Assessing the Quality of 3D Reconstruction in the Absence of Ground Truth: Application to a Multimodal Archaeological Dataset

Benjamin Coupry, Baptiste Brument, Antoine Laurent, Jean Mélou, Yvain Quéau, Jean-Denis Durou

Assessing the Quality of 3D Reconstruction in the Absence of Ground Truth: Application to a Multimodal Archaeological Dataset

Benjamin Coupry[1,*] Baptiste Brument[1] Antoine Laurent[1] Jean Mélou[1]
Yvain Quéau[2] Jean-Denis Durou[1]

[1]IRIT, UMR CNRS 5505, Toulouse, France
[2]Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, Caen, France
[*] benjamin.coupry@toulouse-inp.fr

(a) Biface (b) Fossil (c) Harpoon (d) Blade (e) Axe (f) Shard

Figure 1. Proposed corpus, consisting of six archaeological objects: (a) biface; (b) retouched flint chip with fossil; (c) harpoon with two rows of barbs; (d) dagger blade; (e) edged axe; (f) pottery shard decorated with incisions. All objects are approximately 10 cm large.

Abstract

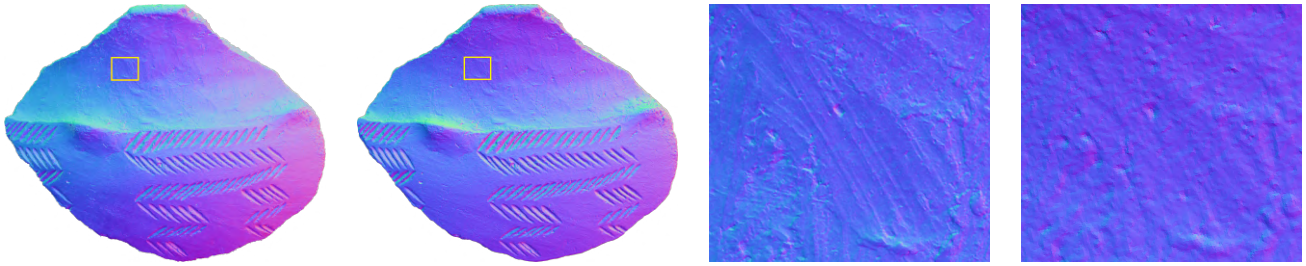
This paper proposes a new dataset of archaeological artefacts for evaluating 3D reconstruction methods, and questions the notion of "ground truth". Indeed, 3D reconstruction of archaeological objects can be carried out using either scanners or photographic methods. It turns out that multi-view stereo (MVS) faithfully reconstructs the overall shape of an object, on a par with a hand-held scanner, while calibrated photometric stereo (CPS) reveals relief details. The restitution of low- and high-frequencies is therefore the prerogative of distinct methods, which indicates that the ground truth, and the metric used for evaluation, should be chosen in view of the target frequencies. This observation led us to combine MVS and CPS, using MVS to calibrate the illumination used by CPS. We demonstrate on our dataset of archaeological objects that this original 3D reconstruction method indeed combines the advantages of MVS and CPS. Our proposed dataset can be accessed here: https://github.com/BenjaminCoupry/the-MAD-project.

1. Introduction

This paper deals with the 3D digitisation of archaeological heritage. Most archaeological objects are opaque and diffusive, which allows us to use Lambert's law to model image formation. But for these objects, there is generally no ground truth. So how can we prove that one 3D digitising method is more accurate than another?

The simplest way to answer this question is to simulate the images of a 3D scene model (relief and colour of objects, and lighting), provided we also know the camera parameters. Since various 3D vision techniques aim to derive the geometry of the 3D scene from such a set of simulated images, it is easy to assess the accuracy of the estimated relief, since the ground truth is indisputable. What is more, both reliefs are expressed in the same reference frame, since the (virtual) camera poses are perfectly known.

What is wrong with this approach, apart from its lack of realism? Our main complaint is that, in practice, it can only be used for photographic 3D digitising methods (photogrammetry). In fact, in our field of interest, that of archaeological heritage, 3D digitisation is very often carried



(a) CPS (b) Scanner (c) Zoom of (a) (d) Zoom of (b)

Figure 2. Two normal fields of a pottery shard (cf. Figure 1-f) estimated: (a) by calibrated photometric stereo (CPS); (b) using a hand-held scanner. The overall shape of the object is less well rendered by CPS, probably due to a lack of accuracy in lighting calibration, which has the effect of "bulging" the relief. On the other hand, surface details are less well rendered by the scanner, which tends to smooth out the relief, as shown by zooms (c) and (d) of the same area highlighted in white in (a) and (b): a fingerprint visible on (c) is much less so on (d).

out using scanners. While the term "scanner" is used to describe a whole range of instruments operating on various principles (structured light projection, time-of-flight measurement, etc.), they all share a common characteristic: they are "closed" commercial instruments, for which it is impossible to simulate realistic 3D datasets.

We therefore prefer a different approach. Having chosen a corpus of representative objects, we carry out a campaign of 3D digitisation using different modalities, then express the results in a single reference frame. While it is possible to compare different 3D reconstructions, there is generally no ground truth for real objects, except for those that have been machined from 3D models. Although such objects have appeared in several databases, the question remains open for archaeological objects. In the absence of ground truth, the 3D reconstruction obtained by a scanner is often considered as such, particularly in the field of archaeology. However, even though the accuracy claimed by scanner manufacturers seems to surpass that of photographic techniques, we challenge this presupposition.

Contributions In this paper, we show that the restitution of an object's overall shape and that of its relief details are generally the prerogatives of distinct methods (cf. Figure 2). This leads us to combine two photographic methods for 3D reconstruction, using the MVS result to calibrate the illumination required by CPS. We show, on a dataset of archaeological objects to be made publicly available, that this original 3D reconstruction method indeed combines the advantages of both MVS and CPS.

