Spatially-Varying Diffuse Reflectance Capture Using Irradiance Map Rendering for Image-Based Modeling Applications

Ladefoged, Kasper Skou; Madsen, Claus Brøndgaard

Published in:
2019 IEEE International Symposium on Mixed and Augmented Reality

DOI (link to publication from Publisher):
10.1109/ISMAR.2019.00-27

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

Spatially-Varying Diffuse Reflectance Capture Using Irradiance Map Rendering for Image-Based Modeling Applications

Kasper S Ladefoged* Claus B. Madsen†

Department of Architecture, Design, and Media Technology - Aalborg University, Denmark

Figure 1: *The figure shows five objects reconstructed using SfM. Left object show reconstructions rendered using the appearance as texture. Right objects shows reconstructions rendered using the reflectance captured using the method proposed in this paper as texture.*

Abstract

Image-based 3D modelling using Structure-form-Motion (SfM) has matured significantly over the last decade. Standard SfM methods create the object's texture from the appearance of the physical object at the time of acquisition. We propose a method for acquiring the diffuse per-point reflectance of the modelled object, as part of the image acquisition work flow, only adding one extra captured image and an irradiance rendering step, making it easy for anyone to digitize physical objects to create 3D content for AR/VR using only consumer grade hardware. Current state of the art of spatially varying reflectance capture requires either large, expensive, and purpose built setups or are optimization based approaches, whereas the proposed approach is model based.

This paper proposes adding a render of irradiance with modelled camera and light source, using off the shelf hardware for image capture. The key element is taking two images at each imaging location: one with just the ambient illumination conditions, and one where the light from an on-camera flash is included. It is demonstrated how to get the ambient illumination to cancel out, and by assuming Lambertian materials, render the irradiance corresponding to the flash-only image, enabling computation of spatially varying diffuse reflectance rather than appearance. Qualitative results demonstrate the added realism of the modelled objects when used as assets in renders under varying illumination conditions, including limited outdoor scenarios. Quantitative tests demonstrate that the reflectance can be estimated correctly to within a few percent even in cases with severe un-even ambient illumination.

*e-mail: ksla@create.aau.dk

†e-mail: cbm@create.aau.dk

Index Terms: Computing methodologies—Computer graphics—Rendering—Reflectance modeling; Computing methodologies—Computer graphics—Image manipulation—Computational photography; Computing methodologies—Computer vision—Computer vision problems—Reconstruction

1 Introduction

Image-based 3D modelling using Structure-from-Motion (SfM) techniques has matured significantly over the last decade. For an overview of SfM see Özyeşil et al 2017 [30]. It is now a widely used technique for obtaining accurate 3D digitization of physical objects at any scale: entire cities, constructions sites, bridges, wind turbines, down to sub-millimeter precision for surface roughness quality control [26].

There is a host of solutions for SfM based 3D modelling, both commercial and free e.g., Bentley ContextCapture [6], Agisoft Metashape [1], Reality Capture [7] and VisualSFM [35] in conjunction with MeshLab [10]. They all basically work the same way: a set of uncalibrated images are acquired from different unknown viewpoints of an object, the images are then fed into the software as an un-ordered set, and the software performs the 3D reconstruction. The SfM solutions provide: 1) the geometry of the object in the form of a sparse 3D point cloud which is typically also triangulated in to a dense 3D point cloud, 2) calibrations of all images used in the reconstruction, i.e., the precise camera orientations and positions as well as calibrations of the internal parameters such as focal length, etc., and 3) a texture map associating each vertex in the point cloud with a uv-coordinate in a texture atlas.

Since this texture atlas is made from the images originally acquired to model the object, these textures represent *appearance*, i.e., they represent what the object looks like in the illumination environment of where it was captured. This might be fine for some applications, but if one wants to really exploit the digitized 3D object for visualization in another visual context, under other virtual

illumination conditions, it would in fact be much better if the texture represented *reflectance* information as opposed to appearance. An example of the difference in using appearance and reflectance resulting from the proposed approach can be seen in Figure 1.

The main contribution that this paper presents is an approach to capturing diffuse reflectance maps as part of the SfM workflow using consumer grade hardware, by rendering the irradiance, making it easy for anyone to digitize physical objects to create 3D content for AR/VR. The approach uses two images taken at each location, rather than just one: an image of the object under the given illumination conditions, plus an image of the object where an on-camera flash is added to the given illumination conditions. The approach does not require direct measurement of incident illumination, instead this is simulated using a approximated model of the on-camera flash.

The paper is organized as follows: Section 2 describes related work. An overview of the proposed approach is given in section Section 3, to provide an intuitive understanding. Section 4 describes assumptions of the proposed approach and contain the theoretical basis for the solution, including irradiance model and surface interaction. The experimental results are presented in Section 5, followed by the discussion in Section 6. Finally, concluding remarks are presented in Section 7.

2 Related Work

The field of reflectance capture can roughly be split into two categories: controlled lighting conditions and uncontrolled lighting conditions. We limit the related work to the methods with controlled lighting. This is due to the approach proposed in this paper being based on the addition of light from an on-camera flash and the need of a pair of images: one containing the ambient, unknown light, only, and one containing both the ambient light and the light from the on-camera flash.

Flash Photography Flash photography is not a new concept, and has been used for a multitude of different applications [13,15,17,31]. In our proposed method flash photography leverages the fact that light is additive. This makes it possible to isolate a known light source that can then be simulated to generate the irradiance on each point of the captured image. This has to the knowledge of the authors not been done before.

Nam et al. 2018 [24] proposed an optimization based method capturing both specular and diffuse properties while refining the geometry, showing very promising results. Our method proposes the use of already existing tools and simple to use methods in order to simplify and utilize the advances in these related areas, thereby creating a very simple and straight forward way to reconstruct a real-world object and calculate the corresponding diffuse albedo. Due to this, the steps needed to reproduce the proposed method is easy and straight forward, with a limited need for specialized knowledge.

Furthermore, due to the inclusion of the calibrated flash, and the resulting calibrated irradiance, the resulting diffuse albedo is correct in relation to the radiometric world, a step that are not discussed nor show in Nam et al. 2018. This enables the resulting diffuse albedo to be usable for augmentation using calibrated lighting, such as calibrated HDRI etc. omitting any user defined scaling necessary for visual consistency.

