care, judgment, dexterity

CRAEFT

D3.2 Advanced digitisation technologies

Project Acronym	Craeft
Project Title	Craft Understanding, Education, Training, and Preservation for Posterity and Prosperity
Project Number	
Deliverable Number	D3.2
Deliverable Title	Deliverable
Work Package	
Authors	Xenophon Zabulis, Nikolaos Partarakis, Petros Stavroulakis, Ioanna Demeridou
Number of pages	92

This project has received funding from the European Commission, under the Horizon Europe research and innovation programme, Grant Agreement No 101094349.

http://www.craeft.eu/

Executive summary

This deliverable introduces cost-efficient and reproducible approaches for the reconstruction of surfaces and transparent objects. For surface reconstruction, a new scanning technology capable of capturing surfaces in both 2D and 2½D at very high resolutions is proposed. For transparent objects, a new scanning method and apparatus for digitisation is proposed.

In Section 1, we introduce the deliverable, scope, terminology and contribution, as well as place emphasis on the cost-effectiveness and ease of use of the proposed approaches. As described in the project proposal, Craeft makes use of results from previously EU-funded projects, namely Mingei and Transparent3D. Thus, in that section, we provide an analysis of which work took place in which project.

Section 2 reviews related work and is separated into the following three domains: surface scanning, surface digitisation, and transparent object digitisation.

In Section 3, we propose an image acquisition approach, for increasing the accuracy of camera motion estimation. Along with this approach, we propose a cost-efficient hardware composition made from printable parts and conventional CNC motors that are programmed to carry out this approach. This approach implemented by the hardware prescribed the acquisition of images at multiple distances and tracks point features across multiple views and distances to increase the robustness of correspondence establishment. This increase along with a global optimisation step, provides benefits for both 2D scanning and 2½D surface digitisation. An approach is formulated for each case, taking into account the particular image acquisition approach and the proprioceptive data from the scanner.

In Section 4, we address the challenge of accurately and practically reconstructing transparent objects using optical 3D metrology. We present the proposed image acquisition, hardware, and 3D volumetric algorithm for the digitisation of transparent, thin-walled objects. A simple-to-create light isolation chamber (white box), a turntable, an LCD monitor, and a conventional camera comprise the proposed setup. The proposed reconstruction method is based on an approximation of the CT scan, which measures the absorption of X-rays by human tissues perimetrically and uses the Radon transform to obtain a volumetric 3D model of the density of the body in each voxel. Similarly, the absolved light even from transparent objects, provides a visual cue that is exploited for the case of thin-walled objects (such as glasses, bottles, and scientific glass items) to obtain their volumetric reconstruction.

In Section 5, we evaluate the three algorithms qualitatively, investigating the applicability of the proposed approaches to the problem and compared to conventional methods.

In Section 6, we quantify the accuracy of the proposed methods, indicating the extensions to the state-of-the-art brought by them.

In Section 7, future work is discussed and technical conclusions are drawn.

Six appendices complement this deliverable. Five of them present a market survey, printable components, and hardware calibration details of the scanner proposed in Section 3, as well as surface scanning results obtained using it that are referenced in the evaluation. The reason that these results

are reported in appendices is explained in Section 1. Also, Appendix D reports on experiments conducted in the context of Transparent3D.

Document history

Date	Author	Affiliation	Comment
2024-06-15	Xenophon Zabulis	ICS/FORTH	Created the basic format of the deliverable
2024-07-01	Xenophon Zabulis	ICS/FORTH	Created the first draft and added references
2024-07-15	Ioanna Demeridou	ICS/FORTH	Inserted figures and images related to the topic. Designed tables.
2024-07-20	Xenophon Zabulis	ICS/FORTH	Added results section. Finalized the references and bibliography.
2024-07-26	Ioanna Demeridou	ICS/FORTH	Added an executive summary
2024-07-28	Xenophon Zabulis	ICS/FORTH	Proofread the entire document and made corrections.
2024-07-29	Xenophon Zabulis	ICS/FORTH	Read the review and adjustments.
2024-08-10	Xenophon Zabulis	ICS/FORTH	Update for reviewer comments.
2024-08-20	Xenophon Zabulis	ICS/FORTH	Finalising document.

Abbreviations

The following abbreviations are used in this manuscript:

C3Dp	Cartesian 3D printer
C2Ds	Cartesian 2D scanner
FoV	Field of View
GPS	Global Positioning System
JPEG	Joint Photographic Experts Group
USD	Unites States Dollar
bAh	milliamp Hour
c	Cent
G	Giga
hrs	hours
mm	millimetre

µm	micrometre
cm	centimetre
kg	kilogram
p	pixel
pt	point
ppi	points per inch
T	tera
V	Volt
W	Watt

Table of contents

1 Introduction

That "traditional craftsmanship is perhaps the most tangible manifestation of intangible cultural heritage"[1] underscores the significance of properly representing materiality in the digital media used to present it. We prioritise their visual and geometrical representation, although other material aspects, such as haptic and acoustic, are acknowledged and discussed as future work in Section 7. The reason for this prioritisation is twofold. First, appearance is a common and significant characteristic of almost every handmade product. Second, the digital representation of handmade products is important in the modern marketplace, as the online presentation of products impacts the prosperity of craft practitioners. Besides, high-quality presentation raises awareness of high-quality work and can make talented practitioners stand out.

The democratisation of digital imaging and 3D reconstruction methods has increased the expectations of consumers for the digital inspection of products. Digitally capturing handmade products is not always simple or economical if a high-quality presentation of the product is warranted. Traditional methods of digitisation, such as photography, require a level of expertise that many craft practitioners do not possess. Furthermore, achieving the high resolution necessary to truly reflect the quality and uniqueness of these products often demands expensive equipment. Besides consumer photography and photographic documentation, the 3D digitisation of artefacts is nowadays often considered the norm. Moreover, the need for high-resolution surface digitisation is encountered in a multitude of applications, such as the documentation of cultural heritage artefacts, the creation of realistic textures for digital design, and the capability of communicating product details using digital means and over the Internet.

To avoid ambiguities in expression we follow the following terminology, in this deliverable.

- Surface scanning is a method that creates a 2D image of the scanned surface that is devoid of perspective distortions, like a map.
- Surface digitisation regards the 2½D reconstruction of a surface that captures variations in elevation, shape, and texture of the scanned surfaces.
- 3D volumetric reconstruction regards a representation of the surface and spatial occupancy of a 3D solid that also captures its appearance characteristics.

For brevity, we sometimes use the term 'transparent' instead of 'semi-transparent'. When doing so, we mean highly transparent or more technically, materials with very low scattering (i.e., low refractive index).

Despite the versatility of off-the-shelf scanning and photogrammetric approaches, some crafts are still underrepresented due to the challenging nature of the materials they employ. Specifically, "non-Lambertian", or otherwise, objects made of or including semi-transparent, translucent, and shiny materials, such as glass and metals, are always a difficult reconstruction target. The reason is that the

[1] https://ich.unesco.org/en/traditional-craftsmanship-00057

scanning principles employed in photogrammetry or time-of-flight scanning do not hold for dielectric materials, such as glass and metals. Specialised methods exist for the scanning of such objects requiring sophisticated equipment and in some cases being pervasive and unsuitable for cultural heritage objects.

Given the above needs, in this deliverable, we are motivated to make the following contributions.

1. The introduction of a new scanning technology capable of capturing surfaces in both 2D and 2½D at very high resolutions and the scaling of a small-scale scanner prototype to a wide-area scanner.
2. A new scanning method and apparatus for the digitisation of transparent objects.

The contributions are aligned with two axes, to make them useful for craft practitioners and researchers.

1. Cost-efficiency by using printable and low-cost visual sensors and hardware equipment. To further enhance cost-efficiency we robustified surface reconstruction algorithms for visual inaccuracies caused by hardware motion and vibratory side-effects, eliminating the need for expensive hardware components and materials.
2. Ease of installation and operation enables non-expert users to produce detailed and accurate digital representations of their work with ease and confidence. Both methods provide hardware-facilitated automatic acquisition of images in the particular way prescribed by our algorithmic methods. Besides the automation of this tedious task, the software methods are automatic and do not require tuning or training, besides the initial system calibration.

As part of the reported work has roots in past projects that FORTH participated in, we provide an analysis of the topics presented in this deliverable and explain which part was performed in which project.

First, in Mingei, FORTH developed a surface scanning device (a surface scanner) as a "spin-off" result, as it was not predicted by the GA but inspired by the recognition of the need to capture the intricate wefts of woven silk. This work digitised a small area surface in ultra-high resolution and has been published in the following open-access journal: Zabulis X, Koutlemanis P, Stivaktakis N, Partarakis N. A Low-Cost Contactless Overhead Micrometer Surface Scanner. Applied Sciences. 2021; 11(14):6274. [doi.org:10.3390/app11146274](https://doi.org/10.3390/app11146274). This deliverable extends this prior work by inheriting and improving the hardware developed in the aforementioned work, in two ways.

1. We scaled the hardware to operate on wider areas and be able to scan large textiles in analogously high resolution.
2. We upgraded the scanning algorithm to produce 2½D surface digitisation, in addition to the previous 2D scans, which led to the following open-access publication: Koutlemanis P, Zabulis X, Stivaktakis N, Partarakis N, Zidianakis E and Demeridou I (2024) A low-cost close-range photogrammetric surface scanner. Front. Imaging. 3:1341343. [doi: 10.3389/fimag.2024.1341343](https://doi.org/10.3389/fimag.2024.1341343).

Both scans are metric and, in both cases, the method can capture (opaque) matte and shiny surfaces. As we plan for a "scanning handbook" later down the road, the results produced for the 2D scanner are included for completeness. However, to concisely report work produced in Craeft, these results are provided in the form of appendices.

Second, as mentioned in the Craeft GA, in Craeft we will improve the results produced in the Transparent3D Marie Curie project. In that project, we investigated the potential of scanning 3D transparent solids. Although some research efforts we investigated in Transparent3D, such as polarized imaging, generative vision methods, and physics-based rendering[2], did not give the anticipated results, we inherited significant theoretical and practical experience, as well as the imaging equipment. Using these resources, in Craeft we propose a more specialized method, tailored for the 3D, volumetric reconstruction of thin-walled, transparent objects such as glasses, bottles, scientific glass items, etc. This work:

1. Has been published in open-access in Stavroulakis, P.I.; Ganetsos, T.; Zabulis, X. Large Scale Optical Projection Tomography without the Use of Refractive-Index-Matching Liquid. Sensors 2023, 23, 9814. doi:10.3390/s23249814.
2. Exhibits industrial interest in glass and plastic manufacturing and quality control and, thus, we initiated the process of applying for a patent (patent prosecution). Currently, this process has reached the PCT (Patent Cooperation Treaty) stage, often referred to as the "international phase" or "PCT phase" of patent prosecution. The PCT ID of this application is PCT/GR2023/000051.

Table 1 summarizes the contributions per project.

Table 1. Contributions per project.

Project	Result	Target	Material	Scale
Mingei	2D scan	Planar surfaces	Opaque, matte or shiny	Small
Craeft	2D scan	Planar surfaces		Large
Craeft	2½D digitisation	Undulating surfaces		Large
Transparent3D	Equipment & Theory	Solids	Semi-transparent	Any
Craeft	3D digitisation	Thin-walled solid		Small

In terms of the temporal occurrence of this work, we report that all activities run concurrently, with one exception. The 2D scanning of wide-area surfaces was initiated last, as the ordering and assembly of the hardware had to take place (as opposed to the small-scale scanner that was already available). As such, the large-scale scanner has only been qualitatively and not quantitatively evaluated. This evaluation will be reported in the next version of this deliverable.

The remainder of this deliverable is structured as follows. Section 2 reviews related work and is separated into the following three domains: surface scanning, surface digitisation, and transparent object digitisation. In Section 3, we proposed an image acquisition approach, the hardware to implement it, and two algorithms for surface scanning and digitisation, respectively, that capitalise upon this approach and hardware. In Section 4, we present the proposed image acquisition, hardware, and 3D volumetric algorithm for the digitisation of transparent, thin-walled objects. In Section 5, we evaluate the three

[2] Although that in Transparent3D we did not find physics-based rendering and simulation suitable for 3D reconstruction of transparent solids, we re-used our experience in Craeft in creating photorealistic material-specific renderings which are useful in the Design Studio and in craft-specific simulations. The experience to select these technologies in Craeft, stems exactly from this EU-funded Marie-Curie.

algorithms qualitatively and, in Section 6, quantitatively. These two sections are subdivided into three parts each, for the 2D, 2½D, and 3D methods, respectively. Finally, in Section 7, we report conclusions and discuss future work.

2. Related work

A recent and comprehensive review of 3D scanning technologies and approaches can be found in The Mingei Handbook on Heritage Craft representation and preservation, at Step 3 “Craft recording”, in Section 3.2, “Digitisation of endurant assets”. This section serves the positioning of the proposed work pertinent to the state of the art and does not provide new information. It is organised in three sections, for surface scanning, surface digitisation, and digitisation of semi-transparent objects, respectively.

2.1 Surface scanning

The generation of image mosaics out of partial images of a surface is a useful task in many applications. Mosaics are useful because they image a larger amount of surface than a single image does. If a mapping between pixel and metric coordinates is achieved, then world measurements can be performed using the mosaic, much like in cartographic maps. Image registration upon general surfaces enables photorealistic maps for Geographical Information Systems, photo-panoramas [2], but also specialized mosaics from usually unseen surfaces such as the inner of pipes [3], the gastrointestinal tract [4], and the human retina [5]. At the core of all methods for image mosaic generation is the problem of image registration.

The simplest case of mosaic generation is met when imaging a planar surface, by multiple and conveniently tessellated overlapping frontal views. Just this case is useful in several domains, such as remote sensing [6], document scanning [7], bioinformatics [8, 9], art [10], and others. Approaches to this problem that are based purely on visual cues are continuously making progress but, given pixel quantization, they exhibit error. For a large number of images, this error accumulates and gives rise to distortions. For this reason, applications that require large mosaics make use of independent information about the location of the camera. For example, photorealistic stitching of remote sensing and aerial images is supported by GPS measurements. We use this principle in the context of overhead scanning, where approximate location measurements are available from the motion mechanism of the scanner.

2.1.1 Visual image registration

The problem of computational image registration dates back at least four decades of study. Methods in the literature are usually called ‘local’ if they use point feature correspondences, or ‘global’ if they use overall image similarity [11, 12]. When combining images in a mosaic, these images are required to have some lateral “overlap” and the registration task is called stitching. Due to restricted overlap, stitching is more accurate when local methods are used.

Given point correspondences across two images of a planar surface, robust registration solutions have been found and, by now, are textbook material [13, 14]. When many images are registered sequentially in a mosaic, the error is accumulated distorting the result. A solution is to employ a “global alignment” or “bundle adjustment” step, which either obtains a more accurate solution or, at least, distributes error so that the shape of the scanned area is retained. Although this improves the result, the error from the registration of many images manifests as local, noticeable distortions, often called “seams”. For a few images, these distortions are small and well-treated by methods that reduce their visual prominence. However, when the number of images is increased by two orders of magnitude, we observed these distortions to become significant and noticeable at close and macroscopic inspection.

2.1.2 Proprioceptive image registration

Another way to register images in mosaics is by scanners, which use mechanisms to drive the sensor to locations where the acquired images would precisely match. The utilised sensors are most often line cameras with intense illumination and less often photographic cameras. A market survey of pertinent solutions can be found in Appendix B.

Contact-based, flatbed scanners provide up to 1000ppi at a significantly high cost. Large-format scanners provide resolution up to 1200ppi and are almost contactless. However, the scanned material should be less than a thickness threshold, i.e. 3cm, to pass through the scanning slit and also exhibit high cost. Film scanners exhibit higher resolution, but require contact, and material transparency, and are limited to the frame size of photographic film.

Large-format scanners are contactless and designed for sensitive documents but are also used for scanning textiles and other similar materials. They reach up to A0 scanning size. The precision required for this mechanical task elevates the cost of the scanning hardware.

Book scanners are contactless and exhibit resolution in the range of 600 − 1200ppi. Their cost ranges from low to very high, though in many cases the elevated cost is due to the mechanics for the automation of page-turning. It ought to be noted though that V-shaped, as opposed to flatbed, book scanners are unsuitable in the case of deformable materials, such as fabrics or sand.

Recently, the need for realistic textures gave rise to flatbed, contactless surface scanners, called “material scanners”. They use camera photography and exhibit resolution up to 1000ppi. Their effective scanning surface is in the order of 30×40 cm^2. Lower-cost material scanners utilize the sensor of the mobile phone [15], the result however exhibits lower resolution and definition, compared to the aforementioned solutions.

2.2 Surface digitisation

The approaches which achieve accurate and reliable surface digitisation in large detail are as follows. Computed Tomography (CT) uses X-rays to provide volumetric scans of objects and their surfaces and has been used for the digitisation of small objects [16]. Magnetic Resonance (MR) tomography has been also used for detailed volumetric scans of small objects [17]. Laser scanning uses the time-of-flight principle to measure distances and reconstruct surfaces. Structured light scanners project light patterns onto surfaces, capture the deformed patterns using a camera and, from these deformations, estimate surface structure. Finally, as this work proposes, photogrammetric reconstruction of small artefacts has also been employed, but with significant constraints as to the scanning area and hardware requirements.

Albeit highly accurate, tomographic methods (CT, MR) exhibit severe costs and can be applied only in specialised laboratories. Moreover, they provide volumetric and accurate structure reconstruction but do not reconstruct surface appearance (texture). Laser scanning and structured light scanning, capture surface appearance and are marginally less accurate and are more cost-efficient because they require less equipment. Photogrammetry is even more affordable and widely available, as it relies on standard cameras only. Moreover, photogrammetry captures best the appearance of surfaces because it images surfaces without the effects of active illumination. Photogrammetry produces high-resolution and highly detailed textured models that accurately represent visual appearance. Compared to laser and structured light methods, photogrammetry can be less accurate, but is passive, in that it does require the projection of energy in the form of radiation (e.g., light, X-rays, etc.) upon the scanned surface. This is important in light-sensitive materials, encountered in historical, archaeological, biological, and cultural contexts.

The literature is first reviewed in terms of the applications where close-range photogrammetric systems are employed, as well as the optics and illumination methods utilised for their implementation. Thereafter, data acquisition quantities and reconstruction methods are reviewed. Last but not least, photogrammetry creates reconstructions “up to scale”, that is, the resultant reconstruction is not in metric units but in arbitrary ones. As the goals of this work include surface measurement, the ways that the scale factor is estimated for close-range photogrammetric methods conclude this section.

2.2.1 Applications

Photogrammetry has been applied in a wide variety of contexts over the last three decades, mainly referring to large-scale reconstruction targets in the range of hundreds to a few meters, such as built and terrain structures captured from aerial views, indoor environments, and cultural heritage objects. More recently, close-range photogrammetry has found applications in the high-resolution reconstruction of structures of much smaller scale, of industrial [18, 19, 20], biological [21], archaeological [22, 23], anthropological [24], and cultural interest [25, 26], in the range of millimetres (see [27] for a comprehensive review of close-range photogrammetry applications).

2.2.2 Motorisation

When using a single camera, camera motion is required to acquire 3D information. Motorisation of camera movement has been used in photogrammetric reconstruction to alleviate user effort and acquire images at numerous, prescribed viewpoints. The main strategies of motion are either circular around the target, using a turntable, or in a Cartesian lattice of viewpoints [28]. Cartesian approaches exhibit the advantages of being unconstrained of the turntable size and that the camera can be moved arbitrarily close to the scanning target. This work follows the latter approach to scan wider surface areas and avoid the occurrence of shadows and illumination artefacts. The most relevant scanning apparatus to this work is [29], which uses a CNC to move the camera.

2.2.3 Optics

In the millimetre and sub-millimetre range, zoom and microscopic, or “tube”, lenses have been used [30, 31, 28], employing tedious calibration procedures and relatively inaccurate results [32]. Instead, macro lenses are more widely used in the photogrammetric reconstruction of such small structures [33]. However, macro lenses exhibit a very limited depth of field. This limitation makes photographs acquired with a macro lens out of focus in the periphery of the image. To compensate for this effect, focus stacking [34] is often employed in macro photography [35].

2.2.4 Illumination

Illumination is necessary for photogrammetric methods so that the surfaces are visible to the camera. Photogrammetry typically operates upon environmental illumination. Ambient illumination is ideal for photogrammetric methods because it prevents the formation of shadows, which confound the visual documentation of the reconstructed surfaces. Most works use setups that insulate the target object from environmental illumination and use a specific light source in conjunction with illumination diffusers to prevent the formation of shadows. In closer relation to this work, some systems use light sources that move along with the camera [36, 37, 38].

Purposefully designed illumination is used in photogrammetric methods to facilitate the establishment of more stereo correspondences. This technique is known as “active illumination”, “active photogrammetry”, or “structured light” [37, 39, 40]. Pertinent methods use active illumination to support the reconstruction of the geometry of the reconstructed surface, by artificially creating reference points on the surface which can be matched across images. The disadvantage is that the projected light alters the visual appearance of the reconstructed surfaces. As such, this work does not use active illumination as it strives to realistically capture the visual appearance of the target surfaces.

2.2.5 Data

The data acquired from the close-range photogrammetric methods vary depending on the hardware and optical apparatus employed.

Works that report scanning areas fall in the range of ≈ [0.14,158.7]cm^2. Specifically, the maximum area reported from these works is (approximately) as follows: 0.196cm^2 [30], 10 cm^2 [39], 15cm^2 [37], 22.5cm^2 [23], 25cm^2 [41], 32 cm^2 [29], 40 cm^2 [36], 57.6 cm^2 [42], 60 cm^2 [22], 91.5 cm^2 [31], 93.4 cm^2 [30], and 158.7 cm^2 [42]. Out of these works, only [30] reports the achieved resolution, which is 3745p⁄mm^2.

Works that report the number of images and pixels processed fall in the range of [9.6,3590.0]MP. The data load is calculated as the number of utilised images times the resolution of each image. Specifically, the maximum number of pixels reported from these works is (approximately), in MP, as follows: 16 × 0.6=9.6 [31], 7 × 10.1 = 70.7 [38], 12 × 10.1 = 121.2 [41], 30 × 24 = 720.0 [40], 72 × 12.3 = 885.6 [29], 40 × 36.3 = 1452.0 [23], 245 × 10.2 = 2,499.0 [21], and 359 × 10 = 3590.0 [36].

2.2.6 Reconstruction

Several works use commercial photogrammetry software to reconstruct the imaged scene, with the most popular being the Pix4D and AgiSoft suites. As discussed in Section 5, these software suites provide less accurate results, as they are agnostic to the image acquisition strategy employed in this work and do not utilise feature tracking.

Some of the reviewed works perform partial reconstructions which result in individual point clouds and then merge them using point cloud registration methods. These methods are mainly based on the Iterative Closest Point algorithm [43, 44], either implemented by the authors or provided by software utilities, such as CloudCompare and MeshLab. The disadvantages of merging partial reconstructions are the error of the registration algorithm and the duplication of surface points that are reconstructed by more than one view. These disadvantages impact the accuracy and efficacy of the reconstruction result, respectively.

2.2.7 Scale

Several ways to generate metric reconstructions have been proposed. One way is to place markers at known distances so that when they are reconstructed, they yield estimates of the absolute scale. However, this method is prone to the localisation accuracy of said markers. An improvement to this approach comes from [38] which uses the ratio of the reconstructed objects over their actual size, which is manually measured. However, it requires the careful selection of the reference points to estimate the size of the object both in the real and the reconstructed objects and is, thus, prone to human error. This can be a difficult task for free-form artworks as they may not exhibit well-defined reference points, i.e. as opposed to industrially manufactured objects. Therefore, the accuracy of this approach is dependent on the accuracy of manual measurements of the actual object and its reconstruction.

Another way to achieve metric reconstruction is to include objects of known size in the reconstructed scene, such as printed 2D or 3D markers [38, 36, 41]. This method does not involve human interaction and is not related to the structure of the reconstructed objects. The main disadvantage of this approach is the production of these markers, as printing markers at a micrometre scale is not achievable by off-the-shelf 2D or 3D printers.

Finally, a way to estimate the scale factor is based solely on the reprojection error of the correspondences of a stereo pair [45]. This approach is independent of the shape of the target object and does not require human interaction. However, this approach is formulated for a stereo pair and not for a single camera and, also, is highly dependent on the calibration accuracy of this pair.

2.3 Digitisation of semi-transparent objects

The accurate and practical optical 3D reconstruction of transparent objects has been an open challenge for the field of optical metrology [86, 87, 88]. The main difficulty in using conventional practical optical metrology tools such as structured light scanning, laser scanning, and photogrammetry to reconstruct transparent objects, is due to the combined phenomena of transmission, reflection, and refraction noticed in transparent objects [89]. This fact does not allow the use of the aforementioned practical optical metrology tools, all of which require high levels of diffuse surface reflection to operate [90].

The typical way of getting around this limitation is to render the transparent object's surface opaque, by spray-coating it with a diffuse reflection layer [91]. The transparent object can then be reconstructed with conventional visible-spectrum optical metrology tools [91]. The process of spraying the object, however, is time-consuming, adds considerable costs to the reconstruction process, and might not be allowed in some applications (sensitive cultural heritage objects, food and drink containers) due to the risk of contamination.

There have been multiple attempts to create alternative spray-less digitization techniques for transparent objects in the past [86, 87]. The proposed techniques include the use of OPT with refractive index-matching liquid [92], the fringe projection for detection of pattern differences [93], infra-red (IR) to detect induced surface heating [94, 95], Infrared Digital Holography [96], ultra-violet (UV) fluorescence [97, 98], a combination of X-ray tomography and photogrammetry [99], shape from polarization [100], a combination of polarization imaging and inverse rendering [101], shape from interaction [102], the visual hull technique [103], various AI-based image processing methods [104, 105] terahertz (THz) Imaging and tomography [106, 107], passive single-pixel imaging [108], and edge estimation computer vision techniques [109].

None of the spray-less techniques suggested above, however, possess the combination of advantages of conventional visible-spectrum optical metrology tools, that of being concurrently cheap-to-use, rapid, practical, non-contact, and easily automated.

In this deliverable, we investigate the possibility of achieving the 3D reconstruction of large thin-walled transparent objects for quality control and digital preservation purposes in cultural heritage applications. The solution proposed retains the aforementioned characteristics of conventional optical metrology tools already in use in these industries today but can additionally operate without the need for opaque coating. To achieve this, the use of OPT [110] without the use of refractive index-matching liquid was investigated.

3. Surface scanning and digitisation

We propose a comprehensive approach to the problems of reconstructing approximately planar surfaces, such as textiles or relief structures. Depending on the application's needs, we provide solutions for both surface scanning and digitisation. The common axis in all the solutions provided is the approach to acquiring the images for both types of reconstructions. The central idea is that we use a low-cost camera to take close-range images of the surface at the detail. Typically, these images cover only a detail of the entire surface, so multiple are acquired to cover it and are, then, combined by reconstruction methods into a coherent result. Central to the success of this operation is the FoV overlap between neighbouring images, so that “correspondences” can be found between them and estimate the camera motion that links them.

Our approach requires the acquisition of images using a specific strategy that includes the acquisition of images that have an auxiliary role, in that they are not going to be directly used in texturing the final reconstruction, but contribute to its overall accuracy. This approach is common for both 2D and 2½D reconstructions and, as such, the same hardware is used for both, but with appropriate parameterisations for each. As this is central to the proposed work, we briefly present it first.

A mechanical device which we call an “overhead surface scanner” automates the tedious part of image acquisition. This device is comprised of the CNC kinematic mechanism of a 3D printer, in which we have replaced the printing head with a camera. The C3Dp provides 3 Degrees of Freedom (see Figure 1, top-left). We use this device to acquire images above the surface, facing it perpendicularly. The locations of image acquisition form a hypothetical pyramid, such as the one shown in Figure 1, top-right. The acquired images are conceptually classified into layers in this pyramid, as defined by the distance from the surface of interest. These images and their spatial organisation are key to the proposed approach. We use images from top layers as auxiliary: not for surface reconstruction to obtain global consistency constraints that contribute to the overall accuracy of our reconstruction.