Section 2 provides an overview of related works. In Section 3, we present our archaeological corpus, the 3D scanning modalities and the metrics. The evaluation of several existing methods on this benchmark, conducted in Section 4, encourages us to propose an original combination of MVS and PS, before drawing our conclusions in Section 5.

2. State of the Art

3D Reconstruction *Geometric* photographic techniques for 3D reconstruction are the best known. The classic pipeline proposed by many softwares comprises two stages. From a collection of images of a 3D scene obtained from different angles, the structure-from-motion [35] technique estimates camera poses and a sparse 3D point cloud. The multi-view stereo [7] technique subsequently uses the estimated poses to perform a dense reconstruction of the scene by maximising photometric coherence between the different views.

Combined with neural volume rendering approaches [40], this second technique can effectively handle complex geometries and self-occlusions, even if the results remain disappointing in certain configurations, notably when the 3D scene is lightly textured [44]. In addition, despite recent efforts in this direction [23], the rendering of relief details remains limited. Finally, the estimation of scene reflectance is a bottleneck for all geometric techniques.

On the other hand, *photometric* photographic techniques for 3D reconstruction, which are less widely used, excel in recovering the high frequencies of the normal field, which means they reproduce relief details well. These techniques are based on the analysis of light quantities captured by the camera. Among these techniques, photometric stereo (PS) estimates geometry and reflectance of a surface from images taken from the same viewing angle, but under different lighting conditions [42]. This 3D reconstruction technique is the only one that can estimate scene reflectance. It involves inverting the image formation model, which links the 3D scene's relief, reflectance (reduced to the albedo, in the Lambertian case) and illumination. Calibrated PS (CPS) assumes known lighting. Under this assumption, with $p \geq 3$ directional lightings not all coplanar, it is possible to estimate the normals and albedo of the scene.

Calibrated PS methods can also be extended towards non-directional illumination sources such as LEDs [31], yet a dedicated (and potentially cumbersome) calibration procedure remains required. To avoid such a calibration procedure, it is possible to use an uncalibrated PS algorithm. It is well known that uncalibrated PS is an ill-posed problem [1, 11], for which a common solution is to introduce an a priori [8, 37, 43]. Another approach, closer to that proposed in Section 4, is to use the 3D scene as a calibration chart: note that in [12], it is assumed that a certain number of points in the scene have the same albedo, an assumption that may be difficult to guarantee.

Neural approaches to uncalibrated PS generally assume that the lighting is directional. A first network determines the lighting [6], which makes the problem calibrated, but this approach fails if the directional lighting assumption is false. The *universal* PS approach, which has been proposed very recently to limit this risk [13, 14], accommodates a wide variety of lighting conditions, learned by deep learning. Today, it is without doubt the most effective approach to solving uncalibrated PS.

Merging Geometric and Photometric Approaches Geometric and photometric 3D reconstruction methods have complementary strengths. The fusion of these two approaches has a long history: the refinement of multi-view geometry through shading has been the subject of numerous works using classic methods [18, 26].

The addition of a second point of view for PS, proposed by Kong et al. [21], is still an active area of research [24]. However, the most promising approach seems to be multi-view photometric stereo (MVPS). This approach was initially presented by Hernandez et al. [12]. Later came uncalibrated MVPS [30] or surface representation using a signed distance function [25].

The advent of neural networks has opened the door to many other approaches. In particular, many works build on the multi-view advances offered by neural approaches [29, 40] and have adapted them to the MVPS case [3, 17, 45].

Ground Truth Estimation Evaluating 3D reconstruction methods requires the use of data accompanied by ground truth. Added to this is the impressive rise of deep learning methods, which require an immense amount of annotated data. The implementation of synthetic datasets allows total control over the characteristics of the 3D scene. Many datasets, for example, have followed the lead of ShapeNet [5], which offers a staggering number of 3D objects. Physical-based rendering tools then made it possible to multiply the characteristics of dataset objects and lighting [10, 13, 14]. It should be noted, however, that some databases present real objects digitised with the aim of extracting more realistic synthetic data [32].

Evaluating 3D reconstruction methods also requires testing on real data. A common method of obtaining ground truth involves the use of a scanner [16]. The classic PS method evaluation dataset DiLiGenT [38] uses this approach, as do its successors, extending this approach to multi-view [22], point sources [27], or flat surfaces [39]. However, the use of the scanner remains open to criticism, since its resolution may remain limited, and 3D-2D registration towards the images induces unavoidable errors. Other works prefer to use CAD models [33] or well-known geometry [9], yet the actual objects slightly differ from the models due to (unavoidable) manufacturing inaccuracies.

Despite their undeniable merits, these datasets accompanied with "ground truth" may not be sufficient for objectively assessing 3D-reconstruction methods. In particular, they intrinsically favour either the high- or the low-frequency part of the 3D reconstruction. In the next section, we present a new evaluation dataset dedicated to archaeology, that comes along with three different "ground truths" and two evaluation metrics, enabling fairer comparisons.

3. Proposed Archaeological Benchmark

3.1. Corpus Description

The corpus contains objects representative of archaeological artefacts, both in terms of the diversity of materials and forms and the recurrence of objects found in excavations. They date from the Middle Paleolithic to the Protohistory, and are remarkable both for their state of preservation and for the great finesse of their decoration (see Figure 1).

The first object is a Mousterian flint biface. It has two major advantages for our study. Its surface retains a cortical part in its centre (the outside of a flint chip) whose characteristics differ significantly from those of the contours. Secondly, the fine removals resulting from debitage reveal a faceted relief with sharp edges. The second object is also a Middle Paleolithic flint. However, it has a fossil on its surface. It was deliberately kept by the cutter to decorate the tool. From our point of view, it is a complex object to reconstruct, due to the fluting of the shell. In addition, this flint has a relatively uniform colour. The third object is a reindeer antler harpoon with two rows of barbs. Dating from the Upper Palaeolithic, its fine surface incisions and barbs are clearly difficult to reconstruct. The fourth object is a Neolithic flint dagger blade. In addition to its uniform hue, the numerous removals intended to revive the active part of the tool are important for tracing the chronology of the technical gestures performed. The fifth object is a protohistoric bronze axe with a weathered surface characteristic of metallic objects. The sixth and final object is a Bronze Age ceramic shard. This fragment comes from unturned pottery. Shaping elements are visible on its surface. Its incised decoration may make 3D reconstruction difficult.