To further differentiate from Nam et al. 2018 the irradiance rendering step and using this as a fundamental parameter to calculate reflectance, has not been investigated or shown possible before.

In-Shoot Artifact Based Methods Lensch et al. 2001 [20] presented an approach of capturing spatially varying appearance by capturing multiple images of the object. The geometry of the object was captured using a laser scanner, and further refined by hand. The position and intensity of the light source was captured using a mirroring sphere and the intensity by a gray card, whilst having a known camera position. This method was extended by Lensch et al. 2003 [19] to include per-texel normal refinement and in Lensch et al. 2003 [21] describing a method for finding the most advantageous measurement directions.

Goldman et al. 2010 [14] used the same principle techniques as in Lensch et al. in [19–21] but mainly differed in the underlying assumption of the BRDF model, where Lensch et al. 2001 [20] expected analytical Lafortune models [18] where Goldman et al. 2010 [14] expected an underlying Ward model [34].

The proposed solution differs by using a static relative position of the camera and light source removing the need for any in-shoot equipment. This is done by calibrating the camera and light source prior to the acquisition and using this to simulate the irradiance on the surface rather than measuring it. Furthermore the geometry is captured using the acquired images.

Multiple Light Sources or Cameras Instead of having a moving camera or light source, a setup consisting of multiple light sources and/or cameras can be conceived. Kohler et al. 2013 [16], Noll et al. 2013 [27], and Noll et al. 2015 [28] were all based on such a device called the "OrcaM". This device allowed both reconstruction and SVBRDF based on the Ward model [34] and non-linear optimization [22]. The device had 633 static light sources mounted equidistantly in a spherical array. Further seven cameras were mounted together with a rotating glass plane which allowed for observations from both upper and lower hemispheres.

Tunwattanapong et al. 2013 [33] created a similar approach with a rotating arc containing 105 controllable LEDs enabling spherical harmonic illumination condition with longer exposure times. Five cameras and a turntable enable capture from different angles of the object.

Chen et al. 2014 [9] introduced a generalized linear light source reflectometry method using two sinusoidal lighting pattern, a camera, and a linear light source. This enabled anisotropic surface reflectance capture of small flat samples. Two prototype acquisition devices were proposed, one with the camera position fixed with respect to the material sample, and one where the camera was fixed with respect to the linear light source.

These methods differ from the approach proposed in this paper as they use highly specialized and inaccessible equipment whilst using a high amount of point like light sources. Furthermore the above mentioned methods are limited by the physical size of the setup. Our approach works with a single on-camera light source that is not limited to be point like. Our approach can be used in any scale setup, provided that the on-camera flash is powerful enough to compete with the natural light in the scene. Additionally the proposed approach can move to the object instead of needing the opposite.

Light-Weight Consumer Grade Setups Aittala et al. 2013 [2] aimed for a light-weight, inexpensive approach, using a rigid setup consisting of a display used for illumination and a camera. The display emitted certain fourier basis patterns, enabling capture of spatially varying BRDF modelled as a Mixture of Gaussian's of near-planer material samples.

Riviere et al. 2016 [32] also captured spatially varying BRDF for near-planar material samples, but they used mobile hardware, specifically the camera and flash unit. Rough specular BRDFs could be captured using this approach, however highly specular materials necessitated the use of an external light source, a screen on a tablet was used for this. Using polarization the diffuse and specular components were split and captured separately.

In the approach presented in Aittala et al. 2015 [3], the similarity of reflectance on several points were exploited, which enabled reflectance capture using only two images. One with a flash, and one without. Based on prior gathered observations, a regularized texture statistic transfer was performed fitting spatially varying BRDF to the respective data. The method was limited to flat material samples.

The methods presented have the limitation of only handling near-planar cases. This is not a limitation of the proposed system as the

irradiance is simulated using the reconstructed geometry and the extrinsic camera parameters following this.

3 Overview of Approach

An image of an object represents the appearance of the object under some given illumination conditions, and each pixel corresponds to a certain small surface patch, a "point", on the object. Given only an image the problem of determining per-point reflectance is ill-posed. Reflectance cannot be estimated from appearance alone. If an image shows a red surface we cannot know if it is a red surface in white light, or a white surface in red light. Popularly speaking the appearance is a product of 1) the reflectance at the point on the object and 2) the total irradiance at that point. We remind the reader that this paper is delimited to diffuse surfaces, as stated previously. The proposed approach are illustrated in Figure 2.

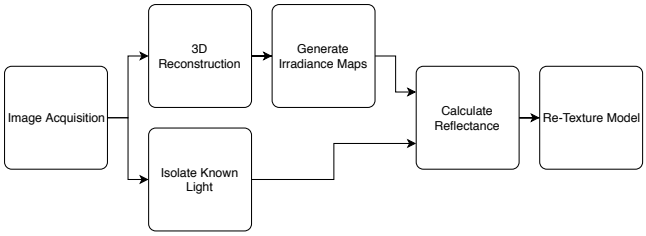

Figure 2: *The figure shows the main process of the proposed approach, starting from the image acquisition and ending with a textured, usable 3D object.*

As described in Section 1 our approach is based on a basic idea of capturing *two* images at each camera location (see Figures 3a and 3b). The first image is the object lit only by the unknown illumination conditions at the scene of capture. Let us call this the ambient illumination image. The second image is the object lit by both the ambient light *and* an on-camera flash. Let us call this the ambient plus flash illumination image. Subtracting the former from the latter results in an image of the object subject to flash-only illumination (see Figure 3c). The unknown ambient illumination cancels out completely.

Let us say that N acquisition points are used for capturing an object. The N ambient illumination images are fed to some SfM modelling software package, resulting in a triangulated 3D model of the scene. It also provides camera positions, orientations and intrinsic parameters of all N images. Since the illumination conditions from the known flash can be modelled, we can render an irradiance image for the i'th camera location, where each pixel represents the irradiance at the corresponding object surface point (see Figure 3d).

Given the irradiance image for the i'th image location it is simple to compute the per-point reflectances simply by dividing the flash-only image by the irradiance image (see Figure 3e). By doing this for each of the N images, these new reflectance images can be used, in place of the original images, in order to generate a diffuse reflectance texture atlas.

