

Practical 3D Reconstruction of Cultural Heritage Artefacts from Photographs – Potentials and Issues

Dieter W. Fellner^{1,2}, Sven Havemann¹, Philipp Beckmann¹ and Xueming Pan¹

¹ Institute of ComputerGraphics and KnowledgeVisualization (CGV), TU Graz, Austria

² Fraunhofer IGD and GRIS, TU Darmstadt, Germany

Abstract

A new technology is on the rise that allows the 3D-reconstruction of Cultural Heritage objects from image sequences taken by ordinary digital cameras. We describe the first experiments we made as early adopters in a community-funded research project whose goal is to develop it into a standard CH technology. The paper describes in detail a step-by-step procedure that can be reproduced using free tools by any CH professional. We also give a critical assessment of the workflow and describe several ideas for developing it further into an automatic procedure for 3D reconstruction from images.

Key words: 3D RECONSTRUCTION, PHOTOGRAMMETRY, 3D ACQUISITION, 3D SCANNING, ARC3D, MESH LAB

1. INTRODUCTION

3D reconstruction from photographs has the potential to revolutionize the digital documentation of Cultural Heritage artifacts. No expensive delicate equipment like 3D-scanners is necessary for capturing data, but an ordinary digital camera is sufficient. Decent cameras are often already part of the equipment of CH professionals for capturing documentary photographs. But in addition to this they can also be used for taking so-called *image sequences*, and the acquired photos can be used for 3D-reconstruction. This requires today a somewhat tedious workflow, but it may soon be automated.

Apparently this technology is not so widely known in the professional Cultural Heritage community. It was for instance presented in Graz, Austria, on a workshop of the *Steirisches Denkmalamt* on the conservation of pre-historic wood. The following presentation was given by Rengert Elburg from the *Landesamt für Archäologie Sachsen* who presented the enormous effort made for excavating a large scale pre-historic well in Saxonia. He started his presentation with the words: “*If only someone had told me three years ago that we should simply take many image sequences!?*”.

To make this readily available for testing in the CH community is the purpose of our contribution.

The quality of the 3D reconstruction results can be greatly improved when following only a few rules. The purpose of this paper is to report on the practical experiences we have made in a number of acquisition campaigns. With any such campaign a great challenge is *sustainability*: Without a faithful documentation of the workflow it is difficult if not impossible to judge the quality, the authenticity, or to re-use intermediate results.

The 3D-COFORM project

3D-COFORM (www.3d-coform.eu) is a 4-year integrated project (FP7-IP) funded by the European community. Its main objective is to make the use of 3D technology a standard in Cultural Heritage, and to develop all necessary technologies for data acquisition (in museums and on archeological campaigns), for data processing, for semantics and metadata processing, for museum presentations.

In particular, a repository-centric approach is adopted with a distributed central database to document paradata (LONDON CHARTER, 2006) describing the digital provenance of all acquired data, processing steps, and results (PAN, 2010).

Especially innovative is that all metadata are recorded in a semantic network following a common standard, the CIDOC-CRM (CROFTS, 2005), which is in fact an ISO standard for describing cultural facts.

At this point the question remaining is how image-based 3D reconstruction works in practice.

2. THE ACQUISITION PHASE

The *Gipsmuseum* of the Institute of Archeology of Karl-Franzens University Graz (KFU) contains more than a hundred 1:1 plaster copies, primarily of ancient Greek and Roman statues. We have selected a set of 24 statues for photogrammetric reconstruction. Each statue has a label showing a number (Figure 1), this number we have taken as object ID for our files and directories (Table 1).



Figure 1: The Gipsmuseum of KFU Graz contains more than 100 plaster copies of marble statues. 24 of them were acquired using photo-based 3D reconstruction using projected light to account for the white surface. The labels of the statues were directly used as data reference in the reconstruction workflow.

Acquisition Phase: Process Description

The Gipsmuseum acquisition process required projecting a random pattern onto the statues: As explained in more detail below, photogrammetric reconstruction proceeds by comparing pairs of images, i.e., individual pixels. If the pixels all have the same color, no distinctions can be made, and thus, no depth value can be computed. Plaster statues are primarily white, this is why we had to apply a high-frequency texture artificially.

The setup (Fig. 2) is somewhat complex because the projector has to be put into a different position for every sequence. The camera tripod ideally moves on a half-circle around a chosen point on the surface. However, the camera must not be in front of the projector, so it must either be higher or lower than the projector. In principle, the camera should be as close as possible, so that a larger field of view can be used, which allows for calculating the disparity more robustly. Sometimes this is not possible because of the space constraints in the museum, of course we may not move the artefacts.



