements. This manipulation leads to a final appearance profile of size Δ_p, different for each pixel p.

Given shape, viewing, and lighting information for each pixel, we cast BRDF estimation of a pixel p into a weighted least squares problem of the form:

$$\Pi_p = \underset{\Pi}{argmin} \sum_{q \in \Omega_p} \sum_{\delta_q \in \Delta_q} \left(w^2_{p,q,\delta_q} \left\| m_{q,\delta_q} - f_r\left(\Pi, N_q, V_q, L_{q,\delta+q}\right) \right\|^2 \right) + \lambda R_p^2 \tag{1}$$

where Ω is a WxW window centered on p, so that q is a pixel in a local neighborhood, and m_{q,δ_q} is a normalized measurement value that factors in incident light intensity, w_{p,q,δ_q} is a per-measurement weight (Sec. 4), and λR_p^2 is a regularization term (Sec. 5). In the following, we define how we compute these values, before describing how we efficiently solve Equation 1 by focusing on multi-lobe BRDFs (Sec. 6) and by approximating the equation with a fixed number of terms and distributing computation to avodi reduntant computations (Sec. 7).

4 WEIGHTING STRATEGY

The weight w_{p,q,δ_q} associated to each measure determines the importance of each error term in Equation 1. In regular single-pixel approaches, it is defined to balance the effects of each of the measurements for the single pixel, which are known to be of the same material since they come from a single surface point [16]. In this work, the measurements that contribute to the fitting also come from a neighboring area in a multi-material object. For this reason, we expand the concept of weight to also incorporate an estimation of the likelihood that they really come from the same material of the computed pixel. We therefore define the weight of a given error term as:

$$w_{p,q,\delta_q} = w^{measurement}_{q,\delta_q} w^{radial}_{p,q} w^{similarity}_{p,q} \tag{2}$$

where $w^{measurement}_{q,\delta_q}$ is a per-measurement weight associated to each entry in the appearance profile of pixel q, $w^{radial}_{p,q}$ is a radially decreasing weight that depends only on the spatial distance between the central pixel p and the neighboring pixel q, and $w^{similarity}_{p,q}$ is a similarity weight that increases with the the distance between the appearance profile of pixel p and the profile of pixel q.

4.1 Measurement weight

In the literature, many metrics have been presented that weight each error term so as to produce a more stable and perceptually valid error metric [16]. In particular, it is customary to weight each term with the cosine input angle $N_q \cdot L_q$ in order to compensate reflection increase at grazing angles. Lavouè et al. [16] have, in addition, shown how non-linearly compressing the measurement m and the BRDF $f_r(...)$ with a cube root to attenuate peak values before computing the fitting error greatly improves the perceptual quality of the fit, as it avoids the overfitting of specularities, and the magnification of their measurement errors, at the expense of all the other regions of the BRDF. Applying such a trasformation, however, would introduce a non-linearity that could avoid the application of optimizations for the fitting of multi-lobe BRDFs (Sec. 6). We, therefore, strive to achieve the same goal by linearly compressing the data through a measurement-dependent weight that penalizes higher-magnitude samples. Given the good perceptual performance of the cube-root transformation, we set the weight to $w_{q,\delta_q}^{measurement} = m_{q,\delta_q}^{-2/3}$, which applies the same compression to the measured value. Using a measurement-dependent weight is also used in statistics to minimize relative rather than absolute errors [35]. Note that, here, we have no instability around zero here, since we trim appearance profile prior to fitting to remove under-exposed and shadow areas. The measurement weight can also be extended by multiplying it with other factors, e.g., the cosine input and output angle [16]. The results presented in this article do not include those factors, since we did not appreciate significant differences in results for our use cases. This is mainly due to the pruning that discards the samples at grazing angles, for which these factors have the largest effect.

4.2 Radial weight

Due to material continuity, a pixel closer to the center has a higher 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:

$$w_{p,q}^{radial} = \max\left(0, 1 - \left(\frac{r_{p_q}}{R}\right)^2\right), \tag{3}$$

where $r_{p,q}$ is the distance between the central pixel p and the other pixel q, while $R = \frac{W}{2}$ 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 combination of this weight with the other weights plays the same role as the spatial closeness factors in bilateral filtering, as it drives the solver towards edge-preserving smoothed solutions.

4.3 Similarity weight

The similarity weight $w_{p,q}^{similarity}$ encodes our estimation of the likelihood that the appearance profiles of pixel p and q come from measurements of the same material at different surface location. The core idea of using a similarity metric is to have a way to effectively extend the sampling of a BRDF without requiring taking multiple images, just by borrowing data from similar pixels. By including with a high weight $w_{p,q}^{similarity}$ the samples from a pixel q with a similar normal of p, we achieve denoising, while if q has a different normal we obtain the even more important effect of increasing the angular sampling, eventually retrieving hard to capture specularities.

Intuitively, we must use a metric that will provide high similarity if the profiles contain close measured values at the same angles, and a low similarity if measurements at the same angles widely differ. Efficiently and reliably computing such a metric for two sets of sparsely sampled measurements is, however, non trivial, as we must compute it for all pixels in the neighborhood and produce likely values also for pixels sampled at different viewing and light angles, hence with little overlap in angular sampling. Our solution relies on the following components: first, we transform variable-sized appearance profiles into compact fixed-size represention that capture the main reflectance characteristics; second, we introduce a distance function among descriptors that are computed from

appearance profiles that share sampling locations; finally, we use a propagation technique to derive the final similarity metric valid also for pixels with different normals.

4.3.1 Pixel descriptors. In order to efficiently similarity without having to traverse variable-sized appearance profiles, we transform the profile of each pixel q into a feature vector $F_q = F_q^{(0)}, F_q^{(1)}, ..., F_q^{(B-1)}$ that contains a compact regularized representation of size B (with $B = 10$ in this paper). Each $F_q^{(i)}$ slot contains the maximum sampled color for each region of the Θ_h angle in the Rusinkiewicz parameterization [30]. Θ_h is a function of the angle between the normalized half vector $H_q = \frac{N_q \cdot L_q}{\|N_q \cdot L_q\|}$ and the normal N_q, 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 directly 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.

Since the feature vector is only dependent on the profile, it can be computed once and for all in parallel for each pixel, prior to the BRDF computation, and can be considered part of the pixel description during the entire optimization process.

4.3.2 Direct similarity weight. Given two feature vectors F_p for the central pixel and F_q for a pixel in the neighborhood, we compute the similarity based on a distance metric $d(...)$ between the two features:

$$w_{p,q}^{direct-sim} = 1 - \min\left(1, d\left(F_p, F_q\right)\right) \tag{4}$$

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 very different, we have no information on how to compute the distance, and, thus, we must return the maximum distance, as, in case of doubts, we prefer to exclude from computation pixels coming from different materials rather then include them. Similarly, the maximum distance must be 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 configuration. In all other cases, we must return a distance directly proportional to the relative difference of spectral values in the same bins using a perceptual simularity metric. For the latter, we use the log2 metric of Sun et al. [32]. To sum up, by assigning to each bin index i a value $b^{(i)} = 0$ if the bin i of p and q both contain values and $b^{(i)} = 1$ otherwise, the distance function at the basis of our direct similarity computation will be:

$$d\left(F_p, F_q\right) = \begin{cases} 1, & \text{if } \sum_{i=0}^{B-1} b^{(i)} = 0 \quad \text{(Lack of overlap)} \\ 1, & \text{if } \min_i \frac{F_p^{(i)}}{\left\|F_p^{(i)}\right\|} \cdot \frac{F_q^{(i)}}{\left\|F_q^{(i)}\right\|} \leq \cos(\chi) \quad \text{(Chroma diff)} \\ 1, & \text{if } \max_i \log^2\left(\frac{\left\|F_p^{(i)}\right\|+\varepsilon}{\left\|F_q^{(i)}\right\|+\varepsilon}\right) \geq \log^2 \rho \quad \text{(Magnitude diff)} \\ \frac{1}{\rho \sum_{i=0}^{B-1} b^{(i)}} \sum_{i=0}^{B-1} b^{(i)} \log^2\left(\frac{\left\|F_p^{(i)}\right\|+\varepsilon}{\left\|F_q^{(i)}\right\|+\varepsilon}\right), & \text{otherwise} \quad \text{(Possibly similar)} \end{cases} \tag{5}$$

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. Thus, the angle χ is 5 degrees, while $\rho = 1.1$, which means maximum allowed relative magnitude deviation of 10%.

4.3.3 Propagated similarity weight. Since we compare only overlapping regions of the BRDF, Equation 4 can effectively achieve denoising, as well as improve the angular sampling of the BRDF, and thus the recovery of non-diffuse behaviors, in case of smooth normal variations due to local roughness or small curvature. However, the function is, by design, very conservative when comparing pixels with different normals, and, therefore, with little or no overlap in angular sampling. While this avoids degrading BRDF recovery by mixing different

materials, it also blocks the improvement that could be possible by including differently oriented areas with the same material. We handle this problem by diffusing simularity measures in the local neighborhood, aiming to determine the similarity of differently-oriented pixels from the iterative combination of the similarity of similarly-oriented pixels. In practice, for each pixel q_i in the neighborhood, we compute, through Equation 4, the direct similarity $w_{q_i,q_j}^{direct-sim}$ with its eight direct neighbors q_j and store it in the edge of a graph having each pixel as a node. We then initialize each node value $w_{q_i,p}^{node}$ to $w_{q_i,p}^{direct-sim}$. We then iteratively replace each $w_{q_i,p}^{node}$ with $\max\left(w_{q_i,p}^{node}, \sqrt{w_{q_i,q_j}^{direct-sim} w_{q_j,p}^{node}}\right)$ until convergence. At the end of this relaxation process, each node value contains the final estimation of the similarity weight $w_{p,q}^{similarity}$.

5 REGULARIZATION

Due to the nature of sparse capturing of mostly flat multi-material objects, it is likely that certain pixels might have under-sampled regions of the BRDF. We therefore include in our cost function Equation 1 a regularization term λR_p^2 that controls the fitting behavior in under-determined cases to avoid disturbing visual artifacts. In particular, if a BRDF is sampled only in the diffuse zone, the fitter 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 regularize the fitting by introduce a single virtual BRDF measurement R_p in a highly specular direction (i.e., for V_p and L_p both collinear with N_p) with a value set to the largest measured value for that pixel among all sampled light directions, and assign it a weight λ. As the weight λ is set to a very low value, this possibly underestimated value only affects the solutions when no measured data is available for those angles. In fact, we chose the value of λ small enough to make the term associated to the virtual measurement much smaller than the terms associated to real ones, but high enough to prevent the final BRDF to explode in magnitude. We have performed several experiments with both only diffuse and specular pixels, and we found that a value of $\lambda = 10^{-4}$ is a good conservative and stable choice across many different heterogeneous datasets.

