Combining Geometric and Photometric 3D Reconstruction Techniques for Cultural Heritage

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

Combining Geometric and Photometric 3D Reconstruction Techniques for Cultural Heritage

Antoine Laurent[a,b,*], Benjamin Coupry[a], Baptiste Brument[a], Jean Mélou[a], Yvain Quéau[c], Carole Fritz[d], Jean-Denis Durou[a]

[a] *IRIT, UMR 5505, ENSEEIHT, 2 rue Camichel, Toulouse, 31071, France*
[b] *TRACES, UMR 5608, UT2J, 5 allée Antonio Machado, Toulouse, 31058, France*
[c] *GREYC, Normandie Univ., UNICAEN, ENSICAEN, CNRS, 6 boulevard du Maréchal Juin, Caen, 14032, France*
[d] *LAMS, UMR 8220, Sorbonne Université, 4 place Jussieu, Paris, 75005, France*

Abstract

There are mainly two families of photographic 3D reconstruction techniques. Photogrammetry techniques work according to the principle of triangulation, from the matching of different views, while photometric techniques link the appearance of a 3D point to the orientation of its normal, relative to that of the incident light. While photogrammetry allows to find the global shape of a 3D scene, if it is sufficiently textured, photometric techniques highlight the details of the relief, as long as the model linking the lighting to the shape and reflectance of the scene is sufficiently realistic. In this work, we compare these different approaches with some others in the context of reconstructing archaeological features. After discussing their advantages and disadvantages, we describe a promising new method combining both families in a multi-view, multi-lighting context.

Keywords: Photometric Stereo, Photogrammetry, 3D reconstruction, Cultural Heritage.

1. Introduction and Research Aim

The creation of 3D digital doubles of cultural heritage meets three needs in the heritage chain: conservation, research and mediation. The development of 3D reconstruction techniques and their deployment within research teams

*antoine.laurent@univ-tlse2.fr

has accelerated sharply since the turn of the millennium, with the increase in computer computing capacity. Most of these methods were already known, but their use has since become much more widespread.

Numerous technical solutions exist for 3D reconstruction, which can be adapted to the variety of heritage. Without going into an exhaustive list, we propose in this article to explore the non-contact solutions commonly used in archaeology, and in particular those that bring out the fine relief.

1.1. Related Work

The 3D scanning solutions can be divided into two categories: dedicated scanners and photographic techniques – which can themselves be separated between geometric and photometric techniques.

Scanners. Lasergrammetry is used to acquire the geometry of the environment with a precision of around 1 to 3 mm, for a scene located ten meters away, depending on the choice of TLS model (terrestrial laser scanning). These solutions are costly and generally produce poor quality textures, despite the advent of HDR (high dynamic range) cameras. However, lasergrammetry is still widely used in caves [14, 25]. Due to their low relative resolution and precision (above one centimeter at the time of writing), dynamic scanners such as SLAM (simultaneous localization and mapping) or ALS (airborne laser scanner) will not be discussed here.

Handheld scanner technologies are effective for small surfaces and objects. They allow one to obtain 3D models directly with the metric. The announced spatial resolution is around 100 µm, with an accuracy of a few tens of microns depending on the model [39], according to the data supplied by the manufacturer. They operate either by triangulation or structured light (white, blue LED, with infrared). RGB sensors record RGB information to colorize the 3D mesh [13]. Furthermore, this technology is adaptable to complicated heritage materials [23].

Photographic techniques. Let us turn now to photographic 3D scanning techniques, of which we will present just a few examples in the next paragraphs.

Reflectance transformation imaging (RTI) reproduces the surface appearance of the object of interest. It involves interpolating images taken from the same viewpoint and under different lighting conditions [37, 38]. The result is real-time dynamic rendering, effective for all types of objects. RTI cannot be considered a photogrammetric technique, but it does enable fine relief details

to be highlighted, and therefore remains very popular among archaeologists, particularly with regard to the study of cave art [18].

Among all the photographic techniques, *geometric* ones are the best known [35], referred to as *photogrammetry*. The classic pipeline proposed by many software packages [1, 4, 24] comprises two stages. From a collection of images of the 3D scene to be reconstructed, obtained from different viewpoints, the *structure-from-motion* (SfM) technique estimates the camera poses and a sparse cloud of 3D points [27]. The *multi-view stereo* (MVS) technique uses the estimated poses to perform a dense reconstruction of the scene by triangulation, maximizing photometric consistency between the views [22].

Due to its low cost and ease of implementation, this approach has achieved great popularity within the archaeological community. Combined with neural approaches to volume rendering [47], MVS can effectively handle complex geometries and self-occlusions. However, the results remain imprecise in certain configurations, particularly when the 3D scene is lightly textured [49]. In addition, the rendering of relief details remains limited. Finally, the estimation of scene reflectance is a bottleneck for all geometric techniques.

Photometric stereo. On the other hand, *photometric* techniques, which are less widely used, excel in recovering the high frequencies, which means that they reproduce relief details well. They are based on the analysis of light quantities registered by the camera's photosensitive receiver. Under certain conditions, the grey level of a digital image is proportional to the luminance emitted by the scene, which is characteristic, at least in part, of its relief [19].

Among them, photometric stereo (PS) estimates the geometry and reflectance of a surface using data similar to those of RTI: photographs taken from the same point of view, but under different lighting conditions [48]. 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 lighting. *Calibrated PS* assumes known lighting conditions. Under this assumption, with $n \geq 3$ non-coplanar lighting vectors, the albedo and normal of each pixel can be deduced from the photographs [48].

Several works have extended calibrated PS to non-Lambertian materials [45]. The assumption of directional illumination has also been relaxed [34]. However, all of these approaches require images to be acquired in a perfectly controlled lighting environment. More recently, many solutions have involved the use of neural networks [29].

A situation that often arises in the digitization of archaeological heritage is that of *uncalibrated PS*, where the lighting is unknown. For example, in the context of prehistoric caves, access to the walls may be restricted for reasons of soil conservation, in which case the lighting source is directed "on the fly" towards the area of interest. Uncalibrated PS is known to be an ill-posed problem [8], for which the usual solution is to add a priori knowledge [26]. Neural approaches to uncalibrated PS generally assume that the lighting is directional. A first network determines the lighting [15], which makes the problem calibrated, but this approach fails if the directional lighting assumption is defeated. *Universal PS*, which has been proposed very recently to limit this risk [30], accommodates a wide variety of lighting conditions, learned by deep learning. Today, this is undoubtedly the most effective approach for solving uncalibrated PS. However, if the 3D scene is too different from the scenes used for learning, geometry inference may be erroneous.