4 From Pixel Values to Diffuse Reflectances

The proposed approach rests on a number of assumptions. First and foremost, we limit ourself to surfaces that act as perfect diffuse reflectors i.e. follows the Lambertian reflectance model.

The object being captured is assumed to be static. Additionally the camera is assumed to have a linear response curve so that for any pixel a certain change in incoming radiance results in a proportional change in pixel value. Furthermore, reflected light from the environment, originating from the physical flash, is assumed not to be present.

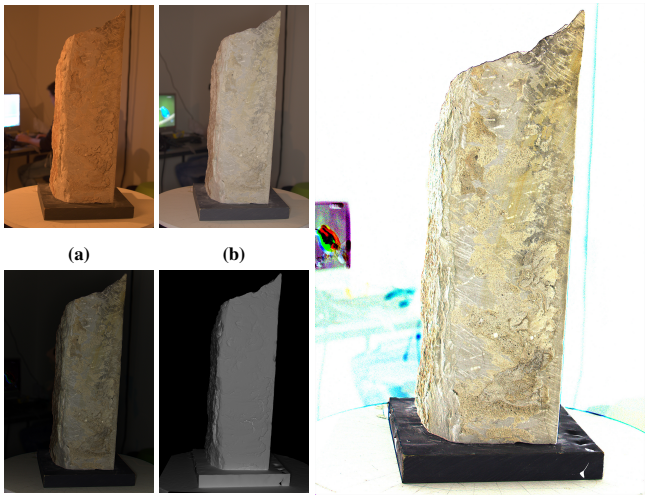

Figure 3: *Image (a) shows an acquisition image with only unknown ambient illumination. The warmness of the image stems from the white balance being set in regards to the on-camera flash. Image (b) contains illumination from both the unknown ambient light and the physical flash. Image (c) shows an image only containing illumination from the on-camera flash. Image (d) shows the irradiance image for the given view point. And lastly, image (e) shows the captured reflectance for the given view point.*

Section 3 provided a brief, visual presentation of the proposed approach to estimate spatially varying diffuse reflectances for reconstructed objects. This section will elaborate on the theoretical background for the approach. Let x denote a point on the surface of an object being reconstructed. Let $L_r(x)$ be the reflected radiance of that point; since we assume the surface is a diffuse reflector, $L_r(x)$ is the same in all directions on the hemisphere, Ω, defined by the surface normal of x. Let $P(x)$ then be the pixel value of the projection of x onto an image. Cameras essentially measure radiance [12], and hence, for a camera with a linear response curve, P will be proportional to $L_r(x)$ with a certain unknown factor of proportionality, S:

$$P(x) = S \cdot L_r(x) \tag{1}$$

Equation (1) is valid for each of the *RGB* color channels of the image, with color channel dependent radiance levels. In fact, all derivations presented in this section are per color channel. The reflected radiance, $L_r(x)$, is the appearance of the object at x. It depends on the material reflectance at x and the light hitting x. The relationship is determined by the Bi-directional Reflectance Distribution Function, BRDF or $f_r(x)$ [4, 12, 25, 29]:

$$L_r(x) = \int_\Omega f_r(x) \cdot L_i(x, \vec{\omega}_i) \cdot cos\,\Theta\, d\,\vec{\omega}_i \tag{2}$$

where $L_i(x, \vec{\omega}_i)$ is the incoming radiance at x from a certain direction $\vec{\omega}_i$. For diffuse surfaces the BRDF is naturally, especially simple, i.e., $f_r(x) = \rho_d(x)/\pi$, where $\rho_d(x)$ is the diffuse albedo at x. Thus, for diffuse surfaces Equation (2) degenerates to:

$$\begin{aligned} L_r(x) &= \frac{\rho_d(x)}{\pi} \int_\Omega L_i(x, \vec{\omega}_i) \cdot cos\,\Theta\, d\,\vec{\omega}_i \\ &= \frac{\rho_d(x)}{\pi} \cdot E(x) \end{aligned} \tag{3}$$

where $E(x)$ is the total irradiance at x. Combining Equations (1)

and (3):

$$P(x) = S \cdot \frac{\rho_d(x)}{\pi} \cdot E(x) \tag{4}$$

In this paper our objective is to estimate the diffuse reflectance, $\rho_d(x)$. The pixel values $P(x)$ we have access to from images; the camera response scale factor S is unknown, and the irradiance $E(x)$ is unknown, as it is the result of the unknown illumination conditions at the scene of acquisition. Let us now return to the acquisition strategy described in Section 3 where it was described that two images were acquired at each camera location: one with just the unknown scene illumination, and one with an on-camera flash in addition to the unknown scene illumination. In the following, subscript u will denote entities for the "unknown scene illumination" case, and $u+f$ will denote entities for the "unknown scene illumination plus flash" case.

$$P_u(x) = S_u \cdot \frac{\rho_d(x)}{\pi} \cdot E_u(x) \tag{5}$$

$$P_{u+f}(x) = S_{u+f} \cdot \frac{\rho_d(x)}{\pi} \cdot \big(E_u(x) + E_f(x)\big) \tag{6}$$

It is important to note, that the camera response scale factor is not necessarily the same in the two exposures, so they have also received subscripts. The reason for this is that since there is more light in the case when the flash is also fired, the camera settings, e.g., exposure time and aperture, may also need to change to get a properly exposed image. As the scaling factors can be different we are not allowed to subtract pixel values without compensating for this. Correcting for the difference in scale factors, subtracting unknown from unknown+flash images, and combining with Equations (5) and (6):

$$\begin{aligned} &\frac{S_u}{S_{u+f}} \cdot P_{u+f}(x) - P_u(x) = S_u \cdot \frac{\rho_d(x)}{\pi} \cdot E_f(x) \\ &\Updownarrow \\ &\frac{\rho_d(x)}{\pi} = \frac{\frac{S_u}{S_{u+f}} \cdot P_{u+f}(x) - P_u(x)}{S_u \cdot E_f(x)} \end{aligned} \tag{7}$$

Equation (7) show us that the diffuse reflectance, $\rho_d(x)$, can be expressed by pixel values from the two images, the camera scale factors and the irradiance generated by the flash. Hence, the unknown scene illumination cancels out and no longer has any influence at all. While we may not know the actual value of the camera scale factors, the *ratio* of the two camera scale factors, $\frac{S_u}{S_{u+f}}$, is known, as we can ensure that we alter the exposure in full f-stops. A full f-stop could be going from an exposure time of $1/250s$ to $1/500s$, and the ratio of the scale factors would then be 0.5. Similarly, if changing the aperture from $f/5.6$ to $f/8$, this would also result in a scale factor ratio of 0.5. The remaining element in Equation (7) is the irradiance generated by the flash, $E_f(x)$. We compute the per-point flash irradiances by rendering an "irradiance image" of the object, where each pixel value is the irradiance from the flash at the object point corresponding to the pixel. The subsequent section will describe how we perform this rendering, and how we calibrate the rendering so that it will provide the $S_u \cdot E_f(x)$ values that are needed to compute the diffuse reflectance using Equation (7).

