Image-based 3D capture of cultural heritage artifacts

An experimental study about 3D data quality

G. Guidi, L. L. Micoli, S. Gonizzi
Dept. of Mechanical Engineering
Politecnico di Milano
Milan, Italy
gabriele.guidi, laura.micoli, sara.gonizzi@polimi.it

M. Brennan, B. Frischer
Dept. of Informatics
Indiana University
Bloomington (IN), USA
matthewrbrennan, bernard.d.frischer@gmail.com

Abstract — **The paper presents an analysis of the 3D data quality generated from small-medium objects by well-known automatic photogrammetry packages based on Structure from Motion (SfM) and Image Matching (IM). The work aims at comparing different shooting configurations and image redundancy, using as high-quality reference the 3D data acquired by triangulation-based laser scanners characterized by a low measurement uncertainty. Two set of tests are presented: i) a laboratory 3D measurement made with the two active and passive approaches, where the image-based 3D acquisition makes use of different camera orientations leading to different image redundancy; ii) a 3D digitization in the field with an industrial laser scanner and two sets of images taken with different overlap levels. The results in the field confirm the relationship between measurement uncertainty and image overlap that emerged in the Lab tests.**

Index Terms— **SfM, Image Matching, Image overlap, 3D data quality, Resolution, Uncertainty, Accuracy.**

I. INTRODUCTION

Digitizing Cultural Heritage is nowadays a well-established activity for 3D documenting of heritage assets. Active and passive 3D techniques have been both used since the early days but, thanks to the last few years improvements of Computer Vision algorithm associated to photogrammetric principles, the passive (e.g. image-based) techniques are becoming predominant in this application area.

In the case of cultural heritage artifacts like statues and smaller objects the two methods involve two rather different sets of tools and processes for obtaining the final textured mesh model.

The well-known 3D processing pipeline based on active technologies gives its output after a 3D acquisition phase made with a dedicated device for medium/small volumes like a triangulation-based laser scanner or a pattern projection range device. The raw 3D data generated by such devices are registered in a single reference system using some 3D data redundancy and the Iterative Closest Point (ICP) algorithm, or using an additional device for capturing position and orientation of the range device at the acquisition stage. Independently of the alignment method used, the final result is a cloud of 3D points, often made by subsets that can be possibly meshed independently, that originates a single mesh representing the scanned artifact with a level of resolution and uncertainty given by the intrinsic performances of the range device. Once the geometrical part of the acquisition is completed, a high quality texture can be added performing a photographic campaign around the same artifact and aligning each image to the mesh with an additional process involving significant manual work. The quality of this 3D output is surely very high both in term of accuracy of the digital model with respect to the physical object, and, if the quality of photographs is high, in terms of texture. The weak part of this process lies in its cost both regarding the equipment used (one laser scanner and one professional camera if we omit the cost for the software by using an open source software like Meshlab for processing the 3D data), and especially regarding time, which cannot fall below a certain threshold given the number of activities to be completed before obtaining the final result.

Photo-modeling solutions are instead just based on a camera, a single piece of equipment that is used both for capturing 3D data and textures at the same time. The photos taken around the object to be surveyed are analyzed with Computer Vision algorithms to detect and describe local features in images (e.g. SIFT). The corresponding local features found on adjacent photos are used for determining their orientation in space through the well-known bundle adjustment method used for years in photogrammetry, and finally each pixel or group of pixels within each image is matched with the corresponding ones belonging to different images, in order to determine their 3D coordinates in space by triangulation. As a result a dense cloud of colored points is generated (x, y, z, R, G, B), all in the same reference system. By meshing this cloud with well-known algorithms (i.e. Delaunay, Poisson, Ball-Pivoting, etc.) a high density mesh with color on each vertex is generated. A final UV parametrization of the mesh makes it possible to map the color points onto an UV space, generating a texture image associated with the 3D geometry.

As already noted by other papers, the total time needed to obtain the final textured model is much shorter in the second case, where most of the operations, even if possibly long, are automatically executed by a computer and do not require a continuous interaction with the operator, as in the first case.

