DIGITAL HERITAGE (2025), pp. 1–10
S. Campana, D. Ferdani, H. Graf, G. Guidi, Z. Hegarty, S. Pescarin, and F. Remondino (Editors)

# A Practical Inverse Rendering Strategy for Enhanced Albedo Estimation for Cultural Heritage Model Reconstruction

R. Pintus[1,2] A. Zorcolo[1] A. Jaspe-Villanueva[3] E. Gobbetti[1,2]

[1] CRS4, Italy
[2] National Research Center in HPC, Big Data, and QC, Italy
[3] KAUST, Saudi Arabia

**Abstract**

*We present a practical single-image framework to address uncontrolled global and local illumination effects in flash photography for improved albedo estimation and color projection onto 3D cultural heritage models. Our approach leverages an inverse rendering pipeline to process a single registered flash photograph and models ambient illumination due to environmental reflections and local interreflections. By compensating for direct and indirect light contributions, we recover a more reliable albedo signal for color projection onto the 3D model. We validate our method through extensive evaluations on two synthetic datasets and real-world acquisitions in conservation and museum settings, demonstrating its effectiveness in improving photometric accuracy and support for relighting, and proper integration of optimized color data into existing 3D models.*

**CCS Concepts**

• ***Computing methodologies*** → ***Reflectance modeling; Computational photography;*** *Appearance and texture representations; Reconstruction;*

## 1. Introduction

The production of high-quality colored 3D digital models plays a crucial role in the documentation, analysis, and dissemination of cultural heritage (CH) assets. These models enable accurate preservation, virtual restoration, and interactive visualization, supporting scholarly research and public engagement [PDC*19, Sco21, FAB*24]. Advances in 3D digitization techniques, including photogrammetry, structured light scanning, and multi-light imaging, have significantly improved the geometric accuracy and color fidelity of digital replicas. However, achieving high spatial resolution and photometric accuracy remains a challenge, particularly when dealing with complex geometries and uncontrolled lighting conditions during acquisition outside laboratory settings.

A particularly challenging scenario arises when shape and color acquisitions are decoupled. This occurs, for instance, for long-term monitoring, where multiple color acquisitions are performed on-site and over time. In this case, once the geometric model has been acquired, only a single or a few images are captured at a particular time, either for efficiency reasons (photogrammetry and multi-light datasets require extensive acquisition time) or when only a small region of an object needs to be updated. In such situations, the newly acquired color data must be seamlessly integrated into an existing 3D model, which is done through 3D image registration and subsequent color projection [PGCD17].

What is contained in the captured photograph, however, is not a surface property, but is influenced by both the surface reflectance and the illumination environment. While projecting the apparent color is often used in many production pipelines [PGCD17], shading and illumination removal is necessary, before projection, to obtain a surface characterization, eventually comparable over time, and to support different applications, such as relighting (i.e., rendering the model under different illumination conditions), which is often used for surface inspection [PDC*19]).

Flash photography in dark environments is commonly used for repeatable surface color extraction and shading removal under controlled lighting. While full reflectance recovery typically requires multiple images (see Sec. 2), under a local illumination prior, surface albedo can be efficiently estimated per pixel from known light intensity, surface normal, and distance to the light. This has enabled the creation of color estimation and projection pipelines that improve over simple apparent color mapping (e.g., [LPC*00, BPV*15]). However, while the purely diffuse reflectance prior is verified for many surfaces of interest in CH, indirect lighting, such as environmental reflections and surface interreflections, violates the local illumination assumption, distorting the projected photometric signal and introducing artifacts. Although dark fabric is sometimes used to minimize environmental reflections, it is not always feasible during on-site captures. Interreflections from concave surface portions, in particular, remain unavoidable.

In this work, we present a practical single-image framework to address indirect global and local illumination effects in flash photography for albedo estimation and projection onto a given 3D geometry. Our approach leverages an inverse rendering pipeline that processes a single registered flash photograph, and it approximates ambient illumination from environmental reflections and local interreflections. By compensating for direct and indirect effects, we recover an image of the pure albedo, free from illumination contributions. This refined signal serves as the input for a more accurate matte color projection onto the 3D model. To validate our method, we conduct extensive evaluations on both synthetic datasets and real-world acquisitions performed in a routine conservation/museum setting.

## 2. Related Work

Color mapping and blending are well-established fields with numerous successful applications, particularly in the CH field. Extensive research explored methods for aligning and integrating color information onto 3D models, ensuring photometric consistency and visual fidelity [PGCD17, PG15]. In this section, we focus on approaches closely related to our work, specifically those addressing material modeling and global/local illumination estimation.

In the CH domain, standard pipelines for creating colored 3D models often involve capturing images under uncontrolled lighting conditions [Rem11], ranging from professional setups with diffuse illumination to in-the-wild photography. The color is directly projected onto 3D models, typically within photogrammetric frameworks [LLC23, FC17, Sch21, C*21], where geometry is derived from dense multi-view stereo [FH15, WWL*21], or through advanced blending algorithms [PGC11]. While widely used, these methods lack a measurable and repeatable color signal. This limitation is particularly critical in CH applications like monitoring and preservation [SBG11], where quantitative surface characterization is essential. Our approach addresses lighting inconsistencies due to an unsupervised capture setup by utilizing controlled flash photography; this ensures a projected color signal that is both more accurate and repeatable. Flash photography has been previously improved for color projections by pipelines that perform a flash characterization (e.g., [DCC*09]), or that remove shading effects under a local illumination prior (e.g., [LPC*00, BPV*15]). We extend those methods to account for indirect illumination effects from ambient contributions from an unknown environment and inter-reflections from the object itself.

When aiming to recover accurate surface appearance, two general strategies are commonly adopted: multi-image methods, particularly based on Multi-Light Image Collections (MLICs), and single-image methods that often rely on recent deep learning frameworks. These approaches attempt to extract not only appearance, but also meaningful optical characteristics and lighting properties.

MLIC techniques [PDC*19] extract both geometry and color parameters using approaches like photometric stereo [JLX*24] or optical modeling [KHM*24]. These typically involve capturing several images from a fixed viewpoint under different lighting directions. Some methods assume uniform or simple materials [ASOS13, AWL13], while others leverage material dictionaries [HS17] or clustering of appearance profiles [TGVG12]. Although effective, these techniques require extensive acquisition and may often assume planar surfaces, making them less suitable for objects with complex geometry. Additionally, recovery methods often neglect global and local indirect illumination effects. In contrast, our method, targeting matte objects as many color projection methods do, uses only a single image, while accounting for both direct and indirect illumination, simplifying acquisition and improving applicability.

Single-image color estimation has gained traction by enabling material inference with minimal acquisition effort [DAD*18, VPS21]. These leverage CNNs, adversarial training, and differentiable rendering [ZGW*23] to infer spatially varying reflectance properties. Hybrid supervision strategies [ZK21] and mobile-friendly pipelines [LSC18] have also been explored. However, these methods heavily rely on training data and often struggle with generalization [SP23], especially when applied to objects with significant geometric variation [GLT*21, SLS23]. Moreover, learned priors can introduce artifacts or physically implausible results [LSBE24]. Instead of relying on data-driven priors, we employ a physically grounded inverse rendering strategy [ZSH*22], which better supports the goal of consistent albedo estimation together with global and local illumination modeling for more accurate surface color projection in CH scenarios.