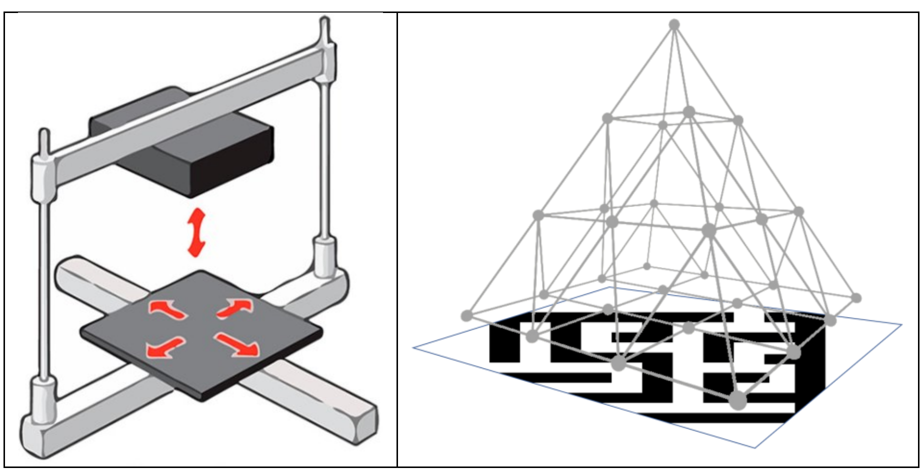


Figure 1. : Abstraction of C3Dp motion (left, from [1]) and proposed scanning locations (right).

The two most important elements relevant to the integrated hardware and software image acquisition mechanism and strategy are:

1. The usage of multiple layers to acquire auxiliary images, which are medially (as opposed to laterally overlapping). This gives us the possibility to establish correspondences across layers which helps us estimate better the location of the camera.
2. The collection of proprioceptive data (CNC motor readings) and use them as well to improve camera localisation accuracy.

For each layer, the locations of image acquisition form a hypothetical grid, as illustrated in Figure 6 (right). These images overlap medially, due to elevations of the camera relative to the scanned surface, and are used to provide additional registration cues. All the camera locations form a hypothetical pyramid or frustum. Camera locations are determined so that images laterally overlap at all boundaries. Across elevation layers, viewpoint locations are configured hierarchically, so that the images from “parent” viewpoints at higher layers medially overlap with “child” viewpoints at lower layers, by a factor of τ_m. This is illustrated in Figure 2, where viewpoint A overviews the area imaged by viewpoints A1, A2, A3, and A4. When τ_m = 25%, as in the example, the hierarchical structure has the form of a quad-tree. The top layer, indicating the upper end of this frustum is user-determined. We refer to this type of overlap across layers as medial overlap.

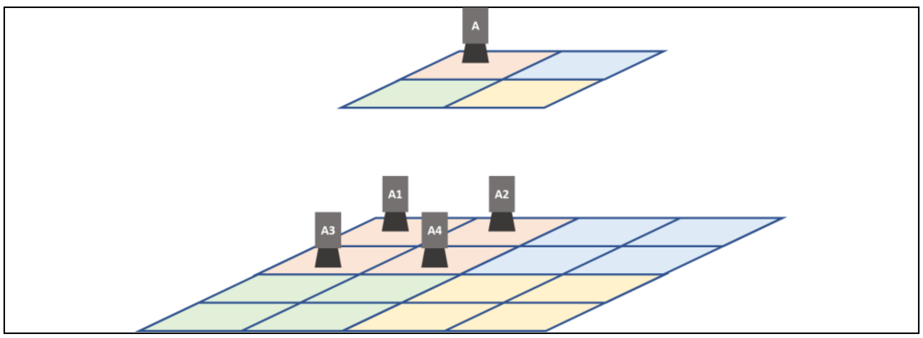


Figure 2. image acquisition across layers.

The approach actively seeks registration cues by utilising proprioceptive data in both correspondence and registration tasks. That is, we use the approximate camera locations to reduce the search for correspondences. In addition, we utilise these locations to globally optimise the geometry of the reconstruction and reduce reconstruction errors.

The acquisition of auxiliary images and the utilisation of proprioceptive data are essential factors in the reduction of the cost of our scanner. The reason is that by being able to compensate for the motor and hardware infrastructure errors we can sustain more mechanical jitter. In turn, this means that we can resort to more cost-efficient materials, such as lower-accuracy motors and lightweight metals for the CNC rig. Reduced weight also means that we can use motors of lower power.

The method of image acquisition for surface scanning and digitisation differs in the overlap of neighbouring images acquired during image acquisition. In comparison to 2D mosaicking which requires (global) image alignment, 2½D surface reconstruction requires more information, to recover surface structure. Specifically, we increase the image overlap across neighbouring images, to increase the number of stereo correspondences among them and achieve decent surface reconstruction. In both cases, the acquisition strategy exploits the proprioceptive information available from the CNC camera motion mechanism, as well as the acquisition of auxiliary images at farther distances.

To demonstrate the versatility of this approach and cover a wide range of applications encountered in crafts, we have implemented it in two spatial scales. The first is the fine-scale, micrometre domain which reveals the care put by the practitioner in intricate details of hand-carved materials, the type of thread and weave used in a textile, or the exact measurements of a missing surface fragment needed for the archaeological preservation of an artefact. The second is a wide-area apparatus for scanning large textiles, maps, etc or scanning multiple pieces in one go. The two devices are shown in Figure 3.

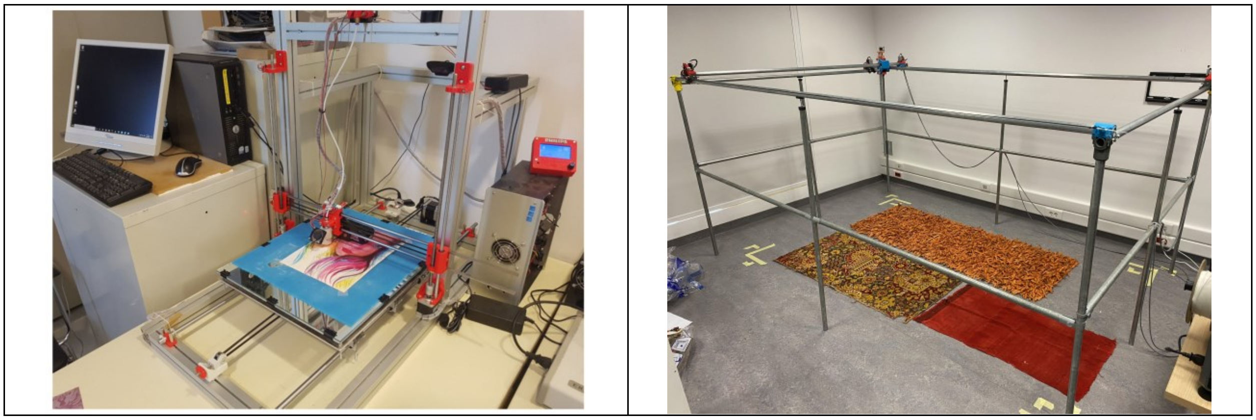


Figure 3. Surface scanners. Fine detail (left) and wide-are (right).

The rest of this section is organised as follows. We describe the materials for the scanner in Section 3.1, the way it is calibrated in Section 3.2, our approach to image acquisition in Section 3.3, the way that correspondences are established in Section 3.4, the way that surface reconstructions are computed in Section 3.5, and the output types in Section 3.5.

In each of the subsections, when needed, we distinguish between the description of the scanning and digitisation approaches, in corresponding subsections.

3.1. Materials

The specifications of the computer used in the experiments were as follows: CPU ×64 Intel i7 8-core 3GHz, RAM 64Gb, GPU Nvidia RTX 8Gb RAM (RTX2060 SUPER), SSD 256Gb, HDD 2Tb. The critical parameters are CPU and GPU RAM as they determine the number of correspondences that can be processed and, therefore, the area that can be reconstructed.

The components required to construct the proposed piece of hardware are cost-efficient and several of them are 3D printable. These components along with instructions for their assembly into the proposed hardware are reported in Appendix A.

The flash moves along with the camera, as in [36, 37, 38] (see Section 2.2.4). In this work, a ring flash is employed which is a circular illumination device that fits around the camera lens, creating a ring of light around the subject. Ring flashes provide even and shadow-free illumination for close-up shots of small subjects and evenly light the subject from all angles. Sensor brightness, contrast, and colour balance were set to automatic. Focused stacking was implemented by the sensor hardware and firmware. A textured substrate is employed to assist reconstruction as it gives rise to more feature point correspondences. The data are stored in a 128Gb SD card. The storage capacity of this card determines the maximum number of images that the system can acquire. The average size of the image file is 2.4 MB.

The apparatus sensor is different for the close and far-range cases.

- **Close-range.** The visual sensor was an Olympus Tough TG-5, with a minimum focus distance of 2mm, a depth of focus of 1cm, a revolution of 4000 × 3000p, in JPEG format, and a FoV of 16.07◦× 12.09◦. The 1cm depth of focus is currently the limit of the surface elevation variability that the proposed system can digitise. Illumination was produced by the sensor's flash.
- **Wide-area.** The visual sensor was a Nikon D850, with a minimum focus distance of 1.5m, a depth of focus of 5m, a resolution of 8256 × 5504p, in RAW format, and a FoV of 45.7∘× 35.4∘. The 5m depth of focus is currently the limit of the surface elevation variability that the proposed system can digitise. Illumination was produced by an array of external LED lights.

Translations on the xx′ and zz′ axes are implemented by moving the camera. Translations on the yy′ axis are implemented through substrate motion. This elevation determines the height of a hypothetical square frustum. The doubling of the elevation for each layer is to create a medial overlap of ≈ 4. In this way, an image at a higher layer oversees 4 images from a lower layer.

3.1.1. Configuration for surface scanning

The number of layers was 4, and they were configured as noted in Table 2. In this table, E is the elevation, n is the total number of acquired images, and n_x and n_y are the numbers of images acquired in the horizontal and vertical dimensions, respectively. Moreover, s_x and s_y are the lengths of the steps that the camera is moved to acquire an image in the horizontal and vertical dimensions, respectively. Finally, s_p is the length of the side of the square surface region that is imaged by an image pixel, at the specific layer.

Table 2. Spatial arrangement of camera locations in a square frustum structure.

Layer	E (mm)	n (#)	n_x (#)	n_y (#)	s_x (mm)	s_x (mm)	s_p (μm)
1	160.0	50	5	10	4.51	3.38	11.296
2	80.0	551	19	29	2.25	1.69	5.648
3	40.0	3102	47	66	1.12	0.84	2.824
4	20.0	14420	103	140	0.56	0.42	1.412

This level of medial overlap was sufficient, for the samples we scanned. Denser or sparser configurations are treated in the same way. For the utilised sensor, the base layer of this pyramid is 5 × 5cm^2 and is covered by 25 × 19 = 475 images. The inequality between the n_x and n_y steps is by preference. Given the rectangular camera FoV, this configuration results in a square scanned area.

We report an optical resolution of 19754 horizontal and 19820 vertical ppi, without interpolation. Mosaic pixels, thus, deviate by a factor of φ = 0.00334 from squareness. The measurement was obtained using the banknote, which is of known dimensions. This means that mosaics are linearly scaled by a factor of φ in the vertical direction. If needed, the mosaic can be resampled to feature square pixels. We report the horizontal as the scanner resolution, thus 19754ppi. The benchmark was obtained using a banknote and mm-grade graph paper.

The use of time and computational resources, as a function of the scanned area, is reported in Table 3.

Table 3. Temporal and computational requirements.

Area (cm^2)	Scan time (hrs)	RAM (Gb)	Storage (Gb)	Computation time (hrs)
25	2	3.4	1.4	3.5
91	8	12	5.5	11.5
176	15.3	25	10.5	38.5

We tested the utilisation of multiple, laterally overlapping pyramids in samples of larger areas. The targets were a 21 × 7cm^2 piece of industrially woven, patterned silk fabric and a 12 × 6.2cm^2 banknote. The results are shown in Figure 4. The pyramids used were X and Y for these cases, respectively. For the fabric, an arrangement of 5 × 2 pyramids was used and, thus, the top layer was a mosaic of 10 images. For the banknote, a 2 × 3 arrangement was used and, thus, the top layer was comprised of 6 images. In this configuration of RAM and scanning resolutions, the maximum scan size is 42 × 14 = 588cm^2.

Figure 4. Larger scans.

3.1.2. Configuration for surface digitisation

For the reconstruction of a 2.5cm^2 surface area, 3,348 images were acquired whose dimensions were 4,000 × 3,000 pixels. In these images, 7.52 · 107 key point features were detected. The computation lasted 3.55hrs and the amounts of RAM utilised were 21Gb and 3.29Gb for the CPU and the GPU, respectively. The result consisted of a mesh with 371,589 nodes and 741,885 triangles and a texture map of 16,384 × 16,384 pixels. The obtained resolution for the geometry of the reconstruction (mesh nodes) is ≈ 714 p/mm^2 or ≈ 679ppi, while its texture is ≈ 257Kp/mm^2 or ≈ 13Kdpi. These measurements can be verified in the Euro coin reconstructions (see Section 6.2.1, below).

Scanning area limitations stem from the memory capacity of the computer. Using the computational means in Section 4.1 the maximum scanning area achieved was 5cm2. In all experiments, the following parameter values were utilised: τ_o = 90%, τ_m = 25%, τ_p = 50 pixels, and τ_c = 25 pixels. Moreover, four (4) elevation layers were used. Given that medial overlap is τ_m = 25% and that the smallest elevation is at the minimum focus distance of the camera (2mm) the elevations were 2mm, 4mm, 8mm, and 16mm.

An example of the scanning dimensions is provided for a 24 × 28cm^2 artwork. In Figure 5, the artwork is (left) and is marked with a rectangle, which indicates the scanned area. The four images right to it are original images, one from each elevation layer; starting from top to bottom and shown in left to right order, respectively.

Figure 5. Artwork and original images, one from each elevation layer.

3.2. Calibration

Our sensor is active and composite. Its active part is a 3D CNC that translates the camera to a prescribed location and distance from the target surface. The sensor is a macro camera. Both components are controlled and orchestrated by a PC using the appropriate drivers. The two components of the sensor are calibrated so that each acquired image is associated with the location and distance where it was prescribed to be acquired. The calibration of the sensor for both scanning and digitisation of surfaces is the same. It has been covered in prior work and can be found in Annex F.

3.3. Image acquisition

A simple utility GUI facilitates the generation of a scan plan given the surface region to be scanned.

Given the sensor's FoV and the proportions of lateral or medial overlap, camera locations are pre-computed and saved in the scan plan file. When acquiring datasets for surface scanning or digitisation, the planned image acquisition locations are depth-first serialized and converted to scanner coordinates.

A corresponding segment of G-code is accordingly generated. This code is transmitted to the CNC and images are acquired. A driver program orchestrates the operation in the computer halting motion when images are acquired and moving the camera across the prescribed locations. In the output dataset, image filenames are associated with the camera locations in the scan plan.

3.2.1. Single image acquisition

At each location, a focus-stacked image is acquired. That is several photographs are taken with the focus point shifted slightly in each image to cover the entire depth of the subject. Using software the images are aligned to ensure they overlay correctly and the in-focus parts of each image are combined to produce a single image with a greater depth of field than any of the individual images. In our case, this operation is embedded in the sensor however they can be performed at the computer as well. The acquired images are not compensated for lens distortion at this stage. Due to the focused stacking operation that is applied individually on each image, lens distortion is estimated at a later stage.

3.2.2. Dataset acquisition

Albeit the CNC apparatus acquires all images in one pass, the acquired images are conceptually classified into layers modulated by imaging distance from the surface of interest. The software interface drives the device to acquire all images and delivers them organised per layer.

Within each layer, viewpoints are organised in a lattice, as illustrated in the example of Figure 6 (left). In the example, n_x = 4 and n_y = 5. Solid lines mark the surface regions imaged by the camera. Dots mark camera centre locations. The horizontal and vertical distances between camera centres are denoted as s_x and s_y, respectively. Transparent regions indicate the lateral overlap between neighbouring images of the same layer. The FoV of the camera and the required amounts of lateral and medial overlaps determine the values of s_x, s_y, and distances between layers. The top layer, indicating the upper end of this frustum is also user-determined. We use the same amount of lateral overlap for the horizontal and vertical dimensions of each later, let τ_o. Layer distances are configured so that a "parent" viewpoint images four times the area of its "child" viewpoints; this leads to a quad-tree arrangement of viewpoints and the doubling of the elevation per layer.

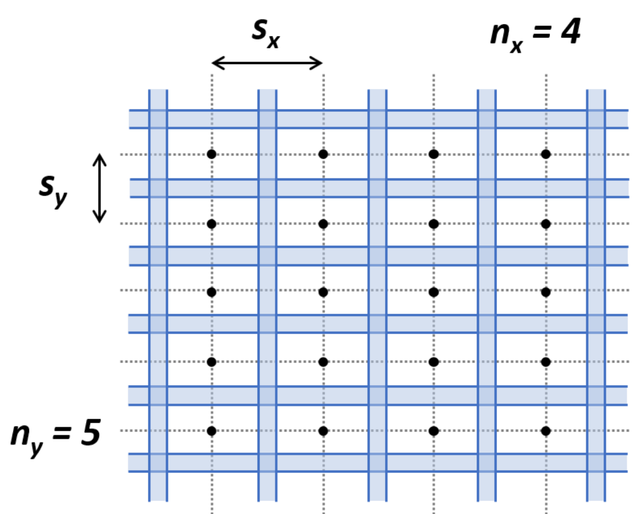


Figure 6. Illustrations of CNC motion parameterization for one layer.

The purpose of lateral overlap is the establishment of lateral (stereo) correspondences, across neighbouring images from the same layer. These correspondences are the basis of every photogrammetric surface reconstruction method. The purpose of medial overlap is the anchoring of error accumulation when registering images and reconstruction results from bottom layers. That is, the proposed approach constrains potential distortions in the final result, by requiring that the surface reconstruction is consistent with the appearance of the surface at larger distances.

The difference in configuration for surface scanning and surface digitisation is that for surface scanning and digitisation, 50% and 80% overlap are used, respectively. Naturally, this has direct implications for the number of images our memory can handle and, consequently, the amount of area we can scan or digitise at a given resolution.

3.4. Correspondence

A stereo correspondence is a match between a point in one image and a point in another image of the same scene taken from different locations. The establishment of stereo correspondences between image pairs is a fundamental, well-studied, and still open problem in Computer Vision. In our case, correspondences are sought in, either laterally or medially, neighbouring images. Moreover, sequences of correspondences or "feature tracks" are also sought. The process is described below, in steps.

3.4.1. Image key points

Image correspondence is key point-based. We selected SIFT [48] as the baseline, but any other more suitable keypoint flavour can be used instead. Key point features with content descriptions are detected in all images. In the implementation, Scale-Invariant Feature Transform (SIFT) [48] features are employed; however, any other type of key point features can be used.

3.4.2. Stereo correspondences

To establish correspondences between images of a stereo pair, we use the detected key points and search for keypoint matches. As the relative pose of the two cameras is approximately known, the search space is significantly reduced due to the use of epipolar constraints. Then, matching is guided by the keypoint descriptors, as follows.

1. **Epipolar constraint.** Initially, individual point matches are sought only within circular neighbourhoods, as predicted by scanner motion. Within these regions, the epipolar constraint is implemented as follows. Let x and x′ the homogeneous coordinates of a point in the image i and image j, respectively. These coordinates represent the 2D location of the same physical point in both images. The epipolar constraint can be expressed as $x'T \cdot F_{i,j} \cdot x < \tau_c$. Threshold τ_c is defined as an input parameter, relevant to imaging factors such as image resolution and imaging distance. The estimated F is used to implement the "epipolar constraint" and discard spurious correspondences. Thereby correspondences along a line rather than the entire image. Correspondences are approved only if the reprojection error is below threshold τ_c. Otherwise, they are discarded.
2. **Keypoint similarity.** To accelerate and robustify the similarity comparison of keypoint descriptors, the cascade hashing method in [49] is employed. To accelerate feature matching, the cascade hashing method in [49] is employed. Cascade hashing significantly reduces the feature search space by discarding non-matching features early and at a low cost. The method creates multiple hash tables, each capturing a different aspect of the feature descriptor. A global hash table captures high-level information about the descriptor. This table helps to quickly eliminate many non-matching candidates. At lower levels, more specific hash tables capture increasingly detailed information about the descriptors. By descending through the levels, the number of potential candidates decreases, and the matching becomes more precise.
3. **Keypoint matching.** Correspondence establishment is symmetrical, as in [50]. To filter erroneous correspondences a symmetrical, or left-to-right check [51] is employed. In other words, to establish a feature correspondence, we require that the same key point feature pair is found both when features of I_i are matched against I_j and vice versa.[5]

3.4.3. Feature tracks

To increase robustness, features are tracked across images, as follows. Let a physical point in the scene that was detected and corresponded in three images or, otherwise, two image pairs. The images of these pairs can be laterally or medially neighbouring. These two correspondences associate the same feature in the three images and comprise a "feature track". We keep tracks comprised of three or more features.

These tracks are used to further reject feature matches, using the fundamental matrices computed for the image pairs of these matches. Using the epipolar constraint, the position of each feature in the track is checked to determine its consistency with the fundamental matrix of every pair. This check is implemented using the “chain” of fundamental matrices that are associated with the images in which each feature is tracked. The check projects a feature from the first image of a track to the last image of this track. If this projection occurs at a distance greater than τ_p then the track and all corresponding feature matches are rejected.

Feature tracks are not constrained in laterally neighbouring images, but are also established using correspondences across layers as well. This is central to achieving overall consistency in surface reconstruction. The reason is that the spatial arrangement of features in images of greater distance to the surface constrains the potential correspondences in images at closer distances. Through this constraint, erroneous correspondences are reduced.

The memory requirements for the succeeding tasks depend on the number of tracked features found. Depending on the number of input images the requirements for memory capacity may be large. It is possible to reduce memory capacity requirements at this stage, by discarding some of the feature tracks. If this is required, then it is recommended to discard the tracks with the fewest features. The reason is that they carry less information and they are, typically, much larger in number.

Despite the action of the constraints above, the correspondences found at this stage still contain errors. This step detects and further reduces spurious correspondences, aiming at a more accurate reconstruction result.

Features are tracked across images, as follows. Let N key point features that corresponded in N images. Let u_i, $i \in [1, N]$ their locations in these images. Points u_i, comprise a “feature track”. If the track contains spurious correspondences, then some of u_i shall not image the same physical point. No specific order is imposed on the elements of the feature track. Feature tracks are not constrained in laterally neighbouring images, but are also established using correspondences across layers as well. This is central to achieving overall consistency in surface reconstruction. The reason is that the spatial arrangement of features in images of greater distance to the surface constrains the potential correspondences in images at closer distances.

Each track is evaluated as follows. Let $F_{1,N'} = F_{1,2} \cdot F_{2,3} \cdot \ldots F_{N-1,N}$ a composite fundamental matrix of the pair of images that contain u_1 and u_N, that is the first and last element of the track. Images I_1 and I_N may not be directly neighbouring. Even if they are, still $F_{1,N'}$ is used instead of the $F_{1,N}$. The reason is to involve all correspondences referenced in the feature track and, this way, detect whether any of them is spurious.

Using $F_{1,N'}$, image location u1 is projected to image I_N, which contains location u_N. The location of this projection is up = $F_{1,N'} \cdot u_1$ and the distance of these two points is $\delta = |u_p - u_N|$. If the correspondences in the feature track are correct, then these points should approximately coincide as only calibration inaccuracy should account for δ. However, if the track contains erroneous correspondences then the two locations are expected to be grossly inconsistent and δ to be greater than the distance threshold τ_p. Thus, if $\delta > \tau_p$, then the feature track and all of its correspondences are discarded.

The memory requirements for the succeeding tasks depend on the number of tracked features found. Depending on the number of input images the requirements for memory capacity may be large. It is possible to reduce memory capacity requirements at this stage, by discarding some of the feature tracks. If this is required, then it is recommended to discard the tracks with the fewest features. The reason is that they carry less information and they are, typically, much larger in number.

3.4.4. Global optimisation

Once candidate correspondences have been established, a global optimisation step discards potentially spurious and inaccurate correspondences and refines camera pose estimates, as follows. The pairs of neighbouring images, let I_1 and I_j, within and across layers are found in the scan plan. The fundamental matrices, F, for all pairs of neighbouring images are computed. Matrix $F_{i,j}$ denotes the fundamental matrix for image pair (i,j). Matrix $F_{i,j}$ defines the geometric constraints that exist between points in image i and their corresponding epipolar lines in image j. Matrices $F_{i,j}$ are estimated from the available correspondences between image pairs (i,j). The estimate of $F_{i,j}$ is obtained using Random Sample Consensus (RANSAC) [52] and least squares fitting as follows to reject spurious correspondences, or "outliers", between images i and j. The remaining, "inlier", correspondences are used to estimate $F_{i,j}$, using least squares.

3.5. Surface reconstruction

A different reconstruction method is used to generate the surface scan (2D image) and the surface digitisation (textured 3D model). Both results are called reconstructions since they recreate either the surface colours or the surface structure and colours. Depending on the operation, the output is an image or a surface. Surface scanning generates an image mosaic and surface digitisation a 3D textured mesh of triangles.

3.5.1. Surface scanning

We call a map the imaged surface in pixel coordinates, in the coordinate frame of the mosaic to be created. The input to image registration is the proprioceptive estimates of the camera centres and the established point correspondences across laterally or medially adjacent images. The output is a set of projective homography transforms H_i, estimated for each image I_i, where i enumerates the images across all pyramid layers. These homographies associate image locations in each I_i to the corresponding locations in the mosaic.

Homography estimation World points C_i are the proprioceptively obtained coordinates for these locations in 3D space. Image points c_i are the image centres of images I_i. Initially, projective homography H_g is estimated across this map and the 3D grid locations Ci, using least-squares. For each pair of adjacent images I_i and I_j, we enumerate the correspondences between them using k and denote their locations in I_i and I_j as f_{ki},f_{kj}, respectively. The computation estimates the homographies, by optimizing the following objective function

$$\sum_i \sum_j (H_i fk_i - H_j f_{kj})^2 + \sum_i (H_g C_i - H_i c_i)^2. \ (1)$$

The first term is the conventional reprojection error metric for point correspondences. In that term, j enumerates the neighbours of I_i. The second term promotes compliance with the scanner coordinates. In Figure 7, the notation is illustrated.

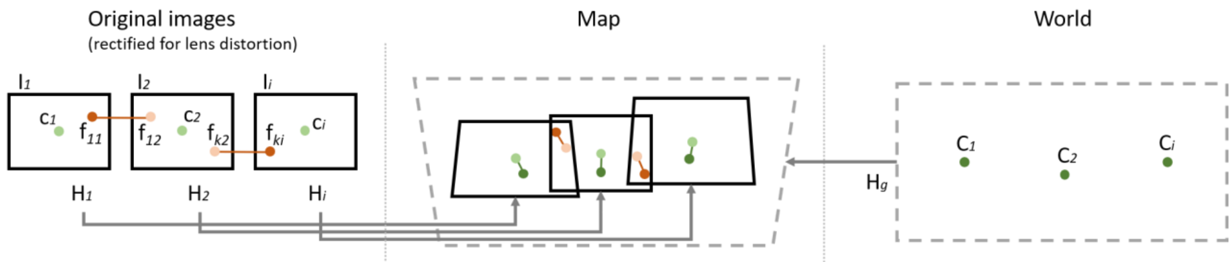


Figure 7. Illustration of objective function notation.

The projective homography has 8 free variables and, thus, the optimized variables are 8 times the number of images. The optimization capitalizes on the adjacency information contained in the pyramid data structure, to create a topological graph, as the one in **Error! Reference source not found.** (right). This graph has points Ci as nodes and, as vertices, their adjacency relations. These relations constrain the search space of the optimization. We employed the work in [53], which is a framework for least-squares optimization of an error function that can be represented by a graph and has been specifically designed for SLAM or bundle adjustment problems.