6 BRDF REPRESENTATION AND OPTIMIZATION APPROACH

While our method is in principle applicable to any parametric BRDF, our weighted least squares formulations permits to apply important optimizations for the common case of a multi-lobe BRDF consisting 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.:

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

The term K_γ^S represents the spectral value of the γ^{th} BRDF component, and it is a multi-spectral color vector of dimensionality S; for an RGB signal $K_\gamma^3 = \left(k^R, k^G, k^B\right)$. The scalar function $f_\gamma\left(\dots\right)$ is the component BRDF; for instance, for a Lambertian component of the material $f_\gamma\left(\dots\right) = const = 1/\pi$. 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 isotropic 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:

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

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,

Equation 1 becomes:

$$\Pi_p = \underset{\Pi}{argmin} \sum\nolimits_{q\in\Omega} \sum\nolimits_{\delta_q\in\Delta_q} \left(w_{p,q}^2 \left\| m_{q,\delta_q} - \sum_{\gamma=0}^{\Gamma-1} K_\gamma^S f_\gamma \left(\Pi_\gamma, N_q, V_q, L_{q,\delta_q} \right) \right\|^2 \right) + \lambda R_\Omega^2 \tag{8}$$

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 [10]. 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 article, in particular, we test the proposed framework with the analytic BRDF f_r set to Duer's variant of the original isotropic Ward model [7]:

$$f_r\left(\alpha, N, V, L\right) = K_d^S f_d + K_s^S f_s\left(\alpha, N, V, L\right) = K_d^S \frac{1}{\pi} + K_s^S \frac{1}{4\pi\alpha^2\sqrt{\left(N\cdot L\right)\left(N\cdot V\right)}} e^{-\frac{(H\cdot X)^2+(H\cdot Y)^2}{\alpha^2(H\cdot N)^2}}, \tag{9}$$

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 $\sum_{i=0}^{\Gamma-1}\left|\Pi_\gamma\right|$ 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. [20]. In our current implementation, we use the locally-biased DIRECT global optimization algorithm for the nonlinear search [14] and a SVD solver for the linear least squares sub-problem.

From the formulation depicted in Equation 6, it is clear that the proposed model can be applied to a wide range of different BRDF other than the chosen Ward model. The only requirement is that the BRDF can be expressed in terms of a sum of products of spectral values by scalar functions. This includes, in particular, models for layered materials in which, for instance, a Lambertian or Disney diffuse BRDF is combined with a microfaced-based specular BRDF (e.g., GGX) by simply summing the contributions weighted by the Fresnel term [10]. Nonlinear search for finding the values of the scalar parameters should be adapted to work in more than one dimension, either by using the multi-dimensional DIRECT solver or replacing it with any other nonlinear global solver, while the linear solver that finds the spectral values will remain unchanged.

7 PRUNING AND SCALABILITY

The method presented in the previous sections can be further optimized by reducing the cost of fitting and by carefully organizing computation phases to handle very large images without memory limitations, while exploiting parallel computation.

7.1 Pruning

Solving the weighted nonlinear least squares problem dominates the fitting cost. Its complexity is directly proportional to the number of included weighted measurements, which determines the overall cost of evaluating the objective function, in general, as well as the size of the linear least-squares problem for the optimized implementation of Sec. 6. In its basic formulation, used in our previous work [22], this number is strictly determined by the the window size and the number of input photographs. For a typical neighborhood with a radius R of 10 pixels (i.e., a window of 21 pixels), we would need to solve a problem $21\times 21 = 441$ times larger than for the equivalent single-pixel fitting.

We can reduce this overhead by noting that not all the terms that determine the value of Equation 8 have the same importance. First of all, the sum is likely vastly dominated by the terms with the largest weight, corresponding to the central pixel and to pixels with the most similar material. Moreover, several of the terms might be redundant, as they provide multiple measurements for the same angular configuration, while others have more impact on reconstruction, as they are alone in sampling regions of the BRDF. We can, therefore, reduce evaluation cost by approximating Equation 8 with a low number B of samples, using weight magnitude and distribution of angular sampling considerations to discard the terms that have likely the lowest impact on reconstruction accuracy.

To perform pruning based on angular distribution and weight magnitude, we subdivide the 2D Rusinkiewicz parameter space into buckets, and order them with θ_h as primary sort key and θ_d as secondary sort key. As for feature vector computation, bucket size is directly proportional to $\sqrt[3]{\theta}_h$, so as to more densely sample the specular region. We then iteratively visit all buckets in order of priority (i.e., from the most diffuse to the most specular), removing samples while the count of remaining ones exceed the target B. For a given visited bucker, we remove the samples with the lowest weight until we reach the minimum sampling rate for the cell. For all our experiments, we use a $10x10$ grid subdivision of the BRDF domain, and a minimum sample count per bucket equal to 3.

As B can be set to a fixed value, the cost becomes independent of neighborhood size W. Speed up is considerable, since we can set, e.g., $B = 50$, i.e., much less than $W \times W \times \Delta$, with negligible effect on solution accuracy (see Sec. 8.1.3).

7.2 Scalable tile-based implementation

The pixel-wise, local nature of our computation allows for a single-pass highly parallel and low-memory implementation of the overall BRDF estimation process. For efficiency of access to stored data, our method uses a tile-based approach. We define and in-core memory budget, and, from that, we subdivide the MLIC into tiles in image space. The tiles slightly overlap (by an amount equal to W), to make the computation of non-overlapping pixels self-contained.

Each tile is computed independently (sequentially on a single machine, or distributed to multiple machines on a cluster). The computation of each tile starts by loading into memory all the appearance profiles, the viewing and lighting environment, and the normal map. All the subsequent computation exploit current multi-core architectures by parallelizing computation with one thread per pixel. We first compute, in parallel for each pixel, the descriptors that summarize the profiles for weight computation, and store them as per-pixel data (Sec. 4.3.1). Given these descriptors, we compute and store, in parallel for each pixel, the distance to their 8 neighbors according to Equation 5, since it is reused many times for similarity computation and is independent of the central pixel location (Sec. 4.3.3). Propagated similarity weight computation needs, thus, to only compute for each neighborhood the distance to the central pixel, as distances to neighbors are known. Moreover, since the distance function is symmetric, we only need to compute 4 distance per pixel to obtain all 8-neighborhood values.

Given the descriptor map and the distance map, we can proceed, in parallel for each pixel, with weight computation (Sec. 4), regularization term computation (Sec. 5), pruning (Sec. 7.1), and optimization (Sec. 6), so as to determine the final BRDF parameters for all pixels in the tile.

8 IMPLEMENTATION AND RESULTS

The proposed approach has been implemented in C++, and integrated in a pipeline for producing relightable image representations from MLICs.

In this article, we report on the validation of the proposed solution by analyzing its performance in the reconstruction of SV-BRDFs from sparsely sampled MLICs. We compare our results with 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 output pixel without considering its neighborhood. Moreover, we also include tests that illustrates improvements over the original version of this method, as presented in our GCH 2021 contribution [22]. For all our tests, we use RGB images and the isotropic Ward analytic BRDF. In this setting, 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. For all our tests we use the same pruning parameters, i.e., we subdivide the BRDF domain with a $10x10$ grid, the minimum number of samples per cell is 3, and the maximum final sample budget is 150.

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, it should be noted that even with lots of images, a single-view MLIC acquisition intrinsically provides an undersampled set of BRDF measurements, due to the fixed view point. Thus, including information coming from differently oriented surface is expected to be beneficial even in that case.

In the following, we first provide a quantitative analysis of the quality achievable when reconstructing BRDFs using both synthetic data, for which ground truth is available, and painting mockups captured in controlled conditions (Sec. 8.1). Then, we report the result of a user test, which provides a perceptual human feedback for the visualization quality of the proposed method for relighting applications (Sec. 8.2). Finally, we show the results obtained in practice for the free-form acquisition of a painting (Sec. 8.3).

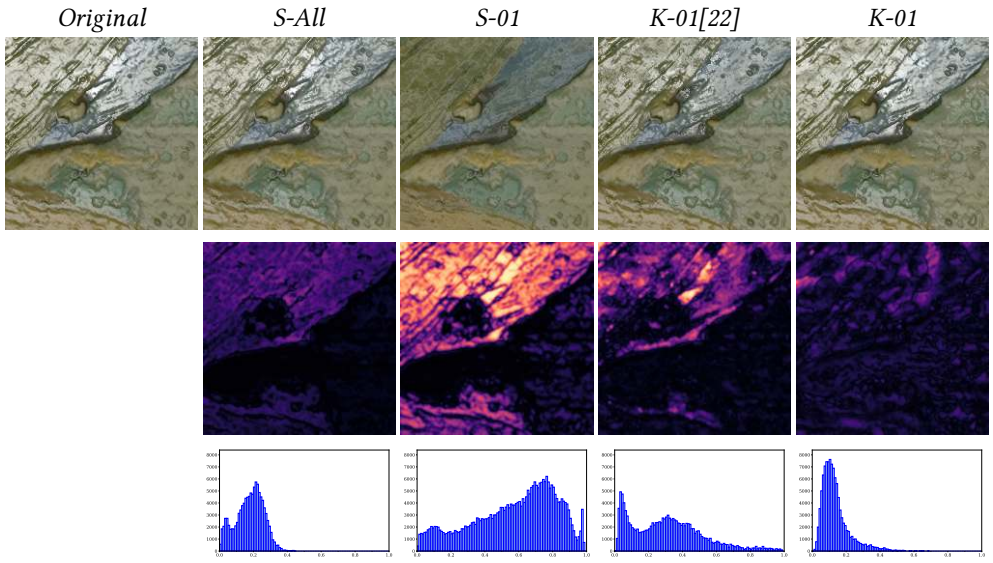


Fig. 2. Synthetic test (example 1). We compare the original image (first column) with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-01*, *K-01*[22], and our *K-01* while removing from the MLIC the orginal image. The original and rendered images are in the first row, while in the last two rows we present the corresponding ∃LIP maps and statistics.

