The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-2/W1-2022
7th International Workshop LowCost 3D – Sensors, Algorithms, Applications, 15–16 December 2022, Würzburg, Germany

3D DIGITIZATION OF HERITAGE: PHOTOMETRIC STEREO CAN HELP

Jean MÉLOU[a], Antoine LAURENT[a,b], Carole FRITZ[b], Jean-Denis DUROU[a]

[a]IRIT, UMR CNRS 5505, Toulouse, France [b]TRACES, UMR CNRS 5608, Toulouse, France

Technical Commission II

KEY WORDS: Cultural Heritage, 3D Digitization, Photometric Stereo, Photogrammetry, Multi-view Stereo.

ABSTRACT:

There are mainly two families of photographic 3D reconstruction. 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 the direction 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 order to avoid errors in the 3D models obtained, all the photographic techniques of 3D reconstruction have benefited, over the years, from algorithmic improvements that make them more and more robust to outliers or unreliable data. Moreover, the complementarity between these two types of approaches having been identified for a long time, many solutions have been proposed to merge them. Our work aims at providing the free and open-source photogrammetry software Meshroom with the benefits it could get from photometric stereo, particularly in the context of the 3D digitization of heritage, knowing that it is the only photometric technique for 3D reconstruction that has really proven itself.

1. INTRODUCTION

The 3D digitization of heritage is motivated by two complementary expectations: on the one hand, digital preservation for conservation, long-term archiving and analysis by professionals; on the other hand, accessibility to the greatest number of people thanks to the most realistic "digital twins" possible. Existing technologies have been deployed to meet specific qualitative expectations, for instance metrological needs (guarantees of measurement accuracy, scale and georeferencing) for archiving in archaeology, the need to optimize photorealism (quality of textures, colors and lighting) for immersive devices in museography, and so on.

To meet these different needs, two families of digitization solutions for surface acquisition are currently used. On the one hand, acquisitions made by *lasergrammetry* allow to reach a fairly good geometric accuracy, but are very expensive and provide poor quality textures. On the other hand, acquisitions made by *photogrammetry*, which work according to the principle of triangulation, from the matching between different viewpoints, require a simple camera, are inexpensive and can produce 3D models with better quality textures, but with increased geometric uncertainty. Photogrammetry and LiDAR capture are therefore often combined to take advantage of the strengths of each, in order to digitize heritage works, but this solution remains globally expensive and requires very high acquisition times.

Among the methods of photographic 3D reconstruction, in addition to photogrammetry techniques, there is another class of techniques, called *photometric*, which relate the appearance of a 3D point to the angle between its normal and the direction of the incident light. Whereas photogrammetry allows to find the global shape of a 3D scene, if it is sufficiently textured, photometric techniques highlight the details of the relief, provided that the model linking the lighting to the shape and reflectance of the scene is sufficiently realistic. In order to limit errors, these different photographic 3D reconstruction techniques have benefited, over the years, from algorithmic improvements that make them increasingly robust to outliers or unreliable data.

Although lasergrammetry and photogrammetry are operational and largely meet the needs of heritage actors, our ambition is to change their practices, by introducing into the basket of the scanning tools a 3D photometric reconstruction technique. A very recent such technique, called *reflectance transformation imaging* (RTI), has been specifically designed to read and analyze the finest engravings. By allowing to simulate any kind of lighting, in particular grazing lighting, this technique emphasizes the details of the relief, as illustrated in Figure 1. However, even if the results of RTI are very realistic, it is only a re-illumination technique, which does not allow us to produce 3D models. The only photometric technique that has really proven itself in 3D reconstruction is *photometric stereo* (PS), the principle of which will be recalled in Section 3.

Figure 1. The planarity defects of the facade are highlighted by a grazing lighting, which is the basic principle of RTI.

The complementarity between photogrammetry and photometry has long been identified. Following a pioneer article by Nehab et al. (Nehab et al., 2005), various solutions have been proposed to merge them (Karami et al., 2022, Ikeuchi et al., 2022). In this paper, we propose a new concrete solution of fusion between both these approaches which can answer the expectations of the heritage actors. Rather than designing a 3D photographic reconstruction software "from scratch", we aim to evolve the photogrammetry software Meshroom (Griwodz et al., 2021). In addition to being free and open-source, this software allows the user to design and archive a sequence of treatments adapted to his needs. The nodal aspect of the interface with the user seems to us particularly adapted to our final objective, which consists in endowing the Meshroom pipeline with a 3D reconstruction branch by photometric stereo.

Our article is structured as follows. In Section 2, we briefly review the state of the art of 3D digitization techniques used for cultural heritage. In Section 3, we recall the principle of photometric stereo. Section 4 describes our contribution, which consists in showing at which level of its pipeline of photogrammetry the Meshroom software could benefit the most from the contribution of photometric stereo. Finally, this preliminary study opens several perspectives which are listed in Section 5.