A benefit of using the aforementioned graph-based method is the robustness of "missing estimates". Such a case was encountered in Section 3.3.2, where we discarded unreliable homography estimates. Another is the case where the apex of the pyramid cannot be reached by the hardware. The latter case is encountered when covering wider areas, using multiple, laterally overlapping pyramids. There is no special treatment for running the method in this way. The difference is the lack of the cues that would have been provided by images acquired from a larger distance.

The high definition of the macro lens at close distances comes at a significant cost, which is its shallow depth of field, or otherwise, the distance range that imaged surfaces are in focus. This locus is a spherical shell centred at the focal point. For the macro lens, this shell is thin and small. The depth of focus is set foveatically so that the centre of the image is best focused. As the imaged surfaces are planar, the image periphery is less focused. Another common issue in mosaics is the occurrence of seams at the stitching boundaries. Both issues are treated the method in [54], applied for 32 spectral bands.

3.5.2. Surface digitisation

The surface of the object is reconstructed as a textured mesh of triangles. This mesh is comprised of two lists. The first is a list of 3D, floating point locations that represent the mesh nodes. The second is a list of integer triplets that contain indices to the first list and indicate the formation of triangles through the represented nodes. A texture image accompanies the mesh, along with a third list of 2D coordinates that represent the texture coordinates of each node; thus the third list has the same length as the first one.

Initialisation Using the obtained feature tracks as a connectivity relation, a connected component labelling of the input images is performed. After this operation, the largest connected component is selected. The rest of the images, which belong to smaller components, are discarded. The discarded components correspond to groups of images that are not linked, through fundamental matrices, to the rest. As such, they cannot contribute to the main reconstruction. They are, thus, discarded to reduce memory capacity requirements.

The reconstruction method requires an image pair, let (i’,j’), as the basis for the surface reconstruction. This pair is selected to exhibit a wide baseline to reduce reconstruction uncertainty and error [55]. The reason is that wider baselines result in more separation between the optical rays, reducing ambiguity in the 3D reconstruction. At the same time, the reliability of this pair depends on the number of feature correspondences existing in this pair. Thus, the initial pair is selected as the product of the baseline with the number of correspondences, or $(i',j') = \mathrm{argmax}_{i,j} (b_{i,j} \cdot k_{i,j})$ where $b_{i,j}$ is the baseline of pair (i,j) and $k_{i,j}$ the number of feature correspondences between I_i and I_j.

Sparse reconstruction A sparse reconstruction is first performed, using the camera poses in the scan plan. The purpose of this reconstruction is to refine these pose estimates. The Open Multiple View Geometry [56] is utilised to obtain a sparse point cloud, which includes implementations of Incremental Structure from Motion pipeline [57] and AC-RANSAC [58].

The purpose of the far-range images is encountered in this step. The large number of images, i.e. tens of thousands, required to photogrammetrically cover wide surfaces in large detail, pose accuracy problems that are not intensely pronounced when reconstruction involves a few hundred images. In the context of thousands of images, even small registration errors may accumulate leading to globally inconsistent surface structures. As correspondences include pairs of points across layers, the reconstruction process is guided to produce a structure that is consistent with far-range views. In the experiments, it is observed that these additional constraints reduce global distortion errors in the final result.

A bundle adjustment, using [59], is then performed in the end. This operation is adapted to optimise only the lens distortion and the extrinsic parameters for each image. The reason is to ease the convergence of the bundle adjustment optimization, as the intrinsic camera parameters have been accurately estimated in Section 3.1.

The method is formulated in Algorithm 1.

Algorithm 1 Incremental Structure from Motion

- **Require:** internal camera calibration (matrix K)
- **Require:** pairwise geometry consistent point correspondences
- **Ensure:** 3D point cloud
- **Ensure:** the camera poses
    - compute correspondence tracks t
    - compute connectivity graph G (1 node per view, 1 edge when enough matches)
    - pick an edge e in G with sufficient baseline (compare F and H)
    - robustly estimate essential matrix from images of e (AC-RANSAC)
    - triangulate t ∩ e, which provides an initial reconstruction
    - contract edge e
- **while** G contains an edge **do**
    - pick edge e in G that maximizes track(e) ∩ {3D points}
    - robustly estimate pose (external orientation/resection) (AC-RANSAC)

```
        o  triangulate new tracks
        o  contract edge e
        o  perform bundle adjustment (uses our initial estimation from the
           scan plan)
  •  end while
```

Finally, the scale factor is estimated at this step. In contrast to the methods in Section 2.2.7, this work estimates scale factors using the extrinsic parameters of the camera. This accuracy is relatively high because they are founded on the accurate motorisation of CNC devices. The initial estimates of camera locations obtained from the scan plan are already in metric units and, thus, so are the refined ones.

Textured mesh generation A mesh of triangles is computed based on the obtained sparse point cloud. In the implementation, the OpenMVS [60] library is utilised.

The "Z-buffering technique" [61] is employed to obtain depth maps, D_i. These maps are images that have the same dimensions as I_i, imaging the scene from the same viewpoint as Ii and with the same intrinsic parameters. In $D_i(u)$, each depth map stores the distance of the surface point imaged at $I_i(u)$ from the optical centre where I_i was acquired. Given the camera extrinsic, the pixels of the depth map are converted into a point cloud in world coordinates. In this way, depth maps Di are aggregated in a dense point cloud, using [62]. Following [63], a global mesh surface that best explains the dense point cloud is generated. Afterwards, this mesh is refined via the variational method in [64].

Next, index maps T_i are computed which store at each pixel the id of the mesh triangle, imaged at that pixel, that is in $T_i(u)$, encoded is the id of the triangle that is imaged at u. Using maps D_i and T_i, texturing the mesh considers the visibility of the triangle in each I_i. When a triangle is imaged in multiple views, then several choices can be made as to which view to select to acquire the texture or how to combine these multiple images of the same surface regions, into a better texture [65, 66].

Despite the accuracy improvements the mesh and the camera pose still contain residual errors. Albeit these errors are relatively small, they are well noticed by the human visual system as texture discontinuities. The phenomenon is more particularly pronounced when a large number of images is utilised. To address these inaccuracies, multiple views are combined using [67], which is a method designed for large numbers of images, as in this case. The generated texture is efficiently packed in a "texture atlas" (or texture image) as in [68].

3.6. Output

For surface scanning, the result is an image mosaic. This is typically stored in a single image. As this image can be very large in dimensions, tile-based rendering is supported, by exporting the result in smaller images ("tiles"). An appropriate tile-based viewer dynamically loads higher-resolution tiles for the specific area viewed, allowing for smooth and detailed exploration without needing to load a massive image all at once.

In surface digitisation, the resultant reconstruction is a textured mesh of triangles. This mesh consists of three lists and one texture image. The first is a list of 3D, floating point locations that represent the mesh nodes. The second is a list of integer triplets that contain indices to the first list and indicate the formation of triangles through the represented nodes. The third is a list of floating point 2D coordinates in the texture image, one for each node. This mesh is stored in Polygon File Format (PLY) format, in two files. The first file contains the mesh representation and uses the binary representation of the format to save disk space. The second file contains the texture, in the Joint Photographic Experts Group (JPEG) or Portable Network Graphic (PNG) image file format.

4. Digitisation of semi-transparent objects

There are methods for creating 3D images or models of real objects, often used in various industries and academic fields. These methods typically involve capturing 2D images and processing them into a 3D model, such as optical computed tomography, which uses light to create a digital volumetric model of an object. However, making accurate 3D models of translucent or transparent objects is challenging due to optical effects like light refraction and reflection, which can distort the final image. This issue is particularly relevant in fields like archaeology and quality control in manufacturing.

The proposed method provides a quick and simple way to create 3D images of transparent objects without using special substances or liquids, keeping the objects clean. It works by illuminating the object, taking pictures from different angles, and processing these images to produce a 3D model. Conventional methods for creating 3D images of translucent or transparent objects often involve covering the object with a substance to reduce transparency or immersing it in a liquid that matches its refractive index. These methods aim to avoid problems caused by light reflection and refraction rather than directly addressing them. However, these approaches add complexity, can be impractical, and might not be suitable for delicate objects, like those of archaeological or cultural significance, due to the risk of contamination or damage.

The novelty in the proposed approach lies in the specific setup being able to take advantage of the cone-beam xCT principle, the Radon Transform, by swapping the detector and light source in the original fan beam architecture. Additionally, we carefully selected the objects to be thin-walled, cylindrically symmetric objects in air, to enable the measurement without the use of index-matching liquid.

The scanning apparatus designed for the digitisation of transparent objects has the following core components: a white chamber, a motorised turntable, an LCD light source, and a high-resolution camera, configured as in Figure 8.

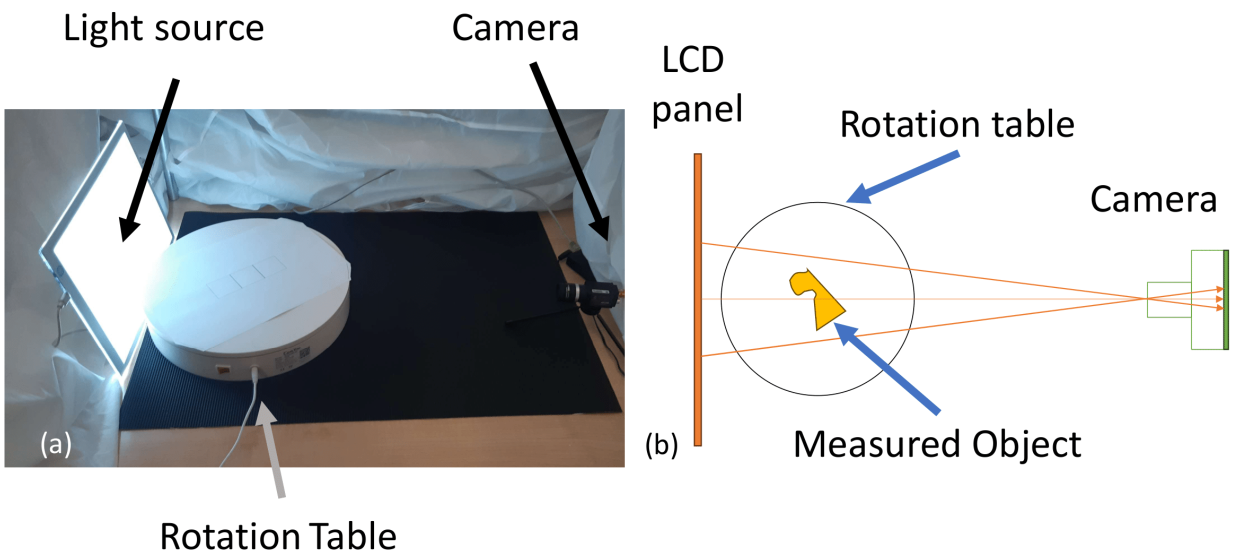
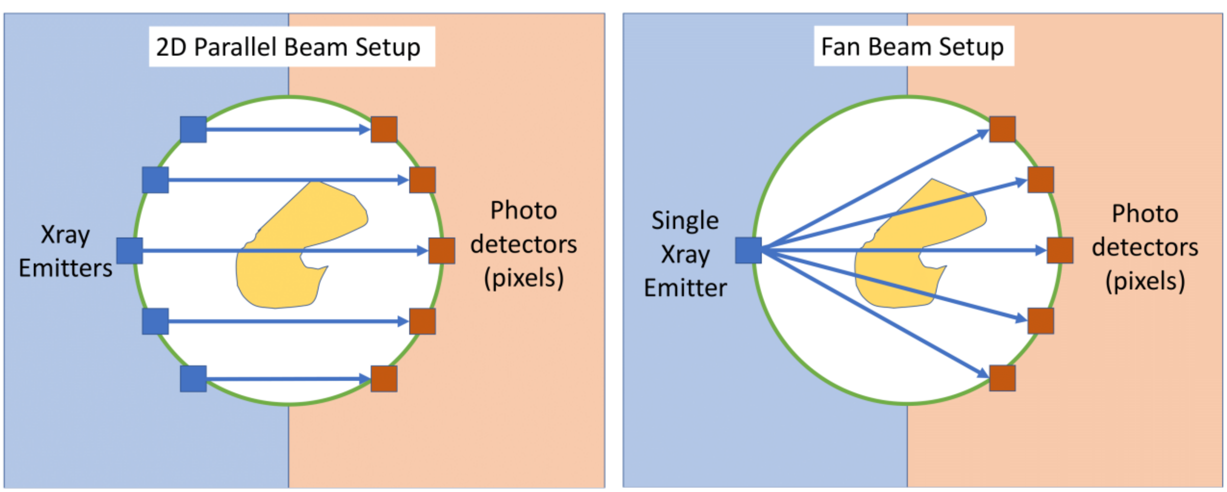


Figure 8. Experimental setup used for the acquisition of object photographs consisting of a screen, a rotation table, and an industrial camera. (a) Photograph and (b) diagram.

The operational principle is the following. In 2D, the available tomographic configurations are a 'fan-beam' and a 'parallel beam' setup. For 3D scanning, they correspond to 'cone' and 'parallel' setups. A diagram illustrating these configuration modes is shown in Figure 9. In this work, the 'cone' configuration was used by replacing the X-ray source with an optical camera, and the X-ray detectors with a field light source (LCD Panel). It is assumed that the camera used can be modelled by the 'pinhole camera model', and therefore we essentially use the 3D 'cone' beam geometry in reverse.



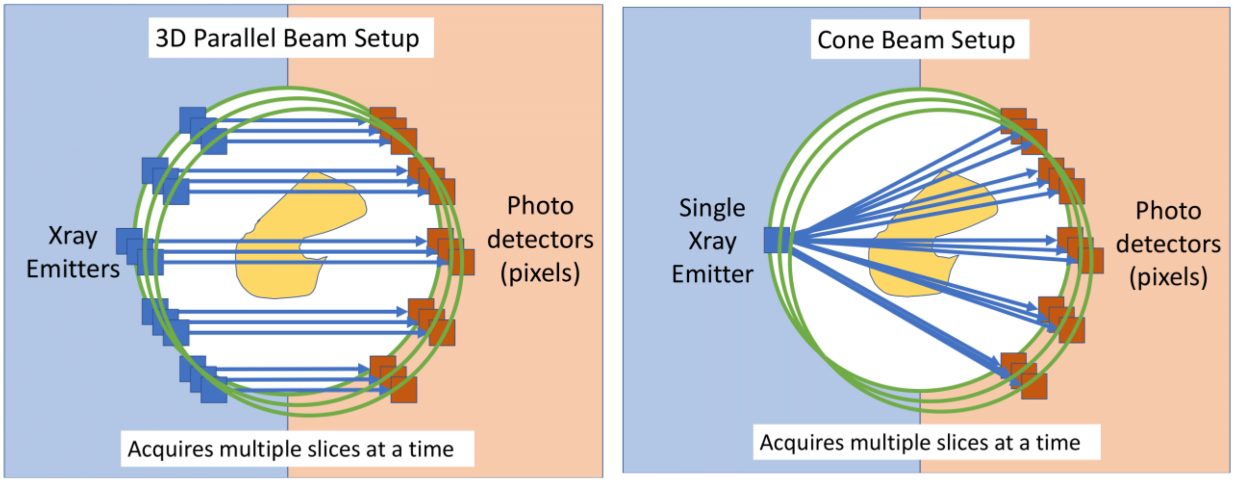


Figure 9. The four standard 3D tomography setups are available in the Astra Toolbox X-ray tomography software [111]. (a) 2D parallel beam setup. (b) 2D fan beam setup. (c) 3D parallel beam setup. (d) 3D cone beam setup.

4.1. Materials

The computational hardware used for our calculations was a single laptop, equipped with a single Intel(R) Core(TM) i7-1065G7 CPU @ 1.30 GHz 1.50 GHz CPU and an onboard NVidia graphics card NVIDIA(R) GeForce(R) MX250 with 2 GB GDDR5 memory. However, for the Astra toolbox calculations, only the CPU mode was used without GPU acceleration. This means that the speed of the computation can be significantly improved from the current time required (2–5 min per object) by the use of a stronger CPU and GPU and by enabling GPU acceleration.

In the ‘cone’ beam setup used, the following parameters are required to be set in the Astra Toolbox namely: CCD x and y pixel distance (set to 1), number of pixel rows in the detector (512), number of pixel columns in the detector (512), explicit projection angles (64 points of view), the distance between source and centre of rotation (70 cm), the distance between the centre of rotation and detector array (20 cm).

Opposite the camera, an LCD screen serves as the light source. The screen emits a uniform and diffuse light, which passes through the object. The choice of an LCD screen as a light source is particularly advantageous for transparent objects, as it provides consistent backlighting, essential for revealing the contours and internal structures. The camera is positioned opposite the LCD screen, aligned to capture the images of the object as it rotates on the turntable.

The scanning process occurs within a white chamber, which eliminates external illumination. This controlled environment ensures that the light interacting with the object is consistent and predictable. At the base of the chamber is a motorised turntable, upon which the object to be scanned is placed, allowing the object to be viewed and captured from multiple angles. The rotational motion is precise, smooth, and controlled, enabling the acquisition of evenly spaced images around the object. The turntable rotates the object incrementally, and at each predetermined position, the camera captures an image. This process continues until a full 360-degree view of the object has been documented.

4.2. Calibration

The same procedure as in Section 3.2 is followed for calibration of the intrinsic camera parameters.

The extrinsic calibration is similar but not the same, because this time the mechanism is rotational (as opposed to translational motion in Section 3).

A checkerboard pattern is placed on the turntable. The target is centrally aligned with the rotational axis of the turntable to ensure consistent reference points during calibration. The turntable, driven by the step motor, is rotated in precise increments. At each rotational step, the camera captures an image of the calibration target. The step motor's movement must be calibrated to ensure that each rotation corresponds to an exact, known angle, which is essential for the subsequent calibration steps.

Before or during the extrinsic calibration, the step motor is calibrated to verify that the rotational increments are accurate and consistent. This involves ensuring that the motor rotates the turntable by the correct degree with each step and that there is no slippage or misalignment. Calibration software can be used to verify the exact angular displacement per motor step.

In each captured image, the calibration software detects feature points on the calibration target. These points are used to determine the spatial relationship between the camera's image plane and the turntable's rotational axis. Based on the known geometry of the calibration target and the step motor's calibrated movements, an initial estimation of the camera's extrinsic parameters is calculated. These parameters include the rotation matrix and translation vector that describes the camera's orientation and position relative to the turntable.

The calibration software performs an optimization process, refining the estimated extrinsic parameters to minimize the reprojection error. This involves adjusting the rotation matrix and translation vector based on the detected feature points and the exact angles provided by the step motor.

Finally, the scale factor is calibrated as follows. A polished glass ball 80 mm in diameter (i.e., a 'lens ball' in specialist photography) was used for calibration, shown in Figure 10. Since the sphere is not hollow and has a convex shape, its outer surface was reconstructed by manually extracting its silhouette from each of the 64 axial rotations by the well-known 'visual hull' 3D reconstruction method [113]. The sphere's reconstruction was then loaded in point cloud processing software (Cloud Compare [114]) and its diameter was measured. The measured diameter was then divided by the true 80 mm sphere diameter to acquire the system's calibrated scaling factor.

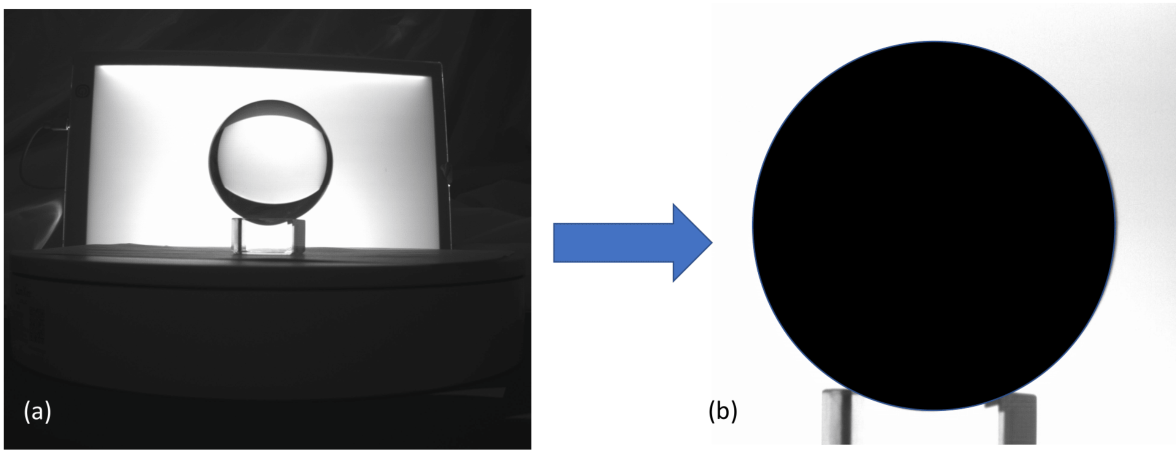


Figure 10. Back-illuminated images (a) from the setup camera were used to extract a reference solid black silhouette (b) and then reconstructed using the visual hull technique.

4.3. Image acquisition

To measure the shape of thin-walled transparent items with OPT without the use of refractive index-matching fluid, the following experimental sequence, similar to that of a typical OPT workflow, was used:

1. A rotation stage, a camera, and a field light source were set up as in the apparatus shown in Figure 8;
2. Light from a field source (LCD Panel) was projected through the object and registered at the camera;
3. The object is placed in the middle of the rotation stage and is rotated to acquire 64 rotational views around the object;
4. The images acquired by the black and white camera are inverted so that areas of high absorption are bright and areas of low absorption appear darker;
5. The images were processed by the Astra Toolbox's [111] X-ray CT reconstruction software into a density volume;
6. Thresholding the voxels of the 3D density volume to remove the low density of air. We are then left with the higher-density voxels of the object;
7. To extract a single surface from the thresholded density volume, we then post-process the object density volume slice-by-slice and line-by-line, to extract only the peak densities on each row of the image plane. Identifying the peaks represents the areas with the densest material and hence those of the sidewall (Figure 11);

8. The peak locations are then scaled using the calibrated scaling factor calculated in Section 11.1 and saved in a file in a point cloud format.

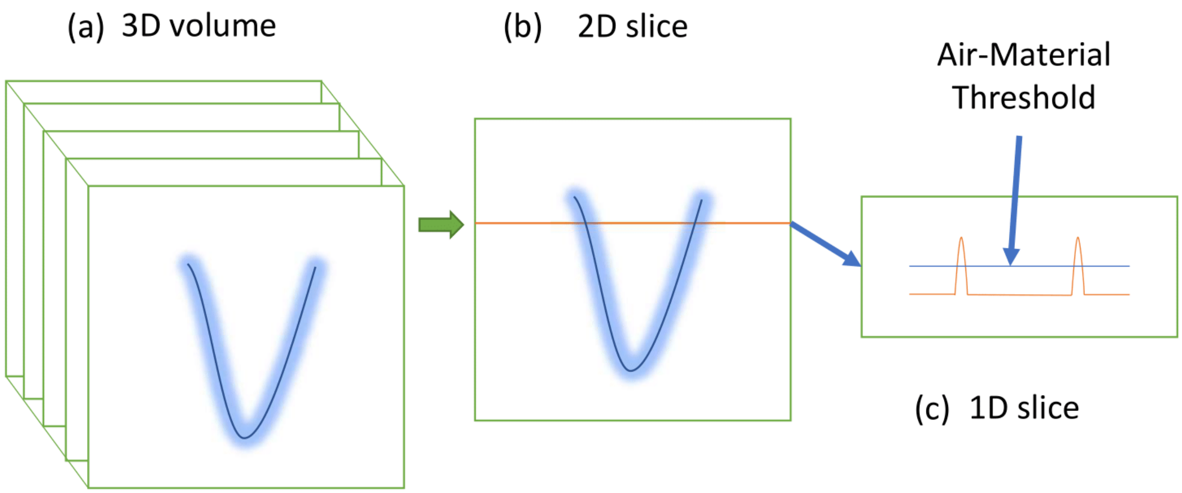


Figure 11. Process of thresholding (a) 3D voxel density to isolate transparent object surface. Each thresholded vertical voxel slice (b), is processed line-by-line (c), to extract the absorption peaks on each row.

4.4. Reconstruction

The filtered backpropagation algorithm [112] is used in Xray-CT to reconstruct the measured volume density from the photographs taken at each rotation angle. The density of each voxel g(x,y,z) was calculated per slice f(x,y) at a particular z height by the integral in Equation (2) [111].

$$ffbp = \int q\theta(x\cos\theta + y\sin\theta)d\theta \quad (2)$$

where θ is the rotation stage's angle, x and y are the particular slice's voxel locations and $q\theta(t)$ is the filtered Fourier transform of the detected image described in Equation (3).

$$q\theta(t) = \int P\theta(\omega)|\omega|ei2\pi\omega td\omega \quad (3)$$

where ω is the frequency in the Fourier domain, the spatial dimension of the 1D absorption measurement of each slice (the row of pixels of the photograph acquired at each rotation).

Using this approach we acquire the correct object shape, but it is scale-less. Therefore, here we need to scale the object appropriately by either measuring its exact distance from the camera or by scaling the results with the measurement of a known object. We selected the latter because it has to be performed only once.

4.5. Discussion

As shown in Section D.1 and Section D.2, the light propagation characteristics of visible light rays for these types of objects are close to what OPT requires to perform an accurate reconstruction. Some errors are still expected due to the minimal but non-zero refraction induced by the object's sidewalls.

5. Qualitative experiments

Qualitative results investigate the applicability of the proposed approaches to the problem and compare them to conventional methods. As in the previous sections, we organise the reported work in subsections per technical method. In some cases, the experiments took place before Craeft. In these cases, we provide the contents of the corresponding appendices.

5.1. Surface scanning

We report qualitative results in which we look for seamlessness, consistency in colour and tone, geometric alignment, and the absence of visible artefacts. The scan should appear as a coherent whole, with individual images blending smoothly without noticeable seams or abrupt changes in lighting or contrast. The colours and tones across the images should be consistent, avoiding any abrupt shifts that could disrupt the visual flow. Additionally, geometric alignment is crucial; images should be correctly oriented and aligned to maintain the integrity of the overall scene. Last, we check for the presence of artefacts like ghosting, blurring, or misaligned edges, which can detract from the quality of the mosaic. The results are organised into two classes, close-range and wide-are respectively.

5.1.1. Close-range

Work performed before initiation of Craeft. See Annex E.1.

5.1.2. Wide-area

We used our wide area apparatus in the application of textile scanning.

In the first example, Figure 12, we demonstrate the resolution of the result. We show a portion of the result in the original resolution. On the top, we show an image portion that occupies ~ 2 x 3 cm2. The fabric is made of coarse wool and, thus, the stray fibres appearing in the images are in the order of 35 μm (microns). On the bottom, we show the mosaic obtained at the same resolution, annotated with the image part shown on the top. A video inspection of this mosaic can be found at https://youtu.be/5knmizTFcFk.

Figure 12. Surface scan detail in the original resolution (top) and overall result (bottom).

In the second example, Figure 13, we test the image definition and demonstrate that the robustness against deviations from planarity is consistent in both wide-area and close-range implementations. To illustrate this, we scanned a rough textile, specifically a "rag rug"—a type of floor mat made by weaving or braiding strips of fabric, often recycled from old clothing or textiles. We selected a sample composed of coarse fabrics, which exhibited a depth variability of at least 0.8 cm. As shown in the example, both the regions of the fabric close to and farther from the camera remain in focus.

Figure 13. Scanning of an anomalous rag rug; detail (top) and overall result (bottom).

Followingly, we tested for robustness against repeated patterns. In Figure 14, we illustrate the success of the approach, which is of particular importance to us, for two reasons. The first is that the majority of textiles, which are our primary application domain for the wide-range scanner, bear repeated patterns. The second is that repeated patterns are challenging due to the ambiguity they introduce in matching points between images. The result illustrates the achieved robustness, despite imaging at multiple distances.

