# Web-based Exploration of Annotated Multi-Layered Relightable Image Models

Alberto Jaspe-Villanueva, CRS4, Italy
Moonisa Ahsan, CRS4, Italy
Ruggero Pintus, CRS4, Italy
Andrea Giachetti, University of Verona, Italy
Fabio Marton, CRS4, Italy
Enrico Gobbetti, CRS4, Italy

We introduce a novel approach for exploring image-based shape and material models registered with structured descriptive information fused in multi-scale overlays. We represent the objects of interest as a series of registered layers of image-based shape and material data. These layers are represented at different scales, and can come out of a variety of pipelines. These layers can include both RTI representations, and spatially-varying normal and BRDF fields, possibly as a result of fusing multi-spectral data. An overlay image pyramid associates visual annotations to the various scales. The overlay pyramid of each layer is created at data preparation time by either one of the three subsequent methods: (1) by importing it from other pipelines; (2) by creating it with the simple annotation drawing toolkit available within the viewer; (3) with external image editing tools. This makes it easier for the user to seamlessly draw annotations over the region of interest. At run-time, clients can access an annotated multi-layered dataset by a standard web server. Users can explore these datasets on a variety of devices; they range from small mobile devices to large scale displays used in museum installations. On all these aforementioned platforms, JavaScript/WebGL2 clients running in browsers are fully-capable of performing layer selection, interactive relighting, enhanced visualization, and annotation display. We address the problem of clutter by embedding interactive lenses. This focus-and-context-aware (multiple-layer) exploration tool supports exploration of more than one representations in a single view. That allows mixing and matching of presentation modes and annotation display. The capabilities of our approach are demonstrated on a variety of cultural heritage use cases. That involves different kinds of annotated surface and material models.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**] Picture and Image Generation; I.3.7 [**Computer Graphics**] Three-Dimensional Graphics and Realism; I.4.10 [**Image Processing and Computer Vision**] Image Representation

General Terms: Cultural Heritage

Additional Key Words and Phrases: 3D visualization, 2D visualization, reflectance transformation imaging, annotations

**ACM Reference Format:**

Alberto Jaspe-Villanueva, Moonisa Ahsan, Ruggero Pintus, Andrea Giachetti, Fabio Marton, Enrico Gobbetti. Web-based Exploration of Annotated Multi-Layered Image Models *ACM J. Comput. Cult. Herit.* 14, 2, Article 24 (May 2021), 31 pages.
DOI:http://dx.doi.org/10.1145/3430846

Author's Addresses: CRS4, 09134 Cagliari (CA), Italy; email: firstname.lastnamecrs4.it www: http://www.crs4.it/vic/ and University of Verona, Verona (VR), Italy; email: firstname.secondnameunivr.it www: http://www.univr.it/

DOI:http://dx.doi.org/10.1145/3430846

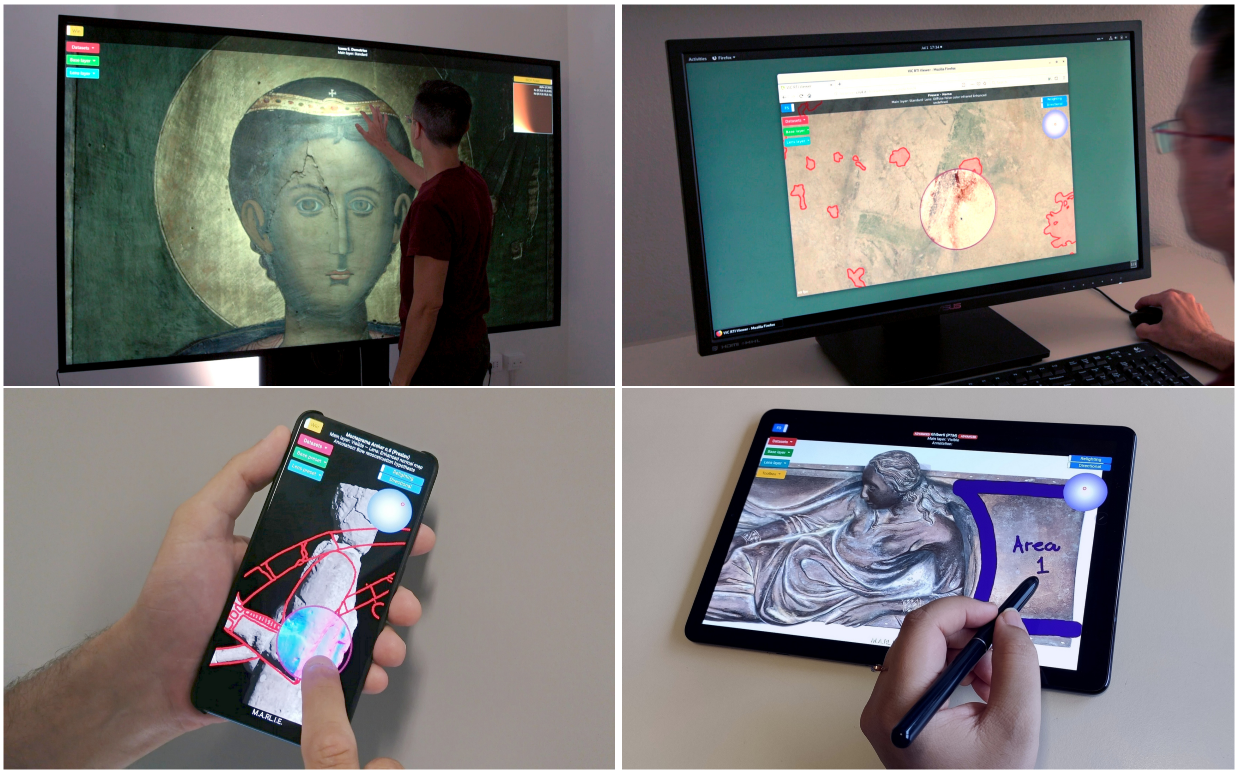



Fig. 1: **Web-based interactive inspection.** Our system allows users to explore annotated heterogeneous datasets coming from a variety of pipelines using a web platform capable to run on several devices. *Top Left*: exploration of a multispectral acquisition of a painting on a large scale display; *Top Right*: analysis of a fragmented statue with annotations in a desktop browser; *Bottom Left*: analysis of an annotated fresco on a mobile phone. *Bottom Right*: analysis and annotation of a bas-relief on a tablet.

## 1. INTRODUCTION

The virtual inspection of digital representations of objects is of fundamental importance in many application fields [Scopigno et al. 2011; Roupé et al. 2016; Pintus et al. 2019]. Particularly in the Cultural Heritage (CH) domain, it supports conservation and preservation strategies by revealing geometric cues or information on the materials used to create an object. Currently, users' interests are shifted towards direct exploration and more interactive methods as compared to the early passive visual presentations in the museums [Falk and Dierking 2000; Kuflik et al. 2011]. In this context, in parallel to generic interactive viewers displaying fully-3D virtual replicas [Potenziani et al. 2015; Sketchfab 2019], relighting interfaces, popularized by Reflectance Transformation Imaging (RTI) viewers [CHI 2019], have emerged as one of the most successful exploration modes. The are easy to acquire, they support meaningful inspection visualization of surface details, and the restriction of camera motion removes one of the main difficulties of 3D exploration applications, hence reducing learning curves [Jankowski and Hachet 2013]. For this reason, interactive inspection of relightable images has been applied to a wide range of items and acquisition scenarios [Malzbender et al. 2001; Krátkỳ et al. 2020; Durr et al. 2014; Pitard et al. 2017; Wainwright et al. 2017], and are highly suitable for both expert and casual users.

Such relighting viewers, however, have so far been restricted to inspecting only few types of models, mainly in the form of raw multi-light image collections or low-frequency analytical relighting representations, such as PTM or HSH [Pintus et al. 2019]. This is due to the fact that, so far, more complex representations, such as spatially-varying normals and BRDFs were too difficult to be reliably extracted from captured data. Recent advanced methods of geometry [Ackermann and Goesele 2015] and appearance [Weinmann and Klein 2015; Dorsey et al. 2010; Guarnera et al. 2016] extraction are increasingly removing some of these limitations. On the other hand, they introduce the need for new tools to support the inspection of image-based representations (e.g., decoupled geometry and appearance parameters).

Moreover, in many use cases, communicating information on real-world objects requires the ability to mix highly photorealistic or illustrative presentations of the object themselves with annotations or semantic labels that provide additional insights on these objects [Ponchio et al. 2020]. These added visual cues makes data understanding easier for the viewer [Vanhulst et al. 2018]. In particular, in order to effectively support a rich and informative experience, the analysis of the shape and/or the material should be complemented with semantic data integration and linking; that added information might be in a textual, visual, abstract or tangible form [Callieri et al. 2013; Jankowski and Hachet 2013]. Much of the extra information requires spatial association, as it describes, or can be related to, different spatial contexts.

For instance, cultural artifacts often exhibit subtle material and shape details, presenting information at multiple scales and levels of abstractions, since both the overall shape and the finest material micro-structure carry valuable information (e.g., on the overall meaning, on the carving process, or on the conservation status). Few relighting viewers support annotations on top of the relightable model, and, in most cases this is restricted to adding separate information through hyperlinks associated to point or area hotspots (see Sec. 2.3).

Our work has the goal of allowing viewers to explore the so-called image-based relightable models (which contain information about surface shape and material) together with additional visual information (organized as registered and multi-scale image overlays). This is achieved through a novel web-based framework for exploring complex image-based object representations on the web.

Off-line content preparation organizes data in a series of registered multiresolution layers of shape and material information. These layers are all image-based, and can come out of a variety of pipelines that produce parametric information, such as RTI representations, or spatially-varying normal and BRDF fields, eventually obtained by fusing multi-spectral data. At authoring time, each layer is associated to an overlay image pyramid, which contains visual annotations to the various scales. Annotations are different at the various resolution levels to provide various levels of abstractions (e.g., with a set of text and drawings associated to the global shape of the object, and a different set to the small-scale feature details). Such a pyramid can be the outcome of other processing steps (e.g., crack detection), or hand-drawn by users. Authoring it is beyond the scope of this paper, which focuses on the exploration framework.

At run-time, the annotated multi-layered datasets are made available to clients using web servers. Users explore the datasets on a variety of devices (from mobile phones to large-scale displays in museum installations) that employ JavaScript/WebGL2 clients. Such clients can perform several tasks on the data, such as layer selection, interactive relighting, enhanced visualization, and annotation display. They can also provide focus-and-context exploration of multiple layers, by using a lens metaphor that provides alternative visual representations for selected regions of interest. Our main specific contributions are the generalization of fixed-view relighting interfaces to several representations outside of pure RTI, the support for multi-scale visual annotations, and the application of advanced focus-and-context visual exploration means to the field of annotated shape and material analysis in CH.

This article is an invited extended version of our contribution to the *2019 Eurographics Workshop on Graphics for Cultural Heritage* [Jaspe Villanueva et al. 2019]. We here provide not only a more thorough exposition and an update on the file format definition, but also significant new material on the image layer generation process, the 2D annotation builder and editing toolbox, and the configurable web viewer. We also provide an additional use-case showing the multi-layered exploration and annotation of a the RTI representation of a relief carving. A source code release of a reference implementation is available from www.crs4.it/vic/download/. The accompanying video illustrates the performance and capability of our approach on four challenging datasets using very different setups and hardware configurations.

## 2. RELATED WORK

Image-based relighting techniques and annotated-model visualization are vast and well-researched subjects, and a complete review of the literature is out of the scope of this paper. We discuss here only a selection of approaches most closely related to ours that can help to better understand the novelty of the proposed solution with respect to the state-of-the-art techniques. In particular, we focus on three main areas: the relighting methods that are based on $2D$-$2.5D$ images; the tools developed to interactively explore such type of data; the annotation strategies employed to enrich the visualization and inspection of such image-based models. For a wider view and coverage of the topic we refer the reader to more general established surveys [Pintus et al. 2019; Mudge et al. 2008; Vanweddingen et al. 2018; Ponchio et al. 2020].