2. COMPARISON OF 3D DIGITIZATION TECHNIQUES FOR CULTURAL HERITAGE

The 3D acquisition of an object is carried out either in a volumetric way or in a surface way. The fragility of archaeological remains means that contact sensors are not taken into account and that the focus is on non-invasive and surface acquisitions. A much more detailed state of the art than this one can be found for example in (Adamopoulos et al., 2020).

2.1 Lasergrammetry

Terrestrial laser scanning (TLS) or LiDAR (Light Detection And Ranging) allows a 360° acquisition of the environment around the sensor. It calculates distances by launching infrared beams, either by measuring the time of flight or by phase shift. Each station captures several hundred thousand measurement points per second. This generates a spherical depth map from the center of the sensor. Each point has polar coordinates and a distance that will allow the creation of points in Cartesian coordinates. Each point has a measure of return signal strength. The TLS takes a series of photographs to generate an RGB view that will either colorize the point cloud or create a photorealistic texture on the mesh (Davis et al., 2017).

The main advantage is the acquisition accuracy, of the order of $1\ mm$ to $4\ mm$ at a distance of $10\ m$, depending on the model. However, the accuracy of the scanner is limited by the size of its incident beam, which forces it to be able to detect, at best, only millimetric details. It is therefore suitable for architecture and interior spaces. The laser beam works in total darkness, which makes it an asset for cave mapping. In order to acquire the most complete volume possible and with a spatial resolution of 3D points, the operator must multiply the stations and therefore the viewpoints. This increases the recording uncertainties between stations, which leads to a thickening of the point cloud. The equipment is expensive (several tens of thousands of euros). Recent models, equipped with a geolocalization system and a tablet, allow a fast and simplified implementation.

Pulsed light scanners (white, blue LED, with infrared) generate depth maps by triangulation, with the advantage of obtaining the metric directly. The announced spatial resolution is around 100 microns, with an accuracy of a few tens of microns depending on the model. Everything depends on the calibration of the sensor and the quality of the measurement defined by the manufacturer. The assembly of the different depth maps is done by geometry and with or without color (Graciano et al., 2017). This mode of operation can make it faulty on objects without asperity and non-textured.

The simplicity of implementation and learning of handheld 3D scanners makes them very functional for most small objects for obtaining sub-millimeter details. However, they are no longer suitable for objects larger than a few square meters. The color sensors are of low resolution, which results in textures closer to a mesh colorization than to a photorealistic texture. Its high cost, from a few thousand to tens of thousands of euros, associated with the purchase of the computer for the calculations, limits its wider use in archaeology (Patrucco et al., 2019).

2.2 Photogrammetry

The acquisition of volume by photogrammetry is obtained from photographs taken from multiple viewpoints (see Figure 2). The implementation simply relies on the overlap and displacement between the shots. Its cost is theoretically low since a digital camera is sufficient. The camera can be used on a vector (drone, pole) to reach remote areas and thus adapt to different types of volume. Photogrammetry is adapted to scale changes and multi-scalar digitization, from general volume to details. In order to make the changeover to metric or for its georeferencing, it is necessary to integrate rules or topographic coordinates.

Photogrammetry is the combination of two 3D reconstruction techniques: *structure-from-motion* (SfM) and *multi-view stereo* (MVS, see Section 4). SfM needs contrasting data for a good aerotriangulation result of the cameras. The same is true for the densification of SfM by MVS, which allows a geometric reconstruction of volumes by generating depth maps. This makes it difficult to digitize poorly textured surfaces. During the shooting, it is necessary to have a controlled and constant light for the creation of a photo-realistic texture without shadows and with the most accurate colorimetry possible. The quality of the model produced depends on the precision of the reference frames, and on the overlap of the shots on the rest of the object.

The reading of fine engravings depends on the quality of the photographic material (size and resolution of the sensor, quality of the lens). The increase in the number of pixels to be taken into account in the calculation will require a substantial computer equipment with sometimes long calculation times for generating the geometry.

2.3 Combining these Techniques

In order to improve the quality of the results obtained by each of these techniques, they can be combined (Cassen et al., 2014). Lasergrammetry and photogrammetry can be combined if they work in the same reference frame. This can be the case, either after alignment by SfM, or by using common benchmarks, or finally by georeferencing. The depth maps are then cross-referenced in the MVS step. Lasergrammetry by a handheld scanner can also be combined with photogrammetry, in such a way that the volumes are generated by the scanner, then the texture uses the data from photogrammetry (Barreau et al., 2022).

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

Figure 2. (a) General view of the photogrammetry setup to digitize in 3D a "statue-menhir". (b-c-d) Three images taken from different viewpoints, under a fixed diffusive lighting.

2.4 Photometric Techniques

The scanning methods presented here aim to provide a volume on which a photorealistic texture is plated. In some cases, fine engravings are visible on the texture, but are not readable on the relief. For certain heritage objects, for example, cave art which mixes engravings and paintings, it is relevant to distinguish the volume from the color in each pixel. For this, the contributions of photometry seem undeniable (Ikeuchi et al., 2022).