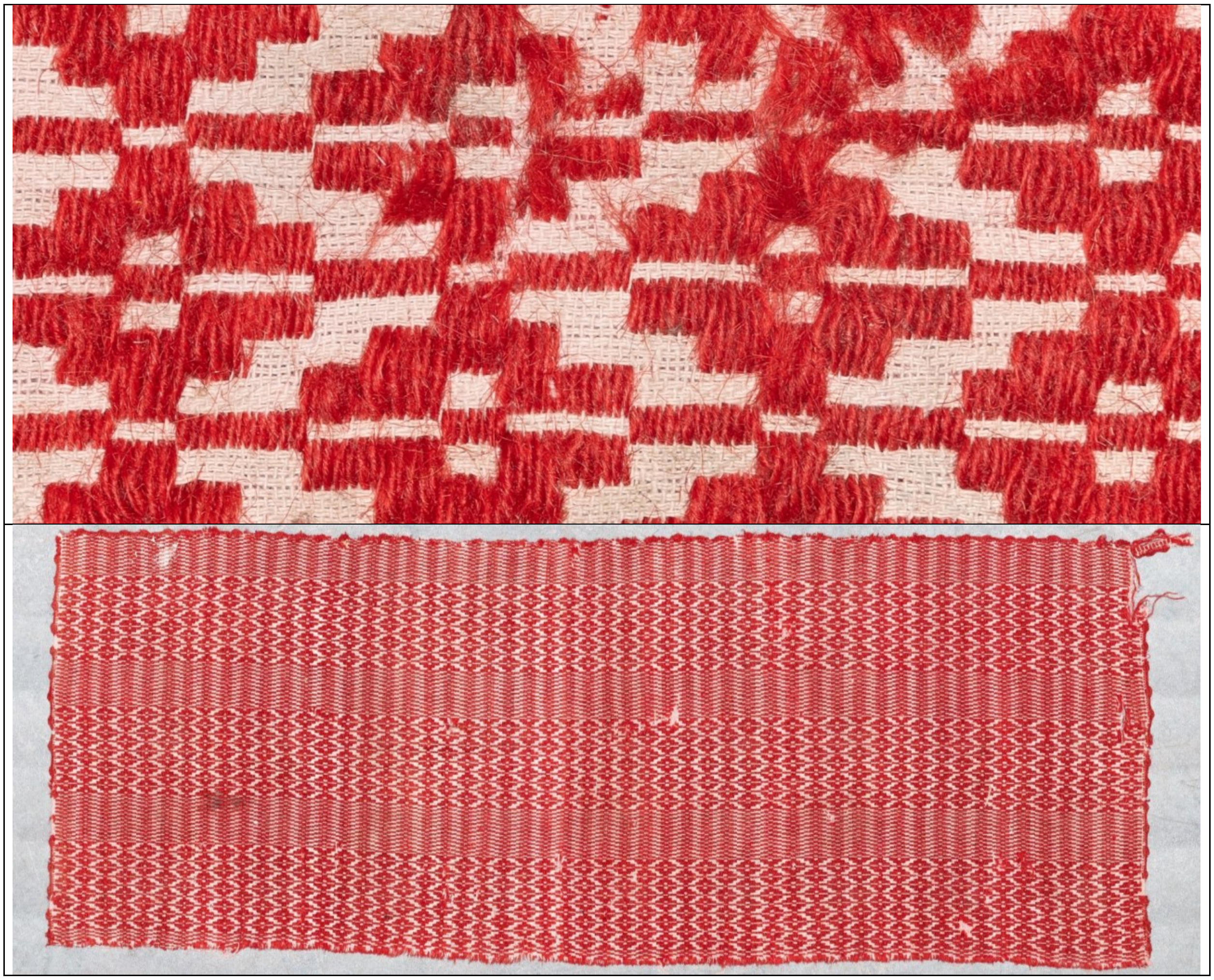


Figure 14. Scanning of a textile with repeated patterns; detail (top) and overall result (bottom).

5.2. Surface digitisation

The purpose of this experiment is twofold. First, to investigate the overall accuracy of the proposed approach. Second, to test its applicability in challenging reconstruction targets in terms of reconstruction appearance when surfaces made from shiny materials are digitised.

5.2.1 Structure and composition

Exploratory experiments with surfaces made from different types of materials are presented. The rationale of this experiment is to assess the capability of the scanning method in materials of different reflectance properties.

Materials The following surfaces were scanned, ranging in levels of shininess and texture. In all cases, the scanner area was $4 \times 3cm^2$.

Two pieces of medium-density fibre (MDF) wood board, let g1 and g2, with carvings each made with a Dremel rotary tool mounted with a 2mm cutting disk. In addition, piece g2 has a coarser marking made by dragging the rotating tool across the surface. Part g2 had some scribbles done on a blue ballpoint pen at its top right corner. The cuts were up to 2.5mm deep. Both pieces were matte. The texture was light but sufficient; the ballpoint pen markings enhanced it (see Figure 15, second from left).

An integrated circuit board, let g3, is made from mainly shiny materials, i.e. copper, solder, silkscreen, and glass fibre. The highest parts of the board were at 1.5mm from the glass fibre surface of the board. As evidenced in the middle column of Figure 15, the board was shiny in general and specifically at the regions of the flat, highly reflective, glass fibre board. Texture was available in general, but absent in some regions of the board.

An artwork (painting) made from 1.75mm threads of polylactic acid, let g4; this is the same object as in Figure 5. For the application of the thermoplastic on the artwork surface a 3D pen of 0.7mm nozzle size was used. The height variability was ≈ 3mm, but a couple of individual threads extended up to 8mm from the surface. Most importantly, this target was not a surface but an assembly of PLA threads; in some cases, one thread could be suspended above others at a height of a few millimetres. The texture is dense but the top of the plastic filament is shiny at all ranges, even the closest ones.

A composite surface, let g5, is made from a piece of aluminium and a piece of synthetic leather, the latter embossed. The synthetic leather material was moderately shiny and textured. The height variability was ≈ 3mm. The aluminium was shiny and with poor texture. Original images from the top and bottom layers are shown in Figure 15, along with corresponding reconstruction results.

No post-processing (e.g., smoothing, hole-filling, normal correction etc) was applied in the results shown below.

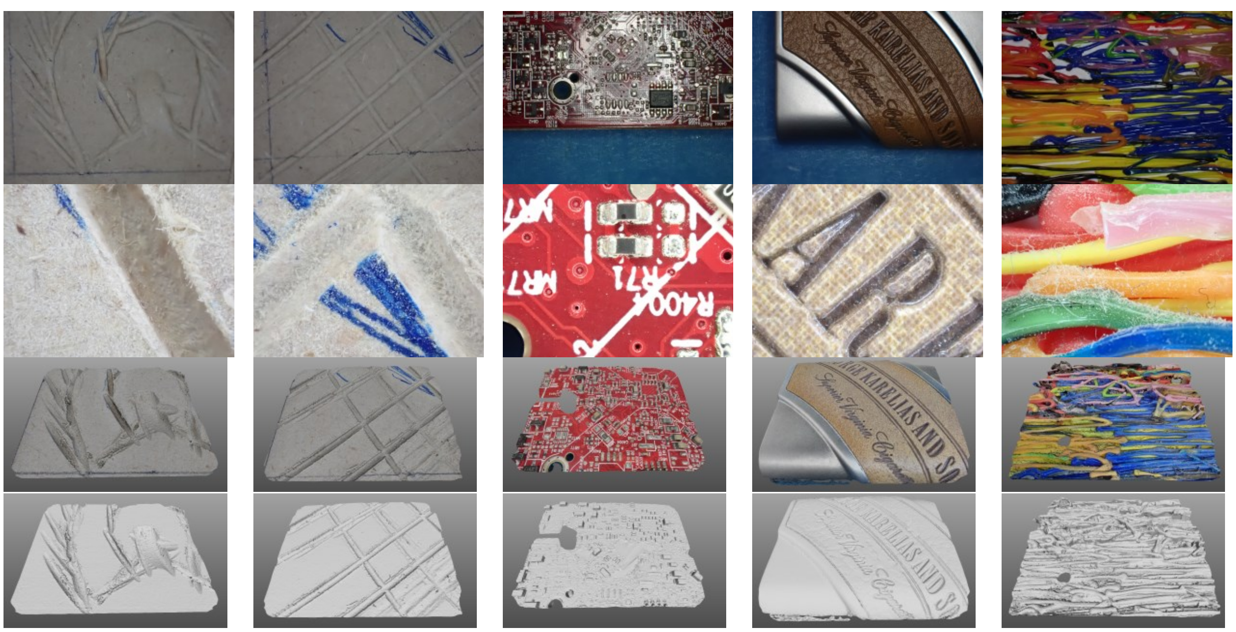


Figure 15. Rows, top to bottom: original images of targets g1 to g5, respectively from left to right, from the top elevation layers; original images of targets g1 to g5 from the bottom elevation layers, in the same order; textured reconstructions of targets g1 to

Observations The reconstruction of g1 and g2 did not exhibit issues. Despite that the MDF surface appears textureless to the eye in far-range images, in close-range imaging the faint structure shown in Figure 15 gives rise to feature detection.

The reconstruction of g3 exhibited a hole at the location of a region that contained solely fibreglass. At the corresponding region, the sparse reconstruction has a relatively lower density of feature points. Similarly, at the same region, the original images exhibit a lack of texture and aperture problems compared to the rest of the image. In the top row of Figure 16, shown are five images that assist this observation, from left to right. The left is a 2D mosaic of images overviewing the surface, computed as in [71]; it is shown as a reference. The second is the top view of the surface reconstruction. The third shows the reconstructed reference points in the sparse reconstruction; the hole region is indicated with a dashed rectangle. The fourth image is a magnification of the third in the area of the dashed rectangle. The fifth is an original image from the closest layer, centred above the hole region.

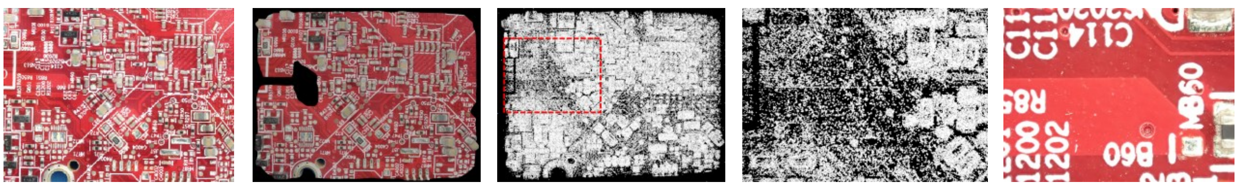


Figure 16. Left to right: 2D mosaic of scanned surface; Sparse reconstruction of g3; Detail at the hole region; Original image centred at the hole region.

Although the reconstruction of g4 appears complete from the top, in Figure 17, missing surface texture and gross geometric inaccuracies are visible, when the reconstruction is viewed laterally. The reason is that the scanned structure is more complex to be characterised than a surface because it contains filament threads suspended one above the other, leaving void space between them. As such, there are artefact locations that are not visible to the camera. Such an occasion is shown in Figure 17, where a pink thread is suspended above a green thread. In the figure, the same detail of the reconstruction is shown from a top and a lateral viewpoint, in two pairs of images. The left pair shows the reconstruction detailed, textured and untextured, from a top view. The right pair shows the same detail, from a lateral view, textured and untextured. In the lateral view, it is observed that the void between the bottom and the suspended filament is reconstructed. The structure however at regions not visible to the camera (underside of the suspended filament) is grossly inaccurate. The same regions are textured with black, as they do not appear in any image in the data.

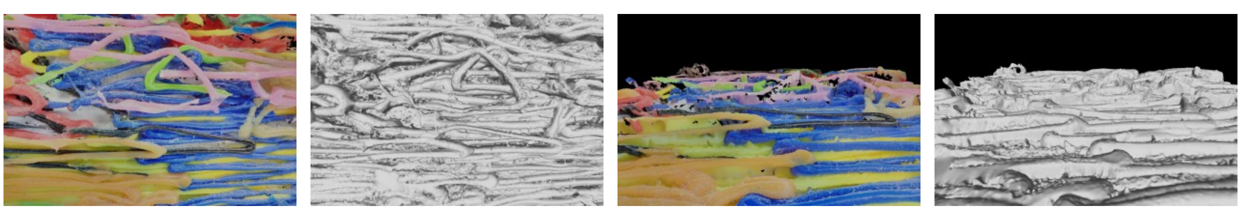


Figure 17. Details of the reconstruction of g4, from a top (left) and lateral view (right).

In g5, the shiny aluminium surface gives rise to texture, due to the minute surface structure that is visible at that close range.

Discussion Automated scanning and close-range photogrammetry are sufficient for the photorealistic and fairly accurate reconstruction at very high detail.

Close-range imaging When imaged in close range, surfaces exhibit less specularities and reveal texture, due to natural wear or inherent structure. Systematic image acquisition makes more probable the imaging of structure without specular artefacts, such as in the case of aluminium in g5. Still, in g3, the impeccability of the industrial and coated fibreglass hinders the method.

Lack of texture and visibility the shortcomings in g3 and g4 are known limitations of photogrammetry. Conventional methods for their solution require the photogrammetric measurement of such regions, require structured lighting to create texture and the acquisition of more views to fully cover the structure. The first is possible as a source of structured light can be mounted on the implemented setup. The second would require a more flexible motorisation mechanism. They are both left for future work.

Scope The image acquisition locations in this work are unable to reach surface regions that are not visible from top views, as in the example of g4. Although some of the structure is reconstructed this approach cannot guarantee the coverage of structures that are more complex than anaglyphs. As such, the recommended domain of the proposed approach is set in surface scanning. This is the reason that the approach is characterised as a surface scanning method in the title of this paper.

5.2.2 Global consistency

This experiment indicates the contribution of far-range images to the accuracy of the reconstruction result. As discussed in Section 3.4.2.0, photogrammetry with large numbers of images is prone to camera pose estimate error accumulation, resulting in reconstruction inaccuracies. To illustrate this point, this experiment uses the same images, as acquired by the proposed method but in two basic conditions and an additional one. In the first condition (C1), the proposed method is utilised. In the second condition (C2), the same images were inputted to the Pix4D photogrammetric suite. In the additional condition, (C3), we used a state-of-the-art variant [72] of the Neural Radiance Fields (NeRFs) method [73], which was also fed with the same images. Indeed, NeRFs are targeted at learning the radiance field, rather than achieving a geometric reconstruction. That is, NeRFs are not destined for measurements but for view synthesis. Still, they are quite useful when 3D visualisation is the application goal.

The reconstruction target was a handcrafted engraving at the end of the handle of a silver spoon. The engraving occupies an area of $45 \times 25mm^2$. In Figure 18 (top row), this item is shown from a top and a side view. It is observed that the spoon is undamaged and its handle at the location of the engraving is straight.

In Figure 18, indicative original images and the obtained reconstructions are shown. The second row of Figure 18 shows one image from each elevation layer. The third and fourth rows show the result of C1 and C2, respectively and in the following way. The left column shows the frontal views of the textured reconstructions. The rest of the columns show only their geometrical, untextured, structure. Specifically, the second column from the left shows frontal views. The second column from the right shows slanted views (30°). The right column shows side views (90°).

Figure 18. Rows, top to bottom: top and side views of a handcrafted silver spoon, on the left and right column, respectively; original images, one from each elevation later, from left to right; reconstruction using the proposed method (C1); reconstruction using an of

Global accuracy improvements are observed in the side views, as in the C2 the surface appears spuriously curved, while the original artefact is straight; see Figure 18 (top row, right). In C1, this effect is reduced. This improvement is attributed to the utilisation of feature tracks over the far-range images and their contribution to better camera pose estimation, which leads to more consistent reconstruction. This is attributed to the contribution of the far-range views constraint the accumulation of camera pose estimation errors.

The comparison between the two reconstruction approaches is not entirely fair, as different pre- and post-processing pipelines are followed in C1 and C2. To add to this unfairness, C2 is not programmed to search for medial correspondences as the proposed image acquisition approach is not guaranteed in generic photogrammetry implemented by off-the-shelf photogrammetric suites. The experiment provides the finding that the additional constraints provided by tracking features across distances are to the benefit of reconstruction quality.

The result of C3 is shown in Figure 19. As expected, the photometric appearance of the NeRF reconstruction; i.e., the colours in the reconstruction look more vivid. However, when its structure is inspected, it is observed that the surface structure is only coarsely captured. In defence of the NeRF method, the recommended image acquisition for its optimal use is different from ours and requires the acquisition of views around the object. However, this would require a significant change in the hardware and increase the cost of the devices, while still not providing fine structural measurements.

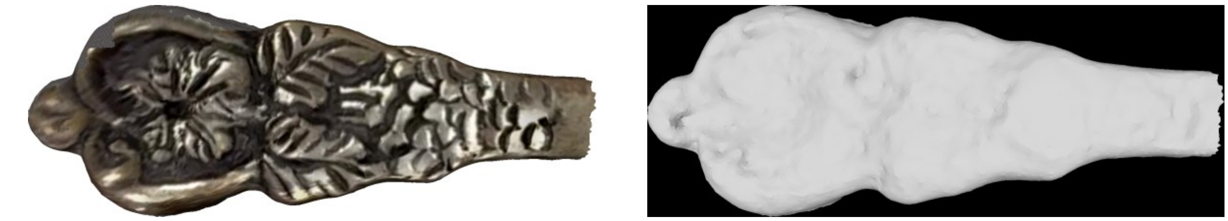


Figure 19. Reconstruction results of the artefact in Figure 19, using [72], for condition C3. Left: top view of the textured reconstruction. Right: top view of the untextured reconstruction.

5.2.3 Shiny, curved, and sharp surfaces

Stereo vision and photogrammetry are usually incapable of reconstructing even moderately shiny surfaces. The reason is that they reflect different parts of the environment from each viewpoint and thereby the “uniqueness constraint” [74] is not met. This incapability has been countered by the employment of additional algorithmic methods, such as photometric stereo [75], which requires additional and high-end hardware and illumination, the application of matting spray [76], or manual editing [77]. A study demonstrating the problems caused by highly reflective surfaces in multiple 3D reconstruction modalities, including photogrammetry, can be found in [78]. However, when the imaging range is very close, even shiny surfaces contain some texture. The purpose of the experiment is to assess reconstruction quality for metallic surfaces and find the limits of the proposed configuration of this type of surface.

Shiny metallic nuts and screws were scanned, because they feature multiple orientations and curvatures. These structural features are susceptible to illumination artefacts because they reflect light from multiple directions. Some of these structures are higher than the depth of focus range of the camera. For this reason, coarse sand was used as a substrate and the targets were partially submerged in it, to study the upper surface of these structures. Original images are shown in the top two rows of Figure 20.

The images exhibit specular reflectances which are reduced with imaging distance. Surface regions that contain such reflectances are constrained to the high curvature parts of the surface such as the creases of the bolts and the railings of the screws. In the images, these specularities are expressed as saturated (white) regions of pixels at the high curvature regions. Still, small imperfections, dust particles, and structural features give rise to some feature correspondences.

The obtained reconstructions are shown, in the same order, in the four bottom rows of Figure 20. It is observed that the reconstructions do not suffer from gross structural errors. It is furthermore observed that surface patterns are reconstructed, such as screw threads and bolt markings. However, when specular reflections are systematic over broad regions of pixels, the reconstruction exhibits artefacts that occur exactly at the high-curvature regions of the surface.

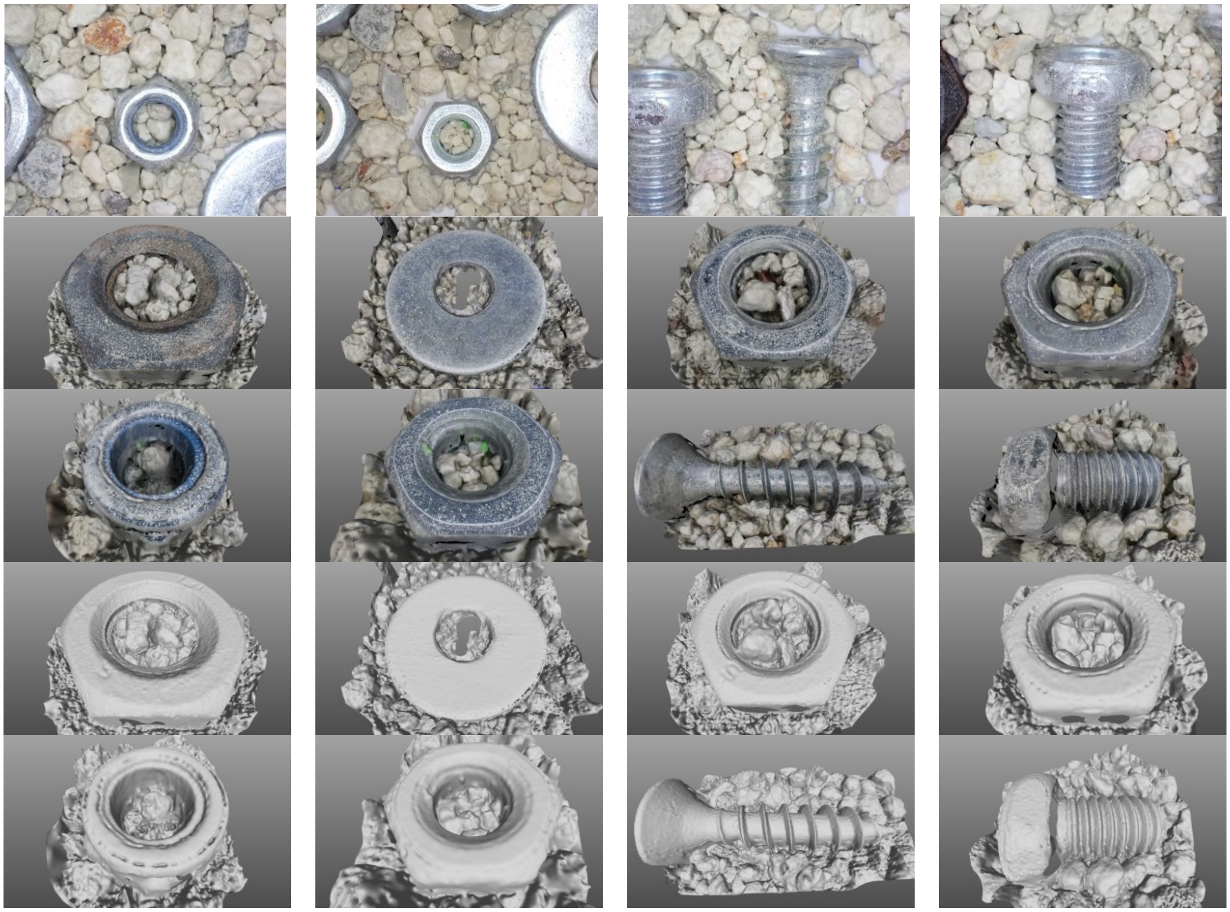


Figure 20. Top two rows: original images of metallic nuts and screws from the top layer. Middle two rows: textured reconstructions of the metallic nuts and screws in the top two rows. Bottom two rows: untextured renderings of the reconstructions in the two middle rows.

To compare against conventional photogrammetry, we have reconstructed the same scene using the Pix4D software and present the results in Figure 21 (left). As can be observed, very little of the scene is reconstructed. To indicate the difference with the proposed method in Figure 21 (right), shown is also the sparse reconstruction of the scene using the proposed method, from the same viewpoint. It is observed that the reconstruction obtained using Pix4D is poor and manages to reconstruct only a few parts of the scene. The reason is the lack of correspondences and the establishment of many erroneous correspondences, due to the shiny material of the targets.

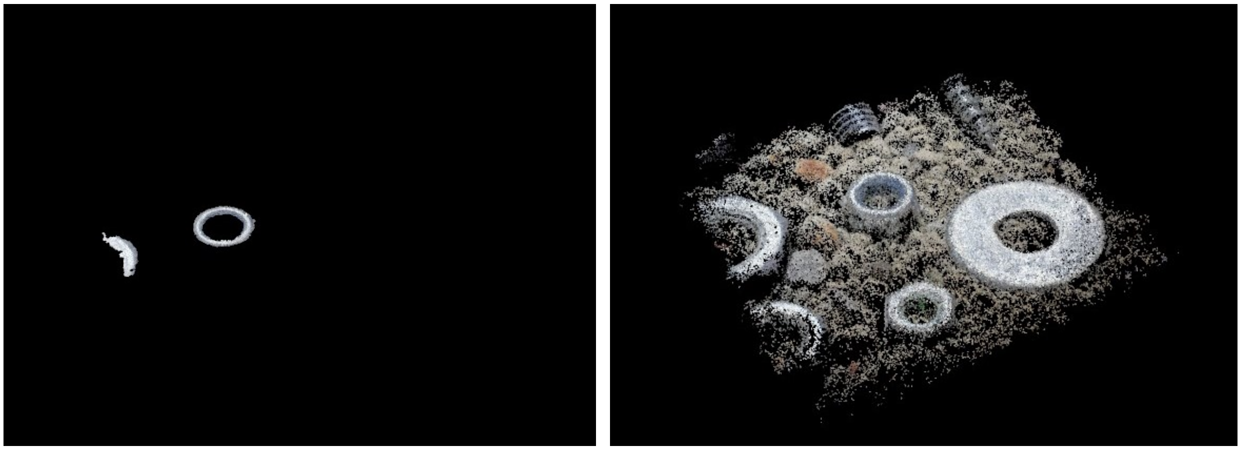


Figure 21. Left: reconstruction of a scene with shiny objects using Pix4D. Right: sparse reconstruction of the same scene using the proposed method.

5.3 Semi-transparent object reconstruction

The types of objects to which this measurement principle is most suited, are hollow, thin-walled, cylindrically symmetric objects, which do not induce considerable refraction as light traverses through them.

Hollow objects can be further subdivided into two categories, namely ‘shelled’ (objects whose internal and external surfaces are identical in shape and one is scaled down relative to the other by the size of the wall thickness) usually made of plastic. The other type is that of ‘non-shelled’ hollow objects (objects whose internal and external surfaces are not identical in shape and therefore have variations of thickness around the object), which are commonly made of glass.

Both types are of interest in manufacturing and cultural heritage applications, which this investigation aims to focus on. To test the category of ‘shelled’ objects that are commonly used in the beverage industry (e.g., soda, and water bottles), we selected a soda bottle and two water bottles (Figure 22a–c). To test the category of hollow ‘non-shelled’ objects, applicable mainly to cultural heritage and drink and food containers made of glass, contemporary glass cups (two liqueur and one wine glass) with and without embossed features were measured. (Figure 22d–f).

