

FROM 2D (TO 3D) TO 2.5D – NOT ALL GRIDDED DIGITAL SURFACES ARE CREATED EQUALLY

Geert J. Verhoeven^{1,*}; Markus Santner²; Immo Trinks³

¹ Ludwig Boltzmann Institute for Archaeological Prospection & Virtual Archaeology (LBI ArchPro), Franz-Klein-Gasse 1, 1190 Vienna, Austria; Geert.Verhoeven@archpro.lbg.ac.at

² Dresden University of Fine Arts, Güntzstrasse 34, 01307 Dresden, Germany; santner@hfbk-dresden.de

³ Vienna Institute for Archaeological Science (VIAS), University of Vienna, Franz-Klein-Gasse 1, 1190 Vienna, Austria; immo.trinks@univie.ac.at

KEY WORDS: Depth map, Digital surface model, Image-based modelling, Metashape, Mural painting, Polymesh, Rasterisation

ABSTRACT:

The surface of most heritage objects holds important clues about their creation. To answer specific research questions about a 16th-century mural painting located in the Bischofstor of Vienna's St. Stephen's Cathedral, the three-dimensional (3D) geometry of the entire painted surface was digitised in minuscule detail using thousands of overlapping photographs. Although this article provides image acquisition and processing specifics, it aims to assess which image-based modelling workflow can achieve the most detailed, noise-free, two-and-a-half dimensional (2.5D) raster surface of this mural painting. Other than their full 3D counterparts and in contrast to the focus of most academic research, 2.5D raster surfaces are ideally suited for visualising and analysing sizeable, detailed surfaces. They are, therefore, still the preferred surface encoding of many heritage projects that want to leverage digital surface approximations to further heritage insights (and not just use them as mere eye-catchers). In the end, only a combination of different 2.5D rasters was able to accurately represent the variable surface of this mural painting with the right amount of spatial detail.

1. INTRODUCTION

1.1 Image-based modelling and digital surfaces

Over the past decades, many different and application-specific surface representations have been developed in the fields of CG (Computer Graphics), CAD (Computer-Aided Design), manufacturing, gaming, GIS (Geographic Information System), or the medical world. These representations can be spatially two, two-and-a-half, or three dimensional (2D, 2.5D, and 3D, respectively), encode the surface continuously or discretely, follow an implicit or explicit mathematical formulation, and utilise a wide variety of data formats. This deep pool of surface representation technologies results from the different types of object surfaces that are around (e.g. smooth versus rough, organic versus analytical shapes) and the special uses these surfaces can have: animation, simulation, gaming, digital preservation, geographical and structural analysis, illustration or manufacturing. Consequently, digital surfaces need suitable data structures, just like in any other field of computer science.

The two main workflows to get a digital surface encoding are to build it from scratch in a CAD, CG, or GIS environment, or to digitise the surface geometry of an existing physical object or scene. Amongst the many application-driven techniques developed by multiple disciplines, one possible approach to surface digitisation is to use a set of overlapping photographs. Because the digital construction of an object's or scene's form and placement is called a (geometric) model and this creation task known as (computer) (geometric) modelling (Hess, 2010), using photographs/images to digitise a surface is also denoted Image-Based (3D) (surface) Modelling or simply IBM.

IBM encompasses different techniques, but they all rely on object or scene photographs taken from different locations to extract digital surface data. IBM has been the focus of both the photogrammetric (Kraus, 2007) and computer vision (Szeliski,

2011) fields. However, over the past decade, hybrid photogrammetric computer vision-based approaches (Förstner and Wrobel, 2016) have become commonplace in cultural heritage documentation. With photogrammetric principles at their core, these hybrid approaches mainly rely on the computer vision algorithms Structure from Motion (SfM) and Multi-View Stereo (MVS) to digitally extract surfaces from overlapping images.

1.2 Digital surface terminology

In the photogrammetric world, IBM pipelines usually end with a dense 3D point cloud (Nocerino et al., 2020). Like the point clouds generated by laser scanners, these (generally unstructured) point datasets represent the surface of a particular physical object or scene digitally but discretely in 3D. However, many applications expect a continuous surface representation. When needed, this cloud of 3D points gets interpolated into a continuous triangle- or quad-based 3D surface polygon mesh (polymesh, mesh, or even 'polys' in colloquial CG talk). Such 3D polymeshes have become a typical IBM output in the cultural heritage sector. (Note that this text uses the term polymesh to make a clear distinction with volumetric meshes.)

In geo-sciences, truly 3D digital surfaces are less used, because most GIS software struggles till this very day with their display and analysis (Verhoeven, 2017). In GIS environments, the prevalent representation schemes for continuous surfaces have 2.5 spatial dimensions: TINs (Triangulated/Triangular Irregular Networks) and the more common elevation grids. As a result, most IBM software can also generate such a uniformly spaced grid that records the elevation on a cell-by-cell basis. These 2.5D elevation grids are also known as height fields or height maps in CG. Although a 2.5D data structure has specific disadvantages when used to represent surfaces digitally, height fields are often better suited than full 3D encodings for the fast execution of certain computational methods to visualise (Kokalj and Hesse, 2017) or analyse (Jordan, 2007) surface features.

If continuous, true 3D surface geometry is absolutely needed, then the choice in representation schemes is usually limited to the Non-Uniform Rational Basis Spline (NURBS) surfaces in CAD packages or the SubDivision surfaces (SubD), T-splines and above-mentioned polymeshes primarily found in CG pipelines. Only a few packages bridge between these worlds (see Verhoeven, 2017). However, since all these representations can digitally encode a surface, they are all denoted Digital Surface Model or DSM in this article.