RTI has been developed since the 2000s for epigraphic purposes (Malzbender et al., 2004). Its field of use has subsequently expanded considerably (Earl et al., 2011). The principle of this method is that the grey level at a pixel is maximum when the normal to the surface is directed towards the light. Using several images, taken from the same viewpoint but under varying lighting, a normal field can be deduced by detecting the maximum value at this pixel through the set of images. This method does not produce accurate normal fields. Rather, it seeks to produce an improved surface simulation to facilitate study since the object can then be relighted. This method is particularly suitable for flat objects and comes close to the use of grazing lights for reading engravings (Díaz-Guardamino et al., 2015).

In order to avoid lighting calibration, which would require the presence of a sphere in each image, a previously calibrated dome may be used. However, this solution has the disadvantage of limiting the size of the object studied to that of the dome. While RTI is a fast and resource-efficient method, it has two major drawbacks. On the one hand, it does not produce volume, which considerably reduces the study of the reconstructed object. On the other hand, its accuracy is limited, since it is based on interpolation. Let us now focus on photometric stereo.

3. PHOTOMETRIC STEREO

3.1 Normal Estimation

If it is calibrated, then the grey level (or color level) of each photosite of the camera sensor is proportional to the luminance emitted by the scene, which depends on the relief of the scene, its reflectance and the illumination. In the case of a Lambertian surface, the grey level at an image point $\mathbf{p} = [u, v]^\top$ is written:

$$I(u,v) \;\propto\; \rho(u,v)\max\{\mathbf{n}(u,v)\cdot\mathbf{s}(u,v), 0\} \qquad (1)$$

where $I(u,v)$ denotes the grey level, $\rho(u,v)$ the albedo, $\mathbf{n}(u,v)$ the outside normal to the surface with unit norm, $\mathbf{s}(u,v) \in \mathbb{R}^3$ a vector characterizing the illumination, in norm and direction. A simplifying assumption is to neglect the self-shadows, i.e. to accept the values of $I(u,v)$ for which $\mathbf{n}(u,v)\cdot\mathbf{s}(u,v) < 0$. Another assumption is to assume $\mathbf{s}$ to be parallel and uniform, in such a way that it no longer depends on (u,v). Moreover, the proportionality coefficient of the relation (1) can be integrated with the albedo, which allows us to rewrite (1):

$$I(u,v) = \rho(u,v)\,\mathbf{n}(u,v)\cdot\mathbf{s} \qquad (2)$$

Even if the albedo $\rho(u,v)$ and the illumination $\mathbf{s}$ are assumed to be known, estimating the 3D shape from (2) is an ill-posed problem, called *shape-from-shading* (Horn, 1970), since in each pixel, the unknown normal has two degrees of freedom, but must satisfy a single equation. The principle of *photometric stereo* (PS) is to overcome this problem by using several images taken from the same viewpoint, but under m different illuminations (see Figure 3).

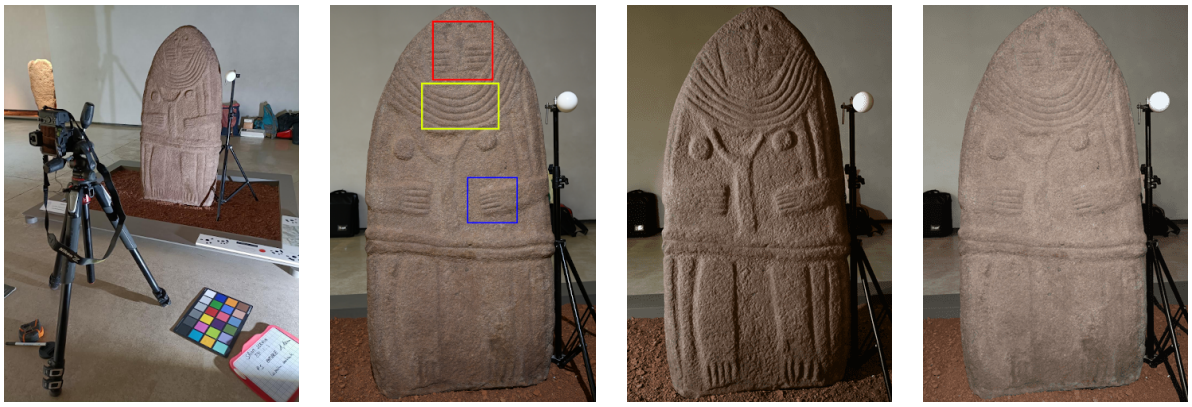
Let us characterize each illumination by a vector $\mathbf{s}^i \in \mathbb{R}^3$, $i \in \{1,\dots,m\}$. In order to obtain the most accurate 3D reconstruction possible, we assume that the illuminations are calibrated. Moreover, by introducing $\mathbf{m}(u,v) = \rho(u,v)\,\mathbf{n}(u,v)$, Equation (2) generalizes to the following linear system:

$$I^i(u,v) = \mathbf{m}(u,v)\cdot\mathbf{s}^i, \quad i \in \{1,\dots,m\} \qquad (3)$$