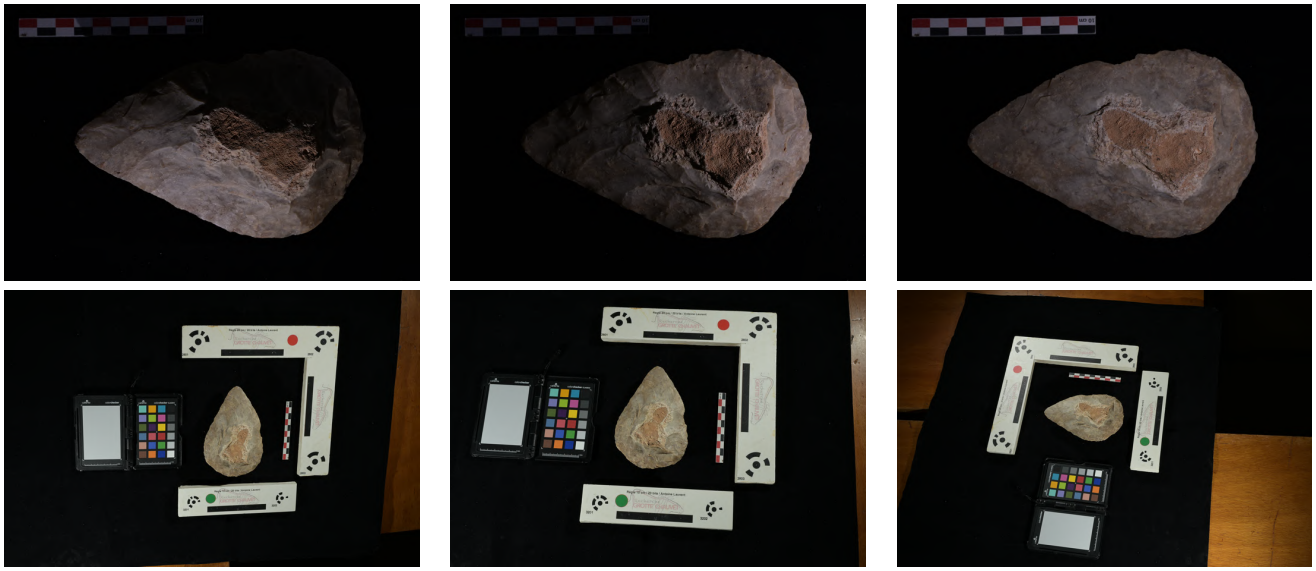


Figure 3. Images of the biface object. First row : 3 PS images (out of 105) captured from the same viewing angle under varying illumination, using an RTI dome. Second row : 3 multi-view images (out of 80), captured from multiple viewing angles under fixed illumination.

3.2. 3D Digitising Modalities

In our study, we used a multi-view photographic acquisition, an RTI (Reflectance Transformation Imaging) dome and a pulsed-light hand scanner. Only the main face of each object in the corpus was digitised. Figure 3 presents examples of captured multi-view and multi-illumination images, for the biface object.

Multi-view acquisition was carried out with a Nikon Z7II camera equipped with 50 mm lens, for a full-format resolution of 8256×5504 pixels. The scene consists of the object in the central position, plus a ruler and a square positioned around it. These two elements, which include targets automatically recognised by the photogrammetry software, are used to characterise the scale and orientation of the object. Photographs are taken under fixed ambient lighting from a flash at a distance of 75 cm, then at a distance of 30 cm. Each series comprises between 45 and 80 photographs. Photogrammetric processing was carried out using Agisoft Metashape 1.8.5 software. After alignment, scaling and calibration of the camera, the mesh is generated in "ultra-high" quality (resolution equal to one pixel) from the depth maps.

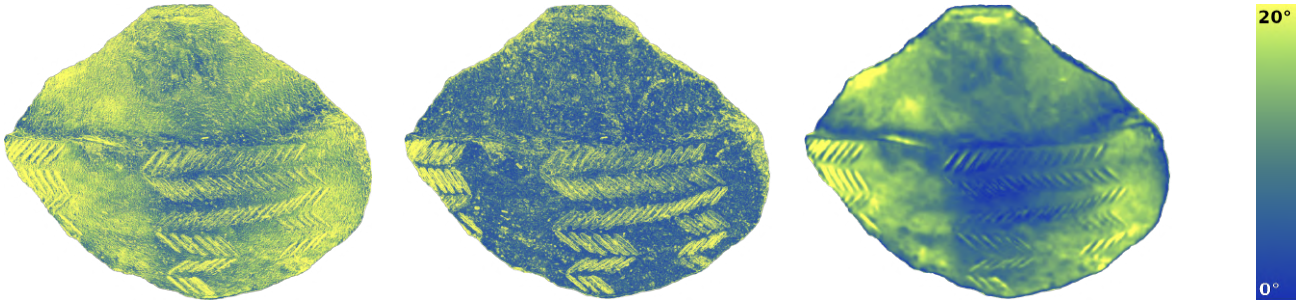
Multi-illumination acquisition is carried out using the same camera. The RTI dome is the Mercurio Imaging CEOS model, 50 cm in diameter and equipped with 105 LEDs. The sequence of images is automatic. The distance between the object and the camera is around 30 cm. When all the LEDs are lit, the photogrammetric software recalibrates one shot with the multi-view images, thus expressing both acquisitions in the same metric reference frame.