Figure 2: The acquisition process. The string helps to keep the camera at the same distance from the chosen surface point. The camera autofocus is switched off both for greater speed and accuracy.

Preparation and Capture

We use a 6 megapixel Nikon D50 camera with a good 18-55 mm zoom lens and a 55-200 mm tele lens. There is a rather informal list of things to do and remember when taking the pictures:

Preparation phase

- Clear memory cards, load batteries
- Projection laptop: Display random image in 1400 x 1050 resolution in presentation mode

- Material: projector, camera, tripod, string, 2nd memory card and 2nd battery pack, long power cable, multi-plug, camera remote control

Acquisition planning

- Take context pictures of the statue from all sides, at normal light, including ID plate
- First take an overview sequence of statue as a whole, then make detail sequences
- Plan sequence of camera positions for the detail sequences, in particular whether to place the camera in front or behind the projector, and whether it should be above or below.

The actual acquisition pattern is highly object dependent, although for some object classes a certain typical schema emerged with experience. For example for a **bust** we used this schema: Three projector positions, from front left, front right, and from behind. For more complex busts the projector is positioned two times, shining from slightly above and slightly below. Questions to clarify are: Is anything occluded? Is any part of the statue not visible? Then, for each projector position, one sequence is taken from below the projector (camera looking upwards), and one from above the projector (camera looking slightly downwards).

Acquisition of a sequence

- Determine one fix point on the surface
- Determine the camera distance (halfcircle radius), make a mark on the string, fix the camera vertically
- Determine the zoom factor: Is the object completely visible at the start, the middle, and end positions on the halfcircle?
- At the midpoint: Determine best focus setting, then lock the lens (switch off autofocus)
- Set camera to manual operation: F-stop is set above middle to (large depth of field), ISO is always 200 (minimum, e.g., lowest noise), then adjust shutter speed accordingly so that image is sufficiently bright. Avoid over- and under-exposure. Use a remote control for the camera.
- Go with tripod along halfcircle, take one photo every 10-15 cm, keep fix point in image center
- Rule of thumb: 60 images per full circle
- Change the height of the camera every 2-3 photos by adjusting the tripod top rod

Considerations on camera settings and lighting

Concerning the F-stop setting, a compromise is required between a large depth of field (max F-stop, e.g. 28) and the avoidance of refraction blur which occurs at a large F-stop because the pinhole is extremely small. This is a very delicate issue. In case of objects with high depth complexity, a large F-stop is unavoidable, but this leads to a uniform blur over the whole depth range. A medium F-stop (e.g., 14) greatly increases the focus sharpness, but only within some range around the

focus depth. Some experimentation is required; we tend to use a medium F-stop whenever possible.

Another delicate issue is lighting. Ideal is a bright, diffuse illumination from all sides, as for instance outside on a cloudy day, where it is impossible to tell where the sun is. – We try to approximate this kind of illumination whenever taking pictures inside (see Fig. 3).

Interestingly, it is desirable to use a closer range with a large depth of field (wide-angle zoom, e.g., 18 mm) instead of a tele (50 mm or more). The larger the focal distance is, the more resembles the image an orthographic projection, and the less significant is the needed perspective distortion. – As explained next, significant depth disparity is most vital for any photogrammetric computation.



Figure 3: Artificial diffuse illumination. The paper cylinder is illuminated from outside to obtain as diffuse illumination as possible in its interior.

Essentials of Dense Matching

Knowing some basic facts about photogrammetry is required for understanding the issues with acquisition of photo sequences. First of all, the position and orientation (pose) of all cameras is computed through a global optimization process, the *bundle adjustment*. It computes 100-200 recognizable *feature pixels* per image, and it compares the positions of pixels that exhibit the same features. This first phase is followed by a second that is the core of 3D-reconstruction from images, *dense matching*. Every pair of successive photos P_1, P_2 in an image sequence is compared pixel by pixel as follows. The positions c_1 and c_2 of the two cameras (optical centers) are two points in space. Together with a 3rd point p in space that lies on the object, they form a triangle (c_1, c_2, p) which defines a plane in space. When looking through one of the cameras, this plane is seen from the “side”, i.e., it forms a line, the so-called *epipolar line*. If q_1 is the pixel in image P_1 showing point p of the object, then the epipolar line $L_1(p)$ goes through this q_1 . And if p projects on another pixel q_2 in P_2 , then this yields another epipolar line $L_2(p)$. The color values of the pixels along $L_1(p)$ and $L_2(p)$ are very similar. A comparison of the “horizontal” displacements allows, like with the disparity of the human eyes, for computing the position of p and, consequently, of the depth (z-coordinate) of both pixels q_1 and q_2 .