## 3. Method

Our method takes as input a combination of geometric, photographic, and calibration data. Specifically, we use a 3D digital model of the object to provide its geometric structure, along with a single flash photograph capturing its appearance. Additionally, we leverage comprehensive metadata related to the camera setup, including intrinsic parameters, the relative position of the flash light relative to the camera body, and the rigid-body transformation that aligns the camera with the 3D model. By integrating these elements, our approach aims to better estimate the body color signal and achieve more accurate color projection.

The image formation model employed in our approach takes into account multiple factors that contribute to the observed pixel color. Specifically, we consider a diffuse Lambertian surface reflectance, which is determined by the albedo $\boldsymbol{\rho}$, the surface normal $\mathbf{n}$, and the light direction $\mathbf{l}$. The illumination consists of two primary components: a direct point light source with intensity $\mathbf{L}_p$, representing the flash, and an ambient illumination component $\mathbf{L}_a$, which originates from a distant environment and is integrated over the hemisphere $\Omega$ of incoming directions $\omega$. This ambient contribution can be independent of the flash or result from multiple bounces of the flash light within the surrounding geometry. Additionally, we account for local interreflections $\mathbf{I}_r$, capturing light bouncing from nearby surfaces. Finally, a residual term $\mathbf{I}_\epsilon$ models any remaining signal not explicitly described by the previous elements. Based on this image formation model, the observed radiance $\mathbf{I}$ at a given surface point $\mathbf{x}$ is expressed as:

$$\mathbf{I}(\mathbf{x}) = \boldsymbol{\rho}(\mathbf{x}) \left[ \mathbf{L}_p(\mathbf{n}\cdot\mathbf{l}) + \int_\Omega \mathbf{L}_a(\omega)(\mathbf{n}\cdot\omega) d\omega + \mathbf{I}_r(\mathbf{x}) \right] + \mathbf{I}_\epsilon(\mathbf{x}) \quad (1)$$

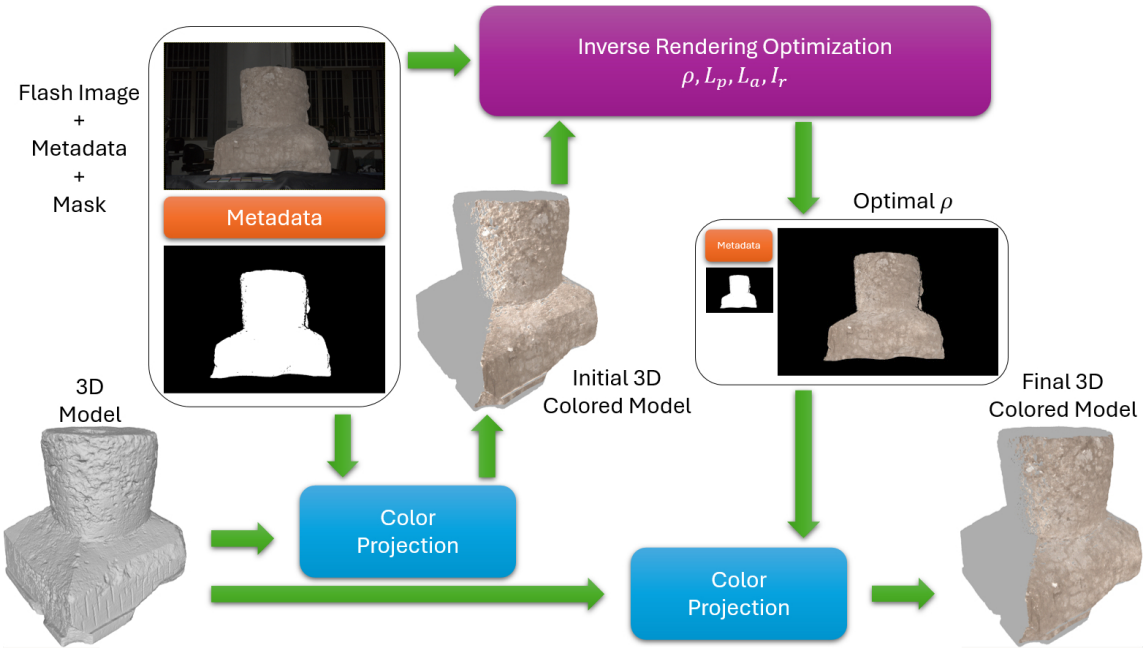


**Figure 1:** *Overview of the proposed pipeline. The process starts with a 3D model and a flash photograph with its metadata and a precomputed mask. The input image is initially projected onto the geometry to provide an initial color estimate. This serves as input for the optimization process, which refines the albedo $\rho$. The final per-vertex color is then updated through an additional photo projection and blending step.*

where the expression $max(0, n \cdot ...)$ is implicitly assumed in the normal dot products and omitted to enhance the readability and clarity of the equation.

The unknown values we want to recover from a single image using the proposed inverse rendering approach include the intensity of the point light $\mathbf{L}_p$, the ambient illumination intensity $\mathbf{L}_a$, the interreflection $\mathbf{I}_r$ and residual $\mathbf{I}_\epsilon$ signals, as well as the albedo $\boldsymbol{\rho}$ at all object points. We derive these signals by leveraging only the known input geometry and flash photograph, and the goal is the optimization of a cost function which is based on the mean squared error (MSE) between the observed flash photograph $\tilde{\mathbf{I}}$ and the modeled radiance $\mathbf{I}$. Since the inverse rendering procedure operates on pixel-based images, the modeled radiance $\mathbf{I}$ is computed per pixel. According to Equation (1), this implies that $\mathbf{I}(p)$ corresponds to the integral of all radiance values $\mathbf{I}(x)$ across the surface patch projected onto pixel $p$ through the camera's projective geometry. Consequently, the minimization problem is defined on a per-pixel basis as follows:

$$\mathbf{I}^* = \underset{\rho, \mathbf{L}_p, \mathbf{L}_a, \mathbf{I}_r}{argmin} \frac{1}{N} \sum_{p \in P} \left\| \mathbf{I}(p) - \tilde{\mathbf{I}}(p) \right\|^2 \qquad (2)$$

where $N$ is the number of pixels $p$ belonging to the set $P$ of image pixels defined by the precomputed mask. The residual $\mathbf{I}_\epsilon$ is equal to $\left(\mathbf{I}^* - \tilde{\mathbf{I}}\right)$, where $\mathbf{I}^*$ is the optimized computed radiance.

Figure 1 illustrates the complete workflow of our proposed approach. The process begins with a 3D model, a flash photograph along with its metadata, and a precomputed mask as inputs. To initialize the color signal, we project the input image directly onto the 3D geometry (see Sec. 3.1 for further initialization details). This initial colored model, along with the reference photograph, serves as the foundation for the optimization process described in Equation 2. Once the optimization is complete, we extract the estimated albedo $\rho$ and use it to update the per-vertex color through a second color projection step.