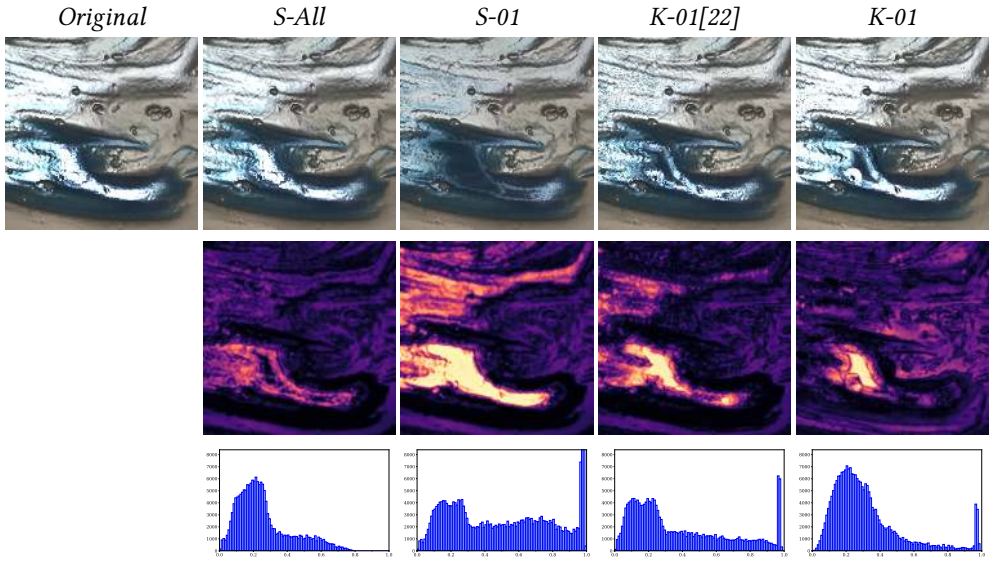


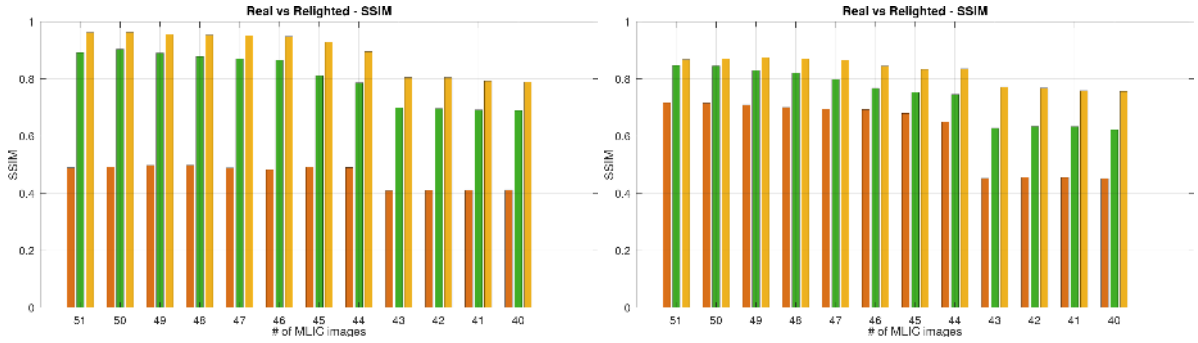
Fig. 3. Synthetic test (example 2). We compare the original image (first column) with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-01*, *K-01*[22], and our *K-01* while removing from the MLIC the orginal image. The original and rendered images are in the first row, while in the last two rows we present the corresponding ꟻLIP maps and statistics.

Synthetic Dataset	Method	ꟻLIP Statistics			
		Mean	Weighted Median	1st Quartile	3rd Quartile
Paint-Texture-16	*S-All*	0.10	0.19	0.12	0.24
	S-12	0.29	0.63	0.41	0.77
	K-12[22]	0.13	0.30	0.13	0.44
	K-12	**0.09**	**0.11**	**0.07**	**0.16**
Paint-Texture-14	*S-All*	0.16	0.22	0.15	0.32
	S-12	0.28	0.55	0.25	0.85
	K-12[22]	**0.18**	0.27	**0.15**	0.58
	K-12	0.19	**0.25**	0.16	**0.37**

Table 1. ꟻLIP statistics for the two synthetic datasets.

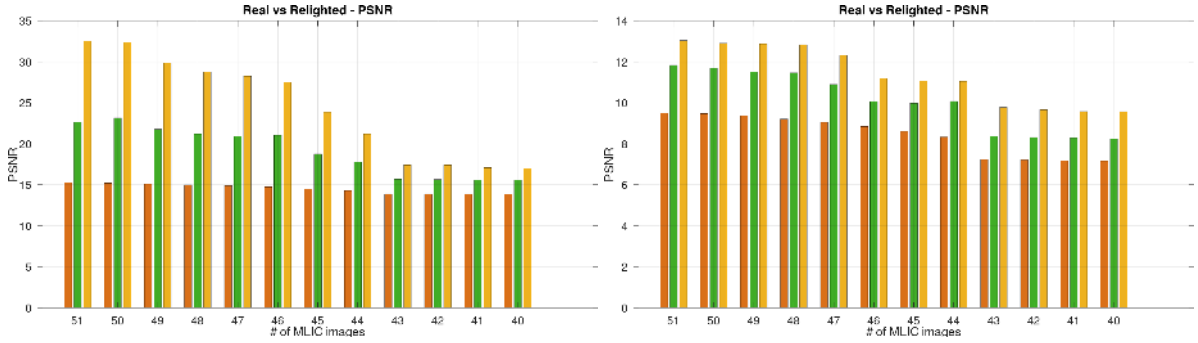
8.1 Quantitative evaluation of reconstructions from sparse MLICs

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 (about 50 for this article) 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 or sixteen images depending on the dataset). For each subset, we compute the SV-BRDF with the single pixel algorithm, with the method proposed by Pintus et al. [22], and with our proposed method; we respectively call *S-X* and *K-X* the single pixel or one of the k-neighbor based computations 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 image pair (reference vs virtual) using three image similarity metrics. First, we use the ꟻLIP map [2], which is a difference evaluator metric that has been shown to approximate well the differences perceived by humans between a rendered image and the corresponding ground truth image. The



(a) *Paint-Texture-16* (b) *Paint-Texture-14*

Fig. 4. Performance on various image counts for synthetic tests (SSIM). We compare the performance of the single pixel strategy *S-X* (orange), our previous method [22] (green), and our solution *K-X* (yellow). 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.

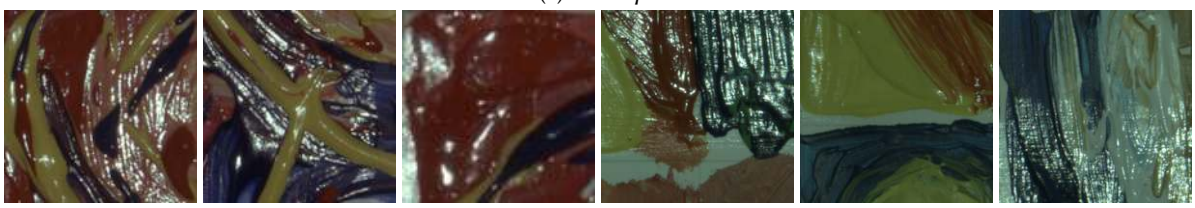


(a) *Paint-Texture-16* (b) *Paint-Texture-14*

Fig. 5. Performance for various image counts for synthetic tests (PSNR) We compare the performance of the single pixel strategy *S-X* (orange), our previous method *K-01*[22] (green), and our solution *K-X* (yellow). The graphs show the worst PSNR value in the relighted MLIC obtained from SV-BRDFs computed after removing from one to twelve images from the original 52 image MLIC.

main idea of ꟻLIP is to present the perceptual difference seen by humans in the particular condition of alternating between two perfectly superimposed images without blank images in between. The metric is thus designed with the goal of discarding overall image differences that, whether numerically high or low, cannot be perceived by the humans, while highlighting differences in very small image regions with point-like shapes, or differences due to edges rather than changes in color magnitude and chromaticity. These perceptual results are complemented by results that quantify image quality reproduction with the Structural Similarity Index (SSIM) [36]. Finally, we also use the PSNR metric to numerically evaluate the accuracy of the computed results. Given the large amount of tests performed, we show the ꟻLIP maps and statistics for some particular and challenging images in the MLIC, while for SSIM and PSNR, as a final value for the quality of the computed SV-BRDF, we summarize results by taking the worst SSIM and PSNR value among all the real vs. virtual image comparisons in the MLIC. In addition,

(a) *Mockups*



(b) *Mockup #0* (c) *Mockup #1* (d) *Mockup #2* (e) *Mockup #3* (f) *Mockup #4* (g) *Mockup #5*

Fig. 6. A series of painting mockups acquired with a RTI dome acquisition setup. Six image crops, with different material and geometrical characteristics, are used to test the performance of the proposed SV-BRDF modeling method.

we also present additional visual comparisons among the relightings obtained by the *S-All*, the *S-X*, and the *K-X* approaches.

8.1.1 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 [33], since they exhibit a detailed shape and appearance similar to the type of real objects we are interested in. As the data is provided through *Diffuse*, *Bump*, and *Normal* maps, we take the *Diffuse* as is, generate the *Gloss* component by converting the *Bump* map to monochrome and rescaling it, and assign a constant highlight color. These maps were used to create a synthetic MLICs using a fixed camera and 52 directional lights. Fig. 2 (top left) shows one original image from the synthetic MLIC of the *Paint-Texture-16*, together with the same image virtually relighted from the SV-BRDF computed with the

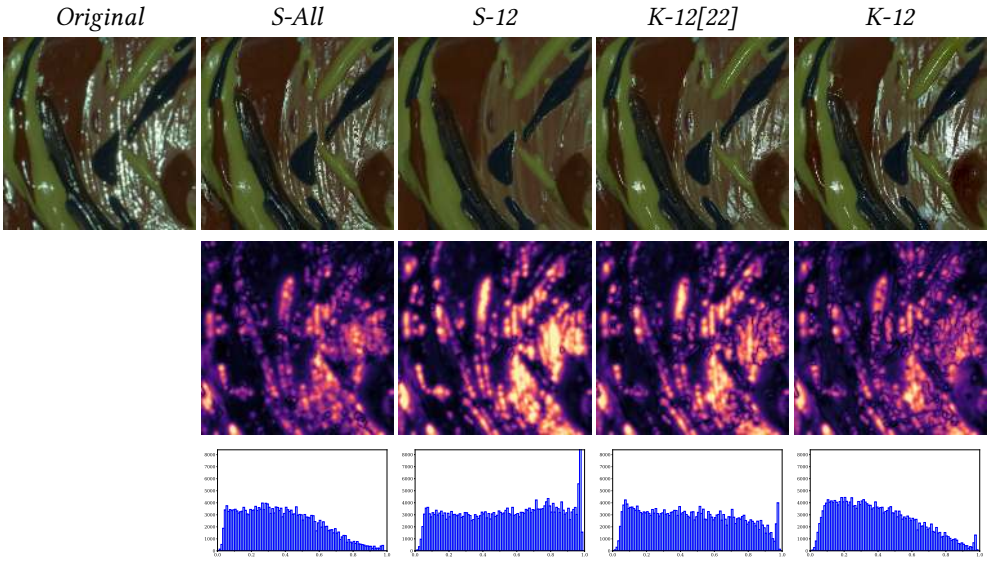


Fig. 7. Painting mockup #0. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-12*, *K-12*[22], and our *K-12*. For the last three virtual renderings, we removed twelve images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

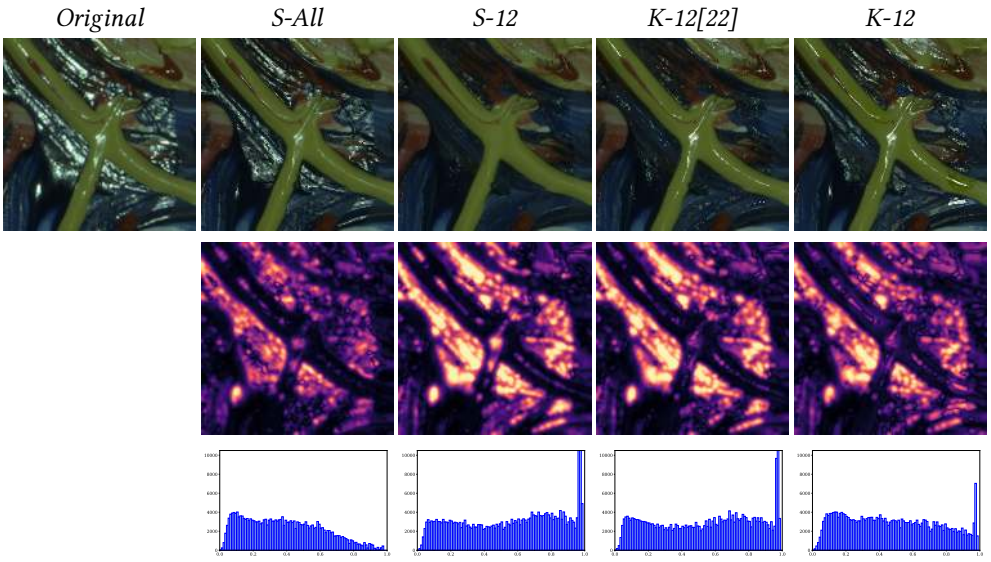


Fig. 8. Painting mockup #1. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-12*, *K-12*[22], and our *K-12*. For the last three virtual renderings, we removed twelve images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