For the time being, PS remains confidential in archaeology, although the community is showing increasing interest in it [17, 40, 42].

None of the methods described above fully addresses the problem of 3D reconstruction. It is therefore customary to combine several methods. For example, some approaches combine multi-view reconstruction and RTI [41, 43], or [31] and [36], which introduce a multi-view and multi-lighting approach. *Multi-view PS* is currently the subject of a great deal of work, thanks to spectacular advances in neural approaches [12, 50].

1.2. Research Aim

To validate the relevance of PS as a 3D heritage digitization technique, we need to compare this method with those currently used in field activities. To do this, a non-negligible obstacle lies in the difficulty of obtaining the ground truth of archaeological remains, then merging the data or projecting them into a common spatial reference frame, so that they can be compared [32].

Several modalities are tested in our study: lasergrammetry (TLS scanner), hand-held scanner (fringe projection), photogrammetry, calibrated PS, and PS calibrated using a photogrammetric result. To achieve quantitative comparisons, we ensure that the same equipment is used for all the acquisitions, and that the different 3D scanning modalities are used in the same session. We compare the methods either under field or controlled conditions, the final objective of this study being to propose a functional, practice-oriented and efficient processing pipeline.

2. Material

2.1. Prehistoric Cave Wall

The first case study is horse number 4 from the "Panneau des Chevaux" in the Chauvet cave (Ardèche, France), a UNESCO world heritage site. This piece of cave art could be examined qualitatively, based on an archaeological survey [20, 21]. On the other hand, as the variety of digital coverage spans almost twenty years, this makes quantitative analysis difficult: photographs and surveys taken between 1999 and 2011, lasergrammetry by TLS (we used the 2016 survey by surveyors Perazio Engineering for conservation), photogrammetry photographs (2024), photometric stereo photographs with calibration sphere (2022) or with RTI dome (2023).

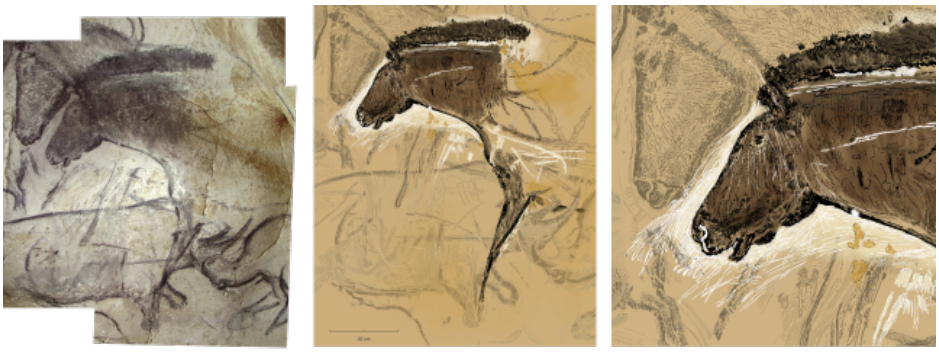

Figure 1: Panneau des Chevaux in the Chauvet cave (Ardèche, France). Left: Photomontage based on low-angle lighting to visualize the engravings. Center: Graphic record of the lines on horse number 4. Right: Enlarged view of the engravings and the shaping of the horse's head. Photomontage and measurements: Carole Fritz and Gilles Tosello.

This major work of cave art, dating back almost 38,000 years, presents a number of specific digitization challenges. It has been preserved because it is located in a cave, but this requires the use of battery-powered lighting and space-saving, non-invasive equipment that can withstand high humidity. Another specificity is constitutive of the work itself, which was created on a limestone wall covered with an ochre-colored clay film. The work is composed of a charcoal-based paint layer and a contrasting accentuation of contours, using fine infra-millimeter engravings. This fast and precise work is particularly evident around the nose and mouth of horse number 4. Finally, the whole is covered with a veil of calcite that acts as a varnish on the wall. The surface area of our study window measures less than one square meter.

2.2. Ceramic Figurine

The second case study is a small ceramic piece (5.7 cm by 4.2 cm and 1.4 cm thick) in the Musée Saint-Raymond (Toulouse, France). Dating from the 4th century BC and originating from Cyprus, it depicts a woman's head with calathos, broken at the neck. "Above the forehead, the hair is pulled symmetrically in wavy bands towards the temples, over the mitra (headband) marked in the center with a recumbent cross. They are surmounted by a high calathos (basket) made of rows of stylized plant elements" [6]. This terracotta object offers a double advantage in terms of quantitative comparison. Its material is close to the Lambertian model [19], which limits the difficulties of digitization. The other is the fineness of the surface detail, which provides information on the shaping and condition of the surface.

Figure 2: Cypriot ceramic head (Musée Saint-Raymond, Toulouse, France).

3. Methods

3.1. Data Acquisition

For the Panneau des Chevaux, the acquisitions were made for different purposes and by different operators. The difficulty lies in repositioning all the data in the same spatial reference frame, corresponding to the cave's topographical canvas.

The main cavity-scale acquisition was carried out using lasergrammetry. The two major acquisition phases took place in 2010 and 2016, at the request of the conservation department. The aim was to obtain an overall digital footprint of the site, with denser stations on the main panels. The 2010 3D point cloud was used, in particular, for the production of the Chauvet 2 replica [46]. The 2016 scan was carried out to meet fine mapping needs for conservation and research. The latter was chosen for our study.

The equipment used is a Zoller+Fröhlich 5010C scanner, which operates on the phase-difference principle. For the purposes of this study, only the station closest to the panel was used, at a distance of around three meters, with 43.3 million 3D points, in order to avoid any errors between stations [10]. The acquisition, which was performed with no light input, contains no RGB color information. In addition, signal intensity information is unusable on this panel, as the laser wavelength (around 1550 nm) is absorbed by the calcite deposit covering the work [11].

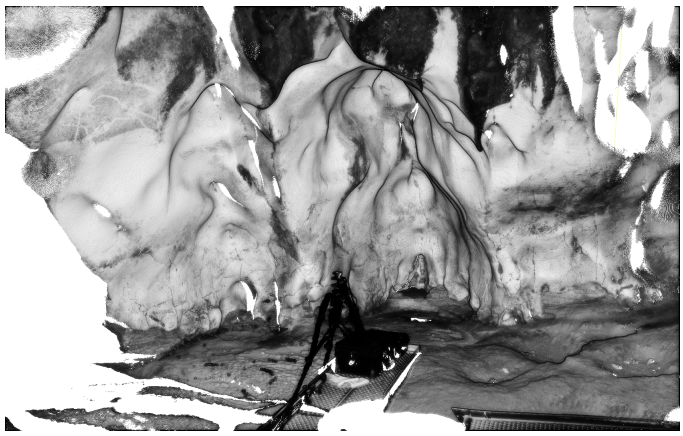

Figure 3: 3D point cloud of the Panneau des Chevaux obtained with a TLS scanner. The grey levels of the signal intensity are weakly representative of the RGB color of the wall. Credit: Perazio Engineering, French Ministry of Culture.