Of course the basic assumption is that the heritage artifact is suitable to be 3D digitized with both methods. This means that for active devices the object surface has to be optically

978-1-5090-0048-7/15/$31.00 ©2015 IEEE

cooperative (i.e., as close as possible to the Lambertian behavior), with enough 3D features to be used for self-referencing the range images in the alignment phase. Similarly, image-based method works well if enough features are present on the texture of the object to be captured, and the light-material interaction optically cooperative.

Therefore the only reasons why an active device might still be useful in cultural heritage 3D acquisition is related to two aspects: i) the visual appearance of the object does makes it impossible to find automatically the image features needed for orienting images (e.g. a completely white plaster statue); ii) the stronger "a-priori" controllability of metrological parameters. In fact, with a triangulation 3D scanner as the ones used in these experiments, once a lens have been defined, the baseline and the light source vs. camera orientation are fixed, giving a rather predictable uncertainty at a given distance. On the other hand in a range map generated with SfM/IM, several parameters (focusing, depth of field, baseline, image overlap, camera orientations, possible movement blurring, image pre-processing, bundle adjustment, matching algorithm, etc.) may influence the final result, giving a much smaller "a-priori" predictability of the final result.

So, in terms of closeness of the digital model to the real object, if the object is properly textured,the two methods can render similar results, with an important difference that may be present. The passive technique gives a metric result only after a scaling step made by assigning a known value to one or more distances recognizable in the scene or measuring some "Ground Control Points" (GCPs) with a complementary method. Contrarily active devices, once calibrated, are intrinsically metric. Therefore in automatic photogrammetry the final scaling may become a critical step whose accuracy have impact on the whole 3D digitization accuracy.

Another advantage of the pipeline based on active devices is that the measurement uncertainty of the device and of the whole process can be precisely checked. Recently this feature has been introduced also for some image matching packages (e.g. SURE), but that was not available on the photogrammetric tools used in this research.

For this reason the final 3D model generated by 3D scanning was metrologically characterized in a stronger way than with those generated with our photogrammetric tools, becoming a suitable "gold standard" for the comparisons.

The purpose of this paper is to explore those two points: i) how much a standard photogrammetric process like the one used in museums, may influence the final result, and ii) how the image redundancy may influence the quality of the 3D points in terms of measurement uncertainty.

II. PREVIOUS WORK

Several authors has already made studies comparing the results attainable by active or passive survey technology, but the comparison has been mostly dedicated to the comparison between photogrammetry, on the one hand, and, on the other hand, terrestrial laser scanning for large volumes, mainly based on Time of Flight (ToF) or Phase Shift (PS) detection. The focus has been first on traditional photogrammetry for buildings and large structures [1, 2], extending more recently the evaluation to SfM-based automatic photogrammetry by comparing the point cloud with laser scanned data [3, 4].

Such comparison at building-scale has been extended to different laser-scanners and SfM packages like architectural structures acquired with a ToF device compared with a cloud generated by SfM/IM [5, 6], or comparing PS devices with SfM/IM [7]. In any case the various papers report comparable results in terms of measurement uncertainty, overstressing the low cost and operational speed of image-based methods, confronted with the metric output generated by laser scanners.

In 2012 Rodríguez-Navarro made a test on a small stone sculpture and an architectural element extending the analysis to a smaller scale, by comparing the results from a triangulation-base Nextengine laser scanner and a point cloud generated by AGISOFT Photoscan [8] which was obtained with less effort using photogrammetry. Such early comparison on small volumes was later extended by consideration of various commercial and open source image-based solutions by Remondino et al. [9], evidencing the not negligible influence of the SfM and Image Matching (IM) algorithms on the final result. However, as demonstrated in digitizing an entire museum [10], the efficiency of image-based 3D digitizing of small objects is nowadays clear, but the metrological quality of the many 3D data collected in that project, although comparable or better than ToF and PS laser scanners, seems not always at the same level of high end triangulation-based range devices. Exploring this specific point is the motivation of the present paper.