Finally, the Artec Space Spider scanner, widely used in archaeology, features manufacturer-guaranteed 50 microns accuracy and 100 microns mesh resolution. All processing is carried out with the dedicated Artec Studio v18 software. During mesh generation, the object closure option is activated, to fill in any information gaps during capture.

The meshes obtained from the scanner were all realigned with the meshes obtained from MVS. The alignment procedure consists of an initialisation by matching 3D FPFH descriptors on a decimated mesh [34], followed by a robust matching of the calculated descriptors, and finally by a robust ICP. Once matching is achieved, the normal map can be computed using standard tools, such as the Blender software. This way, the normal maps can all be defined in the same camera system – that of the multi-illumination setup.

3.3. Frequency-aware Metrics

Geometric 3D reconstruction methods produce reliable results for the global geometry of the object. The reconstructed surfaces are close to those acquired by hand-held scanner, which are generally used as ground truth. However, Figure 2 shows that calibrated PS (CPS) produces much finer results than the scanner, but the overall shape of the object is less well rendered by CPS. This is due to the lack of accuracy in lighting estimation, which has the effect of "bulging" the relief. It therefore seems that the information obtained by the hand-held scanner and by CPS are complementary. The low frequencies of a 3D reconstruction should thus be evaluated against those of the scanner, and its high frequencies against those of CPS.



(a) (b) (c) (d)

Figure 4. (a) Angular error (AE, in degrees) maps between scanner and CPS results on the shard object. (b) High frequencies AE, showing scanner errors. (c) Low frequencies AE, showing CPS errors. (d) Error scale in false colours. The same scale will be used throughout the results.

The normal maps obtained by CPS could be integrated, so that 3D reconstructions could be assessed in terms of depth maps. However, this step can only be carried out up to a scale factor, and potential discontinuities would bias the comparison [4]. Therefore, we rather compare the 3D reconstructions in terms of normals (see Figure 2), which are also used in archaeology [20]. In this way, PS and MVS methods can be compared faithfully, without introducing integration bias.

In order to assess low frequencies accuracy of a 3D reconstruction, thin details must be removed. To this end, the normals are smoothed out using a normalised Gaussian filter defined as:

$$F_\sigma(\mathbf{n}) = \frac{\mathbf{n} * G_\sigma}{\|\mathbf{n} * G_\sigma\|} \tag{1}$$

with $*$ the 2D convolution operator and σ the standard deviation (which is equal to 20 pixels in our experiments). The low-frequency error E_{LF} in a point $x \in \mathbb{R}^2$ is then defined as the angular error between the reference normal map $\mathbf{n}_{\text{LF}}$ (obtained with the scanner) and that to evaluate $\mathbf{n}$, after applying the filter:

$$E_{\text{LF}}(x) = \arccos\left\{F_\sigma(\mathbf{n}_{\text{LF}})(x) \cdot F_\sigma(\mathbf{n})(x)\right\} \tag{2}$$

The high-frequency error, on the other hand, must be free from low frequencies bias i.e., normals should be compared after registration towards a common local tangent plane. Low-frequency transformations preserving the appearance of Lambertian scenes are long known and studied, the bas-relief ambiguity being the most famous example [2]. In a general case, the sought transformation is a pixelwise rotation of normals. Therefore, we register the 3D reconstructed normal $\mathbf{n}$ with the high-frequency reference $\mathbf{n}_{\text{HF}}$ (that of CPS) by locally searching for the rotation $\mathbf{R}$ that minimises the low-frequency error:

$$\mathbf{R}(x) = \underset{\mathbf{R}\in \text{SO}(3)}{\text{argmin}} \sum_{t\in\nu(x)} \|\mathbf{R}\, F_\sigma(\mathbf{n}_{\text{HF}})(t) - F_\sigma(\mathbf{n})(t)\|^2 \tag{3}$$

with $\nu(x)$ a neighbourhood of x large enough to avoid degenerated configurations (in our experiments, $\nu(x)$ represents the 25 closest neighbours of x in the sense of the Manhattan distance). This optimisation problem can be solved using, e.g., SVD [36]. Then, the high-frequency error E_{HF} in x is defined as the angular error between reference normal and the 3D reconstructed one, after applying the registration:

$$E_{\text{HF}}(x) = \arccos\left\{\mathbf{R}(x)\, \mathbf{n}_{\text{HF}}(x) \cdot \mathbf{n}(x)\right\} \tag{4}$$

Figure 4 illustrates these metrics on the shard object. Next, we turn our attention to evaluating several 3D reconstruction methods, with respect to these high-frequency and low-frequency metrics.

4. 3D Reconstruction Assessment

In this section, we first evaluate state-of-the-art photometric and geometric techniques on the proposed archeological benchmark, before proposing an original approach combining the merits of both approaches.

4.1. Photometric vs Geometric Techniques

Our benchmark compares both established and recent methods in the literature. Given that the appearance of the objects in our dataset is close to Lambertian, we consider as photometric reference (CPS) the robust calibrated method based on sparse regression from [15] (calibration is deduced from the acquisition dome CAD model). To emphasise the importance of calibration, we also provide the results of two state-of-the-art uncalibrated PS techniques based on deep learning, namely SDM-UniPS [14] and UniMS-PS [10]. Regarding geometric techniques, the hand-held scanner serves as reference, and we provide the results attained by the recent NeuS2 [41] framework, as well as that of the Metashape software.

For quantitative evaluation, the scanner result serves as reference for evaluating low frequencies, and the CPS result for high frequencies. Evaluation is carried out in terms of angular error (AE) on normal maps. All normal maps are stored as 16bits floating point tensors, and displayed using the standard RGB convention (R for the horizontal direction, G for the vertical one, and B for the camera axis) for qualitative inspection. As some methods provide results in the form of meshes, these have been realigned as described in Section 3.2, before extracting the normal map associated with the PS view.