When this process is carried out for all epipolar lines, the result is one depth value (z -coordinate) for every pixel in every image of the sequence.

Implications for 3D reconstruction

One immediate consequence of the method is that the worst case is in fact a perfect white wall: All pixels along all epipolar lines have the same color, so no disparity at all can be detected. This makes the method especially well suited for Cultural Heritage, since most old objects have rich texture. Matte objects (mud) can be reconstructed much better than shiny materials (metal): Highlights are view dependent, which creates fake color correspondences on epipolar lines of consecutive images, which results in wrong depth values. The computed depth information is most reliable at sharp boundaries of differently colored regions. And of course, blurred but also uniformly colored regions yield bad or no results at all.



Figure 4: Turntable problem. Nearby and far away object parts are out of focus, the background and specular highlights confuse the feature extraction

Fig. 4 (left) shows a particularly bad case, a turntable sequence. In this case the object has a high depth. In order to fill the whole image, the camera needs to get too close to allow for the whole object to be in focus; the closest and farthest parts are blurred. This is even the case at the border of the cup since the background of the cup is already blurred. The problem with turntables is that the image features are contradictory: The background stays in place while the turntable and the object move. So the bundle adjustment fails.

Figure 4 (right) shows the typical effect of non-diffuse illumination. The highlights on the left and right on the bronze surface stay in place when the object is turned, so the color comparison along the epipolar lines fails to compute the right depth.

This can be avoided if, like in Fig. 3, the object is put on a table and the camera is moved around the object, instead of moving the object itself.

Camera variation in all x , y , and z is mandatory

A not so obvious fact is that the computation of the optical centers requires a variation of the camera position in all three spatial directions over the sequence. A circle, for instance, typically varies only in x and y , but not in z . So if all images are taken with the camera on exactly the same height, all optical centers lie on one plane in space. In this case the bundle adjustment computation of the pose must fail, and the sequence

cannot be reconstructed. When using a handheld camera, the small variations of the human hand give sufficient variation in x , y , and z to compute the optical centers even when walking along a line. When using a tripod, care must be taken to assure that there is some variation in all 3 spatial directions.

All in all it must be acknowledged that taking image sequences requires a bit of training. Most effects can be explained when the implications of the method are understood. With some experience, it is in most cases easy to avoid the pitfalls and to obtain really good reconstruction results.

Gipsmuseum trick: Artificial texture

In the Gipsmuseum campaign we were faced with the problem that the plaster surface of many statues was too perfect, i.e., the statues were too white. Our workaround was to apply a random texture to the surface artificially using a projector (Fig. 5).

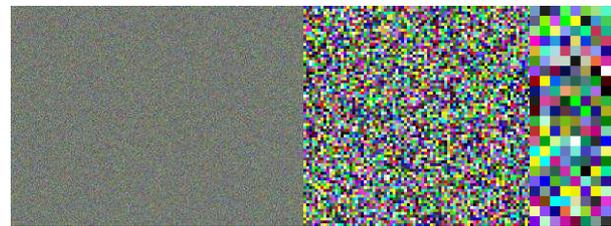


Figure 5: The 1400×1050 random pattern with a uniform noise distribution. The image resolution is half of the camera resolution, both in x and y (1.5 MPixel vs. 6 MPixel), so one projector pixel corresponds to more than one camera pixel.

3. THE PROCESSING PHASE

Each day of an acquisition campaign typically results in 300-500 fotos. These are first stored on a file server as raw data, then they are sorted into sequences. For the Gipsmuseum campaign, we have created one directory for every statue, named after the statue ID on the plate. The plate of statue 83 for instance reads

83. Athena, so-called Lemnia, Dresden, Albertinum, Roman marble copy of a Greek original (from Phidias?), around 450/440 b.c.

So the corresponding directory is GM083, GM standing for GipsMuseum. For every campaign we invent such a two-character short. The GM083 directory is structured as shown in Table 1.

The different files correspond to the individual processing steps. In the following we describe the workflow in more detail.

Sequence preparation

- Store all images from the day into a raw data directory, e.g., GM083/Rawdata/2008-07-23. Raw data are archived and never changed.
- Split up the images belonging to one statue into sequences. Review each individual image in full resolution to remove

the bad images: blurred, over-/under-saturation, parts missing,...