The second acquisition was for photogrammetry, using dense image correlation. Several photo sessions had already been carried out, notably by the Chauvet cave research team. However, in 2024 we took a new series of photographs with a Nikon Z7 II full-frame camera, equipped with a 50 mm focal-length lens, under lighting obtained by a set of LED panels diffusing a homogeneous light on the wall. The camera's main parameters were: open at F/14, ISO blocked at 100, and image size 8256×5504 (45.4 million pixels).

With these parameters fixed, the GSD (ground sample distance) depends only on the distance from the camera to the scene. This distance must be minimal to optimize the resolution of the photographs, but the scene must be fully visible in the camera's field of view, which defines a lower limit for the GSD. In our case, since the photographs are acquired at a distance of 1.8 m, the GSD is estimated at 150 µm/pixel.

The third acquisition, in 2023, was carried out using an RTI dome with 105 LEDs, 50 cm in diameter, with the same camera as described above, but with a 35 mm focal-length lens (see left-hand photograph in Figure 4). With a shooting distance of around 40 cm, the GSD is 50 µm/pixel.

The fourth and final acquisition corresponds to a series of photographs taken in the same pose, but under different lighting conditions (see right-hand photograph in Figure 4). A calibration sphere was positioned close to the wall, in the visible field. The light source was a continuously illuminating LED. Taken with a Canon R5 camera equipped with a 35 mm focal-length lens at around 1.6 m from the wall, the GSD is 200 µm/pixel. This acquisition comprises 21 photographs.

Figure 4: Implementing the digitization of the Panneau des Chevaux using calibrated photometric stereo. Left: Use of an RTI dome, for which the lighting has been pre-calibrated. Right: Use of a white sphere to calibrate lighting a posteriori.

For the Cypriot head, the acquisitions were all made in 2024, in a single day, in the reserves of the Musée Saint-Raymond in Toulouse. The first acquisition was made using an Artec Spider hand-held scanner. According to the manufacturer, this instrument guarantees 3D point accuracy and 3D mesh resolution down to 50 µm and 100 µm, respectively. All other acquisitions were carried out using photographic techniques, with the aim of obtaining the best possible results for each modality.

As in the first case study, we applied photographic 3D reconstruction methods of both geometric and photometric types. All photographs were taken with a Nikon Z7 II camera equipped with a 50 mm focal-length lens. The first acquisition, for photogrammetry, consisted of 48 photographs of the front face taken from an average distance of 22 cm. This gives an estimated GSD of 20 µm/pixel. The second acquisition, for photometric stereo, comprises six sequences of 35 shots from the RTI dome, for a GSD of around 25 µm/pixel. For the first sequence, the scene consisted of five white spheres, in order to pre-calibrate the RTI dome. Five other sequences of the front face of the Cypriot head were then taken, from five different viewpoints.

3.2. Data Processing

This section describes the different treatments applied to the data. We will focus in particular on the photometric stereo processing.

Lasergrammetry. Lasergrammetry acquisition was georeferenced by the aforementioned firm of surveyors. The 3D point clouds were processed with Trimble Realworks software [5] and exported in e57 format. The complete point cloud was converted into a 3D mesh using Metashape 1.8 software [1].

Hand-held Scanning. Hand-held scanner data were processed using Artec Studio software version 18 [2]. All depth frames with an error greater than 100 µm were removed. The 3D mesh, generated with the best possible resolution, was then exported.

Photogrammetry. The photogrammetry software used is Agisoft Metashape. Georeferencing is based on ten or so registration points obtained from the scanner acquisition. In the batch of photogrammetry photographs, it is interesting to insert the photometric stereo photographs taken with all n lighting (cf. the $\mathbf{s}^0$ lighting in Figure 9), since the SfM step provides the different camera poses, which will be useful for expressing the 3D reconstructions of the different photographic techniques in a common reference frame.

Single-view Photometric Stereo. In this study, we use several variants of the single-view photometric stereo (PS) technique. The physical model we adopt is the Lambertian one, which is well suited to archaeology, given that many objects are both opaque and diffusive. This model expresses the grey level $I(\mathbf{p})$ of an image point $\mathbf{p}$ in the following particularly simple form:

$$I(\mathbf{p}) = \rho(\mathbf{P})\,\mathbf{n}(\mathbf{P})^\top \mathbf{s}(\mathbf{P}) \tag{1}$$

In (1), $\rho(\mathbf{P}) \in \mathbb{R}^+$ denotes the *albedo* (the colour) of the $\mathbf{P}$ point in the scene which is projected at $\mathbf{p}$ into the image, $\mathbf{n}(\mathbf{P}) \in \mathbb{R}^3$ denotes the unit outgoing *normal* to the surface at this point, and the vector $\mathbf{s}(\mathbf{P}) \in \mathbb{R}^3$, called the *lighting vector*, characterises the direction and intensity of the luminous flux incident at point $\mathbf{P}$.

The calibrated approach of PS [48] assumes that the lighting vector $\mathbf{s}(\mathbf{P})$ is known, without necessarily assuming that it is uniform, i.e. independent of $\mathbf{P}$. Introducing the new unknown $\mathbf{m}(\mathbf{P}) = \rho(\mathbf{P})\,\mathbf{n}(\mathbf{P})$, Equation (1) becomes linear in $\mathbf{m}(\mathbf{P})$. Moreover, the use of n lighting $\mathbf{s}^i(\mathbf{P})$, $i \in \{1, \ldots, n\}$, can make the estimation of $\mathbf{m}(\mathbf{P})$ well-posed. Indeed, the n grey levels $I^i(\mathbf{p})$, $i \in \{1, \ldots, n\}$, can be grouped together in matrix form:

$$\begin{bmatrix} I^1(\mathbf{p}) \\ \vdots \\ I^n(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} \mathbf{s}^1(\mathbf{P})^\top \mathbf{m}(\mathbf{P}) \\ \vdots \\ \mathbf{s}^n(\mathbf{P})^\top \mathbf{m}(\mathbf{P}) \end{bmatrix} \tag{2}$$

which can be rewritten as:

$$\begin{bmatrix} I^1(\mathbf{p}) \\ \vdots \\ I^n(\mathbf{p}) \end{bmatrix} = \begin{bmatrix} \mathbf{s}^1(\mathbf{P})^\top \\ \vdots \\ \mathbf{s}^n(\mathbf{P})^\top \end{bmatrix} \mathbf{m}(\mathbf{P}) \tag{3}$$