### 3.1. Implementation

The input flash photography data is captured as a raw linear signal, which has been previously calibrated using a color checker to ensure that the flash chromaticity appears as a perfect gray, allowing us to focus solely on determining the intensity of the light. The linear nature of the data is crucial for the computational process, as it ensures more precise modeling. Additionally, we store the data in a high dynamic range (HDR) image format, which helps mitigate numerical issues arising from quantization and ensures more precise calculations.

For the albedo signal, we treat it as a multi-spectral entity; in our case, we use the classic RGB color spectra. During our experiments, we determined that modeling the ambient environmental light as a constant sphere is sufficient for the flash photography setup. This approach yields good minimization results. Given the inherent ambiguity in distinguishing between completely unconstrained ambient light and per-vertex chromaticity, we also constrain the chromaticity of the ambient light to be gray, reducing ill-posed issues in the optimization process.

The initial guess for the color of the surface depends on the scenario. If the image is the first color signal to be projected onto the

3D geometry, and the geometry lacks any color information, we initialize the surface color directly from the original flash photograph, without any light compensation. However, if we are accumulating additional images, the initial guess is based on the already existing color signal on the 3D model, ensuring consistency in the color representation. For the ambient light, we initially set its value to zero, assuming no ambient light at the start. To improve convergence and mitigate the well-known scale ambiguity between the global brightness of the albedo and the intensity of the illuminators, our differentiable rendering-based optimization follows a two-step approach. We begin by performing a preliminary one-dimensional search to estimate an initial value for the flash light intensity. This step leverages prior knowledge of the average distance between the flash and the object, as well as radiometric calibration data obtained via the color checker, as described earlier. These constraints define a plausible absolute radiometric range for the flash intensity. Within this range, we execute a series of lightweight optimization steps to identify a suitable initialization value that best aligns the rendered appearance with the observed image. Once this initialization is established, we proceed with the full local optimization of all unknown parameters, including the flash and ambient illumination components as well as the per-vertex albedo values. This two-stage process ensures a stable and physically meaningful convergence of the inverse rendering pipeline. Our experiments show that this procedure achieves comparable or better results than directly optimizing all variables at once while significantly reducing the number of optimization iterations, leading to faster convergence.

The inverse rendering framework is built around a basic path tracer with six light bounces. This number of bounces strikes a balance between ensuring high-quality minimization and controlling the computational complexity, as increasing the number of bounces indefinitely would significantly increase the computational load. The framework employs the Adam optimizer with a learning rate of 0.05. To improve the stability of the optimization process and mitigate issues related to numerical divergence, ill-posedness, and inherent ambiguities, we enforce specific constraints on the optimized parameters throughout the iterative procedure. Since the flash illumination is pre-processed through white balancing, its chromaticity can be assumed to be neutral. This greatly simplifies the inverse rendering process by eliminating a key ambiguity between the color of the direct illumination and the rest of radiometric components. Additionally, constraining the ambient component to remain achromatic further mitigates potential sources of ambiguity and enhances the stability of the estimation process. Since the optimization operates directly on RGB spectral values, we then regularize both flash and ambient illumination by enforcing achromaticity at each iteration; their RGB components are averaged to maintain a consistent gray chromaticity. Furthermore, we apply clamping to bound the range of the optimized variables: RGB albedo values and ambient illumination components are restricted to the normalized [0, 1] range, ensuring physically plausible reflectance values; the flash intensity is clamped within an absolute range derived from the radiometric calibration, as previously described. These regularization strategies help guide the optimization toward physically meaningful solutions and prevent degenerate behaviors.

## 4. Results

We present an evaluation of our proposed solution through both controlled synthetic experiments and real-world CH assets acquired in a museum setting. The evaluation is conducted on a high-performance computing platform equipped with an Intel® Core™ i9-14900KF CPU with 32 cores, an NVIDIA GeForce RTX 4090 GPU, and 188GB of RAM. The inverse rendering framework is implemented in Python using the Mitsuba library [JSR*22], ensuring physically-based simulation of light transport. Additionally, for the color projection step, we employ an efficient and scalable streaming technique designed to map high-resolution color information onto extremely dense point clouds [PGC11]. RAW color images from the camera are handled using the *dcraw 9.28* library. With the current hardware and software configuration, the initial estimation of the flash intensity—performed via a brute-force one-dimensional search—takes less than one hour, while the subsequent full local optimization completes in about five minutes for each photograph. It is important to note that the flash intensity estimation is required only once per acquisition session, regardless of the number of images processed. As a result, its computational cost becomes progressively amortized as more images are incorporated. While the initial estimation could be significantly accelerated using established techniques commonly employed in production environments, such engineering optimizations lie beyond the research objectives of this work.

The results are structured as follows: first, we assess the method's performance in a controlled synthetic environment, allowing for a precise analysis of its ability to model and compensate for complex illumination effects (see Sec. 4.1). Then, we demonstrate the applicability of our approach on real CH artifacts (see Sec. 4.2), highlighting its effectiveness in practical scenarios where uncontrolled lighting conditions and intricate surface details pose additional challenges.

### 4.1. Synthetic Datasets

To assess the effectiveness of our approach in a controlled environment, we conducted evaluations using two simple synthetic datasets: a sphere and a V-shaped pair of planes. Both models were assigned the same uniform color and we study our approach in a controlled setting by isolating various effects of illumination variations. The sphere was chosen due to its convex topology to analyze mostly how ambient global light influences the color projection process and to evaluate the ability of our method to correctly estimate and compensate for this effect. On the other hand, the V-shaped planes were selected to investigate the impact of interreflections and assess how well our approach can account for these local and indirect lighting effects. In fact, the sharp intersection between the two planes creates a challenging case where light bounces between surfaces, introducing color bleeding that must be modeled and corrected. This setup provides a well-defined scenario where indirect illumination plays a significant role in the observed appearance.

Figure 2 illustrates the ground truth models used in these experiments, and the performance evaluation results. For the *Sphere* dataset, we analyze three different conditions. In the first scenario,

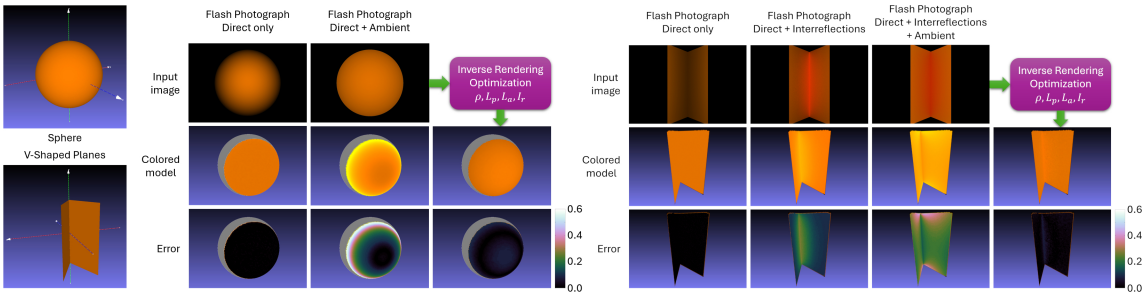


**Figure 2:** *Evaluation on synthetic datasets. The left images show the two synthetic datasets. A* Sphere *is used for ambient light analysis and* V-Shaped Planes *for interreflection correction. The group of images in the middle shows evaluation results for the* Sphere *dataset, while the rightmost ones correspond to the* V-Shaped Planes. *Columns represent direct illumination (leftmost), indirect global or local effects (middle columns), and our optimized approach (rightmost). Rows display input images, colored 3D models, and color projection errors. Our method compensates for local and global illumination effects, ensuring more accurate albedo projection with minimal error, even in challenging scenarios.*