- Send each sequence to the ARC3D web service using the upload client, labeled e.g. as: User scene: GM083, Academic Reference: CGV, Sequence: 01
- After several hours you receive an email with an ftp location of the reconstruction zip to download it to the sequence directory

The Arc3D webservice is free for CH users

The Arc3D webservice (www.arc3d.be) is a great service provided by the group of Prof. Luc van Gool from the Katholieke Universiteit Leuven (Belgium). Several hundred CPU cores are available for performing the dense reconstruction of uploaded image sequences. The service can be used free of charge for non-commercial applications, it is only required to register. The only limitations are that uploaded images must be from the domain of Cultural Heritage, and that KU Leuven reserves the right to archive and use the uploaded image sequences for academic research.

GM083/Rawdata /2008-07-23	63 MB	All photographs taken on a specific date for this statue
GM083-01/DSC_0055.JPG GM083-01/DSC_0078.JPG	13 MB	Images for sequence 01 of statue GM083, only the good photos, not unsharp or shaken, over/under saturated, parts missing, ...
GM083-01/GM083-01.zip	210 MB	zip archive with range maps obtained from ARC3D
GM083-01/GM083-01_c.ply	91 MB	Step 1: Multilayer mesh
GM083-01/GM083-01_cc.ply	75 MB	Step 2: Multilayer mesh interactively cleaned
GM083-01/GM083-01_ccp.ply	17 MB	Step 3: Poisson reconstruction single layer mesh
GM083-01/GM083-01_ccpc.ply	16 MB	Step 4: Poisson reconstruction cleaned
GM083-01/GM083-01_ccpcs.ply	11 MB	Step 5: Poisson reconstruction cleaned and simplified
GM083-01/GM083-01_ccpsc.ply	21 MB	Step 6: Color information re-applied - final result
GM083-01/GM083-01.metadata	3 KB	Workflow information in XML format

Table 1: File structure for the 3D-reconstruction workflow, with file sizes and workflow information

Description of the Arc3D data

The zip archive obtained from ARC3D contains for each and every image:

- **Pose:** position of the optical center (focal point in world coordinates) and orientation, computed using bundle adjustment. This is also called external camera calibration.
- **Internal calibration:** A few parameters describing the radial distortion of the lens and the deviation of the optical center from the image center
- **Range map:** A 2D floating point number grid with the same resolution as the photo. It gives for each pixel the distance from the focal point, which is obtained using dense matching
- **Confidence map:** 2D grid of integer numbers with the same resolution as the photo. For each pixel this is counts on how many photos this same surface point appears. It is a measure for the reliability of the depth value.
- **.v3d info file:** this is a list in XML format of the individual files mentioned above

A 6 MP camera obtains in principle 6 million vertices and consequently 12 million triangles for each range map, since each quadrangular pixel is split into two triangles. A sequence has

typically 20-50 photos, resulting in as many as **240-600 million triangles**, but with very high redundancy: Each surface point is typically contained in 5-15 photos, so the surface is composed of many layers residing at the same position in space.

This multi-layered surface has to be reduced to a single layer (*sequence merge*), resulting in one single-layered mesh per sequence. Finally, all sequences of one statue have to be merged together to obtain one coherent, integrated mesh for this statue. This final step is difficult to perform with standard software since the different sequence meshes all have different scales: The ARC3D webservice has no information at all about absolute scale. Therefore each sequence is scaled differently, basically with respect to the focal length of the camera, which is the only "absolute" value that is available at this stage. So before merging the sequences together, they have to be re-scaled to a common scale.

The Processing Workflow

In the following, the six processing steps are described in more detail. They are illustrated in Figure 7. Since many different processing options exist, and new ones are being developed, the standard workflow used for the Gipsmuseum will be subject to change in the future. It is nevertheless worth describing since it was successfully applied for reconstructing as many as 257 different sequences that were acquired for the 24 statues.