The qualitative results for high frequencies are shown in Figure 5. As expected, the photometric 3D reconstruction methods are much better than their geometric counterparts. This is confirmed by the quantitative results for all the objects shown in Table 1. Figure 6, on the other hand, depicts the qualitative results associated with low frequencies. As expected, geometric methods perform better in this case, which is confirmed quantitatively by Table 2. Although uncalibrated PS techniques partially reduce the bias inherent to lighting calibration, the photometric results remain far from the geometric ones.

	SDM	UniMSPS	MVS	NeuS2	Scanner	Ours
Biface	5.5	5.5	7.1	6.9	8.3	3.1
Fossil	4.5	8.3	6.8	7.1	8.7	4.7
Harpoon	7.2	7.8	8.9	10.1	12.1	5.9
Blade	2.5	4.1	4.6	5.3	6.3	2.4
Axe	5.6	8.3	9.2	11.1	11.4	4.1
Shard	4.8	5.9	8.4	9.6	9.5	4.4

Table 1. High-frequency AE (in degrees) with respect to CPS. Photometric methods (SDM and UniMSPS) largely outperform geometric ones (MVS, NeuS2 and Scanner), which all miss the thinnest geometric details. However, the proposed combination of photometric and geometric 3D-reconstruction (see Section 4.2) further improves the results.

	SDM	UniMSPS	CPS	MVS	NeuS2	Ours
Biface	11.5	5.8	10.8	1.5	2.2	5.0
Fossil	5.5	4.4	7.2	1.1	3.4	3.0
Harpoon	10.8	7.8	12.2	3.3	3.4	5.3
Blade	6.3	6.0	8.2	1.0	1.1	2.8
Axe	5.0	6.2	7.8	1.1	1.3	2.4
Shard	9.1	5.9	10.8	1.0	7.1	3.6

Table 2. Low-frequency AE (in degrees) with respect to Scanner. Our proposed method (see Section 4.2) produces results with low-frequency AE in the range of the geometric methods (MVS and NeuS2), and well below those of uncalibrated (SDM and UniMSPS) and calibrated (CPS) photometric stereo methods.

It should also be emphasised that both families of approaches are not equally simple to use from the end-user perspective. Indeed, while the use of the scanner or geometric reconstruction methods is immediate, the use of calibrated PS requires knowledge of the lighting in the scene, which requires a dedicated, and possibly cumbersome calibration stage. Besides, lighting calibration outside the laboratory requires positioning a calibration chart in the scene at the time of shooting, which is most often a mirror ball or a matte white sphere. However, unless the incident luminous flux is perfectly parallel and uniform, which is difficult to guarantee in practice, this calibration method is subject to several limitations. Firstly, the extrapolation to the entire scene of the illumination estimated at the sphere's location generally results in a low-frequency bias in the normal field estimated by CPS. Such a bias can be limited using non-directional calibrated PS methods [31], yet sphere-based calibration remains tedious. The use of a calibration sphere can also prove problematic for other reasons, since it must be placed as close as possible to the scene, without occluding it or causing shadows. Moreover, in certain contexts such as archaeology, it is inconceivable that the sphere could come into contact with the object or scene to be digitised. For all these reasons, lighting calibration should, if possible, be carried out differently.

Overall, the ideal method would exhibit the high-frequency accuracy of photometric methods, yet be free from low-frequency bias and simple to use. In the next section, we present a practical approach for calibrating the lighting of a scene on the fly, in order to apply PS without any calibration chart, based on a 3D reconstruction reliable in the low frequencies.

4.2. Towards a New PS Method

We have at hand a low-frequency 3D reconstruction, represented by a set of normals $\mathbf{n}(x)$. We then aim at inverting the image formation model in order to retrieve the illumination, and subsequently use this estimate into a calibrated PS algorithm for re-estimating the normals.

Unfortunately, knowledge of the normals is not sufficient for estimating the illumination in each image, because the reflectance values $\rho(x)$ are also unknown. While the reflectance of a calibration sphere of uniform colour is known, admittedly to within a factor, the same cannot be said for any 3D scene. For this reason, in previous works, points with similar reflectance must be identified [8, 12]. Instead, we jointly estimate reflectance and illumination, as also proposed for instance in [28]. This joint optimisation problem becomes tractable under the Lambertian assumption, because the reflectance becomes independent from the incident illumination. Hence, in a PS setup with $p \geq 3$ images and n pixels, we are given np observations to estimate n albedo values and p illumination vectors.