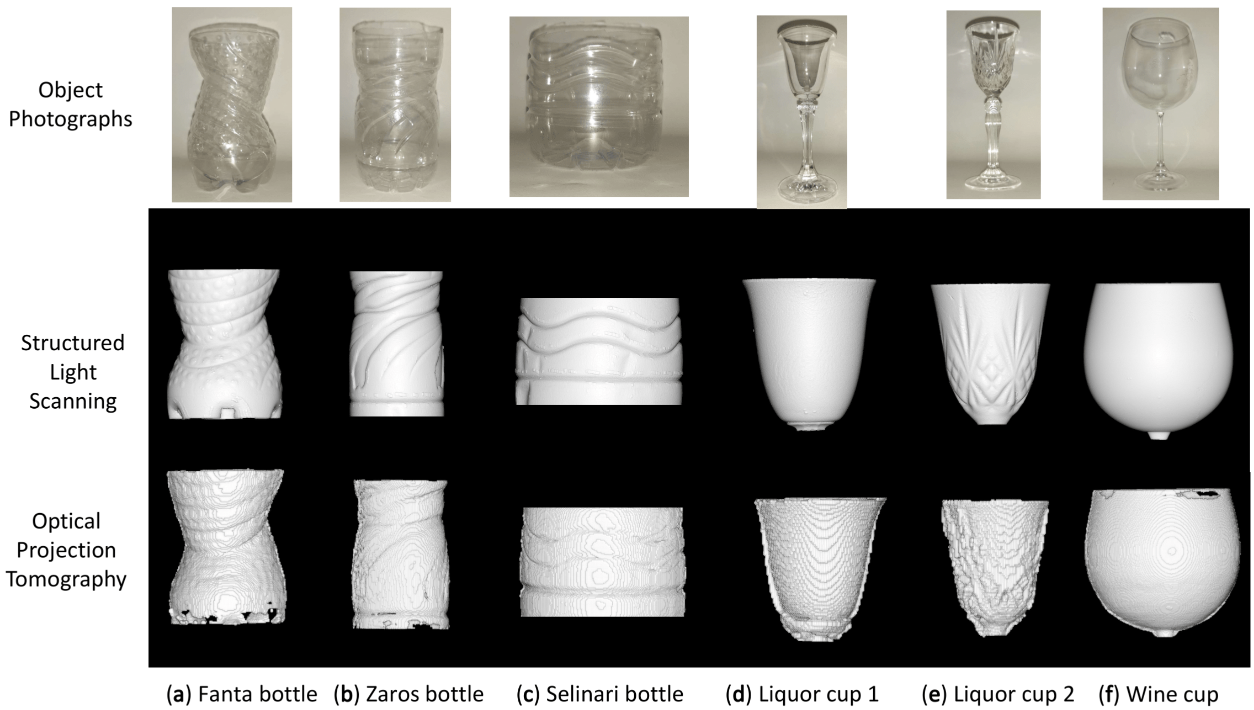


Figure 22. Photographs of transparent objects measured (top row), 3D reconstructions acquired from the "Shining 3D Pro 2X" industrial 3D scanner (middle row), and 3D reconstructions by our OPT method (bottom row). (a) Fanta bottle (b) Zaros bottle (c) Selinari bottle (d) Liquor cup 1 (e) Liquor cup 2 (f) Wine cup.

However 'non-shelled' hollow objects are more difficult to reconstruct since the thickness of the material is not consistent around the whole object and therefore of varying refraction. They contain areas where the light passing through the object encounters thick layers of material and therefore gets refracted significantly (neck, base, bottom of cup area). Additionally, a lot of cultural heritage items also contain embossed features which further add to the variation of material thickness in specific areas of the object. We nevertheless tested such items, to test the limits of the suggested OPT method.

Before reconstructing the hollow 'non-shelled' selected in this work, they were cut down for two reasons. The first is that their sidewall thickness needed to be accurately measured by use of electronic callipers (Table 4) and this could not be done from the mouth area, which is much thicker. The second reason is so that they could fit in the field of view of the camera used in the OPT setup, which could only measure objects of about 150 mm in height (Figure 10). To measure the hollow areas of the glass objects selected (wine and liqueur glasses cup area), they were placed inverted onto the rotation table with the hollow side down and the stem and base pointing up. In any case, only the hollow parts of the glass items were considered, and the stem and base which contain thick material areas by default were ignored in this study.

Table 4. Average thickness measurements of the plastic bottle sidewalls using electronic callipers (average of 8 measurements).

Object	**Thickness** (mm)
Fanta soda bottle	0.2
Zaros water bottle	0.1
Selinari large water bottle	0.1

Liquor glass 1	2.1
Liquor glass 2	2.3
Wine glass	0.7

6. Quantitative experiments

Quantitative results measure the computational performance and the accuracy of surface reconstruction. Quantitative experiments were challenging due to our inability to accurately manufacture targets of known size and with structural features at the fine scale required. For this reason, structures of known size were utilised.

As in the previous sections, we organise the reported work in subsections per technical method. In some cases, the experiments took place before Craeft. In these cases, we provide the contents of the corresponding appendices.

6.1 Surface scanning

Work performed before initiation of Craeft. See Annex E.2.

6.2. Surface digitisation

To measure the accuracy of reconstruction, targets of known size and structural features are utilised. To the best of our knowledge, there exists no benchmark for close-range photogrammetry at the scale dealt with by this work in general, and in particular for the specific image acquisition approach proposed. Therefore, we have used coins as reference targets so that other works can compare the same or analogous structures. Two experiments are reported that assess metric errors and global distortions.

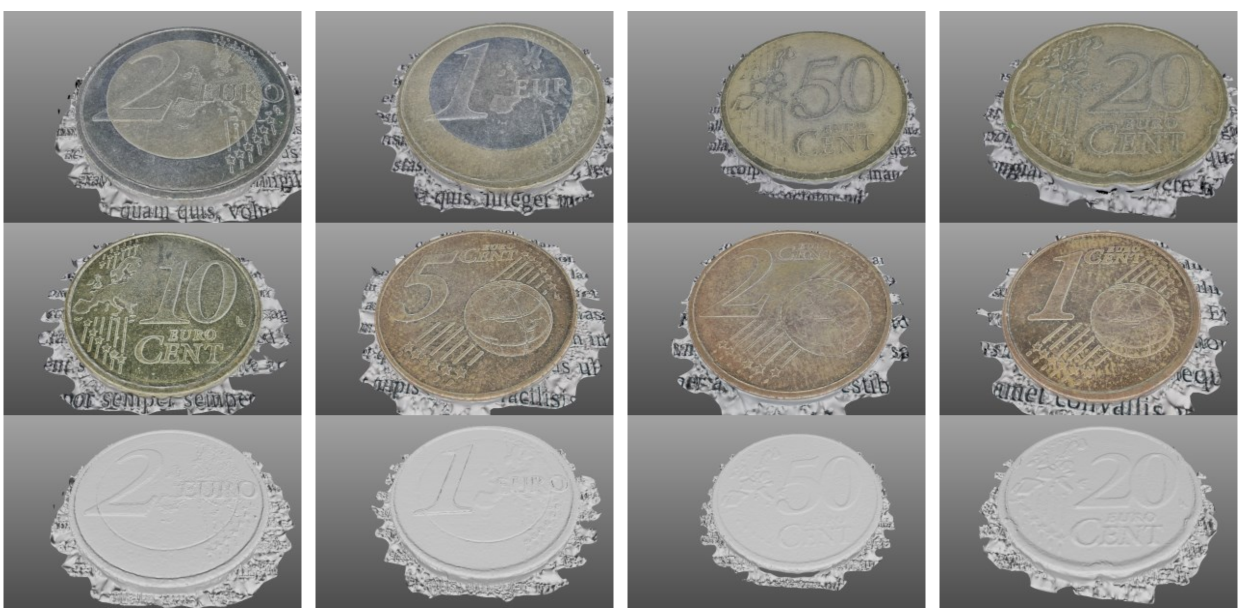
6.2.1. Dimensions

The purpose of the experiment was to measure reconstructed dimensions and compare them with the ground truth.

State-manufactured coins were used because they are of standard size and very accurately manufactured to avoid counterfeiting. The scanned coins belong to the Euro currency. All eight coins from this family were scanned and placed on a planar and textured surface. In addition, to the nominal dimensions provided by the manufacturer, the coins were measured with an electronic calliper. No larger than 0.01mm difference was found in these measurements. The measured dimensions were considered ground truth, as the coins were used and may have suffered distortions. The digital models are in metric units and their dimensions were compared to the ground truth dimensions of the coins. Their discrepancy is the measurement error.

Original images are shown in Figure 23. The top two rows show images from the highest elevation layer and the other two images from the lowest layer. The corresponding reconstructions are shown, in the same order, in Figure 24.

Figure 23. Original images of circular coins. The top couple of rows show an original image of the target from the top layer. The bottom couple of rows show an original image from the closest layer to the target.



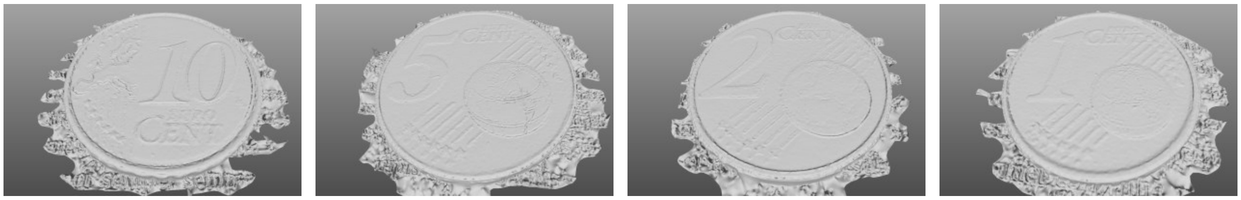


Figure 24. Reconstructions of the circular coins are shown in Figure 25. The top couple of rows show the textured reconstructions. The bottom couple of rows show the untextured reconstructions from the same viewpoints.

The dimensions and errors are reported in Table 5. The first column notes the currency value. Column "Nominal" reports the dimensions provided by the manufacturer (European Central Bank), column "Measured" reports our calliper measurements (ground truth), column "Measurement" reports the dimensions of the digital model, and "Error" is the percentage error. The thickness of the coins was measured as the distance of the supporting plane to the top face of the reconstruction. The average measurement error is ≈ 2.91%.

Table 5. Coin dimensions (diameter × thickness): nominal, measured, and reconstruction errors.

Currency	Nominal (mm)	Measured (mm)	Measurement (mm)	Error (%)
1c	16.25 × 1.67	16.25 × 1.66	16.51 × 1.70	1.58 × 2.19
2c	18.75 × 1.67	18.76 × 1.66	19.45 × 1.68	3.69 × 1.50
5c	21.25 × 1.67	21.26 × 1.67	20.21 × 1.73	4.95 × 3.58
10c	19.75 × 1.93	19.76 × 1.91	19.06 × 1.96	3.56 × 2.87
20c	22.25 × 2.14	22.27 × 2.14	21.51 × 2.23	3.40 × 4.00
50c	24.25 × 2.38	24.27 × 2.38	25.41 × 2.26	4.70 × 4.97
€1	23.25 × 2.33	23.32 × 2.34	22.90 × 2.31	1.80 × 1.43
€2	25.75 × 2.20	25.75 × 2.19	26.06 × 2.16	1.19 × 1.24

6.2.2. Aspect ratio

The purpose of this experiment is to measure any deviation from the circular shape of coin edges to assess global distortions in the reconstruction.

To assess global distortions in the reconstruction, the orthoimages of the coin reconstructions were used. These orthoimages are perpendicular projections of the reconstruction upon a hypothetical plane parallel to the surface. These images are "map accurate" in that they do not contain perspective distortions included in the original photographs. Depth maps computed for orthophotos share also this property.

The depth map for an orthophoto of a frontal view of the reconstruction was computed using Z-buffering [61]. Canny edge detection [79] was performed on that map, detecting depth discontinuities. Circles were robustly detected using RANSAC to eliminate outlier edges. The inlier edges, let Ein, were used to fit circles, using least squares. In Figure 25, shown are the inlier depth edges Ein superimposed on the orthoimages of the textured reconstructions.

Figure 25. Reconstructions of the circular coins are shown in Figure 25. The top couple of rows show the textured reconstructions. The bottom couple of rows show the untextured reconstructions from the same viewpoints.

In Table 6, deviations of the detected edges from the fitted circle are reported, as the mean distance of these edges from the fitted circle and their standard deviation. The small deviations and the visual results indicate that the circles were appropriately fitted. The average error is ≈ .58p.

Given the representativeness of the circle, the aspect ratio of the detected points is computed to assess whether the reconstruction is isotropic. To compute this the leftmost (p1), rightmost (p2), top (p3), and bottom (p4) points of inliers Ein were found. Then, aspect ratio $|p2 - p1|/|p4 - p3|$ is an indicator of anisotropy over the horizontal and vertical surface dimensions. The last column of Table 6 reports these ratios, indicating a mean aspect ratio of 0.998 or 0.0018%. deviation from isotropy.

Table 6. Radius, mean circle fit error, and standard deviation, for each measured coin.

Currency	Radius (p)	Error, std (p)	Aspect ratio, (%)
1c	224.58	0.49 (0.38)	0.995
2c	243.45	0.69 (0.41)	1.000
5c	232.66	0.48 (0.35)	1.000
10c	234.63	0.67 (0.41)	1.008
20c	238.50	0.66 (0.43)	0.985
50c	248.21	0.51 (0.37)	0.997
€1	231.52	0.71 (0.40)	0.997
€2	237.88	0.41 (0.33)	1.000

6.2.3. Surface structure

The purpose of this experiment is to assess the accuracy of the reconstruction of surface structure. To achieve this, we have used the digitisation of a coin by a scanner that is more accurate than photogrammetry and used that digitisation as ground truth. By comparing this higher accuracy model, we obtain a measure of the accuracy of our method. To quantify the error of the proposed method, we used cross-correlation to measure the similarity between two depth maps. To measure the similarity of the depth edges in these maps, we have used the Haussdorff distance [80], which is a metric that quantifies the similarity between two sets of points or shapes.

The proposed method and the Pix4D reconstruction of a 2€coin were compared to a higher-quality scan of the coin. This higher-quality scan was produced by the TetraVision company using an elaborate scanning technique that involved binocular (stereo) imaging and structured light, using the "Atos III Triple Scan" 3D scanner manufactured by the GOM company. This dataset is available in [81]. The coin was clamped in a fixture with reference points to ease the registration of partial scans and the scanning mechanism included an automated tilt and swivel unit to image the coin from different angles. In addition, the coin was covered in a water-based, transparent anti-reflex spray. The three scans were named S1 for the proposed method, S2 for the Pix4D reconstruction, and S3 for the TetraVision reconstruction.

The depth maps of the three scans were produced and the top surface region of the coin was isolated, to compare the same surface regions. In these maps, the following measurements were acquired. First, we computed the cross-correlation between image pairs (S1, S3) and (S2, S3), which were 99.84% and 86.43%, respectively. Second, we performed Canny edge detection [79] on all three maps, using the same parameters. The spatial arrangements of the obtained edges were compared using the Haussdorf distance for the same image pairs. The results were 73.53p and 110.63p for image pairs (S1, S3) and (S2, S3), respectively. In Figure 26, the textureless reconstructions and the edge detection results are shown.

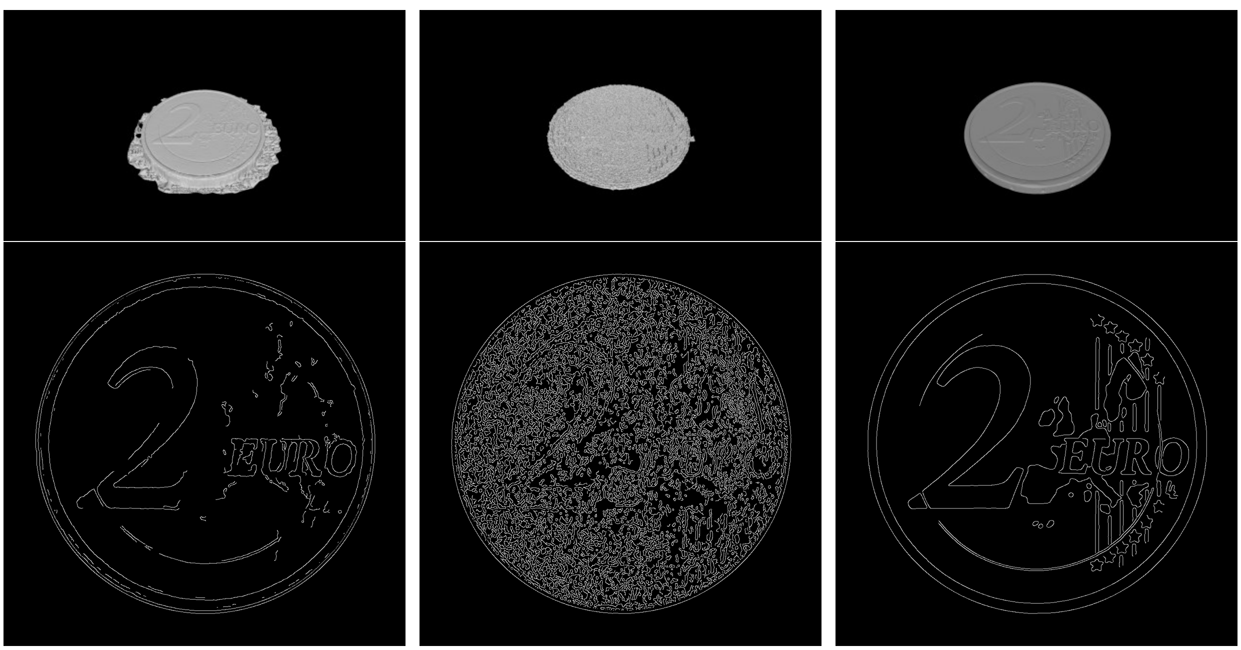


Figure 26. Top: surface reconstructions of a 2E coin. Bottom: edge detections on the depth maps of the reconstructions. Left to right: proposed method, Pix4D, ground truth.

Qualitatively, the comparison of S1 with S2 offers similar observations as those obtained in Section 5.2.2, that is S2 exhibits significant levels of noise. This is also observed by the structure of edges in Figure 26 (bottom, middle). Quantitatively, both the correlation and the Haussdorf measures of dissimilarity indicate the greater accuracy of the proposed method, as it provides more similar results to S3.

6.2.4. Nuts and screws

Metallic nuts and screws were scanned in a second experiment for reconstruction accuracy. Original images are shown in Figure 20 and their reconstructions are shown, in the same order, in Figure 20. These objects are enumerated from 1 to 8 in order of appearance in these figures. The ground truth dimensions of these items were measured using a calliper.

In this experiment, a non-planar substrate was used to place the scanned items. Specifically, coarse sand was used. The reason for this choice was twofold. First, to secure these items against any slip motion due to the substrate motion along the yy′ axis. The second reason is related to the 1cm depth range limit of the camera. As some of the scanned screws exhibit a thickness greater than this limit, they are partially submerged in the sand to restrict the depth range variability of the imaged scene.

In Table 7, we compare the measured dimensions of the nuts and screws with their ground truth dimensions. The order of the measurement is the same as in Figure 20. In this table, the first column reads the enumeration of each object as defined above.

The second column reads the ground truth dimensions of the targets. The third column reads the dimensions of the target measured from the reconstruction. The third column reads the percentage error. The dimensions reported are the diameter of the target and its thickness, in that order. In these rows, the first dimension reads the “height” of the screw (its longest dimension) and the second dimension reads the diameter of its “head”. The average error is ≈ 2.8%.

Table 7. Measurements of nut and screw diameter measurements, manufacturer dimensions, and percentage measurement errors (mm).

#	Dimensions (mm)	Measurement (mm)	Error (%)
1	16.80 × 7.70	17.29 × 7.52	2.89 × 2.36
2	17.80 × 1.50	18.15 × 1.53	1.97 × 2.08
3	12.80 × 6.30	12.63 × 6.54	1.36 × 3.73
4	9.80 × 4.80	10.15 × 4.97	3.59 × 3.54
5	7.80 × 4.80	7.95 × 4.64	1.94 × 3.43
6	6.90 × 3.00	7.18 × 2.89	4.08 × 3.65
7	24.50 × 9.80	25.00 × 10.13	2.03 × 3.33
8	13.00 × 10.00	12.70 × 9.52	2.27 × 4.76

6.3. Semi-transparent object reconstruction

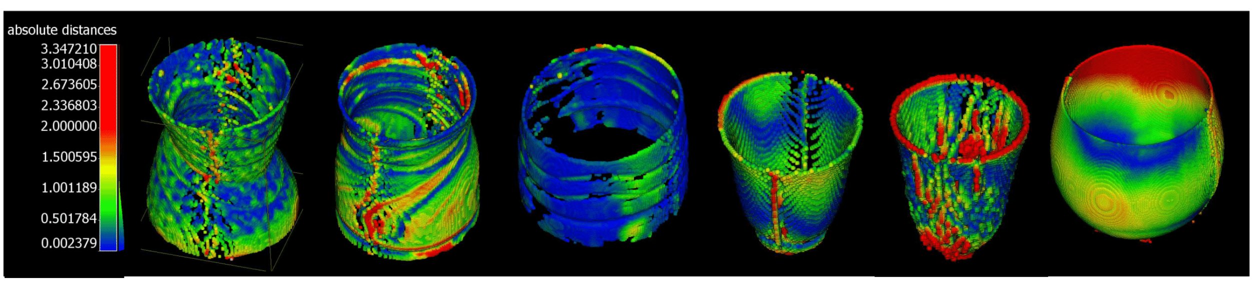
The evaluation of 3D requires volumetric error measures to quantify the deviation between the actual 3D structure of the object and its reconstructed version.

6.3.1. Accuracy

To obtain reference 3D reconstruction results for the outside shapes of the measured items, we scanned the objects with a conventional white light structured light optical scanner used for industrial purposes, called the Shining 3D Einscan Pro 2X scanner (Figure 27). The reference scanner was calibrated to an accuracy of ± 22μm using the calibration plates, which are provided by Shining 3D. To use this scanner, the transparent objects needed to be coated with the 'AESUB blue' opaque spray coating.

Figure 27. Static structured-light experimental setup used for acquisition of the reference 3D reconstructions.

Then, using the CloudCompare [114] point cloud software, both the reference point cloud reconstructions and the point cloud reconstructions created by the OPT process were first aligned by hand and then aligned more accurately via Iterative Closest Point (ICP) to an error tolerance of 10−4. Finally, to extract the dimensional error, the residual point cloud distances were calculated. The point cloud errors are depicted as color textures on the OPT point clouds in Figure 28 and the numerical average of the point cloud distances for each object is reported in Table 8.



(**a**) Fanta bottle (**b**) Zaros bottle (**c**) Selinari bottle (**d**) Liqueur cup 1 (**e**) Liqueur cup 2 (**f**) Wine cup

Figure 28. Use of colour texturing on the OPT reconstructed point clouds of the objects in Figure 33, which visually represent the value of the closest distances of each point in the OPT reconstruction from the reference point cloud.

Table 8. Results of comparing the point clouds acquired by MVS, NeRF, and OPT on the transparent objects shown in Figure 33 to the reference measurements acquired from an industrial-grade 3D scanner.

Object	**MVS** (mm)	**NeRF** (mm)	**OPT/Ours** (mm)
Fanta soda bottle	6.1	3.4	0.3
Zaros water bottle	6.8	5.8	0.4

Selinari large water bottle	8.6	3.2	0.3
Liquor glass 1	5.9	3.7	0.5
Liquor glass 2	3.1	2.2	0.8
Wine glass	10.4	5.6	1.5

There are multiple methods of comparing point clouds [115, 116, 117], using point-to-point, point-to-mesh, and mesh-to-mesh strategies. We opted for using the closest point-to-point distance, rather than comparing point-to-mesh or mesh-to-mesh, because in our case, the reference point clouds created by the structured light scanner were extremely dense, and therefore it was not necessary to create a mesh surface to accurately compare the point clouds, as suggested in [116].

For hollow 'non-shelled' objects, the average distance errors (± 0.92 mm) as expected were higher on average than that measured for the hollow 'shelled' objects (± 0.34 mm). This fact alone, however, was not reflective of the much wider type of errors experienced on these objects due to intense refraction effects, which manifested as distorted embossed shapes (Figure 22e), artificial 'ghost material' partially filling up the hollow areas (Figure 29), and the reduction the object's size (Figure 29).

The maximum precision expected from the specific OPT setup, in general, was calculated by dividing the available camera pixels by the field of view and was found to be 0.5 mm per pixel. Therefore the minimum dimensional error expected on the lateral and vertical distances is half this value, ± 0.2 mm. This sanity check is in line with the measurements we collected (Table 8). The measurement with the lowest error achieved was an average point cloud distance of ± 0.3 mm between the OPT and the reference reconstructions for the 'Selinari' water bottle (Table 8).

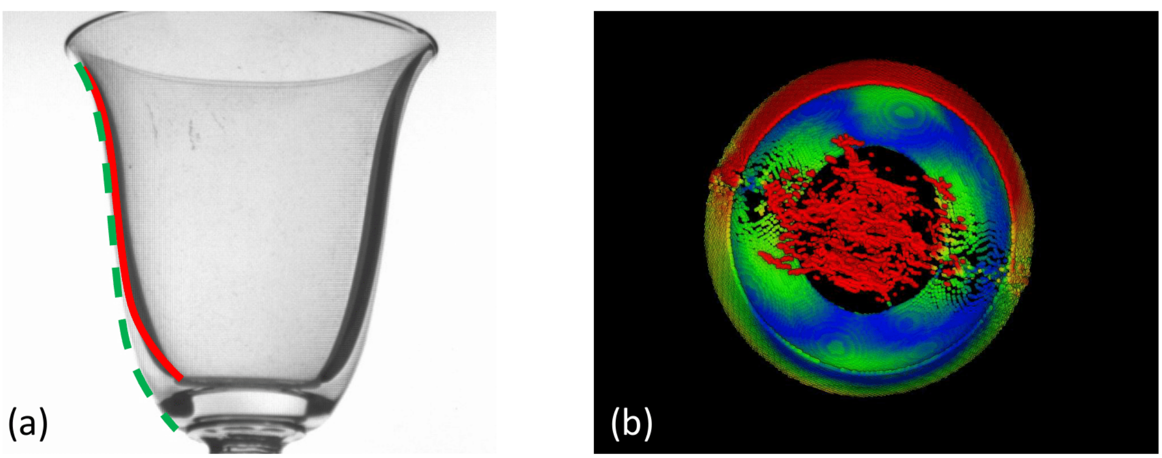


Figure 29. (a) Image from the object in Figure 33, shows a severely refracted liqueur glass outer edge (red line), compared to the actual surface edge (dashed green line), which manifests as an erroneous material density and therefore erroneous surface shape, (b) top view of wine glass 3D reconstruction in Figure 22f showing the erroneously partially filled hollow area of the wine glass cavity.

6.3.2. Comparison with visual hull reconstruction

In this work, we use OPT to extract the shape of thin-walled objects as a single surface, and it is, therefore, worth qualitatively comparing it to another technique, the visual hull, which is very similar and used during calibration. The visual hull technique can operate in the visible spectrum without the use of spray coatings. It is well known that it can extract only the external convex shape via the use of silhouettes [113], which is why it was used to measure our calibration sphere in Section 11.1.

For hollow non-convex objects, however, it is known that this technique cannot be used as it produces a solid convex 3D shell around the object. For example, when this technique was used in [103] to reconstruct a wine glass, the opening of the hollow end was covered. Similarly, it is known that any convex cavities (small craters) around the external surface of the object are 'filled up' due to the nature of the visual hull technique. OPT being a tomography technique, similar to X-ray CT, does not have these drawbacks as it can reconstruct hollow objects and can also deal with convex surface structures.

What is more, OPT can also measure internal surfaces. To demonstrate OPT's ability to measure internally, we placed two cut-offs of plastic bottles, one inside the other, and the reconstructed result is shown in Figure 30. However, since we could not perform a reference measurement for internal surfaces (e.g., using an X-ray CT machine), it was not possible to confirm the achievable accuracy.

