12 February 2025 ·
Qeios

Peer Review

Review of: "Acquisition of Spatially-Varying Reflectance and Surface Normals via Polarized Reflectance Fields"

Ruggero Pintus[1]

1. Visual and Data Intensive Computing, CRS4 Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna, Pula, Italy

In this paper, the authors present a novel approach for acquiring spatially varying reflectance and geometry using a polarization-based method combined with a statistical analysis algorithm. They stated that they introduce a unique acquisition setup and contribute a dataset of polarized reflectance fields, providing a valuable resource for material modeling and appearance research.

Overall, the paper lacks technical soundness and exhibits several inconsistencies. Furthermore, the presentation throughout the writing is quite lacking. For instance, the presentation of an acquisition setup with eight cameras without utilizing them for any computation or results does not make any sense. Additional cameras could have been leveraged to mitigate occlusions, reduce flare, and average specular interreflections from the light rig, yet no contribution in this regard is presented. Furthermore, beyond the fact that certain aspects are unclear or debatable, the approach does not appear to offer any novel contributions. Moreover, while the abstract mentions a dataset, no link to it seems to be provided in the paper (unless it was overlooked). Lastly, it remains unclear what new comprehensive statistical analysis is proposed beyond standard optimization solvers.

===== RELATED WORK =====

The related work section provides a sufficient overview of the state-of-the-art literature relevant to the presented approach. However, it lacks a clear motivation, comparison, and discussion of how the current work differs from or improves upon existing methods. As a result, the novelty of the paper is not immediately apparent, leaving the reader to infer it from the rest of the paper and prior knowledge

 doi.org/10.32388/WP1CH4

of the field. To address this, a dedicated paragraph or additional commentary within the existing sections should be included to better highlight these aspects.

===== METHOD =====

----- Overexposure -----

The sentence “...For a particular surface point, the intensity observation should exhibit continuity within a specific range rather than displaying an emergent highlight...” is misleading. It implies that the presence of a highlight contradicts the continuity of intensity, which is not accurate. First, a highlight is not the same as overexposure, though they may be related in low dynamic range acquisition setups. In high dynamic range acquisitions, highlights can appear without overexposure. Second, the emergence of a highlight is a result of the material’s interaction with light and can occur even with a continuous specular lobe. Therefore, continuity in intensity is maintained both in the presence of highlights and without overexposure.

Similarly, the sentence “...shown in Fig. 2, the intensity variations of three surface points under OLAT illumination clearly reveal the recognition of overexposure through strong pulses in the sequence...” is inaccurate. Overexposure in imaging occurs when excessive light reaches the sensor, causing a loss of detail in highlights, which appear "washed out" or completely white. However, the signal in Fig. 2 remains within the dynamic range of the acquisition device (i.e., 0.175), meaning it is not overexposed and can still be used for computation. While the color reprojection in the figures may make it appear completely white, if the OLAT intensities are correctly measured, then the signal itself is not overexposed.

The type of signal observed in Fig. 2 should not be removed outright but rather it could be considered in two different ways: (1) as an outlier if a Lambertian model is used, or (2) as an inlier if a more complex model that accounts for specularity, such as the Ward model, is employed. Furthermore, removing overexposure is relatively straightforward, as it can be done by setting a conservative confidence threshold below the saturation level. Conversely, in Fig. 2, while those are not overexposures, this is particularly simple only because the mirror ball exhibits either very high or zero values. However, in the case of a typical object with both diffuse and specular components, distinguishing between points affected by specularity and those that are not—without actual overexposure—becomes challenging. In robust frameworks, values are typically sorted from darkest to brightest, and after discarding dark values (shadows) and excessively bright values (overexposures), the lowest 50% are retained. This ensures that at least 50% of the values are inliers, as having fewer than that would lead to an ill-posed problem.

Moreover, the heuristic threshold used is unclear. The authors claim that this value can be easily determined during capture, yet they provide no explanation of how to do so. Stating that the threshold is solely related to the light source is misleading. If the goal is to remove highlights, the threshold depends on the surface's diffuse and specular behavior rather than the light itself. If the aim is to filter out overexposure, it is determined by the capture setup, not the light source. While reducing light intensity can prevent overexposure, highlights will still be present. Overexposure is dictated by the capture setup, particularly whether a low or high dynamic range acquisition is used. The claim that determining this threshold is straightforward is questionable, as it highly depends on the interaction between the object and its materials. A mirror ball is an idealized, very easy, and cooperative test case, but in a more general scenario, the authors should justify why finding this threshold is easy. Without such an explanation, it is reasonable to assume (as usual) that the process of removing specular highlights from the signal is actually quite challenging.