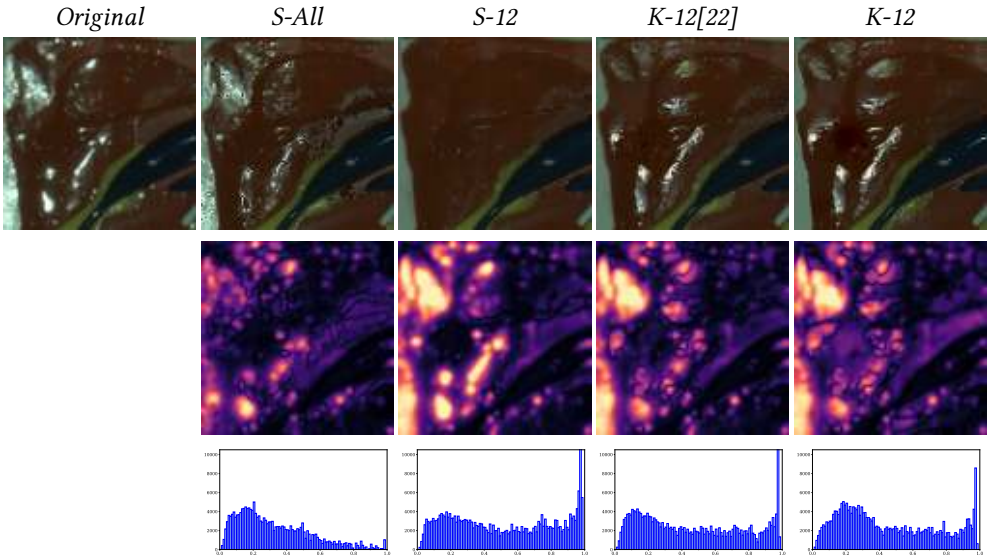


Fig. 9. Painting mockup #2. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-12*, *K-12*[22], and our *K-12*. For the last three virtual renderings, we removed twelve images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

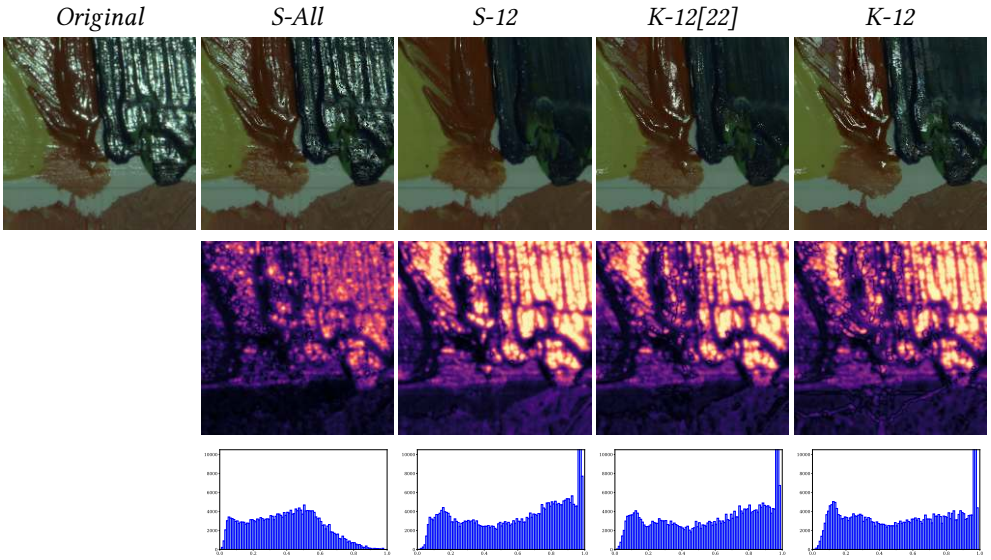


Fig. 10. Painting mockup #3. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-12*, *K-12*[22], and our *K-12*. For the last three virtual renderings, we removed twelve images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

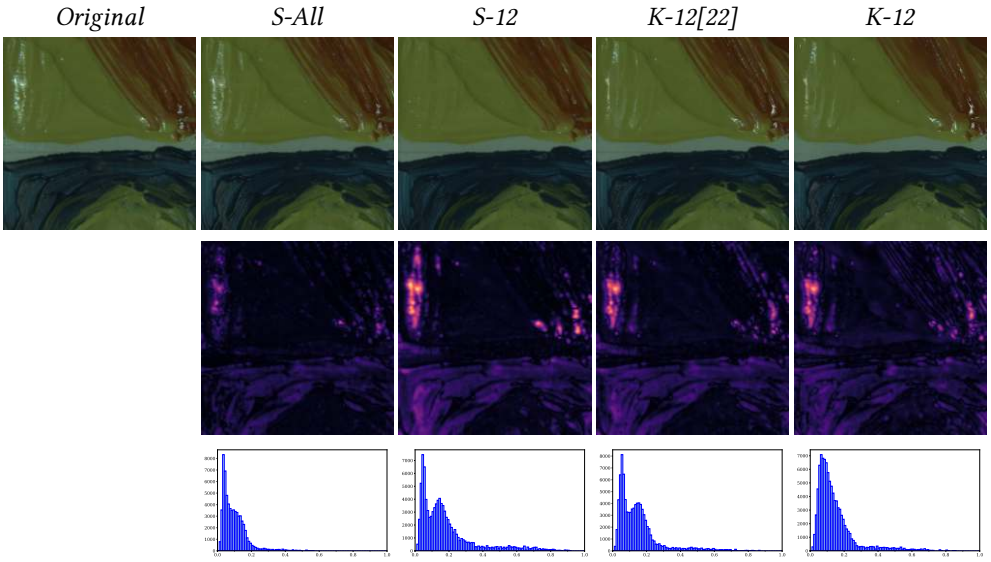


Fig. 11. Painting mockup #4. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-12*, *K-12*[22], and our *K-12*. For the last three virtual renderings, we removed twelve images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

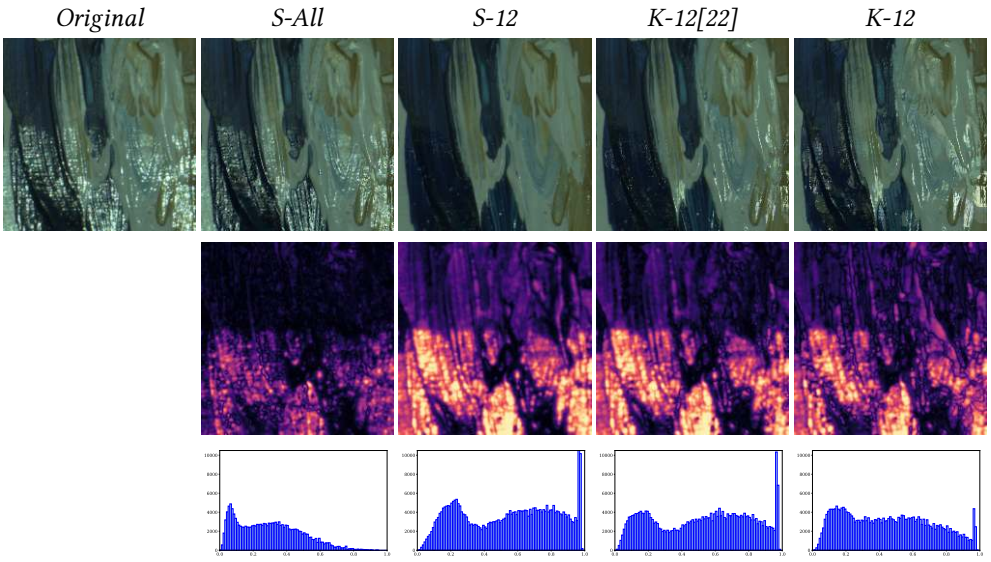


Fig. 12. Painting mockup #5. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-12*, *K-12*[22], and our *K-12*. For the last three virtual renderings, we removed twelve images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

entire MLIC and the single pixel algorithm *S-All* (second row). The selected image exhibits both diffuse and specular behavior. We compare the virtual relighting after computing the SV-BRDF by removing just this image in Fig. 2 (top left) from the MLIC, and by applying the single pixel fitting (*S-01*), our previously presented approach (*K-01* [22]), and the proposed solution (*K-01*). Even with this slightly sparser input (one less image), it is clear how the single-pixel method (*S-01*) column) misses many of the glossy areas, because of insufficient angular sampling. Including information from the neighboring 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 (fourth and fifth column). Moreover, our edge-preserving strategy properly keeps the sharpness of the original image. In the second and third rows of Fig. 2, we present the corresponding ꟻLIP maps, which illustrate how our approach improves the perceptual performances with respect to both the single pixel method and to our previous approach [22]. A similar behavior can be seen in the second synthetic dataset (Fig. 3). The highlight signal in the blue part is completely lost with the standard procedure (*S-01*), while it is largely recovered by our approach (*K-01*, last column). ꟻLIP maps confirm what we perceive in the first row. Tab. 1 numerically presents the ꟻLIP statistics in terms of mean, weighted median, and 1st and 3rd quartile of the ꟻLIP values. Highlighted in bold are the best values, which mostly belong to the approach proposed here. We also quantitatively compare the improvement of our algorithm compared to the standard single pixel technique and the previous neighborhood-based approach [22] by reporting the SSIM and PSNR metrics. Fig. 4 and Fig. 5 respectively show the SSIM and PSNR 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 others, especially for the sparser models.

