ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume X-M-2-2025
30th CIPA Symposium “Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation”, 25–29 August 2025, Seoul, Republic of Korea

Blurred Areas in Very Short Range Photogrammetry: A Flexible Alternative to Focus-Stacking. Case Study of a Complex-Geometry Ornament of Strasbourg’s Cathedral, France

Yannick Faure[1], Pierre Grussenmeyer[1], Tania Landes[1]

[1] Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, 67084 Strasbourg, France
(yannick.faure, pierre.grussenmeyer, tania.landes)@insa-strasbourg.fr

Keywords: Cultural Heritage 3D documentation, Optical Blur Processing, Laplacian Pyramid, Image Masks, Very Short Range Photogrammetry, Focus-stacking.

Abstract

Cultural Heritage documentation of small artifacts with complex geometry processed in photogrammetry has to deal with narrow depth-of-field inherent to optics. The focus-stacking method is traditionally employed to overcome this issue, but has to be deployed on a stable environment in terms of movement and light. It requires a very large number of photographs which are modified in colors and geometries, depending on focus-stacking algorithms, to be able to generate the focus-stacked images needed by Structure from Motion (SfM) processes. This article proposes a more flexible method, suitable for unmovable artifacts or acquisitions in the field where tripods for lights and cameras are excluded. The proposed method uses multiple single-shot photographs, as in traditional photogrammetry of larger items, improved with paired masks on areas suffering from optical blur. These masks are automatically generated using Laplacian Pyramid after an operator reviews the settings to fit the full set of the acquisition pictures. The case study is a Strasbourg’s Cathedral ornament carved in red sandstone in the Flamboyant Gothic style. This delicate sculpture presents a complex geometry with recesses and protrusions. Three processes leading to photogrammetric 3D dense clouds are described and evaluated: the proposed method with masks, the same method but without masks and a focus-stacking approach. These models are compared with a reference acquired using a metrology laser scanner, demonstrating the benefits of masking blur for accuracy and noise. When assessed relative to the focus-stacking approach, the proposed method with masks achieves similar accuracy while offering greater speed and adaptability.

1. Introduction

To perform documentation in 3D, various methods are employed. Currently, the major techniques used are: laser-scanning, mostly used for large spaces such as buildings, or for very thin details with metrology dedicated laser scanners; structured-light for small objects; and photogrammetry for both small and large items, more affordable compared to the other ones and providing textures and 2D documentation as well.

Traditionally, SfM algorithms assume the camera interior orientation to be constant over a whole set of pictures to enable the most accurate automatic self-calibration possible (Luhmann et al., 2014a). In other words, the operator has to choose the desired aperture (in ”manual mode” or ”aperture mode” of the camera) and the right focusing distance to cover its full acquisition in terms of depth-of-field (DoF) and image sharpness, maintain these settings unchanged, especially by switching its lens to ”manual focus” (MF) mode. To ensure stability of the interior parameters, stabilization of the lens and the sensor (when available) also have to be switched off (Historic England, 2017).

As a result, the chosen aperture is usually in the mid-range of its possible values, since a large aperture would lead to a too shallow DoF for photogrammetry applications, while a small aperture (to increase DoF) (figure 1) would lead to a loss of resolution due to the diffraction of light. In this latter case, the softened image is a consequence of the growing diameters of the Airy disks as aperture decreases, gradually loosing in resolving power until reaching the limit of resolution due to diffraction (Rayleigh criterion) (Gallo et al., 2012).

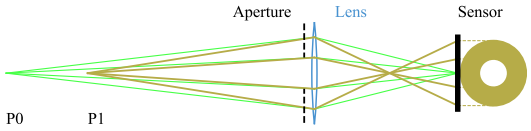


Figure 1. Difference in spot area of a point’s image depending on aperture. For every point which is not in the focal plane such as P0, as aperture decreases and before seeing the diffraction effect, the spot image of every point, such as P1, becomes smaller.

The main challenge here is to find the right compromise between widening the DoF and loosing resolution, knowing the practical consequence of a reduced aperture for the operator is the increasing exposure time, which cannot be balanced by too high ISO sensitivity (leading to noisy photographs).

The very notion of DoF is not binary and depends on what is considered either sharp or blurry in a picture. The transition between these two states is nuanced and subjective, so a consensus is established to evaluate them by calculating the maximal spot diameter projected on the sensor by the image of a point (figure 2). This spot is known as the circle of confusion and depending on the ratio between its diameter and the pixel’s size or the image diagonal, the location of a point is considered in-focus or out-of-focus (Luhmann et al., 2014b).

1.1 Types of blur and processing

Photographs can suffer from various kinds of blur: on one hand, motion blur has a typical directional components very useful to

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-M-2-2025-89-2025 | © Author(s) 2025. CC BY 4.0 License.

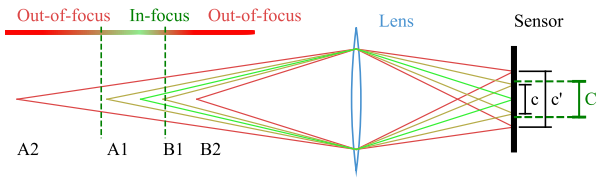


Figure 2. The circle of confusion is a threshold to determine the DoF. Since the diameter of the spot *c* made by A1 or B1 is below the circle of confusion *C*, these points are considered in-focus. Because diameter *c'* exceeds *C*, A2 or B2 are out-of-focus.

handle it (Levin et al., 2009; Xu and Jia, 2010). On the other hand, optical blur is mostly out-of-focus blur, smoothing gradually details both in front of and behind the DoF. Several methods are employed to classify the type of blur, sharpen it using deconvolution or mask it. When traditionally image gradients or spectra are analyzed, some authors use either machine learning on those metrics, or examine neighboring regions sharing similar properties to perform image classification and masking (Liu et al., 2008; Shi et al., 2014; Micusik and Hanbury, 2006).