As soon as $n \geq 3$ lighting vectors $\mathbf{s}^i(\mathbf{P})$ are used, and these vectors are not all coplanar, the matrix of the second member of (3) is of rank 3. The estimation of $\mathbf{m}(\mathbf{P})$ then becomes a well-posed problem, whose least squares solution is written:

$$\mathbf{m}(\mathbf{P}) = \begin{bmatrix} \mathbf{s}^1(\mathbf{P})^\top \\ \vdots \\ \mathbf{s}^n(\mathbf{P})^\top \end{bmatrix}^\dagger \begin{bmatrix} I^1(\mathbf{p}) \\ \vdots \\ I^n(\mathbf{p}) \end{bmatrix} \tag{4}$$

where the exponent $\dagger$ denotes the pseudo-inverse.

Knowing that $\mathbf{n}(\mathbf{P})$ is of norm 1, it is easy to deduce from (4):

$$\mathbf{n}(\mathbf{P}) = \frac{\mathbf{m}(\mathbf{P})}{\|\mathbf{m}(\mathbf{P})\|} \quad ; \quad \rho(\mathbf{P}) = \|\mathbf{m}(\mathbf{P})\| \tag{5}$$

Lighting Calibration for Photometric Stereo. In Model (1), both vectors $\mathbf{n}(\mathbf{P})$ and $\mathbf{s}(\mathbf{P})$ can be interchanged, since they only intervene via their scalar product. Knowing that $\rho(\mathbf{P})$ and $\mathbf{n}(\mathbf{P})$ can be estimated from the knowledge of $n \geq 3$ lighting vectors $\mathbf{s}^i(\mathbf{P})$, supposedly not all coplanar, is it possible, by analogous reasoning, to estimate $\rho(\mathbf{P})$ and $\mathbf{s}(\mathbf{P})$ from knowledge of $q \geq 3$ normals $\mathbf{n}(\mathbf{P}_j)$, $j \in \{1, \ldots, q\}$, supposedly not all coplanar?

If we define a vector equivalent to $\mathbf{m}(\mathbf{P}) = \rho(\mathbf{P})\,\mathbf{n}(\mathbf{P})$, in this case $\overline{\mathbf{m}}(\mathbf{P}_j) = \rho(\mathbf{P}_j)\,\mathbf{s}(\mathbf{P}_j)$, the q grey levels of q points $\mathbf{P}_j$ of known normal can be grouped together in a matrix form similar to (2):

$$\begin{bmatrix} I(\mathbf{p}_1) \\ \vdots \\ I(\mathbf{p}_q) \end{bmatrix} = \begin{bmatrix} \mathbf{n}(\mathbf{P}_1)^\top \overline{\mathbf{m}}(\mathbf{P}_1) \\ \vdots \\ \mathbf{n}(\mathbf{P}_q)^\top \overline{\mathbf{m}}(\mathbf{P}_q) \end{bmatrix} \tag{6}$$

but it is impossible to factor $\overline{\mathbf{m}}$ in (6) as was done for $\mathbf{m}$ in (3), because this vector depends on the point $\mathbf{P}_j$. In other words, whereas n equations (1) at the same point $\mathbf{P}$, subjected to n different illuminations, constitute a system of n linear equations with 3 unknowns, it turns out that q equations (1) at q points of known normal, under the same illumination, constitute a system of q linear equations with $3\,q$ unknowns, which correspond to the coordinates of the q vectors $\overline{\mathbf{m}}(\mathbf{P}_j)$. The problem is therefore ill-posed.

An admittedly very strong assumption, but sufficient to make this problem well-posed, consists in supposing not only that the normal is known at q points $\mathbf{P}_j$, but that these q points have the same albedo ρ and are subjected to the same illumination $\mathbf{s}$. In this case, vector $\overline{\mathbf{m}}$ is independent of $\mathbf{P}$ and can be estimated, in least squares, in a form strictly analogous to (4):

$$\overline{\mathbf{m}} = \begin{bmatrix} \mathbf{n}(\mathbf{P}_1)^\top \\ \vdots \\ \mathbf{n}(\mathbf{P}_q)^\top \end{bmatrix}^\dagger \begin{bmatrix} I(\mathbf{p}_1) \\ \vdots \\ I(\mathbf{p}_q) \end{bmatrix} \tag{7}$$

To deduce $\mathbf{s}$ from $\overline{\mathbf{m}}$, the fact that $\mathbf{s}$ is not a unit vector, unlike $\mathbf{n}$, is not a problem, because we can arbitrarily set $\rho = 1$, which implies $\mathbf{s} = \overline{\mathbf{m}}$, as far

as this is done for all the lighting, in order to guarantee the consistency of all the estimates.

The hypotheses that make the previous estimation problem well-posed are more or less easy to justify [28]. It is possible to know the normal at a certain number of points from multi-view data, thanks to SfM and MVS. It is more difficult to ensure that the lighting $\mathbf{s}$ is independent of the point $\mathbf{P}$ on the surface, unless these points are close to each other. Finally, it is only possible to know q points $\mathbf{P}_j$ of the same albedo thanks to a prior knowledge on the object, which is the case for lighting calibration patterns, for which relief and colour are perfectly known.

If, in order to estimate the illumination $\mathbf{s}$, it is necessary to know the normal at $q \geq 3$ points $\mathbf{P}_j$ with the same albedo ρ, it should be remembered that these normals do not all have to be coplanar, so that the pseudo-inverse matrix which appears in (7) is defined. Consequently, this disqualifies planar patterns. The simplest solution is to use a sphere of uniform colour. The silhouette of the image of a sphere in perspective projection is an ellipse whose eccentricity increases with the distance from the center of the sphere to the optical axis of the camera. Not only does this object have the widest possible variety of normal orientations, but the normal is perfectly known at any visible point on the silhouette. We can also dispense with the need for a calibration pattern, by using an RTI dome whose lighting can be pre-calibrated. Subsequently, the same PS algorithm can be applied to both types of acquisition.

It is also possible to use the 3D scene itself as a light calibration object, provided that the normals at a certain number of points of the scene are known. For example, if the user carries out a multi-view acquisition, followed by a multi-lighting acquisition, it is possible to calibrate the lighting using neither a white sphere nor an RTI dome, since the mesh obtained by photogrammetry is sufficiently reliable to be used as a calibration object [16].