4.1 Computing Flash Irradiance Images

Practically there are multiple factors that should be kept in mind at the acquisition of these image pairs, and throughout the process, in order to simplify and optimise the proposed approach. As the two images in an image pair should be aligned, the camera should be placed on a tripod and a remote trigger should be used. White balance should be calibrated to the on-camera flash, and be kept static, to simplify the flash calibration, which in turn makes the rendering of the irradiance simpler. The camera settings should be manually controlled, with a static small aperture (high f-stop) and low ISO, with a variable shutter speed. It is explained in Section 4 how an image pair using different settings can be used in order to minimize the error coming from limitation in data precision. By having a well exposed ambient light image the contribution of the on-camera flash can be captured more precisely. This paper limits the problem to use single exposure images, hence will not touch upon multi exposure HDR images, though it would be interesting to measure the improved precision related to this. It should be noted that all work in this paper uses 14bit RAW images. Futhermore, special care should be taken to not add gamma, color profiles etc. throughout the process as this would skew the resulting reflectance. In addition, the reconstructed object needs to be scaled in relation to the real world in order to correctly render the irradiance. This can be done using known points in the reconstruction. A ruler or pattern is useful for this.

Once the 3D model has been reconstructed by an SfM method using all the acquired unknown scene illumination images, the model can be used to render irradiance images.

We render these irradiance images using 3Ds Max [5], in conjunction with the V-Ray renderer [8]. We create a scene consisting of the reconstructed object, a virtual camera, and a virtual flash, (see Figure 4a). The virtual flash is modelled to match the emission surface of the actual, physical flash used in acquisition. For the reported experiments we have used a Godox Witstro AR400 ring flash. The virtual camera is placed using camera calibration information produced by the SfM software as part of the reconstruction process. In the reported experiments we have used ContextCapture by Bentley [6] due to showing good results in Nikolov and Madsen 2016 [26]. ContextCapture produces an .xml file with position, orientation, and internal camera parameters (e.g., focal length) for each acquisition image. For each camera location we set up a virtual camera in the 3Ds Max scene, and place the virtual flash accordingly.

The reconstructed object in the virtual scene is given an ideal diffuse material (unity reflectance). From eq. (3) we see that a unity BRDF will render to a reflected radiance in $W/m2 \cdot sr$ which is equal to the irradiance from the flash in W/m2, i.e. the rendered radiance values can be treated as irradiance values. Now, all we need is to determine how powerful the physical flash is. To calibrate the virtual flash to the physical flash we performed the following calibration process.

4.2 Virtual Flash Calibration and Validation

To simplify the flash calibration the falloff cone was assumed to be constant, hence that a measurement at a single distance of the falloff cone would be enough to fully characterize the shape. Furthermore, the distance at which the calibration and validation is conducted at, should preferably be about the distance used in the acquisition process in order to minimize the impact of this simplification. We modelled the physical flash's emission surface geometrically, using the entire surface as the light emitter (see Figure 4b). The simulated flash was validated using a series of images of a moving constant target giving a measured falloff profile. The target was chosen as it had diffuse properties, minimizing specular reflections. This was compared to rendered test images at equal distance to a virtual flat surface. The intensity was calibrated and validated using a known reflectance target at measured extrinsic parameters, though it should be noted that the target was of older date, hence the relfectance on the target contained some unknown error, resulting in some uncertainty in the intensity calibrations. This was however not a problem with the method, and should therefore result in precise intensity levels using any known reflectance. This results in a virtual flash that is calibrated in shape and intensity to the known on-camera flash, which thereby enables the simulation of a per point irradiance on the reconstructed object (see Figure 4c).

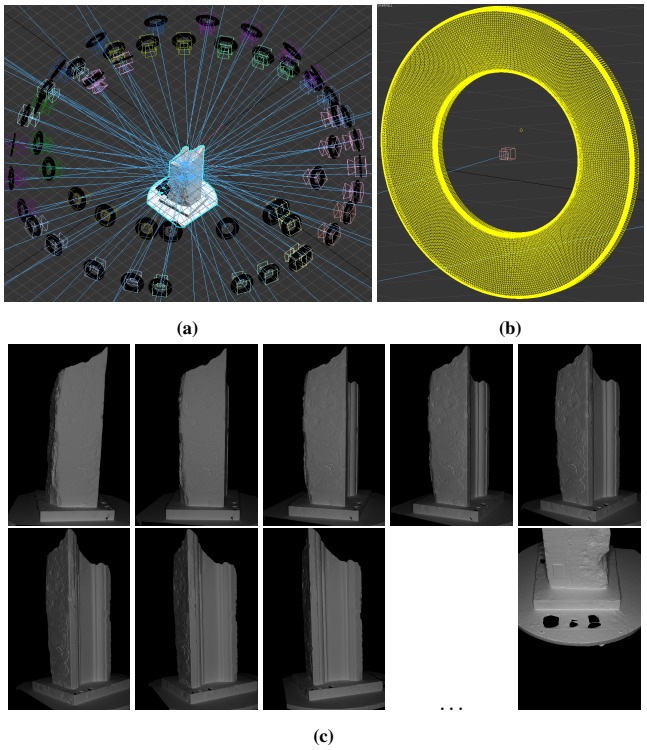

Figure 4: *Image (a) shows a scene with a reconstructed object in the middle and the light source and camera pairs. Image (b) shows a geometric model of the physical flash. Image (c) shows a series of irradiance images from different view points, rendered from the same scene, using the flash model seen above.*