Referring to blur, current focus-stacking algorithms use local values and gradients around each pixel to reconstruct the final picture. Numerous software solutions exist to perform focus-stacking. The most used methods are Difference of Gaussian (DoG) at various scales (also known as Laplacian Pyramid), preserving details however sensitive to noise, and depth maps (worse on details but keep original smoothness and colors) (Burt and Adelson, 1983; Zerene Stacker, 2009). Whereas commercial software solutions have their own ways to name and describe succinctly their algorithms, open-source projects tend to use Laplacian Pyramid as in ChimpStackr (Peeters, 2021), or promising methods based on Daubechies wavelets as in Focus-stack (Aimonen, 2022). On the side of photo editor software alternative to complete this task, in order to digitize small archaeological artifacts for the MUSINT II project, masks were generated on blurred areas with a batch in a commercial software (Marziali and Dionisio, 2017).

Optical blur is often considered to be well approximated by a convolution with a Gaussian kernel applied to a sharp image (Zhuo and Sim, 2011), which led to the Laplacian Pyramid (LP) method. This one deteriorates iteratively the original image (creating Gaussian Pyramid), applies convolution and subtracts pixel values obtained from the previous less-altered image (DoG) (generating Laplacian images, stored in LP). The image size is then downscaled by two horizontally and vertically, and the process is repeated with the new downsized image. Finally, in the reverse order compared to Gaussian Pyramid generation, each LP’s image is upscaled and merged with the following one until reaching original image size. In practice, when convolving with a Gaussian kernel, recomputing pixel values as a weighted average flattens the details. In an already defocused area, the convolution does not significantly modify each pixel value, resulting to almost black pixels on Laplacian difference images. Conversely, in an in-focus area the smoothness generated by the convolution leads to high values on Laplacian images while performing DoG. Thereby, this process emphasises the sharp parts at various scales. We chose to start our pipeline with this method (see §3.3).

1.2 Masking in SfM

The interest in masking parts of photographs is not limited to blur, since the surrounding environment might be undesired in the final reconstruction, introducing errors on tie points due to moving elements (Sieberth et al., 2014), or simply increasing computational time by forcing the detection of feature points (also known as key points) and their matching on the whole images, until finally computing unnecessary bundle triangulation. Even though some SfM software provides tools to draw masks on each picture, this tedious task tends to be replaced by automatic methods. Different attempts are made to create masks on images using camera orientations but they first require alignment (i.e. computing interior and exterior orientations) on unmasked images (Sergeeva and Sablina, 2018; Pan, 2019; Capturing Reality, Epic Games, 2025). Skipping this step saves time and avoids a second-pass alignment with computed masks to improve the accuracy of orientations. Hence, we will focus below on mask generation before any alignment process.

The mentioned methods depend on the scale of the scene to acquire, since background may be from heterogeneous to fully controlled. Applied to buildings and monument scales, deep learning networks for semantic segmentation have been described with interesting results (Stathopoulou and Remondino, 2019; Murtiyoso and Grussenmeyer, 2022). At the other end of the spectrum, movable artifacts placed on a turntable can benefit from convenient backdrops or photo light boxes. In this highly ideal case, masking might even be almost irrelevant as long as no feature point is detected. Nevertheless, it can still be a good idea to mask homogeneous background. Eastwood and al. (2020) focused on emphasizing contours in their study, then selected the one encompassing the maximum area to extract the whole object from background. Although their method does not eliminate blur, they evaluated and demonstrated the accuracy benefits of masking the background.

Between these situations lies the range of application cases we attempt to handle. It’s noteworthy that at very short range, masking the background can be more a beneficial consequence of the processing chain than a main concern as long as it belongs to the defocused area.

1.3 Very short range photogrammetry requirements

We define very short range photogrammetry as SfM processes exerted on objects ranging from a few centimeters to about a meter wide, captured using lenses (macro or non-macro) focusing near the minimum focusing distance (MFD) available for each lens, thereby maximizing magnification.

When confronted with small objects, at very close range, the DoF is drastically reduced, leading to items impossible to acquire with full sharpness. This issue is well known in macro photography as well as in microscopic photography, and induced focus-stacking (also named extended depth-of-field) methods to merge sharp parts from a stack of pictures, sharing the same optical axis but with slightly different focusing (McHugh, 2005) . When close to the subject, the narrowing DoF brings a new difficulty for the operator to ensure that the regions of interest are within the sharp focus range. Although it might be tempting to switch the lens to ”auto-focus” mode (AF), a slight change of focus at a close distance has a greater consequence on principal distance (i.e. focal length) than at a more distant one. In this unwise case, the operator has to anticipate the struggle for SfM algorithms and the ensuing effects in terms of accuracy.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-M-2-2025-89-2025 | © Author(s) 2025. CC BY 4.0 License.