### 2.1 Image-based relighting

Representations suitable for relighting frameworks and visualization tools are computed from a variety of sources, from colored 3D models to multi-light image collections (MLICs). Image-based relighting relies on a pixel-wise representation, which stores for each image location a description that is used at real-time to produce, while panning and zooming, the rendering of areas of the image under a novel, virtual illumination. Per-pixel representations include rearranged/compressed/resampled input data samples acquired by a MLIC capture setup [Macdonald 2015; Ponchio et al. 2018], color layers deriving from data fusion (e.g., [Vandermeulen et al. 2018]), coefficients of a parametric or data-driven reflectance model (e.g., [CHI 2019]), shape (e.g., normal map) and material descriptors (e.g., BRDF or SV-BRDF) data [Gardner et al. 2003]), as well as a variety of derived maps, typically presented as overlays over original images (e.g., [Hameeuw 2015]). In this work, we aim to provide a general multi-layered representation inspection framework which is not limited to pure reflectance fields, but also includes separate shape and material representations. We also emphasize the importance of combining multiple registered representations, and to mix and match them with different display modes for the exploration of models.

### 2.2 Interactive tools

In the last decade, a wide variety of tools for performing visual inspection of image-based data have been presented and, sometimes, deployed for open, public use. The tools target either static exploration of image data (e.g., multi-spectral or stratigraphic data [Moutafidou et al. 2018; Ponto et al. 2009] or multi-light image collections [Vandermeulen et al. 2018; Macdonald 2015]), or dynamic exploration through relighting. The latter is typically done by exploiting specific RTI formats [Pintus et al. 2019]. Our solution extends this approach to a larger variety of cases, including image-based representations that explicitly separate shape and appearance contributions.

While early software tools were designed for desktop use and locally resident data [CHI 2019], recently a lot of web-based solutions have been made available for remote access to relightable models, as well as for use on mobile devices [Palma et al. 2014; Ponchio et al. 2018; Windhager et al. 2018]. Among desktop tools, the most popular in CH applications remains *RTI viewer* [CHI 2019], also due to its coupling with popular acquisition and processing software (*RTI Builder* [CHI 2019]). It provides various ways to visualize/inspect relightable image files both locally or remotely. It employs the classic light direction selector to allow the users to render the studied artwork under novel illuminations. In addition to a photorealistic relighting mode by using Polynomial Texture Mapping (PTM) or Hemi-Spherical Harmonics (HSH) reflectance field models, it also features several non-photorealistic or illustrative enhancements, e.g., diffuse gain, specular enhancement, unsharp masking applied to the normals, the image, the luminance or the model coefficients, static or dynamic multi-lighting, or normal map visualization. While mostly devoted to processing MLIC data, the *APTool* [Giachetti et al. 2019] desktop tool enables a continuous visualization of original data through Radial Basis Functions (RBFs) interpolation. It also provides classic PTM and HSH modes, the display of some feature maps, and the visualization of the per-pixel interpolated reflectance maps. In the field of multi-spectral relighting, *PLDviewer* [KUL 2019], strongly linked to a specific processing framework [Vandermeulen et al. 2018] and a proprietary compressed file, supports photorealistic and illustrative visualizations of various 2D and 3D outputs, e.g., curvature coloring, exaggerated shading, sketching, single pixel reflection maps, and local histogram. Web-based tools for image-based relighting, i.e., WebRTIViewer [Palma et al. 2019], PLDWebviewer, the Digital Materiality Viewer [DHLAB 2019; Fornaro et al. 2017], and Relight [Ponchio et al. 2019; Ponchio et al. 2018], typically support only specific parametric representations of relightable images (in particular, PTM and HSH), and provide interactive relighting and some enhancement capabilities. An important aspect of these tools is their compression strategy used to decrease the data memory footprint for remote access; some rely on classic image compression techniques, while others explicitly exploit the nature of some input data (e.g., MLICs) to perform resampling and compression [Ponchio et al. 2018]. Another important aspect is how the web-tool behaves in a cross-platform scenario. OxRTIViewer [MacDonald et al. 2019] is a web-based viewer devoted only to the PTM visualization. It has been implemented in Scalable Vector Graphics (SVG) with an extremely limited number of dependencies and requirements; it consists of a single HTML file, and no libraries or add-ons are needed. It works across a wide range of browsers and platforms. Rather than studying new ways to blend or fusion different types of visual information through new advanced visualization metaphors, many interactive tools mostly strive for increasing the type of data that they can manage. Some recent tools can handle 3D models, RTI data, large scale images, multi-spectral signals, structure from motion set of points and cameras, together with metadata in a unified platform that allows crow-sourcing annotation and multiple user collaboration [Sophocleous et al. 2017; Pamart et al. 2019]. Of course they can provide interesting authoring tools (see Sec. 2.3), and they are capable of efficiently linking spatial information between different views, different data representations, and textual and visual contents. Nonetheless, the strategies of how this data is conveyed, explored and visualized remain a hot research topic, mostly in the application of such visualization techniques among the various areas of CH investigation.

In typical image viewers, multiple layer presentation is generally supported only by the controlled blending of data coming from multiple modalities. We also support blending of representations, but provide a more general solution by adopting an interactive lens approach [Tominski et al. 2014], whose main purpose is to support multi-faceted data exploration. In our case, we enable display of multiple layers or multiple representations of the same layer, by supporting, on top of a defined visualization mode, an alternative visual representation of a local area of the data, whose size and position is controlled by the user. This method has proved successful for multi-field visualization, but has not been used, so far, in the CH field for relightable image exploration.

### 2.3 Annotations

Visual displays can be categorized into different types based on the relation between the representation and its referent and the complexity of the information represented [Hegarty 2011]. Annotation display falls in the category of visual-spatial displays that dynamically mix the object representation with associated information overlays. Annotations are pivotal components for applications requiring the visual examination of virtual objects of focus. The literature shows evidences of improved usability by using annotations, hence providing end-users an intuitive method to view and interact with the system [Lavrič et al. 2017]. Moreover, user-authored annotations can support analysis and hypothesis process, hence providing key observations with better insights about target object [Zhao et al. 2016]. As noted elsewhere [Ponchio et al. 2020], while this concept has a long story and was managed in the past by the use of 2D technical drawings or thematic maps with an associated legend, nowadays interactive digital instruments open much wider capabilities.

Most of the CH data presentation tools, thus, also support annotations, mostly in the form of integration of interconnected text and model information with bidirectional navigation [Götzelmann et al. 2007; Jankowski and Decker 2012; Callieri et al. 2013; Potenziani et al. 2015]. Some tools are recently emerging that also offer interfaces for multi-user annotation creation [Aioli 2019; Ponchio et al. 2020; Wang et al. 2018; Pamart et al. 2019]. CHER-Ob tool [Wang et al. 2018] allows to visualize both 2D, RTI and 3D models, RGB or multi-spectral color information, and to add annotations as text or images associated to a particular 2D or 3D point of the object under study. It includes an automatic video generation tool that transform all the information contained in the project into a video that summarizes its content. It also provides an easy to use framework for the collaboration of multiple researchers over time. Recently, another web-based visualization solution has been proposed that only focuses on RTI inspection, interaction and annotation, i.e., OxRTIViewer [MacDonald et al. 2019]. This tool provides options for multiple annotations and drawing layers in a crowd-sourcing, collaborating framework, devoted both to the study and interpretations of ancient epigraphy from researchers and scholars, and to direct public engagement. Another interesting framework is Medici2 [Sophocleous et al. 2017]. This provides a web-based environment that is compliant with both 3D models and RTI data, together with the interface to produce and store metadata, and to enable provenance support. Some of these distributed solutions are scalable, flexible, and robust, but they mostly concentrate on the data, lacking a focus on the way the information is conveyed from a visualization perspective. While in these previous works much of the effort has been put on the development of techniques to create, represent, and maintain annotations, in this work we focus more on how to effectively explore and visualize an image-based representation of such data. In particular, we do not place 3D annotations on objects but on the image plane, constructing multi-scale annotations layers. This leads to some limitations, since the annotations carried out on a representation are transferable from one layer to another of the same multi-layered dataset, but they are not automatically transferable to another view of the same object. On the other hand, we also achieve more flexibility, as we expand over present approaches by supporting the authoring and integration of free-form overlays at multiple scales. Free-form annotations are not forced to be just 3D ones, but can be, e.g., sketches of possible reconstructions of missing parts. Note, also, that we only focus on the renderable representation, and, thus, other kinds of annotations (e.g., typical region-based ones [Ponchio et al. 2020]) might be imported by transferring them to image layers.

Annotation techniques, typically, have difficulty in handling the case in which many annotations are present at the same time: by displaying them on top of the object clutter is created, while by not displaying them on top of the object the spatial relation is lost [Balsa Rodriguez et al. 2016]. Clutter is not necessarily linked to the number of items, but to a state in which the items cause confusion in the user so they negatively affect performance [Camba et al. 2014]. Special care should thus be taken to decide when and how an information is presented. This is typically handled by defining annotation categories and explicitly enabling/disabling the visible categories in the interface [Potenziani et al. 2015]. We also support this technique, but enhancing it by selectively enabling or disabling annotations within or outside interactive lenses. This makes it possible to explore an uncluttered model, while moving a lens to selectively discover annotations. Conversely, it allows to explore a fully annotated model, while moving a lens to show uncluttered areas.

## 3. OVERVIEW

Even though the proposed framework is of general use, its creation has been motivated and inspired by a large series of use cases and collaborative efforts in the CH field. From interactions with domain experts, in particular conservators and museum curators, we derived a list of requirements to guide our development. Additional requirements stem from our own past experience of developing interactive systems for cultural heritage [Bettio et al. 2009; Marton et al. 2012; Balsa Rodriguez et al. 2013; Balsa Rodriguez et al. 2016]. The final list of requirements, summarized in Table I, as well as our analysis of related work presented in Sec. 2, were taken as guidelines for our development process, which resulted in the definition of a client-server approach based on a multi-layered representation of data. The overall architecture is depicted in Fig. 2. The mapping between features and requirements is summarized in Table II.

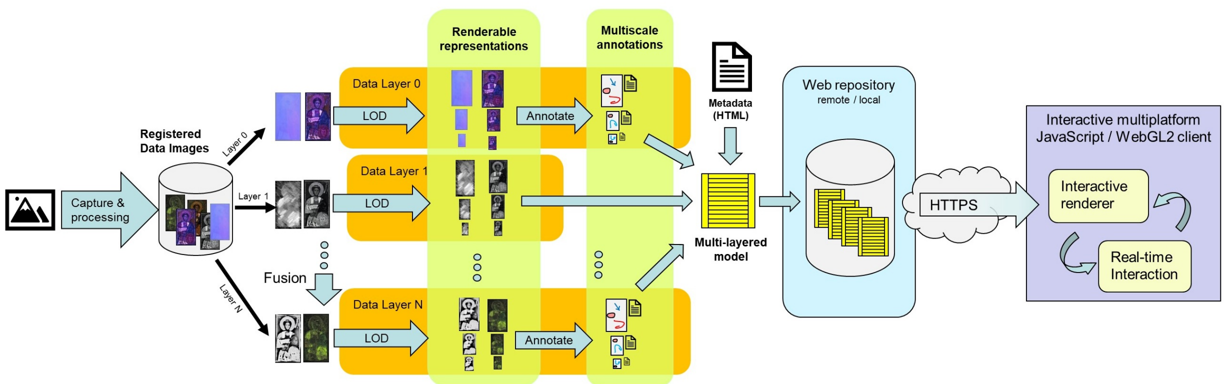



Fig. 2: **Architecture overview.** Off-line data preparation phases assemble different data representations into an image-based multi-layered dataset composed of shape and material layers coupled with per-scale annotation overlays and additional metadata. The dataset is made available on the net using a standard web server. A JavaScript/WebGL2 client loads data and interactively supports layer selection, interactive relighting and enhanced visualization, annotation display, and focus-and-context multiple-layer exploration using a lens metaphor.