Usually, much of the scholarly literature exclusively reserves the term DSM – or its related cousins Digital Elevation Model (DEM), historic Digital Elevation Model (hDEM), Digital Terrain Model (DTM), Digital Ground Model (DGM), Object Height Model (OHM), Canopy Height Model (CHM), Crop Surface Model (CSM), normalised Digital Surface Model (nDSM) – for a continuous, regular 2.5D grid (Wood, 2008). However, since this acronym does *sensu stricto* not explicitly define the exact digital encoding of the surface elevations (random or systematic, continuous or discrete, raster- or vector-based or even hybrid), it seems more accurate to use DSM for any digital representation of surface elevations: sparse and dense point clouds, contour lines, height fields, NURBS, TINs, T-splines, polymeshes, sweep and SubD surfaces, metaballs, wireframes, amongst several others. One just needs to specify what type of DSM is meant. As a note: the use of DEM, DSM, DTM and associated acronyms is also largely unstandardised, but will be the topic of a forthcoming paper.

1.3 Goal: best raster DSM of a mural painting

This article wants to research the different workflows one can commonly follow to derive a 2.5D raster DSM from a set of overlapping 2D images whose interior and exterior orientations are correctly established. More specifically, the article leverages the algorithms embedded in the popular IBM software Agisoft Metashape Professional 1.7.1 (build 11797) to assess the impact of specific MVS steps on the generated 2.5D DSM. This assessment uses a large collection of digital photographs depicting a mural painting located in the *Bischofstor* (Eng. Bishop's Doorway) of the *Stephansdom* (Eng. Saint Stephen's Cathedral) in Vienna, Austria (Figure 1). The mural painting shows a triptych consisting of two painting phases: phase 1 around 1510, phase 2 around circa 1515 to circa 1580 (Kohn, 2001). The older surrounding painting is composed of two wings, each with a standing female Saint above some predella with a putto holding a coat of arms. The younger painting in the centre shows a standing male Saint figure holding a church model. The predella zone below him depicts a kneeling founder and a putto. The mural painting also features two different lime plaster surfaces separated by a striking plaster edge. This lime plaster edge forms the outline of an epitaph (visible on Figure 1B), which was likely mounted during the first painting phase. Although the painted surface contains various large- and small-scale undulations, it can be reasonably well approximated by a height field. This is important because 2.5D raster DSMs effectively discard half a geometrical dimension compared to a full 3D surface encoding, leading to an inevitable loss of geometrical information. However, if the surface lacks quasi-vertical walls, overhangs, and under-cuttings, a height field can often satisfactorily digitally approximate that surface.

Some architectural elements border the mural painting, one of them being a protruding stone cornice above it (see Figure 1). It is impossible for a 2.5D gridded DSM to accurately encode the entire surface of this element since the wavy cornice features an undercut (see Figure 7 on the left). Even without this undercut,

there would be a horizontal stone surface which is considered a vertical wall from the perspective of the painting's plane. Including this architectural element will illustrate how different 2.5D DSM modelling approaches handle unfavourable surfaces.



Figure 1. (A) Photographing from two small scaffolding boards (photo by R. Burszan, adapted by G. Verhoeven). (B) The entire mural painting, whose extent is indicated in yellow on (A).

2. DATA ACQUISITION

2.1 Photographic setup

Photographing the mural painting took place on the 9th of November 2020. Although the painting is located inside the *Stephansdom* souvenir shop, tourists could not hinder the photographic activities because the shop remained closed as a precaution after the Vienna terror attack on the 2nd of November 2020. A Nikon D750 24-megapixel reflex camera was used to acquire conventional, three-band colour photographs digitally representing the amount of reflected visible radiation. The Nikon body was equipped with a Tamron SP 24-70mm F/2.8 Di VC USD lens. The lens' focal length was locked at 24 mm and its focusing ring immobilised with cellophane tape at a focusing distance of about 43 cm (measured from the lens' optical centre). This camera-lens combination was mounted on a stereo bar, with a Godox AD200 Pro pocket flash unit to its left side. The flash unit featured the Godox H200R round flash head with a diffuser dome attached. A Godox X2T radio controller mounted on the D750's hot shoe wirelessly controlled and triggered the flash unit. The flash fired at $1/64^{\text{th}} + 0.5$ stops of its maximal output, resulting in a flash duration of $1/4500$ s. This setup enabled a $1/200$ s exposure time, which ensured that camera-induced motion blur (due to handholding the camera) could not negatively affect the photographs. The latter were captured as 14-bit lossless compressed RAW images at ISO 100. An $f/11$ -aperture provided sufficient depth of field while avoiding unnecessary diffraction blur.

It is worth mentioning that the combination of a powerful flash and short shutter speed cut out any ambient light, which means that there was effectively only one illumination source for image acquisition. The latter is essential to accurately determine the photographs' white balance (see Figure 2). This proper white balance notwithstanding, it must be stressed that the photographs were acquired with 3D modelling rather than colour accuracy in mind. To achieve colour-accurate photographs, one would need a different illumination setup with minimally two tripod-supported flash heads inside large

softboxes. However, such a studio-like illumination setup would immensely complicate and lengthen the image acquisition since one could only photograph the painting via temporary scaffolding featuring narrow wooden boards (see Figure 1A). Moreover, one would risk a sub-optimal image network for 3D surface extraction because the tripod light stands could interfere with camera positioning. These considerations lead to the stereo bar setup with a wirelessly triggered pocket flash on the camera's left side. Figure 2A illustrates that this single flash solution created a noticeable illumination gradient in every photograph, despite using a dome diffuser. Although this gradient was largely removed during the conversion of every RAW photograph into the final JPEG image (Figure 2B; section 3.1), this workflow cannot guarantee colour accuracy.

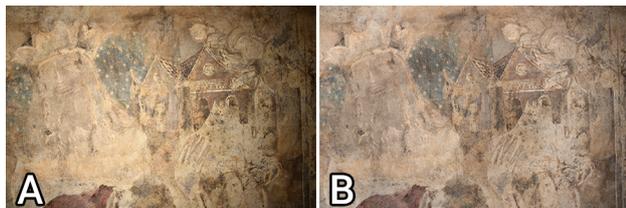


Figure 2. The original photographs' orange tint and illumination gradient (A) are removed in the final images (B).