3.3. Comparison between Different Methods

In order to compare these different 3D reconstruction methods, it is necessary to perform some alignments. On the one hand, to compare results by hand-held scanning and by photogrammetry, the meshes obtained are aligned by ICP (iterative closest point) [9]. On the other hand, to compare a triangular mesh by photogrammetry and a PS normal field, we calculate the normal field derived from the mesh, in the same frame as the PS acquisition.

4. Results and Discussion

This section compares the different approaches to 3D reconstruction presented in this article.

4.1. Panneau des Chevaux

Figure 5 shows three 3D reconstructions of the head of horse number 4, shown as normal fields in RGB representation obtained, respectively, by laser scanning, by photogrammetry using Metashape software, and by calibrated PS using an RTI dome.

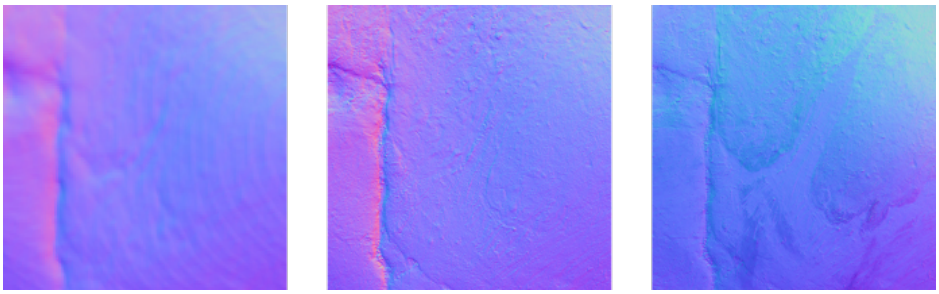

Figure 5: Three 3D reconstructions of the head of horse number 4, shown as normal fields in RGB representation: Laser scanning (left); Photogrammetry using Metashape software (center); Calibrated PS using an RTI dome (right). From a qualitative point of view, the latter technique is clearly the one that best reproduces high frequencies.

The first two methods enable the shape of the wall to be appreciated on a large scale (in the "low frequencies"), but where only the most visible cracks are rendered. It should be noted, however, that photogrammetry shows better resolution than laser scanning. This explains why it is the most popular method for archaeologists, who, when studying cave art, supplement the photogrammetric mesh with low-angle shots to visualize micro-reliefs.

On the other hand, PS seems to provide a very high level of detail, since micro-engravings are observable on the obtained normal field, enabling us to recover, for example, information on the relative chronology of nasal tracings. However, we note a bias in the low frequencies, probably due to insufficiently accurate calibration of the dome lighting. Indeed, the lighting which comes from LEDs located at a short distance from the wall is wrongly assumed to be parallel and uniform. Another significant disadvantage of using an RTI dome for a cave wall is that the visible field is extremely limited.

By way of quantitative comparison, Table 1 gives the mean angular errors between the three normal fields in Figure 5. These values, expressed in degrees, may seem to contradict the previous observation that PS would be more accurate than the other two methods. Firstly, it has already been said that the RTI dome lighting calibration is insufficiently accurate, resulting in a low-frequency bias visible in Figure 5. Secondly, since PS is supposed to render relief details better (the "high frequencies"), the pixel-by-pixel comparison between normals is not meaningful. Indeed, if the scanner result is considered as ground truth, because of its reputation for accuracy, it is difficult to challenge this presupposition, even with a method such as PS, which offers better resolution and, possibly, higher accuracy.

Laser scanning	Photogrammetry	Calibrated PS	MAE
⋆	⋆		6.5°
⋆		⋆	19.7°
	⋆	⋆	20.6°

Table 1: Mean angular errors (MAE) between pairs of normal maps of the Panneau des Chevaux (see Figure 5). The methods chosen for comparison are identified by stars.

The result obtained by PS using a calibration sphere is shown in Figure 6. This acquisition procedure without RTI dome is lighter to implement, and also enables larger wall surfaces to be scanned. However, it is not without its shortcomings. Illumination is estimated at a single point in the 3D scene, where the calibration sphere is located. In the absence of anything better, this estimate is then extrapolated to the entire scene, which again results in a low-frequency reconstruction bias. In addition, it is necessary to position the sphere in the visible field of the camera, as close as possible to the wall. This delicate operation causes shadows on the wall, resulting in artifacts in the estimated normal field, as clearly visible in Figure 6.

In addition to interesting results, PS seems to be an ideal method for use in caves, where it is generally dark and light can be controlled. However, certain difficulties may arise: the confined space of a cave can be restrictive, and if the dome cannot be positioned, lighting directed manually at the object to be digitized may be sufficient as long as a method of estimating lighting on the fly (by a calibration pattern or by the scene) is applicable. On the other hand, caves often have wet or calcite-covered walls, which therefore exhibit specular reflections. The use of robust PS methods is then necessary, in order to treat bright reflections as outliers [45].

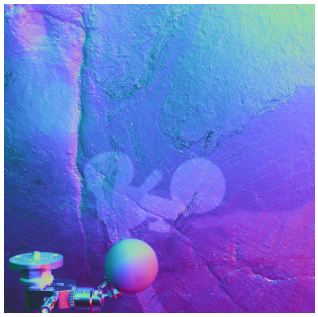

Figure 6: 3D reconstruction of the head of horse number 4 using calibrated photometric stereo, when a calibration sphere is positioned in the scene (see right-hand photograph in Figure 4). Compared with the result obtained with the RTI dome (see Figure 5), the shadow of the sphere projected onto the wall causes artifacts requiring specific treatment.

The Panneau des Chevaux example nevertheless demonstrates the qualitative contribution of PS, with either of these two variants. However, this technique is not intended to produce a real 3D reconstruction, at most a normal field that can be transformed into a depth map by integrating the normals [44]. We would like to propose a real 3D reconstruction method, on a par with other 3D digitizing techniques commonly used by archaeologists, but potentially more accurate.

4.2. Chypriot Head

Figure 7 shows three 3D reconstructions of the Cypriot head, shown as normal fields in RGB representation obtained, respectively, by photogrammetry using Metashape software, by calibrated PS using an RTI dome, and by the already mentioned variant of PS, where lighting is calibrated by the scene itself, based on the photogrammetric result (PG+PS). As before, PS results are more detailed than the photogrammetric one. The comparisons between these normal fields are given in Table 2.