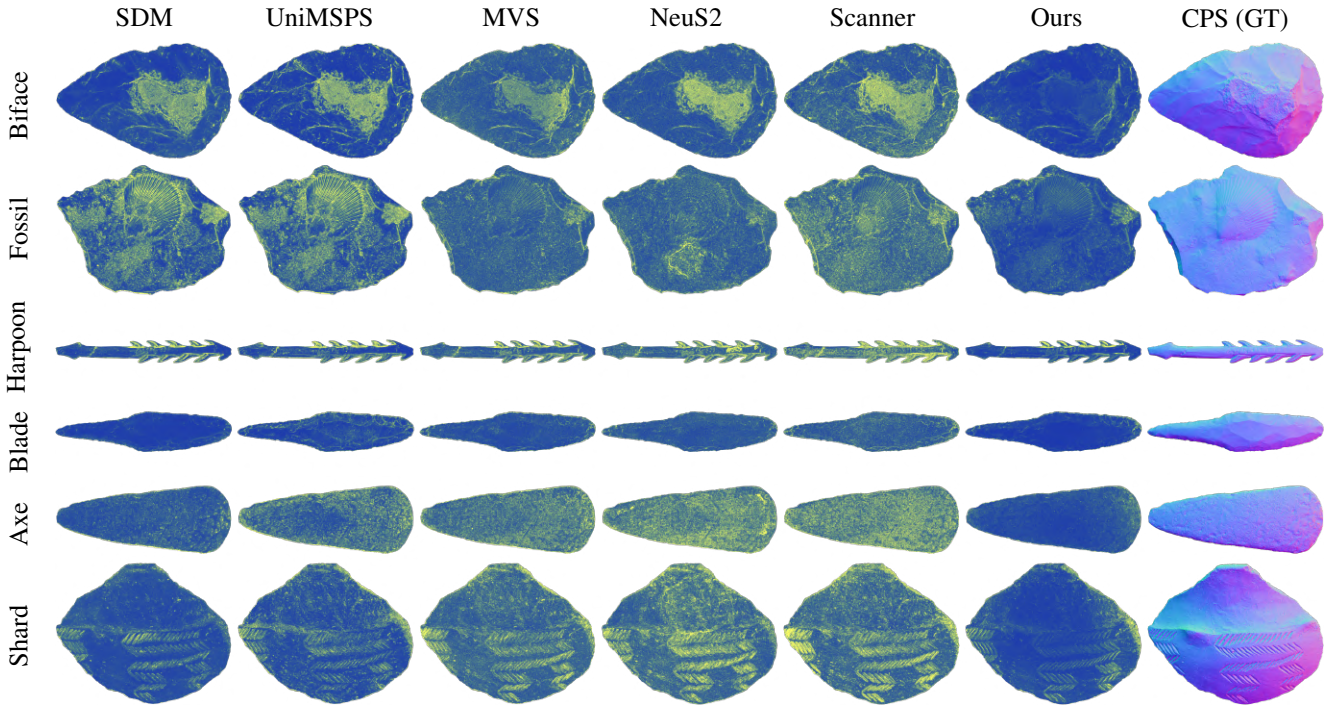


Figure 5. High-frequency AE maps, using CPS (displayed in standard RGB representation) as reference ("high-frequency ground truth"). Fine-scale details are missed by geometric methods (MVS, NeuS2 and Scanner). Moreover, the new PS method that we describe in Section 4.2 (Ours) clearly outperforms the uncalibrated deep learning-based PS methods (SDM and UniMSPS).

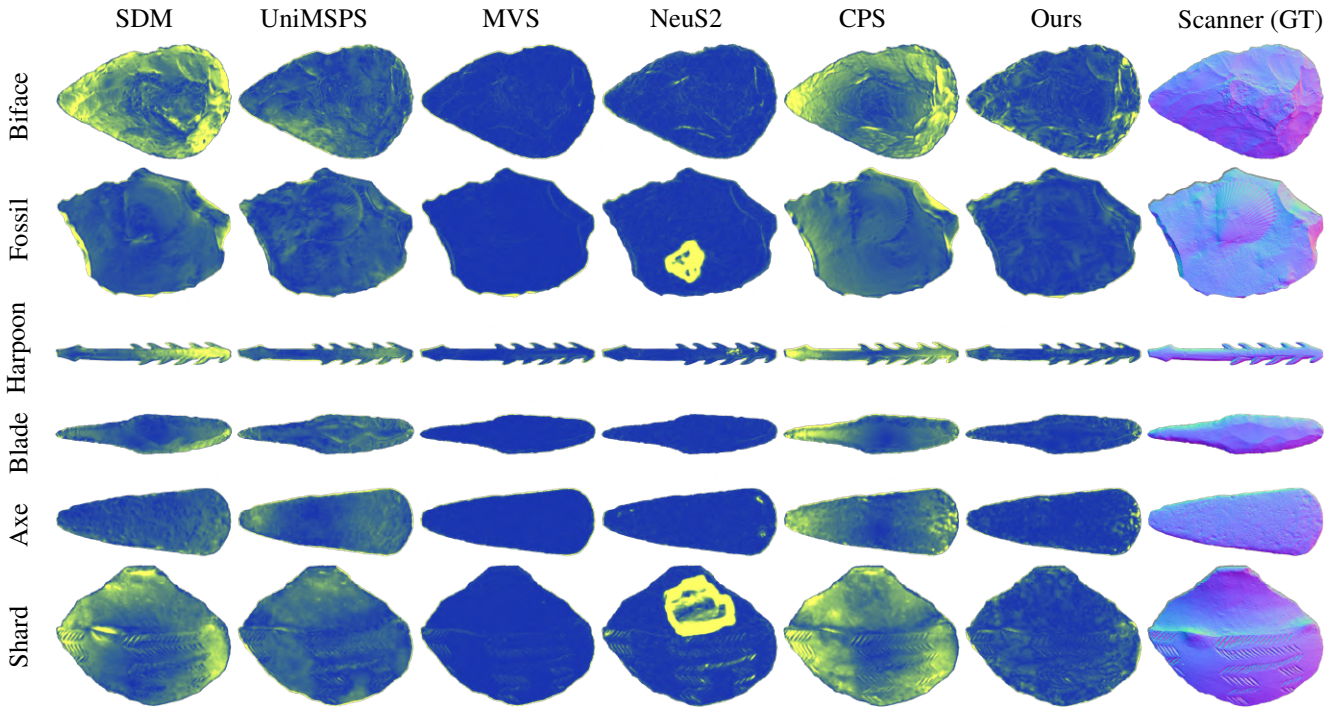


Figure 6. Low-frequency AE maps, using Scanner (displayed in standard RGB representation) as reference ("low-frequency ground truth"). Photometric methods (SDM, UniMSPS and CPS) exhibit a low-frequency bias, which is attenuated a lot with the proposed method (Ours), while preserving fine details. Neus2 exhibits some localised irregularities skewing the MAE.

Assuming directional illumination is characterised by vectors $\mathbf{s}^i \in \mathbb{R}^3$, $i \in \{1, \dots, p\}$, the Lambertian model is written :

$$I^i(x) = \rho(x)\ \max\{0, \mathbf{n}(x) \cdot \mathbf{s}^i\} \tag{5}$$

where $I^i(x)$ is the grey level at point x, $\rho(x)$ the albedo of the conjugate surface point and $\mathbf{n}(x)$ its unit normal.

