The Framework Programme for Research & Innovation
Research and Innovation actions (RIA)

Project Title:

Multimodal Scanning of Cultural Heritage Assets for their multilayered digitization and preventive conservation via spatiotemporal 4D Reconstruction and 3D Printing

Scan4Reco

Grant Agreement No: 665091

[H2020-REFLECTIVE-7-2014] Advanced 3D modelling for accessing and understanding European cultural assets

Deliverable

D5.2 - Protocol for joint acquisition of Low res 3D, microprofilometry and RTI

Deliverable No.		**D5.2**	
Work Package No.	**WP5**	Work Package Title and task type	**Scan4Reco Spatiotemporal Simulation & Reconstruction**
Task No.	**T5.1**	Task Title	**Multi-sensorial Data Fusion**
Lead beneficiary		**UNIVR**	
Dissemination level		**PU**	
Nature of Deliverable		**Report**	
Delivery date		**M16 (31 January 2017)**	
Status		**F: final**	
File Name:		**[Scan4Reco] Deliverable 5.2_final.pdf**	
Project start date, duration		**01 October 2015, 36 Months**	

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under Grant Agreement n°665091

Authors List

Leading Author (Editor)				
Surname		*Initials*	*Beneficiary Name*	*Contact email*
Giachetti		A	UNIVR	andrea.giachetti@univr.it
Co-authors (in alphabetic order)				
#	*Surname*	*Initials*	*Beneficiary Name*	*Contact email*
1	Ciortan	I	UNIVR	irina.ciortan@gmail.com
2	Pintus	R	CRS4	ruggero.pintus@gmail.com
3	Gobbetti	E	CRS4	gobbetti@crs4.it
4	Papachristou	K	CERTH	kostas.papachristou@iti.gr
5	Dimitriou	N	CERTH	nikdim@iti.gr

Reviewers List

List of Reviewers (in alphabetic order)				
#	*Surname*	*Initials*	*Beneficiary Name*	*Contact email*
1	*Urban*	*P*	*FHG*	*philipp.urban@igd.fraunhofer.de*
2	*Galeotti*	*M*	*OPD*	*monica.galeotti@beniculturali.it*

Document history			
Version	***Date***	***Status***	***Modifications made by***
0.1	26/12/2015	ToC and Guideline	Giachetti, Ciortan
0.2	17/01/2017	Draft 1 (UNIVR's & CRS4's contribution)	Giachetti, Gobbetti, Pintus, Ciortan
0.3	24/01/2017	Draft 2 (CERTH's contribution)	Papachristou, Dimitrou
1.0	26/01/2017	Draft for internal review	Giachetti, Papachristou
2.0	31/01/2017	Final (to be submitted)	Drosou

List of definitions & abbreviations

Abbreviation	Definition
CH	Cultural Heritage
DICOM	Digital Imaging and COmmunications in Medicine
RTI	Reflectance Transformation Imaging

Executive Summary

This report describes the protocol developed for the joint use of microprofilometry and Reflectance Transformation Imaging data spatially referenced to a low resolution 3D model acquired with the scanner in order to characterize patches of artworks and samples.

The rationale of the protocol is to support multimodal studies as well as easy retrieval of data and processing of a large number of sample/objects in collections.

While in this project the purpose of the proposed protocol is to provide the basic support for the activities planned in “Task T5.2 - Digital modelling of material appearance & features” and Task T5.5 - Degradation Identification & automatic spatiotemporal Annotation”, the underlying concepts can lead to a proposal for a novel generic way to support any kind of studies of artworks or material samples based on local multimodal patch based acquisitions.

Table of Contents

Table of Figures

Table of Tables

1. Introduction

1.1 Purpose of the deliverable

Demonstration of multimodal study of artworks or material samples for monitoring and diagnosis is one of the main goals of Scan4Reco. The use of diverse information captured by different devices is a key factor for characterizing surface (and volumetric) features of materials as well as to analyze their degradation.

In Scan4Reco one of the processing pipelines needs to jointly process two high resolution surface data sets (RTI and microprofilometry) captured on a potentially larger object on selected approximately flat regions. The object is also reconstructed with a coarse 3D scanning that can be used to support spatial referencing, alignment or capture of the local patches. This joint acquisition steps are described in this deliverable.

In order to fully exploit the information provided by the different devices it is mandatory to develop accurate and repeatable measurement procedures, to record all the available metainformation linking all measurements to the object description, storing all the details about the measurement devices and procedures, and supporting spatial alignment of the data with specific information. Furthermore, collecting multimodal information for time analysis and classification requires also an easy retrieval of selected studies in a collection, and this can be obtained with a careful data organization of data based on the specific kind of studies to be performed on them.

This deliverable, therefore, does not describes the generic protocols used in Scan4Reco to capture single patch based surface measurement (Reflectance Transformation Imaging and microprofilometry) on samples and artworks, already described in D3.4 (a procedure for acquisition of a global 3D scan is here provided in Annex I), but defines methods to store the related information and also a data organization protocol used to store all the metainformation at different levels (artwork, ageing, study) in an easy to handle collection, allowing the association and alignment of patch based measurements and a global low resolution 3D scan. Furthermore it describes example methods for the joint processing of the information, sketching specific procedures to jointly register the data on the basis of the acquisition procedures and data content.

The basic idea of our work is to define a data organization similar to the one used in the medical domain by the DICOM protocol, supporting the concept of multiple studies on a "patient" that corresponds to an artwork or a material sample, organizing a hierarchical data storage with unique and standardized metadata for easy data query and retrieval, and supporting pre-alignment of patches with respect to the global model.

1.2 Relation with other WPs, tasks and deliverables

This deliverable is mainly connected with task 5.1 "Multi-sensorial Data Fusion" and with the tasks 5.5 Degradation Identification and automatic spatiotemporal Annotation. However, the protocol for the multimodal surface acquisition/analysis is related to many components of the global Scan4Reco architecture: apart from the Global scanning and surface acquisition components (HW/VISDEPTH, SW/SCAN, SW/REG, HW/MPROF, HW/MSRTI, HW/HDMSRI, SW/SWSURF, SW/RTIPROC), the protocol is related to the Sample/artwork preparation modules and the global protocol defines the SW/ANNOTATION module: is related to the Assisted Positioning components, that must provide the annotation of patches position and orientation in the global reference frame. Finally all analysis modules working on multimodal data could benefit of the proposed data organization allowing easy query/retrieve of multimodal data for visualization, fusion, model training and testing.

The organized data archive of the data captured in the Scan4Reco multimodal surface analysis acts as a hub connecting all the system components like the PACS system does in digitalized hospitals. This makes the query/retrieval and analysis of data easy.

2. Multimodal Measurements on Cultural Heritage Objects

2.1 Requirements for data fusion

Several authors pointed out the importance of exploiting multimodal 3D capture techniques for artwork documentation. In a recent survey, Remondino et al [1] noted that "there is an increasing use of hybrid sensors and platforms, in order to collect as many different features as possible. The combination of different data sources allows the creation of different geometric levels of detail (LoD) and the exploitation of the intrinsic advantages of each sensor". The use of multimodal information presents several challenges. In [2] is pointed out that "The main difficulty is to face the multiple dimensions of the data acquisition and their registration into a common spatial reference system". Several works in literature deals with multimodal data registration for CH applications [3].

However, prior to study specific solutions for the algorithmic alignment and integration of two or more data sources, we believe that there is a first requirement for the organization of multimodal studies that is the definition of a generic protocol where both the imaging procedures and the multimodal studies are organized.

We need to consider the specific studies for the analyses and monitoring tasks of interest and support them with an organized storage of all the relevant information that can then be exploited by end-users.

A similar protocol or framework is currently missing in the CH domain, but it is actually the first requirement for the joint use of multimodal information for practical tasks.

As previously noted, this can be realized inspired by the Radiology procedures, defining procedures for the annotation of objects, defining standard procedures for each kind of multimodal study on each object, defining the set of acquisitions planned, and storing all the information that is relevant for the study finalization as metadata.

For instance, the surface analysis planned in Scan4Reco for object and material samples, features the rough global modelling of the object, the acquisition of microprofilometry and RTI capture on approximately planar patches at different ageing conditions, and the joint processing of the data.

This kind of study requires

- a standardized annotation of objects and samples
- well defined procedures to capture data and standardized annotations of acquisition protocols
- support for (rough) spatial referencing of patches on the global object

The proposed protocol and framework gives the support for performing the study independently on the specific visualization, fine scale registration or specific techniques for joint processing of data to recover useful parameters.

We underline this fact as in CH as in medical domain is mandatory to support the interoperability of different operators and vendors, and the research on different kind of studies to be realized on the same data.

2.2 State-of-the-art of metadata standards

The enrichment of Cultural Heritage digital collections simultaneously with the growing number of available technologies brings along the imperative need for managing the resources in a clear and organized structure that could ideally converge to a common standard [4]. This need is fulfilled by metadata, which is implemented in order to **"describe, identify, and facilitate the access, usage and management of (digital) resources"** [5]. Metadata is essential for keeping track of the full life-cycle of a Cultural Heritage asset [6], and the intermediary activities involved in generating an output digital model from an input physical object [8]. In other words, metadata opens the gate to interpretation, providing extra-meaning and a legend on how to read and connect raw scientific datasets.

Metadata are particularly useful to support data analysis in the field of 3D scan, where there is a high level of heterogeneity in the choice of technologies, instruments and methodology that can contribute to **data capture** and **data processing** [6], and where it is likely that **data fusion** is applied not singularly, but multiple times at different phases of a 3D reconstruction [8].

The need for order has convinced end-users and stakeholders of Cultural Heritage to add metadata systems to their projects. This resulted in an inflation of **project-tailored metadata schemas** that achieve the ad-hoc purpose for which they have been created, but make the task of standardization difficult. This latter drawback is generally compensated for by **metadata mappings** that employ semantic interoperability by liaising equivalences between various standards. Adding another **abstractization level** [5], **ontologies** can be built upon the skeleton of metadata mappings. [9] presents an overview and assessment of the different metadata schemas developed for cataloguing and documenting Cultural Heritage and then, in a later work [4] they propose a metadata standard, STARC, for 3D objects that brings together the missing parts identified in the previous research. As a result, they highlight a **set of characteristics** that shouldn't be overlooked in the documentation of 3D objects: digital provenance, paradata, software involved, capturing devices, methodologies pursued and full pipeline for deriving the final results.

Indeed, **digital provenance** and **paradata** seem to be generally acknowledged as key features for the metadata of 3D objects in the literature [6], [4], [8]. [6] make the distinction between provenance – the collection of **technical processes** linked to the origins and derivations of the digital resource – and paradata holds information about the **human processes** responsible for the understanding and interpretation of the CH asset. Furthermore, they describe how provenance is covered by the CRMdig schema [10], give credit to the paradata principles of the London Charter [11] and integrate the couple provenance-paradata within the CARARE framework fostered by the Europeana Data Model [12] finding a good compromise between specific projects needs and standardization needs for the 3D ICONS project. Digital provenance is viewed as a **blueprint** of a cultural artefact and storing this information translates into the possibility of tracing down the authentic object [8]. At the same data, paradata furnishes the **intellectual transparency** without which the blueprint would otherwise be illegible.