In order to handle a large variety of use-cases (R11), we designed our system on a web-based platform, with data preparation done off-line and resulting in a repository made available by standard web server to a web client running in a browser on top of WebGL2, a JavaScript API for rendering interactive 3D and 2D graphics within any compatible web browser without the use of plug-ins. WebGL2 does so by introducing an API that closely conforms to OpenGL ES 3.0 that can be used in HTML5 `<canvas>` elements. Since the API enjoys a wide support and exploits hardware acceleration on several platforms, we can support at interactive rates for both remote exploration on mobile clients and local exploration on museum setups, using a full-screen interface (R12). Off-line content preparation organizes data

| | |
|---|---|
| R1 | **Inspection of geometric features at multiple scales.** CH items have a high heterogeneity and present important geometric details at various scales. It should be possible to analyze shape both in conjunction and separately from material, at the micro/meso/macro level. |
| R2 | **Multi-spectral appearance.** The understanding of the material behaviour is of fundamental importance to grasp the highest possible amount of insights for CH item interpretation and conservation. The visualization tool must enable the inspection not only of the visible signal, but also of invisible (e.g., from ultraviolet or infrared light) surface response. Similarly, it is important for CH scholars to visualize and analyze more complex phenomena such as object induced fluorescence. |
| R3 | **Interactive relighting.** It should be possible to select a virtual light direction and to relight the captured object under that novel illumination, in order to understand how the object locally interacts with the light. This should be interactive to get motion parallax. |
| R4 | **Information registered with models.** Descriptive information, textual or visual, should be associated to specific parts of the presented images, since it is generally referring to a region of the cultural object. |
| R5 | **Information abstraction.** Different macro-structural and micro-structural views should be associated with different kinds of information (e.g., descriptions of carvings and decorations, reconstruction hypotheses, comparisons with other objects, which all refer to different scales of the objects). |
| R6 | **Information authoring.** Adding textual and visual information (drawings, images) should be easily supported and should be possible for museum curators and archaeologists without particular training, without requiring intervention of specialized personnel. |
| R7 | **Information layers.** Each object can be acquired and seen from different perspectives in terms of measured data or combination of different acquired signals. For each CH item, a visualization tool must provide the user with the possibility to choose a series of different layers/representations of the data (e.g., visual data, ultraviolet, infrared, normal map, fluorescence, or combinations of these modes). |
| R8 | **Stratigraphy, focus and context.** It is often necessary to analyze multiple layers coming from different modalities to support interpretation. Switching among layers to show, e.g., a stratigraphy of the object is a required feature. However, toggling a single-layer views makes it difficult to understand correlations between separate layers. It is therefore necessary to support the flexible combination of at least two layers at a time. |
| R9 | **Focus on cultural object.** For both the user and the experts, the important information is the visualized object itself, which should therefore not be obstructed by interaction widgets. |
| R10 | **Fast learning curve.** CH domain experts, which do not have a scientific background, have to use visualization and analysis tools with the minimum effort. Moreover, software already exist that are widespread among the CH community. New solutions must take this aspect into account, in order to be compliant with standard interfaces in CH daily work or to add new features that require the smallest and easiest learning process. |
| R11 | **User interface and platform flexibility.** The possible use cases of cultural object analysis are various, ranging from conservation analysis by domain experts to museum exhibits. The tool should be configurable, not constrained a particular hardware, and usable with the largest variety of setups (local/remote) and devices, ranging from common desktops to mobile devices (tablets, smartphones), and large, possibly touch, display for museum or general collaborative applications. |
| R12 | **Seamless interactive exploration.** All the visualization and analysis tasks must be interactive and smooth. The relighting must be continuous and at interactive rate, and all the other inspection operations must give the end-user a smooth and real-time experience. |

Table I. : **Requirements.** List of requirements derived from experts, analysis of related work, and past experience.

in a series of registered multiresolution layers of shape and material information combined with manually added overlay information. For maximum flexibility and to support a simple interface (R10), we assume that all data layers are image-based and co-registered. Focusing on such a purely image-based representation is a design decision that makes it possible to cover a wide variety of use-cases, from the inspection of inherently 2D objects such as paintings to the analysis preferential views of 3D objects, simplifying both data acquisition and inspection, as we can leverage on widely used acquisition pipelines and GUIs for image viewers. Object measurements can come out of a variety of pipelines producing relightable shape and material information, either from processing capture data, or by off-line data-fusion operation that merge layers together (R7). Shape and material data are, at this level, represented in several forms, either explicitly separating shape and material (e.g., normal+BRDF), or fusing them in a reflectance field (e.g., PTM or HSH) (R1, R2)

Annotations are generated in the form of per-layer per-level-of-detail overlay images corresponding to levels of abstraction (R5) of registered information (R4). The final multi-layered dataset is, thus, a list of layers, each composed of a combination of several multiresolution data channels and an optional multiresolution annotation layer. These are further enriched with other additional metadata (such as textual descriptions). Details on formats are further provided in Sec. 4.

In this work, we focus mostly on the exploration of the novel possibilities offered by image-based visualization of multiple, possibly annotated, layers, leaving as future work the implementation of interaction with the individual annotations (see Sec. 8). For exploration, each annotation layer is simply encoded in a pyramid of semi-transparent images registered with the data layers. These layers can be created by procedural means (e.g., edge detectors), standard image editing tools, or simple drawing tools made available within our viewer (R6). A processing script takes care of transforming these data into publishable format. Sec. 5 provides details on data preparation.

At run-time, each annotated multi-layered dataset is made available to clients using a standard web server. A JavaScript/WebGL2 client queries the server to receive metadata as well as image information for each of the layers. The data images are interpreted as parametric fields, depending on the particular data representation associated to the layer (e.g., normal+BRDF or PTM). The user interface is oriented towards panning, zooming, and relighting operations, using the same controls as in standard RTI viewers to reduce the need for training (R3,R9,R10). In order to support stratigraphic analysis and multi-faceted visualization, we make it possible to select the main layer of interest, using a lens metaphor to provide a different visual representations coming out of a second layer in an interactively defined region of interest. Using a lens instead of just supporting layer switching allows us to use the main layer as context during the exploration of the secondary layer, which acts as a focus region. Moreover, clutter is reduced by forcing the display of annotations only inside or outside the lens area (R8). The display of overlay information is synchronized with the zoom level of the current active layers (R8). Further details on viewer design and behavior are provided in Sec. 6.

| | |
|---|---|
| • Web-based design with standard protocols | R11 |
| • Use of hardware acceleration | R12 |
| • Image-based data layers | R10, R7 |
| • Support for both separated and fused data models | R1, R2 |
| • Annotations per-layer, per-level-of-detail, registered overlay images | R5, R4 |
| • Information synchronized with zoom level | R8 |
| • Standard drawing tools for authoring annotations | R6 |
| • Support for common CH relightable acquisition pipelines | R3, R9, R10 |
| • Lens metaphor for focus-and-context with other layers | R8 |

Table II. : Summary of proposed features to fulfil the requirements in table I.

## 4. DATA REPRESENTATION

Our datasets are represented as a set of *layers* associated to general metadata. Each layer is defined as a standalone *image-based renderable representation* (Sec. 4.1) coupled to an *annotation overlay* (Sec. 4.2), and a hypertext description.

### 4.1 Image-based renderable representations

We currently support two main types of *image-based renderable representation*: *shape and material representations*, and *relightable images*. The kind of representation is specified by a *tag*, and each representation has different sets of images defining it, with each image stored in a mipmapped pyramid format.

Shape and material representations explicitly decouple geometry and appearance, represented in images as regularly sampled spatially-varying fields. For geometry, we support *depth images* and *normal maps*, while for materials we currently support a Lambertian + isotropic Ward BRDF representation [Ngan et al. 2005], with coefficients stored in separate image maps: two RGB images for the $K_d$ and $K_s$ components and a 1-component image for the $\alpha$ specular lobe parameter. Both shape and material component are optional in the definition of a layer. When material is missing, the shape component is implicitly associated to a default constant BRDF in the viewer, making it possible to define a pure geometric layer. When shape is missing, the viewer only displays the diffuse color without shading, allowing for the definition of pure image layers. This sort of general representation can come out of a variety of acquisition and processing pipelines. These include single-image capture [Li et al. 2018], multi-light image capture [Hui and Sankaranarayanan 2017], and combinations of scanning and reflectance field capture [Palma et al. 2013].

Relightable images representations, instead, do not separate shape from material. These representations naturally come out of common RTI processing pipelines [CHI 2019; Giachetti et al. 2018]. In our first implementation, we support the common LRGB PTM format, in which a single biquadratic polynomial is used to determine a luminance value that is then used to modulate a companion RGB texture.

Note that, in all cases, we decided to only support RGB layers. Multi-spectral support is provided by the use of multiple layers. A signal of $N$ spectral bands can be organized in $N$ different monochromatic layers (grayscale images), by assigning to each layer a single wavelength. Another way to exploit multi-spectral signals is to prepare a single or multiple layers by re-arranging the $N$ input spectral bands into false-color representations. This is a common best practice in multi-spectral data fusion and manipulation [Zheng et al. 2018], also largely employed in the daily work of scholars in the Cultural Heritage field [Lerma et al. 2011; Salerno et al. 2014]. This type of approach enables the use of advanced global data fusion methods, which are preferable to pixel-wise, local techniques supportable within the real-time constraints of the image viewer.

### 4.2 Annotation overlays

Annotation overlays provide additional information to layers in the form of drawings and hyper-texts associated to specific hierarchical levels of the image-based renderable representations. The images, which can be either rasterized or scalable vector drawings, are either the results of processing steps (e.g., crack identification) or, more frequently, created by users with standard image editors. In the authoring process, each annotation overlay is created in association with a specific zoom level of the corresponding renderable layer, and is supposed to be rendered at the same resolution. This possibility to link a single layer to multiple annotation overlay allows for several levels of abstraction in conveying CH content related to a particular object.

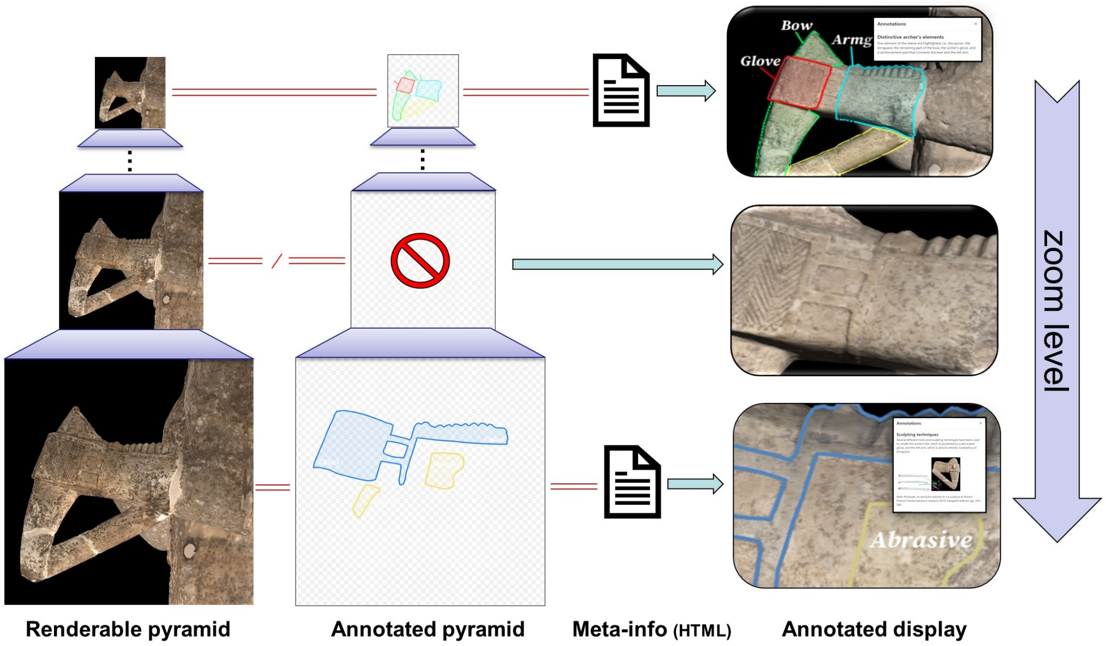



Fig. 3: **Multi-level Annotations.** Each annotation overlay is related to a mipmap level of the *Renderable model*, making it possible to define scale-related levels of abstraction. Not all mipmap levels are required to have an annotation. At run-time, the zoom level selected for display activates the associated annotation. The image is displayed over the renderable layer, while the text dynamically updates a hypertext display widget.