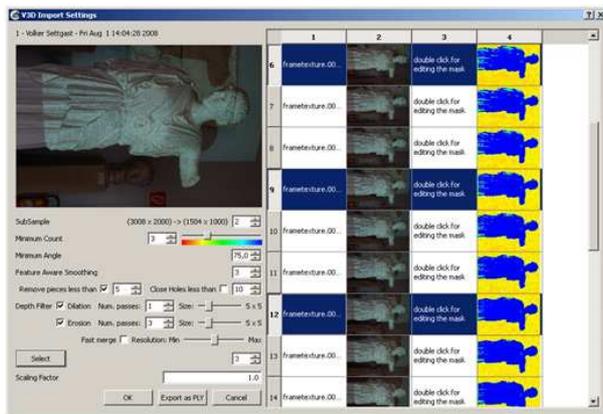


Figure 6: Meshlab opening a .v3d file. On the right, every 3rd image of the sequence is selected, blue parts indicate high quality. Required confidence, grazing angle and sub-sampling are the most important parameters for mesh generation.

The zip file obtained from Arc3D is opened using *Meshlab* to which the .v3d extension is associated on our computers. *Meshlab* is the “swiss army knife” for mesh editing and processing. It is an extensive open source software produced by the group of Roberto Scopigno and Paolo Cignoni from ISTI-CNR in Pisa, Italy. It can inter-operate with Arc3D and read .v3d files.

Step 1: Creating a multi-layer mesh

When opening a .v3d file, *Meshlab* offers a dialogue box with many options for the creation of a multi-layer mesh (Fig. 6). Typically only a small subset of the images is selected for conversion to a mesh layer, for instance every 4th or 5th image of a sequence of 25 images. Because of this limitation good data are potentially ignored. But each range map increases the file size by 20-30 MB, and requires up to 80 MB of RAM. Currently no more than 8 layers can reasonably be processed since on 32 bit Windows, a single process may not use more than 1 GByte of RAM. This is expected to be resolved when 64 bit *Meshlab* becomes available.

Step 2: Manually cleaning the multi-layer mesh

The multilayer mesh must be cleaned because the range maps sometimes contain background parts. The unwanted parts of all meshes can be marked in parallel and deleted, so this is in fact a sequence of interactive move-select-delete actions. *Meshlab* offers many filters for mesh operation. In particular, at the end the filter “*Remove unreferenced vertex*” must be issued since by default, only the indices to the point list are deleted when deleting a triangle, not the points themselves.

Step 3: Poisson reconstruction

Many algorithms have been developed for merging multiple surface layers into a single layer mesh. Each of the methods has its pros and cons, and many work only in an out-of-core fashion. The Poisson reconstruction offers a good compromise. Its disadvantage is the unavoidable smoothing effect (which “washes out” small detail), but this is at the same time also an

advantage since it can cope with very noisy data. A great advantage is that it produces a connected manifold mesh (“water tight”). The Poisson method “blows up a balloon” whose surface snaps to data points. Thus, the resulting mesh is closed by construction. It has high resolution where data points are present but low resolution in the balloon parts. Its surface is sampled at regularly spaced intervals using the well known *marching cube* method, which divides the cell size recursively by a factor of 2 in surface regions with high detail. However, the Poisson mesh resolution is not directly coupled to the resolution of the input mesh, i.e., the range map. With respect to resolution, the limiting factor of this method is the number of these cell divisions, the so-called *octree depth*. An octree depth of 10 gives acceptable results, but crashes sometimes due to RAM limitations. A depth of 11 would be great, but it typically consumes four times as much RAM as depth 10.

Step 4: Cleaning the Poisson mesh

The balloon is cut away interactively, which is easy in most cases since *Meshlab* offers a filter to *remove faces with edges longer than a percentage of the model* (e.g., 0.32%). Exploiting the fact that in regions without data the Poisson mesh has very large triangles, these can be selected and deleted easily. However, the “real” boundaries cannot be detected reliably using this technique, so we additionally apply twice the filter *remove border faces*. Finally, we use *Remove isolated pieces (wrt. Facenum)* to delete clusters < 25 triangles and then do *Remove unreferenced vertex* once in the end.

Step 5: Simplification

Besides a long running time of $O(n^2)$ the marching cubes method has the drawback of creating many irrelevant small faces, and also faces that have one edge that is much shorter than the other two (*lengthy* triangles). They are an artifact of the cubic octree grid. A reduction by 50% remove bad faces without throwing away any useful information. This sort of simplification is offered by the *Quadric edge collapse decimation* filter.

Step 6: Adding color

The Poisson reconstruction in *Meshlab* creates a mesh without any color or material information. We have created a tool to transfer colors from one mesh to another. However, it uses vertex colors instead of a bitmap texture. Thus, the color resolution is actually identical to the mesh resolution. This creates annoying artefacts in surface regions with simple geometry (e.g., flat spots) but complex texture (e.g., written text).

Question: Why storing intermediate results?