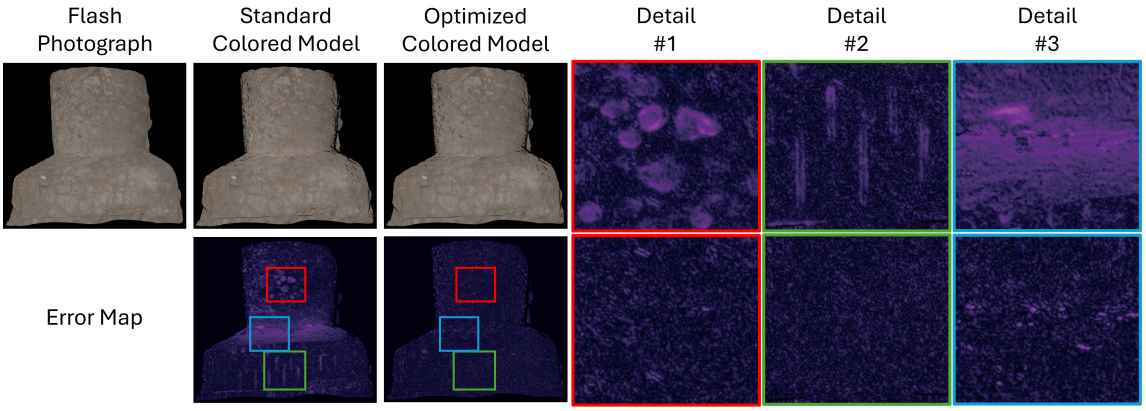


**Figure 3:** *We compare the original flash photograph of the* Torre Quadrata *with two virtual renderings from the same viewpoint. The second column shows a rendering using the standard albedo estimate with only direct flash illumination, while the third column presents the result using the optimized albedo and illumination model, accounting for both direct flash, and indirect global ambient light and local interreflections. ΔLIP error maps highlight discrepancies due to indirect lighting effects, which are effectively compensated by our method, resulting in a more accurate match to the original photograph.*

shown in the left column, we simulate an ideal case where the flash photograph contains only direct illumination. In the second case, presented in the middle column, we introduce indirect lighting effects, resulting mostly from global illumination from the surrounding environment. Finally, in the right column, we apply our proposed inverse rendering optimization to compensate for direct and indirect illumination before projecting the color onto the 3D model. The rows in the figure respectively show the input images used for color projection, the resulting colored models, and the error maps that highlight the difference between the projected colors and the ground truth. For the error maps we used the *CubeHelix* colormap, which is a perceptually uniform colormap designed for information visualization that maintains a monotonic luminance increase and avoids abrupt jumps in brightness that distort perception [Gre11]. In the case of the *Sphere*, the convex nature of the surface makes the contribution of environmental lighting particularly evident, especially in the outer ring regions, where the surface is foreshortened relative to the viewing direction. This effect causes a noticeable darkening in the absence of indirect illumination, as less light is scattered back toward the viewer. In the second scenario, global illumination counterbalances this effect, brightening those regions. Under purely direct illumination, the image formation process can be easily modeled, allowing for an accurate color projection that properly compensates for the interaction between light and the surface. As a result, the difference between the projected color and the original ground truth remains minimal. However, when unknown global indirect illumination effects are present, correcting only the flash contribution leads to increasing errors, particularly in areas with strong foreshortening. In these regions, the expected light intensity is lower, and attempting to counterbalance it without accounting for indirect effects results in an overcompensation, producing significant color projection errors. By incorporating both direct and indirect illumination into the optimization process, our approach effectively compensates for these effects. The adjusted colors are then projected onto the geometry, leading to a final colored model with significantly reduced error and improved accuracy.

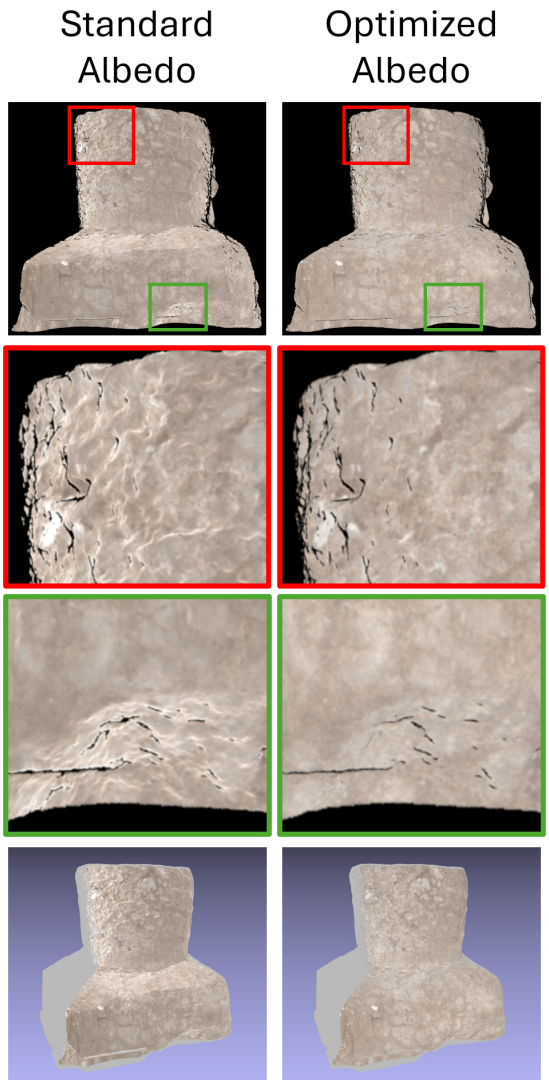


**Figure 4:** *Albedo comparison for* Torre Quadrata*. The optimized albedo (right) is more uniform and less influenced by geometric variations than the direct flash-based estimate (left), reducing baked-in shading effects and albedo overcompensation.*

In the *V-Shaped Planes* dataset, the first scenario remains unchanged, consisting solely of direct illumination from the flash. However, unlike the *Sphere* dataset, here we analyze the effects of indirect lighting in two separate stages: first, we introduce only local interreflections (second column), and then we incorporate global environmental illumination as well (third column). As in the case of the *Sphere*, the final column presents the results obtained with our proposed method. In this setup, interreflections become particularly evident near the junction where the two planes meet, leading to a noticeable increase in brightness due to multiple light bounces between the surfaces. This phenomenon is a well-documented effect of local indirect illumination. Similar to the foreshortened regions observed in the sphere dataset, the naive compensation of illumination in the presence of interreflections results in an overestimation of color brightness, causing an excessive brightening in these areas and an increasing projection error. The introduction of environmental illumination further amplifies this issue, as seen in the third column, where both global and local indirect lighting contribute to additional distortions in the projected color. Despite these challenges, our method effectively models and compensates for both local and global illumination effects before projecting the estimated albedo onto the 3D model. As a result, even in the complex case of the *V-Shaped Planes*, the final projected colors exhibit a significantly reduced error.