Photogrammetry	Calibrated PS	PG+PS	MAE
⋆	⋆		11.4°
⋆		⋆	7.3°
	⋆	⋆	8.7°

Table 2: Mean angular errors (MAE) between pairs of normal maps of the Chypriot head (see Figure 7). The methods chosen for comparison are identified by stars.

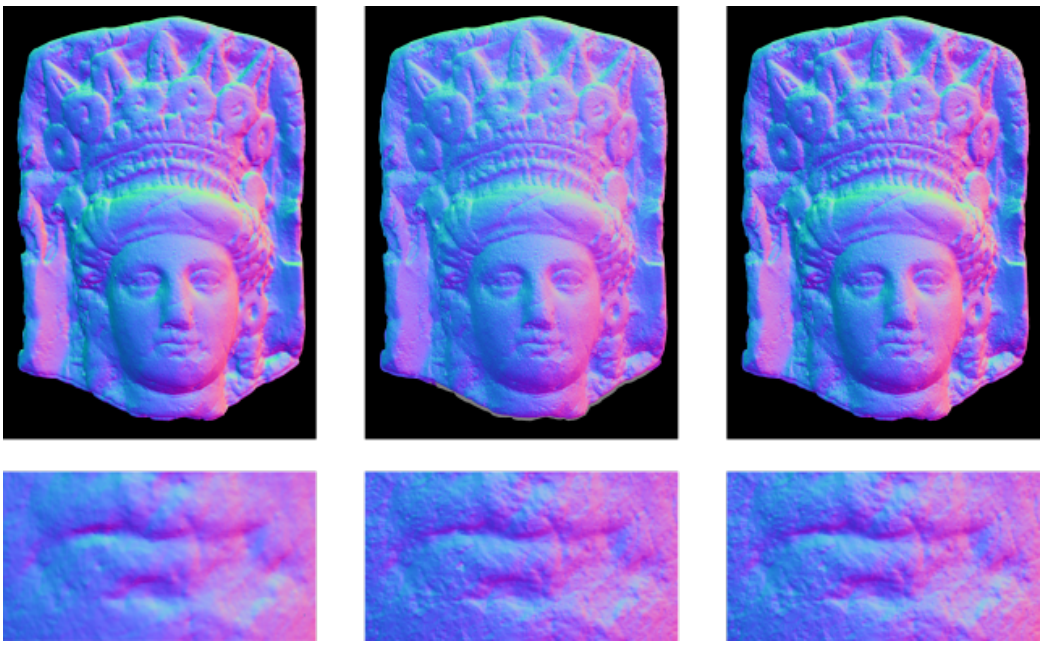

Figure 7: Three 3D reconstructions of the Cypriot head, shown as normal fields in RGB representation: Photogrammetry using Metashape software (left); Calibrated PS using an RTI dome (center); PG+PS (right). The bottom line shows zooms of the mouth area.

The differences between these methods are smaller than before. In particular, using the photogrammetric result to calibrate the lighting used by PS considerably reduces the low-frequencies bias. Indeed, by applying a Gaussian filter (of size 10×10 pixels) to the normal maps to isolate the low frequencies, the mean angular error between the results obtained by photogrammetry and by PS is only 2.0 degrees.

4.3. Towards Multi-view Photometric Stereo

Our approach uses the result of photogrammetry to calibrate the lighting of PS images obtained from a single viewpoint. We can go even further if we have multi-lighting data for each camera pose. This direct extension of our study is called *multi-view photometric stereo* (MV-PS). A first approach is to consider each view independently: the normal fields are integrated into depth maps, which are then merged to produce a 3D mesh. However, integrating the normal fields independently raises a number of problems since, in addition to the scale ambiguity inherent in each integration, which can possibly be

overcome by using the SfM 3D points as control points [39], inaccuracies at depth discontinuities are unavoidable.

A global integration approach seems more promising, as MV-PS methods are currently the subject of a great deal of work, thanks to spectacular advances in neural approaches. The 3D reconstruction by MV-PS presented on the right of Figure 8 was obtained using the RNb-NeuS method [12], which merges the normal fields obtained by PS into a 3D mesh.

Three 3D meshes of the Cypriot head obtained, respectively, by hand-held scanning, by photogrammetry, and by MV-PS, are shown in Figure 8. In order to measure the distances between these meshes [7], we used the CloudCompare software [3]. After registration by ICP of the mesh obtained by MV-PS with that obtained by hand-held scanning, we measured an average distance of 75 µm. Since the mesh by MV-PS is far more detailed than the other two, the question of a reference relief ("ground truth") arises again.

Figure 8: Three 3D reconstructions of the Cypriot head, represented as triangular meshes: Hand-held scanning (left); Photogrammetry using Metashape software (center); Multi-view PS (right), which is the method described on the diagram in Figure 9.

5. Conclusion and Perspectives

In this study, we compared different methods of 3D digitization of archaeological heritage objects. We set out to show that PS renders relief detail better than competing methods, provided that the lighting can be accurately estimated. Otherwise, the results are biased in the low frequencies.

The most promising solution seems to be MV-PS. The main steps of this 3D reconstruction method are shown in Figure 9. This geometric/photometric method does indeed seem to benefit from the advantages of both categories. In the practice of archaeological surveying, it can therefore represent a significant qualitative gain, bearing in mind that acquisitions can be made with equipment that archaeologists are familiar with.

From a colorimetric point of view, while the scanner offers a colorization of the mesh, photogrammetry enables the 3D mesh to be textured in a photorealistic way, i.e. under the lighting conditions prevailing at the time of shooting. On the other hand, MV-PS estimates the intrinsic color of the 3D scene, rather than its apparent color, which guarantees far more realistic renderings.