III. EXPERIMENTS

Two object have been considered for performing the experiments: i) a laboratory test object represented by a female head made of polystyrene, suitable to be captured with both passive and active methods; ii) a full size Roman statue of Caligula conserved at the Virginia Museum of Fine Arts, Richmond (VA), USA.


a) b)

Fig. 1. Objects used for the experiments: a) female head made of polystyrene; b) Roman statue of Caligula conserved at the Virginia Museum of Fine Arts, Richmond (VA), USA.

Several shooting styles were used, together with two different SfM-based commercial packages (AGISOFT

Photoscan and Autodesk Recap), whose results have been quantitatively compared.

The tests have been done using firstly a laser scanner 3D acquisition as test reference. Two different laser scanning approaches have been used in the experiments for generating the "gold reference" model.

In the Lab a Minolta 910 was used, equipped with a Medium lens (focal length=14 mm), working at 80 cm scanner-to-target distance, providing a spatial resolution of 0.4 mm and a measurement uncertainty of 120 µm. The 42 images acquired with such device on the test object had a 40% overlap for allowing a proper ICP alignment of the 3D dataset. Such process was made with the commercial package Innovmetric Polyworks, as the following 3D data merge, meshing and topological cleaning. No smoothing was performed in order to keep the best representation of all the 3D details. The alignment statistics confirmed that the deviation between adjacent range-maps, in the overlap zone, was in the order of the nominal measurement uncertainty, giving a global uncertainty of the digital representation of the object with the respect to the physical object of 120 µm. Once the various range maps were merged the final bounding box was 201 mm x 196 mm x 349 mm, and the mesh size 1.2 millions of polygons (564 kPoints).

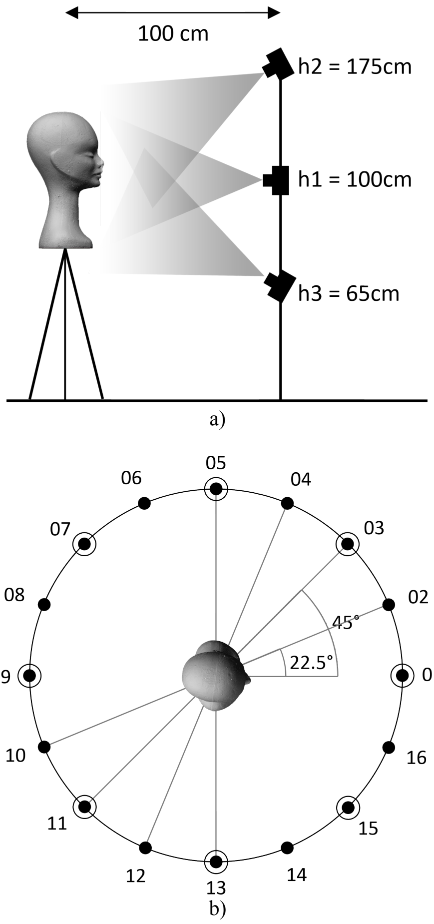

Fig. 2. Laboratory shooting set and positioning of the camera around the object: a) vertical; b) horizontal.

The same object was acquired with images, but owing to the uniform white color, it was "decorated" with signs for increasing the number of features suitable to be automatically detected (Fig. 1a). It has to be considered that this might be a weak point of the test, since a better decoration could have produced more tie points, and a possible better final result. A set of reference targets where then added on the scene for scaling the photogrammetric project.

The photos were taken with a Canon DSLR 5D Mk II, whose full frame sensor (36 mm x 24 mm) provides a 21 megapixel image (5616 x 3744), with a pixel size ps=6.4 µm. The camera was equipped with a f=50 mm macro lens by Canon and used at a distance d=1m from the test object. The sapling distance on the object surface, given by GSD=ps*d/f, result equal to 0.128 mm. The shoot was made with the test object located on a stand at 1m from the ground, and the camera shooting around the object every 22.5°, for a total of 16 images. Three groups of 16 images were taken, with the camera set, respectively, at height of 0.65m, 1m, and 1.75m (see Fig. 2).