Table 1 lists the different files in the sequence directories. This shows that the intermediate files for all processing steps are kept despite of their size. The reason is a conservative strategy: Since we are not sure which files may be used later on, we keep them all for the time being. We can for instance re-do parts of the workflow when new new algorithms or hardware become available. With *Meshlab* 64-bit we might, for instance, re-do all sequences with an octree depth of only 9. – In order to do so, however, we need to make use of the documentation of the workflow that we have collected.

Obtaining the final model

In order to obtain the final model we currently have to resort to commercial engineering software for professional post processing of 3D scans, namely Geomagic Studio from Raindrop Geomagic. The feature that is missing in Meshlab is the alignment of differently scaled model parts. The iterative closest pair (ICP) algorithm in Meshlab assumes that the parts

have the same scaling, which is a true assumption for models from laser scanning. For photogrammetry, there is no absolute scale a priori.

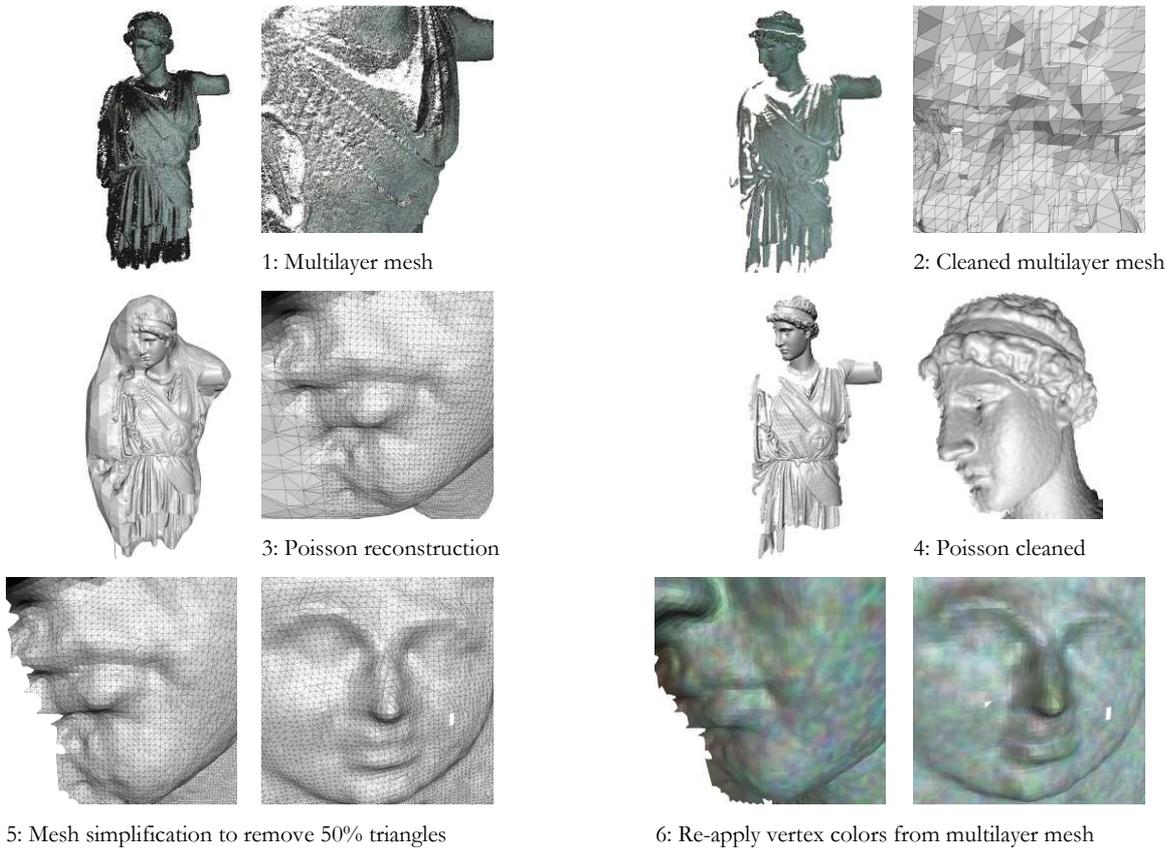


Figure 7: Six typical Meshlab processing steps for reconstructing a 3D mesh from a series of range maps