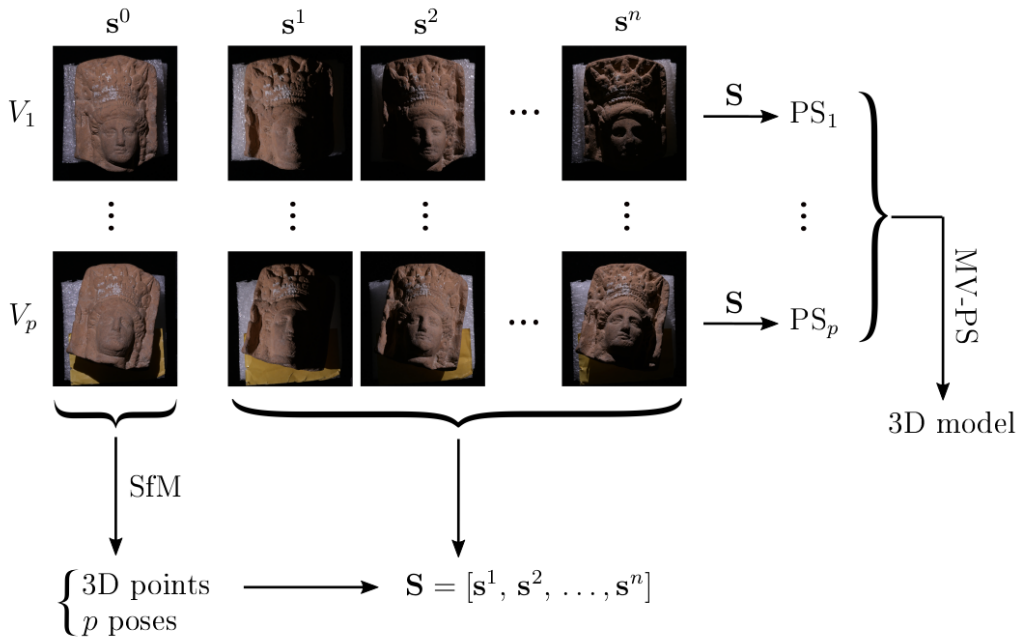

Figure 9: Description of the 3D reconstruction method we recommend, which is multi-lighting and multi-view: it requires n lightings $\mathbf{s}^1, \ldots, \mathbf{s}^n$, and p poses $V_1, \ldots, V_p$. For each pose, another non-directional lighting, denoted $\mathbf{s}^0$, produced when the n light sources are on, is used to apply SfM and thus obtain a sparse 3D point cloud, as well as the p camera poses expressed in a common reference frame. This first 3D reconstruction, non-dense but reliable in the low frequencies, is used to estimate the lighting matrix $\mathbf{S} = [\mathbf{s}^1, \mathbf{s}^2, \ldots, \mathbf{s}^n]$. It then becomes possible to apply calibrated PS for each of the p poses, and derive a complete 3D model using MV-PS.

Nevertheless, a metric study of the contribution of our method remains to be carried out. A first perspective is therefore to extend the test dataset and carry out a quantitative evaluation. However,the numerical comparison of the different approaches raised the question of "ground truth". Indeed, while the scanner (TLS or hand-held) acts as a reference, our approach seems to present more detail. It is important to clarify this question.

The method we recommend involves a number of steps, as shown in the diagram in Figure 9: shooting, lighting calibration, use of various algorithms, etc. A study of the impact of these different steps on the quality of the results of this operating chain seems essential.

Finally, numerical comparison between meshes is not necessarily the only criterion for adopting a particular 3D digitizing method. Other parameters need to be taken into account: material and human costs, transportability, light control, acquisition time, software licensing, etc. With this in mind, one short-term prospect is to integrate all the methods in the proposed operating chain, like the photometric stereo pipeline [33], into the AliceVision Meshroom software [24].

6. Acknowledgements

Our thanks go to Carole Fritz and Gilles Tosello for their work on the unique heritage of the Chauvet cave. All surveys are carried out as part of the scientific team's archaeological operation authorized by the French Ministry of Culture. We would also like to thank the team at the Musée Saint-Raymond in Toulouse, who gave us a warm welcome and access to their collections. Finally, we would like to thank François Baleux, CNRS engineer at the TRACES laboratory, for processing the lasergrammetry data.

References

[1] Agisoft Metashape software (version 1.8.5). `https://www.agisoft.com/`.

[2] Artec Studio 18 software. `https://www.artec3d.com/3d-software/artec-studio`.

[3] CloudCompare software (version 2.11.3). `https://github.com/cloudcompare/cloudcompare`.

[4] RealityCapture software. https://www.capturingreality.com/.

[5] Trimble RealWorks software. https://geospatial.trimble.com/en/products/software/trimble-realworks.

[6] Base Joconde : notice 05630030349. Ministère de la Culture, 2021. https://www.pop.culture.gouv.fr/notice/joconde/05630030349.

[7] Efstathios Adamopoulos, Fulvio Rinaudo, and Liliana Ardissono. A Critical Comparison of 3D Digitization Techniques for Heritage Objects. *ISPRS International Journal of Geo-Information*, 10(1), 2020.

[8] Peter Belhumeur and David Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 28:245–260, 1998.

[9] Paul Besl and Neil McKay. Method for registration of 3-D shapes. In *Sensor Fusion IV: Control Paradigms and Data Structures*, pages 586–606, 1992.

[10] Daniel Blersch, Christoph Held, Markus Mettenleiter, and Christoph Fröhlich. Towards Sustainable Digitization: Technology Solutions for Complex and Challenging Survey Projects. In *3D Research Challenges in Cultural Heritage III: Complexity and Quality in Digitisation*, pages 34–49. 2023.

[11] Yohan Boulard. L'interaction lumière-matière au service de la grotte Chauvet : modélisations et approches expérimentales. Thèse de doctorat en cours dirigée par Philippe Walter et Maguy Jaber.

[12] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5230–5239, 2024.

[13] Stamatis Chatzistamatis, George Tsekouras, and Christos-Nikolaos Anagnostopoulos. The Quality in 3D Acquisition of Cultural Heritage Assets: Challenges and Risks. In *3D Research Challenges in Cultural Heritage III: Complexity and Quality in Digitisation*, pages 65–76, 2023.

[14] Bertrand Chazaly. Relevés de la grotte Cosquer (partie 1) : la grotte, les premiers relevés, un monument et une histoire chaotiques. *XYZ*, 168, 2021.

[15] Guanying Chen, Kai Han, Boxin Shi, Yasuyuki Matsushita, and Kwan-Yee K Wong. Self-calibrating deep photometric stereo networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8739–8747, 2019.

[16] Benjamin Coupry, Jean Mélou, Antoine Laurent, Baptiste Brument, Pierre Gurdjos, Yvain Quéau, and Jean-Denis Durou. Stéréophotométrie avec estimation locale de l'éclairage - Application à la reconstruction 3D du patrimoine archéologique. In *Actes du congrès Reconnaissance des Formes, Image, Apprentissage et Perception*, 2024.

[17] Elisa Crabu, Federica Pes, Giuseppe Rodriguez, and Giuseppa Tanda. Ascertaining the Ideality of Photometric Stereo Datasets under Unknown Lighting. *Algorithms*, 16(8), 2023.

[18] Marta Díaz-Guardamino, Leonardo García Sanjuán, David Wheatley, and Víctor Rodríguez Zamora. RTI and the study of engraved rock art: A re-examination of the Iberian south-western stelae of Setefilla and Almadén de la Plata 2 (Seville, Spain). *Digital Applications in Archaeology and Cultural Heritage*, 2(2–3):41–54, 2015.