As soon as it is possible to illuminate the scene with $m \geq 3$ non-coplanar illuminations $\mathbf{s}^i$, (3) admits a unique solution in the least-square sense, from which ρ and $\mathbf{n}$ are easily deduced:

$$\rho(u,v) = \|\mathbf{m}(u,v)\| \quad \text{and} \quad \mathbf{n}(u,v) = \frac{\mathbf{m}(u,v)}{\|\mathbf{m}(u,v)\|} \qquad (4)$$

Many extensions of this technique have been proposed, for example to the case where the illuminations are not calibrated (Basri et al., 2007). Under the Lambertian assumption, the surface reflectance is entirely characterized by the albedo. However, for a large majority of surfaces, the reflectance shows deviations from the model (2). Even if the case of non-Lambertian reflectances has been studied (Mecca et al., 2016), we prefer to take advantage of the simplicity of the Lambertian model and consider deviations from it, such as specular reflections, as outliers, using robust estimators (Ikehata et al., 2012). Note also that deep learning approaches have recently appeared (Santo et al., 2017), which allow to free oneself from the Lambertian constraint (Chen et al., 2019).



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

Figure 3. (a) General view of the photometric stereo setup to digitize in 3D the same object as in Figure 2. (b-c-d) Three images taken from the same pose, but under a varying lighting estimated using the white sphere.

On the other hand, it is generally difficult to ensure that the illumination is parallel and uniform. More realistic models have been proposed. The methods for solving PS have been generalized to a point light source (Quéau et al., 2018a, Karami et al., 2022), which characterizes a LED illumination rather well. Other approaches modelling the illumination by spherical harmonics are proposed in (Basri et al., 2007, Haefner et al., 2019), which allows to describe quite faithfully the natural lightings.

3.2 Normal Integration

The second step of classical PS is integrating the estimated normals, which allows to find the *depth map* z characteristic of the 3D shape. Note however that some methods allow to estimate z directly, as for example (Smith and Fang, 2016).

If the projection is orthographic, the normal is easily expressed in terms of the depth gradient:

$$\mathbf{n}(u,v) = \frac{1}{\sqrt{1+p(u,v)^2+q(u,v)^2}} \begin{bmatrix} -p(u,v) \\ -q(u,v) \\ 1 \end{bmatrix} \quad (5)$$

where:

$$p := \frac{\partial z}{\partial u} \qquad \text{and} \qquad q := \frac{\partial z}{\partial v} \quad (6)$$

so that $\nabla z(u,v) = [p(u,v), q(u,v)]^\top$. In the perspective case, the change of variable $\tilde{z} = \ln(z)$ reduces the problem to a similar case, with slightly modified expressions of p and q.

Several methods can be used for normal integration, depending on the application requirements in terms of speed, robustness to noise in the estimated normal field, etc. For instance, a standard solution for the recovery of a smooth depth map consists in considering the quadratic variational problem:

$$\min_{z:\,\Omega\to\mathbb{R}} \int_\Omega \left\| \nabla z(u,v) - [p(u,v), q(u,v)]^\top \right\|^2 \mathrm{d}u\,\mathrm{d}v \quad (7)$$

which can be solved, e.g., using discrete sine or cosine transform (Simchony et al., 1990) or iterative methods (Bähr et al., 2017), depending upon the reconstruction domain Ω and the boundary conditions.

4. PHOTOMETRIC STEREO FOR MESHROOM

4.1 Multi-view Stereo (MVS)

For a given camera, the pose information provided by SfM makes it possible to define the central projection π that associates a 3D point $\mathbf{P}$ on the surface with its image $\mathbf{p} = \pi(\mathbf{P})$. This projection is invertible if the depth function z is known:

$$\mathbf{P} = \pi_z^{-1}(\mathbf{p}) \quad (8)$$

where the index z is used to indicate that, without the knowledge of function z, this writing would be ambiguous. Let us suppose that we have $n+1$ images of the same 3D scene, and the corresponding $n+1$ camera poses. The first pose can be chosen as a reference. For a 3D point $\mathbf{P}$ visible in all the images, let us note $\mathbf{p} = \pi(\mathbf{P})$ the projection of this point in the reference image, and:

$$\mathbf{p}_k = \pi_k(\mathbf{P}), \quad k \in \{1,\dots,n\} \quad (9)$$

its projections in the other n images, called "control images". On the other hand, according to Lambert's law:

$$I_k(\mathbf{p}_k) = I(\mathbf{p}), \qquad k \in \{1,\dots,n\} \quad (10)$$

where I_k and I denote, respectively, the grey level functions of the image number k and of the reference image. For a point $\mathbf{p}$ in Ω, which is the set of points of the reference image visible in all the control images, we deduce from (8), (9) and (10):

$$I_k \circ \pi_k \circ \pi_z^{-1}(\mathbf{p}) = I(\mathbf{p}), \quad k \in \{1,\dots,n\} \quad (11)$$