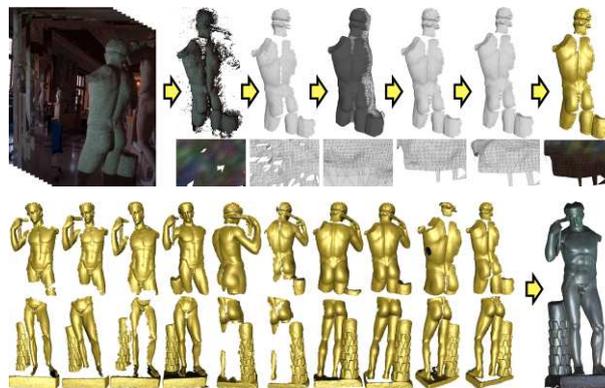


Figure 8: Six standard processing steps are carried out for each sequence (top). But the GM100 statue was assembled from 20 such sequences (middle)! Bottom: The resulting model with 1.5 MTriangles (751 KVertices) is nice but too smooth.

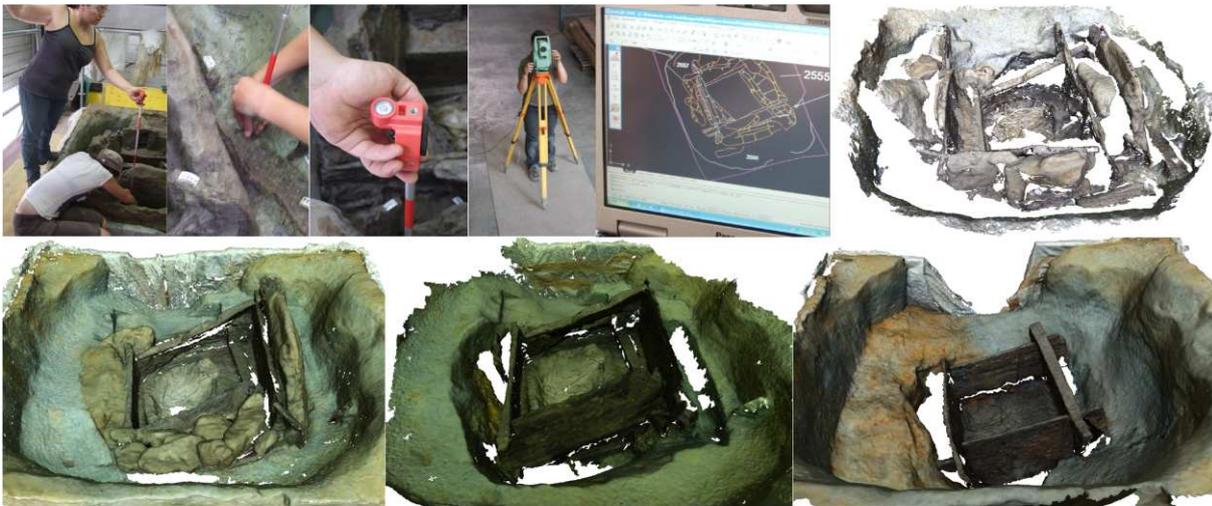


Figure 9: Conventional archeological documentation (top row): A measurement rig with a reflective marker must be kept exactly vertical, then its position is measured as a single point using a total station on a reference position, thus obtaining a point sequence (10 points/minute). The result is a collection of line sequences that are recorded over 2 days using a laptop. – The much more dense 3D reconstruction (top right) was acquired in 5 minutes using a handheld compact camera (Fiji F30). The models in the bottom row were acquired in a more serious way using the 6 MPixel Nikon on a tripod (60 images each).

The ICP algorithm from Geomagic Studio does not take scaling into account either, but it is easier to use (1-click alignment). We currently scale manually, align, look at the misalignment, scale again, and so forth. With a bit of experience, this manual iteration process “converges” after four or five iterations. The result is acceptable only from an aesthetic point of view (Figure 8), but of course this is not a satisfactory engineering solution.

Besides improving ICP another viable solution could be to integrate measurement targets that are scanned together with the object. However, we have not explored this approach so far.

4. CRITICAL DISCUSSION

The described workflow has the same disadvantage as all other post-processing workflows, namely that data gaps are detected too late. With GM100, this is obvious at the top of the head (Fig. 8). The remedy is simple: Go back and acquire additional sequences from the statue. In most cases this implies much additional cost since an acquisition campaign requires expensive travel, obtaining access permissions, and the hardware setup is in most cases not a trivial thing either.

The GM083 example from before is even a much more serious case, since only eight sequences were captured, and during reconstruction it turned out that they provide insufficient overlap to allow for mutual registration. None of the many possible orders that were tried to align one part with another could be successfully continued. So unfortunately, in the end no model could be produced in this case at all (see Figure 10).