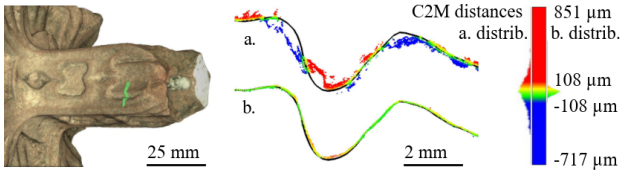


Figure 3. Effect of blur on geometry reconstruction. Distances and their distribution from point clouds to ground truth reference mesh (lines in black). On the left in green color, location of the 50 µm-slices on the ornament. On the right the two comparative slices. The upper dense cloud (a.) was intentionally acquired using slightly blurred photographs whereas the one below (b.) was captured from in-focus photographs.

In SfM, the accuracy of the reconstruction depends on the Ground Sample Distance (GSD) and the sharpness of the pictures, which relies on the lens quality, along with the adjustments made by the photographer. Narrow DoF issue has been previously mentioned, and it’s noticeable that the loss of information in all-blurry images produces inaccurate results with smoothed depth gradients (figure 3), and provides an extra issue to manage in reconstructed model’s texture (Pan, 2019; Kontogianni et al., 2017).

1.4 Focus-stacking methods and limits

With focus-stacking processing chains, SfM algorithms are processed on artificially reconstructed pictures, pseudo-photographs whose interior parameters more or less stable. Regardless of the focus-stacking methods or software used, multiple biases are introduced throughout their process. Optional colorimetric modifications of the original pictures as a pre-process aside, within a stack of photographs, they have to be scaled due to the requirement of backward and forward lens movement to change the focus (inside lens focus or thanks to external motorized micrometer rail), and aligned due to potential misalignment of optical axes depending on the camera setup stability. Since tributary to principal distance, the perspective view would change depending on the method to capture the in-focus pictures and in any case, the field of view captured from a photograph to the next one within a stack will change[1]. Assuming the user’s chosen algorithm works as intended without introducing undesired artifacts such as halos, before handling blur itself, the software still has to recalculate pixel values due to the mandatory rescaling mentioned above. The concept of GSD in photogrammetry requires careful consideration with focus-stacking because of the variable GSD between the first and the last pictures of a stack and because of color interpolation as well. Not to mention that the alignment methods and cropping of the final image (to fit the original image size) vary from one software to another, which can lead to a lack of trust in the capability to recreate the same interior parameters from one stack of pictures to another. Furthermore, any change in the optical axis within a stack has to be avoided to prevent non-reproducible cropping (by focus-stacking alignment and resizing algorithm) from a stack to another. It is reasonable to propose the informed hypothesis that a non-reproducible image cropping would lead to, at least, erroneous calculation of tangential distortions. Thus, stability of the camera setup and tripod is mandatory.

[1] In this study, with parameters described in §3.4, we estimated a change of the field-of-view of 4.3‰ horizontally and vertically from a picture to the following one.

Despite these aspects, SfM software does not seem to struggle that much, since particularly convincing results in terms of geometric accuracy have been shown (Clini et al., 2016; Olkowicz et al., 2019). This method is particularly appreciated in archaeological small documentation and numerous studies have been published refining and assessing processing chains and software (Brecko et al., 2014; Gallo et al., 2014). As mentioned by Marziali and Marziali (2019), among focus-stacking blur processing algorithms, LP tends to increase local contrast, being an appreciable choice for features detection in SfM (see §1.1).

Since most of the current SfM algorithms assume the principal distance to be constant, it is mandatory to establish a reproducible procedure for focus-stacking acquisitions. Specifically, the first and last picture of each sequence need to have exactly the same focus from one stack to another. In practice, keeping the camera fixed on a tripod ensures a stable distance to the artifact, placed on a turntable. Nevertheless, when resetting camera for every higher angle shot, the operator must cautiously evaluate and check the first and last DoF to cover the whole object in its most unfavourable position (its major axis collinear with optical axis). Small artifacts, especially with no prominent shapes, when well centered on a turntable are perfectly adapted to these constraints. For complex artifacts, increasing the number of shots in a stack can maintain a safety margin in depth of field before and below the item, though this significantly raises the final image count for processing.

Although suitable for digitizing movable small artifacts using an automated process, acquisitions are particularly time-consuming and this method suffers from an inherent lack of flexibility (Plisson and Zotkina, 2015). Setting up such acquisitions may be complicated in the field, outside, or on scaffolding.

2. Studied artifact and applicability of the proposed method

The proposed method attempts to cover the spectrum of challenging scenarios, including: unmovable objects (e.g. on buildings or monuments), hazardous or hard-to-access environments (e.g. on scaffolding or cherry pickers), time-constrained acquisitions, and setups limited to portable and space-saving equipment.

For the purposes of this study, we selected a fragment of Strasbourg’s Cathedral to be acquired using multiple methods, including the one to generate the reference model (point cloud and mesh), permitting comparative evaluation of methodologies and results. The artifact is a sculpted ornament in red sandstone in the Flamboyant Gothic style (figure 4). It presents strong similarities with Saint-Lawrence portal’s ornaments crowning lateral statues. It ideally represents the type of sculpted pinnacles and carved tympanums to capture in-situ while heritage documentation.

3. Description of the experiment

We will describe the proposed method using a ring flash and the focus-stacking pipeline separately. Since this study focuses on assessing volumetric accuracy for complex-geometry artifacts, we will only mention colorimetric issues caused by shadows. Processes involving color charts to correct color values have already been described (Molada-Tebar et al., 2019). They can be applied before or after SfM alignment to compute reliable textures. In the first case, they would join the interior and exterior orientation determination and must be knowingly used.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-M-2-2025-89-2025 | © Author(s) 2025. CC BY 4.0 License.