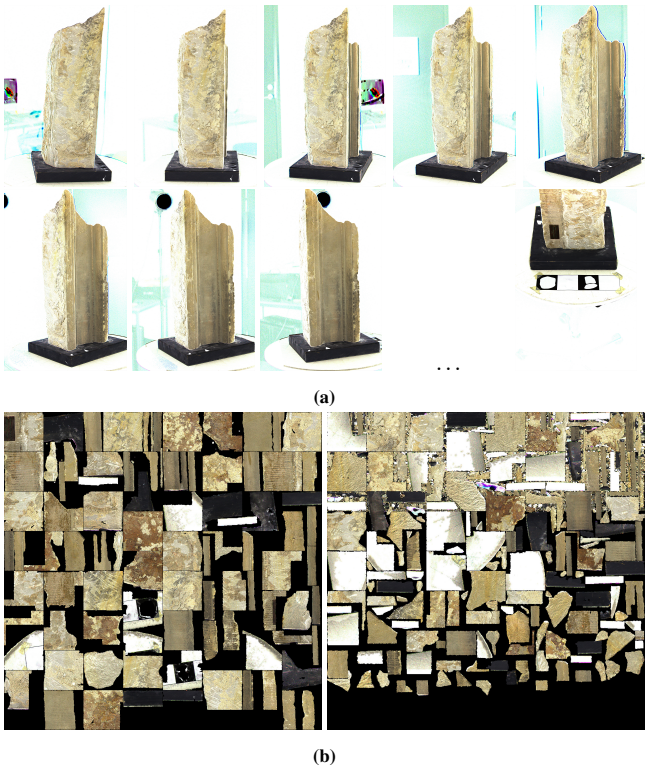

Figure 5: *Figure (a) shows a series of per view reflectance images. Figure (b) shows the texture maps generated from the above reflectance images.*

4.3 Constructing Diffuse Reflectance Texture Maps

With the calibrated flash, and using the appropriate virtual camera settings, the irradiance images represent $S_u \cdot E_f(X)$, as described in Sections 4 and 4.1. Let $P_i(x) = S_u \cdot E_f(X)$ and Equation (7) then become:

$$\frac{\rho_d(x)}{\pi} = \frac{\frac{S_u}{S_{u+f}} \cdot P_{u+f}(x) - P_u(x)}{P_i(x)} \tag{8}$$

The result is a per viewpoint reflectance image (see Figure 5a). In order to generate the corresponding texture map with reflectance instead of appearance, the original images are swapped with the reflectance images. The software is tricked into regenerating the texture by deleting files that indicated the process was done, hence generating a new texture map, from the reflectance images (see Figure 5b).

5 Results

The data sets were captured using a Canon 5Ds paired with a Canon EF 70-300mm f/4-5.6L IS USM lens with a Godox Witstro AR400 ring flash.

As we do not have access to any precise reflectance standard, we conducted two experiments showing that the difference in pixel values created by uneven lighting can be significantly reduced while demonstrating that a material gets the same reflectance estimated regardless of the unknown ambient illumination. Specifically we placed a clay brick in lighting conditions with a powerful light on one side and soft light on the other. Naturally the estimated reflectance should be the same on both sides of the brick.

First, in order to validate that the two sides of the brick had equal mean reflectance, an image of each side was captured under the exact same lighting condition (see Figures 6a and 6b). The measured difference of the two sides for each color channel was: $\overline{R} = 0.001$, $\overline{G} = 0.005$, $\overline{B} = 0.004$, with standard deviations of: Side A:$\sigma_R = 0.020$, $\sigma_G = 0.010.$, $\sigma_B = 0.012$ and Side B: $\sigma_R = 0.021$, $\sigma_G = 0.012.$, $\sigma_B = 0.011$. This confirms that the reflectance on both sides of the red clay brick are very close, both in mean values and standard deviations.

The brick was then placed in a setup with a prominent light source illuminating one side of the brick, leaving the opposite side in shadow (see Figure 6c). We then used the proposed approach to capture the reflectance of the object. The measurements for the ground truth, appearance, and reflectance can be seen in Table 1.

Comparing the Δ, for the calculated reflectance, we observe an improvement in all three color channels. The biggest difference being in the red color channel which is ≈ 3 times closer to the ground truth, going from an error of 0.046 to 0.013. The mean color difference in the green and blue channels are both at 0.001 which is lower than the ground truth's value of 0.005, 0.004 respectively. However the difference between the Δ of the green and blue are closer to the ground truth than the same values in the appearance measurements.

The standard deviation are about halved between the ground truth and the appearance measurements. This reduction is negated in the calculated reflectance, giving standard deviation value that have a maximum error of 0.004 in σ_R on the B side. The corresponding maximum error in the appearance is 0.008 but are in general also higher with a absolute mean error of all the standard deviations of 0.007 compared to 0.001 for the calculated reflectance.

A bust was captured in three different lighting conditions, inside a lab with static lighting, inside in a seminar room with large windows in two of the four walls, and lastly outside on a sunny day in the shadow of a building. By comparing identical areas of the

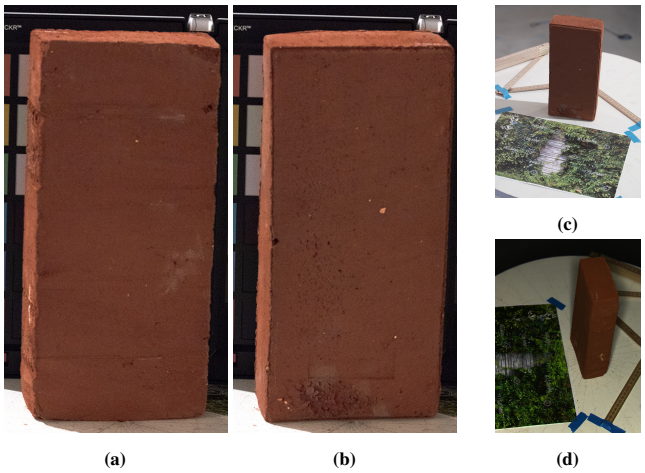

Figure 6: *Images (a) and (b) show the two sides of the brick in the same illumination. Image (c) shows the test setup consisting of the red brick, and piece of paper with a printed texture and marks to calculate scale, and a measuring stick to validate the scale. Image (d) shows the same scene from a different angle, only illuminated by the physical flash.*