MVS searches for the z depth function corresponding to the reference image that maximises the photometric consistency of the reference image with the n control images. Equations (11) are never exactly satisfied, partly because real surfaces are never perfectly Lambertian. Considering z as an unknown function, and denoting by $z_{i,j}$ the value of the depth associated with a pixel $\mathbf{p}_{i,j} \in \Omega$, Equations (11) can be reformulated as a discrete minimization problem:

$$\widehat{z_{i,j}} = \underset{z_{i,j}\,\in\,\mathbb{R}}{\operatorname{argmin}} \sum_{k=1}^{n} \left[I_k \circ \pi_k \circ \pi_{z_{i,j}}^{-1}(\mathbf{p}_{i,j}) - I_{i,j} \right]^2 \quad (12)$$

where $I_{i,j}$ is the greylevel of $\mathbf{p}_{i,j}$ in the reference image.

Note that the function to be minimized in Equation (12) may turn out to be nonlinear, non-derivable and/or non-convex, which makes the optimization potentially difficult. Therefore, the minimization is usually performed by a brute-force search on a set of predefined values of the depth. This strategy, first introduced in (Hernández, 2004), may seem simplistic, but proves to be very efficient for the estimation of depth functions of highly textured 3D scenes (Goesele et al., 2006).

4.2 Comparison between Photogrammetry and PS

Before all, let us emphasize on two major advantages of PS, in comparison with photogrammetry. First, PS not only allows to estimate the relief of the scene, but also its *intrinsic color*, whereas photogrammetry maps the apparent color of the scene at the time of shooting. Second, the production of a colored 3D model is extremely fast, compared to photogrammetry.

Apart from these two major differences, the most important question is which of these two methods of photographic 3D reconstruction is more capable of providing a 3D model that is faithful to the original. The resolution of the 3D models obtained by PS is only limited by the size of the pixels, i.e. the dimensions of the portion of surface corresponding to a pixel, and this, whether the surface is textured or not.

It is not as simple to predict at which resolution photogrammetry can reconstruct a 3D scene. In order not to disqualify photogrammetry definitively, because of its inability to reconstruct a smooth untextured surface, we chose as a test object a "statue-menhir", which is the anthropomorphic representation in sandstone illustrated in Figures 2 and 3.

In order to make the comparison as fair as possible, we used an equivalent number of images. All pictures were taken with a Nikon Z7 II camera with a $35\ mm$ lens. The pictures, which were taken approximately $1\ m$ away from the statue-menhir, contain 45.75 megapixels.

Figure 4 shows the 3D reconstruction of the front side of the statue-menhir by photogrammetry, using 30 views such as those of Figure 2. On the other hand, Figure 5 shows a colored representation of the normal field estimated by PS, using 20 images such as those of Figure 3. Although both these results seem correct at this scale, we will see in Subsection 4.3 that they are not equivalent.

The size of the data used by the two methods is about the same, but this is not the case of the computation time. Indeed, while the result of Figure 4, which was obtained using the software Meshroom, required about one hour on GPU, the estimation of the normal field of Figure 5 and its integration required only a few minutes on a standard laptop.

4.3 Comparison between MVS and PS Depth Maps

The photogrammetry result presented in Figure 4 has been obtained using the software Meshroom with default settings. This free and open-source software separates the different steps of the pipeline and displays them in nodal form at the bottom of its interface (see Figure 6). Therefore, the user is free to add his own nodes, and to integrate them into the chain, as long as the input and output data are consistent with the rest of the pipeline.

Figure 4. 3D reconstruction of the front side of the statue-menhir by photogrammetry, using 30 views such as those of Figure 2.

Figure 5. Normal field of the front side of the statue-menhir estimated by PS, using 20 images such as those of Figure 3.

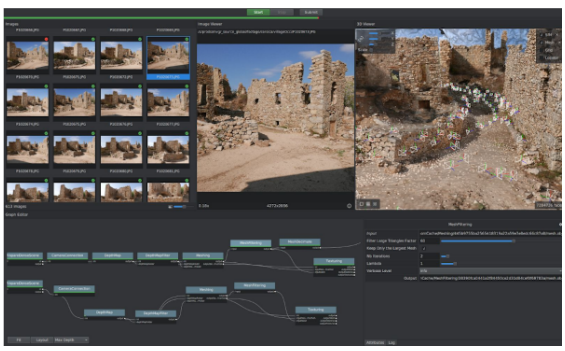


Figure 6. Meshroom interface.

Like most photogrammetry software, Meshroom follows the following sequence:

1. Detection of points of interest in each image.
2. Matching of points of interest between images.
3. Estimation of the different camera poses (see Figure 7) and of a sparse 3D point cloud by SfM.
4. Densification of the point cloud by MVS.
5. Texture plating on the 3D model.