Denoting $\bar{I}^i$ the photograph of the object under the i-th illumination, and I^i its reprojection using the Lambertian model (5), this joint task could be formulated as the minimisation of the following energy:

$$\mathcal{L}(\mathbf{s}^i, \rho) = \mathbb{E}_x \left[\sum_{i=1}^{p} \mathcal{H}\left(I^i(x) - \bar{I}^i(x)\right) \right] \tag{6}$$

with $\mathcal{H}$ the Huber loss, and where $\mathbb{E}_x$ represents the expectation over a random batch of pixels (to limit the computational burden).

However, in the majority of in situ acquisitions, the assumption of directional illumination is unrealistic, because of inherent material constraints and of the distance between the light source and the scene. Actually, this approximation is partly responsible for the majority of low-frequency AE in PS methods – although recent so-called universal techniques [10, 14] considerably attenuated this bias.

To reduce these low-frequency AE, we could consider spatially-varying illumination vectors, both in direction and intensity, and denote them by $\bar{\mathbf{s}}^i(x)$. Unfortunately, evaluating $\bar{\mathbf{s}}^i(x)$ independently for each pixel yields an undetermined problem: an a priori must be introduced to make things tractable. It is reasonable to assume that illumination in a scene has bounded variations. Instead of resorting to an ad hoc regularisation term, we propose to consider $\bar{\mathbf{s}}^i(x)$ as a linear interpolant on a grid, where its parameters $\theta^i \in \mathbb{R}^{3q}$ represent a set of q illumination vectors at q control points, as illustrated in Figure 7.

In practice, the parameters of the interpolation grids are also optimized, to better catch the lighting variations. The loss (6) is thus optimized in terms of all lighting parameters and albedo values. This is achieved using stochastic gradient descent with momentum, namely the Adam algorithm [19]. The median value of each pixel is taken as initialisation for the albedo, a directional least squares estimate for lighting, and a regular grid for the control points.

Once the spatially-varying illumination has been estimated from the low-frequency normals, we can plug it into a calibrated PS algorithm to re-estimate the normals, for instance again in the sense of the Huber loss.

Table 1 shows that these normals are comparable, in terms of high frequencies, with state-of-the-art uncalibrated PS algorithms – although our approach did not resort to any kind of learning. More interestingly, it widely outperforms them in terms of low frequencies (see Table 2). This tends to indicate that despite the remarkable efficiency of modern uncalibrated deep learning-based PS methods, calibrated methods still remain superior, provided that bias is limited in the calibration procedure.

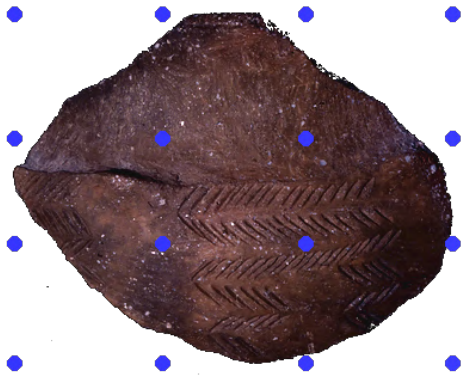


Figure 7. Interpolation grid for local lighting estimation, superimposed on one of the shard images. Each blue point represents one of the q control points.

5. Conclusion and Perspectives

In this paper, we questioned the notion of "ground truth" for evaluating 3D reconstruction methods, with a particular focus on archaeology. We proposed a novel corpus which comes along with several reference 3D reconstructions, and discussed their respective abilities to capture the low and high frequencies.

This led us to propose a new 3D reconstruction method, which relies on geometric techniques for estimating spatially-varying illumination that can be provided to calibrated PS. This method, although dedicated to Lambertian (or quasi-Lambertian) materials that are commonly encountered in archeology, achieves state-of-the-art results while remaining conceptually simple. In the future, we plan to extend this method to cope with parametric non-directional illumination models, to cope for instance with near point light sources such as LEDs.

Extending it towards more complex materials would represent a natural extension of our work, which could be achieved for instance by generalising our framework towards recent PS models based on learning. Finally, making our acquisition corpus publicly available is of interdisciplinary interest, linking the 3D vision community and archaeologists. In the future, it would be useful to continue enriching this dataset with new objects featuring complex geometries and varied materials.

References

[1] Peter N Belhumeur and David J Kriegman. What is the set of images of an object under all possible illumination conditions? *IJCV*, 28:245–260, 1998. 3

[2] Peter N Belhumeur, David J Kriegman, and Alan L Yuille. The bas-relief ambiguity. *IJCV*, 35(1):33–44, 1999. 5

[3] Baptiste Brument, Robin Bruneau, Yvain Quéau, Jean Mélou, François Lauze, Jean-Denis Durou, and Lilian Calvet. RNb-NeuS: Reflectance and Normal-Based Multi-View 3D Reconstruction. In *CVPR*, 2024. 3

[4] Xu Cao, Hiroaki Santo, Boxin Shi, Fumio Okura, and Yasuyuki Matsushita. Bilateral normal integration. In *ECCV*, 2022. 5

[5] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 3

[6] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong. Self-calibrating deep photometric stereo networks. In *CVPR*, 2019. 3

[7] Yasutaka Furukawa, Carlos Hernández, et al. Multi-view stereo: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 9(1-2):1–148, 2015. 2

[8] Heng Guo, Zhipeng Mo, Boxin Shi, Feng Lu, Sai-Kit Yeung, Ping Tan, and Yasuyuki Matsushita. Patch-based uncalibrated photometric stereo under natural illumination. *PAMI*, 44(11):7809–7823, 2021. 3, 6

[9] Heng Guo, Jieji Ren, Feishi Wang, Boxin Shi, Mingjun Ren, and Yasuyuki Matsushita. DiLiGenRT: A Photometric Stereo Dataset with Quantified Roughness and Translucency. In *CVPR*, 2024. 3