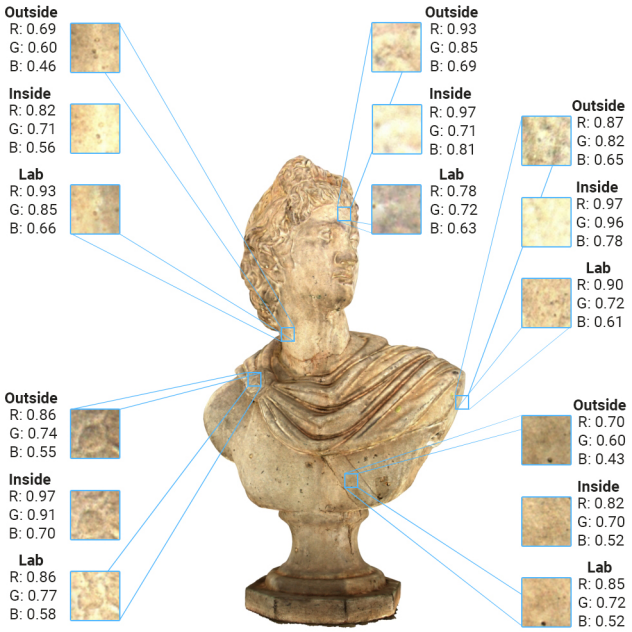

Figure 7: *The figure shows the patches used to investigate the proposed methods invariance to unknown lighting conditions and general reproducibility. The position of the five patches are marked, with the three patches and their mean R, G, and B reflectance values.*

reconstructed objects (see Figure 7 for the location of the areas) and finding the maximum error of these areas, the proposed methods invariance to lighting environments can be evaluated.

The mean reflectance and standard deviation was calculated for each patch and then used to find the biggest error, in this case the biggest difference between the three lighting scenarios. The biggest error in the mean reflectance was between for the patch on the arm between the outside and laboratory light scenarios with an error of 0.23. The mean error across all the patches are however lower at

Ground Truth						
	$\overline{R}$	$\overline{G}$	$\overline{B}$	σ_R	σ_G	σ_B
A	0.237	0.106	0.047	0.020	0.010	0.012
B	0.236	0.111	0.051	0.021	0.012	0.011
Δ	0.001	0.005	0.004			

Apperance						
	$\overline{R}$	$\overline{G}$	$\overline{B}$	σ_R	σ_G	σ_B
A	0.155	0.026	0.009	0.014	0.005	0.004
B	0.108	0.018	0.005	0.018	0.004	0.003
Δ	0.047	0.008	0.004			

Calculated Reflectance						
	$\overline{R}$	$\overline{G}$	$\overline{B}$	σ_R	σ_G	σ_B
A	0.302	0.138	0.060	0.019	0.013	0.013
B	0.288	0.137	0.061	0.025	0.012	0.010
Δ	0.014	0.001	0.001			

Table 1: *The measured values of the two vertical face pairs on the brick. $\overline{X}$ being the mean value of the given color channel, and σ_X being the standard deviation of the given color channel.*

0.16. The standard deviation from the patches are closer to each other, with the highest error observed on the neck patch with 0.06 The mean error in standard deviation is 0.03 (See Table 2 for specific numerical values).

Comparing the three versions of the bust visually (see Figure 8) there is only minor observable differences which can be credited to the difference in quality of the reconstructed mesh itself.

In order to assess the visual improvements of using the calculated reflectance based texture maps, as opposed to appearance based texture maps, the reconstructed objects are illuminated with a series of lighting conditions (see Figure 9) on the very last page of the paper.

Using the appearance for the texture maps is the same as using the direct output from an SfM pipeline, this would be the naive use of a reconstructed object. Inspecting the objects where the appearance is used as texture maps (see the objects on the left of each image in Figure 9) the object is darker. Additionally, shadows present from the unknown lighting at the time of acquisition are still present as part of the texture maps, this is especially noticeable on the Roman Beam in Figure 9.

These effects are especially noticeable when compared to the corresponding renders with the calculated reflectance used as the texture maps where these artifacts are no longer present. Furthermore, the objects using the calculated reflectance reacts and blends better with the lighting environments.

6 Discussion

The proposed approach shows significant improvements in both the difference of mean color values between side *A* and *B*, and the standard distribution of the before mentioned color values. The difference of mean color values is showing both a reduction in error and the relation between each other compared to the appearance based measurements that have a tendency towards the red channel. Comparing results from the brick and the patches, there is an increase in error. This could be due to the need of reconstructing the mesh and estimating camera positions for each lighting scenario. This could lead to error propagating to the irradiance generations and hence the reflectance calculation. This error would however decrease with the improvement of the SfM pipeline.Looking at the quantitative results, there is still room for improvements. The error in Δ of the green and blue channels, compared to the ground truth, though the relations between them are improved, could be due to not handling inter-reflections from the surface that the object of interest was placed

	Broach		Chest		Neck		Arm		Forehead	
	Mean	STD	Mean	STD	Mean	STD	Mean	STD	Mean	STD
Outside	0.74	0.07	0.70	0.06	0.77	0.05	0.81	0.07	0.71	0.07
Inside	0.86	0.06	0.68	0.05	0.90	0.03	0.70	0.08	0.91	0.04
Laboratory	0.72	0.08	0.58	0.05	0.78	0.09	0.58	0.08	0.82	0.06
Max Error	0.14	0.02	0.12	0.01	0.13	0.06	0.23	0.01	0.20	0.03

Table 2: *The table shows the mean reflectance value and standard deviation for each patch of three capturing environments. The max error is calculated to show the worst case for each patch.*

Figure 8: *The figures show the results of capturing the same object in three difference environments. Left: Captured inside a lab with static lighting only. Middle: Captured inside seminar room with mixed static and dynamic lighting coming from outside facing windows. Right: Captured outside on a sunny day in the shadow of a building.*

on, the precision of the flash calibrations, or inaccuracies in the 3D reconstruction. The inter-reflections could be handled developing an iterative method of refining the estimated irradiance, similar to Dong et al. 2014 [11], Xia et al. 2016 [36], or Madsen et al. 2011 [23]. The impact of each step would need further investigation.