Figure 4. a) Photograph of the sculpted artifact; b) Saint-Lawrence portal, Strasbourg’s Cathedral, France © Claude Truong-Ngoc 2020; c) Typical case for very close range photogrammetry on scaffolding, Heritage restoration site of Saint-Lawrence portal’s. Photo courtesy and © G. Calmels 2025.

3.1 Photographic equipment

The camera used for the acquisitions is a Canon EOS R5 with a Canon mount adapter EF-EOS R and a Canon EF 28 mm f/2.8 IS USM lens, chosen for its quality across the entire image, its MFD (23 cm) and its wide angle-of-view (AoV) capturing side elements at minimal distance. With a pixel-image size of 4.39 µm we obtained a GSD of 36 µm at MFD.

We used a Sigma EM-140 DG ring flash in ”manual mode” at 1/64 intensity on both tubes. Having the optical axis surrounded by light, using a ring flash allows to have very few shadows except when hollows are too deep. Despite its category among macro flashes and as many other ring types, this one is not stricto sensu a ring but either two lateral tubes with light diffusers. This aspect sometimes led to 90° rotation of the flash device depending on reliefs, adjusted while shooting the artifact to minimize shadows as much as possible.

Focus-stacking acquisitions were performed without flash but with four stable LED-lights and the camera fixed on tripods. The LED panels were repositioned for every higher-angle shot to ensure the most diffuse lighting and shadow-free images from the camera’s perspective.

The artifact was imaged in two passes (placed as in figure 5), and flipped during the process.

Figure 5. Equipment for both photographic acquisitions. On the left, with ring flash for on-the-fly shots. On the right, focus-stacking-ready setup.

3.2 Acquisitions on the fly with ring flash

In light of the foregoing, the compromise made for aperture was f/16 (in a range of f/2.8-22). Camera was switched to ”manual exposure” mode and lens to ”manual focus” at MFD. Sensor and lens stabilization were switched off. The use of a flash allowed us to work at ISO 100 sensitivity at 1/160 s. This high speed permits to work on-the-fly (without tripod) in thorny positions, being able to move perfectly freely around the artifact, notably forward and backward to adjust the regions in-focus.

Since the aim of this acquisition was to mimic in-situ documentation, rearranging the table was excluded, forcing us to move around the artifact. The vari-angle LCD monitor was particularly appropriated to capture the object from difficult angles.

Undeniably, this capture technique is dynamic and requires undivided attention. As in traditional photogrammetry for larger and more complex scenes, it is necessary to evaluate the best strategy to adopt to fully cover the artifact. The additional challenge for the operator lies in understanding that the entire field of view will not be part of the computation. Figure 6 (a.), with DoF in green, proposes a flexible approach to maximize sharp areas. Planes of focus are estimated to be generally slightly inside the object to also maximize the use of the in-focus space ahead of it. Figure 6 (b.), with DoF in red, shows a too rigid approach, with optical axes far too tangent to the in-focus surface, resulting in sharp overlapping regions that are not wide enough to ensure proper alignment during SfM process. The blurred areas are intended to be masked, so the overlap must be thought dynamically relative to sharp areas already shot to avoid holes. Basic strategies are not very complicated to assimilate, as long as the concept of DoF is well understood. It’s also easy to multiply shots with a slightly different perspective or distance, especially for depressions, notches and prominent areas. The key is to change scale and consider the acquisition as the one for a building or a monument. For these acquisitions, we took 279 pictures in 47 minutes on one side, then 292 photographs in 40 minutes on the other side.

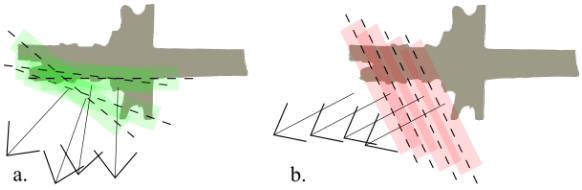


Figure 6. Diagrams of two different acquisition strategies on a slice of the artifact. A flexible approach (a.), with DoF in green. A too rigid one (b.), with DoF in red. Planes of focus are symbolized by dashed lines.

3.3 Mask processing

We employed the LP method as the first step of our pipeline to generate masks on blurred areas because it emphasizes details, but also led by the desire of simplicity and reproducibility. Indeed, for this LP step of the mask pipeline, the operator has only one parameter to choose: the number of iterations to perform. Remembering the desired masks have to be thresholded to be binary at the end, few iterations would be more restrictive, while many of them would be less relevant to segregate what is blur or not. Then morphological operations (dilatation then erosion) are applied to white zones. Figure 7 shows the steps on two different samples pulled out from the same picture, with increased contrast on (a.) and (b.) to enhance visibility in this article. For this figure, on (c.), (d.), (e.) binary masks are layered on original image to provide a clearer representation of each step. In the proposed pipeline, the threshold value and the radius for white areas operations could also be changed by the operator if necessary before computing masks with the same parameters on the whole folder.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-M-2-2025-89-2025 | © Author(s) 2025. CC BY 4.0 License.