These two synthetic datasets have been carefully designed to evaluate our approach under two geometrically challenging scenarios: one featuring highly foreshortened surface regions, and the other characterized by a deep concave structure prone to multiple light bounces. To isolate and clearly observe the effects of lighting and the impact of our correction strategy, the datasets are deliberately kept simple, most notably by employing uniform surface colors. This design choice allows for a more transparent assessment of our method's effectiveness in managing complex illumination phenomena and correcting related artifacts.

### 4.2. Cultural Heritage Datasets

In this section, we evaluate our method on a real-world CH dataset featuring two remarkable statues from the Mont'e Prama complex, a collection of Neolithic stone sculptures discovered in Western Sardinia. Created by the Nuragic civilization between the 10th and 7th centuries BC—an exact timeframe still debated—these sculptures rank among the most significant archaeological finds in the Mediterranean. Carved from local limestone, they reflect the unique artistic and architectural heritage of this ancient culture. The selected artifacts represent stylized models of Nuragic towers, the imposing megalithic structures that once shaped Sardinia's landscape. Referred to here as *Torre Quadrata* and *Nuraghe*, these statues, with their intricate carvings and weathered surfaces, offer an ideal testbed for assessing our method under real-world conditions.

Fig. 3 showcases a comparative evaluation of our method applied to the *Torre Quadrata* object. We juxtapose the original flash photograph with two virtual renderings generated from the same viewpoint, using identical intrinsic and extrinsic camera parameters. The first rendering (second column) is obtained using the albedo estimated by a standard pipeline that assumes only direct flash illumination, disregarding indirect lighting effects. This results in an image formed by combining the estimated albedo with the simulated flash lighting. The second rendering (third column) is generated using our optimized albedo, obtained through differentiable rendering, along with an illumination model that accounts for both the direct flash light and the optimized indirect ambient illumination and interreflections. To quantitatively assess the accuracy of our approach, we compare the original photograph and the two renderings using three image quality metrics: PSNR, SSIM, and ꟻLIP. PSNR (Peak Signal-to-Noise Ratio) [HTG08] provides a pixel-wise measure of similarity, with higher values indicating greater fidelity. SSIM (Structural Similarity Index) [WBSS04] and ꟻLIP [ANAM*20] are perceptual metrics that evaluate image differences based on human visual perception, capturing structural and contrast-based variations more effectively than PSNR. The

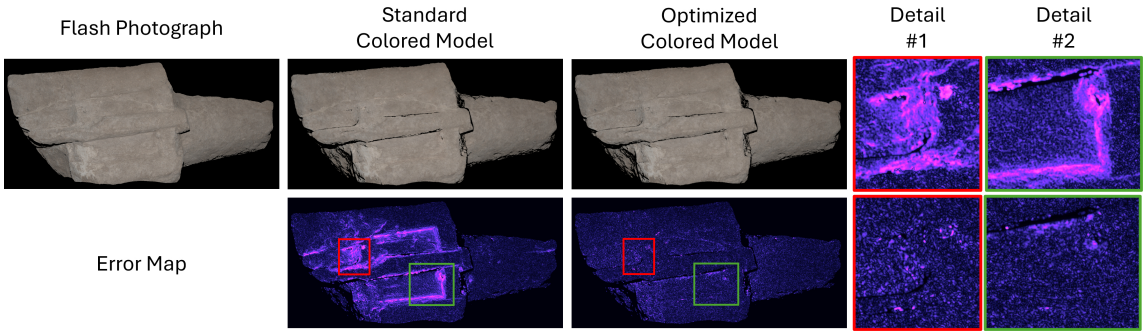


**Figure 5:** *We compare the original flash photograph of the* Nuraghe *with two virtual renderings from the same viewpoint. The second column shows a rendering using the standard albedo estimate with only direct flash illumination, while the third column presents the result using the optimized albedo and illumination model, accounting for both direct flash, and indirect global ambient light and local interreflections. ꟻLIP error maps highlight discrepancies due to indirect lighting effects, which are effectively compensated by our method, resulting in a more accurate match to the original photograph.*

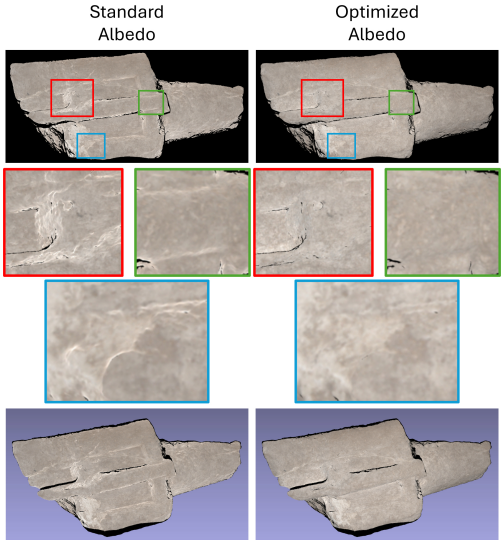


**Figure 6:** *Albedo comparison for* Nuraghe. *The optimized albedo (right) is more uniform and less influenced by geometric variations than the direct flash-based estimate (left), reducing baked-in shading effects and albedo overcompensation.*

ꟻLIP error maps, shown in both full images and detailed insets, highlight the regions where significant discrepancies arise. The ꟻLIP error maps were computed by masking out dark pixels that were not colored by the color projection routine, ensuring that only the relevant regions of the image were considered in the error evaluation. Most of the differences stem from local indirect lighting effects, particularly interreflections within small surface cavities (Details #1 and #2) and areas illuminated by large neighboring surface regions (Detail #3). Our method, as illustrated in the error maps of the second row and last three columns, successfully compensates for these indirect lighting effects, leading to a more accurate albedo estimation and a virtual rendering that aligns more closely with the original flash photograph.

Another way to evaluate the improvement in color reconstruction quality is by analyzing the pure albedo signal, rendered without any illumination, as shown in Fig. 4. On the left, we display the albedo computed using only direct flash illumination. In this case, brightness is significantly overcompensated in depth edges, where the surface normal deviates from a front-facing orientation relative to the camera. This leads to residual shading artifacts embedded in the albedo. In contrast, the optimized albedo, shown in the right column, displays a more uniform and consistent appearance that is less influenced by geometric variations, highlighting the effectiveness of our approach in mitigating these artifacts. The last row of Fig. 4 offers an alternative viewpoint of the pure albedo, further illustrating how the optimized version minimizes baked-in shading effects, providing a more accurate and faithful color representation of the surface.