Painting Mockup	Method	ꟻLIP Statistics			
		Mean	Weighted Median	1st Quartile	3rd Quartile
Mockup #0	*S-All*	0.21	0.36	0.20	0.53
	S-12	0.30	0.56	0.30	0.79
	K-12[22]	0.26	0.45	0.24	0.69
	K-12	**0.24**	**0.37**	**0.21**	**0.56**
Mockup #1	*S-All*	0.21	0.36	0.19	0.55
	S-12	0.30	0.60	0.30	0.83
	K-12[22]	0.29	0.58	0.28	0.81
	K-12	**0.27**	**0.45**	**0.24**	**0.69**
Mockup #2	*S-All*	0.17	0.27	0.15	0.45
	S-12	0.28	0.53	0.24	0.83
	K-12[22]	**0.23**	0.43	**0.20**	0.75
	K-12	0.25	**0.38**	0.21	**0.70**
Mockup #3	*S-All*	0.22	0.40	0.22	0.54
	S-12	0.37	0.70	0.35	0.91
	K-12[22]	0.34	0.68	0.34	0.90
	K-12	**0.33**	**0.58**	**0.28**	**0.84**
Mockup #4	*S-All*	0.06	0.08	0.04	0.13
	S-12	**0.09**	0.14	**0.06**	0.22
	K-12[22]	**0.09**	0.12	**0.06**	0.18
	K-12	**0.09**	**0.11**	0.07	**0.17**
Mockup #5	*S-All*	0.15	0.28	0.13	0.43
	S-12	0.34	0.57	0.28	0.79
	K-12[22]	0.29	0.55	0.26	0.76
	K-12	**0.27**	**0.44**	**0.22**	**0.66**

Table 2. ꟻLIP statistics for the six real painting mockups.

8.1.2 Real-world painting mockups. In order to evaluate our solution in a real-world, but controlled, scenario, we consider six painting mockups with heterogeneous spatially-varying material distribution over the surface (Fig. 6). Mockups were realized on painting paper, with standard acrylic colors. Our aim was to have different kind of color mixtures and geometries, using different painting styles. Regarding color mixtures, we ranged from fully

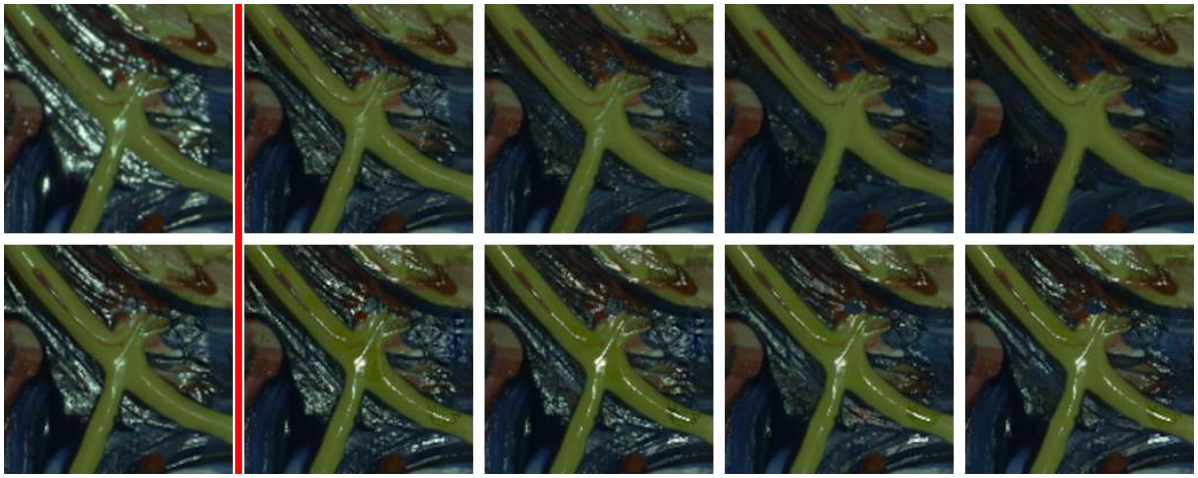
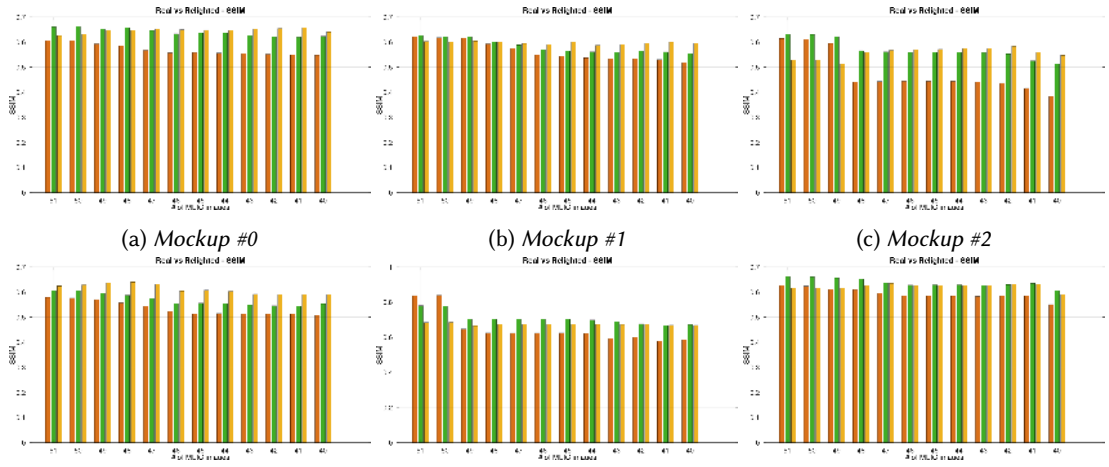


Fig. 13. Relighting of painting mockup #1. From column two to five, we visually compare the relighting results obtained with the *S-X* (Up) and *K-X* (Bottom) method, by removing (from left to right) 1, 5, 11, and 12 images. On the first column we show the original image (Top) and the result of *S-All* (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.

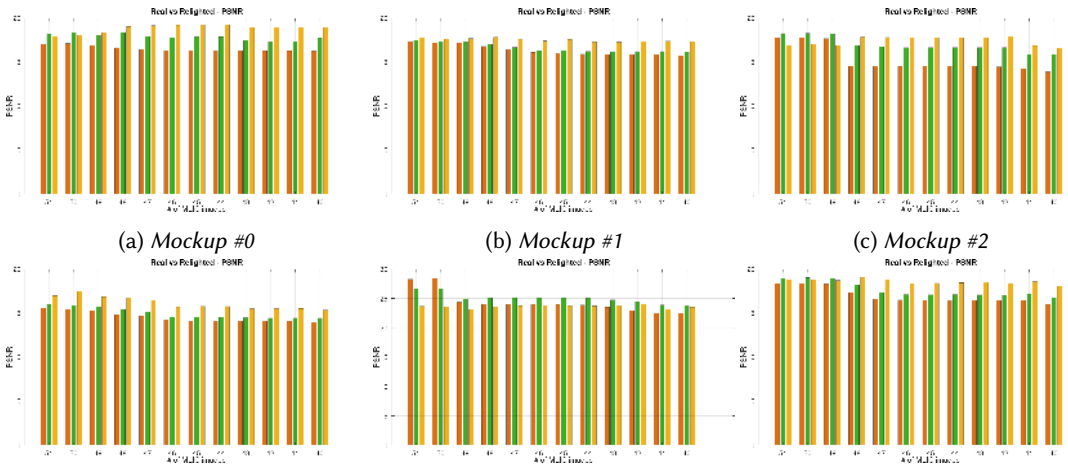


(a) *Mockup #0* (b) *Mockup #1* (c) *Mockup #2*

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

Fig. 14. Performance for various image counts on painting mockups (SSIM). We compare the performance of the single pixel strategy *S-X* (orange), our previous method *K-X*[22] (green), and our solution *K-X* (yellow). 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.

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. Finally, when



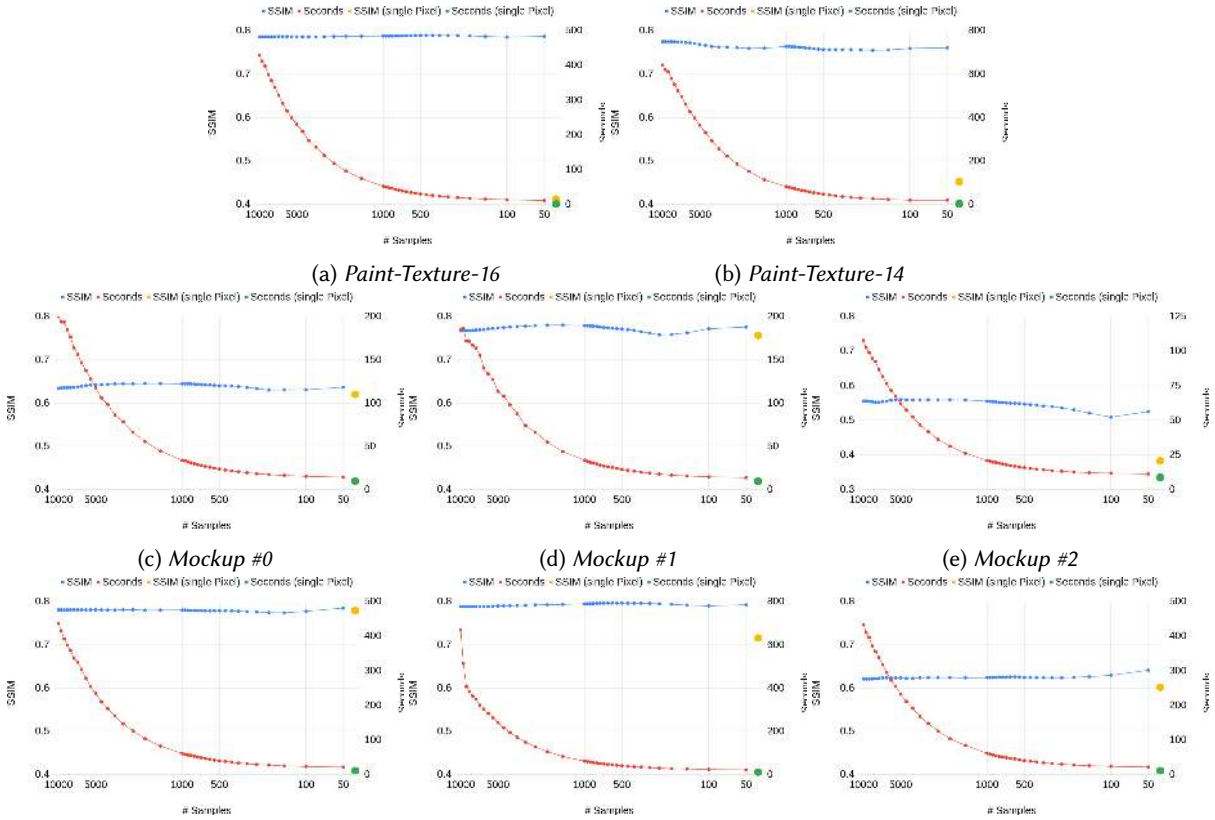
(a) *Mockup #0* (b) *Mockup #1* (c) *Mockup #2*

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

Fig. 15. Performance for various image counts on painting mockups (PSNR). We compare the performance of the single pixel strategy *S-X* (orange), our previous method *K-X*[22] (green), and our solution *K-X* (yellow). The graphs show the worst PSNR 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.