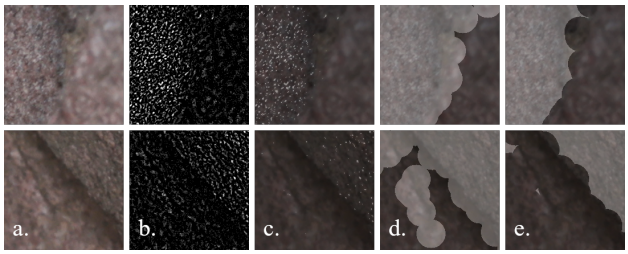


Figure 7. The customizable steps in mask generation from original image (a.), via LP image (b.), LP thresholded (c.), dilated white areas (d.) to eroded white areas (e.).

We developed an interface in Python using OpenCV and Qt to display in real-time the effect of our parameters’ choices. Testing with our pictures and other sets captured with the same camera, we found pertinent to maintain a low level for LP (generally one level), apply a threshold of 5 (in a range from 0 to 255), dilate white areas by 51 pixels then erode them by 52 pixels. Eroding a bit more than dilating (1 pixel in our test cases was sufficient) allows to throw out tiny isolated pixels irrelevant for SfM alignment or bundle triangulation. Figure 8 shows some examples. With those parameters, mask generation took 2 seconds per picture, ending this process in respectively 10 minutes and 11 minutes for the 279 and 292 pictures. We noted the morphological operations could drastically increase the processing time for high radius values. If struggling to generate masks on blurred areas with the aforesaid method, we would advise attempts on LP levels and threshold value before boosting dilate and erode radius.

Figure 8. Example of photographs and associated masks (enhanced contrast for this view).

3.4 Focus-stacking acquisition

Following previous studies and due to complex geometry (particularly in crown section of the artifact), we decided to capture it by rotating the turntable every 15 degrees around the vertical axis and shot vertically from 4 stances 10, 22, 40 and 60 degrees from horizontal plane.

The focus-staking acquisition was controlled by the microsoftware embedded in camera, in order to have the setup as portable as possible, consequently independent from any computer during the acquisition. The function called Focus Bracketing allows to choose the number of shots (in a range of 2-999) and the focus increment (in a range of 1-10). The latter was set to 1, implying the smallest steps from one shot to another. As expressed by Canon R5 Manual, the shift between shots is automatically adjusted to suit the aperture value. ”Exposure Smoothing” was switched off to avoid adding an extra image processing bias.

Lens had to be put in AF mode to enable the necessary motorized shift, however the AF from shutter button was switched off to prevent undesired variations in focal length. ”Drive Mode” was put to ”Shoot in 10 sec” to prevent instability while triggering the shutter button.

Test stacks were made to determine the correct number of shots necessary to capture the scene, including a safety range ahead and behind the object to anticipate the worst placements while rotating (see §1.4). We decided to take 20 pictures per stack at aperture f/11, ISO 100 sensitivity, 1/10 s, focusing from MFD (23 cm) to 60 cm, placing the center of the artifact at about 35 cm from the camera. The mean GSD was shifted to 55 µm in a range of 36-94 µm.

We noticed the object was small enough to be entirely contained in the field of view of every picture. Had this not been the case, the protocol with a turntable would have been rethought. Furthermore, shooting at further distances with a shifted focus range would introduce variations in principal distances difficult to rigorously reproduce throughout the entire acquisition. To solve this issue, it can be considered to systematically start with the first shot at MFD, take enough shots to cover the range of interest, then automatically discard with a script the first n-pictures of every stack. The other solution is to remote-control the camera with a computer, but this would not fit the requirement of computer-free acquisitions. For these ones, we obtained 3880 pictures (1940 for each side), flipping midterm the artifact. They took respectively 1h11 and 1h15.

3.5 Focus-stacking computation

We attempted to use various software solutions including open-source ones. Although ChimpStackr (using LP) gave us results qualitatively[2] as good as Zerene Stacker (Pmax computation), it needed mandatory pre-aligned photographs. In our tests, Focus-stack open-source software produced intermediate images suited for ChimpStackr. However, each final reconstructed result needed to be cropped, which could have been easily done with fixed margins in Python.

Given the lack of batch to run these time-consuming operations using multiple open-source software, emphasized by the will to assess our results to previous studies, we chose Zerene Stacker, using Pmax computation, which appears analogous to LP method and suitable for SfM applications (Gallo et al., 2014). Preceded by a Python script to move images in folders corresponding to each 20-picture stack, we obtained 194 reconstructed images (97 on each side) computed in 4h22.

3.6 SfM processing

SfM workflows were executed for each side of the artifact separately. Metashape was used to process ring flash acquisitions (with masks and without it) and focus-stacked ones (Agisoft, 2025). Every single picture was qualitatively well aligned[3] in each case, with no obvious mistake.

After alignment step, we checked the correct determination of the centers for the 4 markers used as stable points for 4 scalebars, then imported in-situ measurements. The aligned sparse cloud was scaled accordingly (”Update” function in Metashape) without ”Optimizing cameras”, since marker detection errors by our script would have been integrated, potentially deforming the geometry.

[2] Differences between Zerene Stacker (Pmax) and ChimpStackr (LP) were visually imperceptible. However, at bigger scale, slight translations of the center of the recombined images exist as well as different color values distribution visualized locally on samples. The specific influence of these effects on focus-stacked images in SfM processes wasn’t assessed in this study.

[3] Metashape alignment parameters: Accuracy: high, Key point limit: 50 k, Tie point limit: 20 k, Masks applied to: key points (irrelevant for mask-free case). Because of the turntable, ”Exclude stationary tie points” was only turned on for focus-stacked case.

Finally, dense clouds were generated in high quality with mild depth filtering enabled.

Attempts were made with RealityCapture, switching on ”Group Calibration by Exif” parameter to limit unsuitable computations and biases (on purpose, assuming the interior orientation of our acquisitions was constant). With ring-flash photographs, although alignment computation[4] worked without omitting pictures in both cases, only the set with masks managed to reconstruct a model, while tries without masks ended systematically by computer crash.

With focus-stacked images, approximately 90 (out of 97) were aligned. Changing alignment parameters or ungrouping calibrations (as if acquisitions were shot in AF mode) did not improve this score. We did not proceed further, deeming these results insufficient.

The only model fully computed using RealityCapture was the ring flash one with masks, on both sides, built in high quality with no downscale factor for depth maps.

3.7 Dense clouds edition and registration

Dense clouds were edited in CloudCompare to crop out sections touching the table (CloudCompare, 2025). Remaining dense clouds were manually pre-aligned, then finely registered using Iterative Closest Point (ICP) algorithm, considering arbitrary one of the point clouds as reference. Various attempts were made modifying ICP’s final overlap parameter: 50%, 70% and 90%. Distances with M3C2 plugin were calculated[5], to choose the best compromise between mean error and standard deviation (STD) (Lague et al., 2013). For Metashape’s dense clouds we chose a final overlap of 90% (Root Mean Square (RMS) difference 10^{-7}, without scaling), and 70% for RealityCapture dense cloud.

Since histograms of distances distributions closely mimic a Gaussian model without fitting it rigorously, we compared only recombined point clouds to the metrology laser scanner model.

Before merging and subsampling, parts of the point clouds which were facing the table (too dark and already captured by the corresponding counterpart) were cropped out. Spatial subsampling was performed relative to the finest GSD available across the entire acquisitions (see §5). We obtained 2 point clouds for each case, keeping 1 point every 36 or 72 µm (respectively 1 and 2 GSD).

4. Reference model

The reference model was acquired in two times with the Faro Edge ScanArm HD, which claims an accuracy of 25 µm with a repeatability of 25 µm (2 σ). Scans were made in ”high density”, ”high precision” mode, with ”remove overlap” and ”reduce noise” options. We chose to keep and export raw point clouds.

[4] RealityCapture key settings: Max features per image: 100 k, Image overlap: high, Max feature reprojection error: 1, Background feature detection: no, Preselector features: 20 k, Detector sensitivity: ultra, Distortion model: Brown3 with tangential2.

[5] M3C2 key parameters: Cloud #2: Reference, Normal: 200 µm, Projection: 100 µm, Max depth: 500 µm, Normals prefered orientation from barycenter.

Scan point clouds were registered and merged using the same procedure as detailed in §3.7. Final scan cloud was subsampled compared to announced scanner accuracy of 25 µm. A mesh was produced using PoissonRecon in CloudCompare with an octree depth of 10. This meshing was particularly suited for this model, since there was no hole in the scan cloud. Slices along different axes were made to qualitatively assess and validate this mesh as the reference. We also computed Cloud to Mesh (C2M) distances between the scan cloud and the mesh (see §5).

5. Results analysis

We propose evaluating the results using both conventional absolute distance metrics and GSD-relative measurements, the latter notion providing an intuitive accuracy assessment by combining distance to photographic equipment.

Each recombined photogrammetric dense cloud was registered with the reference scan cloud using ICP algorithm, with ”scaling” option enabled[6]. A final overlap of 50% tended to show the lowest statistical dispersion, regardless of the case considered. Gaussian statistical metrics were excluded from these results, because the distances distributions mimic but did not fit rigorously a Gaussian model. Indeed, to force such a fit would require discarding what may be considered as outliers, yet they are precisely one of the elements we aimed to evaluate, as it evokes at least noisy point clouds.

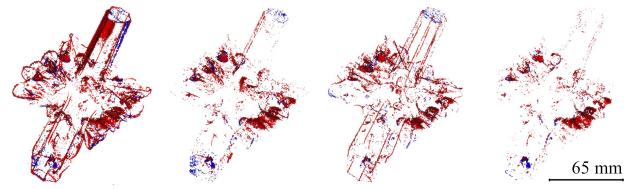
We determined our minimum GSD of 36 µm as our standard of measurements, even for focus-stacking point cloud to ensure comparable results (see §1.4). C2M distances were computed between point clouds and reference mesh, then point clouds were split twice to quantify the proportion exceeding 2 GSD (72 µm) and 3 GSD (108 µm) (table 1). The case processed without masks appears to have much more dispersion than any other. With-masks processed point clouds show quantitatively comparable deviations to those processed using focus-stacking.

	sub. 36 µm		sub. 72 µm	
	2 GSD	3 GSD	2 GSD	3 GSD
MS -Mask	11 %	4 %	12 %	5 %
MS +Masks	6 %	1 %	6 %	1 %
MS FS	6 %	1 %	7 %	2 %
RC +Masks	5 %	1 %	6 %	1 %

Table 1. Proportion of points exceeding tolerance of 2 and 3 GSD (respectively 72 and 108 µm) assessed on subsampled point clouds (36 and 72 µm). MS: Metashape; RC: RealityCapture; ”-Mask”: ring flash without masks; ”+Masks”: ring flash with masks; FS: focus-stacking.

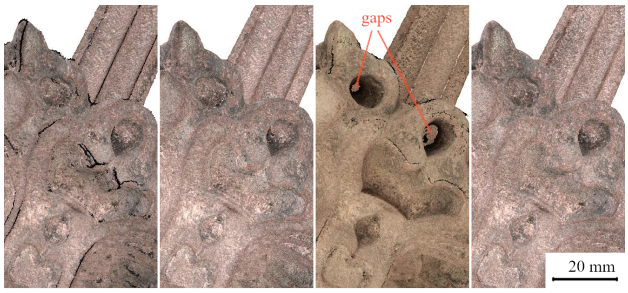
Qualitative comparisons of off-tolerance parts show particularly noisy and protruding dark edges in without-masks case, thin salient dark edges in focus-stacking case and struggles inside crown hollows in every case (figures 9, 10). Despite these differences in deep depressions, the reconstructed point clouds from the proposed method did not show any gap. On the contrary, focus-stacking point cloud suffered from insufficient overlaps in these hollows and did not manage to reconstruct them entirely. Acquisition protocol of such complex-geometry

[6] Scaling was reproducible among our point clouds with mean rescaling = 0.99777 and σ = 0.00029. This rescaling can be plausibly attributed to in-situ accuracy measurements (Mean absolute difference with in-situ measurements: 0.3 mm).



a. MS without mask b. MS with masks c. MS focus-stacking d. RC with masks

Figure 9. Distances to the reference mesh exceeding the 3 GSD tolerance (displayed on point clouds extracted from 36 µm subsampled point clouds). In red, dist. $>$ 108 µm; In blue, dist. $<$ -108 µm.



a. MS without mask b. MS with masks c. MS focus-stacking d. RC with masks

Figure 10. Samples from dense clouds showing the edges issues (a. and c.) and projected shadows (c.). For this view, Eye-dome Lighting (EDL) was disabled, flattening geometry on purpose in order to see undesired dark parts.

artifact using focus-stacking should be less systematized and more dynamic, with at least many more pictures taken from slightly different angles.

About shadows, the without-masks cloud had mixed values, with some darker parts standing beside bright ones. This phenomenon was anticipated, since defocused areas in photographs taken with flash are often darker regions. Focus-stacking cloud was more concerned about projected shadows, pointing out the difficulties to maintain a perfect diffuse lighting on every part of the object with complex geometry. Figure 10 (c.) shows how values influence point cloud colors, speaking undesirably for relief. In our case study, with-masks cases had the best rendering.

About colorimetry, since their color scale and white balance required harmonization due to the different lighting sources, we did not evaluate this aspect.

Processes *(H) human guided*	Proposed method		Focus-stacking
	MS	RC	MS
Setting-up environment *(H)*	0h05		0h15
Photographic acquisition *(H)*	1h27		2h26
Pre-SfM process *(H)*	0h10		0h10
Pre-SfM process	0h21		4h22
SfM process alignment	3h20	0h15	0h25
SfM process dense cloud	4h06	3h56	3h01
Editing dense cloud *(H)*	1h00		
ICP registration	0h01		
Total human time	2h42		3h51
Total machine time	7h48	4h33	7h49
Total processing time	10h30	7h15	11h40

Table 2. Comparison of processing times for SfM using Metashape (MS) or RealityCapture (RC).

Due to the shadows, edge effects, undesired noise (laborious to clean) and proportion of points off-tolerance, the without-masks cloud did not meet the expected specifications. We excluded it from the comparison of processing times (table 2). The interpretation of these times should also consider the incomplete reconstruction of hollows in the focus-stacking point cloud.

Finally, by computing C2M distances between the reference scan cloud and the reference mesh, we observed the spatial dispersion. The distances distribution was Gaussian with a mean distance of 1 µm and STD of 47 µm, which could fit the order of magnitude of this sandstone grain, or be due to the technical limit of our scanner. We could not clarify this point, reaching the limits of our available equipment.

6. Conclusions

In our case study, the SfM point clouds generated through adequate acquisitions, supported by rigorous methodological principles, demonstrated strong potential as a cost-effective alternative to more expensive digitization techniques.

Despite the reduced depth-of-field at very short range, using SfM to digitize small complex-geometry artifacts offers significant practical, technical and economic advantages. The proposed method using a ring flash and a fixed focusing distance (MF), as for traditional SfM bigger size acquisitions, followed by applying masks on blurred areas, appeared as accurate as focus-stacking based pipelines already described in the literature.

Equipment’s portability makes the proposed solution with masks perfectly suitable for complex-geometry artifacts, unmovable ones or ones that are slightly too large to be captured with focus-stacking setup (classically dependent on a turntable). Compared to focus-stacking methods, the proposed one requires less time to proceed the acquisition on site, is less systematic, more demanding and dynamic for the photographer.

This mask generation pipeline, starting by LP, demonstrated both flexibility and speed. One might argue that still in-focus but soft, low-contrast surfaces could be undesirably masked. However, this concern is inherent to SfM workflows, where such areas are typically excluded during alignment and reconstruction. The present study demonstrates the importance of blur masking in photogrammetric reconstruction workflows.

Since this study focused specifically on geometric accuracy, with our lighting setup, the proposed method tended to show a better quality around point clouds’ edges compared to focus-stacking method. Further work will address this aspect in detail.

Acknowledgments

We thank staff members from the Photogrammetry & Geomatics Group and students from INSA Strasbourg for their support.

References

Agisoft, 2025. Metashape. www.agisoft.com (4 April 2025).

Aimonen, P., 2022. Fast and easy focus stacking. github.com/PetteriAimonen/focus-stack (4 April 2025).

Brecko, J., Mathys, A., Dekoninck, W., Leponce, M., Vandenspiegel, D., Semal, P., 2014. Focus stacking: Comparing commercial top-end set-ups with a semi-automatic low budget approach. A possible solution for mass digitization of type specimens. *ZooKeys*, 464, 1-23.

Burt, P., Adelson, E., 1983. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, 31(4), 532-540.

Capturing Reality, Epic Games, 2025. Depth map and mask export. rchelp.capturingreality.com/en-US/tools/depthandmask.htm (4 April 2025).

Clini, P., Frapiccini, N., Mengoni, M., Nespeca, R., Ruggeri, L., 2016. SfM technique and focus stacking for digital documentation of archaeological artifacts. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B5, 229–236.

CloudCompare, 2025. www.cloudcompare.org (4 April 2025).

Gallo, A., Bruno, F., Muzzupappa, M., Russa, M. F. L., 2012. Multi-view 3d reconstruction of small stone samples deteriorated by marine organisms. *2012 18th International Conference on Virtual Systems and Multimedia*, 181–187.

Gallo, A., Muzzupappa, M., Bruno, F., 2014. 3D reconstruction of small sized objects from a sequence of multi-focused images. *Journal of Cultural Heritage*, 15(2), 173-182.

Historic England, 2017. Photogrammetric applications for cultural heritage. Guidance for Good Practice. *Swindon. Historic England*.

Kontogianni, G., Chliverou, R., Koutsoudis, A., Pavlidis, G., Georgopoulos, A., 2017. Enhancing close-up image based 3d digitisation with focus stacking. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5, 421–425.

Lague, D., Brodu, N., Leroux, J., 2013. Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). *ISPRS Journal of Photogrammetry and Remote Sensing*, 82, 10-26.

Levin, A., Weiss, Y., Durand, F., Freeman, W. T., 2009. Understanding and evaluating blind deconvolution algorithms. *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 1964–1971.

Liu, R., Li, Z., J., J., 2008. Image partial blur detection and classification. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, 1–8.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2014a. *Close-Range Photogrammetry and 3D Imaging*. De Gruyter, Berlin, Boston, 151–152, 191, 339.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2014b. *Close-Range Photogrammetry and 3D Imaging*. De Gruyter, Berlin, Boston, 116.

Marziali, S., Dionisio, G., 2017. Photogrammetry and macro photography. The experience of the MUSINT II Project in the 3D digitizing process of small size archaeological artifacts. *Studies in Digital Heritage*, 1, 298.

McHugh, S., 2005. Focus stacking & depth of field. www.cambridgeincolour.com/tutorials/focus-stacking.htm (11 June 2025).

Micusik, B., Hanbury, A., 2006. Automatic image segmentation by positioning a seed. *Computer Vision – ECCV 2006, 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part II*, 3952, 468–480.

Molada-Tebar, A., Marqués-Mateu, A., Lerma, J. L., 2019. Correct use of color for Cultural Heritage documentation. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-2/W6, 107–113.

Murtiyoso, A., Grussenmeyer, P., 2022. Automatic point cloud noise masking in close range photogrammetry for buildings using AI-based semantic labelling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVI-2/W1-2022, 389–393.

Olkowicz, M., Dabrowski, M., Pluymakers, A., 2019. Focus stacking photogrammetry for micro-scale roughness reconstruction: a methodological study. *The Photogrammetric Record*, 34, 11-35.

Pan, R., 2019. Defocus blur estimation in calibrated multi-view images for 3D archaeological documentation. *Digital Applications in Archaeology and Cultural Heritage*, 14, e00109.

Peeters, N., 2021. Free, multi-platform focus stacking software. github.com/noah-peeters/ChimpStackr (4 April 2025).

Plisson, H., Zotkina, L. V., 2015. From 2D to 3D at macro- and microscopic scale in rock art studies. *Digital Applications in Archaeology and Cultural Heritage*, 2(2), 102-119. Digital imaging techniques for the study of prehistoric rock art.

Sergeeva, A. D., Sablina, V. A., 2018. Using structure from motion for monument 3D reconstruction from images with heterogeneous background. *2018 7th Mediterranean Conference on Embedded Computing (MECO)*, 1-4.

Shi, J., Xu, L., Jia, J., 2014. Discriminative blur detection features. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2965–2972.

Sieberth, T., Wackrow, R., Chandler, J., 2014. Motion blur disturbs – the influence of motion-blurred images in photogrammetry. *The Photogrammetric Record*, 29.

Stathopoulou, E.-K., Remondino, F., 2019. Semantic photogrammetry – Boosting image-based 3d reconstruction with semantic labeling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W9, 685–690.

Xu, L., Jia, J., 2010. Two-Phase Kernel Estimation for Robust Motion Deblurring. *Proceedings of European Conference on Computer Vision*, 6311, 157-170.

Zerene Stacker, 2009. How To Use Zerene Stacker. www.zerenesystems.com/cms/stacker/docs/howtouseit (4 April 2025).

Zhuo, S., Sim, T., 2011. Defocus map estimation from a single image. *Pattern Recognition*, 44(9), 1852-1858. Computer Analysis of Images and Patterns.