[19] Jean-Denis Durou, Maurizio Falcone, Yvain Quéau, and Silvia Tozza. A comprehensive introduction to photometric 3D-reconstruction. In *Advances in Photometric 3D-Reconstruction*, pages 1–29. Springer, 2020.

[20] Carole Fritz. Grotte Chauvet-Pont d'Arc : approche structurelle et comparative du Panneau des Chevaux. In *Actes du congrès de l'UISPP - L'art Pariétal Paléolithique dans son Contexte Naturel*, pages 69–86, 2004.

[21] Carole Fritz and Gilles Tosello. The Hidden Meaning of Forms: Methods of Recording Paleolithic Parietal Art. *Journal of Archaeological Method and Theory*, 14:48–80, 2007.

[22] 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.

[23] Alejandro Graciano, Lidia Ortega, Rafael Segura, and Francisco Feito. Digitization of religious artifacts with a structured light scanner. *Virtual Archaeology Review*, 8(17):49–55, 2017.

[24] Carsten Griwodz, Simone Gasparini, Lilian Calvet, Pierre Gurdjos, Fabien Castan, Benoit Maujean, Gregoire De Lillo, and Yann Lanthony. AliceVision Meshroom: An open-source 3D reconstruction pipeline. In *Proceedings of the 12th ACM Multimedia Systems Conference*, pages 241–247, 2021.

[25] Pierre Grussenmeyer, Albane Burens, Emmanuel Moisan, Samuel Guillemin, Laurent Carozza, Raphaëlle Bourrillon, and Stéphane Petrognani. 3D Multi-scale Scanning of the Archaeological Cave "Les Fraux" in Dordogne (France). In *Progress in Cultural Heritage Preservation*, pages 388–395, 2012.

[26] Heng Guo, Zhipeng Mo, Boxin Shi, Feng Lu, Sai-Kit Yeung, Ping Tan, and Yasuyuki Matsushita. Patch-based uncalibrated photometric stereo under natural illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7809–7823, 2021.

[27] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[28] Carlos Hernández, George Vogiatzis, and Roberto Cipolla. Multiview photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):548–554, 2008.

[29] Satoshi Ikehata. PS-transformer: Learning sparse photometric stereo network using self-attention mechanism. *arXiv preprint arXiv:2211.11386*, 2022.

[30] Satoshi Ikehata. Scalable, detailed and mask-free universal photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13198–13207, 2023.

[31] Ali Karami, Fabio Menna, and Fabio Remondino. Combining Photogrammetry and Photometric Stereo to Achieve Precise and Complete 3D Reconstruction. *Sensors*, 22(21):8172, 2022.

[32] Bahador Khaleghi, Alaa Khamis, Fakhreddine Karray, and Saiedeh Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28–44, 2013.

[33] Antoine Laurent, Jean Mélou, Benjamin Coupry, Thomas Sagory, Carole Fritz, and Jean-Denis Durou. Intégration de la stéréophotométrie pour le patrimoine dans la suite logicielle open-source Alicevision Meshroom. In *Actes des Journées du Consortium 3D SHS*, pages 1–10, 2023.

[34] Daniel Lichy, Soumyadip Sengupta, and David Jacobs. Fast light-weight near-field photometric stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12612–12621, 2022.

[35] Thomas Luhmann, Stuart Robson, Stephen Kyle, and Jan Boehm. *Close-Range Photogrammetry and 3D Imaging*. Walter de Gruyter GmbH & Co KG, 2023.

[36] Lindsay MacDonald, Isabella Toschi, Erica Nocerino, Mona Hess, Fabio Remondino, and Stuart Robson. Accuracy of 3D reconstruction in an illumination dome. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41:69–76, 2016.

[37] Thomas Malzbender, Dan Gelb, and Hans Wolters. Polynomial texture maps. In *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*, pages 519–528, 2001.

[38] Thomas Malzbender, Dan Gelb, Hans Wolters, and Bruce Zuckerman. Enhancement of Shape Perception by Surface Reflectance Transformation. In *Proceedings of the 9th International Fall Workshop on Vision, Modeling and Visualization*, 2004.

[39] Jean Mélou, Antoine Laurent, Carole Fritz, and Jean-Denis Durou. 3D digitization of heritage: photometric stereo can help. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 48, pages 145–152, 2022.

[40] Daisuke Miyazaki, Kenji Hara, and Katsushi Ikeuchi. Median Photometric Stereo as Applied to the Segonko Tumulus and Museum Objects. *International Journal of Computer Vision*, 86(2):229–242, 2010.

[41] María Mercedes Morita, Daniel Alejando Loaiza Carvajal, Ivana Leticia González Bagur, and Gabriel Mario Bilmes. A combined approach of SFM-MVS photogrammetry and reflectance transformation imaging to enhance 3D reconstructions. *Journal of Cultural Heritage*, 68:38–46, 2024.

[42] Matthieu Pizenberg, Abderrahim Elmoataz, and Yvain Quéau. Image Processing for Cultural Heritage Accessibility: Digitizing the Bayeux Tapestry. Preprint. https://hal.science/hal-03684035, June 2022.

[43] Samantha Porter, Nadine Huber, Christian Hoyer, and Harald Floss. Portable and low-cost solutions to the imaging of Paleolithic art objects: A comparison of photogrammetry and reflectance transformation imaging. *Journal of Archaeological Science: Reports*, 10:859–863, 2016.

[44] Yvain Quéau, Jean-Denis Durou, and Jean-François Aujol. Normal integration: A survey. *Journal of Mathematical Imaging and Vision*, 60:576–593, 2018.

[45] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3707–3716, 2016.

[46] Gilles Tosello, Alain Dalis, and Carole Fritz. Copier pour montrer, connaître avant de copier. Entre recherche et médiation, le fac-similé d'art préhistorique. *Collection EDYTEM. Cahiers de géographie*, 13:89–98, 2012.

[47] 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 *Proceedings of the 35th Conference on Neural Information Processing Systems*, 2021.

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

[49] Qingshan Xu, Weihang Kong, Wenbing Tao, and Marc Pollefeys. Multi-scale geometric consistency guided and planar prior assisted multi-view stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4945–4963, 2022.

[50] Dongxu Zhao, Daniel Lichy, Pierre-Nicolas Perrin, Jan-Michael Frahm, and Soumyadip Sengupta. Mvpsnet: Fast generalizable multi-view photometric stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12525–12536, 2023.