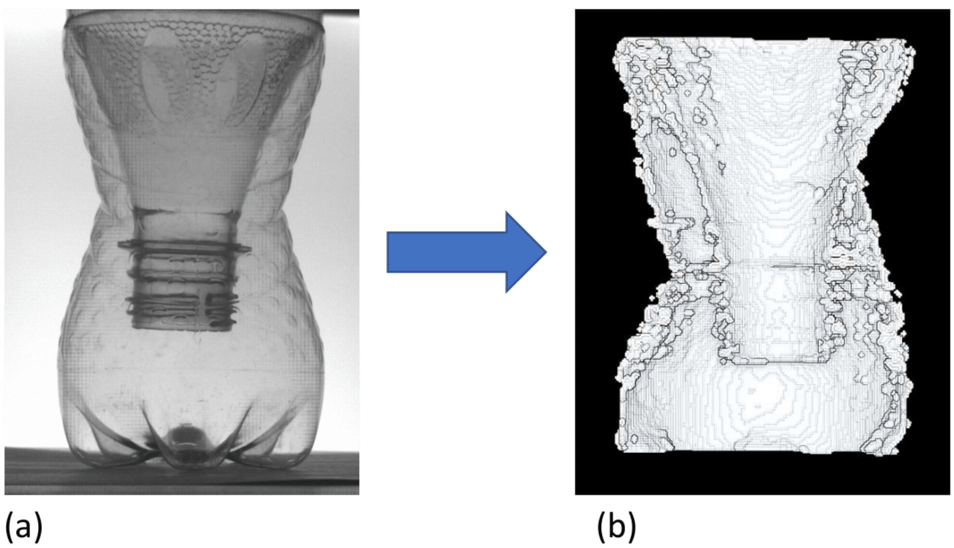
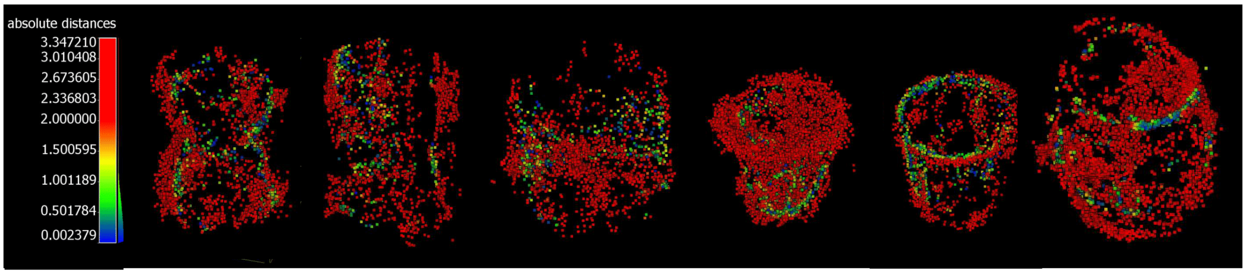


Figure 30. Demonstration of OPT's ability to reconstruct internal surface structures, photo of a mixed bottle structure (a), and 3D reconstruction sliced in half (b), showing the internal structure.

6.3.3. Comparison with Multi-View Stereo (MVS) photogrammetry

MVS requires a surface texture to operate, which is why it is well known to perform very poorly on transparent objects [102]. We attempted to reconstruct the objects in this study using MVS as it is one of the most commonly used camera-based reconstruction techniques today, to contrast its reconstruction quality to that achieved by the OPT technique.

In Table 8 the prohibitive RMS errors involved in reconstructing transparent objects using MVS, when compared to our reference 3D reconstruction performed via structured light can be noticed. In Figure 31, these large errors are visualized by aligning the point clouds to the reference reconstructions, and colouring each point of the MVS reconstructed the point cloud with its minimum distance to the reference point cloud.



(**a**) Fanta bottle (**b**) Zaros bottle (**c**) Selinari bottle (**d**) Liqueur cup 1 (**e**) Liqueur cup 2 (**f**) Wine cup

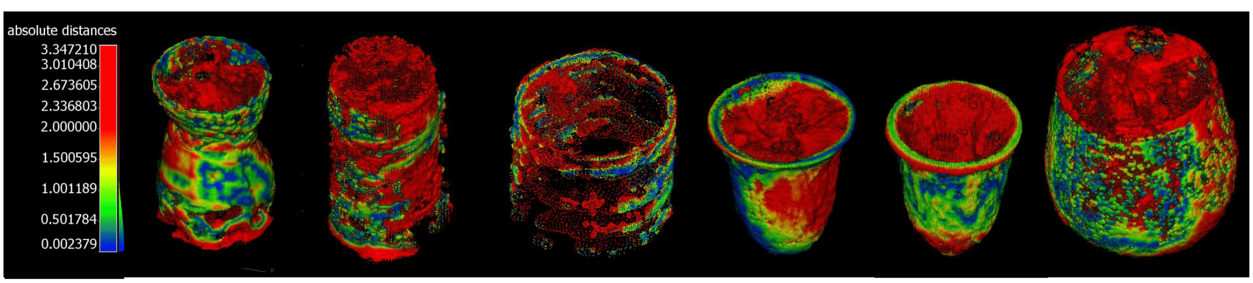
Figure 31. Use of colour texturing on the MVS reconstructed point clouds of the objects in Figure 33, which visually represent the value of the closest distances of each point in the MVS reconstruction from the reference point cloud.

6.3.4. Comparison with Neural Radiance Fields (NeRF)

NeRFs are a relatively new reconstruction technique [118]. It uses artificial intelligence to build a non-linear relationship between the input, which is a single continuous 5D coordinate (the spatial location (x,y,z) and viewing direction (θ,φ)), and the output, which is is the volume density and view-dependent emitted radiance at that spatial location.

It is primarily used for rendering purposes but it can also recover the voxelized 3D shape of the object. Due to the complication of light transport between views the reconstruction of transparent objects is not fully successful. It is however more successful than the multi-view stereo shown in Section 13.3.

In this section, the point clouds created via NeRF with the data acquired by the reference reconstructions in Figure 32 are compared. It can be observed clearly in Figure 32 that NeRF performs better than MVS but worse than OPT. The numerical averages of the errors shown in Table 8 also confirm this observation.



(**a**) Fanta bottle (**b**) Zaros bottle (**c**) Selinari bottle (**d**) Liqueur cup 1 (**e**) Liqueur cup 2 (**f**) Wine cup

Figure 32. Use of colour texturing on the NeRF reconstructed point clouds of the objects in Figure 33, which visually represent the value of the closest distances of each point in the MVS reconstruction from the reference point cloud.

6.3.5 Discussion

A summary of the comparisons performed to our reference 3D reconstructions is found in Table 8 where we numerically compare the average point cloud error achieved by OPT, MVS, and NeRF. It is seen that OPT is much more accurate in reconstructing the external surface of these transparent objects most of the time, with its accuracy being an order of magnitude better than the other two.

If we perform a qualitative study in the 3D reconstruction quality of the same reconstructed object, shown in Figure 33, we notice that OPT retains the most surface details and also has the highest level of reconstruction completeness. Secondly, we notice NeRF with an acceptable level of completeness but without the ability to reconstruct any of the surface details, and lastly, MVS, which has both very poor completeness and reconstruction fidelity.

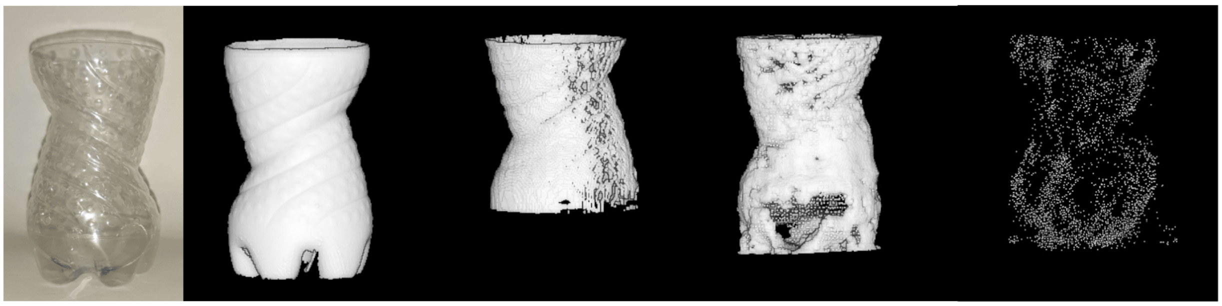


Figure 33. Qualitative comparison of different reconstruction methods results from left to right: (a) Object photograph, (b) structured light reconstruction, (c) OPT reconstruction (ours), (d) NeRF reconstruction, (e) MVS reconstruction.

Compared to other reconstruction methods which have been suggested for the reconstruction of transparent objects mentioned in the introduction, the cost of OPT is minimal, as the only two things required are a field illumination source such as a large LED, a means of rotation, and a black and white camera. The speed of the method has to be divided into acquisition speed and data processing speed, which can be done asynchronously if required. Since the acquisition is performed by cameras, it can potentially be performed on the level of milliseconds. The data processing speed demonstrated here can also be improved many times over by the use of parallel GPUs and with the use of more professional hardware. Regarding the potential for complete automation, it can be completely automated either by adding a robotic arm or via the use of a conveyor belt.

When it comes to the specific class of transparent objects considered in this study, OPT therefore does seem to have the potential to provide near real-time 3D reconstruction at an incredibly low price and with much higher accuracy than any of the techniques that have preceded it so far.

7. Discussion

7.1. Future work

This deliverable proposes methods that advance the state-of-the-art in the reconstruction of surfaces and transparent objects. The proposed approaches are cost-efficient, easily configured, and operated. The proposed methods complement our palette of methods for the reconstruction of tangible heritage.

Nevertheless, the proposed reconstruction methods capture mainly the geometry and secondary the visual appearance of the targeted objects. However, there are more aspects of materiality to capture, such as acoustic, haptic, and other sensory properties. A more advanced digitisation should capture such properties. Some of them are going to be addressed in the next version of this deliverable, in particular, haptic properties, by utilising the 2½D scanner to capture the physical texture of surfaces and, then, haptically render it.

In addition, although the proposed methods capture well the geometry and texture and are, also, robust to reflection properties such as shininess, they do not fully capture the visual appearance of the scanned objects. The reason is that the acquired representations (even when textured) do not capture the interaction of the object with light. This interaction varies according to the material properties of an object and its parts and is different, say, for opaque and transparent objects. The material-specific simulation of light with material bodies of arbitrary properties is studied in D3.1 Craft-specific action simulations. In this respect, in the next version of this deliverable, we will integrate the reported work, with physics-based rendering to visualise the acquired digital assets more realistically.

7.2. Technical conclusions

7.2.1. Surface scanning

The quantitative experiments show that registration errors are in the order of 10p in mosaics comprised of ≈ 4Tp. The proposed approach employs auxiliary images to strengthen image registration cues and fuses proprioceptive data to produce mosaics of the scanned surface with a resolution of 19.8Kppi.

The obtained mosaics were inspected for distortions due to departures from the planarity assumption. We found the limitation of the current configuration to be sharp steps of over 3mm.

We performed no correction as to the global optimization of image intensities. Setting the camera acquisition mode to automatic brightness adjustment adapts the dynamic range of image acquisition to the content of each image. This can be observed in the mosaics of the top layers, where brightness differences across surfaces of the same luminance. Compensation methods tailored for mosaics exist in the literature, e.g. [82, 83]. On the other hand, reflectance calibration even by simple means, i.e. “grey card”, supports the veridical measurement of lightness. It remains to be studied whether images of higher dynamic range are required to capture the brightness variations encountered in all images.

We did not control the stacking process provided by the sensor. By assigning this control to the embedded system accompanying the sensor we may be wasting potential sensitivity to depth variations. Control of bracketing techniques would provide better focus and, thus, more image features. In addition, it can be supported even by weak depth cues, such as depth from defocus [84], or stereo vision.

There are two main factors relevant to scaling the proposed approach for larger scans. The size of the C3Dp bed and memory of the computer that runs the optimization, are as per Section 3.4.1. Larger setups can be achieved using open-source platforms, such as the MPCNC to more precisely control motion over areas up to 2 × 2m2.

We conclude that the resultant device and approach offer a useful imaging modality for several applications, in a cost-efficient manner.

7.2.2. Surface digitisation

A surface reconstruction approach and its implementation are proposed in the form of a surface scanning modality. The proposed approach employs image acquisition at multiple distances and feature tracking to increase reconstruction accuracy. The resultant device and approach offer a generic surface reconstruction modality that is robust to illumination specularities, is useful for several applications, and is cost-efficient. The measurement error of the device ranges between 2.8% to 5% of the size of the scanned object.

This work can be improved to relax the limitation imposed by the restricted depth of focus range (1cm in our case) of the optical sensor. By revisiting the focused stacking method, it is possible to acquire images at several distances and use the depth from the focus visual cue [85] to coarsely approximate the elevation map of the surface. This approximation can be then used to scan the surface in a second pass, guiding the camera elevation appropriately so that the surface occurs within its depth of focus.

Finally, as active illumination supports photogrammetry, we aim to characterise the illumination types that are non-destructive for sensitive surfaces, per type of material. Although active illumination alters colour appearance, it is planned to acquire two images per viewpoint. The first, using structure light to facilitate photogrammetry. The second, from precisely the same viewpoint, would be to accurately capture the colour appearance.

7.2.3. Semi-transparent surfaces

The use of OPT over traditional X-ray CT has many benefits. X-ray CT reconstructions are cumbersome, slow, expensive, and present health risks to the operators. On the other hand, one of the main downsides of OPT is the necessary use of index-matching liquid. This work demonstrates the use of OPT without the need to use index-matching liquid by successfully reconstructing a specific class of large hollow and thin-walled objects.

For glass objects in the context of cultural heritage on the other hand, which on average have thicker sidewalls and also have some areas with thick optical paths, considerable refraction is produced. Areas that suffer more from this were the lower parts of the hollow areas, which have thicker sidewalls than the rim, the glass joint between the stem and the vessel, and the embossed designs on the glass surface. The average shape error of the glass objects measured was higher than that of plastic objects, as expected, at ± 0.92 mm.

The relatively small numerical difference in average dimensional error of ≈0.6 mm between the 'shelled' and 'non-shelled' hollow objects however, does not accurately reflect the rather large qualitative difference of the reconstruction results achieved on the 'non-shelled' hollow objects, as it was noticed that hollow areas were being filled with 'ghost material', embossed features were being distorted, and objects were appearing smaller than their true size. The dimensional errors experienced on 'non-shelled' hollow objects were in part expected as they are due to the non-use of refractive-index matching material. Not using 'refractive index matching liquid' induces a large amount of refraction for hollow objects with thick sidewalls, as predicted analytically in Section 11.

The advantage of using OPT for cultural heritage and industrial applications, where the use of opaque spray coatings is forbidden, is first and foremost that there is no need to use opaque spray coatings to perform the 3D measurement. A second great advantage is the ability to reconstruct internal structures (provided they too are thin-walled), something conventional optical tools cannot do. Additionally, it retains all the characteristics which make conventional optical metrology tools (structured light, laser scanning, and photogrammetry) attractive to industrial and cultural heritage applications, namely: it is cheap and easy to use, safe for human exposure, is camera-based, and it has the potential to perform extremely fast data acquisition, whilst also having a high degree of reproduction fidelity and accuracy.

The disadvantage of using OPT is that a narrow class of objects are possible to reconstruct with high accuracy, namely hollow 'shelled' objects such as plastic bottles. When the sidewalls of the object start to either become much larger or divert too much from being cylindrically symmetric, refraction effects become important enough to distort the shape of the objects considerably and therefore reduce the accuracy of the technique.

When comparing the results with other established 3D reconstruction techniques namely MVS and NeRF, we can see the qualitative difference both in completeness and surface feature detail. A side-by-side comparison of the point clouds created is shown in Figure 33.

In summary, it is demonstrated to the best of our knowledge for the first time, that it is possible to use OPT without the use of refractive index-matching liquid for the reconstruction of objects with sizes larger than 10 mm. The only condition that needs to be met to achieve high accuracy is that the object's sidewall must be thin and consistent enough not to induce significant amounts of refraction. This technique could therefore potentially be used for dimensional quality assurance of hollow 'shelled' objects such as plastic bottles in the beverage packaging industry. On the other hand, for the reconstruction of 'non-shelled' objects such as glass cultural heritage objects, the reconstruction technique is shown to be much more prone to errors due to the higher levels of refraction, which produce severe dimensional and aesthetic distortions to embossed features, as well as artificial object shrinking and the artificial filling of some parts of the hollow object's volume.

ANNEX A. Surface scanning and digitisation apparatus

The proposed approach is implemented using the following materials.

A.1. Off-the-net and off-the-shelf components

The proposed Cartesian 2D scanner (C2Ds) is a device that is attached next to the printing head of a C3Dp. The C3Dp is not otherwise modified, thus the attachment can be removed without affecting its operation. The motion mechanism belongs to the C3Dp. This mechanism moves the printing plate laterally, in two dimensions, and the printing head only vertically. The C3Dp is commanded to reach the imaging locations, by a micro-controller.

The C2Ds were built on top of an adaptation of the Prusa i3 series C3Dp, which were chosen due to their wide adoption, low cost, and ease of construction. The operating volume is $24.89 \times 21.08 \times 6.86 cm^3$. The selected parts for the C3Dp are cited in Appendix B.

The visual sensor was an Olympus Tough TG-5, which has a minimum focus distance of 1cm, 4000 × 3000p resolution, and a FoV of $16^{\circ} \times 12^{\circ}$.

The motor is controlled by the Marlin open-source firmware. Marlin is widely used and runs on the cost-efficient 8-bit Atmel AVR microcontrollers. The reference platform for Marlin is the Arduino Mega 2560 with RAMPS 1.4, which is directly compatible with the equipment used for implementing the printer. This firmware runs on the motherboard and manages real-time controls for heaters, steppers, sensors, lights, LCDs, buttons, etc. The control language is a derivative of G-code. G-code commands issue simple instructions, such as "set heater 1 to 180", or "move to XY at speed F".

The power supplies shipped with 3D printers usually generate up to 350W on 12V output. In our implementation, a more robust solution was preferred to accommodate the power requirements of the visual sensor. To this end, a 650W ATX power supply was used. The electronics and the sensor are connected to the 5V output while the stepper motors are connected to the 12V output.

A.2 Motor and motion

The implementation of a C3Dp contains some free variables, such as the quality of materials, the torque of motors, etc. To reduce the effect of vibrations and increase motion accuracy, the device was implemented as follows.

Aluminium frames of 40 × 40mm thickness were used for the truss. The backside of the print bed was enhanced with an aluminium frame, to increase its weight. High-quality, heat-hardened steel rods of 12mm thickness and high-quality linear bearings were used for the motorized part of the printing bed. Printable components of the apparatus were printed using PET-G and a 60% infill rate to enhance their stiffness and reduce the possibility of heating deformation, due to intensive use.

To provide enough torque for this implementation, motors were standard Nema 17-sized high-torque stepper motors. The motor's motion was transmitted via 6mm non-elastic timing belts, integrated with steel threads for enhanced stiffness. Motors are driven by the Texas Instruments DRV8825 Stepper Motor Controller ICs. The controller supports up to 1⁄32 micro-stepping. The device is operated through a microcontroller built on top of the Arduino Mega 2560. For the wiring of the C2Ds, the RAMPS 1.4 Arduino Mega Pololu Shield was used.

On account of the achieved mechanical robustness, the printing bed was increased by a factor of 4.764 from its specification to 50 × 60cm^2. The bed was coated with a, 5mm thick, aluminium sheet, to ensure a flat slide for the placement of samples.

A.3 Imaging

The camera faces the imaged surface perpendicularly. To mount the camera a sensor base was designed using the TinkerCad software. The design, shown in Figure 34 (left), was exported in STL format and printed on the C3Dp; see Figure 34 (right). The design of the mount is compatible with the print head and is placed on its backside, thus allowing both heads to be mounted concurrently.

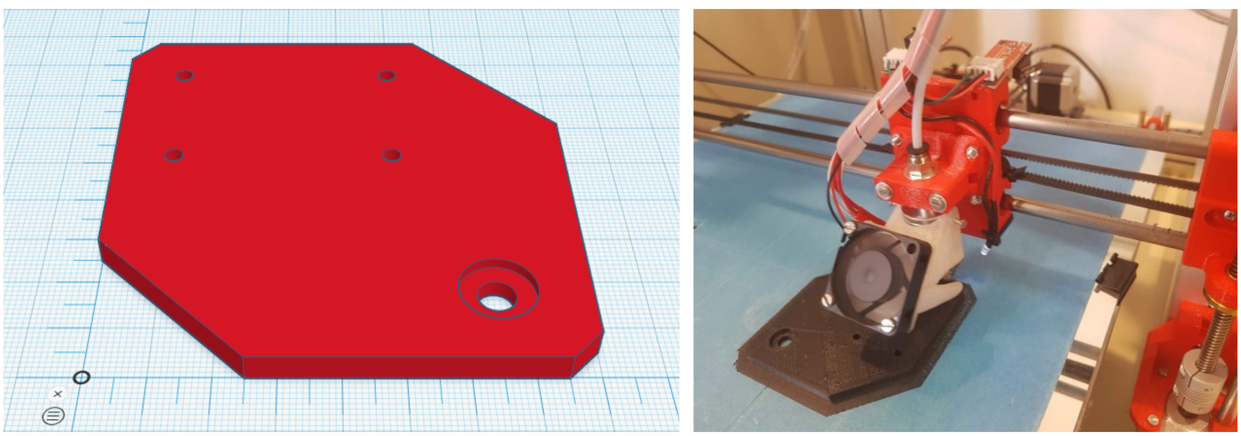


Figure 34. Design of the camera base (left) and photograph from its printing on the C3Dp (right).

The time required to capture the required number of images exceeds the duration of typical consumer-grade batteries, i.e. ≈ 1300mAh. To avoid interrupting the scan for changing and the consequent sensor displacements, power was continuously provided as follows. A printed case emulating the battery was wired to a power supply of the appropriate voltage and current. The second component in Figure 35 guides and stabilizes wirings. The models can be found in the supplementary material.

Figure 35. Designs of a battery emulator and photographs of its implementation.

A.4 Communication and control

The scanner control software runs on a personal computer wirelessly connected to the controller. This computer runs software mediating the image acquisition process. The software executes a scanning plan, containing the locations of image acquisition, encoded as 3D coordinates.

Specifically, the print bed and sensor are drawn to designated relative position, due to a G-Code command to the controller, e.g. “G1 X2.57 Y1.93 Z18.20”. The signal for image acquisition is then sent.

The command sequence in Table 9 (top) is sent to trigger the acquisition of an image. Image acquisition failures and delays are treated as follows. The software checks if the image has been indeed acquired, using the sequence in Table 9 (bottom). This sequence acquires the list of stored images, to be compared with a previously collected one. These two sequences are repeated until the image is acquired. Each filename is stored to later conveniently rename the acquired images.

When all the pictures are acquired, the files are manually copied from the memory card of the sensor and automatically renamed by the software. The renaming includes the coordinates of image acquisition in the filename, in the form of Z-Y-X.jpg, such as ‘018.20-001.93-002.57.jpg’. These 3D coordinates are only readings of the C3Dp controllers, they are not regarded as absolute measurements but fused with visual cues, in Section 3.4.1.

Table 9. Sensor communication command sequences.

Image acquisition
http://192.168.0.10/exec_shutter.cgi?com=1st2ndpush http://192.168.0.10/exec_shutter.cgi?com=2nd1strelease
File listing
http//192.168.0.10/get_imglist.cgi?DIR=/DCIM/100OLYMP

A.5 Cost

The cost of materials is reported in Table 10. On the date of submission, the total cost was 952 USD.

Table 10. Costs of utilised materials.

Component	Quantity	Price (USD)
Visual sensor, Olympus Tough TG-5	1	430
Microcontroller (Arduino Mega 2560)	1	42
Controller RAMPS 1.4	1	9
DRV8825 Stepper Motor Driver	5	20
Stepper motors	5	100
Extruder	1	10
30mm × 30mm aluminum truss	3m	120
Metal rods 12mm	2 pieces, 12mm × 100cm	24
Metal rods 10mm	4cm × 60cm	28

Aluminum sheet 5mm	2 pieces, 60 × 40cm	50
Lead Screw T8 540mm	2	20
Nut for Lead Screw T8 Lead 8mm	2	4
Timing Belt XL 44”	1	10
Aluminum GT2 Timing Pulley	2	4
Aluminum Flex Shaft Coupler 5 − 8mm	2	4
Aluminum GT2 Timing Pulley Idler	1	2
ATX PSU 650W	1	60
Filament	1⁄2kg	15
Total		952

ANNEX B. Printed components

Table 11. Printed components.

Name	Source
Z-axis leadscrew	https://www.thingiverse.com/thing:519391
Controller case	https://www.thingiverse.com/thing:2047732
Bowden extruder	https://www.thingiverse.com/thing:2243325
x-carriage	https://www.thingiverse.com/thing:2514659
z-axis	https://www.thingiverse.com/thing:1692666
y-axis belt holder	https://www.thingiverse.com/thing:1030200
y-belt tensioner	https://www.thingiverse.com/thing:3404464
y-axis motor holder	https://www.thingiverse.com/thing:2808408

ANNEX C. Market survey

All prices are approximate and estimated on the day of submission. Unreported prices require asking for a quote and all exhibit a larger price than the others in the same table.

Table 12. Flatbed A0 scanners.

Name	Optical resolution (ppi)	Price (USD)
Kurabo K-IS-A0FW	1000	50K
Microtek LS-4600	600	60K

Table 13. Film scanners.

Name	Optical resolution (ppi)	Price (USD)
Plustek OpticFilm 8100	7200	400
Epson Perfection V550 Photo	12800	600
UScan+ HD LTE	2400	Quote

Table 14. Large format scanners.

Name	Optical resolution (ppi)	Price (USD)
Colortrac SmartLF Sci	1200	5-12K
Colortrac SmartLF SC 42 Xpress	1200	7K
Contex IQ Quattro X - 44”	1200	7K
Contex IQ Quattro 4450 / 4490	1200	7-8K
Image Access WideTEK 48CL	1200	6K
ROWE 850i - 55”	1200	22K
Image Access WideTEK 60CL	1200	12K
CRUSE ST Light 300	300	30K
CRUSE ST Light 600	600	30K
CRUSE Synchron Table (ST)	830	30K
CRUSE CS 85/145 ST-T	600	30K
CRUSE CS 82 ST-T 2450	600	30K

Table 15. Book scanners.

Name	Optical resolution (ppi)	Price (USD)
Suprascan double A0	600	Quote
Suprascan Quartz A0 LED HD	600	Quote
Sma Scanmaster 0 3650	600	Quote
book2net Hornet	400	Quote
Czur ET16	275	355
Fujitsu Scansnap SV600	280	500
SMA ScanMaster 2	1200	10K
SMA RoboScan 2	600	10K
IID Bookeye 5 V3	600	10K
Bookeye 4 V3 Kiosk	600	10K
Bookeye 4 V2 Semiautomatic	600	10K
Bookeye 4 V2 Professional Archive	600	10K
Zeutschel OS Q1	600	10K
Zeutschel OS HQ	1000	10K
Zeutschel OS Q0	600	10K

Table 16. Material scanners.

Name	Optical resolution (ppi)	Price (USD)
Vizoo xTex	1000	Quote
xrite TAC7	385	Quote

Table 17. Scanning services.

Name	Optical resolution (ppi)	Price (USD per sample)
Materialcapture	600	100 – 300
Muravision	920	100 – 300
Overnight scanning	600	100 – 300

ANNEX D. OCT Appendix

Optical Projection Tomography (OPT) [110, 119], as the name suggests, is a tomographic method, whose principle of operation is similar to that of X-ray computed tomography (CT). It reconstructs objects in 3D by measuring light absorption through an object from different points of view, with the difference being that X-ray CT uses the X-ray spectrum of light to perform the reconstruction and OPT uses the visible spectrum of light.

For visible light to replicate the functionality of X-ray light in X-ray CT, two principles need to be adhered to: the first is that light must be able to pass through the object and be detected on the other end (i.e., not absorbed completely so that the absorption is measurable), and the second is that the light that passes through the object is not 'bent' or refracted as it passes through the object [110].

The two conditions above are normally achieved in OPT for biological imaging where the measured object is very small in size (1–10 mm in thickness) so that visible light can shine through it [110], and by use of refractive index-matching liquid so that its refractive index (RI) matches that of the environment. No refraction occurs as light passes through [120]. This refractive index matching process is called 'optical or tissue clearing' because it has the effect of making the biological tissue less refractive to visible light and easier to photograph clearly.

For light rays in the visible spectrum, glass material used for creating drinking cups found in cultural heritage collections nominally has an RI between ≈1.3 and 1.7 [121]. Polyethylene terephthalate (PET) on the other hand, a very common plastic material, has a nominal RI of ≈1.58 [121]. Both materials therefore have an RI that is very different to that of the surrounding medium of air (RI ≈ 1) and therefore we would normally expect considerable refraction when visible light rays propagate through these materials.