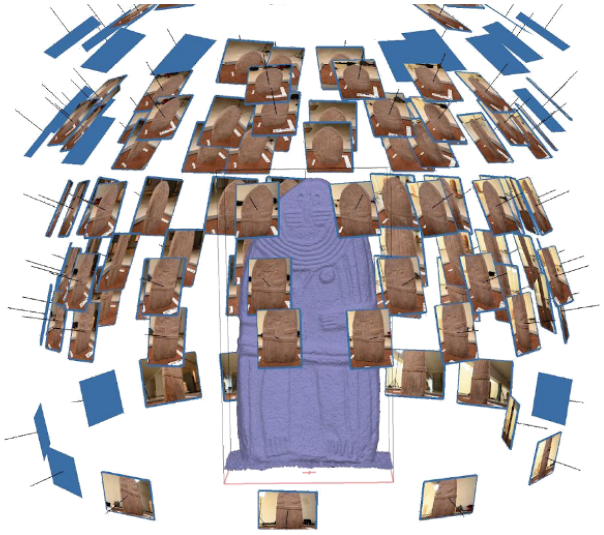


Figure 7. Camera poses estimated by Meshroom SfM, using 122 images such as those of Figure 2 (visualization by Metashape).

Let us now focus on step 4 of this pipeline. In reality, the Meshroom MVS step proceeds in two successive sub-steps:

4.a A depth map is estimated for each view, which means that each image is successively used as a reference image (see Section 4.1).

4.b Then, after removal of a number of 3D points deemed unreliable, a 3D mesh is built from these depth maps and from the knowledge of all the camera poses, which have been estimated by SfM.

Figure 8 allows us to make a qualitative comparison between different depth maps corresponding to the three rectangles highlighted in blue, yellow and red in Figure 3-b. The depth maps on the first line are obtained at the end of sub-step 4.a. A frame is clearly visible, which is due to an undersampling of the images, whose purpose is to accelerate the calculations. The depth maps on the second line of Figure 8 are artificially generated from the 3D model at the end of sub-step 4.b, by reprojecting the mesh on the pixel grid. The results of both these lines clearly show that the relief is much better restored after sub-step 4.b than after sub-step 4.a.

Finally, the third line of Figure 8 show the depth maps obtained by PS, after integration of the estimated normals using discrete cosine transform (Simchony et al., 1990). Since one depth is estimated per pixel, it was expected that the resolution of the relief would be qualitatively better than using MVS, before or after meshing.

Of course, a quantitative comparison of these depth maps would be welcome, but we come up against a well-known difficulty in computer vision: which relief can be considered as "ground truth"? In (Quéau et al., 2018a), a 3D model obtained by laser scanning with structured light projection is used as ground truth, but how to guarantee the validity of such an assertion?

4.4 Integration of PS in Meshroom

The comparison of Subsection 4.3, between the depth maps obtained by MVS and PS, although qualitative, still allows us to say that there must be a benefit to bypassing Meshroom sub-step 4.a, so as to provide depth maps obtained by PS as input to sub-step 4.b. Of course, this will require increasing the number of images. Instead of taking n photographs under n different poses, or m photographs under the same pose, but under m different lightings, it will be necessary in this case to take m photographs under each pose, that is to say mn photographs.

As already mentioned, the idea of merging photogrammetry and photometric stereo is not new. Our aim is not to propose a new fusion method, but simply to replace a link in the Meshroom pipeline, which is a very optimized photogrammetry software. Instead of computing a depth map per view by MVS, which inevitably causes biases, even for a highly textured scene like the example we have chosen, we advocate computing each of these depth maps by PS.