Looking at the standard deviations, the appearance measurements are lower, which indicates a reduction in variation of the texture, flattening of the colors, or loss of contrast. This is improved in the calculated reflectance with a absolute mean error of all the standard deviations of 0.001 compared to 0.007 for the appearance. This should give a better representation of the variations in the texture of the object.

In the patch test, the maximum mean error of 0.23 combined with the low maximum standard deviation error of 0.06 indicates that the error stems from some kind of scaling factor. This could originate from error in the camera estimation from the reconstruction leading to an global under or over simulation of the irradiance. This is further supported as there is no noticeable color shift in the three scenarios, pointing towards this scaling factor instead of the unknown lighting being the reason. If the unknown lighting was the reason, a shift towards either blue or red between the the reconstructions would have been observable due to differentiating color temperatures. Furthermore, it is shown that the proposed method also works for certain outdoor lighting scenarios.

Inspecting the 3D objects, utilizing the reflectance calculated by our approach, in different synthetic lighting scenarios, a significant improvement can be observed. The mismatch in intensity and error stemming from inconsistent lighting at the time of acquisition, are present in the 3D objects using the appearance as texture but are not observable in the 3D objects using the calculated reflectance as textures. Furthermore the before mentioned improvements in standard deviations can be observed by the increased variation in the textures between the two.

7 Conclusion

The presented approach captures spatially varying diffuse reflectance in an SfM pipeline by capturing one additional image per image position, and simulating the per pixel irradiance for each position. Quantitatively, as well as qualitatively, it is shown to be an improvement compared to appearance based texture maps. Using only a minimal amount of extra consumer grade equipment and acquisition work for SfM techniques, this enables anyone to easily digitize physical object to use as 3D content for a multitude of purposes, f.eks. AR or VR. Additionally the proposed technique is applicable to unconstrained, unknown illumination condition scenarios, including limited outdoor scenarios. Future work would include incorporating specular properties and handling inter-reflections.

Acknowledgments

This work is funded by the DARWIN project under the Innovation Fund Denmark, case number: 6151-00020B. The GPU for the calculations in this work was provided by the NVIDIA GPU Grant, which is gratefully acknowledged.

REFERENCES

[1] Agisoft. Agisoft: Metashape. http://www.agisoft.com/, 2010. Accessed: 2018-09-06.

[2] M. Aittala, T. Weyrich, and J. Lehtinen. Practical svbrdf capture in the frequency domain. *ACM Trans. Graph.*, 32(4):110–1, 2013.

[3] M. Aittala, T. Weyrich, J. Lehtinen, et al. Two-shot svbrdf capture for stationary materials. *ACM Trans. Graph.*, 34(4):110–1, 2015.

[4] T. Akenine-Möller, E. Haines, and N. Hoffman. *Real-Time Rendering (2nd Edition)*. A K Peters/CRC Press, 2002.

[5] Autodesk Inc. 3ds max.

[6] Bentley. Bentley: Contextcapture. https://www.bentley.com/, 2016. Accessed: 2018-09-06.

[7] CapturingReality. Capturingreality: Reality capture. https://www.capturingreality.com/, 2016. Accessed: 2018-09-06.

[8] Chaos Software. V-ray.

[9] G. Chen, Y. Dong, P. Peers, J. Zhang, and X. Tong. Reflectance scanning: estimating shading frame and brdf with generalized linear light sources. *ACM Transactions on Graphics (TOG)*, 33(4):117, 2014.

[10] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia. MeshLab: an Open-Source Mesh Processing Tool. In V. Scarano, R. D. Chiara, and U. Erra, eds., *Eurographics Italian Chapter Conference*. The Eurographics Association, 2008. doi: 10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136

[11] Y. Dong, G. Chen, P. Peers, J. Zhang, and X. Tong. Appearance-from-motion: Recovering spatially varying surface reflectance under unknown lighting. *ACM Transactions on Graphics (TOG)*, 33(6):193, 2014.

[12] P. Dutre, P. Bekaert, and K. Bala. *Advanced Global Illumination, Second Edition*. A K Peters/CRC Press, 2003.

[13] E. Eisemann and F. Durand. Flash photography enhancement via intrinsic relighting. In *ACM transactions on graphics (TOG)*, vol. 23, pp. 673–678. ACM, 2004.

[14] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, 2010.

[15] Z. Hui, A. C. Sankaranarayanan, K. Sunkavalli, and S. Hadap. White balance under mixed illumination using flash photography. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–10. IEEE, 2016.

[16] J. Kohler, T. Noll, G. Reis, and D. Stricker. A full-spherical device for simultaneous geometry and reflectance acquisition. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 355–362. IEEE, 2013.

[17] D. Krishnan and R. Fergus. Dark flash photography. *ACM Trans. Graph.*, 28(3):96–1, 2009.

[18] E. P. Lafortune, S.-C. Foo, K. E. Torrance, and D. P. Greenberg. Non-linear approximation of reflectance functions. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 117–126. ACM Press/Addison-Wesley Publishing Co., 1997.

[19] H. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics (TOG)*, 22(2):234–257, 2003.

[20] H. P. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel. Image-based reconstruction of spatially varying materials. In *Rendering Techniques 2001*, pp. 103–114. Springer, 2001.

[21] H. P. Lensch, J. Lang, A. M. Sá, and H.-P. Seidel. Planned sampling of spatially varying brdfs. In *Computer Graphics Forum*, vol. 22, pp. 473–482. Wiley Online Library, 2003.

[22] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[23] C. B. Madsen and B. B. Lal. Estimating outdoor illumination conditions based on detection of dynamic shadows. In *International Conference on Computer Vision, Imaging and Computer Graphics*, pp. 33–52. Springer, 2011.

[24] G. Nam, J. H. Lee, D. Gutierrez, and M. H. Kim. Practical svbrdf acquisition of 3d objects with unstructured flash photography. In *SIGGRAPH Asia 2018 Technical Papers*, p. 267. ACM, 2018.

[25] F. E. Nicodemus. Directional reflectance and emissivity of an opaque surface. *Applied optics*, 4(7):767–775, 1965.

[26] I. Nikolov and C. Madsen. Benchmarking close-range structure from motion 3d reconstruction software under varying capturing conditions. In *Euro-Mediterranean Conference*, pp. 15–26. Springer, 2016.