In this way 3*16=48 images were captured all around the object. The images were processed with Photoscan, first using all the images, and then, in a second stage, taking only every other image. The second method corresponded to one shot every 45°, and utilized half of the acquired images.

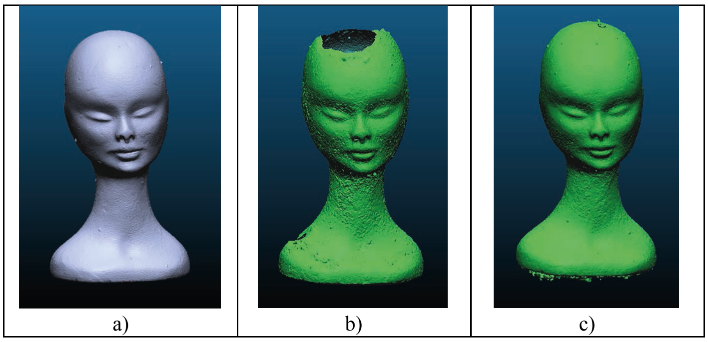
a) b) c)

Fig. 3. Meshes originated by: a) laser scanning; b) SfM/IM with half the images (24); c) SfM/IM with all the available images (48).

Both the orientation and the matching phase were repeated from scratch starting a fresh project with 48 and 24 images respectively. The dense cloud was generated by Photoscan in the so-called "medium" mode, which involves a matching of 4x4 image sub-windows. This imply a spacing of the 3D points in the dense cloud 4 times larger than the GSD, namely 0.512 mm, close to the spatial resolution of the laser data.

TABLE I. LAB DATA COMPARISONS

	Mean (mm)	*Std. Dev. (mm)*
Laser scanner vs. SfM/IM – 24 Images	0.41	1.51
Laser scanner vs. SfM/IM – 48 Images	0.30	0,73

As shown by Fig. 3a, the visual appearance of the shaded mesh originated by laser scanning suggests that the corresponding 3D data are better in quality of the other two meshes, built on the dense cloud generated with IM. In particular it can be noticed that the mesh produced with one image every 45° (less image overlap), did not produce any result on the top part of the head and generated a definitely irregular geometry in the area of the neck (Fig. 3b), while the set of images with more redundancy produces a visually better surface. This apparent behavior is confirmed by the numbers in Table I, where the results of a mesh-to-mesh comparison with Cloud Compare is reported, using the laser data as reference.

It can be noticed that i) the scaling phase might be critical, giving a not negligible mean deviation; ii) the standard deviation of error is reduced by half using the denser image set.

For the test in the field the statue shown in Fig 1b was used. It is a full size marble statue 2032 mm tall including its pedestal, 673 mm wide and 495 mm deep, conserved at the Virginia Museum of Fine Arts. The statue was 3D digitized for a previous project aiming at digitally restoring the original colors (Digital Caligula Project, NEH grant RZ-51221). For this reason the statue was first 3D digitized with an industrial range device made by a laser triangulation head mounted on an arm CMM (Faro Arm). This device measures the 6DOF of the scanning head making it possible to reduce measurement errors with respect to other stand-alone systems since no ICP alignment is made, leading to a global mesh uncertainty of 57 µm, with a lateral resolution of 0.1 mm.

The photogrammetric survey has been carried out after an important physical restoration of the sculpture. This second 3D digitization was carried out in two separate phases. First a Nikon D90 camera was used, equipped with a 12 megapixel DX CMOS sensor covering an area of 24 mm x 16 mm. In order to avoid high ISO settings the camera was used on a tripod, and for this positioning limitation 76 different poses were used. Due to the sub-optimal results obtained a second shooting was made with a Nikon D600, with a 24 megapixel FX CMOS sensor, covering the full frame of 36 mm x 24 mm. Such camera has a higher sensitivity sensor and it was possible to shoot with a hand-held camera, framing the sculpture from 250 different poses with a far higher image overlap.

In both cases the GSD was in the range 0.4-0.6 mm.