That said, I have significant concerns about the approach of handling overexposure by replacing overexposed values with computed ones. In general, in multi-view and multi-light setups, where reflectance is sampled from various viewpoints and lighting directions, overexposed or underexposed signals—being less reliable—are typically treated as outliers and excluded from computation rather than being substituted with artificial (and potentially incorrect) values. Given that each pixel or object point is observed from multiple cameras, a large and redundant set of measurements is available. Thus, removing unreliable values rather than replacing them with synthesized ones would likely be a more robust, reliable, and feasible strategy. Otherwise, the introduced replacement values could become another source of error.

Finally, as mentioned earlier, highlight information is not necessarily as problematic as it might seem. While a saturated numerical value prevents precise computation, it still provides a lower bound on the light intensity under those conditions. This information could be leveraged to better constrain the estimation of material properties, particularly in the context of SVBRDF computation.

----- Interreflections -----

I also have several concerns regarding the interreflection removal (or mitigation). First, no example of interreflection mitigation has been provided. Equation (4), which approximates interreflections, is defined only in the lower hemisphere, but interreflections can also occur when light comes from the

doi.org/10.32388/WP1CH4

upper hemisphere, with the opposite reflection originating from the lower hemisphere. The formulation is heavily dependent on the capture setup, making it not universally applicable. Additionally, this approach only considers specular interreflections caused by light sources reflecting light from other sources.

From a mathematical perspective, equation (4) requires further clarification. First, it is not universally true that when light comes from the lower hemisphere, the object is not illuminated by the principal direction (as stated in the sentence: "...When the lighting arises from the lower hemisphere, the object becomes unobservable due to the absence of incoming radiance..."). This depends on the setup. In fact, Fig. 3 contradicts this, as the light producing the interreflection on the right also illuminates the sphere on the left (and it is slighly below the horizontal plane). Additionally, it is unclear why, besides the weighting max function, the interreflection term has the dimensionality of the square of the image—i.e., multiplying the image minus the average ambient noise by the image itself. How is the average ambient noise computed, and why is it included in this calculation?

Furthermore, claiming that interreflections can be alleviated simply by imposing $n \cdot \omega i > 0$ is incorrect. While it's true that if $n \cdot \omega i$ is less than zero with a non-zero signal, we can infer some global effect and decide to discard unreliable measurements for a particular pixel illuminated by a specific light, interreflection removal is a much more complex issue. It cannot be addressed merely by enforcing light-point visibility. Moreover, depending on the light setup or the light source itself (bodies such as a hemisphere surrounding a point light), interreflections can still occur even when point-light visibility is present. This might not be the case for the light rig and light bodies used in the experiment, but this rule is not universally applicable.

----- Lens Flare -----

It is straightforward to demonstrate that $\omega i \cdot \omega o<0$ does not necessarily imply that $n \cdot \omega i<0$. In fact, it is quite easy to show that there exists a configuration of view direction, normal, and light direction where the following conditions hold:

- $\omega i \cdot \omega o<0$
- $n \cdot \omega i >0$
- $n \cdot \omega o >0$ (visible)

This is also evident in Fig. 3(c). The upper left part of the sphere is entirely affected by lens flare, yet highlights are visible. By definition, the presence of highlights indicates that there is a path where the

doi.org/10.32388/WP1CH4

point is visible from both the light and the viewing direction. Although this point is influenced by the lens flare, the signal is dominated by the highlight due to the mirror ball. In a more general case, such as with a diffuse material, this point would be visible from both the light and the camera, and the lens flare would constitute the largest part of the reflectance signal.

Therefore, claiming that lens flare is addressed by simply imposing point visibility from the light source is incorrect.

----- Occlusion -----

Similarly, while it seems reasonable that "...The measured occlusion map can be used to eliminate shadows from the albedo map...", the exact method of doing so is unclear. Moreover, the occlusion is computed in an unconventional way. Typically, occlusion computation should not only depend on the point in question, but also on its surrounding neighborhood (as a simple first approximation). Therefore, equation (5) is difficult to understand and justify. This equation only takes into account the view direction (omega_i), the outgoing direction (omega_o), and the normal (n) of the point, without considering the neighborhood. As a result, equation (5) computes occlusion based on the intersection of the upper hemisphere of the point (with its normal) and the viewing solid angle, without accounting for other parts of the scene that could cause occlusion. For instance, a visible point on the rightmost part of a sphere, which should have no occlusion, would still be considered highly occluded. Thus, based on equation (5), what is computed is more related to the concept of surface foreshortening than to local occlusion. It would be more appropriate for the section to focus on weighting based on local surface foreshortening rather than referring to it as occlusion.

----- Final remarks on method -----

In conclusion, the preprocessing steps presented are problematic, debatable, and not technically robust or generally applicable. The authors should work on improving the clarity of these sections to enhance reader understanding and ensure reproducibility. Additionally, even if the preprocessing were to work under specific setups, light configurations, and object shapes, it remains challenging to identify any novel contributions that would justify publication.