Similar results are observed for the *Nuraghe* dataset, as shown in Fig. 5 and Fig. 6. The comparative evaluation of virtual renderings and the albedo analysis follow the same reasoning as for the *Torre Quadrata* object, with our method effectively compensating for indirect lighting effects and producing a more accurate color reconstruction, as reflected in the error maps and optimized albedo.

The quantitative performance of our method is further demonstrated in Tab. 1, which presents the *PSNR* and *SSIM* values for both the *Torre Quadrata* and *Nuraghe* examples. As shown in the table, our approach achieves superior results in both metrics, outperforming the standard pipeline in terms of both pixel-wise similarity (PSNR) and perceptual quality (SSIM).

| Method | *Torre Quadrata* | | *Nuraghe* | |
|---|---|---|---|---|
| | PSNR | SSIM | PSNR | SSIM |
| Standard | 44.07 | 0.98 | 39.26 | 0.95 |
| Optimized | 46.73 | 0.99 | 41.74 | 0.96 |

**Table 1:** PSNR *and* SSIM *values for the* Torre Quadrata *and* Nuraghe *datasets, showing improved performance with our method compared to the standard pipeline.*

In another experiment, we also assess the capability of our

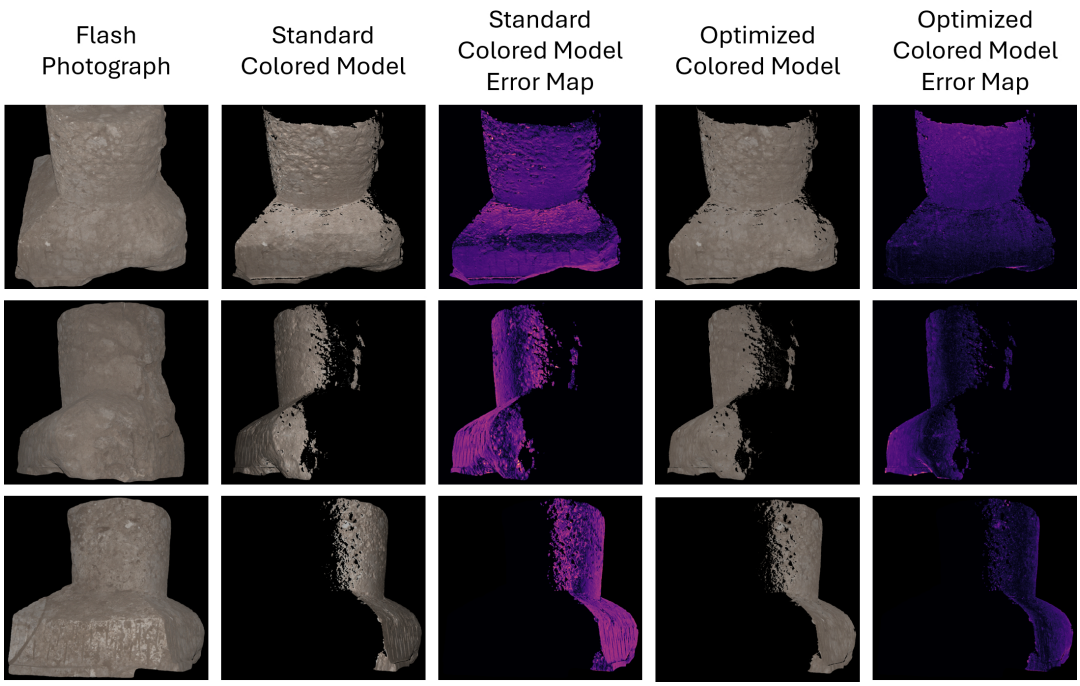


**Figure 7:** *Appearance prediction across novel viewpoints and light conditions. The first column shows original flash photographs, while the second and fourth columns present virtual renderings using the standard and optimized methods. ∄LIP error maps (third and fifth columns) highlight the improved consistency and accuracy of our approach.*

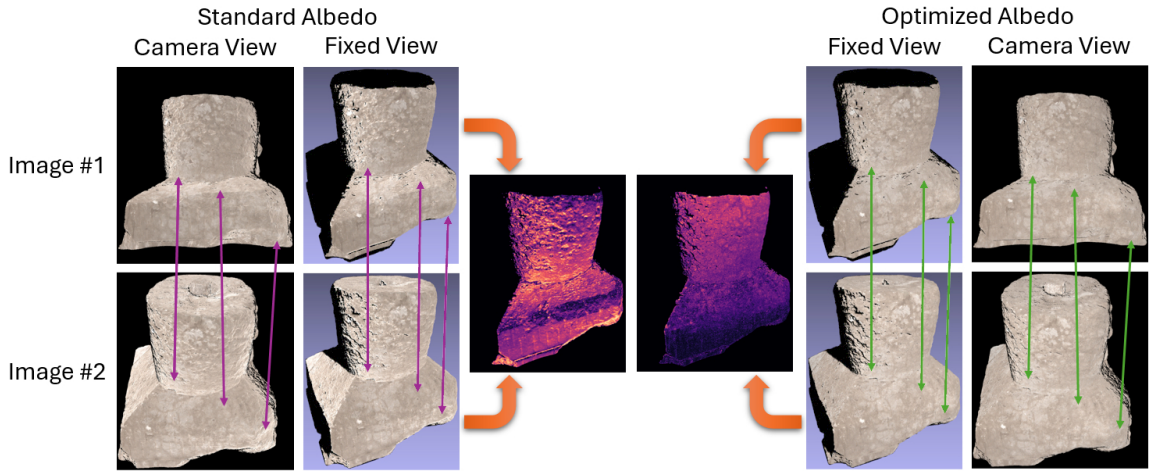


**Figure 8:** *Evaluation of repeatability across viewpoints. The outermost columns show the estimated pure albedo from two flash photographs captured from corresponding camera perspectives, using both the standard and our optimized method. The standard approach reveals inconsistencies in areas that should exhibit the same albedo (indicated by purple arrows), whereas our method achieves significantly better consistency across views (green arrows). The remaining columns illustrate the color projections onto the 3D model from a fixed viewpoint, along with the corresponding ∄LIP error maps, underscoring the improved stability and repeatability of our approach.*

method to predict the appearance of the colored object under novel viewpoints and illumination conditions. Specifically, we evaluate whether the colored model, obtained by projecting the optimized albedo from a specific single flash photograph (i.e., the one presented in Fig. 3 and Fig. 4), can better reproduce the actual colors captured from different viewpoints in other acquired flash photographs. Naturally, we conduct this experiment using that single projected image to prevent error mitigation effects that could result from blending multiple images onto the same surface region. So, in Fig. 7 we present three new flash photographs acquired from view-

points not used during the optimization process, ensuring that these images are entirely unseen by the optimization procedure. For each viewpoint, we first show the original flash photograph, followed by two virtual renderings of the model, i.e., one generated using the standard approach, where the albedo is estimated under the assumption of direct flash illumination only, and another obtained using our optimized solution, which incorporates both the direct flash light and the modeled indirect contributions from global and local illumination. By examining the virtual renderings and corresponding ꟻLIP error maps, we observe that our approach consistently improves the prediction of object appearance across multiple novel viewpoints and illumination settings. Unlike the standard method, which struggles with indirect lighting effects, our optimized model provides a more stable and accurate reconstruction, better aligning with the actual captured images. These results highlight the increased robustness of our approach in handling varying illumination conditions and demonstrate its effectiveness in enhancing color fidelity under different viewpoints.