2.2.1 Metadata architecture

Roughly, there are two main approaches for modelling metadata [6][12]: **object-centric** and **event-centric**. In the first approach, the object is in the spotlight and dedicated attention is given to creating a thoroughly descriptive characterization of the cultural heritage asset and its attributes. From this category the most widespread standard is the **Dublin Core metadata set**. In the second approach, the focus is relayed to the sequence of events that the cultural object had taken part into. A representative standard of the event-centric metadata modelling is **CRMdig**, which defines the following event instances: project creation event, the ob-

ject acquisition event, the detailed sequence event and lastly, a capture event for every "end-member" multimedia file. This rationale is in fact an analogue of the intuitive directory structure of a typical acquisition campaign, as remarked by [7], where the commonplace folder levels identified are as follows:

- Project root directory – enclosing all the relevant data about the project
- Object directory – contains data for each of the captured objects
- Sequence directory – for each captured object, one folder for each series of measurements

Similar to the hierarchical folder structure is also the organization based on **wrapper elements**. Wrappers group relative information, usually lacking a stand-alone semantics because their semantics is driven from the constitutive elements [5]. The STARC schema [4] stands as example of a wrapper based structure. It is made up of one global wrapper named PROJECT and is divided into four main wrappers (Project Information, Cultural heritage Asset, Digital Resource Provenance, Activities) that further branch into sub-wrappers.

For large-scale projects, where storage is more problematic and the amount of data becomes harder to monitor with the above-mentioned architectures, the solution could be switched to a **repository-centric approach**, introduced by the work of [8]. They present the design principles and implementation choices of a distributed object repository for cultural heritage objects. The focus is given here to the administration and architecture of a central repository server, which can facilitate database queries and reporting.

2.2.2 Generating Metadata

The so-far tackled theoretical aspects have sketched a preamble on how demanding metadata collection could be in practice. Apart from being a resource-consuming process (time, human resources, financial resources, etc.), it is also very prone to errors as it is semi-manual and user intervention is necessary. For this reason, it is recommended adhering to an automation of the metadata generation process. The non-triviality of this task has given it a quiet close-to-silent voice in the Cultural Heritage literature. However, [13] proposes a Metadata Generator by using generic dynamic input forms for annotating the digital provenance and the semantic links. They opt for key-value pairs to store essential information into intermediate formats, that they later overlay on Resource Description Framework templates to generate final rich metadata formats.

2.3 CH objects targeted by Scan4Reco

The Cultural Heritage assets multi-modally acquired in the Scan4Reco project can be split in two categories:

- Samples developed in controlled laboratory conditions, which recreate the appearance and material composition of real CH artefacts.
- Signed artworks originating from cultural institutions, museums, private collections.

In Scan4Reco, three sample datasets have been created, with different basis material (silver plates, bronze plates and egg-tempera patches on cardboard) on which several coatings and treatments are initially applied and then artificial aging is monitored over fixed moments in time.

These data will allow quantitative tests or assessment and the possibility to model the spatio-temporal changes in appearance triggered along the applied processes. Onsite demonstration will be based on acquisitions performed on sighed artworks.

Multimodal surface analysis in the two kinds of objects is different, as for small surface patches it is possible to capture full shapes both with RTI and microprofilometry and microprofilometric scan can be used as reference 3D model. Scan4Reco material samples, however, have the necessity of recording the ageing information useful for ageing modelling.

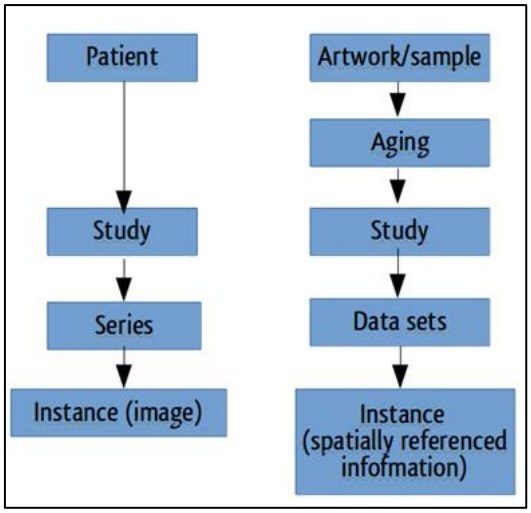


Figure 1: Query/retrieve Information models for DICOM archives (left) and for our archive (right).

3. The proposed protocol for organized data annotation

3.1 General architecture

The specific procedures for local surface capture with RTI and microprofilometry are just a part of an object study that we organized with a generic protocol for annotation and data organization. The basic idea is to follow a procedure similar to the one used for medical imaging studies in hospitals. Each object of study (artwork or material sample) is the equivalent of the patient and when its study is started, it must be annotated registering its basic information as it is done for patients in Radiology Information Systems (RIS). We defined therefore a set of tag-value couples defining this information.

As for each object in Scan4Reco we will have also different artificial ageing procedures, we also defined ageing-level meta-information that can be stored in another group of tag-value couples. This information differentiates natural and artificial aging procedures including also all the relevant information on the artificial aging devices and protocols used.

Medical doctors plan sets of diagnostic imaging captures with different modalities and there is a standard for execution and storage of different imaging studies in PACS (Picture Archive and Communication System) systems. And for each disease treatment there are typical sets of imaging studies planned, executed and stored and an organized information flow. Multimodal data are collected and radiologists /medical doctors can retrieve data, visualize and process them on radiology workstations. The creation of a full DICOM-like protocol including definitions of all services and objects required in the CH domain clearly goes beyond the scope of our work, but we believe that it should be developed by the community.

We just organize artworks and sample studies in the same way as in patient study, similarly to what is done in the DICOM3 protocol. For each uniquely annotated **object (artwork or sample)**, we can acquire several imaging studies and store them with all the possible metadata. These data are stored at the study level of our hierarchical organization and can be retrieved using the standard information model for Query/Retrieve that in DICOM is based on Patient (object), Study and Series. We added in the hierarchy also an aging level to account to possibly different natural or artificial aging procedures performed on the objects. Data stored in our archives are not necessarily standard images, but typically data (surfaces, clouds, grids) spatially referenced in a global coordinate system. Note that this unique spatial referencing should in general be provided by the study protocol and it is not mandatory in the metadata set. We therefore developed specific methodologies to recover the useful

information by adopting ad hoc strategies in the acquisition phase and using specific computational solutions.

3.2 Artworks and samples metadata annotation

Our metadata schema is modelled by an **event-centric design** with the following folder levels: *Object – Aging moments – Acquisition series.* Within the acquisition series, we have defined specific metadata standards for each of the three acquisition methods: RTI, microprofilometer and Low resolution 3D scanning. The Acquisition metadata fields are grouped into several wrappers: Setup specifications, Hardware specifications, Software specifications, Output files and Spatial Reference.

3.2.1 Object level

The object level includes metadata which gives information about the characteristics of the physical cultural asset, be it a sample or an artwork. The fields are described in Table 1.

Table 1: Metadata fields for Cultural Heritage object.

Field	Description
UID	Represents a unique combination of alphanumeric characters that references to a sample or an artwork in the project's database.
TYPE	The type of artifact: to be chosen between artwork or lab simulated mock-ups (samples).
NAME	The text that describes the name of the sample. It also identifies the sample/artwork, but not necessarily unique.
AUTHOR	The person who created the sample or the artwork.
SOURCE	Where the object comes from: museum collection, research laboratory, cultural institution, etc.
DATE_OF_CREATION	The date and time when the physical object (sample or artwork) was completed. Format: dd/MM/YY hour:minutes
DETAILS_OF_CREATION	This might refer on how the sample was created, the history, chronology or steps of creation.
EXTENT	Refers to the physical dimensions of the sample or artwork (height, width, thickness).
SEMANTICS_DESCRIPTION	The description of semantics has to briefly guide through the content of the sample or artwork. Example 1: the sample is made up of 95% uncoated silver (left) and 5% coated silver (right). Example 2: the artwork is an Icon depicting Virgin Mary.
MATERIALS	The constitutive elements that make up the sample/artwork and their type: metals, pigments, support, etc.
CONDITION	The status of the sample/artwork that might regard the following characteristics: novelty, previous restoration, visible aging effects. Example 1: new, out-of-the-oven sample, with no aging effect. Example 2: painting was partially restored on <date>, but there is still visible a red pigment discoloration.
TREATMENT	Restoration method that an artwork has undergone or the chemical treatment applied to a sample at the moment of creation.

3.2.2 Aging level

Since each object will be acquired at definite moments in time, the aging folder level is focused on describing the type of aging (natural or artificial). Aging-related fields group is shown in

Table 2.

Table 2: Metadata fields for Aging procedure.

Field	Description
UID	Unique identifier of the aging process, that can be made up of a unique combination of alphanumeric characters, which ought to be representative of the aging method applied and the time frame. Example 1: UV Exposure – t1.
TYPE	The category of aging process: artificial (human-provoked) or naturally (due to the passing of time, without human intervention).
DATE	The data and time when the aging treatment was applied.
LOCATION	Geographical place where the aging process has taken place.
METHOD_NAME	The name of the aging mechanism involved.
SCIENTIST_IN_CHARGE	The person in charge of the conducting or supervising the aging process.
DESCRIPTION	Description of the aging method.
AGENTS	The bio-chemical agents responsible for the aging.
EXPECTED_EFFECTS	The theoretical expected change in the appearance, structure and geometry of the physical object onto which the aging effect was applied.
ATTRIBUTES_AFFECTED	The intrinsic properties of the object that were affected by the aging process.

3.2.3 Annotation tool

In order to optimize the metadata documentation process, we have developed an annotation tool with form fields and generic formats in this sense. Similarly to (Schrottner, 2012), we use tag-value pairs, with generic formats and as recommended in (D'Andrea, 2013), we adhere to the use of controlled vocabularies, especially for materials and techniques, as a way to limit the errors and help the user. Figure 2 shows a snapshot of the user interface allowing an easy annotation of aging and object properties.

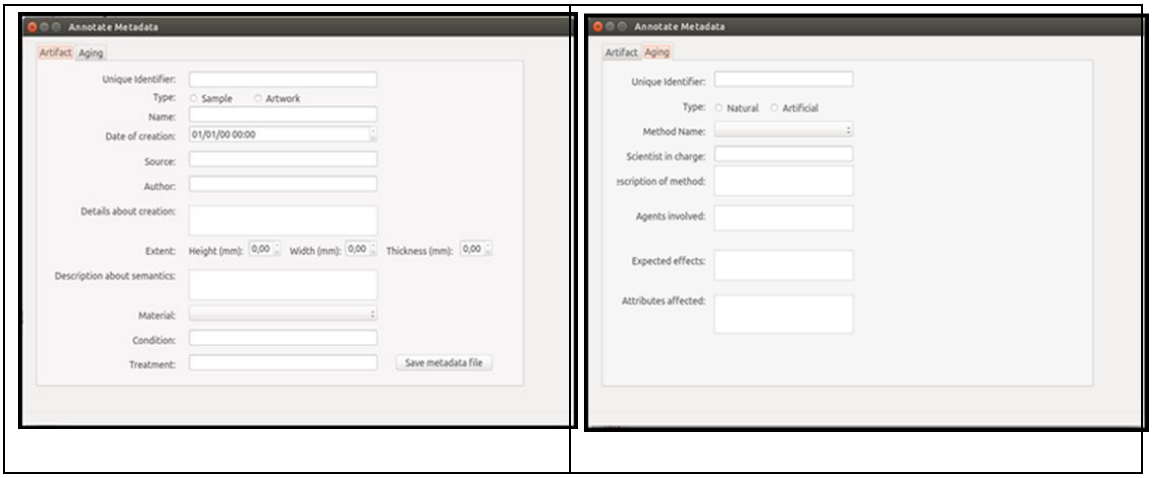


Figure 2: Tool for annotating metadata. Top: Tab for inserting metadata fields for the CH Object. Bottom: Tab for filling in metadata for the Aging treatment.

3.2.4 Study level

The study level includes all the information related to the single study. Each study includes metadata and measurement files. In our organized archive, each study corresponds to a data folder including the study files. For a RTI study, metadata are divided in groups (as in DICOM headers) related to setup, hardware settings, software settings, measurement files, spatial reference (

Table 3). The data files (appearance profiles and cropped image stacks) are stored in the same folder and described in the metadata list.

Table 3: Metadata fields for RTI acquisition. The fields grouped into wrappers are color-coded: Set-up, Hardware, Software, Spatial Reference, data files and encoding.

Field	Description
ACQUISITION_ID	Unique ID of the acquisition
ACQUISITION_DATE	Date and time when the acquisition was performed.
ACQUISITION_TYPE	The type of study applied to the object, such as: RTI, Microprofilometry, Low-res scanning, etc
LOCATION_NAME	Geographical place where the acquisition was performed. Example: Verona, Sardinia, etc.
OPERATOR_NAME	The name of the operators
SETUP_TYPE	Highlight or Dome.
SETUP_DESCRIPTION	Description of the dome or the objects included in the capture scene, number of spheres, the use of frame, calibration targets, etc
CAMERA_OBJECT_DISTANCE	Approximate distance from the camera to the object in cm.
CAMERA_ID	The camera's id is composed of the CameraName_LocationAbbreviated. Example: NikonD810_VR
CAMERA_CHARACTERIZATION	Pointer to the file/folder comprising information about the camera, technical specifications , measurements, etc.
LENS_ID	The lens id is comprised by the brand and focal length of the lens. Example: Nikkor_50mm
LENS_CHARACTERIZATION	Pointer to the file/folder comprising information about the lens, technical specifications , measurements, etc.
FILTER_ID	The name of filter used, if any.
FILTER_CHARACTERIZATION	Pointer to the file/folder comprising information about the filter, technical specifications , measurements, etc.
LIGHT_SOURCE_ID	The name and type of light source used.
LIGHT_SOURCE_CHARACTERIZATION	Pointer to the file/folder comprising information about the light source, technical specifications , measurements, etc.
SPECTRUM_RANGE	UV, Visible or Near Infrared
CAMERA_ISO	The sensitivity to light of the camera sensor.
CAMERA_SHUTTER_SPEED	The exposure to light of the camera sensor.
CAMERA_APERTURE	The opening letting the light reach the camera sensor.
CAMERA_CENTER_POSITION	The spatial coordinates of the optical camera center.
CAMERA_AXIS_DIRECTION	The direction of the optical axis of the camera (X,Y, Z) relative to the object plane.
CAMERA_INTRINSICS	A 3x3 matrix comprising the intrinsic parameters of the camera.
IMAGE_WIDTH	The number of columns of the image.
IMAGE_HEIGHT	The number of rows of the image.
CROP_COORDINATES	The spatial coordinates (x,y, w, h) of the crop area relative to the full image, on which the appearance profile was computed.
APPEARANCE_PROFILE_INFO_FILE	A text file auxiliary to reading the appearance profile name, giving information about the dimension and type of appearance profile file. The name of the appearance profile info file is typically the sample name, with the .info
APPEARANCE_PROFILE_DATA_FILE	Name of the binary file that contains for each pixel, 9 numeric coefficients characterizing the light direction, together with the reflectance value, encoded either as a single value (monochrome intensity) or a 3-channel value (corresponding to the R, G, B chromaticity). The name of the appearance profile file is typically the sample name with .apx or .aph
CROPPED_IMAGES	Pointer with a folder with the original images cropped in the region of interest of the sample
POINTER_TO_THE_RAW_DATA	Pointer to a UID of a folder with the original image acquisition

Similarly, microprofilometric data are characterized by a large amount of information that is organized and grouped as shown in Table 4. Measurement files can be actually stored in different formats (text, raw binary, tiff), specified for each instance in the related metadata.

Low resolution 3D scans are stored in a study as well, coupling the mesh file with a metadata header as shown in Table 5.

Table 4**:** Metada fields for **microprofilometric** Acquisition. The fields grouped into wrappers are color-coded: Setup, Hardware, Software, Spatial Reference, data files and encoding.

Field	Description
ACQUISITION_ID	Unique ID of the acquisition
ACQUISITION_DATE	Date and time when the acquisition was performed.
ACQUISITION_TYPE	The type of study applied to the object, such as: RTI, Microprofilometry, Low-res scanning, etc
LOCATION_NAME	Geographical location of where the acquisition was performed. Example: Verona, Sardinia, etc.
OPERATOR_NAME	The name of the person who conducted the acquisition.
SETUP_DESCRIPTION	Description of the particularities of the setup.
INSTRUMENT_SPECIFICATIONS	Pointer to the file containing the technical specifications of the instrument.
PROBE_NAME	The name of the probe.
PROBE_DIRECTION	The positioning of the probe, relative to the sample – vertical or horizontal.
PROBE_FREQUENCY	The frequency of the CCD sensor inside the probe. Measured in Hertz.
PROBE_POWER	The executing power of the probe.
PROBE_FINEPOWER	The maximum possible power of the probe.
PROBE_COARSEPOWER	The worst possible power of the probe.
LENS_NAME	The name of the lens.
LENS_SERIAL_NUMBER	An identifier of the type of lens.
LENS_PART_NUMBER	An identifier of the type of lens.
LENS_MIN	The minimum distance from the lens, after which the surface of the object becomes optically optimal.
LENS_MAX	The maximum distance from the lens, after which the surface of the object becomes optically optimal.
XAXIS_SPEED	The speed of the scanning probe when moving on the x-axis (mm/s)
YAXIS_SPEED	The speed of the scanning probe when moving on the y-axis.
LASER_POWER	The power of the light source.
LASER_X_SPOT_SIZE	Measured full width at half maximum (FWHM) at standoff position
DRIVER	The software programme used for interaction with the device.
ACQUISITION_X_ORIGIN	The x-axis origin of the coordinate system at the start of the measurement.
ACQUISITION_Y_ORIGIN	The y-axis origin of the coordinate system at the start of the measurement.
PIXEL_SIZE	The physical distance between the pixel centers in x and y directions, measured in milimeters. Also referred to as the stage step between the signal triggers for acquisition.
RANGE_X	The coverage of the measured area scanned by the microprofilometer on the x-axis. The unit of measure Is milimeters.
RANGE_Y	The coverage of the measured area scanned by the microprofilometer on the y-axis. The unit of measure Is milimeters.
WIDTH	The number of measured points on the x-axis.
HEIGHT	The number of measured points on the y-axis.
OFFSET	The distance (in milimeters) chosen by the operator with the purpose of giving the instrument time for stabilization and for reducing any possible vibration effect. The distance is traversed by the instrument at the beginning and the end of each line of measurements.
ORIGIN	XYZ coordinates of the origin in the reference 3D space
ORIENTATION	Unit vectors giving the orientation of the device in the reference 3D space
SNR_FILE	A binary file that contains a 1D array of signal to noise ratio metric, computed for every measurement. The file can be later reshaped into a 2D array according to the the number of rows and columns of the acquisition. The snr file name is typically AcquisitionID with .snr estension
SNR_FILE_FORMAT	Data type format used for encoding the signal-to-noise-ratio values.
TAG_FILE	A binary file that contains a 1D array of values storing the indexes of successful measurements. The file can be later reshaped into a 2D array according to the number of rows and columns of the acquisition. The tag file name is identical to the Acquisition ID, with .tag as extension.
TAG_FILE_FORMAT	A data type format used for encoding the indexes of successful measurements.
DIST_FILE	A binary file that contains a 2D array of the raw distance values, corrected for missing values.The mask file name is identical to the Acquisition ID, with .mpf extension.
DIST_FILE_FORMAT	The file format use to store the depth matrix. Typically raw 32 bit float, but also tiff, xyz point cloud.
MASK_FILE	A binary image which masks out the measurements that fall out of the optimal range of the lens or that have a corresponding SNR value Lower than 500.
MASK_FILE_FORMAT	The file format used to store the binary mask information

Table 5: Metadata fields for low-resolution 3D scanning. The fields grouped into wrappers are color-coded: Setup, Hardware, Software, Spatial Reference, data files and encoding.

Field	Description
ACQUISITION__ID	Unique ID of the acquisition
ACQUISITION_DATE	Date and time when the acquisition was performed.
ACQUISITION_TYPE	The type of study applied to the object, such as: RTI, Microprofilometry, Low-res scanning
LOCATION_NAME	Geographical location of where the acquisition was performed. Example: Verona, Sardinia
OPERATOR_ID	The unique identifier of the operator who performed the acquisition.
OPERATOR_NAME	The name of the who conducted the acquisition.
SETUP_DESCRIPTION	Description of the particularities of the setup.
SCANNING_DEVICE	The scanning device used for capturing the object and the cloud of points corresponding to the object's geometry.
RESOLUTION	The resolution of the scanning device.
ACCURACY_VALUE	A numeric value that indicates the accuracy of the 3D scanning procedure.
TOOL_VERSION	A numeric value that represents the tool's version used for scanning.
ORIGIN	XYZ coordinates of the origin in the reference space
ORIENTATION	Unit vectors of the acquisition space in the reference space
OUTPUT_FILE	3D scan file name
OUTPUT_FILE_FORMAT	Mesh/Point cloud format

4. Joint multimodal fusion procedures for surface analysis

The Scan4Reco surface analysis protocols aim at capturing and processing data related to the superficial layer of a sample/mock-up or an artwork. The signals used to analyze the surface are the data coming from RTI and microprofilometric acquisition components.

RTI and microprofilometric acquisition techniques have been already described in the deliverable D3.4, and a procedure for global 3D scanning for patch referencing is provided here in Annex I.

However, after the capture and before the actual processing and analysis tasks, a series of procedures have to be put in place in order to register those signals and convert them in a single, fused, manageable representation.

Those procedures to associate spatial referencing are different whether we perform LAB acquisitions of samples or ONSITE artwork acquisitions. In the first case, we capture planar samples of small size and the microprofilometer gives the reference coordinate system with no need of further rough scan. In the second, RTI and microprofilometry data are aligned to a reference frame defined by a global 3D scan; that available holistic rough 3D data will be used to guide the positioning system, and to align further, different planar patch scans. Note that several patch studies can be done for each object/sample, both differing for location or any other acquisition parameter (e.g. different wavelengths in RTI, different lenses in MP).

In this context, we first need a procedure to align the RTI image stack to the microprofilometry data, which come from samples (LAB) or from small planar patches on the surface geometry of an artwork (ONSITE). In addition, in the case of the ONSITE environment, we also need a procedure to register RTI and/or microprofilometry data of patches onto the global reference frame acquired by the global scanning architecture components (HW/VISDEPTH,SW/SCAN); the latter gives the possibility of aligning novel, future RTI or microprofilometry scans on the global reference system.

In the definition of our joint registration procedures, we rely on the information coming from the general annotation/storage protocol already defined. For each object at each aging stage we have a series of annotations, aging information and metadata, and we use a subset of them to assist and speed up the fusion procedure. The result of the registration pipeline is additional information stored in our organized multimodal archive which is useful for the successive processing and analyses. The input metadata for RTI are the intrinsic parameters of the camera sensor, light geometry and intensity calibration data, and the pixel coordinate of the cropped image regions used to build the Appearance Profile stack. Similar metadata comes from the microprofilometry data format (e.g., probe step size, range map dimensions, etc.). The main scope of the joint registration procedures is to produce a series additional metadata (e.g., extrinsic parameters of sensors) that express the 2D and 3D data

of RTI, microprofilometry and global scan in a consistent, unique reference frame. For the organized archiving and communication, resulting metadata of each acquisition/processing are stored as indicated in Section 3.3. The RTI/microprofilometry alignment protocol will define the origin and orientation of the RTI camera sensor with respect to the microprofilometry depth range map, while the registration of RTI (or microprofilometry) to the global 3D holistic scan will provide the position and orientation of the acquired small planar patches onto the global geometry.

4.1 RTI vs microprofilometry alignment protocol

The goal of the fusion protocol to register the RTI and microprofilometry data is to compute the position and orientation of RTI camera sensor relative to the reference depth map from microprofilometer. It is based on an image-to-geometry registration procedure that performs a Mutual Information analysis between the two signals. This alignment protocol is useful both in the LAB environment capture and in ONSITE acquisitions. In the first case, we don't have any other signal involved in the surface acquisition and characterization, so that the final reference frame for further computations will be the microprofilometer geometry. In the ONSITE environment, the fused data from RTI and microprofilometry will produce a first representation of a small planar local patch that will be aligned afterwards to the global 3D scan.

As mentioned before, the input information for this fusion procedure is the microprofilometry depth map and some data/metadata from RTI, i.e., intrinsic parameters of the camera sensor and an image with a view point equal to the RTI data (e.g., normal map). Since the RTI image stack has been captured from a fixed viewpoint and varying lighting conditions, it is sufficient to align just one image to the 3D geometry in order to register all the information in the stack onto the sample or planar patch geometry. The image used for the registration could be one of the images in the stack, or, more likely, an image computed by an image stack processing routine (e.g., the normal map field).

The basic registration process consists in two steps: a rough manual alignment and an automatic refinement.

The manual registration aims at building the first link between the 2D information of RTI image stack and the 3D data from the microprofilometer depth map. Since the intrinsic camera parameters has been already computed for the RTI data in a previous calibration step, the user now has to select a small set of 2D-3D correspondences (at least three, but likely more for a better, more robust first estimation). A graphical user interface (see Figure 3) will provide the user with two images, one from the RTI and the other from microprofilometry data, and he can select corresponding pixels based on the shape of the sample or the patch in the artwork surface. The set of correspondences will be processed to obtain an initial estimation of the extrinsic parameters, which will be prone to errors due to the manual nature of the task.

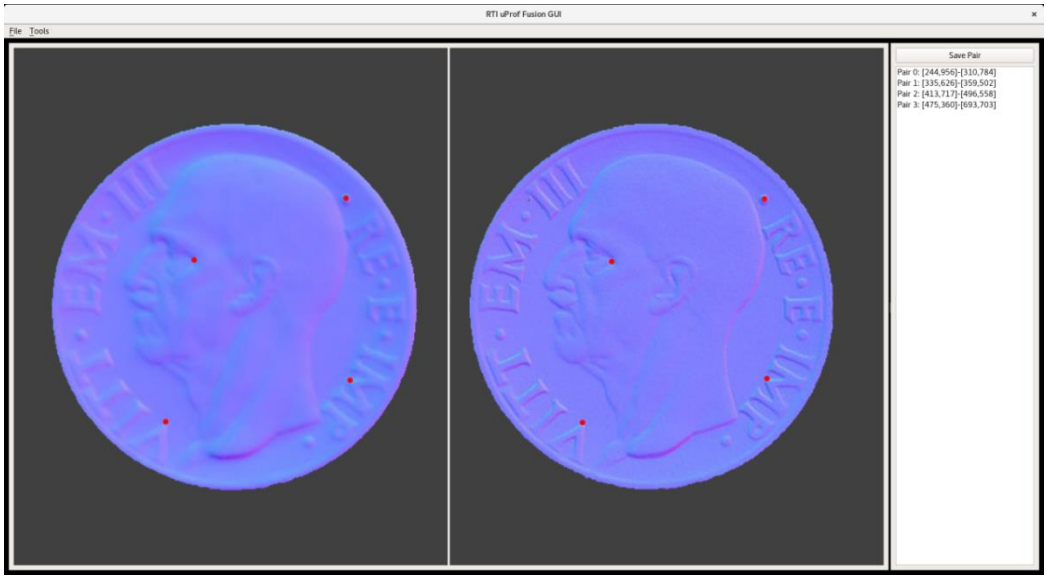


Figure 3: Graphical user interface to manually provide correspondences between signals from RTI (left) and microprofilometer (right). In this example the chosen signal has been the normal map field, and four correspondences have been selected.

The automatic refinement of the registration is based on the Mutual Information paradigm. An optimization framework will search for the optimal set of values in a region of the space of extrinsic parameters close to the initial rough estimation. For each possible extrinsic parameters candidate it will project the microprofilometry data into the RTI camera sensor view point. It then computes a per-pixel correlation of the RTI and microprofilometry signals, in order to obtain a cost function evaluation. The scope of the optimization routine is to find a set of extrinsic parameters that minimize that cost function. A crucial point in Mutual Information approaches is the choice of signals being compared. In our case, a possible (but not unique or mandatory), useful choice is the normal map field, which is available from both RTI and microprofilometry data. In the first case it can be computed by using classical Photometric Stereo algorithms, while, for the microprofilometer, it can be computed by applying finite difference methods for partial derivatives computation across the discrete depth field. In the case of normal field a possible per-pixel cost function is a simple dot product between corresponding normal. Figure 4 shows some angular errors at convergence in the case of some samples or some very small testing objects (ancient coins).

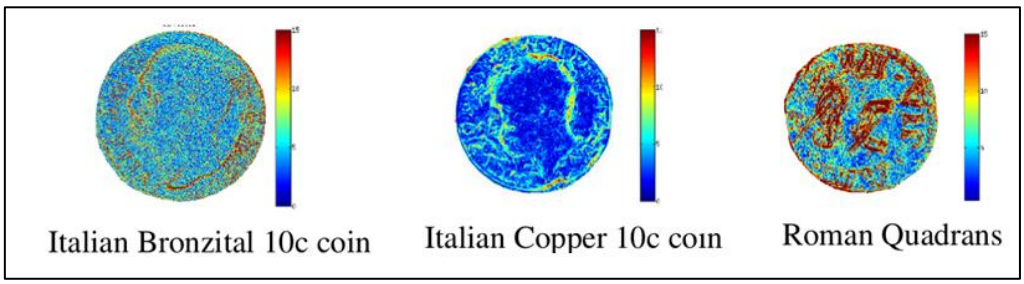


Figure 4: Angular errors. Color-coded angular errors (degree) of RTI estimated normals wrt microprofilometry.

4.2 Protocol for registration of RTI and microprofilometry to global geometry

The protocol for the registration of the RTI (or microprofilometry) data onto the global 3D geometry applies only in the case of ONSITE acquisitions of artworks. As stated by the end-user requirements (EUR/MO/02, EUR/PA/02 and EUR/PA/03), the main goal of this protocol is to map local/punctual measurements onto a global 3D proxy, by finding the position and

relative orientation of a small, local, planar patch with respect to the global, holistic 3D scan of an artwork. It should be noted that this location can be approximate, as the coarse 3D scan serves only as a spatial reference to locate positions on measurements (GIS-like approach).

The main source of this information comes from the component devoted to probe positioning (SW/PROBPOS), which will give the position of the sensor (RTI or microprofilometry), and the corresponding viewing direction, which will provide registration information at acquisition time. As a fallback procedure, in case acquisition-time registration is not available, a post-hoc registration will be performed. The two approaches are described in the following subsections. In both cases, the registration information will be stored as metadata in the RTI/microprofilometry archived information, using the protocol described in this document.

4.2.1 Acquisition-time registration

The SW/PROBPOS module will in general use the mechanical positioning system (e.g., through encoders), or passive/active techniques based on computer vision in order to convey that kind of information.

In fact, the positioning system will place the patch based sensor (RTI or microprofilometry) in front of the region of interest based on the 3D model. Moreover, it will be assumed in that all patch-based sensor systems will be calibrated with known shape and sensor position in their intrinsic reference frames. This allows for the computation of the local plane in the global 3D geometry, and it allows for the registration of this plane to the average plane of microprofilometer data. The RTI to microprofilometry alignment in section 4.1 will be the bridge with which RTI will be aligned to the global 3D.

It should be noted that if the approach of pure mechanical registration results impractical, another way to register measures taken with local sensors (e.g., microprofilometer or RTI) and global 3D geometry data could be the exploitation within SW/PROBPOS of external cameras or depth sensors that record the entire acquisition setup, i.e., both objects and sensors, and the use of manually selected or automatically detected landmarks. In this way, possibly using also markers to facilitate the accurate landmark detection, it should possible to recover the joint positioning of the microprofilometer and the landmarks of the 3D shape. The correspondence of the landmarks allows the reconstruction of the transformation linking the microprofilometer and the 3D scan reference systems (see Figure 5).

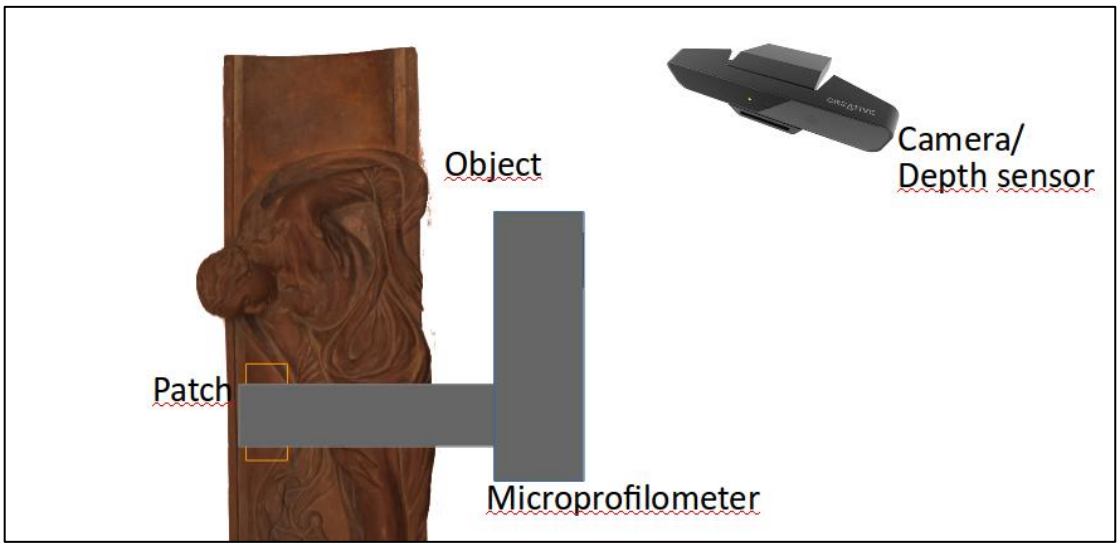


Figure 5: Using external camera/sensor to provide reference 3D space alignment metadata for microprofilometer studies. A similar approach can be used also with RTI acquisition.

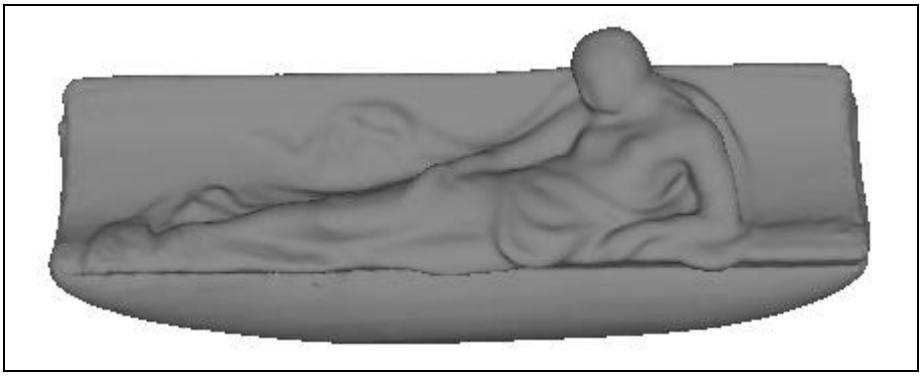
4.2.2 Post-hoc registration

If the information from the SW/PROBPOS positioning system will not be available, a manual alignment approach to register data should be provided. This will allow the user to manually annotate landmarks (e.g., samples corners) on the global rough geometry and the patch data from RTI or microprofilometry. Those correspondences will be used both to link the patch based signal to the proper position on the artwork surface, and to align the orientation of the patch to the local plane approximation of the surface. Such a procedure could be more effective if the low-resolution scan is also enriched with a color signal, which will make it simpler to visually detect correspondences/measurement position in the low resolution 3D scan.

4.2.3 Detection of flat and uniform surfaces

In the ONSITE acquisition, the global 3D scan can be used to select optimal patches for local acquisition based on shape and material features. For this reason, an algorithm has been implemented for the automatic detection of flat and uniform surfaces on a 3D model. Concretely, the algorithm extracts fixed-size square patches on the surface of the 3D model with dimension provided as input by the user.

Initially, the algorithm performs planar segmentation on the point cloud of the 3D model using the RANSAC model selection scheme [13]. In brief, random triplets of points from the 3D model are sampled and used to define plane equations. Additional points of the 3D model are added to these triplets if they are considered inliers, namely if their distance from the corresponding plane is below a pre-defined threshold. Once RANSAC concludes, the 3D model has been segmented in several planes and the final step is to examine in each plane whether square patches of the targeted size can be enclosed. This is achieved by random sampling on the points of each planar segment, followed by geometrically testing whether the fixed-size square fits into the planar segment. This test is performed for each sampled point separately by aligning a reference square to the orientation of the planar segment, translating it so as it centre coincides with the sampled point and examining whether all points of the square fall into the planar segment. An example of this procedure is depicted in Figure 6.



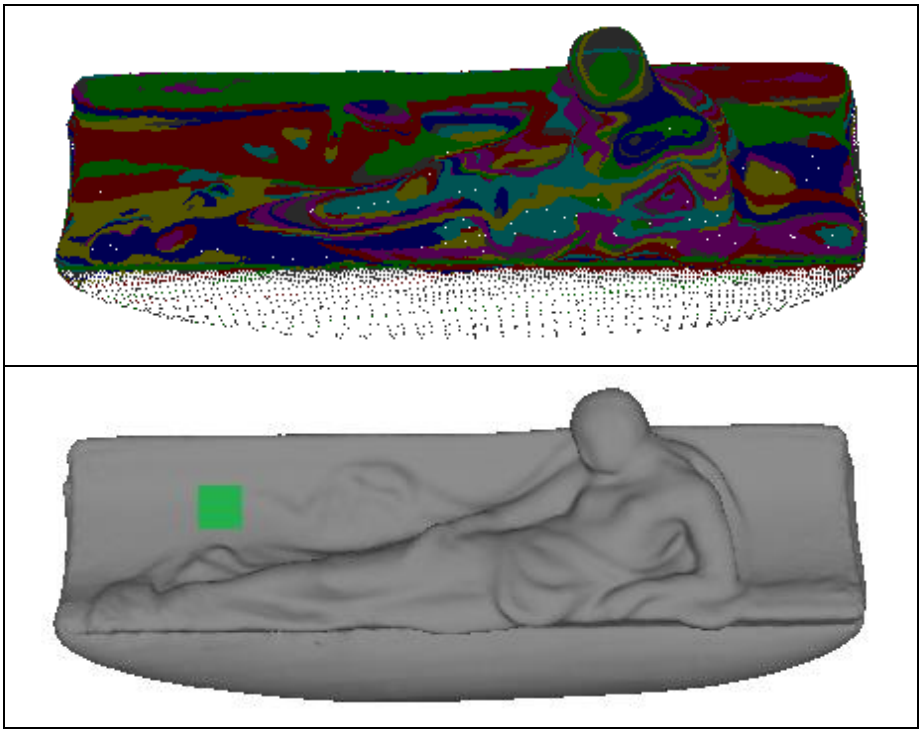


Figure 6: (Top) The reconstructed 3D model of an object. (Middle) After plane segmentation is performed the object is divided in several planes with each color corresponding to a plane. (Bottom) The green square has been computed after the patch detection algorithm concludes.

5. Example

5.1 Project sample studies

Using the joint processing protocol we stored information for the acquired project samples. This means that, for each sample (for the moment at time t0), several patch acquisitions have been realized and annotated with the full metadata organization related to the object, ageing and study levels. microprofilometric point cloud is here also the 3D reference model and has been acquired as described in D3.4. The RTI pipeline has been executed as well on all samples, acquiring multiple images (visible and IR), correcting intensity, estimating Appearance Profiles and cropping them in ROIs including single metallic samples. Finally. annotating in the 2D images the sample corner positions, as camera intrinsic are known, it is immediately possible to recover and annotate the camera optic centre position and the axis orientation (that here should be approximately vertical) in the reference system. As in the examples shown before, we can then match more accurately the samples information minimizing the normal surface difference. This allows the use of features coming from the different measurements to derive richer descriptors for ageing modelling as well as for studying materials reflectance.

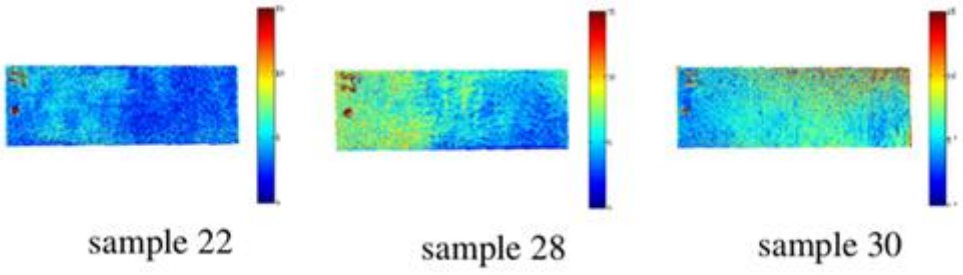


Figure 7: Angular differences estimated from spatial registration of Scan4Reco silver samples after patch alignment.

6. Conclusions

We have described the protocol developed for the joint use of microprofilometry and Reflectance Transformation Imaging data spatially referenced to a low resolution 3D model acquired with the scanner in order to characterize patches of artworks and samples.

While in this project the purpose of the proposed protocol is to provide the basic support for the activities planned in Task 5.2 Digital modelling of material appearance and features and 5.5 Degradation Identification and automatic spatiotemporal Annotation, the underlying concepts can lead to a proposal for a novel generic way to support any kind of studies of artworks or material samples based on local multimodal patch based acquisitions.

References

[1] Remondino, Fabio. "Heritage recording and 3D modeling with photogrammetry and 3D scanning." *Remote Sensing* 3.6 (2011): 1104-1138.

[2] Pamart, Anthony, et al. "Toward a Multimodal Photogrammetric Acquisition and Processing Methodology for Monitoring Conservation and Restoration Studies." (2016).

[3] DICOM official site www.nema.org

[4] Ronzino, P., S. Hermon, and F. Niccolucci. "A metadata schema for cultural heritage documentation." *V., CApellini (ed.), Electronic Imaging & the Visual Arts: EVA* (2012): 36-41.

[5] Stasinopoulou, Thomais, Lina Bountouri, Constantia Kakali, Irene Lourdi, Christos Papatheodorou, Martin Doerr, and Manolis Gergatsoulis. "Ontology-based metadata integration in the cultural heritage domain." In *International Conference on Asian Digital Libraries*, pp. 165-175. Springer Berlin Heidelberg, 2007.

[6] D'Andrea, Andrea, and Kate Fernie. "CARARE 2.0: a metadata schema for 3D Cultural Objects." In *Digital Heritage International Congress (DigitalHeritage), 2013*, vol. 2, pp. 137-143. IEEE, 2013.

[7] Schröttner, Martin, Sven Havemann, Maria Theodoridou, Martin Doerr, and Dieter W. Fellner. "A generic approach for generating cultural heritage metadata." In *Euro-Mediterranean Conference*, pp. 231-240. Springer Berlin Heidelberg, 2012.

[8] Pan, Xueming, Philipp Beckmann, Sven Havemann, Katerina Tzompanaki, Martin Doerr, and Dieter W. Fellner. "A Distributed Object Repository for Cultural Heritage." In *VAST*, vol. 2010, pp. 105-114. 2010.

[9] Ronzino, Paola, Nicola Amico, and Franco Niccolucci. "Assessment and comparison of metadata schemas for architectural heritage." *Proc. of CIPA* (2011).

[10] Doerr, Martin, and Maria Theodoridou. "CRMdig: A Generic Digital Provenance Model for Scientific Observation." In *TaPP*. 2011.

[11] Denard, Hugh. "A new introduction to the London Charter." *Paradata and Transparency in Virtual Heritage Digital Research in the Arts and Humanities Series(Ashgate, 2012)* (2012): 57-71.

[12] Doerr, Martin, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, and Herbert van de Sompel. "The europeana data model (edm)." In *World Library and Information Congress: 76th IFLA general conference and assembly*, pp. 10-15. 2010.

[13] Schröttner, Martin, Sven Havemann, Maria Theodoridou, Martin Doerr, and Dieter W. Fellner. "A generic approach for generating cultural heritage metadata." In *Euro-Mediterranean Conference*, pp. 231-240. Springer Berlin Heidelberg, 2012.

[14] Szeliski, Richard. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[15] Scan4Reco Deliverable D2.2-System Architecture Definition.

[16] Motorized Rotary Stage [Zaber X-RST120AK-DE50-KX13AF] - (*http://www.zaber.com/products/product_detail.php?detail=X-RST120AK-DE50*)

[17] Intel® RealSense™ Camera F200 - (*http://www.intel.com/content/www/us/en/support/emerging-technologies/intel-realsense-technology/intel-realsense-cameras/intel-realsense-camera-f200.html*)

[18] Khoshelham, Kourosh. *Accuracy analysis of Kinect depth data*. In ISPRS workshop laser scanning. Vol 38, no 5, 2011.

[19] Ed Olson. *AprilTag: A robust and flexible visual fiducial system*, Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), 2011

[20] Kazhdan, Michael, and Hugues Hoppe. *Screened poisson surface reconstruction.* ACM Transactions on Graphics (TOG) 32.3 (2013): 29.

[21] Callieri, Marco, et al. *Processing sampled 3D data: reconstruction and visualization technologies.* Taylor and Francis, 2011.

[22] Edward Angel and Dave Shreiner. *Interactive Computer Graphics: A Top-Down Approach with Shader-Based OpenGL* (6th ed.). Addison-Wesley Publishing Company, 2011.

[23] Allène, Cédric, Jean-Philippe Pons, and Renaud Keriven. *Seamless image-based texture atlases using multi-band blending.*In 19th IEEE International Conference on Pattern Recognition, 2008.

Annex 1: Global 3D scanning procedure

In this annex, the low resolution 3D scanning module is described. The 3D scanning module is responsible for the global coarse-resolution 3D representation of the color and geometry of an artwork (i.e. metallic objects, paintings) using a depth sensor. The corresponding hardware and software modules of the Scan4Reco are the following

- HW/VISDEPTH (depth densor)
- SW/SCAN (3D scanning module)

as illustrated in Figure 8 and described in section 5 of Deliverable 2.2 [15].

The extracted 3D representation by the 3D scanning module is used as the basis of the local measurements performed by the RTI and micro-profilometer hardware probes described in section 4.1.

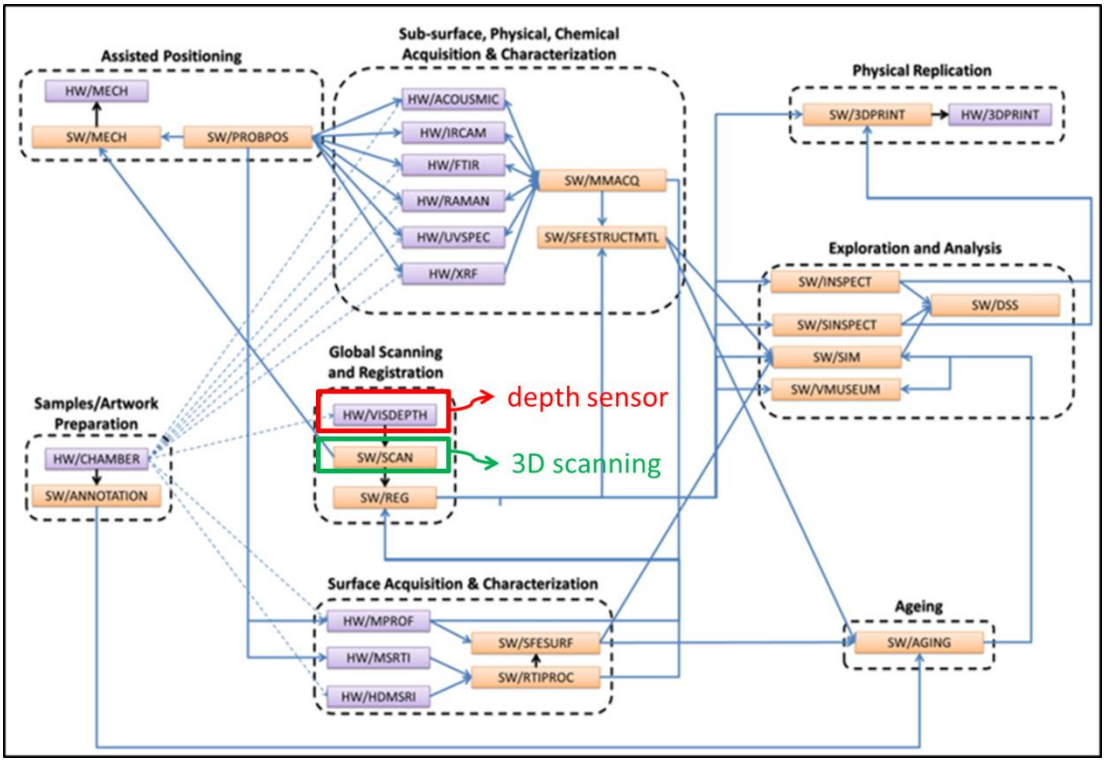


Figure 8: Scan4Reco Architecture Diagram. The depth sensor and the 3D scanning module used for a global 3D reconstruction of an artwork correspond to the HW/VISDEPTH and SW/SCAN modules of the Scan4Reco architecture.

An overview of the algorithm for 3D Scanning is illustrated in Figure 9. Initially, the artwork object (i.e. metallic objects, paintings) is recorded by a low-cost depth sensor for capturing color and depth information, from multiple views using a rotary stage. Then, the depth maps are processed so as to discard areas based on normal information and/or Sobel filtering. All the recorded and processed views are transformed in the same coordinate system using the kinematics of the rotary stage. Then, voxelization is performed on the accumulated point clouds in order to keep points with a high confidence factor and discard the rest. Poisson surface reconstruction is applied on the remaining points to extract the mesh of the object. Finally, texture mapping is performed on the meshed 3D model in order to reconstruct the texture of the artwork object.

The main contributions of our method are

- Development of a 3D reconstruction pipeline from multiple views by incorporating feedback from motion encoders

- A method for the alignment of the depth sensor's coordinate system with the rotary stage
- Improvement of depth information utilizing morphological and edge detection filters and surface normals information
- Confidence factor based voxelization of multiple aligned views

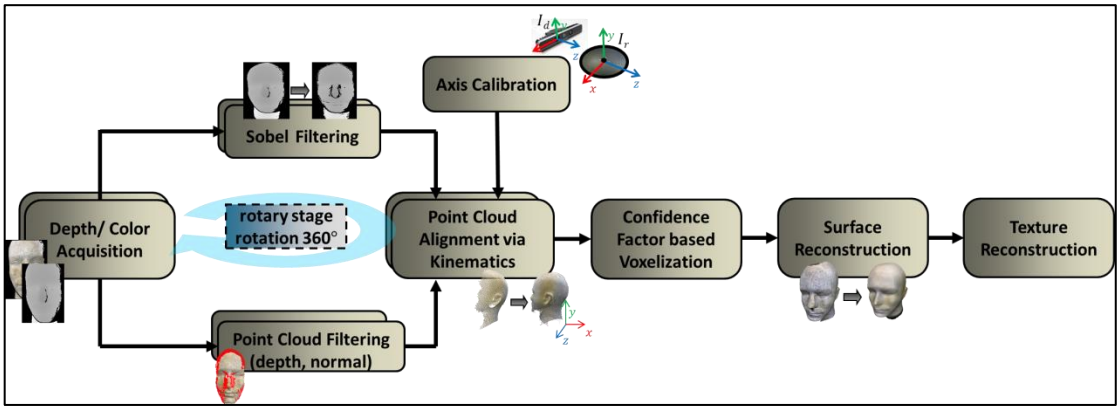


Figure 9: 3D Scanning Overview: The artwork is recorded by a depth sensor from multiple views using a rotary stage. The recorded point clouds are processed so as to discard areas based on Sobel filtering and normal information and, then, are aligned using the calibration of the depth sensor and rotary stage. Finally, voxelization, 3D surface and texture reconstruction are performed on the accumulated point clouds to generate the 3D model.

Depth/Color Acquisition

As illustrated in Figure 10, an artwork is rotated (360 degree rotation) by a controlled rotary stage [16] and is recorded by a depth sensor [17] to produce a series of depth and color maps. Depth map ($\boldsymbol{D}$) contains the estimated depth information in mm for each (x, y) pixel of the image. For each estimated depth value, the corresponding color map provides the registered color information in RGB format. In Figure 11, some recorded depth and color maps of a replica statue for various angles of the rotary stage are given.

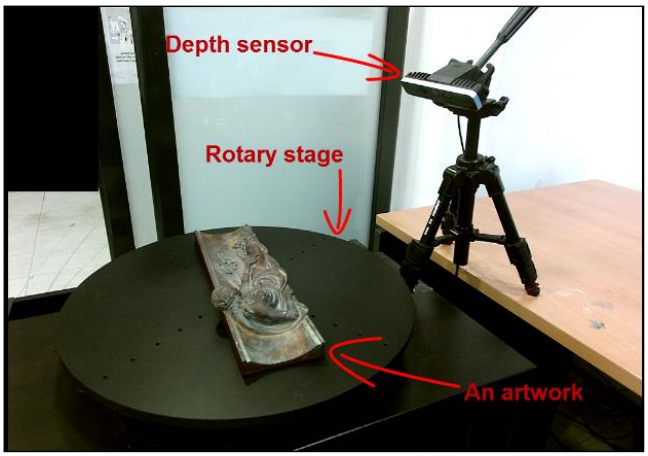


Figure 10: 3D Scanning Setup: an artwork is rotated by a controlled rotary stage [16] and is recorded by a depth sensor [17].

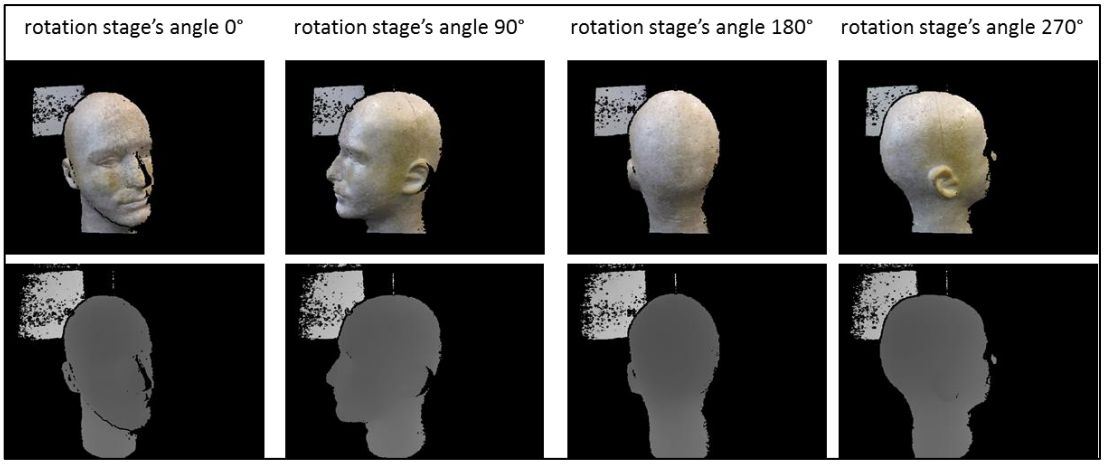


Figure 11: Recorded depth and color maps of a replica statue for various rotation angles of the rotary stage.

Depth & Point Cloud Processing

Depth estimation from such a low-cost camera is usually inaccurate in areas where depth changes abruptly. This fact affects the alignment and surface reconstruction negatively [18]. Such areas can be detected by computing the gradient $\boldsymbol{G}(x, y)$ at each pixel (x, y) in the depth image $\boldsymbol{D}$. An approximation of $\boldsymbol{G}(x, y)$ can be obtained as follows

$$\widehat{\boldsymbol{G}}(x, y) = \frac{1}{2}\widehat{\boldsymbol{G}}_h(x, y) + \frac{1}{2}\widehat{\boldsymbol{G}}_v(x, y)$$

by applying a Sobel edge detection filter on the depth image $\boldsymbol{D}$. $\widehat{\boldsymbol{G}}_h(x, y)$ and $\widehat{\boldsymbol{G}}_v(x, y)$ are the horizontal and vertical gradient approximations respectively. If $\widehat{\boldsymbol{G}}(x, y)$ exceeds a threshold t_h, then $\boldsymbol{D}(x, y)$ is discarded. An example of this procedure is given in Figure 12.

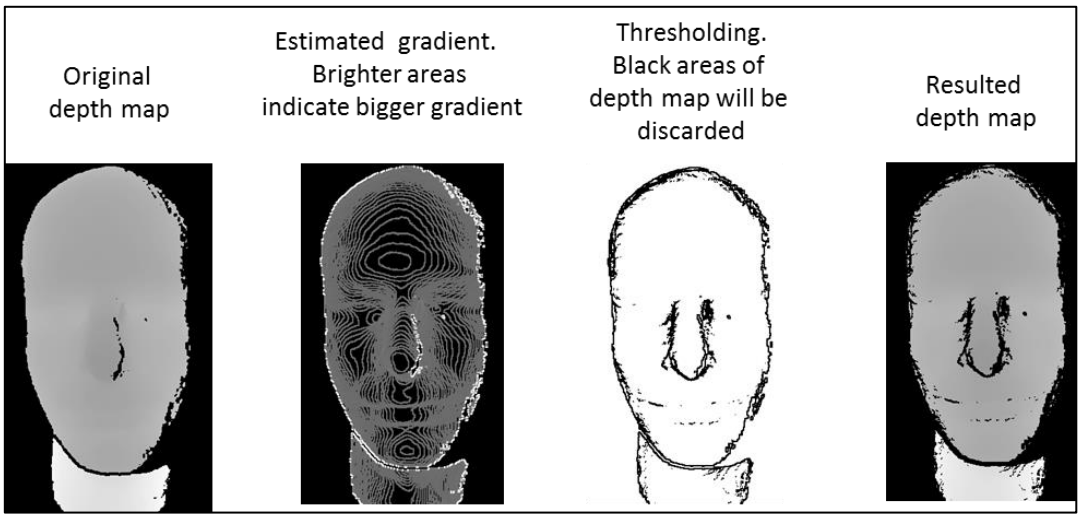


Figure 12: An example of applying the Sobel filter on a depth map.

Another way of removal of inaccurate areas is the usage of surface normal information. In particular, for each 3D point $\boldsymbol{p}_i$ corresponding to depth pixel $\boldsymbol{D}(x, y)$, its normal vector $\boldsymbol{n}_i$ is estimated, by estimating the normal of a plane tangent to the surface of the k (e.g. 30) nearest neighbors of $\boldsymbol{p}_i$. Firstly, the covariance matrix $\boldsymbol{C}$ of the k nearest neighbors of $\boldsymbol{p}_i$ is calculated

$$\boldsymbol{C} = \frac{1}{k}\sum_{l=1}^{k}(\boldsymbol{p}_l - \overline{\boldsymbol{p}}) \cdot (\boldsymbol{p}_l - \overline{\boldsymbol{p}})^T,$$

where $\overline{\boldsymbol{p}}$ representes the 3D centroid of the k nearest neighbors of $\boldsymbol{p}_i$. After the eigenanalysis of $\boldsymbol{C}$, the eigenvector corresponding to the smaller eigenvalue is the normal $\boldsymbol{n}_i$, whereas the orientation of $\boldsymbol{n}_i$ is defined in relation to the depth sensor's position.

A 3D point $\boldsymbol{p}_i$ is removed if the angle $\varphi = \cos^{-1}\frac{\boldsymbol{r}\cdot\boldsymbol{n}_i}{|\boldsymbol{r}||\boldsymbol{n}_i|}$ between the sensor ray direction ($\boldsymbol{r}$) and the normal $\boldsymbol{n_i}$ to the point $\boldsymbol{p}_i$ is bigger than a threshold t_{hn}. For example, in Figure 13, the red points have angle φ bigger than $40°$.

Figure 13: An example of removal of inaccurate regions using normals: the angle between the depth sensor ray direction and normal vector is bigger than 40° in the red regions.

Point Cloud Alignment

Each new point cloud $\boldsymbol{P}_\theta$, corresponding to a rotation of the rotary stage by an angle θ, is transformed in a common coordinate system as defined in the first recording and is added to the overall point cloud $\boldsymbol{P}_{all}$ that holds all the aligned point clouds. More specifically, the point $\boldsymbol{p}_i \in \boldsymbol{P}_\theta$ is transformed as follows

$$\boldsymbol{p'}_i = \boldsymbol{R}\,\boldsymbol{p}_i + \boldsymbol{t}$$

where the rotation and translation matrices are calculated as described in following section. An example of 45 consecutive aligned point clouds in a common coordinate system is illustrated in Figure 14.

Figure 14: The final point cloud consisting of 45 consecutive point clouds aligned in the same coordinate system.

Axis Calibration

The coordinate systems of the depth sensor (I_d) and the rotary stage (I_r) are illustrated in Figure 15. When the rotary stage rotates by an angle θ, the total translation and rotation correspond to the following:

$$\boldsymbol{t} = \boldsymbol{p}_c - \boldsymbol{R}_{dr}\boldsymbol{R}_\theta\boldsymbol{R}_{dr}{}'\boldsymbol{p}_c\,,$$
$$\boldsymbol{R} = \boldsymbol{R}_{dr}\boldsymbol{R}_\theta\boldsymbol{R}_{dr}{}',$$

where $\boldsymbol{p}_c$ is the center of the rotary stage, $\boldsymbol{R}_{dr}$ is the rotation matrix between the coordinate systems of the depth sensor and the rotary stage as shown in Figure 15. In particular the coordinate system for the sensor is defined by its principal axis whereas for the rotary stage

it is defined by the normal vector ($\boldsymbol{n}$) and the centre $\boldsymbol{p}_c$. $\boldsymbol{R}_\theta$ is the rotation matrix of the rotary stage on Y-axis returned from the kinematics of the stage and defined as:

$$\boldsymbol{R}_\theta = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}.$$

It should also be noted that the normal vector ($\boldsymbol{n}$) and the center ($\boldsymbol{p}_c$) of the stage are estimated following a calibration procedure.

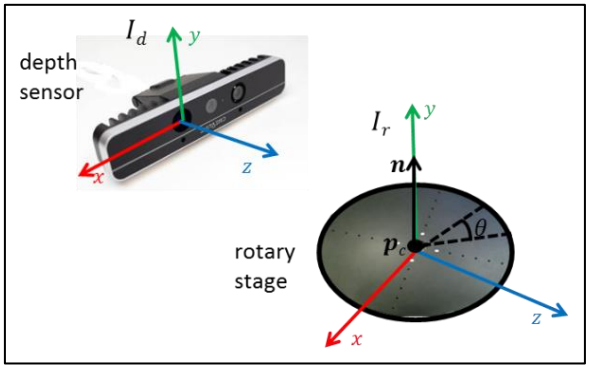


Figure 15: The coordinate systems of the depth sensor and rotary stage. $\mathbf{p_c}$ and $\mathbf{n}$ denote the center and the normal vector of the rotary stage.

Concretely the stage normal is computed as a vector orthogonal to the point cloud of the stage using least squares while the center is extracted either manually (i.e. set by the user) or following a point tracking procedure. In the point tracking procedure, 3D points $\boldsymbol{T_i}, i \in \{1,2,\dots,n\}$ on the rotary stage are tracked using specific tags [19] rotating the stage n times with the same known angle θ, as illustrated in Figure 16. The center $\boldsymbol{p}_c$ of the stage can be estimated by solving the following system:

$$\begin{aligned} < \overrightarrow{\boldsymbol{T_1 p_c}}, \overrightarrow{\boldsymbol{T_1 T_2}} > &= \left\| \overrightarrow{\boldsymbol{T_1 p_c}} \right\| \left\| \overrightarrow{\boldsymbol{T_1 T_2}} \right\| \cos\theta \\ < \overrightarrow{\boldsymbol{T_2 p_c}}, \overrightarrow{\boldsymbol{T_2 T_3}} > &= \left\| \overrightarrow{\boldsymbol{T_2 p_c}} \right\| \left\| \overrightarrow{\boldsymbol{T_2 T_3}} \right\| \cos\theta \\ &\dots \\ < \overrightarrow{\boldsymbol{T_{n-1} p_c}}, \overrightarrow{\boldsymbol{T_{n-1} T_n}} > &= \left\| \overrightarrow{\boldsymbol{T_{n-1} p_c}} \right\| \left\| \overrightarrow{\boldsymbol{T_{n-1} T_n}} \right\| \cos\theta. \end{aligned}$$

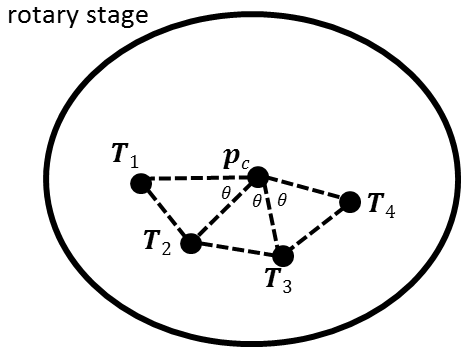


Figure 16: 3D points ($T_1, T_2, ..$) are tracked on the rotary stage. Here, p_c is the center of the stage and θ is the rotation angle.

Confidence Factor based Voxelization

After alignment, the resulting point cloud is dense and noisy, containing many points from different views (point clouds) for the same region. We create a 3D voxel grid over the accumulated point clouds and, then, in each voxel, we keep the points corresponding to the view (point cloud) with the higher confidence factor (f_v) and discarding the rest. The confidence factor ($f_v{}^i$) of the view (point cloud) $\boldsymbol{P_i}$ is defined as

$$f_v{}^i = \frac{1}{N}\sum_{l=1}^{N} f_c{}^{pl},$$

where N is the total number of the points which belongs to the $\boldsymbol{P_i}$ in the voxel and $f_c{}^{pl}$ is the confidence factor of the point $\boldsymbol{p}_l$:

$$f_c{}^{pl} = \cos\theta,$$

where θ is the angle of sensor ray direction ($\boldsymbol{r}$) and normal $\boldsymbol{n}_l$ of $\boldsymbol{p}_l$. An example is illustrated in Figure 17.

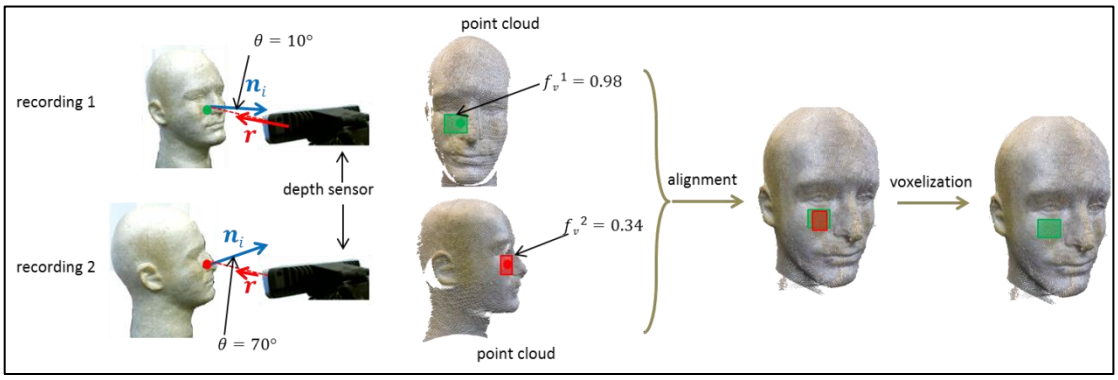


Figure 17: A voxelization example using a confidence factor for each point: The statue has been recorded by a depth sensor (recordings 1 & 2). The red and green regions correspond to the same region. Using the angles between the sensor ray direction (r) and the normals $\mathbf{n_i}$ of the points $\mathbf{p_i}$ belonging to the red and green regions, the two respective confidence factor ($\mathbf{f_v}^1$ & $\mathbf{f_v}^2$) values are calculated. After voxelization, the points of the green region (recording 1) are only kept since $\mathbf{f_v}^1 > \mathbf{f_v}^2$.

The reconstruction accuracy before and after voxelization of a bronze artwork and a replica statue are illustrated in Figure 18 and Figure 19, respectively. Various parts are computed using the Euclidean distance before and after filtering using the confidence factor based voxelization method and compared to the real values. As can be seen, the accuracy is improved by performing voxelization on the aligned point clouds.

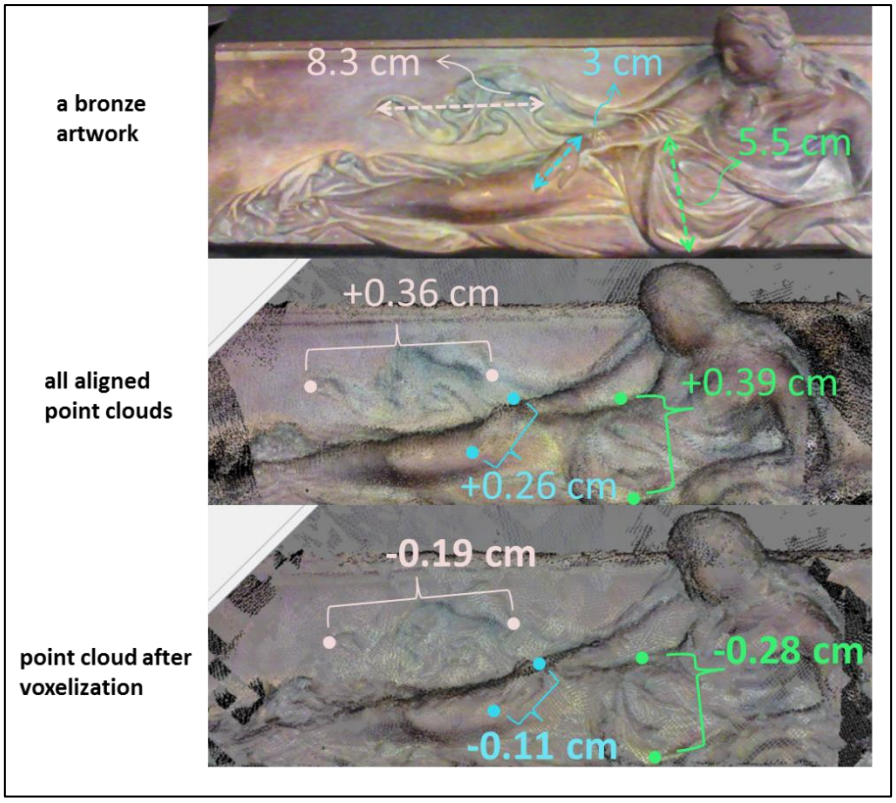


Figure 18: The Euclidean distance between hips and length of a cloth's part and hand of a bronze artwork are computed before (middle) and after (bottom) filtering and compared to the real values (top).

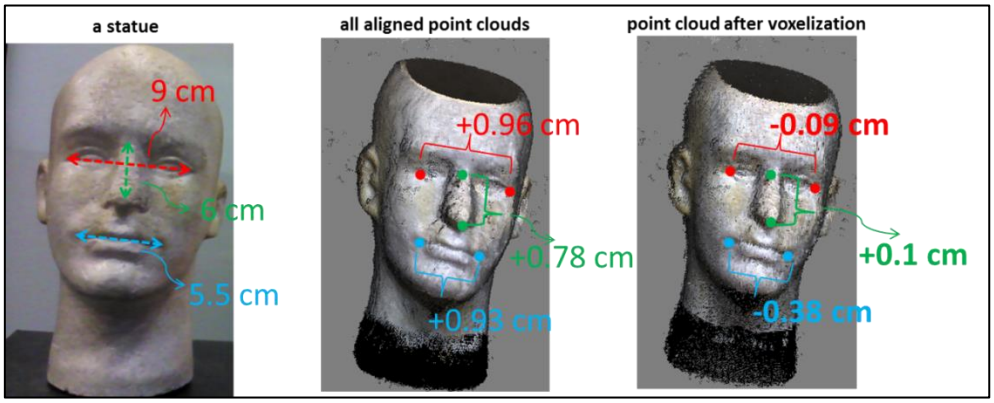


Figure 19: The Euclidean distance between eyes and the length of nose and mouth of a replica statue are computed before (middle) and after (right) filtering and compared to the real values (left).

Poisson Surface Reconstruction

For reconstructing 3D surfaces from the accumulated point clouds, the Poisson Surface Reconstruction (PSR) algorithm [20] has been incorporated in our 3D scanning procedure. PSR's goal is to compute a 3D indicator function χ so as to extract the appropriate surface, as depicted in Figure 20. PSR is based on the observation that the oriented point samples ($\vec{V}$) can be viewed as samples of the gradient ($\nabla\chi$) of the indicator function χ, (Figure 20). Thus, given a set of oriented point samples ($\vec{V}$), PSR finds the indicator function χ whose gradients best match the samples ($\vec{V}$) by minimizing:

$$\min_{\chi}|\nabla\chi - \vec{V}|.$$

Applying the divergence operator, the minimum is obtained by solving the Poisson problem:

$$\Delta\chi \equiv \nabla \cdot \nabla\chi = \nabla \cdot \vec{V}\,.$$

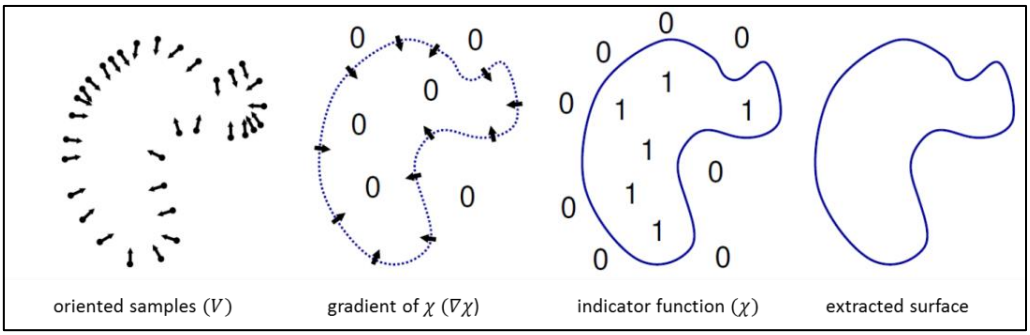


Figure 20: Poisson surface reconstruction example in 2D [20].

PSR algorithm builds a watertight surface and is robust to noise caused by sampling inaccuracy and point cloud misalignment. Figure 21 illustrates the 3D reconstruction of 45 consecutive aligned point clouds (see Figure 14) using the PSR algorithm.

Figure 21: 3D surface reconstruction of a point cloud consisting of 45 consecutive aligned point clouds (see Figure 14) using the Poisson Surface Reconstruction algorithm [20].

Texture Reconstruction

In this step, texture mapping [21] from a set of color images on the 3D reconstructed model is performed. In particular, having a set of N color images $C = \{c_1, c_2, \dots, c_N,\}$ captured from various views with a known relative transformation, the goal is to find for each face f_i of the 3D model, a texture region located in a single color image where f_i is more clearly visible. Initially, back-face culling [22] is performed, in order to define for each f_i the subset of camera views where the face is visible. Namely, if we denote as $\boldsymbol{n}_{f_i}$ the surface normal of face f_i and $\boldsymbol{r}_{c_j}$ the principal axis of camera view c_j, we examine whether f_i is visible in c_j by checking that the sign of their inner product is non-negative, i.e.

$$\boldsymbol{n}_{f_i} \cdot \boldsymbol{r}_{c_j} \geq 0.$$

From the extracted subset of camera views, for each face, we select the view that maximizes $\boldsymbol{n}_{f_i} \cdot \boldsymbol{r}_{c_j}$ assuming that this view is the one where the face is more clearly visible (see Figure 22). The rationale of this step is that the maximum of the inner product corresponds to the view where the image plane of the camera is closer to being parallel to the face and consequently more accurate color information is obtained (i.e. perspective distortion is minimized in the texture-to-surface mapping and the projection's area is maximized [23]). An example of this texture reconstruction algorithm is illustrated in Figure 23, while the color images used in the texture mapping are given in Figure 24.

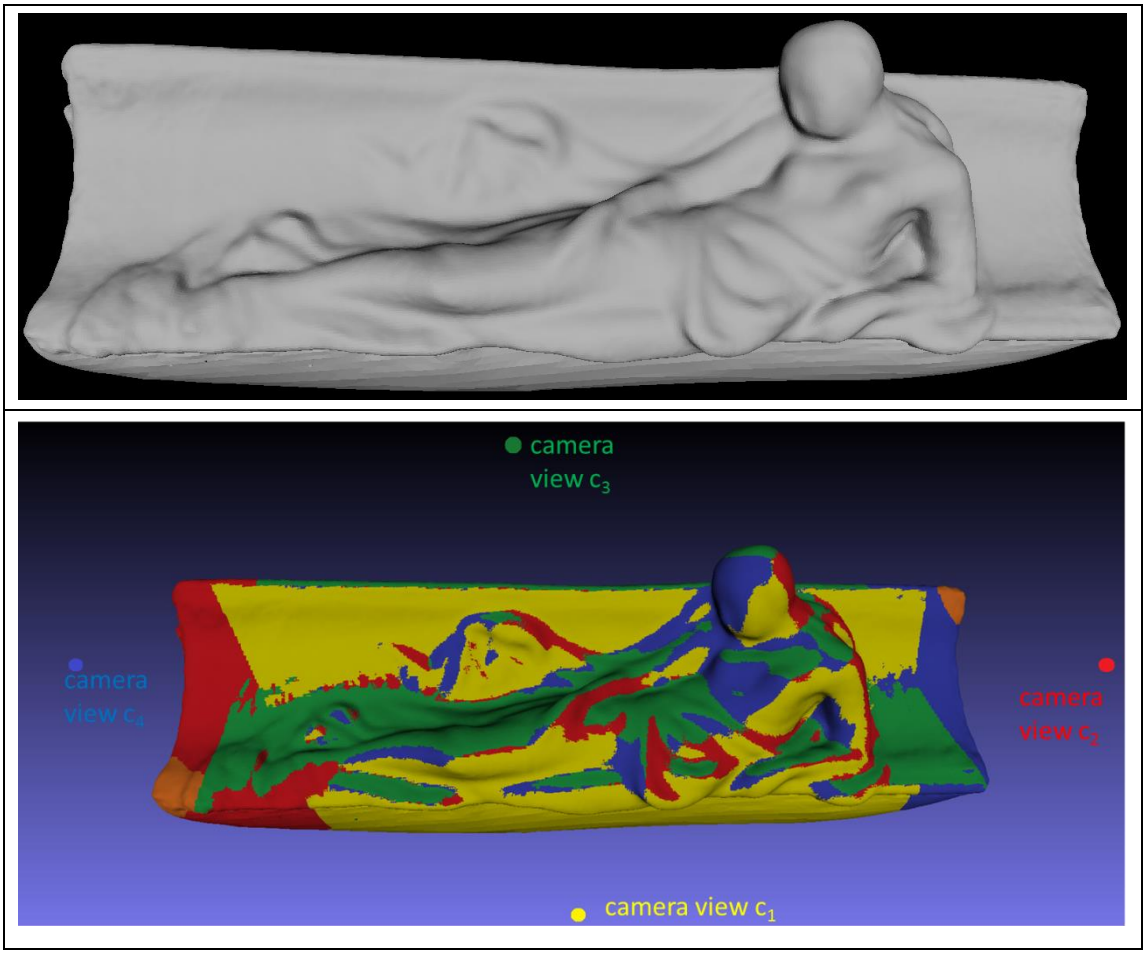


Figure 22: (Top) A 3D reconstructed model. (Bottom) For each camera view c_i, we see the corresponding faces that are more clearly visible. Visibility is measured using the inner product of the face normal and the camera principal vector.

Figure 23: Texture reconstruction example of the 3D model of Figure 22: On the top we see the colored point cloud without texture mapping whereas on the bottom after the described algorithm concludes. For comparison purposes, the object is depicted on the bottom.

Figure 24: The color images used in the texture reconstruction example of Figure 23.