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 the mockups have been acquired by a custom light dome with a radius 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 with the virtually relighted ones, using the *S-All*, *S-12*, *K-12* with the method of Pintus et al. [22], and our *K-12* approaches. Again, for the last three we removed from the MLIC twelve images. From Fig. 7 to Fig. 12 we present the relighted images, together with the ꟻLIP maps and statistics. Tab. 2 summarizes ꟻLIP statistics. Images, error maps and numbers confirm that the presented method is more capable than others in evaluating SV-BRDF for relighting purposes, even in extremely sparse sampling condition. Moreover, Fig. 13 shows that the removal of even one original image may have 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-12* results exhibit many of the highlights present in the original photograph, while the *S-12* strategy results in a more diffuse surface. In particular, in the first column of Fig. 13, 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. We can see how, even with twelve discarded images, our approach keeps the glossy signal, 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 and PSNR statistics in Fig. 14 and Fig. 15, computed with the single pixel, the technique in Pintus et al.[22], and the proposed solution. Our method can generally provide a better and more stable performance than the others, even when the input MLIC provides a really sparse BRDF sampling.



(a) *Paint-Texture-16* (b) *Paint-Texture-14*

(c) *Mockup #0* (d) *Mockup #1* (e) *Mockup #2*

(f) *Mockup #3* (g) *Mockup #4* (h) *Mockup #5*

Fig. 16. Speed-up vs Quality. We launched the SV-BRDF reconstruction with different levels of pruning, i.e., from a sample budget of 10000 to 50 samples. While the computational time decreases as expected, our pruning strategy removes only redundant samples, and preserves the BRDF sampling coverage without affecting the final reconstruction quality. Green and yellow dots are respectively the time and SSIM computed in the same condition by the single pixel approach. Our pruning strategy performs a clever sample selection, so that we can achieve a computation time similar to the single pixel method, while producing better reconstruction results.

8.1.3 Speed-up vs Quality. In order to analyze the behavior of our pruning technique (Sec. 7), we repeated the reconstruction for all the presented synthetic and real-world samples by keeping the neighborhood radius constant ($R = 10$, leading to a window of $21x21$ pixels), and varying the pruning budget B from very little pruning ($B = 10000$) to aggressive pruning ($B = 50$). For all the runs, we remove the top 12 images from the original MLIC, i.e., those with more front illumination, or, in other words, those that sample the more high-frequency part of the BRDF field. We present both the computational times and the SSIM metric computed with the first removed image, i.e., that with the most frontal illumination, which typically contains more highlights. We also compare those resulting values with the single pixel approach. The results summarized in Fig. 16 confirm the fact that our pruning strategy drastically reduce computational cost while keeping the accuracy constant. As the sample budget decreases, the computation time tends to that of the single pixel approach (green dot). It is interesting to note that, with a similar number of samples, the proposed method generally provides an improved SSIM with

respect to the one produced by the single pixel method (yellow dot). This is because the pruning approach is designed to select samples that are better distributed than the samples from the central pixel (i.e., the sample is likely to replace redundant diffuse samples with specular samples from the neighborhood).

8.2 Qualitative evaluation

As a complement to our experiments with metrics that approximate the image difference perceived by humans, 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.

8.2.1 Goal. The user tests aim at comparing our neighbor-aware method (*K-X*) and the single pixel approach (*S-X*). In performing this comparison, we are interested in answering two questions. First, we want to investigate whether relightings from our models are perceived more similar to a reference ground-truth image than the ones derived from single-pixel reconstructions. This test serves to further validate the results obtained with perceptual image diffeerence methods. Second, we would also like to assess whether, without any other information or reference image, relightings of models reconstructed with our method appear to users more or less realistic than the ones derived from single-pixel reconstructions. This test serves to detect whether, by merging data from multiple pixels we create some unnatural situations that are perceived by users, or, conversely, whether those situations are created by single-pixel approaches.

(a) *Static content without reference* (b) *Static content with reference*

Fig. 17. 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.

8.2.2 Setup. The test is conducted through a web-based anonymous questionnaire carried out by a number of volunteers. The questionnaire consists in a generic and eight specific sections. The first section characterizes users, while 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 sections include both static and dynamic data. Static data consists in rendered images, 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. Both 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. 8.1).

8.2.3 Tasks. Each user was first presented with two synced parallel videos (or static images) without any reference (e.g., Fig. 17a). 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. 17b), and users are asked to choose the option that looks more similar to the reference. In all these

scenarios, the users are requested to select one option, or to mark the test as undecided. We set limitless time for the experiment, as we want users to inspect and select their options carefully.

8.2.4 Participants. We recruited 105 (14 males and 11 females, with ages ranging from 20 to 80, median 31 years) among university students and their families and researchers working at our center and at collaborating institutions. Participants include humanities and STEM students, art historians, curators, conservation scientists, restorers, researchers in computer science, software engineers, technicians, students, teachers, professors, retired people from different backgrounds, and others (e.g., attorneys, managers, counsellors, bioengineers, administrative staff, etc.). About 13% do not use a computer regularly. This wide 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.

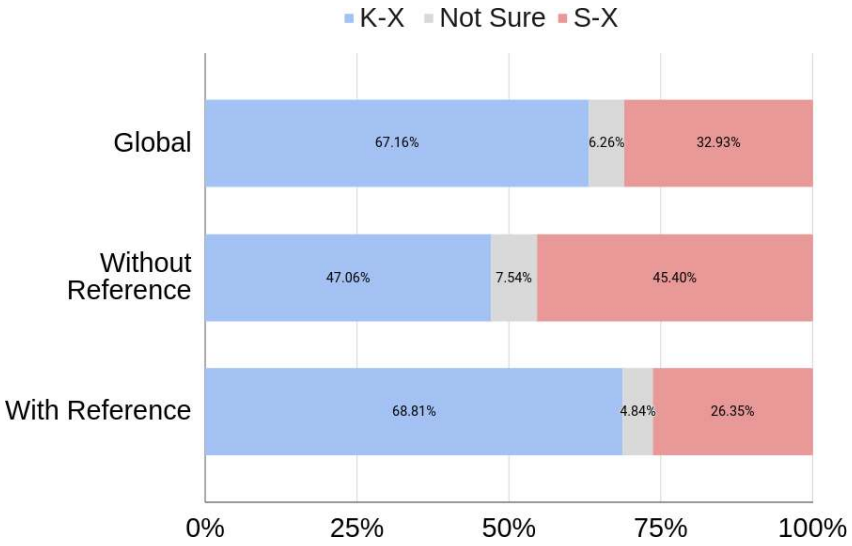


Fig. 18. We present here the global scores across all types of user tests, the scores of only the tests without and with reference. Globally, more than two third of the votes go to the proposed *K-X* solution, while about 30% to the standard single pixel approach *S-X*. Although the proposed solution 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 or the participant couldn't make a clear choice.

8.2.5 Evaluation of results. We start by analysing the scores provided by the participants in the two main sections of our test, which reflect the cumulative results related to our two goals. Beside providing a global cumulative statistics across all the different types of visual tests (just for an overview and for completeness), the important information conveyed by Fig. 18 is that, when no clue has been provided (without reference case), both methods got similar scores, with a 7% of uncertainty, while the choice between *K-X* and *S-X* is more clear when we provide a reference video or image (Fig. 18, third row). This means that both *S-X* and *K-X* approaches are capable of producing similar rendered images, with the same level of visual quality. Conversely, when compared to the original photograph, the proposed method has been recognized as the most reliable in resembling the real-world visual outcome. Since these values are cumulative (videos and images together), Fig. 19 split the four cases, i.e., video with and without reference, and images with and without reference. Here, the statistics confirm that,

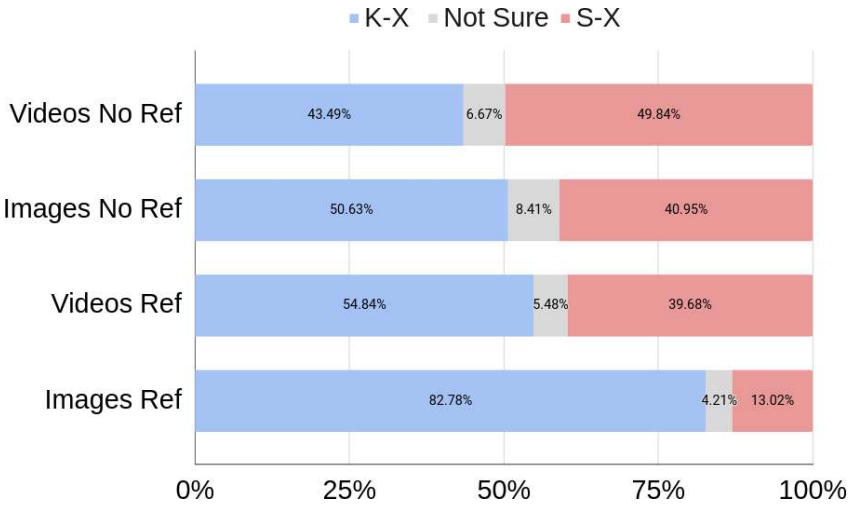
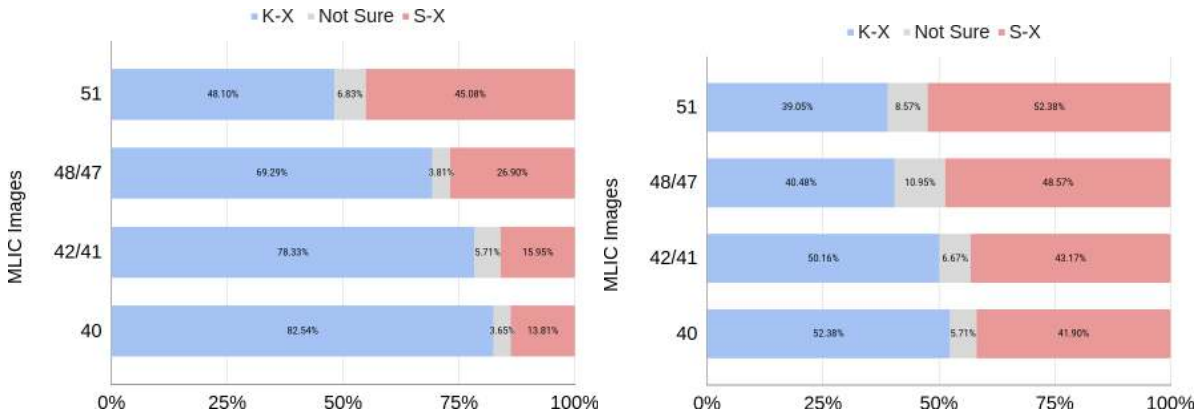


Fig. 19. Comparison between statistics related to only dynamic (videos) or static (images) tests, and, for each of them, the case with and without reference video/image. Two trends are clear here. Without reference both methods are capable of render convincing images, while with reference the proposed method produces more similar images. In addition, the statistics show how it is more difficult to spot differences in the dynamic test than in static rendered images. When the user is more confident about the answer, the votes for the proposed *K-X* method get to more than 80%.