The final experiment evaluates the improvement in repeatability achieved by our method compared to the standard approach. We acquire two flash photographs from different viewpoints, capturing approximately the same region of the object, and process them using both pipelines. Fig. 8 presents a comparative analysis of the resulting pure albedo maps. The outermost columns show the pure albedo estimated from each corresponding viewpoint using the standard and optimized methods. In the standard pipeline, significant discrepancies appear in regions that should exhibit a consistent albedo (highlighted by purple arrows), whereas our optimized approach ensures greater consistency across views, as indicated by the matching regions marked with green arrows. Beyond the per-view albedo comparisons, the remaining part of Fig. 8 visualizes the assigned colors on the 3D model rendered by a fixed viewpoint. To further assess consistency, we compute ꟻLIP error maps between the overlapping image regions, masking non-overlapping areas. The results demonstrate that our method enhances the repeatability of color estimation across different viewpoints. While the final model will be refined through multiple overlapping photo blending, providing a more consistent and accurate albedo estimation at this stage establishes a stronger foundation for blending, ultimately yielding a higher-quality colored model.

## 5. Conclusions

In this work, we presented an improved inverse rendering pipeline for flash photography-based albedo estimation, addressing the challenges posed by indirect illumination effects. Our method refines albedo estimation by accounting for both global and local light interactions, leading to more accurate color reconstruction and improved consistency across multiple viewpoints. Through controlled synthetic experiments, we demonstrated how our approach effectively compensates for interreflections and shading artifacts, significantly reducing color distortions that arise in standard pipelines. We further validated our method on real-world CH assets, showing its robustness in complex, uncontrolled environments. Quantitative evaluations using PSNR, SSIM, and ꟻLIP metrics confirmed that our optimized albedo estimation produces more faithful representations compared to the standard approach. Additionally, we assessed the stability of our method across different viewpoints and illumination conditions. Results showed that our approach consistently improves repeatability, ensuring that the same surface regions exhibit minimal color variations across different acquisitions. This property is particularly beneficial for photometric blending techniques, as it provides a more reliable input for generating high-quality, fully textured 3D models. The proposed solution contributes to a more reliable and detailed reconstruction of heritage artifacts, advancing the broader field of digital preservation and 3D modeling.

While our method has been designed for single-image optimization, its extension to a multi-image framework presents a compelling avenue for future research. By incorporating multiple views, each contributing to the optimization of vertex colors under varying illumination conditions, we could enhance robustness and accuracy. However, this shift would also introduce computational challenges, necessitating the development of more efficient optimization strategies. Further improvements could also be achieved by refining the illumination modeling process. In particular, integrating advanced techniques to better account for specular reflections and material variations would enhance the fidelity of the reconstructed material. Finally, a more sophisticated treatment of environmental lighting remains an open challenge. While our current model assumes a uniform ambient contribution, exploring non-uniform illumination conditions and inferring geometric constraints of the surrounding environment could strengthen the relationship between flash lighting and global illumination effects, further improving the accuracy of the relighting process.

## Acknowledgements

We thank the *Soprintendenza Archeologia, Belle Arti E Paesaggio Per La Città Metropolitana Di Cagliari E Le Province Di Oristano E Sud Sardegna* for access to artworks for digitization and fruitful collaboration. This work received funding from Sardinian regional authorities under the XDATA project (art 9 L.R. 20/2015). The authors also acknowledge the contribution of the Italian National Research Center in High-Performance Computing, Big Data, and Quantum Computing (Next Generation EU PNRR M4C2 Inv1.4).

## References

[ANAM*20] Andersson P., Nilsson J., Akenine-Möller T., Oskarsson M., Åström K., Fairchild M. D.: FLIP: A difference evaluator for alternating images. *Proc. ACM Comput. Graph. Interact. Tech. 3*, 2 (Aug. 2020). doi:10.1145/3406183. 6

[ASOS13] Ali A., Sato I., Okabe T., Sato Y.: Efficient modeling of objects BRDF with planned sampling. *IPSJ Transactions on Computer Vision and Applications 5* (2013), 114–118. doi:10.2197/ipsjtcva.5.114. 2

[AWL13] Aittala M., Weyrich T., Lehtinen J.: Practical SVBRDF capture in the frequency domain. *ACM Trans. Graph. 32*, 4 (July 2013). doi:10.1145/2461912.2461978. 2

[BPV*15] Bettio F., Pintus R., Villanueva A. J., Merella E., Marton F., Gobbetti E.: Mont'e Scan: Effective shape and color digitization of cluttered 3D artworks. *J. Comput. Cult. Herit. 8*, 1 (Feb. 2015). doi:10.1145/2644823. 1, 2

[C*21] Cave, et al.: OpenMVS: A multi-view stereo library, 2021. Accessed: 2025-05-28. URL: https://github.com/cdcseacave/openMVS. 2

[DAD*18] DESCHAINTRE V., AITTALA M., DURAND F., DRETTAKIS G., BOUSSEAU A.: Single-image SVBRDF capture with a rendering-aware deep network. *ACM Trans. Graph. 37*, 4 (July 2018). doi:10.1145/3197517.3201378. 2

[DCC*09] DELLEPIANE M., CALLIERI M., CORSINI M., CIGNONI P., SCOPIGNO R.: Flash lighting space sampling. In *Computer Vision/Computer Graphics Collaboration Techniques* (Berlin, Heidelberg, 2009), Springer-Verlag, pp. 217–229. doi:10.1007/978-3-642-01811-4_20. 2

[FAB*24] FURFERI R., ANGELO L. D., BERTINI M., MAZZANTI P., VECCHIS K. D., BIFFI M.: Enhancing traditional museum fruition: current state and emerging tendencies. *Heritage Science 12*, 1 (2024), 20. doi:10.1186/s40494-024-01139-y. 1

[FC17] FU H., CHUANG Y.-Y.: VisualSFM: A visual structure-from-motion system, 2017. Accessed: 2025-05-28. URL: http://ccwu.me/vsfm/. 2

[FH15] FURUKAWA Y., HERNÁNDEZ C.: Multi-view stereo: A tutorial. *Found. Trends. Comput. Graph. Vis. 9*, 1–2 (June 2015), 1–148. doi:10.1561/0600000052. 2

[GLT*21] GUO J., LAI S., TAO C., CAI Y., WANG L., GUO Y., YAN L.-Q.: Highlight-aware two-stream network for single-image SVBRDF acquisition. *ACM Trans. Graph. 40*, 4 (July 2021). doi:10.1145/3450626.3459854. 2

[Gre11] GREEN D. A.: A colour scheme for the display of astronomical intensity images. *Bulletin of the Astronomical Society of India 39* (2011), 289–295. URL: https://arxiv.org/abs/1108.5083, arXiv:1108.5083. 5