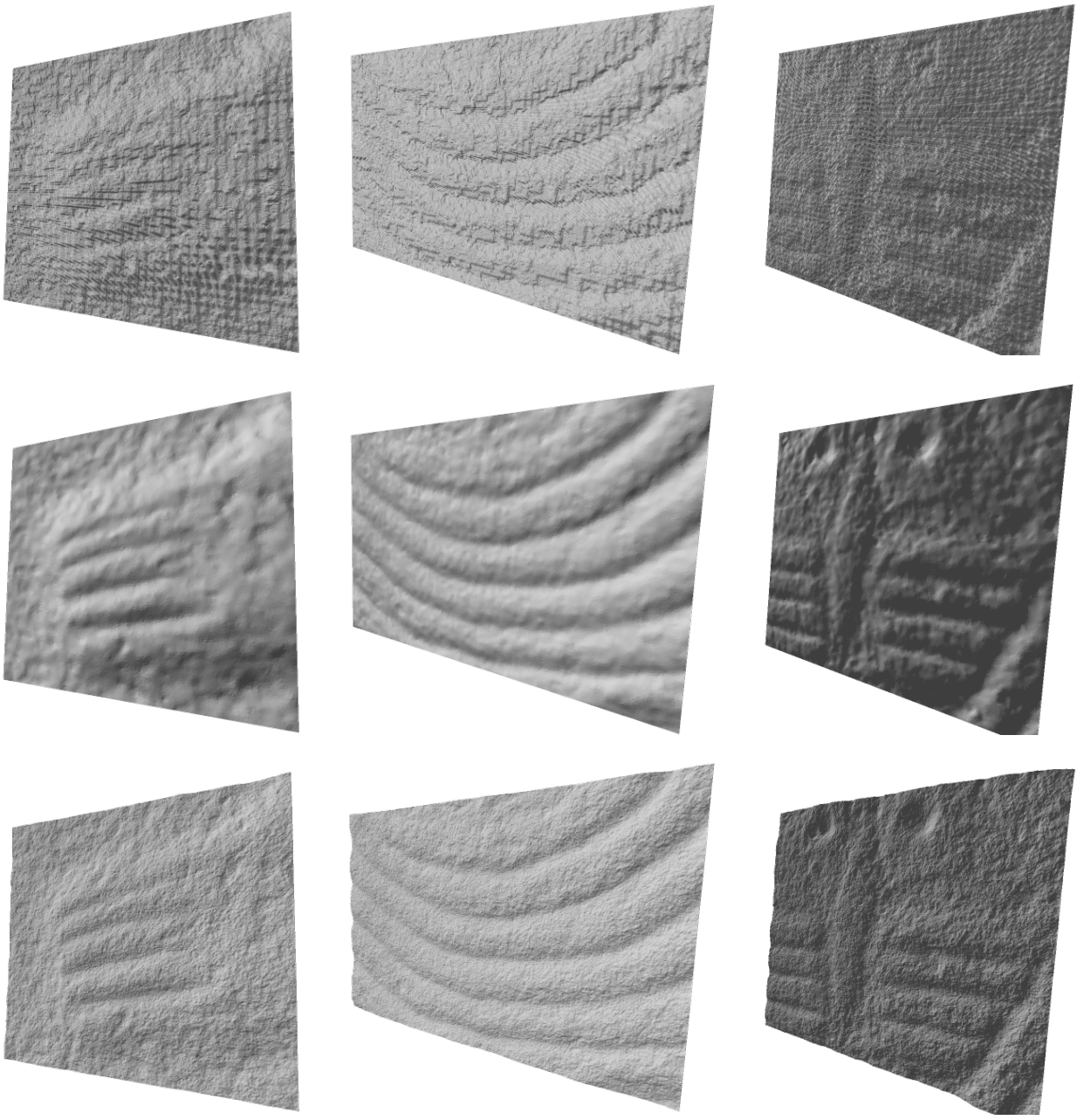
It is well known that PS renders well the high frequencies of the relief, but not the low ones. This is mainly due to the illumination estimation, which is generally too coarse to describe accurately the incident flux at each point of the scene. One solution is to use a more sophisticated lighting estimation method (Karami et al., 2022). Another way, simpler to implement, consists in using the 3D points obtained by SfM as hard constraints during the integration of the normals, which is indeed possible with some integration methods (Quéau et al., 2018b). By doing so, we will also be able to compute scaled depth maps, which would otherwise be impossible, and yet is essential in the eyes of archaeologists.

5. CONCLUSION AND FUTURE WORK

In this paper, we have shown that it is possible to graft PS onto an optimized pipeline of photogrammetry, rather than trying to merge both approaches "from scratch". Among the benefits of this graft, it is not useless to recall that the intrinsic color will be able to be estimated, in addition to the relief, and that the computation time will be drastically reduced, since the MVS step of Meshroom cannot be carried out without GPU.

Let us recall that our goal is to help the heritage actors in their digitization tasks. The constraint of taking m photographs per camera pose can be alleviated by using an RTI dome. Such an equipment could accelerate the acquisition time, since the lighting of the LEDs is synchronized with the camera triggering. Moreover, an RTI dome can be pre-calibrated, which would have the advantage of not having to place a white sphere in the vicinity of the scene, which is sometimes difficult, for example when it is precious or fragile.

Finally, it is not possible to end without mentioning the learning approach which, thanks to ever more powerful neural networks, will probably be able to supplant, in the medium term, all or part of the approaches mentioned above (Wang et al., 2021).



Blue rectangle Yellow rectangle Red rectangle

Figure 8. Qualitative comparison of depth maps corresponding to the three rectangles highlighted in color in Figure 3-b. First and second lines: depth maps obtained at the sub-steps 4.a and 4.b of Meshroom MVS. Third line: depth maps obtained using PS. Although this is only a qualitative judgment, the results of the third line best reproduce the granular appearance of the stone.

REFERENCES

Adamopoulos, E., Rinaudo, F., Ardissono, L., 2020. A critical comparison of 3D digitization techniques for heritage objects. *ISPRS International Journal of Geo-Information*, 10(1), 10.

Bähr, M., Breuß, M., Quéau, Y., Boroujerdi, A. S., Durou, J.-D., 2017. Fast and accurate surface normal integration on non-rectangular domains. *Comput. Visual Media*, 3(2), 107–129.

Barreau, J.-B., Gagnier, A., Gaugne, R., Marchand, G., Gómez, J. C., Gouranton, V., Colleter, R., 2022. Use of Different Digitization Methods for the Analysis of Cut Marks on the Oldest Bone Found in Brittany (France). *Appl. Sciences*, 12(3), 1381.

Basri, R., Jacobs, D., Kemelmacher, I., 2007. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3), 239–257.