(a) *With reference* (b) *Without reference*

Fig. 20. 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 case with the reference image (a), in the first row the uncertainty is the highest and the two methods get similar results. The more images we remove, the more the user appreciates the results produced by the proposed algorithm, and exhibits a lower uncertainty. Conversely, without a reference video or images, the two results are substantially similar, with a higher level of user uncertainty.

without reference, the two compared videos, or the two images, got similar scores. More interesting is the outcome of the experiment with the reference image or video. In static tests, the user has a higher chance to better analyze the subtle differences between the two rendering, and the statistics show how the proposed solution got up to

82% support from the participants, while the score of the *S-X* approach is just about 13%. Conversely, in the dynamic, but not user-controlled, comparison, it is more difficult to spot subtle difference while the video is playing, so that, although with reference, the *K-X* and the *S-X* solutions respectively take 54% and 39% scores. In the last graph (Fig. 20, 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. Fig. 20a shows that, as expected, 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 82% when we remove twelve images from the original MLIC. Although the case without reference shows again that the participants gave a similar score to both methods, nonetheless Fig. 20b exhibits a similar trend, when the more images we remove, the more the proposed solution has been selected as the best. It does seem that, by borrowing data from neighboring pixels, we do not create situations that lead to images perceived as unnatural. 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 more photo-realistic, with significant improvements with respect to single-pixel approaches.

Retablo	Method	ꟻLIP Statistics			
		Mean	Weighted Median	1st Quartile	3rd Quartile
Retablo Crop #0	*S-All*	0.29	0.37	0.26	0.47
	S-16	0.43	0.56	0.38	0.79
	K-16[22]	0.42	0.55	0.37	0.76
	K-16	**0.41**	**0.50**	**0.36**	**0.67**
Retablo Crop #1	*S-All*	0.38	0.44	0.33	0.55
	S-16	0.59	0.72	0.54	0.86
	K-16[22]	0.58	0.71	0.53	0.85
	K-16	**0.45**	**0.52**	**0.38**	**0.68**
Retablo Crop #2	*S-All*	0.22	0.27	0.19	0.39
	S-16	0.40	0.54	0.36	0.73
	K-16[22]	0.40	0.54	0.36	0.72
	K-16	**0.34**	**0.45**	**0.28**	**0.66**

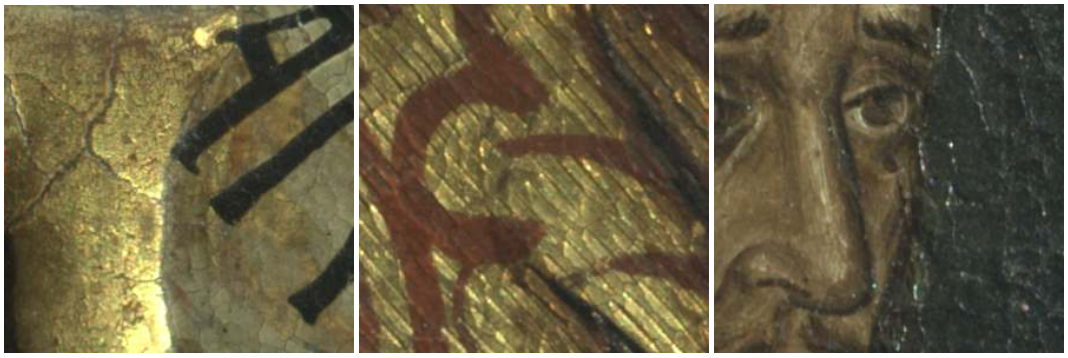
Table 3. ꟻLIP statistics for the three painting crops.

8.3 Painting acquisition, reconstruction, and relighting

We have applied our technique to the reconstruction of BRDFs from free-form acquisitions of a number of paintings. In this article, we report on the results obtained for the capture, reconstruction, and relighting of a panel part of the *retable of St. Bernardino* (1455), a polyptych originally from the chapel of St. Bernardino in the St. Francesco church in Cagliari, Italy, and currently preserved and displayed at the Pinacoteca Nazionale in Cagliari. The analyzed panel, measuring 34x25cm, is painted in oil on a wooden support and depicts the prophet Daniel.

The painting (Fig. 21) has been acquired with a free-form RTI setup, with a 36.3 Mpixels DSLR FX Nikon D810 Camera with a 50*mm* AF Nikkor Lens and a handheld white LED (5500K) covering the entire visible spectrum. We have acquired 60 images for the MLIC. The acquired data has been calibrated with four glossy spheres (for light direction), and with a gray frame positioned around the object (see Fig. 1 left), using the camera and light calibration method recently presented by Pintus et al. [25]. In order to assess the method as we did for synthetic datasets and real mockups, we removed sixteen images from the dataset, thus comparing reconstructions starting from a 44 image MLIC with a reconstruction from a 60 image MLIC.

(a) *Acquired painting (component of the Retable of S. Berndardino)*



(b) *Retablo Crop #0* (c) *Retablo Crop #1* (d) *Retablo Crop #2*

Fig. 21. A painting (panel of the Retable of Saint Bernardino) acquired with a free-form RTI acquisition setup. Three image crops, with different material and geometrical characteristics, are used to test the performance of the proposed SV-BRDF modeling method.

To avoid blurring the results by averaging over the entire painting statistics for areas with different features, we focus on three image subregions/crops (Fig. 21), selected for their characteristics. Crop#0 includes a gold-colored

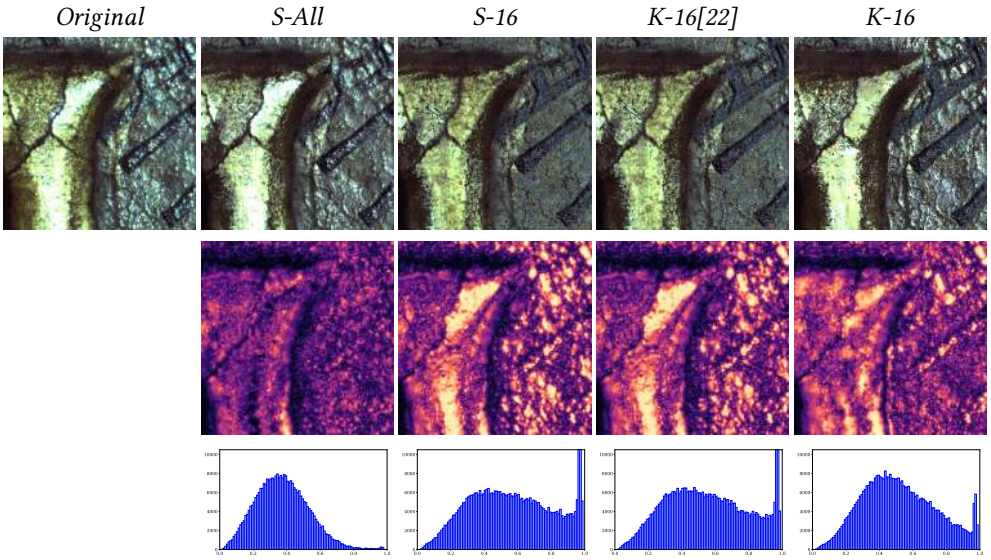


Fig. 22. Retablo crop #0. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-16*, *K-16*[22], and our *K-16*. For the last three virtual renderings, we removed sixteen images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

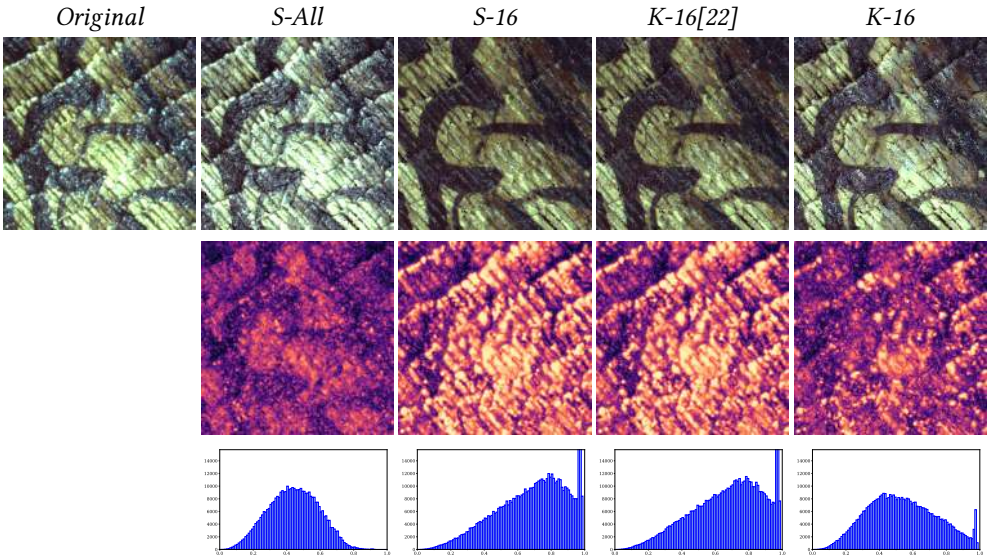


Fig. 23. Retablo crop #1. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-16*, *K-16*[22], and our *K-16*. For the last three virtual renderings, we removed sixteen images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

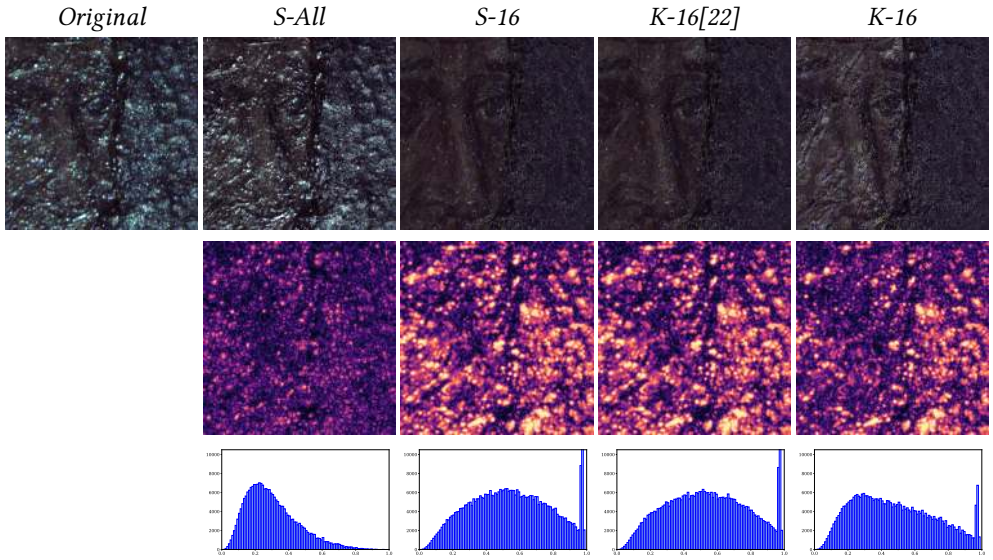


Fig. 24. Retablo crop #2. We compare the original image with its virtual relighting obtained from the SV-BRDF computed with *S-All*, *S-16*, *K-16*[22], and our *K-16*. For the last three virtual renderings, we removed sixteen images from the MLIC. In the last two rows we present the corresponding ꟻLIP maps and statistics.

Fig. 25. Exploration of a painting (panel of the polyptych retable of Saint Bernardino (1455), Cagliari, Italy). The user explores the dataset using visualization lenses and relighting on a large multitouch screen. Left: the outside of the lens shows the normal map, while the inside shows a relighting from SV-BRDF. Right: the user interactively controls a spot light to relight the painting.

relief with several cracks and a flat white area. Crop#1 is a decoration detail in a mostly flat area with cracks, and Crop#2 is a detail of the face, with lots of paint mixing and several damages visible.