### 4.3 Multi-layered data assembly

The full multi-layered data representation is assembled in a single description necessary to define a scene to initialize the viewer. All the elements, composed of scene-level data layers, are declared as JavaScript JSON structures, with the their own properties and values. They are assembled into a comprehensive scene structure, which is parsed by the viewer at initialization time. In this way, the viewer can setup its interface, initialize its internal data, and download the required information. Similar to other popular solutions [Potenziani et al. 2015], we chose to use JSON over XML because it is a human-readable format simple to understand, write and parse. An example of multi-layered dataset is the following:

```
{
  "name": "Sample dataset",
  "info": "<h1> ... </h1> ... ",
  "width": 2969, "height": 5469,
  "layers": [
    {
      "id": "VIS", "name": "Visible",
      "obj": {
        "shader": {
          "type": "LRGB_PTM",
          "scale": [2,2,2,2,2,2],
          "bias": [164,166,101,87,78,0]
        },
        "maps": {
          "coeff_0_1_2": "L100.png",
          "coeff_3_4_5": "L101.png",
          "coeff_6_7_8": "L102.png"
        }
      },
      "overlay": {
        img: "uv_plus/overlay.gif",
        range: [0,3],
        shortInfo: [
          "Short annotation level 0",
          "Short annotation level 1", ...
        ],
        longInfo: [
```

```
          "Long annotation level 0",
          "Long annotation level 1", ...
        ]
      }
    },
    {
      id: "UV1", "name": "UV(385nm) + Fluorescence",
      "obj": {
        "shader": {
          "type": "SVBRDF",
          "glossLimits": [0.01, 0.5],
          "inputColorSpace": "linear"
        }
        "maps": {
          "kd": "vis/kdMap.jpg",
          "ks": "vis/ksMap.jpg",
          "gloss": "vis/glossMap.jpg",
          "N": "shared/N.jpg"
        }
      },
      "overlay": {
        img: "vis/overlay.png", range: [2,5]
      }
    }
  ] ...
}
```

An important feature of the scene description is that layers are allowed to share image resources. Rather than to explicitly declare the sharing as in scene graphs, we decided to implement the sharing implicitly by resource URI matching (see Sec. 6). In the above example, for instance, the normal map is shared between the two layers. Moreover, for the renderable representations we assume that the full pyramid is available, while, for the overlays, only some levels of detail, within a specified range, may be available on the server. Note that the specified image filenames are templates, and the actual file names stored on the server and queried by the client are a combination of the base name of the image, a two digit level of detail, and the extension.

## 5. DATA PREPARATION

Data preparation is performed off-line to create the datasets for exploration. We assume that the input of our data preparation has been already collocated in a common and consistent reference frame, by performing any sort of data fusion or registration procedure (see Sec. 5.1). In the following, we describe the approaches that we use to create data maps, which encode the measured data (Sec. 5.1), and annotations layer, which encode related information not present in the object itself (Sec. 5.2).

(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 4: **Image Layers Creation.** These are a set of image layers created by the shape and material modeling pipeline from MLICs. The shape information is conveyed with a normal map field ( a). The parameter of the visible Ward model have been organized in a diffuse color ( b), a specular color ( c), and a surface roughness ( d). Invisible spectrum has been sampled as well, both in the infra-red IR ( e) and ultra-violet UV ( f) signals (here the diffuse components are shown). Finally, false color images can be produced by the computed images; here we show both an image obtained by merging the visible and the UV components ( g), and the UV fluorescence ( h).

## 5.1 Creating Data Maps

The elements that compose the input handled by the proposed web-based exploration approach are in the form of a set of 2D data maps, which contain information related to shape and material characteristics of the object under study. The main prerequisite for the preparation of these maps is that the input must have a form that allows the production of 2D/2.5D registered layers. To make this concept more clear, some acquired data are intrinsically registered, while others need an explicit alignment procedure. For instance, classic RTI [CHI 2019], Multi-spectral-RTI [Watteeuw et al. 2016], multi- or hyper-spectral image stacks acquired by a single camera [Eichenholz et al. 2010; Beule et al. 2007] are given by the sensor already with a pixel-level registration, and the data they natively produce is ready to be processed by our data maps creation pipeline, and used in our visualization framework (see also Sec. 5.1.2 and Sec. 5.1.3). Sometimes, due to different light paths in the sensor optics that depends on each single filtered wavelength, multi-spectral signal has to undergo a sub-pixel refinement of the wavelength alignment, although it has been acquired by the same steady sensor. Another implicitly registered case is the optical signal "baked" in the 3D shape, so that, once a 2D view of the object has been chosen and fixed, the extracted rasterized images of its different attributes are implicitly aligned (see Sec. 5.1.1). When different multi-modal acquisitions have been performed by different sensors, from different point of view, and at different times (e.g., technical photography, infrared reflectography, etc.), an explicit registration is needed, to bring all the data consistently within the same reference frame; this step will give us a data that is aligned at pixel level and viewed from a fixed and specific direction. A lot of different solutions exist to perform this operations, based on different rationales, i.e., from photogrammetry, to mutual information, or to other hybrid multi-modal data fusion techniques [Pamart et al. 2017; Shen et al. 2014]. In any case, all those examples have pretty standard and well known approaches to provide a properly registered signal. Thus, they can easily provide the input for our data preparation. Moreover, they are daily used both by scholars and investigators in the CH community [Chane et al. 2013; CHI 2019; Pamart et al. 2019; Gunawardane et al. 2009], and also by general non-experts in many application fields that are far from the computer science domain.

Our framework can generate those data maps from a variety of sources, ranging from 2D view from colored 3D model to Multi-light Image Collections (MLICs). Given the mathematical representation and the organization of their parameters into a set of data maps, the rendering engine will be capable of taking a new light direction and producing in real-time virtual relighting of the object. In the next three sections we explain several complementary alternatives to prepare the data for our viewer. Those data preparation strategies show a high level of interoperability of the proposed approach with the standard pipelines and practices in the field of digital models and Cultural Heritage. Each type of data maps that is created with one of the following procedures can be performed by common, widespread tools (see for instance Meshalb [Cignoni et al. 2008], RTIBuilder, or RTIViewer [CHI 2019]), daily used by non-experts in the Cultural Heritage field. These data creation pipelines are stand-alone independent procedures, each of them leading to a set of images that can be used as data layers for our visualization system.

5.1.1 *Creation of data maps from 3D models.* Independently of the 3D representation (it could be per-vertex colored point-cloud, textured triangle mesh, etc.), we start by choosing the proper 3D view point and the corresponding camera settings, from which we want to visualize the object and generate the data maps. Then, we render the 3D model to project surface and material attributes to images. In this way we produce, for instance, the surface normal map, the diffuse color map, etc. To do so, we use Meshlab [Cignoni et al. 2008], a free and powerful tool for 3D model visualization. Those resulting images are a perfect set of layers for our inspection tool. Depending on the type of extracted layers, of course, a different rendering strategy will be applied to produce meaningful visualization. An example of layered visualization of data obtained with this procedure has been presented in Section 7.2.

5.1.2 *Creation of relightable image models from MLICs.* In the case of Multi-light Image Collections (MLICs), which are defined as stacks of photos of a scene acquired with a fixed viewpoint and a varying surface illumination, a first supported representation is the relightable image, which strives to directly model the reflectance field without separating shape and material contribution. In particular, we can import common Polynomial Texture Mapping (PTM) [Malzbender et al. 2001] and Hemi-Spherical Harmonics (HSH) [Elhabian et al. 2011; Mudge et al. 2008] formats generated by the CHI [CHI 2019] capture software and store them as a series of 3-channel data maps. In these cases, it is possible to store the $N$ parameters of PTM or HSH into $N/3$ 8-bit RGB images, by just using the classic approach proposed by Malzbender et al. [Malzbender et al. 2001], i.e., the computation of the parameters' scales and biases, and then the 8-bit quantization of the scaled and biased parameters.

To ensure interoperability with the standard relightable image preparation pipelines (e.g., Polynomial Texture Maps from RTIBuilder [CHI 2019]), we can accept image layers coming from more sophisticated and non-standard processes. In general, if the output of a MLIC processing step can be represented by a series of 2D images, that result can be an input of our visualization algorithm. This versatility of the proposed solution open the way for improved data preparation strategies. For instance, in order to increase repeatability of the measurements, we can handle data layers obtained by MLICs acquired with non-standard setups, and processed with advanced calibration techniques [Giachetti et al. 2018; Sun et al. 2013; Angelopoulou and Petrou 2014; Xie et al. 2015], ranging from flat fielding techniques to analytical modeling of light direction and intensity. The calibrated measurements (and sometimes lights) at each pixel are finally approximated with the standard PTM or HSH analytical model, which maps an input 2D-space light direction into a reflectance value, using the standard least-squares approach.

The addition of the calibration step is of course optional, and removing it does not prevent the use of relightable images (PTM, HSH) within our visualization tool. However, it shows the level of modularity of the proposed method and its versatility across different processing scenarios. Section 7.4 proves how our tool can properly and seamlessly work with data coming from the classic PTM/RTI pipeline.

5.1.3 *Creation of surface and material maps from MLICs.* Another alternative way to produce data maps from MLICs is to explicitly derive surface and material maps. As previous procedures, our visualization tool is compliant with this output obtained by using several different approaches (standard and non-standard ones). For instance, we can compute shape information, in terms of a normal map field, and color information, as an albedo signal, from standard tools (e.g., RTIViewer [CHI 2019]), which are easy to use even by non-experts, and are widespread among the Cultural Heritage community. On the other hand, although not mandatory, to increase the quality of the final result, it is advisable to use a calibrated pipeline. Again, the relaxed requirements for producing the input for our tool make it very flexible, with the chance of integrating it in very different processing scenarios. If we don't use standard tools to create shape and material data maps, we have to first estimate the surface shape information in terms of a normal map field, typically by applying Photometric Stereo [Woodham 1980] to the MLICs data. A layer is obtained that provides a per-pixel surface normal map. Given the original images and the normal map, we can compute the information about the material optical behaviour in terms of a Spatially-Varying Bidirectional Reflectance Distribution Function (SV-BRDF). The parameters of this function has been obtained through its fitting with the measured data, and those parameters (together with the normal map image) will be the data maps used by the rendering system. One common choice is to use the the isotropic Ward model [Ngan et al. 2005] as the analytical SV-BRDF. The parameters of the Ward function are the diffuse color, the specular color, and the surface roughness, which are organized in a series of data maps. With classic RGB input images there will be seven parameters (3 data maps), while they are $2B + 1$ in the case of multi-spectral images with $B$ different spectral bands ($(2B + 1)/3 + 1$ data maps). We show a case study that employs this data map creation procedure in Section 7.1.

5.1.4 *Derived data maps.* In addition to the above representations, which aim to directly represent the objects, we produce some other derived data maps that are useful for data analysis. In particular, in the case of multi spectral acquisitions, we combine visible and non visible images to build false-color data maps to convey specific material behaviour not exposed in common visible images (see Sec. 4.1). Fig. 4 presents an example of a set of data maps obtained by the SV-BRDF computation pipeline and false color computation.

### 5.2 Creating annotations

A detailed description of the authoring process is orthogonal to the topic of this paper, which focuses on the exploration of models. We will describe, however, our generic annotation creation process.

Since our annotation layers in the viewers are pyramids of semitransparent images, they can be imported from any workflow that can produce such a representation, including exporting them from external annotation tools, getting the images from processing steps (e.g., edge detectors), or creating them using interactive means that, in the end, are capable to store image layers. Since there is not yet a standard for semantic annotation, in order to make our solution compliant with the highest number of external operations, the only constraint we require is that those external annotation tools (either vectorized mapping methods, textual processing tools, or photo editing interfaces) have to be capable of transforming and exporting the created annotations into a 2D rasterized format with transparency (e.g., PNG with alpha channel). This is a very small and barely limiting constraint that applies in the vast majority of scenarios, and aims at highly facilitating the interoperability of the proposed visualization system.

In terms of interactive creation, we can either use external editors, or use a simple interface that expands our advanced viewer (see Sec. 6) with editing capabilities.

5.2.1 *Using external editors.* Any image editor that can produce a drawing or an image could be used with our method. In particular, we have adopted the following process, fully based on open source solutions. We choose one map/overlay from the layer under study, e.g., the diffuse or specular color, the normal map, etc. We open it into an editor, we use GIMP [GIMP 2019], and we draw the annotation as a transparent layer. After we finish that or more annotation overlays, we save them as PNG images with transparency (alpha channel), and we output a dataset description with the levels associated to each annotation (see Sec. 4.3). The same dataset description also links each level to an optional hypertext description. Note that it is not necessary, and sometimes not recommended either, to have all the levels assigned to an annotation; on the contrary, if it is needed that an annotation remains for different levels of zoom, it is advisable to map that to more than one level. In Fig. 3 we show three different levels of an original layer map (in this case the diffuse color of a Ward model of a statue), the two annotations linked to the first and third level, and, finally, the superimposed visualization of them at three different levels of zoom. Note that for the middle zoom, since no annotation has been inserted, only the original map is displayed. As described before, our architecture allows to show different annotations depending on the view, in order to show the ones more appropriate for the zoom level. For this, the creation of the annotation pyramid is based in patches, which are sub-images with transparent background framed in a global coordinate system. They are resolution independent and can be applied at different levels of the annotation mipmap.