Cassen, S., Lescop, L., Grimaud, V., Robin, G., 2014. Complementarity of acquisition techniques for the documentation of Neolithic engravings: lasergrammetric and photographic recording in Gavrinis passage tomb (Brittany, France). *Journal of Archaeological Science*, 45, 126–140.

Chen, G., Han, K., Shi, B., Matsushita, Y., Wong, K.-Y. K., 2019. Self-calibrating deep photometric stereo networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 8739–8747.

Davis, A., Belton, D., Helmholz, P., Bourke, P., McDonald, J., 2017. Pilbara rock art: laser scanning, photogrammetry and 3D photographic reconstruction as heritage management tools. *Heritage Science*, 5(1), 1–16.

Díaz-Guardamino, M., Sanjuán, L. G., Wheatley, D., Zamora, V. R., 2015. 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.

Earl, G., Basford, P., Bischoff, A., Bowman, A., Crowther, C., Dahl, J., Hodgson, M., Isaksen, L., Kotoula, E., Martinez, K. et al., 2011. Reflectance transformation imaging systems for ancient documentary artefacts. *Electronic visualisation and the arts (EVA 2011)*, 147–154.

Goesele, M., Curless, B., Seitz, S. M., 2006. Multi-view stereo revisited. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2, IEEE, 2402–2409.

Graciano, A., Alvarado, L. O., Sánchez, R. J. S., Higueruela, F. R. F., 2017. Digitization of religious artifacts with a structured light scanner. *Virtual Archaeology Review*, 8(17), 49–55.

Griwodz, C., Gasparini, S., Calvet, L., Gurdjos, P., Castan, F., Maujean, B., De Lillo, G., Lanthony, Y., 2021. AliceVision Meshroom: An open-source 3D reconstruction pipeline. *Proceedings of the 12th ACM Multimedia Systems Conference*, 241–247.

Haefner, B., Ye, Z., Gao, M., Wu, T., Quéau, Y., Cremers, D., 2019. Variational uncalibrated photometric stereo under general lighting. *Proceedings of the IEEE International Conference on Computer Vision*, 8539–8548.

Hernández, C., 2004. Stereo and Silhouette Fusion for 3D Object Modeling from Uncalibrated Images Under Circular Motion. Thse de doctorat, École Nationale Supérieure des Télécommunications.

Horn, B. K. P., 1970. Shape from Shading: a Method for Obtaining the Shape of a Smooth Opaque Object from One View. PhD Thesis, MIT.

Ikehata, S., Wipf, D., Matsushita, Y., Aizawa, K., 2012. Robust photometric stereo using sparse regression. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 318–325.

Ikeuchi, K., Morimoto, T., Kamakura, M., Kuchitsu, N., Kawano, K., Ikeda, T., 2022. Kyushu Decorative Tumuli Project: From e-Heritage to Cyber-Archaeology. *International Journal of Computer Vision*, 1–18.

Karami, A., Menna, F., Remondino, F., 2022. Combining Photogrammetry and Photometric Stereo to Achieve Precise and Complete 3D Reconstruction. *Sensors*, 22(21), 8172.

Malzbender, T., Gelb, D., Wolters, H., Zuckerman, B., 2004. Enhancement of Shape Perception by Surface Reflectance Transformation. *VMV*, 183.

Mecca, R., Quéau, Y., Logothetis, F., Cipolla, R., 2016. A single-lobe photometric stereo approach for heterogeneous material. *SIAM Journal on Imaging Sciences*, 9(4), 1858–1888.

Nehab, D., Rusinkiewicz, S., Davis, J., Ramamoorthi, R., 2005. Efficiently combining positions and normals for precise 3D geometry. *ACM Transactions on Graphics*, 24(3), 536–543.

Patrucco, G., Rinaudo, F., Spreafico, A., 2019. A New Handheld Scanner for 3D Survey of Small Artifacts: the Stonex F6. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Quéau, Y., Durix, B., Wu, T., Cremers, D., Lauze, F., Durou, J.-D., 2018a. LED-based photometric stereo: Modeling, calibration and numerical solution. *Journal of Mathematical Imaging and Vision*, 60(3), 313–340.

Quéau, Y., Durou, J.-D., Aujol, J.-F., 2018b. Variational methods for normal integration. *Journal of Mathematical Imaging and Vision*, 60(4), 609–632.

Santo, H., Samejima, M., Sugano, Y., Shi, B., Matsushita, Y., 2017. Deep photometric stereo network. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 501–509.

Simchony, T., Chellappa, R., Shao, M., 1990. Direct analytical methods for solving Poisson equations in computer vision problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5), 435–446.

Smith, W. A. P., Fang, F., 2016. Height from photometric ratio with model-based light source selection. *Computer Vision and Image Understanding*, 145, 128–138.

Wang, X., Wang, C., Liu, B., Zhou, X., Zhang, L., Zheng, J., Bai, X., 2021. Multi-view stereo in the Deep Learning Era: A comprehensive review. *Displays*, 70, 102102.