However, for very thin-walled, cylindrically-symmetric objects, such as plastic bottles and wine glasses, two effects come into play, which minimize refraction: the first is that the wall thickness is so thin that the refraction effect is expected to be insignificant (Section D.1), and the second is the fact that the objects are mostly cylindrical, which means that parallel light rays will experience two, almost opposite, beam shifts as they pass through, 'self-correcting' their path (Section D.2). Therefore, for these types of objects, it is valid to assume that OPT can be achieved directly in air, without the use of index-matching liquid. We quantify these phenomena via simulation in the following sections.

D1	Measuring Visible Light Ray Beam Shift for Small Sidewall Thicknesses
It is assumed in this work that the refraction occurring at any part of a transparent object can be approximated to that of a flat thin slab of transparent material. This is a valid approximation because the sidewalls are very thin and the surface texture contours are much larger (≈1 mm–10 cm) than the scale of the wavelength of light used (400–600 nm). Therefore, at the scale of the photon, every light ray striking object will essentially experience the surface as a thin flat slab of material. To measure the parallel beam-shift effect of a light ray through a thin slab (Figure 36), we use Snell's law of refraction [89] (Equation (4)) both on the incoming and also on the outgoing surface of each sidewall. $n1\sin(\theta 1) = n2\sin(\theta 2)$ (4)	

Using the flat slab approximation and Snell's law it is possible to calculate the amount of parallel beam-shift for various angles of incidence and slab thickness. The calculations were performed for the glass and plastic materials used in the objects we have selected to measure, with an RI of 1.51 for BK7 glass (Figure 37) and 1.58 for PET (Figure 38). As can be seen in Figure 37 and Figure 38, the beam shift increases with the angle of incidence and reaches a maximum that is almost equal to the slab's thickness at an angle of incidence of 89∘ from the surface normal.

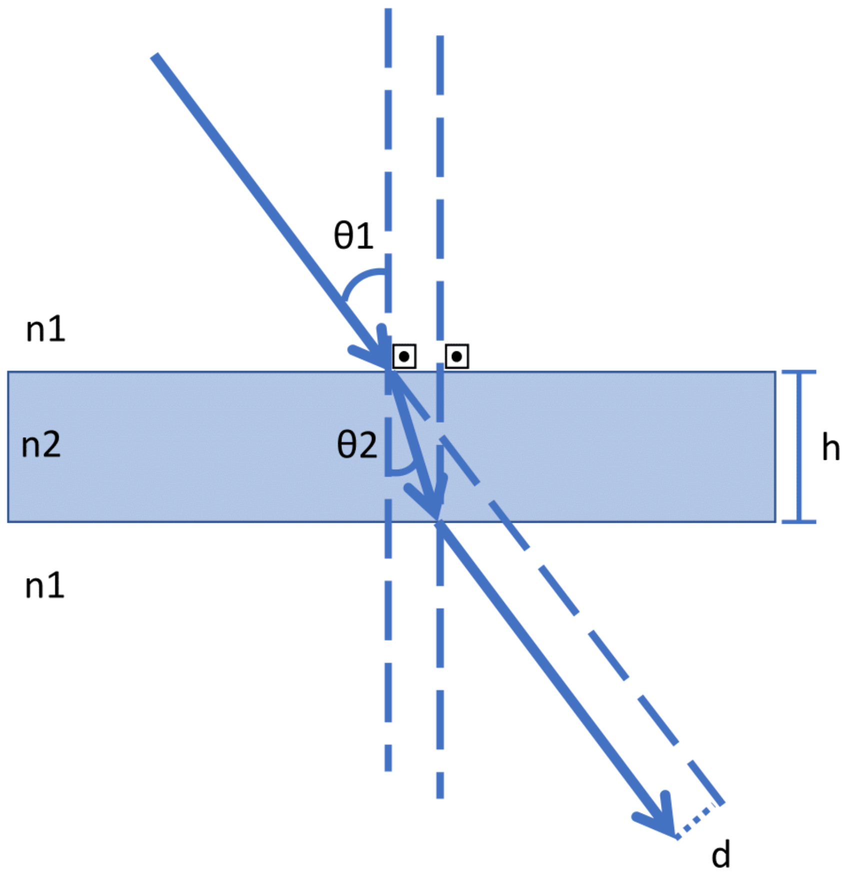


Figure 36. Effect of parallel beam shift predicted by Snell's Law of refraction, when a beam is incident on a slab of material of RI n2 in a medium of RI n1.

The systematic analysis which was carried out therefore shows that, for the glass and plastic materials that were selected, the beam shift will be smaller than that of the sidewall thickness at any angle of incidence and therefore at any point on the object. For the hollow objects selected, the sidewalls (0.15–2 mm) are small compared to their diameter (52–80 mm), so refraction is not expected to affect the reconstructed shape considerably. Especially so, when combined with the effect described in Section D.2.

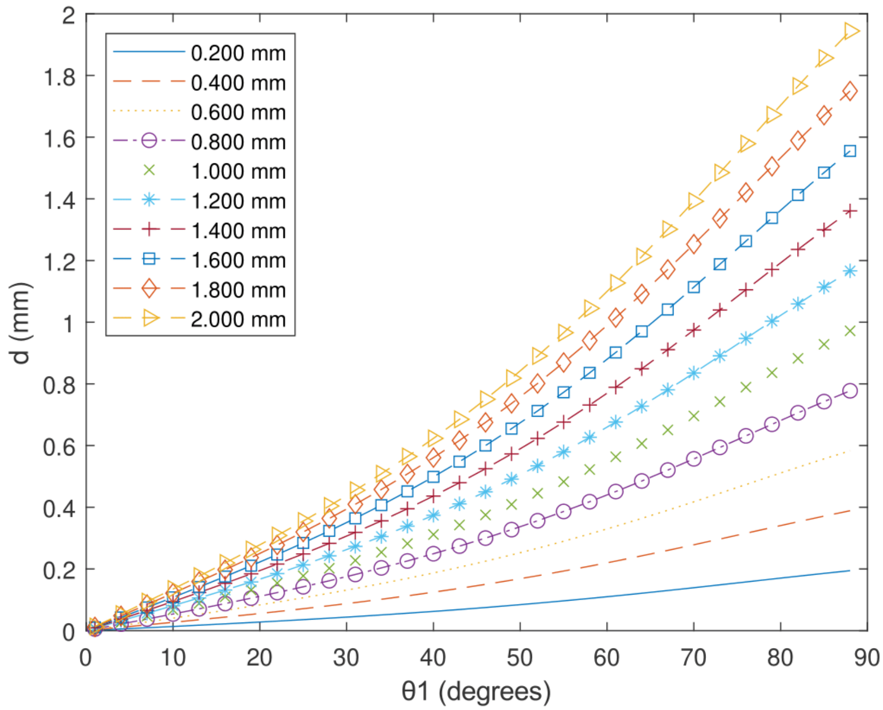


Figure 37. Beam shift distance d as the angle of incidence $\theta 1$ increases between 1and 89ᵒ as calculated by Snell's law, for PET of RI 1.58 and different thicknesses h between 0.2 and 2 mm.

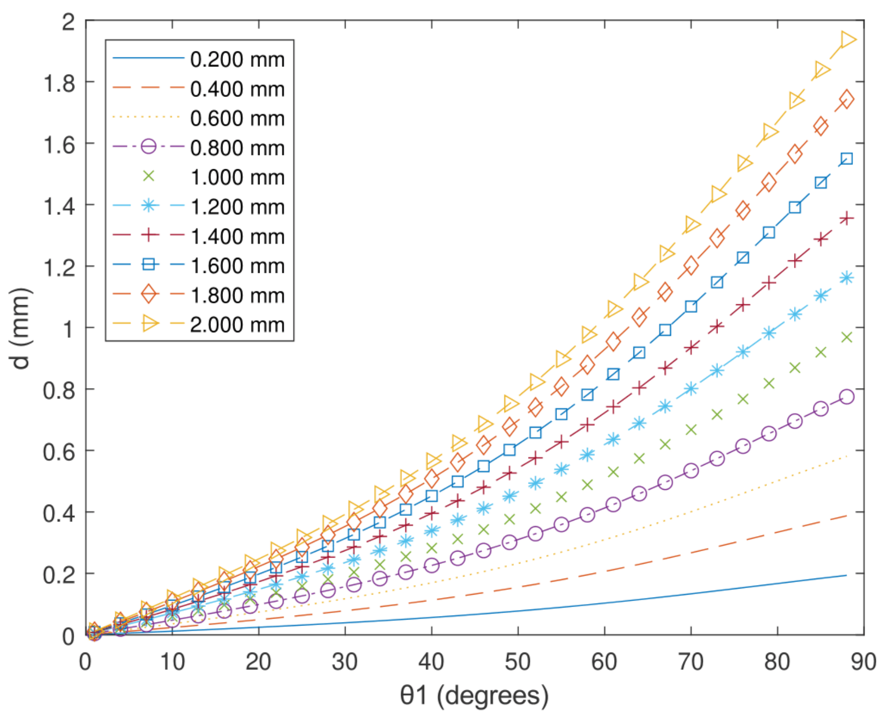


Figure 38. Beam shift distance d as the angle of incidence $\theta 1$ increases between 1 and 89ᵒ as calculated by Snell's law, for BK7 glass of RI of 1.51 and different thicknesses h between 0.2 and 2 mm.

D2	**Light Rays Propagating through Hollow Circular Objects**

The second effect that makes hollow cylindrical objects especially measurable using OPT, is that parallel light rays experience two opposing beam shifts as they propagate through these objects: a first parallel beam shifts towards the centre of the object as the ray enters the hollow transparent object, and a second beam shift away from the centre of the object as the ray exits the object. This effect is simulated for multiple parallel beams entering a hollow thin-walled circular disk made of PET with an RI of 1.58 in Figure 39.

When the camera is placed sufficiently far away, the rays that reach the lens are those which are mostly initiated in parallel. This can be seen in the performed ray tracing simulations (Figure 39), where it can be seen that parallel beams become slightly convergent after passing through the object. This is advantageous as these rays can be collected by a single camera placed ‘far away’ from the object, without the need to ‘stitch an image’ such as in other large-scale OPT approaches [122]. In our experiment, we placed our camera 70 cm from the object which, compared to the object diameters of 5–8 cm is at least a ≈9:1 distance-to-size ratio. The field light source selected was a white light LED panel and it was also placed as far back as possible whilst concurrently being able to illuminate the whole object (in our case 20 cm away) So the total distance between the light source and camera was 90 cm.

In non-hollow cylindrically-symmetric objects filled with PET, on the other hand, the light beams pass through a lot more optical material, which causes larger amounts of refraction. The overall beam paths are therefore changed by a far greater amount, resulting in an intense lensing effect that does not enable easy reconstruction using OPT (Figure 39).

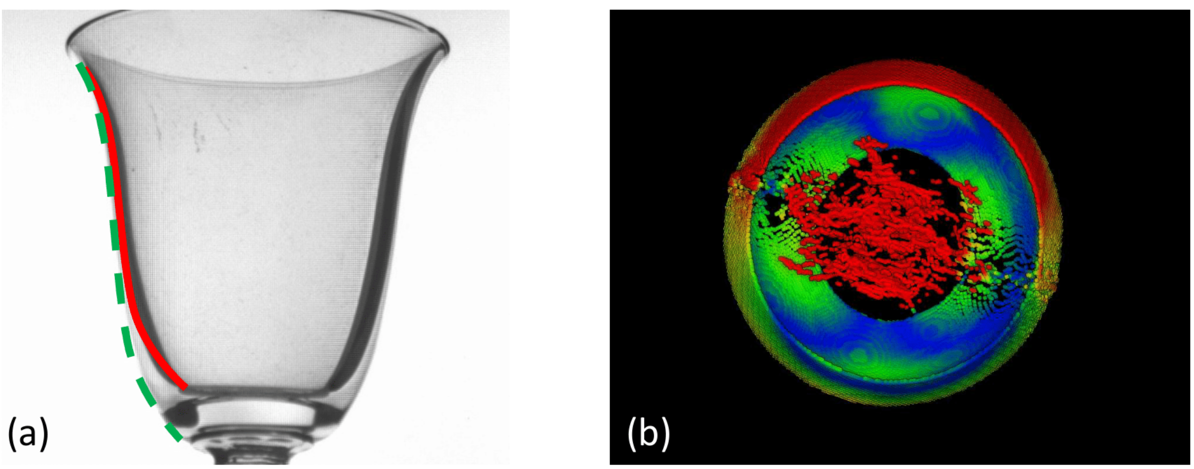


Figure 39. Refraction of visible light rays from a solid disk object (a) and a hollow disk object with a thin size wall (b) both with an RI of 1.58. For a thin-walled hollow object, the rays experience less refraction. Created using [123].

ANNEX E. Surface scanning results

E.1 Qualitative

The goal of the experiments was to assess mosaic registration accuracy and to explore tolerance to departures from the assumptions of Lambertian reflectance and surface planarity. Indicative samples were drawn from applications relevant to sensitive materials, found in art, biology, document analysis, and textiles. The selected materials exhibit variability as to their reflectance properties and their 3D surface texture. We included shiny and rough materials, in the samples. We did not include highly transparent, highly specular, or highly reflective materials.

Macroscopic images of the samples, acquired by a conventional camera, are shown in Figure 40. In the figure, from left to right and top to bottom, samples 1 − 5 are paintings; sample 6 a blank piece of cotton canvas; sample 7 is a scarcely handwritten A4 page; sample 8 is a stamped and signed passport; samples 9 and 10 are blank and printed graph paper, respectively; samples 11-13 are pieces of silk fabric; sample 14 is a leaf; samples 15 and 16 are fine and coarse-grained sand, respectively; sample 17 is an assortment of coins; sample 18 is a banknote. Samples 13 and 18 were scanned entirely. For the rest, a $5 \times 5 cm^2$ region was scanned. In Figure 41, 2048 × 2048p regions from images of the finest layer are shown, in the same order as in Figure 40.

Figure 40. Samples.

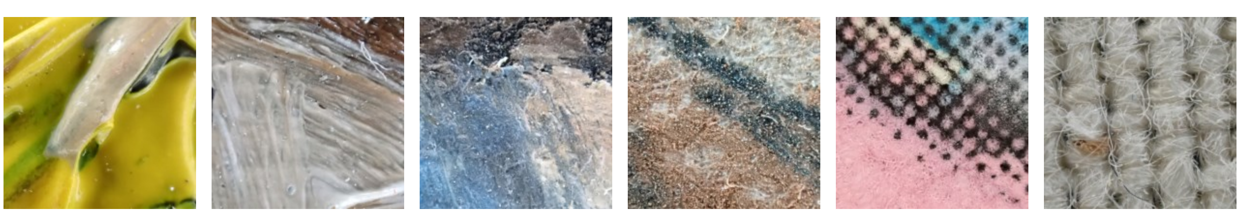
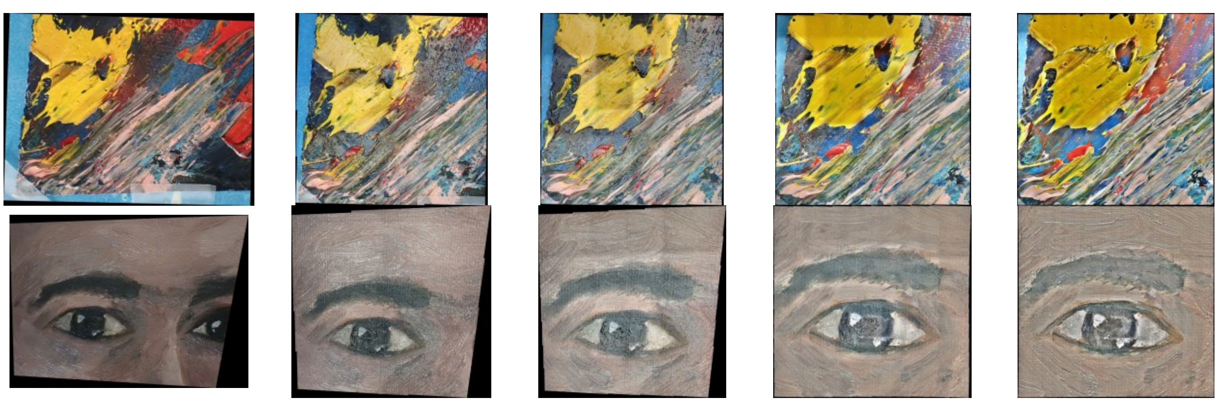


Figure 41. Sample details.

Our primary investigation regarded the translational component of the estimated homographies. The estimated homographies lead to image shifts that are always less than 4p. In other words, none of the estimated homographies suggests an update of Ci that would cause an image shift not larger than 4p. In turn, this suggests that no homography estimate is in gross contradiction with the proprioceptive readings. Second, we observe that the method is robust to the occurrences of missing information that were encountered in the experiments. The information missing was either a single pyramid apex in the use of multiple pyramids or registration failures due to a lack of reliable point correspondences. In all cases, a complete mosaic is provided for the layer of the highest detail.

E.1.1 Paintings

To study paintings, we acquired painted samples on canvas, Canson paper, and regular printing paper. The samples exhibited various degrees of surface roughness. The colours were made from soft pastel or oil. The results are shown in Figure 42. To better investigate the effect of height variations, in the example of the top row, the impasto painting technique [69, p. 100] was utilised. This technique involves painting in overlapping layers and gives rise to surface anomalies. The average height step of these anomalies was 0.75mm. The example is centred upon a 1.5mm protrusion. In the remaining rows of Figure 42, the examples are sorted in the declining level of surface roughness.



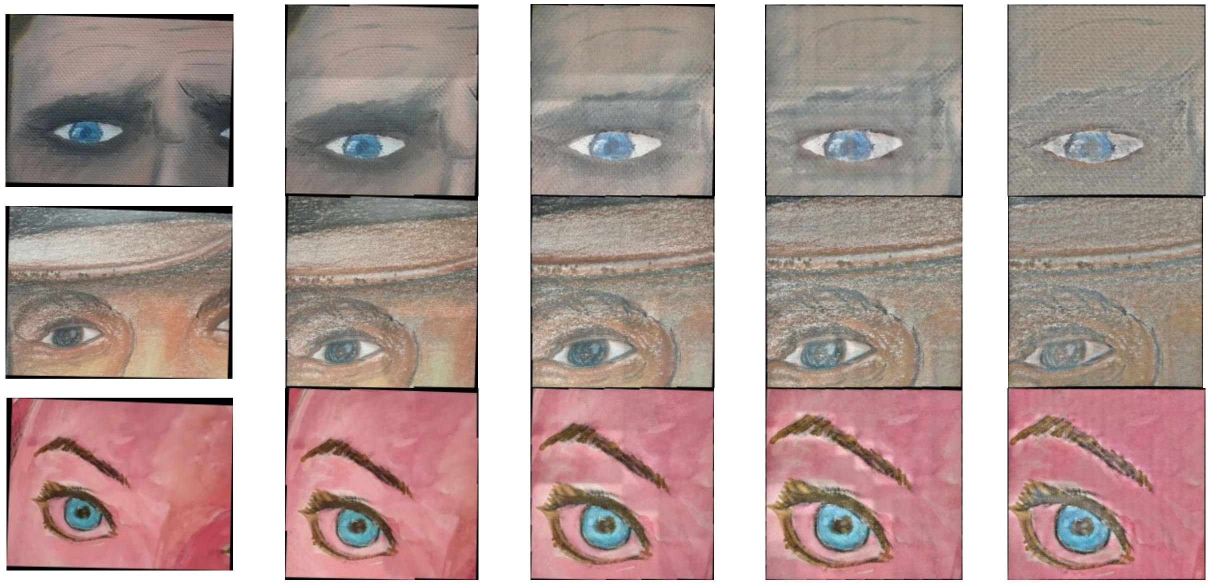


Figure 42. Paintings.

E.1.2 Paper and canvas

A white sheet of paper and a white piece of canvas tested the application of the method in documents and plain fabrics. Canvas is a textile with repeated structure, but only macroscopically, as at the imaging resolution cotton fibres provide unique textures. Cotton plies are in the range of 12 − 20μm. The results are shown in Figure 43. Under investigation were potential effects due to blank surface space. No such effects were observed, as paper when closely inspected reveals rich texture. The same was the case for cotton canvas used for painting. Still, in the third layer from the top, a homography estimate was discarded.

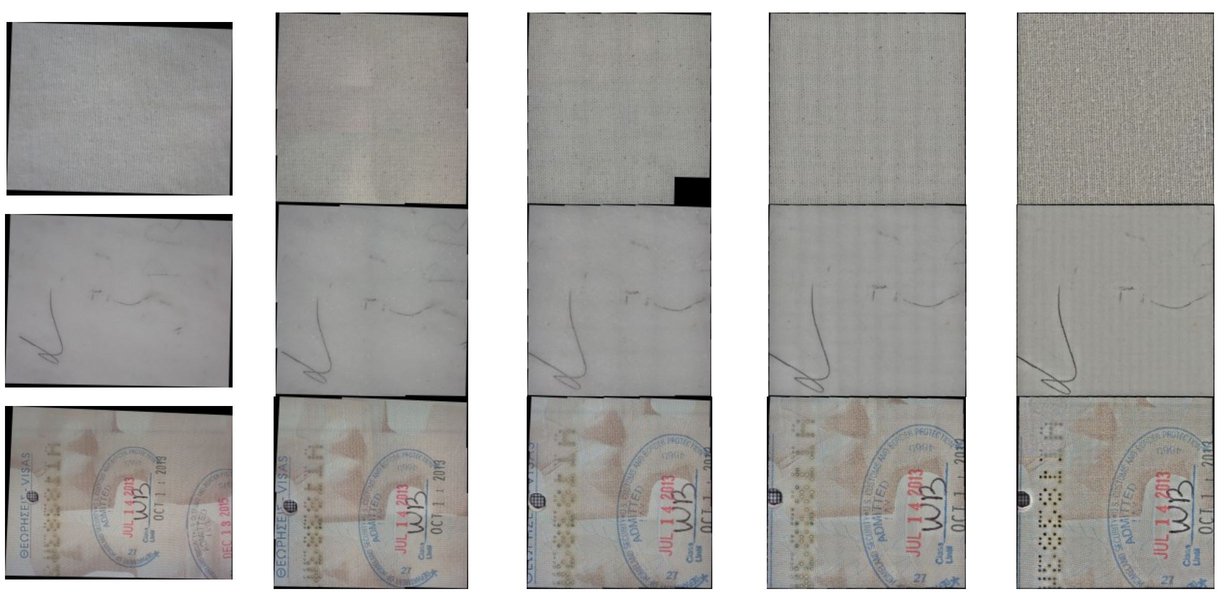


Figure 43. Canvas, handwritten paper, and passport with stamp and handwriting.

E.1.3 Repetitive patterns

To test the sensitivity of feature-based image registration to repetitive patterns, conventional graph paper was used. The results are shown in Figure 44. As in the case of the canvas, in the fine mosaic layers, sufficient uniqueness cues are found to abstain from gross misregistration errors. Nevertheless, a failure is observed in the auto-focus function of the sensor. At the second layer from the top, some images were out of focus, possibly due to the sensitivity of the auto-focus mechanism of the sensor to repeated texture. The homography estimate was considered unreliable homographies and discarded. We repeated the experiment this time using printed text, using the “Liberation Serif - Regular” font, at 6pt. In this condition, said effect did not occur.

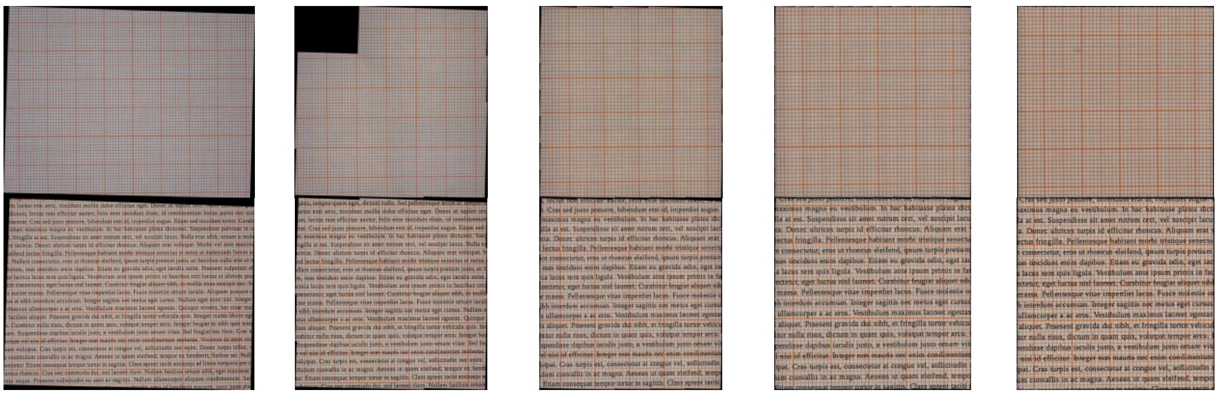


Figure 44. Repetitive patterns.

E.1.4 Fabrics

Though the study of fabrics is related to heritage [70] and industrial applications, ways to digitize textiles and fabrics are constrained in the products reviewed in Appendix C. These approaches do not scan the fabric in sufficient resolution to reveal the fine crafting of some textiles. We chose patterned silk fabrics handwoven on a Jacquard loom because this type of weaving allows for intricate patterns on the fabric. We chose silk as the most challenging material, because its fibrils are sleek, reflecting light from many angles, attributing it to its characteristic sheen. Also, silk is one of the finest plies. Silk fibres from the Bombyx Mori, as in the example, are in the range of 5 − 10μm (a human hair is ≈ 50μm). We scanned two samples woven with the same two-colour pattern but with alternating colours. The results are shown in Figure 45.

Figure 45. Fabrics.

E.1.5 Fiona

To image biological tissue, a leaf was scanned. The result is shown in Figure 46.

Figure 46. Fiona.

E.1.6 Approximately planar surfaces

To observe the effects of surface roughness, we scanned two types of sand and an assortment of coins. The results are shown in Figure 47, in increasing order of surface roughness. The coins ranged from 1.67 – 2.33mm in thickness. For the coarse-grained sand, grains height steps between adjacent grains exceeded well 3mm.

Figure 47. Approximately planar surfaces.

We did not detect artefacts from fine-grained sand or coins. Though not easily found, the case of coarse-grained sand exhibits some tractable mismatches, as the image combination method cannot compensate for the lack of accurate registration between the overlapping image regions. They are shown in Figure 48, where each image shows a region of ≈ 1cm^2.

Figure 48. Registration failures examples.

E.2 Quantitative

Ideally, when viewed at the same resolution, mosaic layers that image the same surface region should be identical. The similarity between mosaic layers was quantified by cross-correlation, in the domain of [-1, 1]. We computed this metric between consecutive layers, as well as between the top and bottom layers. In Table 18, we report the correlation values. The first row shows the correlation coefficient between the coarsest and the finest layer. The remainder columns show the correlation values for consecutive layers.

Table 18. Correlation coefficients between mosaic layers.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
0-4	.5	.3	.1	.2	.6	.6	.5	.6	.9	.9	.2	.5	.6	.3	.7	.8	.3	.5
0-1	.9	.8	.9	.9	.9	.8	.7	.8	.5	1.0	.9	.8	.9	.9	.9	.9	.5	.8
1-2	.8	.7	.7	.9	.9	.4	.6	.9	.5	1.0	.9	.9	.9	.9	.9	.9	.6	.9
2-3	.6	.3	.6	.8	.9	.5	.6	.9	1.0	1.0	.8	.8	.9	.8	.9	.9	.6	.9
3-4	.7	.5	.4	.4	.8	.7	.7	.7	.9	1.0	.2	.7	.8	.7	.8	.8	.7	.9

To measure systematic distortions we used the images of coins (from Section 5.1.6), which are circular structures. We performed Canny edge detection [79] in the finest mosaic layer and selected the edges corresponding to the circular creases of the 2c and a 5c coin. The selected edges were used to fit circles, using least squares without a robust selection of inliers. In Figure 49, the selected edges are shown on the left pair of images. In Table 19, we report deviations of the detected edges from the fitted circle.