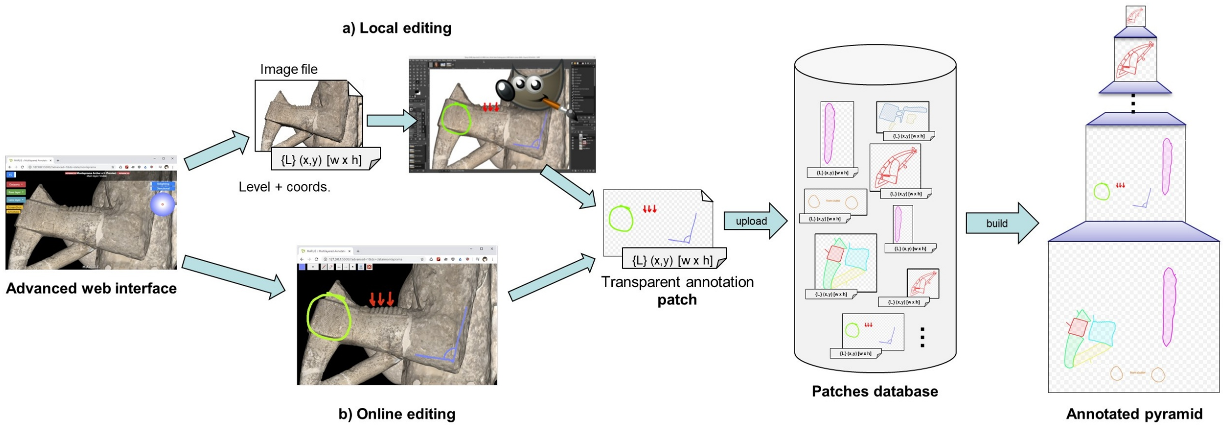



Fig. 5: **Simple annotations workflow with the advanced viewer.** Annotation authoring starts by selecting, through panning and zooming, the area of the model to annotate. That area is a variable-sized rectangle in image coordinates mapped to a fixed-sized rectangle in screen coordinates. It is annotated by creating a semitransparent image with the annotation. The image can be created by external editors or by an in-viewer editing toolbox. The individual annotated rectangles are uploaded to a database, and a server-side process rebuilds the full annotation pyramid by combining the individual rectangles at the correct scales.

5.2.2 *Adding annotations using the advanced viewer.* The creation workflow of new annotations within our viewer is shown in Fig. 5. It is based on the concept that individual annotations are done always on screen-sized rectangles, i.e., at the resolution at which the user will typically see them. The overall annotation database is obtained by combining together all the annotated rectangles (see Sec. 5.2.3). In the viewer, the user navigates over the model and individuates an area of the model to annotate by panning and zooming. When the view is selected, the system computes the parameters associated to that framing that defines the geometric information of the patch in the image reference frame: its coordinates, size and the suggested level of detail depending on the zoom level. Then, the user creates the annotation patch directly inside the web browser using the given toolkit. It is done by overlaying a transparent canvas over the view and enabling tools that allow the user to draw the annotation. We implemented a simple drawing tool for onsite annotation as a proof-of-concept. It allows painting with basic tools such a brush with different colors and sizes, an eraser, clear the whole canvas, undo/redo and download the patch. This approach allows to directly send the annotation patch and its meta-information to the server without manual procedures from the users, as well as to provide them with custom, domain-oriented annotation tools.

5.2.3 *Server-side combination of annotated patches.* In the server side, the database of annotation patches for a dataset consists of a set of pairs of image rectangles and their metadata. In our current implementation, we encode the geometric metadata of the patch directly in the filename as *L_X_Y_W_H_O_INFO.png*, being $L$ the level to apply the patch, $(X, Y)$ and $(W, H)$ the top-left corner coordinates, width and height of the patch for the highest resolution level. The parameter $O$ is a simple number representing the order of application of the patch, for dealing with overlapping patches. Finally, the $INFO$ field is left to the user for referencing the annotation. This encoding into the filenames showed to be convenient as it allows the database to be just a folder with image files, which is very easy to manage and update both locally and remotely. Yet, the data management could be easily replaced with a more sophisticated system such as a relational or an object DB if required. A computational process converts this database into the actual annotation pyramid used in the visualization. The annotation pyramid builder is written in NodeJS, being easily integrated in a HTTP server, and allows to convert the folder of patches into the annotation pyramid as a set of images that will be served and loaded as a mipmap in the client side. In first place, it orders all the patches by level, thus obtaining the ranges of the pyramid. Then an empty transparent image is created for the lowest level, and for the other levels, all are created by halving their size. Then every patch is alpha blended at its corresponding level, scaling the coordinates and resolution by the inverse of two elevated to the level number.

## 6. EXPLORING LAYERED MODELS LOCALLY AND ON THE WEB

Our client-server viewer has been designed to provide a flexible exploration experience of a multi-layered dataset. At the core of the viewer is a basic visualization engine, which supports layer display and layer combination (Sec. 6.1). The engine is used as a basis in customized viewers, which, by configuring and selecting user interface components, target specific application use cases (Sec. 6.1). All viewers support both single-touch and multi-touch interaction so as to support a large variety of devices (Sec. 6.3), and can work for both standalone machines or while connected to a remote server (Sec. 6.4).

### 6.1 Visualization engine

The visualization engine is developed as a JavaScript library that manages the memory and the display. It is initialized with the layer structure defined in Sec. 4.3. Using a hashed container, it first determines the unique set of image resources required for a given multi-layered object. A cache manager, is then, responsible to move the resources to RAM by fetching data from the server, and then to GPU texture memory, as required. The RAM cache is larger than the GPU cache, which is limited to maintaining in core at most all the textures required by all currently active layers, at the minimum the textures required by a single layer. At data fetching time, all the available mipmap pyramid is fetched. When loading to texture memory, all the needed mipmap levels are transferred (if available) or created using the hardware supported *generateMipmap* primitive. Automatic mipmap creation is only done for renderable layers, and not for overlay layers, which implement levels of abstraction.

The rendering process maintains a stack of active layers, which are rendered back to front and alpha-blended using OpenGL to implement layer compositing as well as interactive lenses. For each layer, the maintained status includes, in addition to layer data and image caching status, the viewing parameters and the rendering parameters. All layers are rendered by activating a specific vertex and fragment shader pair, which depends on the layer type.

For each currently active layer, first the GPU texture cache is updated. Vertex and fragment shaders are then activated and configured, while rendering is triggered by drawing a geometry corresponding to the area of the viewport that should be updated (with the full viewport for regular layers, and the lens geometry for lens layers). The vertex shader just computes the data required to compute the 3D and texture coordinates based on the current viewing parameters, while the fragment shader performs the actual rendering, exploiting samplers bound to each of the required texture resources, and rendering parameters associated to uniform variables. In order to enable relighting, all shaders take as input a light direction and intensity. Moreover, in order to support transparency, they also take as input a layer opacity value. In addition, per-mode specific parameters are associated to each shader (see Sec. 6.2). The end result of the shader is a RGBA pixel value, which is composited by alpha blending with the currently stored frame buffer vale to implement transparency. After the renderable representation of a layer is generated, the associated annotation layer is also rendered (if activated) and composited with alpha blending, using the same geometry as the renderable layer. The vertex and fragment shader pair selects the correct level of detail without doing further processing. By iterating the rendering of all active layers and associated annotations, the full image is produced.

### 6.2 Configuration and interaction components

The visualization engine is exploited as a low-level library to implement several viewer kinds. We currently support two specifically targeted implementations. The first one is an advanced viewer which exposes in addition to tools for tweaking all the shading parameters for each displayed layer, a toolbox for rapidly generating annotations (see Fig. 6 left) . The second, the main focus of this paper, is a simplified viewer which only uses visualization modes fully configured using presets (see Fig. 6 right).

In both viewers, we only enable the concurrent visualization of a base layer and a single lens layer. The idea behind supporting only one layer of interactive lenses is to provide, on-demand, an alternative visual representation of the data underlying a local area of the screen. By selecting which layer and presentation mode is used as a base to fill the frame and which layer and presentation mode is, instead, rendered within the possibly semitransparent lens, many common inspection use cases can be covered, from focus-and-context visualization of multiple acquisition modalities, shapes and material representation, and attribute enhancements to the locally-controlled selective display of annotations.

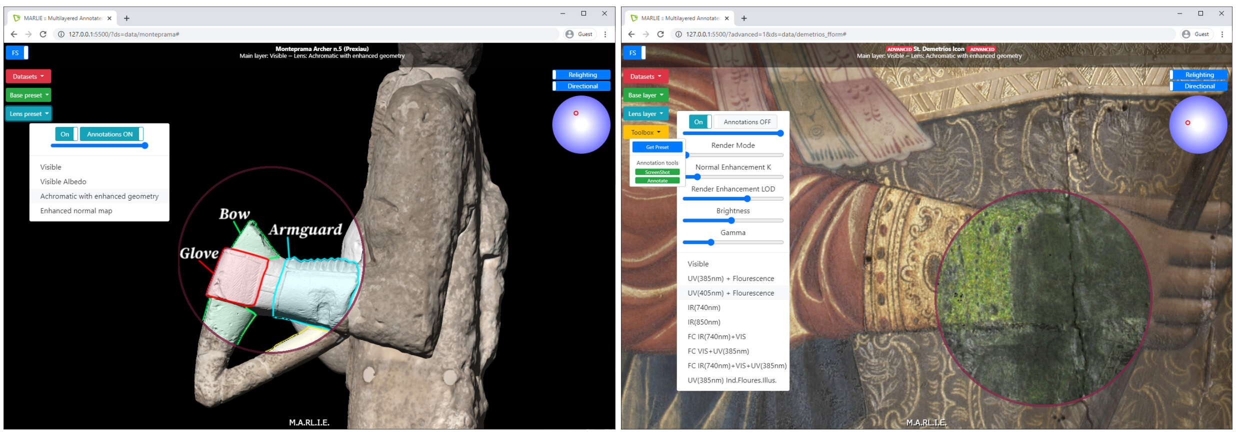



Fig. 6: **User interfaces.** Left: simplified configuration. Right: advanced configuration.

6.2.1 *Advanced viewer.* The advanced configuration of the interface is intended for expert users to both directly explore datasets using an unrestricted user-interface and to prepare datasets for browsing with the simplified user interface based on presets. In particular, user interface tools make it possible to explore and tweak the space of parameters for the rendering, as well as to generate annotations patches as described in Sec. 5.2.2. Thus, expert users can explore and analyze the datasets, without restrictions, by using their knowledge of the acquisition method or representation model to modify the rendering parameters, saving the best combinations as presets. For both the base layer (full frame) and the lens layer (inset), the user can choose any map of the dataset and modify, using GUI widgets, the parameters common to the supported relighting models (such as render mode, normal enhancement factor, brightness, gamma correction, transparency of the lens, etc.). These settings are implemented as sets of values for shader-specific uniform variables, which defines a particular view of the model. When the (expert) user finishes the tweaking of the parameters, resulting in a combination with a particular interest, the set of values is loaded into a "preset", which can be named and exported in a JSON file, to be stored in a preset library ready to be used by the simplified interface.

6.2.2 *Simplified viewer based on presets.* The simplified viewer configuration, instead, just points to a multilayered dataset description and to a preset library, which provides two lists of configurations, one for the base layer and one for the lens layer. Each entry in the configuration list consists of a label associated to a rendering configuration, comprising of a layer id in the multilayered dataset, a preset id in the preset library, and, for the lens layer, an opacity value. At start-up time, the simplified viewer extracts, from the configuration of the multilayered dataset, the list of needed layers to pass to the visualization engine, and creates two context menus pointing to the associated configurations, one for the base layer and one for the lens layer. Such a restricted setup makes interaction more constrained but very simplified, making it possible to target a non-expert audience, as well as dissemination or exhibitions use-cases in which the various rendering modes are carefully authored at data preparation time,


Fig. 7: **BRDF slice.** Visualization of BRDF slice under the cursor.

## 6.3 Pointer-based and multi-touch interaction

The minimalistic interface, using only the base and context menu, is complemented by only two other interaction widgets: one for enabling/disabling the lens and another for switching from relighting mode to measurement mode. All other actions are performed by single or multi-touch interaction. The single touch interface is meant to be operated with a 2D mouse, while multi-touch input is the default operation on most interactive surfaces. To reduce training, we use analogs of common 2D RST gestures to operate the interface, tuning the behavior depending on the current mode and the location of gesture start. Two-finger pan and pinch are mapped to translation and scaling of the lens when the gesture starts inside the lens, and to pan an zoom of the image otherwise. One-finger dragging is mapped to light direction control using the standard trackball interface when in relighting mode, and to appearance display when in measurement mode. The appearance display is only active when the current layer contains a BRDF, and displays the $\Phi_d = 90$ slice of the BRDF under the cursor, which is a good indicator of the main reflectance features of the underlying material (see Fig. 7).

### 6.4 Online and offline deployment

The framework has been designed to easily support both online and local deployment. On the server side, there is currently no need for dedicated servers with specific server-side computation. All communication can be handled by a standard web server (we use Apache in our implementation) accessing a storage location with image data, metadata, and configurations. The exact same setup can be achieved on a single mini-PC without internet connection, using a local client transparently communicating with a local server. This makes deployment easier also for small institutions without any IT infrastructure (like most museums and galleries) and limits the problems related to the need to export possibly protected models outside the institutions (as needed, e.g., for cloud-based systems such as SketchFab [Sketchfab 2019]. Moreover, the simplified touch-based interface is well adapted for walk-up-and-use interaction modalities, making this work suitable for kiosks or other installations, where the system can be simply run within a browser in full-screen mode.

## 7. RESULTS

An experimental software library and viewer applications have been implemented both using JavaScript and the WebGL2 environment. We have extensively tested our system with a number of complex heterogeneous datasets. In this paper, we discuss the results obtained on three different use cases using our simplified interface based on presets: the multi-spectral exploration of a painting (Sec. 7.1), the analysis of a fragmented statue coming out of a 3D scanning pipeline (Sec. 7.2), and the analysis of a damaged fresco coming out of an RTI pipeline (Sec. 7.3). As a final example, we also discuss the visualization using the advanced interface of a bas-relief captured with MLIC acquisition and represented with PTM-based relightable images (Sec. 7.4). The accompanying video shows the system in action on all the datasets on several platforms, which were selected to demonstrate both the portability of the code and the adaptability to several different use cases. In particular, we use a museum setup based on a 98 inch UHD 4K touchscreen display driven by Chrome 75 running on Windows 10, a desktop setup using Firefox 66 on the Arch Linux distribution, a tablet setup using a Teclast T10 Master device with Chrome 74 on Android 7.0, a smartphone setup using a V Mobile J7 device running Chrome mobile 75 on android 7.0, and a tablet-with-pen setup using a Samsung Galaxy Tab S4 with Chrome 84 on Android 9.0.

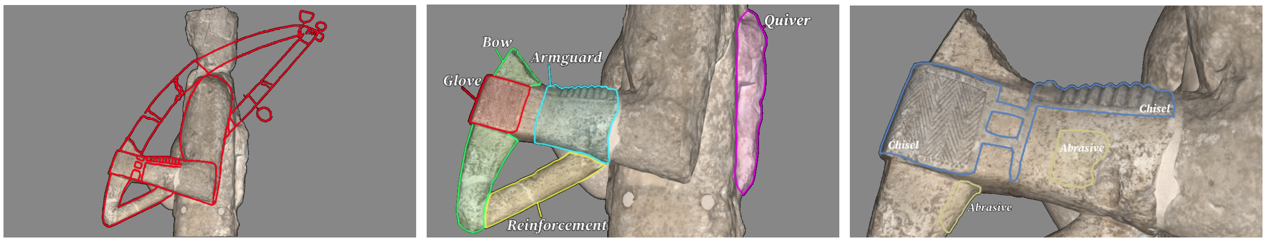
### 7.1 Multi-spectral data exploration

In this use-case we show how the proposed exploration tool is capable of efficiently visualizing and inspecting a spatially-varying multi-spectral signal of a painting. The artwork is the Icon St. Demetrios ($17^{th}$ - $18^{th}$ century). The icon is interesting from the material analysis point of view; it is an egg tempera painting on a wood support, and it contains regions where gold leaf has been employed for artistic reasons.

The painting data has been acquired by a custom free-form Multi-spectral RTI setup, i.e., a camera and a hand-held light source. We have chosen a standard high-resolution Nikon D810 DLSR camera without the hot mirror (the IR filter), and a Nikkor Lens $50mm$ $f/1.8D$, which ensures a proper transmittance characteristics across the spectrum from near-UV to near-IR. We capture seven monochromatic signals, by exploiting a set of five different LEDs light sources, and the three Bayer filters in the camera color filter arrays (CFAs); the light sources are two ultra-violet (UV), a white (VIS) and two infra-red (IR) emitters. The UV LEDs are centered at $385nm$ and $405nm$ with a full width at half maximum (FWHM) of about $20nm$; the IR emitters are centered at $740nm$ and $850nm$ with approximately $40nm$ FWHM;

Fig. 8: **Illustrative lens.** Different uses of the interactive lens and the context+focus metaphor. (Top) a crack visualization is more meaningful by adding, to the pure visible signal, an almost opaque lens showing an enhanced achromatic layer emphasizing shape variations.(Bottom) Areas of different materials are better understood by adding, to the visible signal, an interactive lens showing UV-induced fluorescence, which clearly differentiates the right side from the left side material.

on the other hand the VIS LED is a white light ($5500K$) that covers the entire visible spectrum. The entire captured scene contains both the painting and a series of calibration objects, i.e., a white planar target and four dark glossy spheres. For each wavelength, we have acquired a total number of about 60




(a) Reconstruction Hypothesis (b) Five statue's elements (c) Left arm decoration

Fig. 9: **Annotated 3D scan model.** Three different levels of annotation have been presented for a statue of an archer: (a) a hypothesis of the reconstruction of the missing archer's bow; (b) five archer's parts highlighted with contours and text annotation, i.e., the quiver, the armguard, the remaining part of the bow, the archer's glove, and a reinforcement part; (c) motifs of the decorated glove and armguard.

images, with a light direction domain that covers the hemisphere above the painting. Reflective and diffuse targets have been used to determine calibrated per-pixel light direction and intensity [Giachetti et al. 2018], from which shape and material property maps have been derived.

The final multi-layered representation of the Icon St. Demetrios dataset consists, first of all, in a single shared normal map computed with a robust photometric stereo method [Drew et al. 2012], and several material layers computed by applying a BRDF fitting procedure to the acquired data, i.e., a visible light layer, two pure monochromatic infrared layers, and the two ultraviolet layers that contain both the invisible reflectance and some induced fluorescence. The model used for BRDF fitting was Lambertian + Ward. In addition, through data fusion, we included three false-color layers by combining visible and infrared, visible and UV, and all the other seven signals merged together.

Due to environmental conditions, this painting is a clear example of an artwork that needs to be constantly monitored through periodic acquisitions, visualizations, and analysis. In fact, the artwork is affected by a series of damages and conservation issues. The painting contains cracks, layer separations/detachments, fungi, retouching areas from previous restorations; the protective layer has undergone oxidation, soot or dust deposits, fine cracks, and discoloration. All of those defects involve the appearance of the object and its geometric behavior, so it is a good example of how the proposed visualization tool can assist the scholars in examining the painting, and taking a decision on future interventions. Moreover, by employing calibrated pipelines (which ensure a high level of repeatability in the measurement) and proper registration procedures (to align different acquisitions), our tool can also easily enable the visualization of multiple datasets taken at different times, e.g., before and after restoration, or some natural or artificial aging procedure. Together with all the proposed visualization metaphors, this might help scholars and strongly support their study of the life on an artwork in a multi-temporal setting.

As illustrated in the accompanying video, we employ for this use-case our simplified interface based on presets, which permits a quick switch between predetermined rendered configurations. In this case, the viewer configurations permits, for the base layer, to switch between any of the visible, uv, ir, and fused layers. For the lens, used for detail exploration, we provide a choice between induced fluorescence, diffuse IR, enhanced achromatic display (diffuse grey BRDF with unsharp-masked normal map), and false-color normal map.

In Figure 8 (Top) we show how the proposed interactive lens tool is capable of revealing the information about a crack on the surface in an extremely clearer way compared to a plain visualization. Here, the context layer is the full material rendering, while within the lens a enhanced monochromatic signal has been activated. The presence of dust, fungi, or other superficial contaminant could be often analyzed through ultraviolet imaging. In this case (Fig. 8, Bottom), by only inspecting the visible signal (left), we could not spot any difference in material behavior across the shield zone of the painting. Conversely, giving, as before, the visible signal as a context, the superimposed lens (right) displays the surface fluorescence induced by the ultraviolet light. With this tool it is possible to see how the left part of the face has a completely different chromatic value (more gray) than the right part (more green). In this illustrative representation of the induced ultraviolet fluorescence, the more gray is the surface, the less fluorescence we have.

## 7.2 Exploration of an annotated 3D scan model

Our second use-case illustrates the exploration of an annotated view of a 3D model. The exploration has been conducted with one of the statues from the Mont'e Prama collection [Bettio et al. 2015], which includes 25 life-size human figures, depicting archers, boxers and warriors, and 13 building models representing typical Nuragic towers. All the statues have been acquired using a combination of laser scanning and flash photography to recover shape and albedo of complete 3D models. The example discussed here is Archer $n.5$ (*Prexiau*).

The 3D model coming out of reconstruction has been loaded in Meshlab [Cignoni et al. 2008], in which we have selected a camera view from which to export a diffuse color map and a normal map. The viewpoint has been chosen to display a set of important features from the CH point of view, and has been sampled at the resolution of approximately 0.25mm/pixel, which corresponds to the acquisition resolution. In this view, it is possible to see the quiver (with the resin box and the small sword), and the left arm, which holds the bow. Three different annotations have been included as well. Since the bow has not been entirely preserved (only the lowest part is visible in the statue), at the coarsest resolution level, i.e., in the overall view, the annotation aims at depicting a hypothesis of reconstruction of the missing bow (Fig. 9a). It should have lied on the archer shoulder, and the presence of a tight bowstring remains in some details in the back of the statue. At the middle level, five element of the statue are highlighted (Fig. 9b), i.e., the quiver, the armguard, the remaining part of the bow, the archer's glove, and a reinforcement part that connects the bow and the left arm. By zooming further, the annotation presents in detail the archer's fist, which is protected by a decorated glove, and the left arm almost entirely covered by an armguard (Fig. 9c). As shown in the accompanying video, the annotation display naturally adapts to the viewed area, and level-specific hypertext description can be accessed on demand to gain further information. A particularly interesting usage of the tools has been, moreover, the activation of annotation overlays inside the lens, connected with their deactivation in the main view (as shown for instance in Fig. 6 left). This permits to have a clean context presentation, without any clutter due to annotation, combined with an enhanced and annotated view just in the small lens area, which acts as a focus dynamically moved by the user.

## 7.3 Exploration of an annotated RTI model

The last example of usage of our simplified viewer based on presets concerns the analysis of a damaged fresco with annotations. At the third-floor of *Casa dei Canonici* in Credarello (Italy), in 1947 the painter Ettore d'Iseppi stripped from the wall a huge fresco-paint (1,40 x 11 meters) with the technique of the *strappo*, dividing the surface in five smaller two-meters-paints. The paint is unique in the iconography of the late XVI-century Verona for his size and his secular subject. Unfortunately, we have no photographs of the paint before the strappo. After that operation, the paint was probably so damaged to request the restoration by d'Iseppi. He added chalk-plaster where paint was lost during the strappo, and he heavily re-painted the whole surface to hide the damage. Now the painting is hard to read: the shapes are vague with no depth, the color is flat and faint, the paint layer is fragile and easy to detach. A preliminary

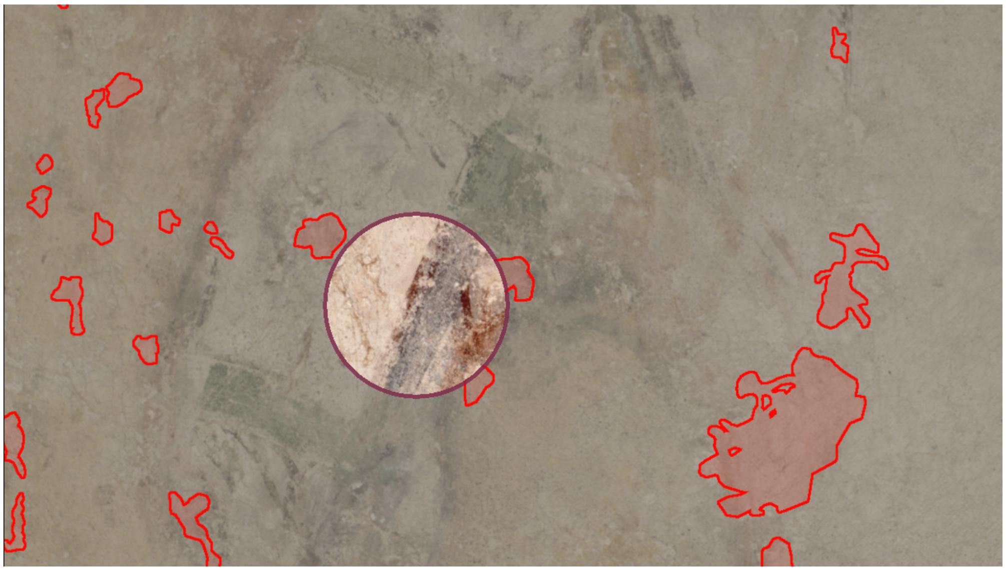


Fig. 10: **Annotated RTI model.** The area outside the lens shows the relighted and annotated version of a damaged fresco detail. The annotation refers to the regions where new materials, due to re-painting or plaster deposition, have been added. Inside the lens, a sharpened FCIR image makes the difference between pigments more evident.

survey is necessary to study the actual state of the paint and design a careful conservation job. The first task is to map the distribution of re-paintings and plasters before to plan the removal of the plaster and the stabilization of the paint layer.

The fresco was thus measured with the multi-spectral RTI technique, using the same setup employed for the example in Sec. 7.1, but using only the VIS and IR lighting. The relighting model was a pure Lambertian one.

The visualization technique has been a powerful tool in surveying. The possibility to relight the surface while enhancing normal variations through unsharp masking clearly permits to observe the variation of the surface shape due by presence of plasters and over-painted layers. This presence is also aided by the analysis of the appearance (in this case, pure albedo). In particular, the analysis was performed by producing a false-color-infrared image by merging the infrared channel with the visible channel (FCIR). An annotation overlay is added to identify the regions where new material (re-paint/plaster) has been added (see Fig. 10). As shown in the video, using a lens layer displaying enhanced FCIR on top of the visible layer with annotations makes it possible to analyze the areas where the restorer placed the annotation, perceiving through lens movement the subtle changes in shading and color due to plaster addition and changed materials.

### 7.4 Exploration of the RTI model of a relief carving using the advanced interface

As a final example, we present here the use of our advanced viewer to visualize a multi-layered model obtained through a standard RTI pipeline, more precisely, by employing MLIC acquisition, calibration and the computation of a PTM-based relightable image. The Cultural Heritage object is a patinated bronze statue, a replica of a bronze panel representing a female figure, from the Paradise door of the Baptistery of Florence, by Lorenzo Ghiberti. The original artwork has been copied by the Opificio delle Pietre Dure (Florence, Italy). The item was cast with a Cu90-Sn10 alloy, and then patinated using iron(III) chloride and coated with a protective varnish. It was placed in central Florence and exposed

(a) (b)

Fig. 11: **RTI model of a bas-relief.** We show a multi-layered visualization ( b) of a relightable image computed by Polynomial Texture Mapping. The relighted image in the context has been overlaid in the focus with a lens displaying the fitted luminance signal. We also present, from top to bottom, the nine PTM coefficients organized by three RGB images ( a); the first two images are the six coefficients of the polynomial that fits the monochromatic luminance of the model, while at the bottom we have the RGB chromaticity.

to outdoor urban environment from 2004 to 2016, to test preservation strategies against weathering. It has been selected due to the fact that its dark and shiny surface is a very challenging surface for multi-light RTI data processing. A MLIC of 54 images has been acquired by using a free-form, hand-held light source acquisition setup, and then the data maps have been computed by using the standard RTIBuilder software [CHI 2019]; the data maps are images that store the 6 coefficients of the polynomial interpolation of the luminance and an image for the chromaticity map.

We present here an example of multi-layer-based inspection and analysis of the geometric/optical response of the bas-relief surface (see Fig. 12). By using the advanced interface together with the lens tool, we show how we can inspect surface response not visible by using standard rendering, while using the standard rendering as a context.

In Fig. 12a, we show the typical PTM-based relighting (in this case a raking light coming from the right). By employing a lens tool, we overlay a focus area, where we apply different enhancing methods, tuned by the parameters in the advanced interface; in particular, here we use unsharp masking of PTM coefficients [Brognara et al. 2013] visualized with the chromatic information (Fig. 12b) or by rendering only the luminance signal (Fig. 12c). Rather than applying the classic unsharp masking technique to the final image, in this approach each single polynomial coefficient is enhanced, and the resulting coefficients are then employed to compute the final appearance by giving a virtual light. This tends to extract not only high frequency in the albedo signal, but also to highlight geometry-based discontinuities. This visualization mode, thus, can bring out a spatially-varying behaviour of the acquired signal, which is not visible in other cases. For instance, it is not contained in the normal map field from PTM (Fig. 12e) or Photometric Stereo (Fig. 12f) (here visualized within the SV-BRDF context). Moreover, it is not visible either if we separate normals and BRDF, and we try to enhance it by applying unsharp masking to the normal signal (Fig. 12g). We also apply unsharp masking directly to the final relighted image color (Fig. 12h), and those changes in surface response remain invisible. These images, and the accompanying video, also show how enhanced depictions can be significantly different than the standard ones. Using a lens, and slowly moving it around an area, it is possible to quickly see the change in representation and put it in context.

Since our advanced interface supports annotation editing, authors can mark those findings on the object in the precise location and at the specific scale used during inspection. In Fig. 12d, we present a possible annotation applied to this specific case, where the scholar marks regions that exhibit different signal with different colors.

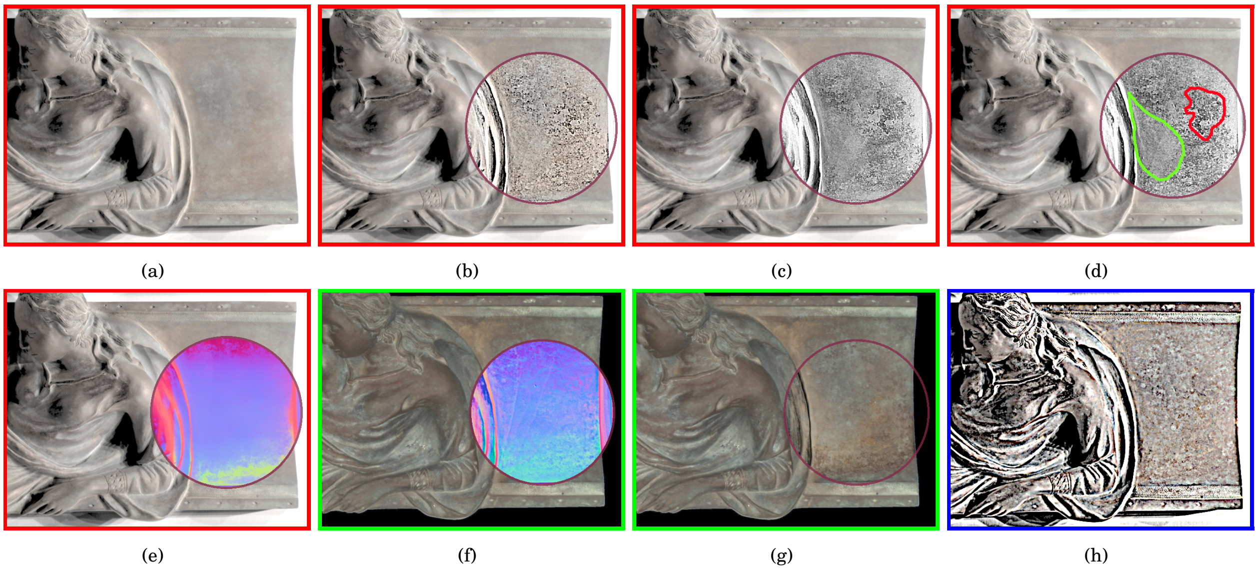


Fig. 12: **Advanced interface, lens and annotation of a RTI model of a bas-relief.** We present the inspection and annotation of a spatially-varying surface behaviour. To highlight a particular object response we employ the combined use of the lens tool and the parameters in the advanced interface properly tuned to increase detail perception. In the sub-figures with the red outline, we show the PTM LRGB standard rendering without the lens ( a) and with the lens that contains: ( b) luminance coefficient unsharp masking plus chromaticity; ( c) only luminance coefficient unsharp masking; ( d) annotations of regions with different signals; ( e) PTM normal maps. In the sub-figures with the green outline, instead, we show a SV-BRDF layer with a lens that contains: ( f) Photometric Stereo normals; ( g) normal-based enhanced SV-BRDF rendering. Finally, in the image with the blue outline, we show the results of classic color unsharp masking applied to the relighted PTM image ( h).

## 8. CONCLUSIONS

We introduced a novel approach for letting users explore detailed shape and material models integrated with structured descriptive information at multiple scales. Our approach extends the classic relighting viewers by combining and extending state-of-the-art results in several areas. First of all, we generalized fixed-view relighting interfaces to sets of image-based renderable representations. Moreover, we support multi-scale visual annotations at several levels of abstractions. Finally, we extend the application of advanced focus-and-context visual exploration means to the field of shape and material analysis in cultural heritage, namely through the use of interactive lenses.

Our results demonstrate how our approach, implemented as a web-based tool, can be applied to a variety of exploration use cases. The same viewer, running within a web browser, supports interaction setups ranging from casual inspection on mobile phones to exhibit-scale exploration on large-scale displays. The illustrated examples include the explorations of statues and paintings created from different acquisition pipelines and producing representations include analytical relighting models to decoupled shape and material data.

The presented approach is very versatile and can be applied to a variety of representations, including plain images, relightable images storing parametric representations of the reflectance fields, and explicit shape and material models in terms of spatially-varying normal and BRDF fields. As demonstrated by our results, such data can come out of MLIC pipelines, as well as 3D scanning and photogrammetry pipelines, and eventually also 3D modeling. The user interface, despite the additional capabilities, remains similar to the relighting interfaces already popular in the CH field, thus reducing the learning curve.

Being image-based, the multi-layered representation is easy to author. In particular, in addition to using the viewer to perform annotations, authors can use standard image editing tools for content preparation, leading to a simple but effective and flexible procedure to create rich visual information in forms of overlays at different levels of abstraction.

Our current implementation was, by design, focused on the exploration of the novel possibilities offered by image-based visualization of the annotations. We do not currently support the direct interaction with the individual annotations. Missing features include highlighting annotations on mouse-over, selection of individual annotations, triggering events on click of an annotation, and the possibility to selectively turn on and off individual annotations rather than merged layers. Such features could easily be incorporated in our tool, on top of the current implementation, using currently available approaches. in particular by preserving, in parallel to our merged annotation layer, a data structure with the individual annotation regions, as done, e.g., by Ponchio et al. [Ponchio et al. 2020]). In addition to extending our implementation with such techniques, we also plan to explore in the future new alternatives to interact with single annotations, that directly integrate with the presented structure and tools. In particular, we are working on integrating individual annotation data in our annotation pyramid as segmentation masks, and on enhancing the lenses to provide novel interaction means with the information layers (e.g., by triggering events), as well as to guide the lens towards areas containing annotations.

The image-based data representation, and the related user interface, avoiding the complexity of full 3D exploration is simple and effective, but, at the same time, limits by design the application of the method to single-view exploration. The approach is thus meant to complement, rather than replace, fully-3D solutions that permit to freely move around objects using 3D cameras. This image-based approach covers, however, a reasonably large portion of CH use-cases, where preferential view directions of cultural objects are, by design, very common.

Moreover, this work only targets the problem of model exploration with image/text overlays. Using other associated multimedia information (e.g., hyperlinks or video) and/or supporting very complex narratives are orthogonal problems not treated in this work, and can be integrated in the system with standard means (e.g., a canvas overlay). Similarly, the current proof-of-concept implementation uses single-image streaming, and is thus limited in terms of image size. A more scalable implementation would involve integrating a tiled pyramid image service, such as IIIF [IIIF 2019].

Finally, the current evaluation focuses mostly on demonstrating the capabilities of our approach in several uses cases. More work is required to objectively assess the effectiveness of the user interface in a variety of settings. Addressing this would require cognitive measures that are beyond the scope of the paper, and are an important avenue for future work.

Despite the limitations of our current implementation, our results indicate that our solution is of immediate practical interest for many use cases. Our web-based implementation, moreover, simplifies a widespread adoption for both local and remote-viewing settings. A source code release of our reference implementation is available from the download web site at www.crs4.it/vic/download/. We believe that this could be a very helpful instrument to support the CH community in the creation and dissemination of advanced contents on the web. We are currently working on expanding its scalability by using tiled images.

Finally, while the applications presented here are in the CH domain, the overall approach is of interest for other potential applications. As illustrated in a recent survey [Pintus et al. 2019], the exploration of single-view, multi-light, image data has found important applications in as diverse fields as natural sciences, industrial monitoring, and medical imaging, and is often done by exploiting relightable images and combinations of multi-modal acquisitions. Our approach, based on multi-layered focus-and-context inspection, should also be appropriate to many of those cases. In particular, we are currently working on its application to the visualization of simulation data at a urban scale.

## Acknowledgments

The authors thank CRBC Sassari, Accademia delle Belle Arti di Verona, Opificio dell Pietre Dure, and Ormylia Foundation for the access to the artworks for the purpose of digitization. Stefano Marziali, from the Accademia delle Belle Arti di Verona, is also warmly thanked for his collaboration to the analysis and for helpful comments and suggestions. The project received funding from the European Union's H2020 research and innovation programme under grant agreement 813170 (EVOCATION), and from Sardinian Regional Authorities under projects VIGECLAB and TDM (POR FESR 2014-2020 Action 1.2.2).

## REFERENCES

Jens Ackermann and Michael Goesele. 2015. A survey of photometric stereo techniques. *Foundations and Trends in Computer Graphics and Vision* 9, 3-4 (2015), 149–254.

Aioli. 2019. Aioli - a reality-based 3D annotation platform for the collaborative documentation of heritage artefacts. (2019). http://www.aioli.cloud/ [Online; accessed 28-Aug-2019].

Maria E Angelopoulou and Maria Petrou. 2014. Uncalibrated flatfielding and illumination vector estimationfor photometric stereo face reconstruction. *Machine vision and applications* 25, 5 (2014), 1317–1332.

Marcos Balsa Rodriguez, Marco Agus, Fabio Bettio, Fabio Marton, and Enrico Gobbetti. 2016. Digital Mont'e Prama: Exploring large collections of detailed 3D models of sculptures. *ACM Journal on Computing and Cultural Heritage* 9, 4 (September 2016), 18:1–18:23.

Marcos Balsa Rodriguez, Enrico Gobbetti, Fabio Marton, and Alex Tinti. 2013. Compression-domain Seamless Multiresolution Visualization of Gigantic Meshes on Mobile Devices. In *Proc. ACM Web3D International Symposium*. 99–107.

Fabio Bettio, Enrico Gobbetti, Fabio Marton, Alex Tinti, Emilio Merella, and Roberto Combet. 2009. A point-based system for local and remote exploration of dense 3D scanned models. In *The 10th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*. 25–32.

Fabio Bettio, Alberto Jaspe Villanueva, Emilio Merella, Fabio Marton, Enrico Gobbetti, and Ruggero Pintus. 2015. Mont'e Scan: Effective Shape and Color Digitization of Cluttered 3D Artworks. *ACM JOCCH* 8, 1 (2015), 4:1–4:23.

Pieter De Beule, Dylan M Owen, Hugh B Manning, Clifford B Talbot, Jose Requejo-Isidro, Christopher Dunsby, James Mcginty, Richard KP Benninger, Daniel S Elson, Ian Munro, and others. 2007. Rapid hyperspectral fluorescence lifetime imaging. *Microscopy Research and Technique* 70, 5 (2007), 481–484.

Cristian Brognara, Massimiliano Corsini, Matteo Dellepiane, and Andrea Giachetti. 2013. Edge detection on polynomial texture maps. In *International Conference on Image Analysis and Processing*. Springer, 482–491.

Marco Callieri, Chiara Leoni, Matteo Dellepiane, and Roberto Scopigno. 2013. Artworks narrating a story: a modular framework for the integrated presentation of three-dimensional and textual contents. In *Proc. ACM Web3D*. 167–175.

Jorge Camba, Manuel Contero, and Michael Johnson. 2014. Management of visual clutter in annotated 3D CAD models: a comparative study. In *International Conference of Design, User Experience, and Usability*. Springer, 405–416.

Camille Simon Chane, Alamin Mansouri, Franck S Marzani, and Frank Boochs. 2013. Integration of 3D and multispectral data for cultural heritage applications: Survey and perspectives. *Image and Vision Computing* 31, 1 (2013), 91–102.

CHI. 2019. Cultural Heritage Imaging website. (2019). http://culturalheritageimaging.org [Online; accessed-March-2019].

Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. 2008. MeshLab: an Open-Source Mesh Processing Tool. In *Eurographics Italian Chapter Conference*, Vittorio Scarano, Rosario De Chiara, and Ugo Erra (Eds.). 129–136.

DHLAB. 2019. RTI tools at DHLAB Basel. (2019). https://github.com/dhlab-basel/rti.js [Online; accessed-March-2019].

Julie Dorsey, Holly Rushmeier, and François Sillion. 2010. *Digital modeling of material appearance*. Elsevier.

Mark S Drew, Yacov Hel-Or, Tom Malzbender, and Nasim Hajari. 2012. Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing* 30, 4-5 (2012), 317–331.

Nicholas J Durr, Germán González, Daryl Lim, Giovanni Traverso, Norman S Nishioka, Benjamin J Vakoc, and Vicente Parot. 2014. System for clinical photometric stereo endoscopy. In *Advanced Biomedical and Clinical Diagnostic Systems XII*, Vol. 8935. International Society for Optics and Photonics, 89351F.

Jason M Eichenholz, Nick Barnett, Yishung Juang, Dave Fish, Steve Spano, Erik Lindsley, and Daniel L Farkas. 2010. Real-time megapixel multispectral bioimaging. In *Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues VIII*, Vol. 7568. International Society for Optics and Photonics, 75681L.

Shireen Y Elhabian, Ham Rara, and Aly A Farag. 2011. Towards accurate and efficient representation of image irradiance of convex-Lambertian objects under unknown near lighting. In *2011 International Conference on Computer Vision*. IEEE, 1732–1737.

H. J. Falk and L. D. Dierking. 2000. *Learning from Museums: Visitor Experience and the Making of Meaning*. Rowman & Littlefield.

Peter Fornaro, Andrea Bianco, Aeneas Kaiser, and Lukas Rosenthaler. 2017. Enhanced RTI for gloss reproduction. *Electronic Imaging* 2017, 8 (2017), 66–72.

Andrew Gardner, Chris Tchou, Tim Hawkins, and Paul Debevec. 2003. Linear light source reflectometry. *ACM TOG* 22, 3 (2003), 749–758.

Andrea Giachetti and others. 2019. apTool project. (2019). https://github.com/giach68/apTool [Online; accessed-March-2019].

Andrea Giachetti, Irina Ciortan, Claudia Daffara, Giacomo Marchioro, Ruggero Pintus, and Enrico Gobbetti. 2018. A Novel Framework for Highlight Reflectance Transformation Imaging. *Computer Vision and Image Understanding* 168 (2018), 118–131.

GIMP. 2019. GNU Image Manipulation Program. (2019). https://www.gimp.org/ [Online; accessed 13-June-2019].

Timo Götzelmann, Pere-Pau Vázquez, Knut Hartmann, Andreas Nürnberger, and Thomas Strothotte. 2007. Correlating Text and Images: Concept and Evaluation. In *Proc. Smart Graphics*. Berlin, Heidelberg, 97–109.

Dar'ya Guarnera, Giuseppe Claudio Guarnera, Abhijeet Ghosh, Cornelia Denk, and Mashhuda Glencross. 2016. BRDF representation and acquisition. *Computer Graphics Forum* 35, 2 (2016), 625–650.

Prabath Gunawardane, Oliver Wang, Steven Scher, Ian Rickard, James Davis, and Thomas Malzbender. 2009. Optimized Image Sampling for View and Light Interpolation.. In *VAST*. 93–100.

Hendrik Hameeuw. 2015. Mesopotamian clay cones in the ancient near east collections of the royal museums of art and history. *Bulletin van de Koninklijke Musea voor Kunst en Geschiedenis* 84 (2015), 5–48.

Mary Hegarty. 2011. The Cognitive Science of Visual-Spatial Displays: Implications for Design. *Topics in Cognitive Science* 3, 3 (2011), 446–474.

Zhuo Hui and Aswin C Sankaranarayanan. 2017. Shape and Spatially-Varying Reflectance Estimation from Virtual Exemplars. *IEEE TPAMI* 39, 10 (2017), 2060–2073.

IIIF. 2019. International Image Interoperability Framework. (2019). https://iiif.io/ [Online; accessed-Jube-2019].

Jacek Jankowski and Stefan Decker. 2012. A dual-mode user interface for accessing 3D content on the world wide web. In *Proc. WWW*. 1047–1056.

Jacek Jankowski and Martin Hachet. 2013. A Survey of Interaction Techniques for Interactive 3D Environments. In *Eurographics STAR*. 65–93.

Alberto Jaspe Villanueva, Ruggero Pintus, Andrea Giachetti, and Enrico Gobbetti. 2019. Web-based Multi-layered Exploration of Annotated Image-based Shape and Material Models. In *The 16th Eurographics Workshop on Graphics and Cultural Heritage*. 33–42.

Vít Krátkỳ, Pavel Petráček, Vojtěch Spurnỳ, and Martin Saska. 2020. Autonomous Reflectance Transformation Imaging by a Team of Unmanned Aerial Vehicles. *IEEE Robotics and Automation Letters* 5, 2 (2020), 2302–2309.

Tsvi Kuflik, Oliviero Stock, Massimo Zancanaro, Ariel Gorfinkel, Sadek Jbara, Shahar Kats, Julia Sheidin, and Nadav Kashtan. 2011. A visitor's guide in an active museum: Presentations, communications, and reflection. *JOCCH* 3, 3 (2011), 11:1–11:25.

KUL. 2019. PLD software KU-Leuven. (2019). https://portablelightdome.wordpress.com/software [Online; accessed-March-2019].

Primož Lavrič, Ciril Bohak, and Matija Marolt. 2017. Collaborative view-aligned annotations in web-based 3D medical data visualization. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. IEEE, 259–263.

José Luis Lerma, Talal Akasheh, Naif Haddad, and Miriam Cabrelles. 2011. Multispectral sensors in combination with recording tools for cultural heritage documentation. *Change Over Time* 1, 2 (2011), 236–250.

Zhengqin Li, Kalyan Sunkavalli, and Manmohan Chandraker. 2018. Materials for masses: SVBRDF acquisition with a single mobile phone image. In *Proc. ECCV*. 72–87.

Lindsay MacDonald, Taylor Bennett, Christopher Ramsey, and Charles Crowther. 2019. New RTI Technology for Palaeography. In *Proc. EVA London*. 174–181.

Lindsay William Macdonald. 2015. *Realistic visualisation of cultural heritage objects*. Ph.D. Dissertation. UCL (University College London).

Tom Malzbender, Dan Gelb, and Hans Wolters. 2001. Polynomial texture maps. In *Proc. SIGGRAPH*. 519–528.

Fabio Marton, Marco Agus, Enrico Gobbetti, Giovanni Pintore, and Marcos Balsa Rodriguez. 2012. Natural exploration of 3D massive models on large-scale light field displays using the FOX proximal navigation technique. *Computers & Graphics* 36, 8 (2012), 893–903.

Anastasia Moutafidou, Georgios Adamopoulos, Anastasios Drosou, Dimitrios Tzovaras, and Ioannis Fudos. 2018. Multiple Material Layer Visualization for Cultural Heritage Artifacts. In *Proc. GCH*. 155–159.

Mark Mudge, Tom Malzbender, Alan Chalmers, Roberto Scopigno, James Davis, Oliver Wang, Prabath Gunawardane, Michael Ashley, Martin Doerr, Alberto Proenca, and Joao Barbosa. 2008. Image-Based Empirical Information Acquisition, Scientific Reliability, and Long-Term Digital Preservation for the Natural Sciences and Cultural Heritage. In *Eurographics (Tutorials)*. 2:1–2:30.

Addy Ngan, Frédo Durand, and Wojciech Matusik. 2005. Experimental analysis of BRDF models. *Rendering Techniques* 2005, 16 (2005), 2.

Giampaolo Palma and others. 2019. WebRTI Viewer. (2019). http://vcg.isti.cnr.it/rti/webviewer.php [Online; accessed-March-2019].

Gianpaolo Palma, M Baldassarri, MC Favilla, and R Scopigno. 2014. Storytelling of a coin collection by means of RTI images: the case of the Simoneschi Collection in Palazzo Blu. (2014). https://mwf2014.museumsandtheweb.com/paper/storytelling-of-a-coin-collection-by-means-of-rti-images-the-case-of-the-simoneschi-collection-in-palazzo-blu/ [Online; accessed 13-June-2019].

Gianpaolo Palma, