HANDLING CRITICAL ASPECTS IN MASSIVE PHOTOGRAMMETRIC DIGITIZATION OF MUSEUM ASSETS

E.M. Farella, L. Morelli, E. Grilli, S. Rigon, F. Remondino

3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy
Web: http://3dom.fbk.eu – Email: <elifarella><lmorelli><grilli><srigon>< remondino>@fbk.eu

Commission II

KEYWORDS: massive museum digitization, photogrammetry, depth of field, masking, filtering, cleaning

ABSTRACT:
In recent years, a growing interest in the 3D digitisation of museum assets has been pushed by the evident advantages of digital copies in supporting and advancing the knowledge, preservation and promotion of historical artefacts. Realising photo-realistic and precise digital twins of medium and small-sized movable objects implies several operations, still hiring open research problems and hampering the complete automation and derivation of satisfactory results while limiting processing time. The work examines some recurrent issues and potential solutions, summing up several experiences of photogrammetric-based massive digitisation projects. In particular, the article presents some insights into three crucial aspects of the photogrammetric pipeline. The first experiments tackle the Depth of Field (DoF) problem, especially when digitising small artefacts with macro-lenses. On the processing side, two decisive and time-consuming tasks are instead investigated: background masking and point cloud editing, exploring and proposing automatic solutions for speeding up the reconstruction process.

1. INTRODUCTION

3D digitisation of Cultural Heritage (CH) assets is a widespread practice supporting the knowledge, conservation, and promotion of sites and artefacts of cultural and historical value. The recent pandemic period has highlighted and increased the role and importance of digital technologies in this field as tools to overcome physical inaccessibility and experience heritage access differently (Tausch et al. 2020, Raimo et al., 2021). Among CH settings, museum collections are immense, fascinating and fragile treasures. With a view to innovating traditional museum exhibitions and increasing the attractiveness of the preserved collections, the demand for artefact 3D digitisation is constantly growing. However, the 3D reproduction ("digital twin") of museum assets is generally a hard-working task, since:

- virtualisation projects typically include vast collections;
- lighting conditions, artefact material and shape, available spaces, among others, constrain the data acquisition phase;
- artefacts differ in size and materials (reflective, textureless, thin, etc.), and digitisation equipment should be flexible enough to tackle data acquisition in various situations;
- reasonable times must be planned for processing and delivering 3D results.

This work sums up experiences and lessons learnt from massive photogrammetric surveying of museum artefacts to realise digital libraries, virtual interactive exhibitions, and AR/VR applications.

In the image-based 3D reconstruction process, many factors can affect the quality of the produced 3D results. While image quality, sharpness, and camera network define the achievable accuracy and completeness of the model, image orientation and dense image matching quality impact the level of noise of the dense reconstruction and the efforts required to generate a photo-realistic and precise 3D model.

The work investigates three key aspects of the acquisition and processing pipeline through experiments and analyses, with the aim to reduce operational time while guaranteeing adequate and satisfactory results: depth of field (DoF – Section 3), background masking (Section 4) and point cloud cleaning (Section 5). Figure 1 shows the general photogrammetric pipeline, where acquisition and processing phases are paired to achieve high-quality results, both in terms of geometry and texture.

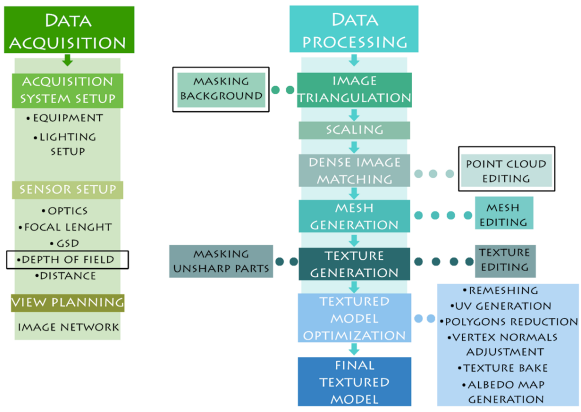

Figure 1. A schematic photogrammetric pipeline (acquisition and processing) for the production of textured 3D models. The steps discussed in this paper are shown in brackets.

2. RELATED WORKS

When massively digitising museums' artefacts for conservation, preservation, restoration, visualisation and dissemination purposes, three main requirements are generally to be satisfied:

- a faithful, complete and precise reconstruction of the object's shape and geometry, limiting occlusions and avoiding loss of information;
- a high-resolution and truthful texture for optimised (low-poly) geometries, to virtually inspect in web-based and/or augmented/virtual reality applications (AR/VR);
- a limited acquisition and processing time per object to perform massive digitisation activities.

Both range and image-based techniques (Cignoni and Scopigno, 2008; Remondino et al., 2013; Gonizzi-Barsanti et al., 2014; Guidi et al., 2015; Russo et al., 2015; Roncella et al., 2021) demonstrated to fulfil the first need, although passive methods are preferable for the second requirement (Gattet et

al., 2015; Menna et al., 2016; Menna et al., 2017; Collins et al., 2019; De Paolis et al., 2020; Rodríguez-Martín and Rodríguez-Gonzálvez, 2020; Apollonio et al., 2021; Roncella et al., 2021).

However, some phases of the photogrammetric workflow can be critical in massive 3D digitisation, and a few tricks can be helpful to limit digitisation times and improve the quality of the results.

One of the key acquisition aspects is to ensure proper sensor settings to achieve sufficient image quality and sharpness while meeting the planned 3D model's resolution. The DoF is a typical issue when acquiring medium and small-sized artefacts, especially with macro lenses. Effective but time-consuming solutions, like focus stacking or shape from focus (Niederöst et al., 2003), can solve this well-known problem in macro photography, but they are hardly applicable in massive digitisation. Further processing approaches rely on masking unsharp and out-of-focus areas by also resorting to automatic defocus estimating algorithms (Verhoeven, 2018), which present several limitations. However, DoF effects in 3D reconstruction is still a topic scarcely investigated (Menna et al., 2012; Sapirstein, 2018; Lastilla et al., 2019; Webb et al., 2020). Knowing the employed digital camera and its specifications (focal length, aperture, object distance), various smartphone apps and web-based tools are available to assess which camera settings are required to achieve a desired level of sharpness in the images.

A second crucial point for significantly reducing manual efforts in an image-based 3D digitisation pipeline is background masking. This operation is frequently necessary (i) when a turntable is used, and the background is static, (ii) for jointly processing images where the artefact has been flipped (e.g. front and back side) or (iii) to limit the area where Multi-View Stereo (MVS) algorithms are applied to decrease the computational time and unwanted 3D points. Removing backgrounds with manual or semi-manual procedures is a pretty easy task, and some commercial software offer tools for lightening the needed effort. Few operations are required to generate masks in these cases, although images must be edited one by one. Therefore, this process remains highly time-consuming, with thousands of pictures to mask. Since image masking is an image segmentation problem, deep learning methods and Convolutional Neural Networks (CNNs) were also explored (Long et al., 2015; Liu et al., 2020). The main issue of these fully automatic methods is the amount of input data required for training models and the availability of proper image datasets for handling the image classification and semantic segmentation tasks. Moreover, as artefacts are different from one another, segmentation models based on U-Net, Mask R-CNN, etc. (He et al., 2020; Knyaz et al., 2020; Grilli et al., 2021; Minaee et al., 2021) are not suitable.

Another key and demanding step in the reconstruction process is the dense point cloud cleaning. Noise and outlier removal from MVS point clouds is necessary to generate clean polygonal models. Point cloud filtering is a wide investigated field with many algorithms for tackling this process (Han et al., 2017). Traditional methods can be mainly categorised as statistical-based, neighbourhood-based, projection-based and PDEs-based, although further hybrid procedures were investigated. Recently, learning-based methods were also presented for point cloud denoising (Duan et al., 2019; Hermosilla et al., 2019; Erler et al., 2020; Luo and Hu, 2020; Rakotosaona et al. 2020; Luo and Hu, 2021). Although promising, these techniques are frequently sensitive to outliers and generally fail with a high level of noise in the data.

3. DEPTH OF FIELD (DOF)

3.1 The Depth of Field (DoF) problem

While assuring to meet the planned spatial resolution, an adequate DoF is crucial to prevent unsharp areas in the images, leading to noisy 3D reconstructions. The DoF defines the range of acceptable image sharpness around the plane of sharp focus, and it depends on the average object distance, the focal length c, the scale number S, the F-number, and the value of the Circle of Confusion (CoC) (Luhmann et al., 2019) (Eq.1):

$$DoF = \frac{2 * CoC * F_{number} * (1 + S)}{S^2 - (\frac{CoC * F_{number}}{c})^2} \qquad \text{Eq.1}$$

Following Menna et al. (2012), the CoC should not be set larger than the required resolution: it is the diameter of the blur spot measured on the sensor, calculated as the ratio GSD/S. Images must be "acceptably" sharp for the image-based pipeline, although the range of acceptability embodied in the circle of confusion has not been quantified yet (Verhoeven, 2018). In massive digitization, selecting adequate camera settings for avoiding or limiting unsharp areas can prevent further image pre-or post-processing efforts. The choice of the capturing parameters should balance the final image quality, the need to maximize the frame with the object view (as required by image-based modelling applications), and return an accurate product according to the planned GSD or required spatial resolution. Once acquisition distance and focal length are fixed to meet these requirements, the DoF problem can be controlled by selecting proper lens aperture parameters. Smaller aperture values return deeper DoF, but a limit to the aperture choice is imposed by blurring effects caused by diffractions, which affect and decrease the image quality.

3.2 Experiments

Two artefacts were considered: a small and complex statue portraying Moses (6x6x15 cm) and a flat and reflective 2-euro coin (25 mm diameter). They both simulate archaeological objects' shape, size and materials causing image acquisition problems when very high-resolution details are needed. For both objects, ground-truth 3D data acquired with a triangulation-based laser scanner, were available (spatial resolution of 0.015 mm and 0.01 mm, respectively). The performed tests focused on verifying the influence of lens aperture settings (i.e. DoF ranges) on the final 3D models' quality when:

- the acquisition distance is bound by the planned spatial resolution while maximising the imaging frame with the surveyed artefact;
- the image masking of unsharp areas is avoided for limiting the processing times.

In the experiments, a Nikon D750 (full frame CMOS sensor, 5.95 µm pixel size) coupled with a Sigma 105 macro f2.8 was employed in both cases. All images were captured at ISO 100, keeping the focus fixed (on the central part of the objects). The planned GSD was 0.05 mm and 0.025 mm, respectively. The same camera network and a consistent illumination (modifying shutter speed values) were kept during the acquisitions while changing the aperture settings (F5.6, F11, F16 and F22). As known, the higher the F-number, the larger is the DoF. At the same time, with high F-numbers, some diffraction effects can decrease the lens's resolving power (Verhoeven, 2018).

F-number	**F5.6**		**F11**		**F16**		**F22**	
DoF [mm]	**6.1**		**12.2**		**17.2**		**24.3**	
	Mean	*St.dev.*	*Mean*	*St.dev.*	*Mean*	*St.dev.*	*Mean*	*St.dev.*
C2C	0.1931	0.4157	0.1218	0.3104	0.0989	0.2561	**0.0908**	0.2498
C2Mesh	-0.0455	0.4626	-0.0316	0.3396	**-0.0251**	0.2826	-0.0255	0.2733

Table 1. Metric evaluation (Cloud-to-Cloud and Cloud-to-Mesh comparisons [mm]) of the dense reconstructions achieved with several aperture settings for the Moses statue (105 mm focal length, 500 mm object distance).

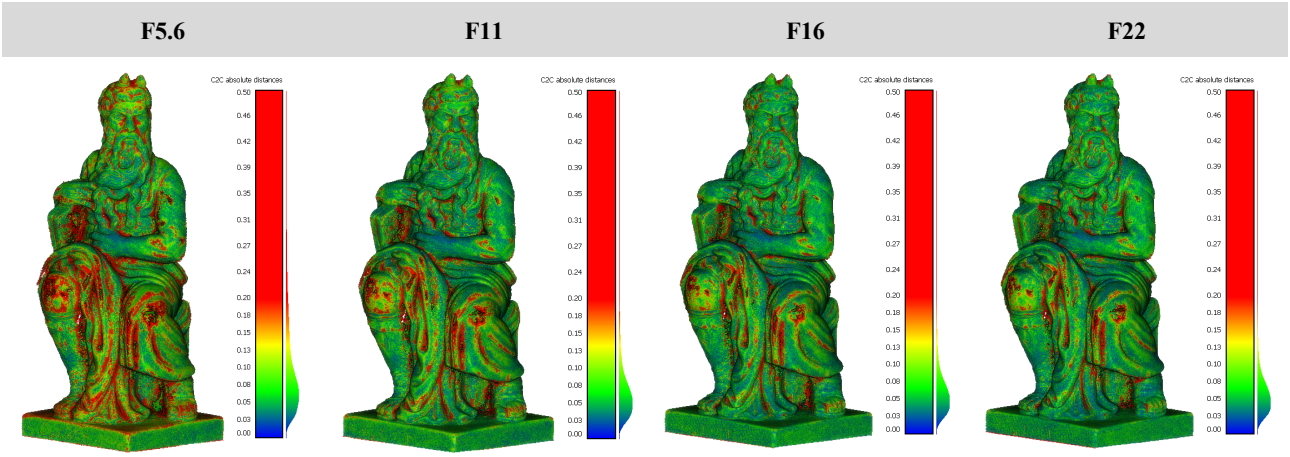

Figure 2. Cloud-to-Cloud comparison [mm] with F5.6, F11, F16 and F22 for the Moses statue. Range [0.00/0.50 mm]. 3D data derived from images acquired with a higher F-number present fewer discrepancies from the ground-truth data.

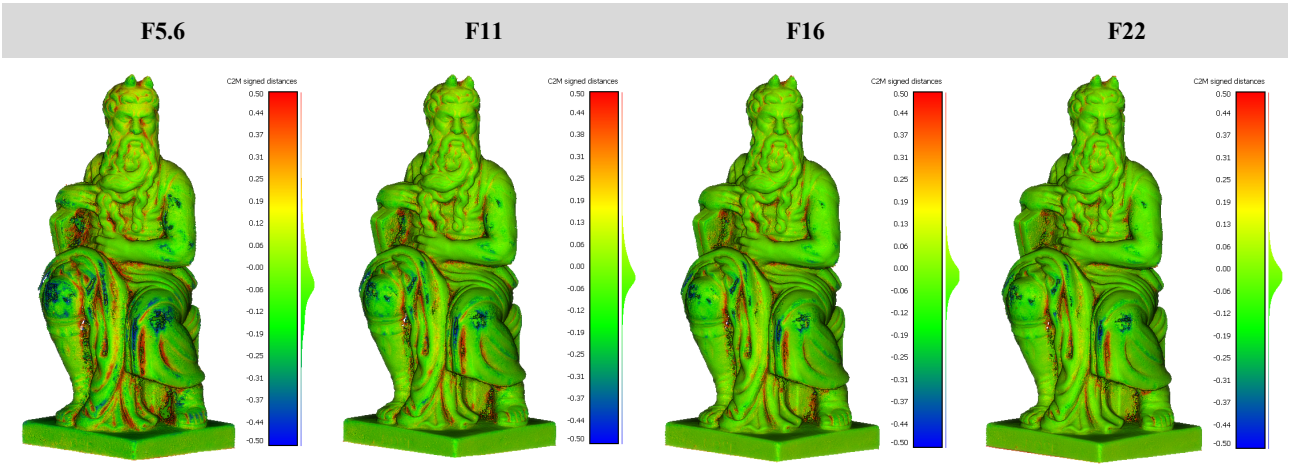

Figure 3. Cloud-to-Mesh comparison [mm] with F5.6, F11, F16 and F22 for the Moses statue. Range [-0.50/0.50 mm]. Results confirm fewer divergences from the ground truth with higher F-numbers.

The Moses statue was surveyed using a turntable, capturing images every 15° from two different standpoints and keeping an acquisition distance of 500 mm to maximise the frame with the object view. The MVS was performed at half-image size. As expected, 3D results (Table 1, Figures 2-3) show that, without masking unsharp areas produced by the DoF effect, the dense point cloud reconstructions achieved with lower F-numbers are noisier, thus requiring a longer editing time for the following steps. In these experiments, with larger F-numbers, the reduced resolving power of the images did not affect the final resolution, returning cleaner results.

The 2-euro coin was surveyed keeping an acquisition distance of 200 mm and testing the same aperture settings of the Moses case. Table 2 and Figures 4-5 report the achieved metrics. Also in this case, higher aperture settings provided better results. However, a slight worsening of metrics is notable in the largest F-value (F22) case, compared to the F16 setting.

Both examples, featuring objects with similar characteristics of museum artefacts, confirm that higher aperture settings and thus deeper DoF can significantly improve the reconstruction results when digitising such small objects with macro-lenses. However, as proved by the 2-euro coin example, some threshold aperture values should be considered for tiny objects, since the decreased image quality could affect the quality of the reconstruction.

It should be noted that, when handling massive digitisation, limited times for image acquisitions could influence the choice of the capturing parameters, as smaller apertures typically lead to longer acquisition processes.

F-number	**F5.6**		**F11**		**F16**		**F22**	
DoF [mm]	**0.6**		**1.2**		**1.7**		**2.3**	
	Mean	*St.dev.*	*Mean*	*St.dev.*	*Mean*	*St.dev.*	*Mean*	*St.dev.*
C2C	0.4084	0.5587	0.2756	0.3034	**0.0666**	0.0653	0.0930	0.1243
C2Mesh	0.2331	0.6525	-0.0128	0.4103	**0.0151**	0.0993	-0.0156	0.1614

Table 2. Metric evaluation (Cloud-to-Cloud and Cloud-to-Mesh comparison [mm]) for the dense reconstruction of the 2-euro object performed with several aperture settings (105 mm focal length, 200 mm object distance).

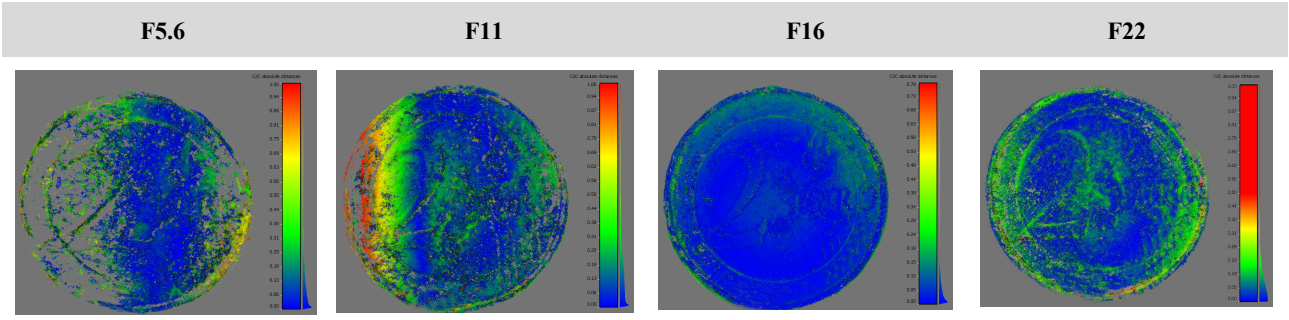

Figure 4. Cloud-to-Cloud comparison [mm] with F5.6, F11, F16 and F22 for the 2-euro case study. Range [0/1 mm].

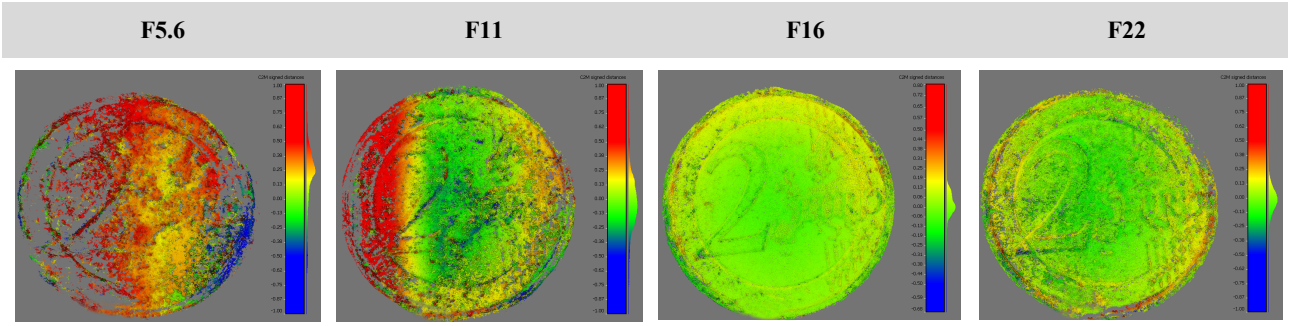

Figure 5. Cloud-to-Mesh comparison [mm] with F5.6, F11, F16 and F22 for the 2-euro case study. Range [-1/1 mm].

4. BACKGROUND MASKING

4.1 Developed masking approaches

Masking backgrounds is a demanding but frequently necessary task in massive digitisation to allow and facilitate tie point extraction and dense image matching (MVS) steps. Typical acquisition setups for acquiring small and medium-sized museum assets include turntables and uniform backgrounds. The artefact, when possible, is tilted and turned upside down on the turntable for entirely capturing the object geometry, and the 360° acquisition is repeated from fixed camera positions and several viewing angles. The image orientation step, in these cases, can be performed by first separately processing groups of images captured from the same camera position and then co-registering them, or jointly processing all the (masked) images. The second solution is preferable for avoiding image orientation issues, although masking the background in all the pictures is then required for orienting cameras.

Automatic techniques can help handle demanding and time-consuming masking tasks. The solutions developed and explored in this work include (Figure 6):

1. **Unsupervised learning approach:** it is based on the K-Means Clustering (Likas et al., 2003), which divides an unlabelled dataset into K groups of data points (referred to as clusters) based on their similarities. Since colour similarities are essential features in image segmentation, image datasets can be pre-processed to improve the results before running the clustering. RGB pictures are converted in the CIELAB (or L*a*b*) colour space, with L* as the Lightness channel, while chromatic a* axis extends from green to red and b* axis from blue to yellow. The predominant chromatic object range determines the selection of the suitable channel. A binary mask is finally generated from the output of the K-Means algorithm on the converted dataset. Some morphological operations (erosion and dilation) refine the masks removing small unwanted items.
2. **Supervised learning approach:** pixel features and user annotations are used to train a Random Forest classifier (Breiman, 2001) and assign a class label (object and background) to every image pixel. Employed pixel features are based on colours, edge filters and texture descriptors, extracted in a multi-scale approach. A model that extends the semantic segmentation to the entire set of images or even to similar datasets is finally generated from small representative annotations on one/a few dataset images. At last, the segmentation output is converted into binary masks, again refined with some morphological operations.
3. **Depth map-based approach:** the procedure relies on preliminary low-resolution data processing for generating and exporting a depth map per image. Then, the low-resolution dense reconstructions are roughly cleaned by removing points belonging to the background. Once the depth maps are generated, they are exported and adjusted, increasing their brightness and contrast values. Subsequently, two image filters can be applied: (i) a Gaussian blur filter (also known as Gaussian smoothing) to reduce image noise and details; (ii) a

Posterisation filter to ease the number of image colours and return sharper edges. From depth maps to image masks, the post-processing phase is carried out in a single round. It is worth noting that depth maps could also be quickly predicted using monocular approaches and deep learning networks (Ming et al., 2021). Although promising for handling the traditional ill-posed problem, these methods are not very suitable for masking artefact datasets. Further investigations are needed.

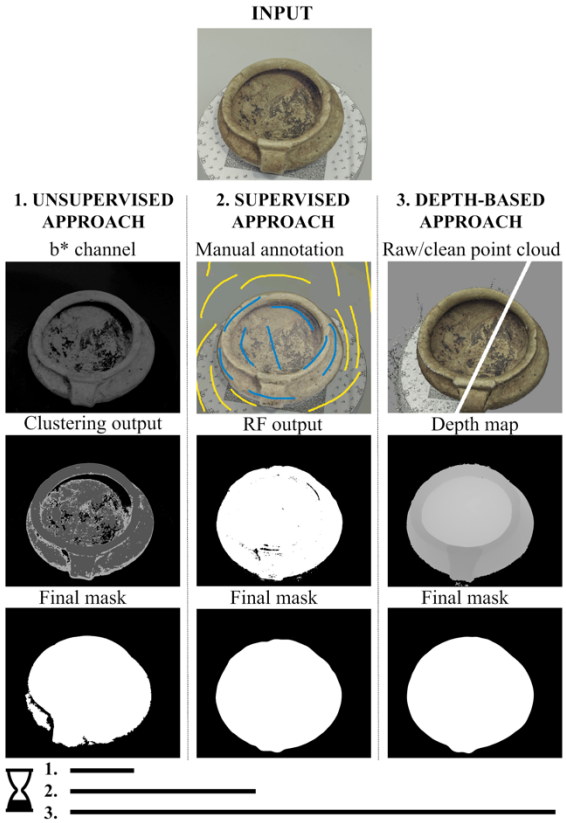

Figure 6. The proposed automated solutions to mask unwanted object's backgrounds. The respective processing times is also reported.

4.2 Experiments

The aforementioned masking techniques were tested on two sets of images depicting different heritage artefacts. The first object, a prehistoric wooden bowl of 17 cm diameter and 10 cm height, was surveyed using a rotating platform and taking pictures from four viewing angles. The artefact was flipped on the turntable during the acquisitions for capturing its entire geometry. The second test case is a small Roman bronze statue, some 18 cm high, including the base. Images were acquired from two different camera standpoints by rotating the object (constant background). In both cases, image masking was needed to orient all images correctly and derive cleaner dense point clouds.

Unsupervised techniques proved to be helpful and effortless for the operator if the object and background are clearly distinguishable. However, artefacts frequently share a similar pattern or colour with the background scenario. In these situations, the algorithm will include a portion of the background in the same segment, affecting the generation of correct masks (Figure 7).

With supervised learning approaches, patches and features should be carefully selected for training the classifier (e.g. Random Forest) and creating the model for the entire dataset. However, once verified that the pre-trained model is efficient, mask generation is rapid, the manual effort is reduced, and results are sufficient for orientation purposes (Figures 8). Lastly, the depth map processing method generally delivers the most precise masks, although the low-resolution data processing and the rough point cloud cleaning require additional operational times (Figure 8).

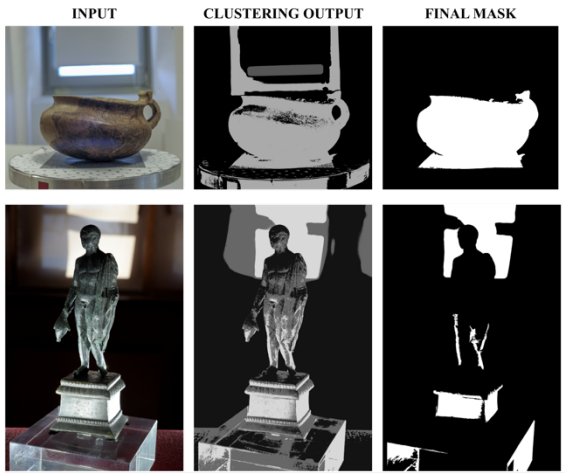

Figure 7. Mask generation with an unsupervised approach with incomplete and inoperative results.

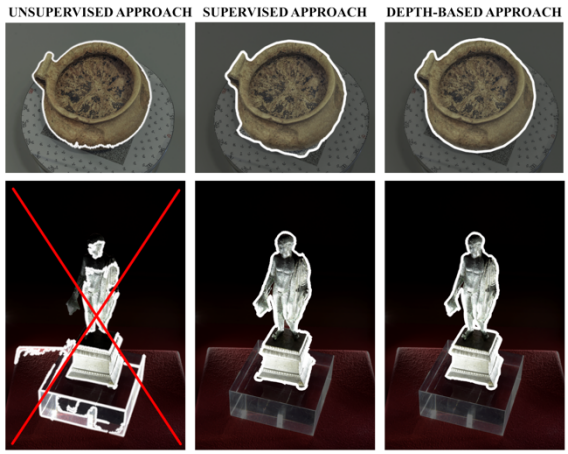

Figure 8. Masking results achieved with different approaches: unsupervised (K-Means Clustering), supervised (Random Forest) and depth-based.

Figure 9 shows MVS results using masks generated with supervised and depth map methods: the latter approach seems to allow dense point cloud with less noise and unwanted 3D points.

The scripts for testing the presented masking methods are available on the 3DOM-FBK Github page[1].

5. POINT CLOUD CLEANING

5.1 Tested point cloud denoising methods

Point cloud cleaning is a tedious but unavoidable task for generating accurate geometries for meshing and optimization steps (Figure 1).

[1] https://github.com/3DOM-FBK/Mask_generation_scripts

Figure 9. Dense point cloud results applying image masks within the MVS reconstruction process.

Among the available methods, those tested in this work include:

1. **Statistical Outlier Removal** (S.O.R.) (CloudCompare, 2021; Rusu and Cousins, 2011): it is a statistical analysis technique that removes points exceeding the average distance from their neighbours plus some standard deviations.
2. **PointCleanNet** (PCNet - Rakotosaona et al. 2020): it is a data-driven learning-based approach that firstly classifies and discards outlier samples, and then projects noisy points onto the original clean surfaces through estimated correction vectors.
3. **Score-Based Point Cloud Denoising** (Luo and Hu, 2021): a neural network is employed to estimate the score of $p*n$, where p is the distribution of a noise-free point cloud convolved with a noise model n. The predicted score is used for denoising data, increasing the log-likelihood of each point via gradient ascent, and iteratively updating each point's position without removing any point from the dataset.

5.2 Experiments

Firstly, a synthetic dataset (a cube with an edge length of 20 cm) was created and used to estimate the efficiency of the aforementioned methods in removing outliers and reducing noise in unordered point clouds. Random noise was added on the cube faces, with a standard deviation ranging from 1% to 5% of the object edge length. Cleaning methods were then applied, and from the results (Figures 10-11) clear messages may be derived:

- while the S.O.R. and the PointCleanNET methods remove most of the points detected as outlier and noise, the Score-Based Denoising algorithm merely adjust and update points' position closer to the main object's surface, and this could affect its performance when data include numerous outliers;
- while S.O.R. and PointCleanNET maintain relatively constant and similar performance as the noise level increases, the Score-based denoising metrics get worse significantly as the noise level rises;
- like most learning-based methods, the computing requirements and calculation times are much higher with respect to statistical techniques.

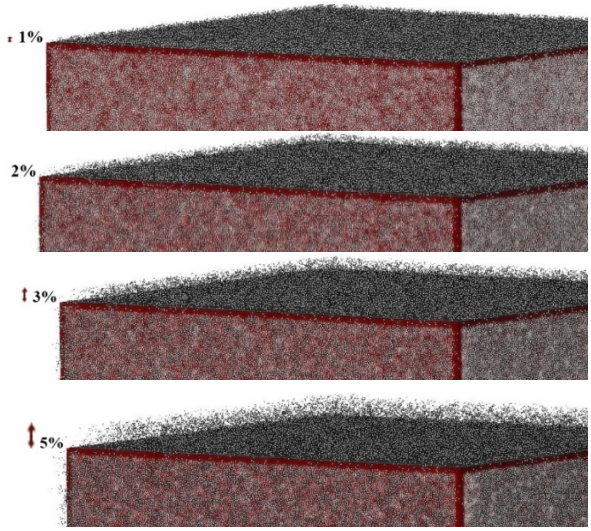

Figure 10. A close view of the synthetic (noisy) cube used for the experiments. From top to bottom, the standard deviation of the random noise corresponds to the 1%, 2%, 3%, 5% of the object edge length (20 cm).

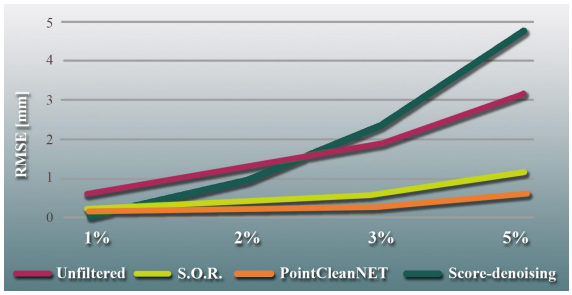

Figure 11. Average RMSE of the plane fitting on the six faces of the synthetic cube, testing several denoising approaches and noise intensity levels (1-5% of the cube length).

Secondly, the three methods were tested on the dense point cloud of the Moses statue, affected by some noise due to image imperfections, triangulation and matching inaccuracies. A visual comparison of the cleaning results is shown in Figure 12. The S.O.R. filter removed about 13% of the original points, while the PointCleanNet only 1%. In both cases, no significant cleaning improvement on the statue is evident. On the other hand, results based on the Score-based denoising method seem to visually outperform the other methods, delivering a cleaner and smoother dense reconstruction. The best result achieved with the Score-based denoising method, apparently in contrast with the results on the synthetic cube (Figure 11), can be explained by the low noise level affecting the Moses 3D reconstruction. However, further investigations are desirable when dealing with reality-based 3D data of artefacts.

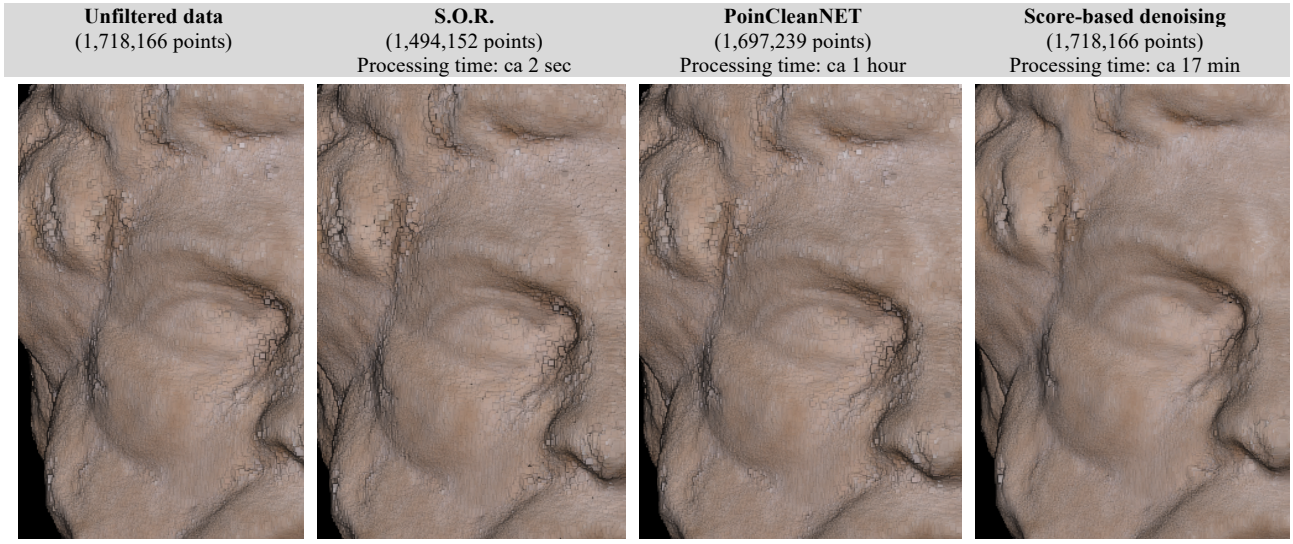

Figure 12: A visual comparison of denoising methods tested on the dense point cloud of the Moses statue. It should be noted that this method is not removing any 3D point from the cloud, it is only moving them based on a neural model.

6. CONCLUSIONS

The paper presented some experiments, analyses and lessons learnt on decisive steps of the photogrammetric workflow while surveying medium and small museums' artefacts. Based on past experiences, the three investigated aspects are among the most critical and demanding in massive digitisation.

The tests on DoF effects (Section 3) when images are acquired with macro-lenses proved that adequate capture settings are crucial for delivering precise and clean geometric results. Since masking blurry areas is a time-consuming task in massive digitisation, high aperture values and thus unsharp images should be prevented. On the other hand, smaller apertures, and thus lower image quality and longer acquisition times, should be carefully considered in the parameters' selection. DoF effects in 3D reconstructions are still under research, and further investigations in this field are planned.

The paper also explored some automatic solutions to speed up and lighten the image masking tasks (Section 4). Among the proposed approaches, the unsupervised method, based on colour similarities, proved to frequently fail, especially when objects and backgrounds share analogues chromatic ranges. On the contrary, both supervised and depth map-based approaches delivered quite accurate masks, sufficient for supporting 3D reconstruction processes. While the supervised technique is generally less accurate on edges and subsequent reconstructions are thus noisier, the method based on depth maps requires longer processing times, despite the quite optimal results.

Finally, the last experiments addressed the dense point cloud cleaning and denoising task (Section 5). A statistical-based approach and two deep learning techniques were compared for investigating their performance. Results show that the learning-based approaches are promising, although the actual processing times and computational requirements are still too high for handling massive digitisation. Further tests need to be performed as these methods were tested only on simple objects.

ACKNOWLEDGEMENTS

Authors are thankful to the Cultural Heritage Directorate of the Autonomous Province of Trento (Italy) and the Museo Alto Garda (MAG) in Riva del Garda (Italy) for giving access to their collections within the joint JUDIT project activities (https://judit.fbk.eu/). Authors are also thankful to Microgeo s.r.l. (https://www.microgeo.it/) for supporting us in the collection of the ground truth data (Moses and coin).

REFERENCES

Apollonio, F.I., Fantini, F., Garagnani, S. and Gaiani, M., 2021. A Photogrammetry-Based Workflow for the Accurate 3D Construction and Visualization of Museums Assets. *Remote Sensing*, *13*(3), p.486.

Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp. 5-32.

Cignoni, P. and Scopigno, R., 2008. Sampled 3D models for CH applications: A viable and enabling new medium or just a technological exercise? *Journal on Computing and Cultural Heritage (JOCCH)*, *1*(1), pp.1-23.

CloudCompare, 2021: http://www.cloudcompare.org/

Collins, T., Woolley, S.I., Gehlken, E. and Ch'ng, E., 2019. Automated low-cost photogrammetric acquisition of 3D models from small form-factor artefacts. *Electronics*, *8*(12), p.1441.

De Paolis, L.T., De Luca, V., Gatto, C., D'Errico, G. and Paladini, G.I., 2020. Photogrammetric 3D Reconstruction of Small Objects for a Real-Time Fruition. *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pp. 375-394.

Duan, C., Chen, S., and Kovacevic, J.. 2019. 3D Point Cloud Denoising via Deep Neural Network Based Local Surface Estimation. Proc. *IEEE ICASSP*, pp. 8553–8557.

Erler, P., Guerrero, P., Ohrhallinger, S., Mitra, N.J. and Wimmer, M., 2020. Points2surf learning implicit surfaces from point clouds. Proc. *ECCV*, pp. 108-124.

Gattet, E., Devogelaere, J., Raffin, R., Bergerot, L., Daniel, M., Jockey, P. and De Luca, L., 2015. A versatile and low-cost 3D acquisition and processing pipeline for collecting mass of archaeological findings on the field. *ISPRS Int. Archives of*

Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(5).

Gonizzi-Barsanti, S., Remondino, F., Jiménez Fernández-Palacios, B. and Visintini, D., 2014. Critical factors and guidelines for 3D surveying and modelling in Cultural Heritage. *Int. Journal of Heritage in the Digital Era*, Vol. 3(1), pp. 142-158.

Guidi, G., Barsanti, S.G., Micoli, L.L. and Russo, M., 2015. Massive 3D digitization of museum contents. Proc. *Built heritage: Monitoring conservation management*, pp. 335-346.

Grilli, E., Battisti, R., Remondino, F., 2021. An Advanced Photogrammetric Solution to Measure Apples. *Remote Sensing*, 13(19):3960.

Han, X.F., Jin, J.S., Wang, M.J., Jiang, W., Gao, L. and Xiao, L., 2017. A review of algorithms for filtering the 3D point cloud. *Signal Processing: Image Communication*, *57*, pp.103-112.

He, K., Gkioxari, G., Dollár, P., Girshick, R.B., 2020. Mask R-CNN. *IEEE Trans. PAMI*, 42, 386-397.

Hermosilla, P., Ritschel, T., Ropinski, T., 2019. Total Denoising: Unsupervised Learning of 3D Point Cloud Cleaning. Proc. *IEEE ICCV*, pp. 52–60.

Knyaz, V.A., Kniaz, V.V., Remondino, F., Zheltov, S.Y., Gruen, A., 2020. 3D reconstruction of a complex grid structure combining UAS images and deep learning. *Remote Sensing*, 12(19), 3128.

Lastilla, L., Ravanelli, R., and Ferrara, S., 2019. 3D high-quality modeling of small and complex archaeological inscribed objects: relevant issues and proposed methodology. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, 42(2/W11), pp. 699–706

Likas, A., Vlassis, N. and Verbeek, J.J., 2003. The global k-means clustering algorithm. *Pattern recognition*, 36(2), pp.451-461.

Liu, L.Y.F., Liu, Y. and Zhu, H., 2020. Masked convolutional neural network for supervised learning problems. *Stat*, *9*(1), p.e290.

Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2020. *Close-Range Photogrammetry and 3D Imaging*. De Gruyter: Berlin, Germany.

Luo, S. and Hu, W., 2020. Differentiable manifold reconstruction for point cloud denoising. Proc. *28th ACM International Conference on Multimedia*, pp. 1330-1338.

Luo, S. and Hu, W., 2021. Score-Based Point Cloud Denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4583-4592).

Menna, F., Rizzi, A., Nocerino, E., Remondino, F., Gruen, A., 2012. High resolution 3D modeling of the Behaim globe. *ISPRS Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 39(5), pp. 115-120.

Menna, F., Nocerino, E., Remondino, F., Dellepiane, M., Callieri, M. and Scopigno, R., 2016. 3D digitization of an heritage masterpiece-a critical analysis on quality assessment. *ISPRS Int. Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *41*.

Menna, F., Nocerino, E., Morabito, D., Farella, E.M., Perini, M. and Remondino, F., 2017. An open source low-cost automatic system for image-based 3D digitization. *ISPRS Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, p.155.

Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N., Terzopoulos, D., 2021. Image segmentation using deep learning: A survey. *IEEE Trans. PAMI.*

Ming, Y., Meng, X., Fan, C., Yu, H., 2021. Deep learning for monocular depth estimation: A review. *Neurocomputing*, 438, pp. 14-33.

Niederöst, M., Niederöst, J. and Skucka, J., 2003. Automatic 3D reconstruction and visualization of microscopic objectsfrom a monoscopic multifocus image sequence. *ISPRS Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(5/W10).

Rakotosaona, M.J., La Barbera, V., Guerrero, P., Mitra, N.J. and Ovsjanikov, M., 2020. Pointcleannet: Learning to denoise and remove outliers from dense point clouds. *Computer Graphics Forum*, Vol. 39, No. 1, pp. 185-203.

Remondino, F., Menna, F., Koutsoudis, A., Chamzas, C. and El-Hakim, S., 2013. Design and implement a reality-based 3D digitization and modelling project. Proc. *Digital Heritage International Congress*, Vol. 1, pp. 137-144.

Rodríguez-Martín, M. and Rodríguez-Gonzálvez, P., 2020. Suitability of automatic photogrammetric reconstruction configurations for small archaeological remains. *Sensors*, 20(10), p.2936.

Roncella, R., Bruno, N., Diotri, F., Thoeni, K. and Giacomini, A., 2021. Photogrammetric Digital Surface Model Reconstruction in Extreme Low-Light Environments. *Remote Sensing*, *13*(7), p.1261.

Tausch, R., Domajnko, M., Ritz, M., Knuth, M., Santos, P. and Fellner, D., 2020. Towards 3D digitization in the GLAM (Galleries, Libraries, Archives, and Museums) sector: Lessons learned and future outlook. *IPSI BgD Transactions on Internet Research (TIR)*, *16*(1), pp.1-9.

Raimo, N., De Turi, I., Ricciardelli, A. and Vitolla, F., 2021. Digitalization in the cultural industry: evidence from Italian museums. *International Journal of Entrepreneurial Behavior & Research*.

Rusu, R.B. and Cousins, S., 2011. 3D is here: Point cloud library (PCL). In *2011 IEEE international conference on robotics and automation* (pp. 1-4). IEEE.

Sapirstein, P., 2018. A high-precision photogrammetric recording system for small artifacts. *Journal of Cultural Heritage*, Vol. 31, pp. 33-45

Verhoeven, G., 2018. Focusing on out-of-focus: assessing defocus estimation algorithms for the benefit of automated image masking. *ISPRS Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, pp.1149-1156.

Webb, E.K., Robson, S. and Evans, R., 2020. Quantifying depth of field and sharpness for image-based 3D reconstruction of heritage objects. *ISPRS Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* 439(B2), pp. 911-918.