Figure 10: The eight parts of GM083 unfortunately provide insufficient overlap for 3D reconstruction

The great vision: Automatic 3D-reconstruction

The described 3D-reconstruction workflow still involves many manual steps, and the resulting models are typically not comparable in terms of quality to models produced by 3D scanning.

However, it is important to mention that the method as such has huge potential for further development and optimization. This becomes especially apparent when considering the state of the art in archeological documentation (Fig. 9).

In the case of the prehistoric well at Wohlsdorf, a drastic speed-up could still be realized, and a much more complete and meaningful model was produced in a fraction of the time and cost required for scanning – which ranges in the area of 1000 Euros for one day of scanning and post-production done by a professional measurement engineer. The model in the top right of Fig. 9 was acquired in five minutes using a handheld compact camera followed by about 1 hour of post processing. The archeologists were enthusiastic about the result.

Potential for optimization: Image acquisition

The acquisition could be dramatically simplified with cleverly constructed camera rigs. Instead of moving a conventional SLR camera on a tripod, a shutterless high-quality industry camera

could be moved using servo controlled step motors along a pre-defined trajectory. The lighting situation could be captured along with the images using photographs of a polished steel ball, in order to cope with highlights moving over the surface.

The need for more faithful sequence processing

Maybe the biggest drawback of the presented workflow is that the triangles produced by Poisson reconstruction have no direct relation to the input triangles (2 triangles/pixel). It is therefore quite difficult to assess the authenticity of the resulting mesh, as there is no direct relation between input data and output mesh. The deviation of the reconstructed surface from the input photographs can only be measured using a posteriori analysis.

Furthermore, Poisson reconstruction averages at each surface point all available mesh layers. This is a disaster for sharp features, e.g., corners or creases in the surface. Even if each individual layer has a nice sharp crease, averaging all layers smoothes away the detail even before Poisson washes it out.

This and the sparse selection of layers (e.g., 8 out of 25) make that only a tiny fraction of the highly redundant captured information finds its way into the final model. For the 20 sequences of the GM100 statue (Fig. 8) for instance, in total 577 images were taken with 6 MPixel each, which amounts to 3.46 GPixel or 7 GTriangles. Even taking the high redundancy and the background pixels into account, the 750K vertices of the model are in fact only 0.1 promille of the input data.

Potential for automatic sequence reconstruction

We are certain that individual image sequences could be reconstructed in a completely automatic way in a much higher

ACKNOWLEDGEMENTS

We gratefully acknowledge the funding from the European Commission for the FP7-IP 3D-COFORM under grant No. 231809. With this support, we are confident to provide solutions for the mentioned problems soon.

REFERENCES

- CROFTS N., DOERR M., GILL T., STEAD S., STIFF M.: *Definition of the CIDOC Conceptual Reference Model*, version 4.2 ed. CIDOC Documentation Standards Working Group, June 2005. Also *ISO/PRF 21127*, available from cidoc.ics.forth.gr.
- LONDON CHARTER INITIATIVE (HUGH DENARD): *The london charter*, June 2006. www.londoncharter.org.
- PAN, X., BECKMANN, P., HAVEMANN, S., TZOMPANAKI, K., DOERR, M., FELLNER, D.W., *A distributed Object Repository for Cultural Heritage*, Proc. VAST 2010 conference, Eurographics Press, 2010

resolution than today. A careful analysis of the 6-stage workflow revealed that all parts of the work that are carried out manually are rather schematic. So the chances of finding algorithmic solutions are good.

A greater challenge is the assembly of a complete model from the parts. We envisage reducing this problem to the sequence processing problem. The idea is to introduce overview and detail sequences. An overview sequence captures the complete model, but in low resolution. A number of detail sequences captures only parts, but these in higher resolution. Now an additional bundle adjustment step could use image features to relate the detail sequences to the overview sequence. This way the 4x4 matrices for the transformation from detail to overview coordinate systems could be obtained.

5. CONCLUSION

This paper has presented a practical recipe for the reconstruction of 3D models from image sequences. It uses state of the art tools, most of which are available for free to Cultural Heritage professionals. Besides pointing out the great potential for CH documentation we have also presented a critical assessment and highlighted ideas with a huge potential for improvement.

Our greatest hope is that we could stimulate a wider take-up of this great technology in the CH community. We firmly believe that in a few years time, all mentioned problems will be solved.

Until then remember: Take many many images!