The two image sets were processed with both AGISOFT Photoscan and Autodesk Recap, generating four meshes that have been compared with the high quality laser scanned mesh through the Polyworks Inspector software. The results of such comparison are reported in Table II.

TABLE II. MUSEUM DATA COMPARISONS

	Mean (mm)	***Std. Dev. (mm)***
Photoscan, D90, images taken with tripod	1,18	3,38
Photoscan, D600, handheld camera	0,66	2,22
Recap, D90, images taken with tripod	-0,30	3,25
Recap, D600, handheld camera	0,34	1,80

IV. DISCUSSION AND CONCLUSIONS

As shown by the numbers in Table II a residual mean error due to a sub-optimal scaling of the statue is always present, which introduces a lack of global accuracy that could be easily restored giving more importance to the scaling phase which is often neglected in cultural heritage applications.

Given the shooting conditions the standard deviation of error is lower of 3.5 mm in any case, confirming a value closer to the typical measurement uncertainty of ToF/PS laser scanners rather than triangulation based active devices.

The results obtained with larger overlap gives a significant reduction in measurement uncertainty, giving a 34% and 45% standard deviation reduction with Photoscan and Recap respectively. This confirms the 52% uncertainty reduction found in the Lab experiment doubling the number images in the photogrammetric block.

This result suggest a direct relationship between image redundancy and measurement uncertainty of the 3D data originated by SfM-based photogrammetry.

REFERENCES

[1] R. Kadobayashi, N. Kochi, H. Otani, and R. Furukawa, “Comparison and evaluation of laser scanning and photogrammetry and their combined use for digital recording of cultural heritage,” *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. 35, no. 5, pp. 401–406, 2004.

[2] P. Grussenmeyer, T. Landes, T. Voegtle, and K. Ringle, “Comparison methods of terrestrial laser scanning, photogrammetry and tacheometry data for recording of cultural heritage buildings,” in *ISPRS Congress*, 2008, pp. 213–218.

[3] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, “On benchmarking camera calibration and multi-view stereo for high resolution imagery,” *2008 IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008.

[4] M. Golparvar-Fard, J. Bohn, J. Teizer, S. Savarese, and F. Peña-Mora, “Evaluation of image-based modeling and laser scanning accuracy for emerging automated performance monitoring techniques,” *Autom. Constr.*, vol. 20, no. 8, pp. 1143–1155, Dec. 2011.

[5] D. Skarlatos and S. Kiparissi, “Comparison of laser scanning, photogrammetry and SfM-MVS pipeline applied in structures and artificial surfaces,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I–3. pp. 299–304, 2012.

[6] A. Koutsoudis, B. Vidmar, G. Ioannakis, F. Arnaoutoglou, G. Pavlidis, and C. Chamzas, “Multi-image 3D reconstruction data evaluation,” *J. Cult. Herit.*, vol. 15, no. 1, pp. 73–79, Jan. 2014.

[7] F. Fassi, L. Fregonese, S. Ackerman, and V. De Troia, “Comparison Between Laser Scanning and Automated 3D Modelling Techniques To Reconstruct Complex and Extensive Cultural Heritage Areas,” in *3D-ARCH 2013 - 3D Virtual Reconstruction and Visualization of Complex Architectures*, 2013, vol. XL–5/W1, pp. 73–80.

[8] P. Rodríguez-Navarro, “Automated Digital Photogrammetry versus the systems based on active 3D sensors,” *EGA*, vol. 17, no. 20, pp. 100–111, 2012.

[9] F. Remondino, M. G. Spera, E. Nocerino, F. Menna, and F. Nex, “State of the art in high density image matching,” *Photogramm. Rec.*, vol. 29, no. 146, pp. 144–166, 2014.

[10] G. Guidi, S. G. Barsanti, L. L. Micoli, and M. Russo, “Massive 3D Digitization of Museum Contents,” in *Built Heritage: Monitoring Conservation Management*, L. Toniolo, M. Boriani, and G. Guidi, Eds. Springer, 2015, pp. 335–346.