[27] T. Noll, J. Kohler, G. Reis, and D. Stricker. Faithful, compact and complete digitization of cultural heritage using a full-spherical scanner. In *Digital Heritage International Congress (DigitalHeritage), 2013*, vol. 1, pp. 15–22. IEEE, 2013.

[28] T. Nöll, J. Köhler, G. Reis, and D. Stricker. Fully automatic, omnidirectional acquisition of geometry and appearance in the context of cultural heritage preservation. *Journal on Computing and Cultural Heritage (JOCCH)*, 8(1):2, 2015.

[29] U. S. N. B. of Standards and F. E. Nicodemus. *Geometrical considerations and nomenclature for reflectance*, vol. 160. US Department of Commerce, National Bureau of Standards, 1977.

[30] O. Özyeşil, V. Voroninski, R. Basri, and A. Singer. A survey of structure from motion*. *Acta Numerica*, 26:305–364, 2017.

[31] G. Petschnigg, R. Szeliski, M. Agrawala, M. Cohen, H. Hoppe, and K. Toyama. Digital photography with flash and no-flash image pairs. *ACM transactions on graphics (TOG)*, 23(3):664–672, 2004.

[32] J. Riviere, P. Peers, and A. Ghosh. Mobile surface reflectometry. In *Computer Graphics Forum*, vol. 35, pp. 191–202. Wiley Online Library, 2016.

[33] B. Tunwattanapong, G. Fyffe, P. Graham, J. Busch, X. Yu, A. Ghosh, and P. Debevec. Acquiring reflectance and shape from continuous spherical harmonic illumination. *ACM Transactions on graphics (TOG)*, 32(4):109, 2013.

[34] G. J. Ward. Measuring and modeling anisotropic reflection. In *ACM SIGGRAPH Computer Graphics*, vol. 26, pp. 265–272. ACM, 1992.

[35] C. Wu. Visualsfm : A visual structure from motion system. http://ccwu.me/vsfm/, 2011. Accessed: 2018-09-06.

[36] R. Xia, Y. Dong, P. Peers, and X. Tong. Recovering shape and spatially-varying surface reflectance under unknown illumination. *ACM Transactions on Graphics (TOG)*, 35(6):187, 2016.

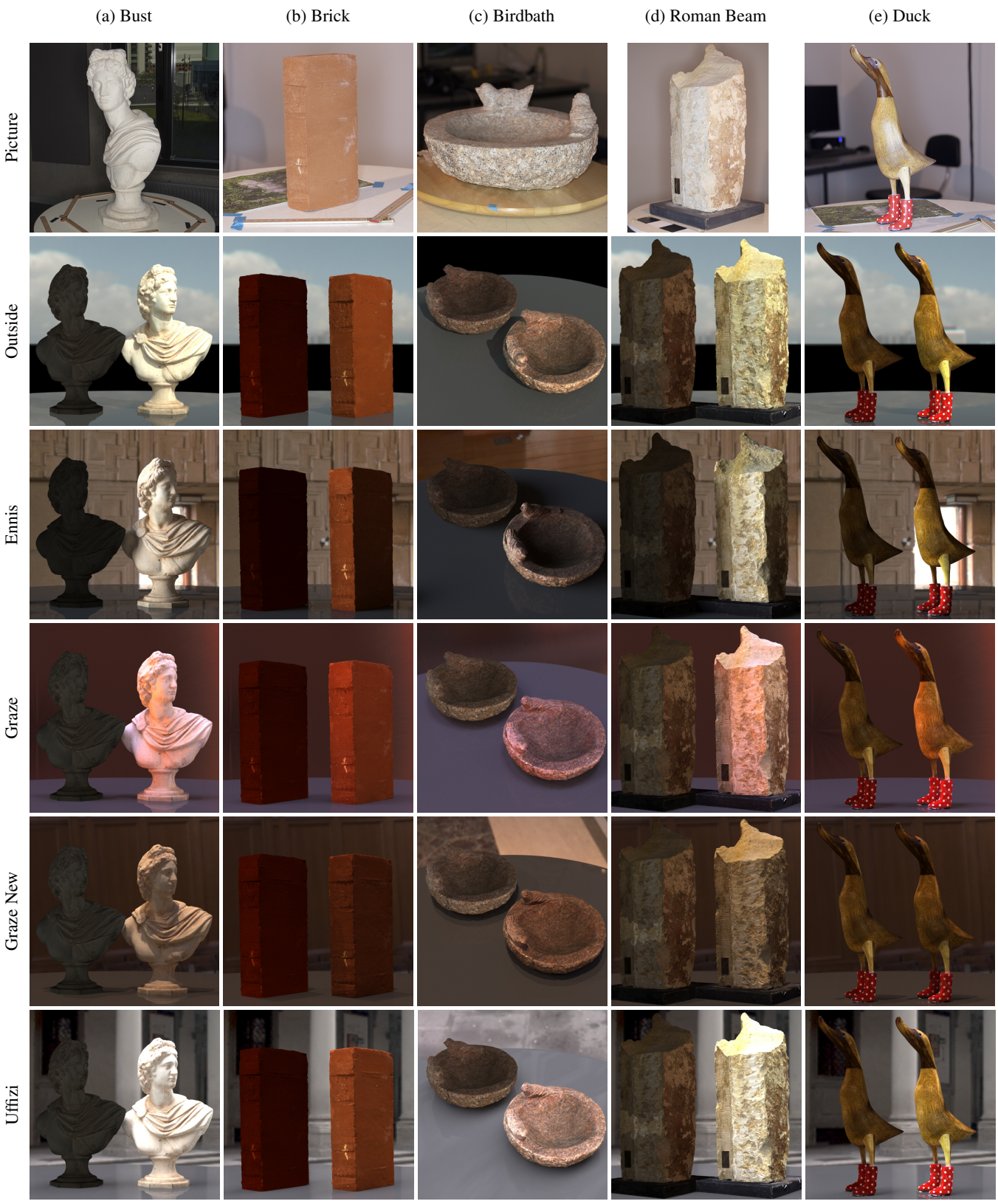

Figure 9: *The figure shows the results of our method. Each test object (column) is shown with a picture from the acquisition (row 1) followed by five different lighting environments (row 2-5). In all the images, the left object is using the appearance as texture. This represents the output of an SfM pipeline. The right object in each image are using the reflectance calculated by our proposed method.*