For each image crop we compare one original image with the virtually relighted ones, by using the *S-All*, *S-16*, *K-16* with the method of Pintus et al. [22], and our *K-16* approaches. Fig. 22, Fig. 23, and Fig. 24 show the rendered images with the ꟻLIP maps and statistics, and, as before, we also provide a table (Tab. 3) with the corresponding numerical values of the mean, weighted median, 1st and 3rd quartile of the error. As shown, for all the tested image crops, the proposed solution is capable of better retrieving the SV-BRDF than previous approaches, and produces relightable images that have a lower perceptual error while using a small number of images. This makes it possible to apply the method in practice, as it reduces capture time by generating higher quality results with sparse captures.

Fig. 25 shows two frames from an interactive exploration sequence using a lens-based user interface [4]. In the left image, the user interactively controls a lens that shows the normal map outside the lens, and the relighted image inside it. In the right image, the user interactively relights the model by controlling a virtual spot light.

9 CONCLUSIONS

We have presented a practical solution for exploiting information from neighboring pixels to efficiently infer subtle per-pixel analytical 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, reusing existing calibration and normal computation components and just replacing BRDF fitting.

In order to provide a controlled evaluation of the approach with quantitative measures compared to ground truth, we have performed tests using synthetic data, for which ground truth is available, as well as with painting mockups of various characteristics acquired using a light dome providing calibrated capture. We have, moreover, applied the method to a real-world use case of free-form acquisition of a painting.

Our qualitative and quantitative results show 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.

As the current approach can be used as a plug-in replacement of single-pixel BRDF fitting, we have integrated it into an existing end-to-end pipeline for capture, calibration, and model reconstruction. The pipeline is mainly used for generating models for an interactive system geared towards visual inspection, starting from rapid free-form acquisitions with a moderate number of photos. We are currently applying it in collaboration with cultural institutions for the analysis of paintings.

Besides improving our proof-of-concept implementation, our current work is focused on further optimizing the approach by dynamically adapting neighborhood size prior to the computation and pruning of weights, to avoid wasting computational efforts on data that is discarded during pruning. With this approach, we aim at further reducing the gap between single-pixel and multi-pixel. We are also working on further refining the similarity metric and on testing the behavior with more complex multi-lobe BRDFs. Finally, we plan to extend the concept of exploiting local similarity for the computation of other high frequency representations (e.g., high-order HSH), as well as for the joint optimization of normals and materials.

In addition, it should be noted that this work has focused on how to better exploit the data acquired for a given multi-light acquisition to produce a better result than current single-pixel approaches, rather than on determining what is the minimum amount of images in the MLIC that can provide a meaningful BRDF reconstruction.

This is because this number strongly depends both on the distribution of the light constellation in the MLIC and on the normal distribution in the object itself and in the neighborhood. Several papers (for instance, the work of Nielsen et al. [21]) have studied the optimal minimal BRDF sampling in a single pixel approach, so we can consider their outcomes as a maximum value for the minimum amount of required images if the capture setup is controllable. It could be worth performing another completely independent investigation about how our

method can decrease that value further. Nonetheless, our results are directly applicable to improve the quality of results even in very common cases where the capture setup is inflexible (e.g., fixed light domes) or where it is not possible to precisely position the lights at prescribed angles (e.g., when using free-form manual capture).

ACKNOWLEDGMENTS

The project received funding from the European Union's H2020 research and innovation programme under grant 813170 (EVOCATION), and from Sardinian Regional Authorities under projects VIGECLAB and SVDC. We thank the Pinacoteca Nazionale in Cagliari for access to artworks for the purpose of digitization, and for their collaboration to the project. The authors thank in particular Gerlinde Tautschnig and Manuela Puddu for their continuous support.

REFERENCES

[1] Miika Aittala, Tim Weyrich, and Jaakko Lehtinen. 2013. Practical SVBRDF capture in the frequency domain. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 110–1.

[2] Pontus Andersson, Jim Nilsson, Tomas Akenine-Möller, Magnus Oskarsson, Kalle Åström, and Mark D. Fairchild. 2020. ꟻLIP: A Difference Evaluator for Alternating Images. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 3, 2 (2020), 15:1–15:23. https://doi.org/10.1145/3406183

[3] Ronen Basri, David Jacobs, and Ira Kemelmacher. 2007. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision* 72, 3 (2007), 239–257.

[4] Fabio Bettio, Moonisa Ahsan, Fabio Marton, and Enrico Gobbetti. 2021. A novel approach for exploring annotated data with interactive lenses. *Computer Graphics Forum* 40, 3 (2021), 387–398. https://doi.org/10.1111/cgf.14315

[5] CHI. 2019. Cultural Heritage Imaging website. http://culturalheritageimaging.org [Online; accessed-July-29-2021].

[6] Mark S Drew, Yacov Hel-Or, Tom Malzbender, and Nasim Hajari. 2012. Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing* 30, 4-5 (2012), 317–331.

[7] Arne Dür. 2006. An Improved Normalization for the Ward Reflectance Model. *Journal of Graphics Tools* 11, 1 (2006), 51–59.

[8] Shireen Y Elhabian, Ham Rara, and Aly A Farag. 2011. Towards accurate and efficient representation of image irradiance of convex-Lambertian objects under unknown near lighting. In *Proc. ICCV*. IEEE, 1732–1737.

[9] Andrea Giachetti, Irina Mihaela Ciortan, Claudia Daffara, Giacomo Marchioro, Ruggero Pintus, and Enrico Gobbetti. 2018. A novel framework for highlight reflectance transformation imaging. *Computer Vision and Image Understanding* 168 (2018), 118–131.

[10] Dar'ya Guarnera, Giuseppe Claudio Guarnera, Abhijeet Ghosh, Cornelia Denk, and Mashhuda Glencross. 2016. BRDF representation and acquisition. *Computer Graphics Forum* 35, 2 (2016), 625–650.

[11] Zhuo Hui and Aswin C Sankaranarayanan. 2016. Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 10 (2016), 2060–2073.

[12] Jacek Jankowski and Martin Hachet. 2013. A Survey of Interaction Techniques for Interactive 3D Environments. In *Eurographics STAR*.

[13] Alberto Jaspe Villanueva, Moonisa Ahsan, Ruggero Pintus, Andrea Giachetti, and Enrico Gobbetti. 2021. 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. https://doi.org/10.1145/3430846

[14] Donald R Jones and Joaquim RRA Martins. 2021. The DIRECT algorithm: 25 years Later. *Journal of Global Optimization* 79, 3 (2021), 521–566.

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

[16] Guillaume Lavoué, Nicolas Bonneel, Jean-Philippe Farrugia, and Cyril Soler. 2021. Perceptual quality of BRDF approximations: dataset and metrics. *Computer Graphics Forum* 40, 2 (2021), 327–338.

[17] Jason Lawrence, Aner Ben-Artzi, Christopher DeCoro, Wojciech Matusik, Hanspeter Pfister, Ravi Ramamoorthi, and Szymon Rusinkiewicz. 2006. Inverse shade trees for non-parametric material representation and editing. In *ACM Transactions on Graphics (TOG)*, Vol. 25-3. 735–745.

[18] Lindsay William Macdonald. 2015. *Realistic visualisation of cultural heritage objects.* Ph.D. Dissertation. UCL (University College London).

[19] Tom Malzbender, Dan Gelb, and Hans Wolters. 2001. Polynomial texture maps. In *Proc. ACM SIGGRAPH*. 519–528.

[20] Addy Ngan, Frédo Durand, and Wojciech Matusik. 2005. Experimental analysis of BRDF models. In *Rendering Techniques*. 117–126.

[21] Jannik Boll Nielsen, Henrik Wann Jensen, and Ravi Ramamoorthi. 2015. On optimal, minimal BRDF sampling for reflectance acquisition. *ACM Transactions on Graphics (TOG)* 34, 6 (2015), 1–11.

[22] Ruggero Pintus, Moonisa Ahsan, Fabio Marton, and Enrico Gobbetti. 2021. Exploiting Neighboring Pixels Similarity for Effective SV-BRDF Reconstruction from Sparse MLICs. In *The 19th Eurographics Workshop on Graphics and Cultural Heritage*. 103–112. https://doi.org/10.2312/gch.20211412

[23] Ruggero Pintus, Tinsae Dulache, Irina Ciortan, Enrico Gobbetti, and Andrea Giachetti. 2019. State-of-the-art in Multi-Light Image Collections for Surface Visualization and Analysis. *Computer Graphics Forum* 38, 3 (2019), 909–934.

[24] Ruggero Pintus, Andrea Giachetti, Giovanni Pintore, and Enrico Gobbetti. 2017. Guided Robust Matte-Model Fitting for Accelerating Multi-light Reflectance Processing Techniques. In *Proc. British Machine Vision Conference*. 32.1–32.15. https://doi.org/10.5244/C.31.32

[25] Ruggero Pintus, Alberto Jaspe Villanueva, Antonio Zorcolo, Markus Hadwiger, and Enrico Gobbetti. 2021. A Practical and Efficient Model for Intensity Calibration of Multi-Light Image Collections. *The Visual Computer* 37, 9 (2021), 2755–2767.

[26] Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Simon-Frederic Desage, Serge Samper, and Maurice Pillet. 2017. 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.

[27] Federico Ponchio, Massimiliano Corsini, and Roberto Scopigno. 2018. A compact representation of relightable images for the web. In *Proc. ACM Web3D*. 1–10.

[28] Gilles Rainer, Wenzel Jakob, Abhijeet Ghosh, and Tim Weyrich. 2019. Neural BTF Compression and Interpolation. *Computer Graphics Forum* 38, 2 (2019), 235–244.

[29] Peiran Ren, Yue Dong, Stephen Lin, Xin Tong, and Baining Guo. 2015. Image based relighting using neural networks. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 111:1–111:12.

[30] Szymon M Rusinkiewicz. 1998. A new change of variables for efficient BRDF representation. In *Rendering techniques*. 11–22.

[31] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. 2016. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 3707–3716.

[32] Tiancheng Sun, Henrik Wann Jensen, and Ravi Ramamoorthi. 2018. Connecting Measured BRDFs to Analytic BRDFs by Data-Driven Diffuse-Specular Separation. *ACM Trans. Graph.* 37, 6, Article 273 (Dec. 2018), 15 pages.

[33] Every Texture. 2021. 3D Texture Database. http://https://everytexture.com/ [Online; accessed-July-29-2021].

[34] David Tingdahl, Christoph Godau, and Luc Van Gool. 2012. Base materials for photometric stereo. In *Proc. ECCV*. 350–359.

[35] Chris Tofallis. 2009. Least squares percentage regression. *Journal of Modern Applied Statistical Methods* (2009), 1–12.

[36] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[37] X-Rite. 2020. Scanner TAC7 Total Appearance Capture. https://www.xrite.com/categories/appearance/tac7 [Online; accessed-January-17-2020].

[38] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. 2018. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 126.

[39] Mingjing Zhang and Mark S Drew. 2014. Efficient robust image interpolation and surface properties using polynomial texture mapping. *EURASIP Journal on Image and Video Processing* 2014, 1 (2014), 25.