[10] Clément Hardy, Yvain Quéau, and David Tschumperlé. Uni MS-PS: a Multi-Scale Encoder Decoder Transformer for Universal Photometric Stereo. *CVIU*, 248, 2024. 3, 5, 8

[11] Hideki Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 11(11):3079–3089, 1994. 3

[12] Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. *PAMI*, 30(3):548–554, 2008. 3, 6

[13] Satoshi Ikehata. Universal photometric stereo network using global lighting contexts. In *CVPR*, 2022. 3

[14] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. In *CVPR*, 2023. 3, 5, 8

[15] Satoshi Ikehata, David Wipf, Yasuyuki Matsushita, and Kiyoharu Aizawa. Robust photometric stereo using sparse regression. In *CVPR*, pages 318–325, 2012. 5

[16] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *CVPR*, 2014. 3

[17] Berk Kaya, Suryansh Kumar, Carlos Oliveira, Vittorio Ferrari, and Luc Van Gool. Multi-View Photometric Stereo Revisited. In *WACV*, 2023. 3

[18] Kichang Kim, Akihiko Torii, and Masatoshi Okutomi. Multi-view inverse rendering under arbitrary illumination and albedo. In *ECCV*, 2016. 3

[19] Diederik Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 2015. 8

[20] Žiga Kokalj and Maja Somrak. Why not a single image? Combining visualizations to facilitate fieldwork and on-screen mapping. *Remote Sensing*, 11(7), 2019. 5

[21] Hui Kong, Pengfei Xu, and Eam Khwang Teoh. Binocular uncalibrated photometric stereo. In *Advances in Visual Computing: Second International Symposium*, 2006. 3

[22] Min Li, Zhenglong Zhou, Zhe Wu, Boxin Shi, Changyu Diao, and Ping Tan. Multi-view photometric stereo: A robust solution and benchmark dataset for spatially varying isotropic materials. *TIP*, 29:4159–4173, 2020. 3

[23] Zhaoshuo Li, Thomas Müller, Alex Evans, Russell H Taylor, Mathias Unberath, Ming-Yu Liu, and Chen-Hsuan Lin. Neuralangelo: High-Fidelity Neural Surface Reconstruction. In *CVPR*, 2023. 2

[24] Fotios Logothetis, Ignas Budvytis, and Roberto Cipolla. A neural height-map approach for the binocular photometric stereo problem. In *WACV*, 2024. 3

[25] Fotios Logothetis, Roberto Mecca, and Roberto Cipolla. A differential volumetric approach to multi-view photometric stereo. In *ICCV*, 2019. 3

[26] Daniel Maurer, Yong Chul Ju, Michael Breuß, and Andrés Bruhn. Combining shape from shading and stereo: A joint variational method for estimating depth, illumination and albedo. *IJCV*, 126(12):1342–1366, 2018. 3

[27] Roberto Mecca, Fotios Logothetis, Ignas Budvytis, and Roberto Cipolla. Luces: A dataset for near-field point light source photometric stereo. *arXiv preprint arXiv:2104.13135*, 2021. 3

[28] Jean Mélou, Yvain Quéau, Jean-Denis Durou, Fabien Castan, and Daniel Cremers. Variational re-

flectance estimation from multi-view images. *JMIV*, 60(9):1527–1546, 2018. 6

[29] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 3

[30] Jaesik Park, Sudipta N Sinha, Yasuyuki Matsushita, Yu-Wing Tai, and In So Kweon. Robust multiview photometric stereo using planar mesh parameterization. *PAMI*, 39(8):1591–1604, 2016. 3

[31] Yvain Quéau, Bastien Durix, Tao Wu, Daniel Cremers, François Lauze, and Jean-Denis Durou. Led-based photometric stereo: Modeling, calibration and numerical solution. *JMIV*, 60:313–340, 2018. 3, 6

[32] Jeremy Reizenstein, Roman Shapovalov, Philipp Henzler, Luca Sbordone, Patrick Labatut, and David Novotny. Common Objects in 3D: Large-Scale Learning and Evaluation of Real-life 3D Category Reconstruction. In *ICCV*, 2021. 3

[33] Jieji Ren, Feishi Wang, Jiahao Zhang, Qian Zheng, Mingjun Ren, and Boxin Shi. Diligent102: A photometric stereo benchmark dataset with controlled shape and material variation. In *CVPR*, 2022. 3

[34] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (FPFH) for 3D registration. In *ICRA*, 2009. 4

[35] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-Motion Revisited. In *CVPR*, 2016. 2

[36] Peter H Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966. 5

[37] Boxin Shi, Yasuyuki Matsushita, Yichen Wei, Chao Xu, and Ping Tan. Self-calibrating photometric stereo. In *CVPR*, 2010. 3

[38] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *CVPR*, 2016. 3

[39] Feishi Wang, Jieji Ren, Heng Guo, Mingjun Ren, and Boxin Shi. DiLiGenT-Pi: A Photometric Stereo Benchmark Dataset with Controlled Shape and Material Variation. In *ICCV*, 2023. 3

[40] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. In *NeurIPS*, 2021. 2, 3

[41] Yiming Wang, Qin Han, Marc Habermann, Kostas Daniilidis, Christian Theobalt, and Lingjie Liu. NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction. In *ICCV*, 2023. 5

[42] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. 2

[43] Zhe Wu and Ping Tan. Calibrating photometric stereo by holistic reflectance symmetry analysis. In *CVPR*, 2013. 3

[44] Qingshan Xu, Weihang Kong, Wenbing Tao, and Marc Pollefeys. Multi-scale geometric consistency guided and planar prior assisted multi-view stereo. *PAMI*, 45(4):4945–4963, 2022. 2

[45] Wenqi Yang, Guanying Chen, Chaofeng Chen, Zhenfang Chen, and Kwan-Yee K Wong. Ps-nerf: Neural inverse rendering for multi-view photometric stereo. In *ECCV*, 2022. 3