[HS17] HUI Z., SANKARANARAYANAN A. C.: Shape and spatially-varying reflectance estimation from virtual exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence 39*, 10 (2017), 2060–2073. doi:10.1109/TPAMI.2016.2623613. 2

[HTG08] HUYNH-THU Q., GHANBARI M.: Scope of validity of PSNR in image/video quality assessment. *Electronics Letters 44* (2008), 800–801. doi:10.1049/el:20080522. 6

[JLX*24] JU Y., LAM K.-M., XIE W., ZHOU H., DONG J., SHI B.: Deep learning methods for calibrated photometric stereo and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence 46*, 11 (Nov. 2024), 7154–7172. doi:10.1109/TPAMI.2024.3388150. 2

[JSR*22] JAKOB W., SPEIERER S., ROUSSEL N., NIMIER-DAVID M., VICINI D., ZELTNER T., NICOLET B., CRESPO M., LEROY V., ZHANG Z.: Mitsuba 3 renderer (v3.01), 2022. Accessed: 2025-05-28. URL: https://mitsuba-renderer.org. 4

[KHM*24] KAVOOSIGHAFI B., HAJISHARIF S., MIANDJI E., BARAVDISH G., CAO W., UNGER J.: Deep SVBRDF acquisition and modelling: A survey. *Computer Graphics Forum* (2024). doi:10.1111/cgf.15199. 2

[LLC23] LLC A.: Metashape, 2023. Accessed: 2025-05-28. URL: https://www.agisoft.com/. 2

[LPC*00] LEVOY M., PULLI K., CURLESS B., RUSINKIEWICZ S., KOLLER D., PEREIRA L., GINZTON M., ANDERSON S., DAVIS J., GINSBERG J., SHADE J., FULK D.: The digital michelangelo project: 3d scanning of large statues. *Proc. ACM Comput. Graph. Interact. Tech.* (2000), 131–144. doi:10.1145/344779.344849. 1, 2

[LSBE24] LUO X., SCANDOLO L., BOUSSEAU A., EISEMANN E.: Single-image SVBRDF estimation with learned gradient descent. *Computer Graphics Forum* (2024). doi:10.1111/cgf.15018. 2

[LSC18] LI Z., SUNKAVALLI K., CHANDRAKER M.: Materials for masses: SVBRDF acquisition with a single mobile phone image. In *ECCV* (Berlin, Heidelberg, 2018), Springer-Verlag, p. 74–90. doi:10.1007/978-3-030-01219-9_5. 2

[PDC*19] PINTUS R., DULECHA T. G., CIORTAN I. M., GOBBETTI E., GIACHETTI A.: State-of-the-art in Multi-Light Image Collections for Surface Visualization and Analysis. *Computer Graphics Forum* (2019). doi:10.1111/cgf.13732. 1, 2

[PG15] PINTUS R., GOBBETTI E.: A fast and robust framework for semiautomatic and automatic registration of photographs to 3D geometry. *J. Comput. Cult. Herit. 7*, 4 (Feb. 2015). doi:10.1145/2629514. 2

[PGC11] PINTUS R., GOBBETTI E., CALLIERI M.: Fast low-memory seamless photo blending on massive point clouds using a streaming framework. *J. Comput. Cult. Herit. 4*, 2 (Nov. 2011). doi:10.1145/2037820.2037823. 2, 4

[PGCD17] PINTUS R., GOBBETTI E., CALLIERI M., DELLEPIANE M.: Techniques for seamless color registration and mapping on dense 3D models. In *Sensing the Past: From artifact to historical site*, Masini N., Soldovieri F., (Eds.). Springer-Verlag, 2017, pp. 355–376. isbn: 978-3-319-50518-3. URL: https://publications.crs4.it/pubdocs/2017/PGCD17. 1, 2

[Rem11] REMONDINO F.: Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sensing 3*, 6 (2011), 1104–1138. doi:10.3390/rs3061104. 2

[SBG11] STANCO F., BATTIATO S., GALLO G. (Eds.): *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*, 1st ed. CRC Press, Boca Raton, FL, 2011. doi:10.1201/b11049. 2

[Sch21] SCHOENBERGER J. L.: COLMAP: A general structure-from-motion and multi-view stereo pipeline, 2021. Accessed: 2025-05-28. URL: https://github.com/colmap/colmap. 2

[Sco21] SCOPIGNO R.: Mixing visual media for cultural heritage. In *Emerging Technologies and the Digital Transformation of Museums and Heritage Sites* (Cham, 2021), Shehade M., Stylianou-Lambert T., (Eds.), Springer-Verlag, pp. 297–315. doi:10.1007/978-3-030-83647-4_20. 1

[SLS23] SHI Z., LIN X., SONG Y.: An attention-embedded GAN for SVBRDF recovery from a single image. *Computational Visual Media 9*, 3 (2023), 551–561. doi:10.1007/s41095-022-0289-1. 2

[SP23] SARTOR S., PEERS P.: Matfusion: A generative diffusion model for SVBRDF capture. In *SIGGRAPH Asia 2023 Conference Papers* (New York, NY, USA, 2023), Association for Computing Machinery. doi:10.1145/3610548.3618194. 2

[TGVG12] TINGDAHL D., GODAU C., VAN GOOL L.: Base materials for photometric stereo. In *ECCV* (Berlin, Heidelberg, 2012), Springer-Verlag, p. 350–359. doi:10.1007/978-3-642-33868-7_35. 2

[VPS21] VECCHIO G., PALAZZO S., SPAMPINATO C.: Surfacenet: Adversarial SVBRDF estimation from a single image. In *Proceedings of the International Conference on Computer Vision* (2021), pp. 12820–12828. doi:10.1109/ICCV48922.2021.01260. 2

[WBSS04] WANG Z., BOVIK A., SHEIKH H., SIMONCELLI E.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (2004), 600–612. doi:10.1109/TIP.2003.819861. 6

[WWL*21] WANG X., WANG C., LIU B., ZHOU X., ZHANG L., ZHENG J., BAI X.: Multi-view stereo in the deep learning era: A comprehensive review. *Displays 70* (2021). doi:10.1016/j.displa.2021.102102. 2

[ZGW*23] ZHANG L., GAO F., WANG L., YU M., CHENG J., ZHANG J.: Deep SVBRDF estimation from single image under learned planar lighting. In *Proc. ACM SIGGRAPH* (New York, NY, USA, 2023), Association for Computing Machinery. doi:10.1145/3588432.3591559. 2

[ZK21] ZHOU X., KALANTARI N. K.: Adversarial Single-Image SVBRDF Estimation with Hybrid Training. *Computer Graphics Forum* (2021). doi:10.1111/cgf.142635. 2

[ZSH*22] ZHANG Y., SUN J., HE X., FU H., JIA R., ZHOU X.: Modeling Indirect Illumination for Inverse Rendering . In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Los Alamitos, CA, USA, June 2022), IEEE Computer Society, pp. 18622–18631. doi:10.1109/CVPR52688.2022.01809. 2