The computation of Anisotropy and Roughness presented in the paper is well-established. Aside from the existence of several alternative methods, the primary issue lies in how to solve the equation and ensure its robustness to noise in the dataset. However, no description of the (new?) solver is provided, nor is there any discussion on how to optimize or accelerate the one- or two-dimensional search for these Ward model parameters.

===== RESULTS =====

The colormap used to depict the normal map appears unusual. Either the normal maps are incorrectly computed or the mapping itself is flawed, or at least not standard. For instance, in figure 4a2, the normal pointing to the right should correspond to a vector (1, 0, 0), and the one pointing to the left should correspond to (-1, 0, 0). These would typically map to RGB values of (1, 0.5, 0.5) and (0, 0.5, 0.5), respectively. However, in the figure, they are mapped to (1, 0, 0) and (0, 0, 1), which correspond to red and blue. The authors should provide an explanation for their normal color mapping.

The process for computing the reference image is unclear. First, the environment map used for the reference differs from that used in the rendering (see backgrounds). The backgrounds of the two images are noticeably different, despite both spheres reflecting the same background in the rendering. Additionally, a reference image should ideally be a real photograph, which is then compared to a rendering obtained by virtually computing the appearance using the same light position and material properties.

The paper does not clarify how the errors are computed, and it remains unclear whether there are any ground truth values used for comparison. It is also not surprising that removing saturated values would lead to an improvement in the final result. Additionally, despite the authors stating that the paper "provides a valuable training dataset for emerging research in inverse rendering," there is no reference or link provided to any dataset, such as a webpage where the dataset can be accessed.

===== PRESENTATION =====

The paragraph numbering is unclear and inconsistent throughout the paper. Some sub-paragraphs are not numbered, such as "Ablation Study" and "Overexposures," while others, like "Optimization" and "Polarized OLAT Capture," are numbered. Please ensure more consistency in the numbering format.

The authors state "...highly accurate surface normals (within 0.1mm/px)...". However, the accuracy of a normal is typically measured by angular error, not in mm/px. The value in mm/px refers to the resolution, which depends on the camera resolution and capture setup (working distance, object size, etc.), not on how accurately the normal is computed.

Some sentences misrepresent the literature they reference. The authors should verify the methods before citing them. For example:

- The authors state, "...Recent advancements in neural renderings, such as NeRF [1] and Gaussian Splatting [2]...". It is important to note that 3D Gaussian splatting is not a neural rendering technique. It does not use a neural network; instead, it models scenes explicitly with a set of 3D Gaussians that store color, opacity, and anisotropic covariance.
- The authors say, "...Pioneer techniques [8][9][10] combine multi-view 3D reconstruction and photography under diverse illumination to measure the geometry and spatially-varying Bidirectional Reflectance Distribution Function (SVBRDF) of real-world objects. Once measured, the models become renderable from any perspective, enabling a faithful representation of digital models in virtual environments...". None of these methods involve multi-view 3D reconstruction; they apply photometric stereo, using one viewpoint and many different illumination conditions (MLICs, Multi-Light Image Collections). Additionally, none of these methods allow rendering from "any perspective"; rather, they are limited to one perspective. Only [8] computes the BRDF, while the others compute only the surface normal map (geometry).

In addition to visually presenting the OLAT in terms of cross-polarized and parallel-polarized images, I would also include the visual presentation of Lambda_d and Lambda_s images to demonstrate how effectively the two signals have been separated.

In Table 1, should it include only entities that are not defined elsewhere, or should it also include indirect entities? For example, the captured OLATs are defined by the symbols of the captured images (see the equations in the first part of section 4.1). Similarly, equation (3) defines the diffuse and specular reflection sequences. Either the latter should be included in the table, or both the captured OLATs and the reflection sequences should be defined outside the table. While this is ultimately a decision for the authors, consistency would be preferable.

When discussing the initial estimate of the surface normal, the authors mention that "...The derivation and synthesized results can be found in the supplementary material...." However, I could not locate the supplementary material, nor did I find any derivation in the appendix. Only the derivation of diffuse and specular albedo is included. Please clarify this point.

In section 4.1, provide a clearer explanation of the light arrangement. From Figure 1, it appears that the hexagonal lights for cross-polarization differ from those for parallel-polarization, with a 60-

doi.org/10.32388/WP1CH4

degree rotation between the two configurations (see the zoomed-in details in Figure 1). However, the sentence "On each light board, the white lights are arranged in a hexagonal pattern, with cross-polarizers and parallel polarizers placed alternately at the front" suggests that the lights are the same and the polarizers are alternately placed in front of the light board. Please clarify this to avoid confusion.

In section 4.1, it would be helpful to specify the placement of the 8 cameras.

The term OLAT is first introduced in Table 1. It would be better to explain it in the text of the "Data Format" section rather than in the image.

For stylistic reasons, I would recommend avoiding the use of references like [XX] as the subject or object in sentences.

Regarding the references, please check the hyphenation, as many of them are incorrect.

Declarations

Potential competing interests: No potential competing interests to declare.