2.2 Camera network

Many projects that use IBM for heritage digitisation lack an explicit goal other than the mere generation of a 3D model. However, to achieve a goal-oriented 3D or 2.5D digital surface with a specific accuracy, completeness, and level of spatial detail, photographs must be acquired according to guidelines. One of the main parameters is the spatial resolution of the images, as it defines the geometrical details that will be visible in the digital surface. Spatial image resolution is a complicated concept (Verhoeven, 2018), but one of its key determinants is the Ground Sampling Distance or GSD of the images (see Figure 3).

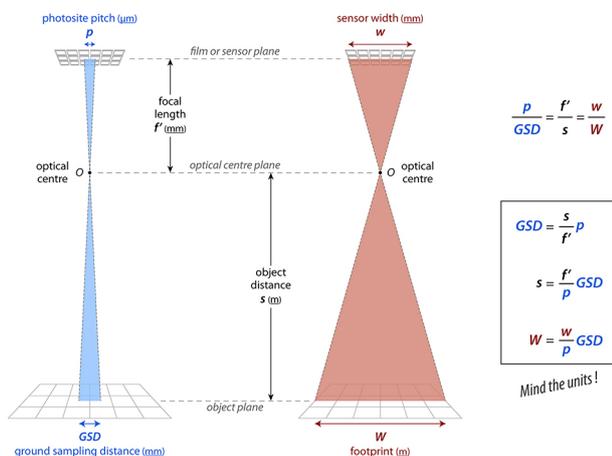


Figure 3. Deriving the purely geometry-based equation that is fundamental to all kinds of optical imaging.

The aim of digitising this painting's surface was to create a raster DSM that could show all geometrical surface details equal to, or larger than, 0.25 mm. Given that two to three pixels are needed to resolve a line feature (i.e. 1 to 1.5 times the Nyquist rate), a 250 μm spatial resolution necessitates a GSD of 125 μm or smaller. With a photosite pitch p of circa 6 μm for the D750's image sensor and a 24 mm lens focal length f' ,

Figure 3 indicates that a 42.9 cm object distance s between the camera's optical centre and the fresco achieves a 0.1 mm (or 107 μm) nadir GSD (which explains the 43 cm focusing distance mentioned above).

Because the image-based digital surface extraction relies on a combination of SfM and MVS approaches, a high image overlap is of the utmost importance. The whole painting was covered by 25 columns of photographs, each counting approximately 100 images (see Figure 4). These 2500 images cover roughly 2.4 m by 3.6 m, which equals circa 92 % in-column (or longitudinal) and 85 % cross-column (or lateral) overlap. Note that these values represent global averages; longitudinal and lateral overlaps between adjacent photographs are not fixed but vary a little throughout the image collection due to the hand-held image acquisition (see Figure 4).

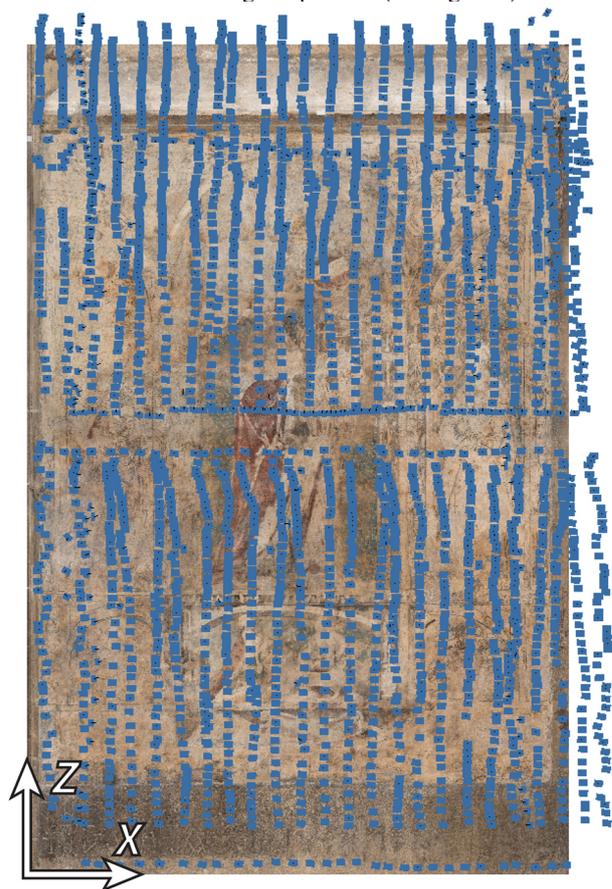


Figure 4. The network of 2790 photographs. The two horizontal gaps indicate the position of the scaffolding boards. Another two columns of photographs to the far left would be desirable, but the scaffolding and bordering wall prevented this.

The camera featured a landscape orientation and had its optical axis perpendicular to the painting for all 2500 photographs. Almost 300 images complemented this image set: convergent images (i.e. with an inclined optical axis) and photographs for which the camera was rotated 180°, 90° clockwise, or 90° anti-clockwise around its optical axis. Some of these images also featured an increased object distance. This change in image scale (*intra*- and *inter*-image) and camera rotation is essential to increase the accuracy of the interior and exterior orientations computed by the SfM algorithm. Figure 4 depicts the network of all 2790 photographs, illustrating both the image density and the slightly varying image baselines.

3. DATA PROCESSING

3.1 RAW development

After tagging all RAW files with metadata, they were 'developed' into 8-bit JPEGs using Adobe's Photoshop Lightroom Classic 10.1. During this conversion, white balancing was performed via a set of photographs depicting an X-Rite ColorChecker Passport Photo. These images also supported the generation of a camera DNG profile with X-Rite's ColorChecker Camera Calibration 2.0.0 software. After applying this profile to all photographs, lens vignetting plus the illumination gradient caused by the single flash unit (see Figure 2) were removed. The geometrical image distortion induced by the lens was left intact, as it gets modelled and taken care of in the SfM step of the IBM pipeline.

3.2 Exterior and interior orientation estimation

Numerous SfM-MVS software packages exist, but this image set was processed using Agisoft Metashape Professional 1.7.1. Within Metashape, the SfM step used a maximum of 40 000 interest points and a 4000-tie point limit for every photograph. The computed exterior orientations can be observed in Figure 4.

Since the output of any SfM algorithm is expressed in an arbitrary coordinate reference system, the sparse point cloud is equivalent to the real-world scene up to a global scaling factor, three rotations, and three translations. An absolute scale was established for the IBM model using four distances measured *in-situ* with a fibreglass tape fully graduated in millimetres. The horizontal, vertical, and diagonal distances respectively exceeded the painting's maximum width, height, and opposite corner distance to minimise scaling errors. A root-mean-square scaling error of 0.8 mm means that one can measure real-world distances on the 3D construction with millimetre accuracy.

Afterwards, the scaled SfM output was translated and rotated until it aligned with a Cartesian XYZ coordinate system having its origin at the painting's lower left. The X-axis runs parallel with the lower part of the painting, while the Z-axis points in the direction of the longest painting side (Figure 4). Thus, the Y-axis indicates the painting's depth. Since the painting's surface is assumed to be almost perfectly vertical, most of the digital surface is located at $Y = 0$.

The scaled, translated, and rotated SfM solution served as input for the subsequent MVS stage. After deactivating all photographs with a larger-than-43 cm object distance or a very inclined optical axis, the different MVS tests were run on the remaining 2684 photographs. The bounding box was set to exclude most of the painting's bordering architectural elements, apart from the stone inscription underneath it and the protruding cornice above it (see Figures 1B and 4).

3.3 Different roads to a 2.5D raster DSM from 2D images

With the interior and exterior camera orientations from the SfM step as input, the dense image matching executed during the MVS step can yield a discrete or continuous digital surface. Different classes of MVS algorithms exist (Aanæs et al., 2016), but Metashape relies on a depth map-based method. Such approaches compute a depth map for each input view and then merge them into a 3D point cloud or a volumetric representation of the scene. Both types of intermediary products can be converted into a 2.5D raster DSM with one or more additional steps. Metashape provides five different ways (see Figure 5) to

end up with a gridded DSM from a set of depth maps (this is, when considering only the relevant options).

Since the depth map generation is slightly different according to the final envisioned product (see section 4), there are more unique ways of getting to these 2.5D raster DSMs. For instance, one could compute a polymesh from depth maps, and a second polymesh from depth maps that were initially extracted to generate a dense point cloud. One assumes both polymeshes to be identical, but they are not. However, the recorded differences are so minor that this workflow (and others that should deliver identical point clouds and polymeshes) are considered equal. The next paragraphs detail these five primary DSM generation methods. Note that the remainder of this article will use "raster DSM" as shorthand for a 2.5D uniformly gridded DSM.

3.3.1 3D point cloud (→ 3D polymesh) → 2.5D raster

The algorithms which Metashape leverages to generate and merge image-specific depth maps are unknown. However, the logs of Metashape suggest that its MVS approach for generating a dense 3D point cloud is similar to the work of Shen (2013). Shen adapted the PatchMatch method for stereo depth map generation by Bleyer et al. (2011) to develop an accurate, photoconsistent and efficient MVS solution for large image sets.

Extracting a dense 3D point cloud from merged depth maps is still the most common approach in the photogrammetric world (Nocerino et al., 2020). When needed, this cloud of 3D points gets interpolated into a continuous polygonal 3D surface mesh, usually constructed from triangles or quads. It is unclear if Metashape's meshing algorithm leverages visibility information, as the software seems to rely on a Screened Poisson surface reconstruction for point cloud meshing. Proposed by Kazhdan and Hoppe (2013), this meshing method incorporates positional constraints into their own, well-known Poisson surface meshing algorithm (Kazhdan et al., 2006) to combat oversmoothing when meshing a point cloud with normals. Such a meshing operation is disjoint from the MVS steps, as the 3D point cloud gets triangulated without applying any visibility constraints or photoconsistency checks. The dense point cloud and its derived polymesh can be converted to a raster DSM (see Figure 5), thus establishing raster DSM creation methods 1 and 2.

3.3.2 3D polymesh → 2.5D raster

In version 1.5.0 (build 7125), Agisoft introduced depth maps-based polymesh generation for Metashape (Agisoft LLC, 2021), thus omitting the need to compute an intermediate point cloud. Because this approach works directly on the depth maps, it might recover finer meshed surface detail from the same set of input images compared to a meshed dense point cloud. It has, however, the disadvantage that all non-masked image regions are constructed in 3D. In contrast, one could first filter, clean, and classify the dense point cloud to, for example, only mesh terrain points.

Similar to the work of Li et al. (2016), Metashape's depth maps-based polymesh construction likely integrates the individual depth maps into an optimised Truncated Signed Distance Function, a volumetric scene representation using a voxel grid. Afterwards, a modified version of the original marching cubes algorithm (Lorensen and Cline, 1987) – likely similar to the one proposed by Splietker and Behnke (2019) – extracts a coherent 3D polymesh representation of the scene's surface. A raster DSM based on the resulting polymesh makes up the third possible pathway (Figure 5). Note that the facet count in the final polymesh is not reduced for this approach (see section 4.2 for some important considerations).

Metashape also offers a separate photoconsistent refinement tool to iteratively recover additional surface detail on any triangular polymesh generated inside the software, like the technique described in Vu et al. (2012). Although one could "refine" the meshed point cloud, the fourth path to a raster DSM constitutes the refinement of the depth maps-based polymesh.

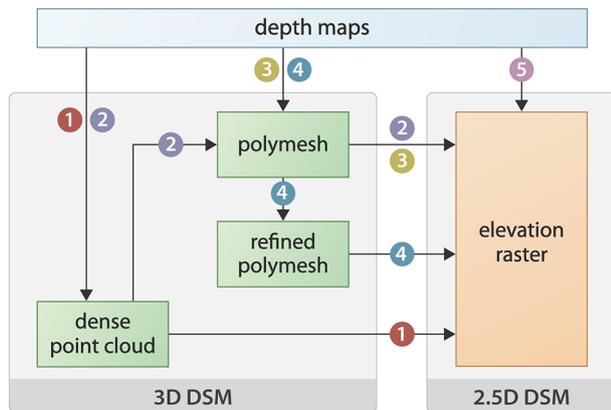


Figure 5. The five relevant pathways one can follow in Agisoft Metashape Professional 1.7.1 to create a 2.5D raster DSM.

3.3.3 2.5D raster directly

Similar to the direct integration of depth maps to generate a polymesh, Metashape can also leverage depth maps to yield a raster DSM since version 1.6.0 (build 9617) (Agisoft LLC, 2021). This is the fifth and last relevant path to a raster DSM (see Figure 5).

3.4 Important note

Before presenting the raster DSMs from these five MVS pipelines, it is important to note that Agisoft Metashape (Professional) uses all image pixels during the SfM step when enabling setting "High", while the MVS step necessitates an "Ultra high" setting since "High" would only utilise 25 % of the pixel count (due to 2x subsampling in image width and height). Failing to leverage all image pixels in the MVS step would render any of the pre-determined GSD requirements invalid. With the entire pixel count as input, the raster cell size of the DSM should approximate the 0.1 mm GSD of the images. This is essential, because the main idea is to end up with a digital surface that conveys all surface details of 0.25 mm or larger.

4. TESTS AND COMPARISONS

This comparison of different roads to a raster DSM is only valid if Metashape's depths maps are always identically generated, irrespective of the product they are intended for. This was tested by computing a 3D point cloud, a 3D polymesh and a 2.5D raster DSM from depth maps. The stored depth maps of the first two approaches were then also used to generate a raster DSM. Mutual subtraction of these three rasters in Global Mapper 19 resulted in difference surfaces with values usually below 50 µm, indicating that the three raster DSMs can be considered quasi, but not entirely, identical. These small variations were also attested when comparing 32-bit floating-point TIFF versions of the three depth maps from the same image. Moreover, a visualisation of these differences highlights their spatially variant nature (see Figure 6). When faced with these results, Agisoft ensured that their depth map algorithm is deterministic, explaining that these minor differences are likely caused by the order of arithmetic operations inside a parallelised computing architecture (pers. comm. Dmitry Semyonov, Agisoft LLC).

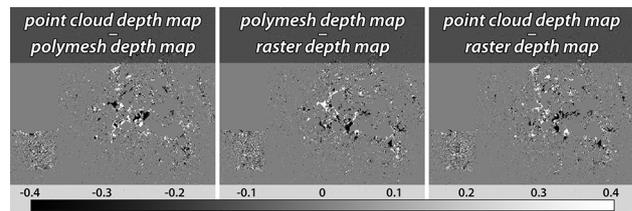


Figure 6. Visualisation of the differences between 32-bit floating-point depth maps (of the same image) computed to generate a point cloud, a polymesh, and a raster.

Even though Metashape produces slightly different depth maps depending on the product they are intended for, the differences are visually almost indiscernible. They are, therefore, ignored in this study.

4.1 Each workflow has pros and cons

Applying the approaches mentioned above yielded five different raster DSMs from the central upper part of the mural painting (Figure 7). Restricting the comparison to this zone ensured reduced processing times. Because the area includes parts of the relatively flat mural painting, the wavy stone cornice with its undercut (see the outer left inset of Figure 7 for its vertical profile), and the small stone ridge below the cornice, these test results are still representative for the complete scene.

A few things can be readily noted when observing the first Region Of Interest (ROI 1). Both the point cloud-based (method 1) and depth maps-based (method 5) DSMs have issues in the most protruding part of the cornice and vertical surfaces of the ridge. DSM 1 contains much noise in both locations, while DSM 5 contains noise and lacks the lower part of the cornice. There are three different surfaces at the level of the cornice undercut (i.e. three different elevation values) that have to be 'summarised' in the raster DSM. Metashape prioritises the lowest elevation values so that the protruding part of the cornice gets partly removed in favour of the wall surface behind it. The polymesh-based approaches (methods 2, 3, and 4) convey the front of the stone cornice and the ridge much better. The horizontal, vertical, and slanting surfaces are well-defined and clean; only the raster DSM of the meshed dense point cloud (i.e. method 2) still contains a few noisy locations.

However, a closer look at the raster DSMs in ROI 2 reveals that the three polymesh-based methods encode less spatial detail than the other methods. The fuzziest, least detailed DSM results from method 3. Although it might be surprising that a raster DSM extracted from the depth maps-based polymesh features fewer surface details than a raster DSM from a triangulated dense point cloud (method 2), these results are in line with the assessment of polymesh generation methods reported by Nocerino et al. (2020). Moreover, Metashape estimated a 69 µm raster cell size for method 3, about 30 % smaller than the other methods. This finer grid only increased the file size, not the amount of digital surface detail.

The latter is sometimes slightly increased when deriving the raster DSM from the refined polymesh (i.e. method 4). Metashape's photoconsistent refinement should iteratively recover additional surface details on a polymesh. However, the results of the mural painting show that this processing step is not effortless and leads to variable results. First, refining a polymesh based on the total pixel count (i.e. "Quality" setting "Ultra high") was not possible in Metashape 1.7.1, so the results in this paper were generated with version 1.7.2 (build 12070).

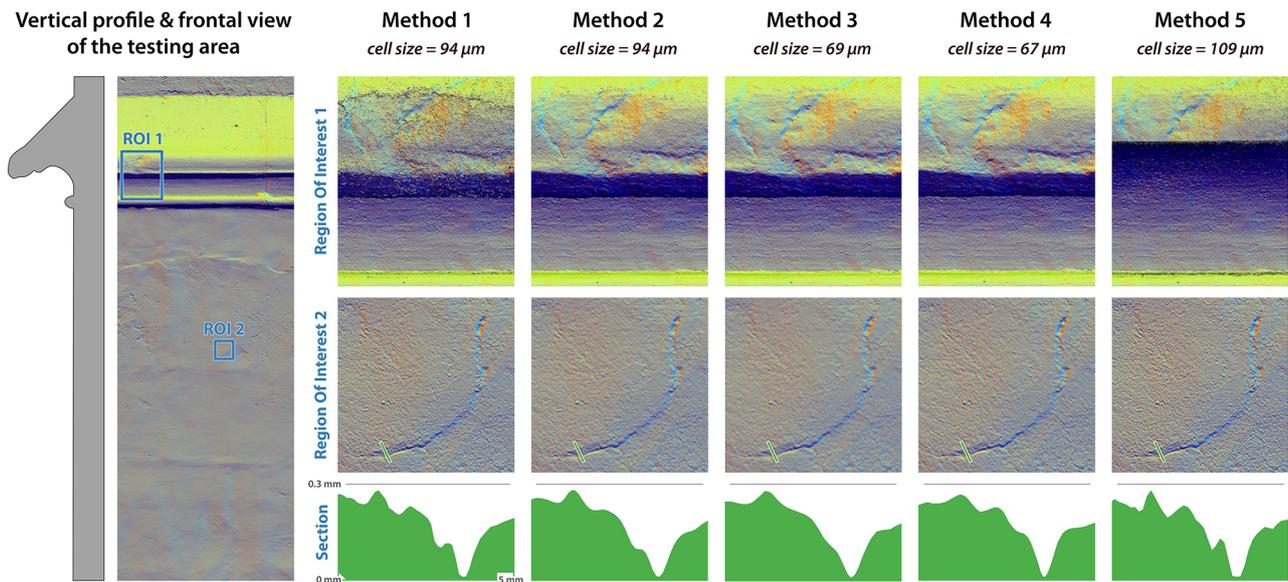


Figure 7. Comparison of raster DSMs resulting from five different methods. The DSM is visualised using a multiple hillshade technique: a composite of three different hillshades, computed with a 35° elevation for the illumination source and azimuths of 315°, 22.5°, and 90° for the red, green, and blue channels, respectively. The multiple hillshades were linearly histogram stretched to increase contrast. All visualisations are computed with the Relief Visualization Toolbox 2.2.1 (Zakšek et al., 2011).

Second, polymesh refinement is highly time and video memory (i.e. VRAM) intensive. The user also must balance the iteration count (more iterations might recover more surface detail at the expense of processing time) with a "smoothness" parameter (less smoothing might recover more surface detail at the expense of noise). Here, all refinements used five iterations and a 0.5 smoothness factor.

Third, refinement results largely depend on the number of input images. The cleaner and more detailed surface geometry visible in Figure 7 (method 4, ROI 2 and section) was only achievable after deactivating 80 % of the images. Upon refinement with all images, barely any additional surface detail could be recovered. Finally, including slanted views in the polymesh refinement often increases the surface noise, which is likely related to the increased blur circles of the objects points.

Overall, polymesh refinement can yield cleaner raster DSMs with improved surface detail compared to other polymesh-based rasterisations. However, the issues mentioned above prevent it from being a no-brainer. In addition, methods 1 and 5 still outperform method 4 in terms of derived surface details.

4.2 Avoid polymesh decimation

Polymeshes with 10s to 100s of millions of polygons will choke many downstream applications like visualisation engines or mesh-repair software. Most IBM software, therefore, features automatic or partial tuneable polymesh decimation functions, although standalone applications with more intelligent ways to reduce excessive polygon counts exist as well. Despite reducing the number of polymesh facets, decimation processes should also preserve the surface model's essential morphology to some extent (the extent being goal-dependent).

Metashape offers three automatic decimation methods, whereby a "Face count" setting "High" yields the highest polygon count to approximate the initial undecimated polymesh. Since it is activated by default (besides being the "best possible" auto-decimation setting in the software), it is reasonable to assume that most polymeshes generated in Metashape have undergone

this non-neglectable amount of decimation. Although the resulting polymesh might still be appropriate for the intended purpose, Figure 8 illustrates that it is unsuitable as basis for a raster DSM for which the cell size should equal the image GSD.

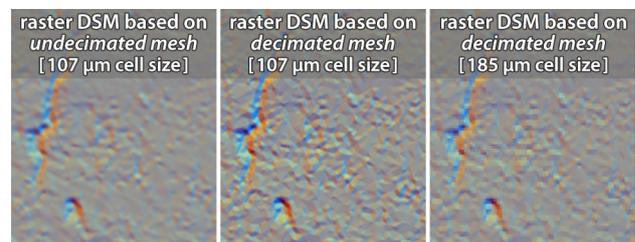


Figure 8. A raster DSM derived from an undecimated polymesh (left) versus its auto-decimated version (centre and right). The latter two rasters have cell sizes that correspond to the image GSD and the size proposed by Metashape, respectively.

After auto-decimating the polymesh, Metashape proposes to use a cell size for the raster DSM of 185 μm (Figure 8, right inset), thereby indicating that quite some surface information has been lost. This is effectively visualised when turning the decimated polymesh into a raster DSM with a 107 μm cell size. At that moment, the fine raster starts to encode the edges of the decimated mesh polygons, which show up as surface artefacts (see the central inset of Figure 8). Although a simple auto-decimation might still have its place in specific workflows, its use should be discouraged when a raster DSM is the final goal. Enormous polymeshes might be the result, but these can – if not needed as a product – be avoided by going the dense point cloud- or depth maps-based raster DSM route.

On a positive note, Metashape's suggested rasterisation setting effectively accounts for this reduced amount of surface detail, as a raster DSM with the proposed cell size (Figure 8, right inset) does barely suffer from these artefacts. However, surface features are less detailed than those of a raster DSM based on the undecimated polymesh (Figure 8, left inset).

5. DISCUSSION

5.1 Towards a final 2.5D raster... and beyond

The direct depth maps-based approach provides the most spatially resolved 2.5D digitisation of the painted surface. On the other hand, the surrounding architectural elements are encoded the cleanest in the polymesh-from-depth-maps-pipeline. The ideal raster DSM is thus a combination of both results. To this end, one could extract the best part of each raster DSM in GIS and patch them together. However, the result would feature a well-visible seam where the DSMs meet. This was avoided by smoothly blending both raster DSMs in Adobe Photoshop 22 for which the Avenza Geographic Imager 6.3 plugin enabled reading and writing support (Figure 9). The final raster DSM features 22.1k cells in width by 34.5k cells in height, equivalent to approximately 762 megapixels.

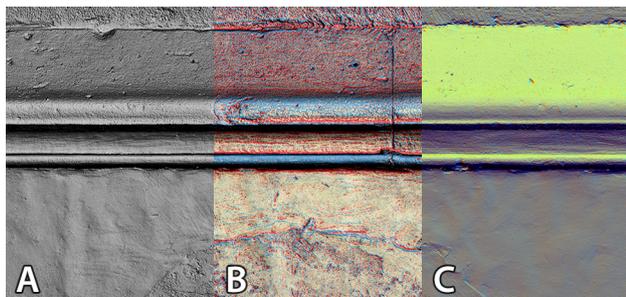


Figure 9. Three visualisations of the upper central part of the final raster DSM: A) contrast-enhanced omnidirectional hillshade (30° elevated light source); B) red-yellow-blue colour-coded local relief model (radius 30 cells, [-80, 80] μm cut-off values] fused with the omnidirectional hillshade of A; C) multiple hillshade (see Figure 7 for the parameters). Note that the seam between both raster sources is undetectable.

Even though a quest for the best possible digital surface approximation is an interesting academic feat, it does next to nothing to understand the cultural heritage asset it represents. Digital surface data do not speak for themselves; they only get purpose if they are a means to digitally preserve the surface geometry of a heritage asset or facilitate its analysis. What concerns the 16th-century mural painting, the detailed raster DSM has opened new cultural-historical research avenues, leading to insights that could not be gained with more common methods of painting research. Those results are presented in Verhoeven et al. (2021) while Figure 9 displays three different DSM visualisations that helped obtain these new insights.

5.2 Lack of scholarly focus on 2.5D surfaces

Although visualising and (spatially) analysing the resulting digital surface might be primordial, there is still a need to get better, more robust 2.5D extraction methods. Compared to the focus on 3D surface generation, there is a surprising lack of scholarly focus on 2.5D raster surface creation. This underrepresentation is even more striking upon realising that many (cultural heritage) mapping projects and most spatial analyses are still mainly 2D and 2.5D in nature. At present, neither commercial nor open-source GIS packages can natively work well with large polymeshes featuring real-world coordinates.

For a functional 3D package that treats the third spatial dimension (elevation or height) equally to the two horizontal ones, one must resort to the CAD, CG and gaming worlds and their various data structures. If the latter is a polymesh, then

also CAD software must be ruled out (apart from exceptions like Rhino by Robert Mc Neel & Associates – Verhoeven, 2017). This leaves only CG software and gaming engines as valid candidates, but these packages seldom go beyond the mere visualisation and modification of polymeshes, featuring rather underwhelming spatial analysis and mapping tools. So, either the 3D-capable tools must drastically improve, or the generation of 2.5D raster surfaces must become more robust and tuneable.

5.3 The risk of over-assessing

Distinct solutions have been developed for the creation and merging of depth maps. It is, therefore, not uncommon to find large varieties in the accuracy, roughness, and completeness of the 3D polymesh DSMs generated by MVS pipelines (Nocerino et al., 2020). The same variability can also be attested in the creation of raster DSMs. Since this variety is algorithm and dataset dependent, a possible criticism on this paper could be that the assessment is based only on one specific surface. However, this criticism is, in the authors' opinion, not valid for two reasons. First, Agisoft is a commercial company and will, as a result, never disclose its core algorithms. Without this knowledge, overly detailed assessments make little sense as one can only guess the causes (and remedies) for the observed phenomena. Second, software changes constantly. Over the past years, there have been major algorithmic changes in how Metashape (Professional) computes polymeshes (Agisoft LLC, 2021). Besides, various implementation details and parameters were continuously tuned (pers. comm. Dmitry Semyonov, Agisoft LLC). All these factors make a relevant assessment of the MVS pipeline by Metashape (or any other proprietary IBM software) very hard. At any moment, the core algorithms or their implementation details can vary. This, in turn, might have an enormous influence on the resulting point cloud, polymesh and raster DSMs, which renders any previous MVS assessment quasi useless. This limited shelf life does not make all *intra*- and *inter*-software comparisons senseless. When done in a rigorous manner (and preferably with disclosed algorithms), these comparisons provide temporal reference points upon which the choice for a specific software can be based.

Due to the proprietary and fast-changing nature of Metashape, this article does not aim to be such a reference point. Instead, the presented results should make users aware that not all image-based 2.5D raster DSMs are created equally. It always pays off to test different approaches on an image subset if one aims to generate the most complete, most detailed, and least noisy raster DSM from the available imagery.

5.4 Need for reference data

Most methodologically proper investigations into the accuracy and spatial detail recording capabilities of an IBM pipeline compare the output to some reference data. Thus, one might wonder why the assessments performed in (and advocated by) this paper are all purely visual. The reasons for this are threefold. First, no better-resolved and highly accurate reference surface was available for the *Stephansdom* mural painting. This will be the case for most heritage recording projects. One could, of course, argue that this paper did not present the mere documentation of a heritage asset, but a data processing exercise and that the latter necessitates the use of a reference surface.

This brings up the second reason, which ties into the previous section. Software and algorithmic implementations constantly change, and their specifics are, in the case of Metashape, even unknown. A more quantifiable assessment could become

worthless with a new Metashape release (which happened while writing this article). When comparing software pipelines, a pure visual assessment often suffices to spot general trends in software outputs, which was precisely the aim of this paper.

Third, it is tough to find a technique that can deliver a reference surface when dealing with small sampling distances (the raster DSM contains 100 data points per mm²). This issue was also raised by Sapirstein (2018). Scanners for industrial metrology seem appropriate to generate reference surfaces, but they are typically limited to small surface patches. Covering only a 1 m² area would already require the alignment of numerous scans, thereby constraining the final accuracy of the generated surface. Besides, the meshing and mesh-repair operations that turn their raw 3D point clouds into a polymesh further downgrade the theoretical precision and accuracy of every captured surface point. It would, of course, be valid to know if the differences in smoothness between the raster DSM profiles from method 1, method 4, and method 5 (see Figure 7) correspond to authentic surface detail or merely digital surface noise. The depth maps-based raster DSM (i.e. method 5) exhibits the sharpest bends in its surface profile. It might represent the lime plaster's structure at that scale, but this remains hitherto unclear. However, the raster DSM from method 5 proved the most suitable to interpret the painting's surface visually. From that point of view, the answer to the previous question becomes suddenly less relevant.

6. CONCLUSION

Different image-based modelling roads can lead to the creation of digital, rasterised surface representations. However, these spatially two-and-a-half dimensional rasters do not equally represent the geometrical surface details encoded in the image collection from which they are extracted. This paper assessed the options to generate a rasterised and geometrically detailed digital surface from about 3000 images of a 16th-century mural painting in Vienna's *Stephansdom*. Five approaches embedded in Agisoft's Metashape Professional were tested, their pros and cons reported, and complemented by theoretical and practical reflections. In the end, combining a depth maps-based and a polymesh-based raster was necessary to digitally approximate the subtle and more pronounced relief differences exhibited by this mural painting and its neighbouring architectural elements.

REFERENCES

- Aanæs, H., Jensen, R.R., Vogiatzis, G., et al., 2016. Large-Scale Data for Multiple-View Stereopsis. *Int. J. Comput. Vis.* 120(2), 153-168. doi.org/10.1007/s11263-016-0902-9.
- Agisoft LLC, 2021. Agisoft Metashape Change Log. www.agisoft.com/pdf/metashape_changelog.pdf (25-03-2021)
- Bleyer, M., Rhemann, C., Rother, C., 2011. PatchMatch Stereo - Stereo Matching with Slanted Support Windows. In: *Proc. of the British Machine Vision Conference 2011*. BMVA Press, Durham, pp. 14.1-11. doi.org/10.5244/C.25.14.
- Förstner, W., Wrobel, B.P., 2016. *Photogrammetric Computer Vision. Statistics, Geometry, Orientation and Reconstruction*. Springer International Publishing, Cham.
- Hess, D.R., 2010. *Blender foundations. The essential guide to learning Blender 2.6*. Focal Press/Elsevier, Amsterdam.
- Jordan, G., 2007. Digital Terrain Analysis in a GIS Environment. In: *Digital terrain modelling*. Springer, Berlin, pp. 1-43. doi.org/10.1007/978-3-540-36731-4_1.
- Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. In: *Proceedings of the 4th Eurographics symposium on Geometry processing*, Eurographics Association, Switzerland, pp. 61-70. doi.org/10.1145/2487228.2487237.
- Kazhdan, M., Hoppe, H., 2013. Screened poisson surface reconstruction. *ACM Transactions on Graphics* 32(3), 1-13. doi.org/10.1145/2487228.2487237.
- Kohn, R., 2001. Der Wiener Stephansdom als „österreichisches Pantheon“. *Wiener Geschichtsblätter* 56(4), 347-355.
- Kokalj, Ž., Hesse, R., 2017. *Airborne laser scanning raster data visualization*. Založba ZRC, Ljubljana.
- Kraus, K., 2007. *Photogrammetry. Geometry from images and laser scans*, 2nd edition. Walter de Gruyter, Berlin - New York.
- Li, F., Du, Y., Liu, R., 2016. Truncated Signed Distance Function Volume Integration Based on Voxel-Level Optimization for 3D Reconstruction. *Elec. Imaging* 2016 (21), 1-6. doi.org/10.2352/ISSN.2470-1173.2016.21.3DIPM-398.
- Lorensen, W.E., Cline, H.E., 1987. Marching cubes: A high resolution 3D surface construction algorithm. In: *Proc. of SIGGRAPH '87*. ACM, New York, pp. 163-169. doi.org/10.1145/37401.37422.
- Nocerino, E., Stathopoulou, E., Rigon, S., Remondino, F., 2020. Surface Reconstruction Assessment in Photogrammetric Applications. *Sensors* 20(20), 5863. doi.org/10.3390/s20205863
- Sapirstein, P., 2018. A high-precision photogrammetric recording system for small artifacts. *J. Cult. Herit.* 31, 33-45. doi.org/10.1016/j.culher.2017.10.011.
- Shen, S., 2013. Accurate multiple view 3D reconstruction using patch-based stereo for large-scale scenes. *IEEE Trans. Image. Process.* 22(5), 1901-1914. doi.org/10.1109/TIP.2013.2237921.
- Splietker, M., Behnke, S., 2019. Directional TSDF: Modeling Surface Orientation for Coherent Meshes. *Proc. of IROS 2019*, pp. 1727-1734. doi.org/10.1109/IROS40897.2019.8968264.
- Szeliski, R., 2011. *Computer vision. Algorithms and applications*. Springer, New York.
- Verhoeven, G.J., 2017. Mesh Is More. *J. Archaeol. Method Theory* 24(4), 999-1033. doi.org/10.1007/s10816-016-9305-z.
- Verhoeven, G.J., 2018. Resolving some spatial resolution issues – Part 1: Between line pairs and sampling distance. *AARGnews* 57, 25-34. doi.org/10.5281/zenodo.1465017.
- Verhoeven, G.J., Santner, M., Trinks, I., 2021. Bildbasierte 3D Modellierung der Wandmalerei in der Bischofstorvorhalle im Wiener Stephansdom. *ÖZKD* 75(1/2), 163-174.
- Vu, H.-H., Labatut, P., Pons, J.-P., Keriven, R., 2012. High accuracy and visibility-consistent dense multiview stereo. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 34(5), 889-901. doi.org/10.1109/TPAMI.2011.172.
- Wood, J., 2008. Digital Elevation Model (DEM). In: *Encycl. of geographic information science*. SAGE, LA, pp. 107-109.
- Zakšek, K., Oštir, K., Kokalj, Ž., 2011. Sky-View Factor as a Relief Visualization Technique. *Remote Sens.* 3(2), 398-415. doi.org/10.3390/rs3020398.