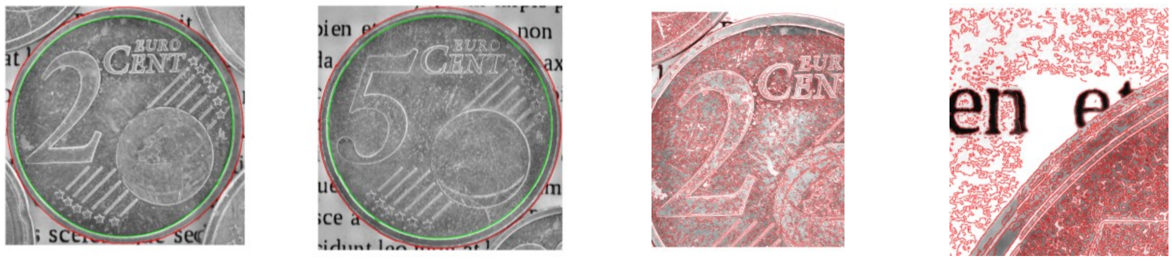


Figure 49. Edges belonging to the inner and outer creases of two coins (left) and edge detection details (right).

Table 19. Mean circle fit error and standard deviation.

Circle	Radius (p)	Error (p)
2c outer	3464.8	3.05 (2.62)
2c inner	3216.8	4.09 (3.50)
5c outer	3936.4	4.86 (3.17)
5c inner	3630.8	5.20 (3.36)

This example facilitates observations at the locations of lateral image overlap, where seams are typically observed. In general, due to accurate registration and the effect of the method in [54], seams are usually easily observable in fine scales. In the experiment, the focus distance was automatically adjusted. As the coins are elevated from the background, near the coin boundary the sensor focuses either on the paper background or the coin. When the focus is placed on the background, the image region where the coin appears is blurry and, instead, the background is focused. Though the structure distortion is minute, the difference in the focus of the blended images is observable when image edges are detected as in the right pair of images in Figure 49. In turn, the different amounts of image blur at the boundaries of the blended image give rise to edge dislocations. In Figure 50, a mosaic of 2 × 2 images is shown for each sample in the experiments (minified for document scale). In the supplementary material, these images can be found in their true resolution. We observe that although no highly frequent seams are observed, a global brightness difference is observed between stitched images.

Figure 50. Mosaics of 2 × 2 original images.

ANNEX F. Calibration of surface scanning and digitisation apparatus

F.1 Camera

The proposed approach requires estimates of intrinsic and extrinsic camera parameters. Intrinsic parameters regard the resolution and optical properties of the visual sensor. Extrinsic parameters regard its position and orientation in space and represent the estimation of camera motion relative to the target surface.

The camera is calibrated to estimate its intrinsic parameters and its lens distortion, using the methods in [46, 47]. The intrinsic parameters represent the location of the optical centre, the skew of the optical axis, and the focal length of the camera. The FoV of the camera is also derived from these parameters. In addition, lens distortion is also estimated in this step. The aforementioned parameters are independent of the camera location and, thus, are estimated once before mounting the camera.

Extrinsic parameters represent the orientation and the location of the camera, as a rotation and translation of the camera concerning some world coordinate system. An initial estimate of extrinsic camera parameters is provided by the CNC device and specifically from the readings of the stepper motor controllers that move the camera. The area of the scan table and the elevation range are measured to estimate the motor step length per dimension, given that these motors produce equal motion steps. The number of horizontal and vertical steps are denoted as n_x and n_y, while the length of the horizontal and vertical steps s_x and s_y, respectively. Step s_x and s_y, are defined in metric units. Thereby the obtained initial estimate of the extrinsic parameters is also in metric units. This initial estimate of camera locations is refined at a later stage of the method.

F.2 CNC

CNC calibration is the process of verifying and adjusting the accuracy and precision of the 3D CNC machine to ensure it moves the camera according to specifications. This process involves ensuring the machine can move to the exact coordinates specified in the design, verifying that the machine can repeatedly return to the same position over multiple cycles, checking and correcting the alignment of the spindle and the table, and adjusting for mechanical inaccuracies, such as backlash or lead screw errors. CNC calibration often involves the use of precision measuring instruments to measure deviations and make necessary adjustments.

F.3 Implementation

The benefit of calibrating the camera concerning the CNC measurements is that each acquired image is associated with a coarse estimate of its extrinsic parameters. In terms of delivering a “plug and play” device, these calibration steps are done once at manufacturing (or installation) time. As a device, the sensor receives a “scan plan” with image locations and provides a dataset with the image acquired at those locations and appropriately indexed by location.

References

1. B. Schmitt, C. Zirbes, C. Bonin, D. Lohmann, D. Lencina, and A. Netto. A comparative study of Cartesian and Delta 3D printers on producing PLA parts. Materials Research, 20:883–886, 2017.
2. M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. International Journal of Computer Vision, 74(1):59–73, 2007.
3. Z. Zhu, J. Fu, J. Yang, and X. Zhang. Panoramic image stitching for arbitrarily shaped tunnel lining inspection. Computer Aided Civil Infrastructure Engineering, 31(12):936–953, 2016.
4. R. Carroll and S. Seitz. Rectified surface mosaics. International Journal of Computer Vision, 85(3):307–315, 2009.
5. C. Hernandez-Matas, X. Zabulis, A. Triantafyllou, P. Anyfanti, and A. Argyros. Retinal image registration under the assumption of a spherical eye. Computerized Medical Imaging and Graphics, 55:95–105, 2017. Special Issue on Ophthalmic Medical Image Analysis.
6. V. Eremeev, A. Kuznetcov, G. Myatov, Oleg Presnyakov, V. Poshekhonov, and P. Svetelkin. Image structure restoration from Sputnik with multi-matrix scanners. In L. Bruzzone, editor, Image and Signal Processing for Remote Sensing, volume 9244 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, page 92440F, 2014.
7. M. Pilu and F. Isgro. A fast and reliable planar registration method with applications to document stitching. In Proceedings of the British Machine Vision Conference, pages 67.1–67.10. BMVA Press, 2002.
8. C. Blasse, S. Saalfeld, R. Etournay, A. Sagner, S. Eaton, and E. Myers. Premosa: extracting 2d surfaces from 3d microscopy mosaics. Bioinformatics, 33(16):2563—2569, 2017.
9. T. Levanic. ATRICS - a new system for image acquisition in dendrochronology. Tree-Ring Research, 63(2):117–122, 2009.
10. D. Conover, J. Delaney, and M. Loew. Automatic registration and mosaicking of technical images of old master paintings. Applied Physics A, 119:1567–1575, 2015.
11. Z. Xiong and Y. Zhang. A critical review of image registration methods. International Journal of Image and Data Fusion, 1(2):137–158, 2010.
12. Barbara Zitova and Jan Flusser. Image registration methods: a survey. Image and Vision Computing, 21:977–1000, 2003.
13. R. Szeliski. Computer Vision - Algorithms and Applications. Texts in Computer Science. Springer, 2011.
14. R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, USA, 2 edition, 2003.
15. A. Salvi. Ryour smartphone is a material scanner, 2017.
16. B. Illerhaus, E. Jasiuniene, J. Goebbels, and P. Loethman. Investigation and image processing of cellular metals with highly resolving 3D microtomography (uCT). In U. Bonse, editor, Developments in X-Ray Tomography III, volume 4503, pages 201 – 204. International Society for Optics and Photonics, SPIE, 2002.
17. K. Semendeferi, H. Damasio, R. Frank, and G. Van Hoesen. The evolution of the frontal lobes: a volumetric analysis based on three-dimensional reconstructions of magnetic resonance scans of human and ape brains. Journal of Human Evolution, 32(4):375–388, 1997.
18. T. Luhmann. Close range photogrammetry for industrial applications. ISPRS Journal of Photogrammetry and Remote Sensing, 65(6):558–569, 2010.
19. M. Rodríguez-Martín and P. Rodríguez-Gonzalvez. Learning based on 3D photogrammetry models to evaluate the competences in visual testing of welds. In IEEE Global Engineering Education Conference, pages 1576–1581, 2018.

20. M. Rodríguez-Martín and P. Rodríguez-Gonzálvez. Learning methodology based on weld virtual models in the mechanical engineering classroom. Computer Applications in Engineering Education, 27(5):1113–1125, 2019.
21. M. Fau, R. Cornette, and A. Houssaye. Photogrammetry for 3D digitizing bones of mounted skeletons: Potential and limits. Comptes Rendus Palevol, 15(8):968–977, 2016.
22. D. Gajski, A. Solter, and M. Gašparovic. Applications of macro photogrammetry in archaeology. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B5:263–266, 2016.
23. S. Marziali and G. Dionisio. Photogrammetry and macro photography. the experience of the MUSINT II project in the 3D digitizing process of small-size archaeological artifacts. Studies in Digital Heritage, 1(2):298–309, 2017.
24. B. Hassett and T. Lewis-Bale. Comparison of 3D landmark and 3D dense cloud approaches to hominin mandible morphometrics using structure-from-motion. Archaeometry, 59(1):191–203, 2017.
25. L. Inzerillo. Smart SfM: Salinas Archaeological Museum. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 42:369–374, 2017.
26. J. Fernández-Lozano, G. Gutiérrez-Alonso, M. Ángel Ruiz-Tejada, and M. Criado-Valdés. 3D digital documentation and image enhancement integration into schematic rock art analysis and preservation: The castrocontrigo neolithic rock art (NW Spain). Journal of Cultural Heritage, 26:160–166, 2017.
27. P. Lussu and E. Marini. Ultra close-range digital photogrammetry in skeletal anthropology: A systematic review. PloS one, 15(4):e0230948, 2020.
28. F. Lavecchia, M. Guerra, and L. Galantucci. The influence of software algorithms on photogrammetric micro-feature measurement's uncertainty. International Journal of Advanced Manufacturing Technology, 93:3991–4005, 2017.
29. A. Gallo, M. Muzzupappa, and F. Bruno. 3D reconstruction of small sized objects from a sequence of multi-focused images. Journal of Cultural Heritage, 15(2):173–182, 2014.
30. M. González, J. Yravedra, D. González-Aguilera, J. Palomeque-González, and M. Domínguez-Rodrigo. Micro-photogrammetric characterization of cut marks on bones. Journal of Archaeological Science, 62:128–142, 2015.
31. G. Percoco and A. Salmerón. Photogrammetric measurement of 3D freeform millimetre-sized objects with micro features: an experimental validation of the close-range camera calibration model for narrow angles of view. Measurement Science and Technology, 26(9):095203, 2015.
32. M. Shortis, C. Bellman, S. Robson, G. Johnston, and G. Johnson. Stability of zoom and fixed lenses used with digital SLR cameras. In ISPRS Symposium of Image Engineering and Vision Metrology, pages 285–290, 2006.
33. L. Galantucci, M. Guerra, and F. Lavecchia. Photogrammetry applied to small and micro-scaled objects: A review. In J. Ni, V. Majstorovic, and D. Djurdjanovic, editors, International Conference on the Industry 4.0 Model for Advanced Manufacturing, pages 57–77, Cham, 2018. Springer International Publishing.
34. L. Galantucci, F. Lavecchia, and G. Percoco. Multistack close-range photogrammetry for low-cost submillimeter metrology. Journal of Computing and Information Science in Engineering, 13(4), 2013.
35. A. Mathys and J. Brecko. Focus Stacking, pages 213–216. Amsterdam University Press, 2018.
36. L. Galantucci, M. Pesce, and F. Lavecchia. A stereo photogrammetry scanning methodology, for precise and accurate 3d digitization of small parts with sub-millimeter-sized features. CIRP Annals, 64(1):507–510, 2015.

37. G. Percoco, M. Guerra, Sanchez S., and Luigi M. Galantucci. Experimental investigation on camera calibration for 3D photogrammetric scanning of micro-features for micrometric resolution. International Journal of Advanced Manufacturing Technology, 91:2935–2947, 2017.
38. G. Percoco, F. Modica, and S. Fanelli. Image analysis for 3D micro-features: A new hybrid measurement method. Precision Engineering, 48:123–132, 2017.
39. D. Sims-Waterhouse, S. Piano, and R. Leach. Verification of micro-scale photogrammetry for smooth three-dimensional object measurement. Measurement Science and Technology, 28(5):055010, 2017.
40. D. Sims-Waterhouse, P. Bointon, S. Piano, and R. Leach. Experimental comparison of photogrammetry for additive manufactured parts with and without laser speckle projection. In Optical Measurement Systems for Industrial Inspection X, volume 10329, page 103290W. International Society for Optics and Photonics, SPIE, 2017.
41. F. Lavecchia, M. Guerra, and L. Galantucci. Performance verification of a photogrammetric scanning system for micro-parts using a three-dimensional artifact: adjustment and calibration. International Journal of Advanced Manufacturing Technology, 96:4267–4279, 2018.
42. L. Galantucci, M. Pesce, and F. Lavecchia. A powerful scanning methodology for 3D measurements of small parts with complex surfaces and sub-millimeter-sized features, based on close-range photogrammetry. Precision Engineering, 43:211–219, 2016.
43. P. Besl and Neil D. McKay. A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2):239–256, 1992.
44. Yang C. and G. Medioni. Object modelling by registration of multiple range images. Image and Vision Computing, 10(3):145–155, 1992.
45. M. Lourakis and X. Zabulis. Accurate scale factor estimation in 3d reconstruction. In International Conference on Computer Analysis of Images and Patterns, volume 8047, page 498–506, Berlin, Heidelberg, 2013. Springer-Verlag.
46. [46] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In IEEE International Conference on Computer Vision, volume 1, pages 666–673, 1999.
47. J. Heikkila and O. Silven. A four-step camera calibration procedure with implicit image correction. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1106–1112, 1997.
48. D. Lowe. Distinctive image features from scale-invariant key points. International Journal of Computer Vision, 60(2):91–110, 2004.
49. J. Cheng, C. Leng, J. Wu, H. Cui, and H. Lu. Fast and accurate image matching with cascade hashing for 3D reconstruction. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, 2014.
50. N. Ayache. Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception. Texts in Computer Science. MIT Press, Cambridge, Massachusetts, 1991.
51. P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine vision and applications, 6(1):35–49, 1993.
52. M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6):381–395, 1981.
53. R. Kuemmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard. g2o: A general framework for graph optimization. In Proceedings of the IEEE International Conference on Robotics and Automation, pages 3607–3613, Shanghai, China, 2011.
54. P. Burt and E. Adelson. A multiresolution spline with application to image mosaics. ACM Transactions on Graphics, 2(4):217–236, 1983.
55. G. Csurka, C. Zeller, Z. Zhang, and O. Faugeras. Characterizing the uncertainty of the fundamental matrix. Computer Vision and Image Understanding, 68(1):18–36, 1997.

56. P. Moulon, P. Monasse, R. Perrot, and R. Marlet. OpenMVG: Open multiple view geometry. In International Workshop on Reproducible Research in Pattern Recognition, pages 60–74. Springer, 2016.
57. P. Moulon, P. Monasse, and R. Marlet. Adaptive structure from motion with a contrario model estimation. In K. Lee, Y. Matsushita, J. Rehg, and Z. Hu, editors, Asian Conference in Computer Vision, pages 257–270. Springer, 2013.
58. F. Espuny, P. Monasse, and L. Moisan. A new a contrario approach for the robust determination of the fundamental matrix. In Image and Video Technology – PSIVT 2013 Workshops, pages 181–192, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.
59. M. Lourakis and A. Argyros. SBA: A software package for generic sparse bundle adjustment. ACM Transactions on Mathematical Software, 36(1), Mar 2009.
60. D. Cernea. OpenMVS: Multi-view stereo reconstruction library. https://cdcseacave.github.io/openMVS, 2008. [Online; accessed 6-June-2021].
61. E. Catmull. A subdivision algorithm for computer display of curved surfaces. The University of Utah, 1974.
62. C. Barnes, E. Shechtman, A. Finkelstein, and Dan B. Goldman. Patchmatch: A randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics, 28(3), 2009.
63. M. Jancosek and T. Pajdla. Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. International scholarly research notices, 2014, 2014.
64. V. Hiep, P. Labatut, J. Pons, and R. Keriven. High accuracy and visibility-consistent dense multiview stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34:889–901, 08 2011.
65. R. Gal, Y. Wexler, E. Ofek, H. Hoppe, and D. Cohen-Or. Seamless montage for texturing models. Computer Graphics Forum, 29(2):479–486, 2010.
66. B. Wang, P. Pan, Q. Xiao, L. Luo, X. Ren, R. Jin, and X. Jin. Seamless color mapping for 3D reconstruction with consumer-grade scanning devices. In European Conference on Computer Vision Workshops, pages 633–648, 2018.
67. M. Waechter, N. Moehrle, and M. Goesele. Let there be color! large-scale texturing of 3D reconstructions. In European Conference on Computer Vision, pages 836–850, Cham, 2014. Springer International Publishing.
68. Jylänki, Jukka. A thousand ways to pack the bin-a practical approach to two-dimensional rectangle bin packing, 2010.
69. A. Weyer, Roig P., D. Pop, J. Cassar, A. Ozkose, V. Jean-Marc, and I. Srsa, editors. EwaGlos - European Illustrated Glossary of Conservation Terms for Wall Paintings and Architectural Surfaces. Michael Imhof Verlag, Stettiner Strasse 25, Petersberg, Germany, 2011.
70. M. Pérez, P. Casanova-Salas, P. Twardo, P. Twardo, A. León, D. Mladenic, B. Massri, R. Troncy, T. Ehrhart, G. Cicero, M. Vitella, M. Gaitan, J. Gimeno, E. Ribes, M. Fernández, and C. Portales. From historical silk fabrics to their interactive virtual representation and 3d printing. Sustainability, 12(18), 2020.
71. X. Zabulis, P. Koutlemanis, N. Stivaktakis, and N. Partarakis. A low-cost contactless overhead micrometer surface scanner. Applied Sciences, 11(14), 2021.
72. B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. 3D Gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4), July 2023.
73. B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In European Conference in Computer Vision, pages 405–421, 2020.
74. D. Marr and T. Poggio. Cooperative computation of stereo disparity. Science, 194(4262):283–287, 1976.

75. A. Karami, F. Menna, and F. Remondino. Combining photogrammetry and photometric stereo to achieve precise and complete 3D reconstruction. Sensors, 22(21), 2022.
76. A. Petruccioli, F. Gherardini, and F. Leali. Assessment of close-range photogrammetry for the low-cost development of 3D models of car bodywork components. International Journal on Interactive Design and Manufacturing (IJIDeM), 16(2):703–713, 2022.
77. C. Nicolae, E. Nocerino, F. Menna, and F. Remondino. Photogrammetry applied to problematic artefacts. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 40:451–456, 2014.
78. D. Michel, X. Zabulis, and A. Argyros. Shape from interaction. Machine Vision and Applications, 25(3):1077–1087, 2014.
79. J Canny. A computational approach to edge-detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.
80. F. Hausdorff. Grundzüge der Mengenlehre. Leipzig Viet, 1914.
81. TetraVision. 3D scan of a 2 euro coin. https://www.tetravision.be/portfolio-item/3d-scan-of-a-2-euro-coin/, 2024. Accessed on December 6, 2024.
82. X. Zhou. Multiple Auto-Adapting Color Balancing for Large Number of Images. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL7:735–742, 2015.
83. Y. Shin, M. Park, Y. Jeon, Y. Moon, S. Lee, and K. Yoon. Tone correction with dynamic objects for seamless image mosaic. In K. Kutulakos, editor, Trends and Topics in Computer Vision, pages 104–117, Berlin, Heidelberg, 2012. Springer.
84. A. Pentland. A new sense for depth of field. IEEE Transactions on Pattern Analysis and Machine Intelligence, 9(4):523–531, 1987.
85. P. Grossmann. Depth from focus. Pattern Recognition Letters, 5(1):63–69, 1987.
86. Ivo Ihrke, Kiriakos N. Kutulakos, Hendrik P. A. Lensch, Marcus Magnor, and Wolfgang Heidrich. Transparent and specular object reconstruction. Computer Graphics Forum, 29(8):2400–2426, 2010.
87. F. Meriaudeau, R. Rantoson, K. M. Adal, D. Fofi, and C. Stolz. Non-conventional imaging systems for 3D digitization of transparent objects: Shape from polarization in the IR and shape from visible fluorescence induced UV. in 3rd International Topical Meeting on Optical Sensing and Artificial Vision: OSAV'2012, pages 34–40, 2013.
88. A. Karami, R. Battisti, F. Menna, and F. Remondino. 3D digitization of transparent and glass surfaces: State of the art and analysis of some methods. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2022:695–702, 5 2022.
89. Max Born and Emil Wolf. Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light. Cambridge University Press, 7th edition, 1999.
90. P. I. Stavroulakis and R. K. Leach. Invited review article: Review of post-process optical form metrology for industrial-grade metal additive manufactured parts. Review of Scientific Instruments, 87:041101, 4 2016.
91. Influence of Coating Spray on Surface Measurement Using 3D Optical Scanning Systems, volume Volume 1: Additive Manufacturing; Manufacturing Equipment and Systems; Bio and Sustainable Manufacturing of International Manufacturing Science and Engineering Conference, 06 2019. V001T02A009.
92. Borislav Trifonov, Derek Bradley, and Wolfgang Heidrich. Tomographic reconstruction of transparent objects. In ACM SIGGRAPH 2006 Sketches, SIGGRAPH '06, page 55–es, New York, NY, USA, 2006. Association for Computing Machinery.

93. Haiyun Guo, Haowen Zhou, and Partha P. Banerjee. Use of structured light in 3D reconstruction of transparent objects. Appl. Opt., 61(5):B314–B324, Feb 2022.
94. Landmann, Martin, Speck, Henri, Dietrich, Patrick, Heist, Stefan, Kühmstedt, Peter, and Notni, Gunther. Fast 3D shape measurement of transparent glasses by sequential thermal fringe projection. EPJ Web Conf., 238:06008, 2020.
95. Fabrice Mériaudeau, L. Alonso Sanchez Secades, Gonen Eren, A. Ercil, Fred Truchetet, Olivier Aubreton, and David Fofi. 3-D scanning of nonopaque objects by means of imaging emitted structured infrared patterns. IEEE Transactions on Instrumentation and Measurement, 59(11):2898–2906, 2010.
96. Haochong Huang, Enhui Yuan, Dongshun Zhang, Dexin Sun, Meihui Yang, Zhiyuan Zheng, Zili Zhang, Lu Gao, Spozmai Panezai, and Kunfeng Qiu. Free field of view infrared digital holography for mineral crystallization. Crystal Growth and Design, 23:7992–8008, 11 2023.
97. Rindra Rantoson, Christophe Stolz, David Fofi, and Fabrice Meriaudeau. Optimization of transparent objects digitization from visible fluorescence ultraviolet-induced. Optical Engineering, 51(3):033601, 2012.
98. Matthias B. Hullin, Martin Fuchs, Ivo Ihrke, Hans-Peter Seidel, and Hendrik P. A. Lensch. Fluorescent immersion range scanning. ACM Trans. Graph., 27(3):1–10, Aug 2008.
99. Peter Fried, Jonathan Woodward, David Brown, Drew Harvell, and James Hanken. 3D scanning of antique glass by combining photography and computed tomography. Digital Applications in Archaeology and Cultural Heritage, 18:e00147, 2020.
100. Mathias Ferraton, Christophe Stolz, and Fabrice Mériaudeau. Optimization of a polarization imaging system for 3D measurements of transparent objects. Opt. Express, 17(23):21077–21082, Nov 2009.
101. Mingqi Shao, Chongkun Xia, Dongxu Duan, and Xueqian Wang. Polarimetric inverse rendering for transparent shapes reconstruction. ArXiv, abs/2208.11836, 8 2022.
102. Damien Michel, Xenophon Zabulis, and Antonis A. Argyros. Shape from interaction. Machine Vision and Applications, 25:1077–1087, 5 2014.
103. M.Mikhnevich and D.Laurendeau. Shape from silhouette in space, time and light domains. In Proceedings of the 9th International Conference on Computer Vision Theory and Applications, pages 368–377. SCITEPRESS - Science and Technology Publications, 2014.
104. J. Munkberg, J. Hasselgren, T. Shen, J. Gao, W. Chen, A. Evans, T. Mueller, and S. Fidler. Extracting triangular 3D models, materials, and lighting from images. arXiv:2111.12503, 2021.
105. Zhengqin Li, Yu-Ying Yeh, and Manmohan Chandraker. Through the looking glass: Neural 3D reconstruction of transparent shapes. In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 1259–1268, 4 2020.
106. Hichem Guerboukha, Kathirvel Nallappan, and Maksim Skorobogatiy. Toward real-time terahertz imaging. Advances in Optics and Photonics, 10:843, 12 2018.
107. B. Recur, A. Younus, S. Salort, P. Mounaix, B. Chassagne, P. Desbarats, J-P. Caumes, and E. Abraham. Investigation on reconstruction methods applied to 3d terahertz computed tomography. Optics Express, 19:5105, 3 2011.
108. Anumol Mathai, Ningqun Guo, Dong Liu, and Xin Wang. 3d transparent object detection and reconstruction based on passive mode single-pixel imaging. Sensors, 20:4211, 7 2020.
109. C.J. Phillips, M. Lecce, and K. Daniilidis. Seeing glassware: from edge detection to pose estimation and shape recovery. In Robotics: Science and Systems, 2016.
110. James Sharpe, Ulf Ahlgren, Paul Perry, Bill Hill, Allyson Ross, Jacob Hecksher-Sørensen, Richard Baldock, and Duncan Davidson. Optical projection tomography as a tool for 3D microscopy and gene expression studies. Science, 296(5567):541–545, 2002.

111. Wim van Aarle, Willem Jan Palenstijn, Jeroen Cant, Eline Janssens, Folkert Bleichrodt, Andrei Dabravolski, Jan De Beenhouwer, K. Joost Batenburg, and Jan Sijbers. Fast and flexible x-ray tomography using the Astra toolbox. Opt. Express, 24(22):25129–25147, Oct 2016.
112. L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. J. Opt. Soc. Am. A, 1(6):612–619, Jun 1984.
113. A. Laurentini. The visual hull concept for silhouette-based image understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(2):150–162, 1994.
114. Cloud compare software. https://www.cloudcompare.org/. Accessed: 2023-03-07.
115. Petra Helmholz, David Belton, Nick Oliver, Joshua Hollick, and Andrew J. Woods. The influence of the point cloud comparison methods on the verification of point clouds using the Batavia reconstruction as a case study. In IKUWA6 Shared Heritage: Proceedings of the Sixth International Congress for Underwater Archaeology, 2020.
116. Gergana Antova. Application of areal change detection methods using point cloud data. IOP Conference Series: Earth and Environmental Science, 221(1):012082, Jan 2019.
117. S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1 (CVPR'06). IEEE, 2006.
118. Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In ECCV, 2020.
119. Edite Figueiras, Ana M. Soto, Danilo Jesus, M. Lehti, J. Koivisto, J. E. Parraga, J. Silva-Correia, J. M. Oliveira, R. L. Reis, M. Kellomäki, and J. Hyttinen. Optical projection tomography as a tool for 3D imaging of hydrogels. Biomed. Opt. Express, 5(10):3443–3449, Oct 2014.
120. Ali Ertürk, Klaus Becker, Nina Jährling, Christoph P Mauch, Caroline D Hojer, Jackson G Egen, Farida Hellal, Frank Bradke, Morgan Sheng, and Hans-Ulrich Dodt. Three-dimensional imaging of solvent-cleared organs using 3DISCO. Nature Protocols, 7:1983–1995, 11 2012.
121. Matweb material properties. https://www.matweb.com/search/DataSheet.aspx?MatGUID=a696bdcdff6f41dd98f8eec3599eaa20&ckck=1. Accessed: 2023-03-07.
122. Karen J. I. Lee, Grant M. Calder, Christopher R. Hindle, Jacob L. Newman, Simon N. Robinson, Jerome J. H. Y. Avondo, and Enrico S. Coen. Macro optical projection tomography for large-scale 3D imaging of plant structures and gene activity. Journal of Experimental Botany, page erw452, 12 2016.
123. Ray simulator for optics. https://phydemo.app/ray-optics/simulator/. Accessed: 2023-03-07.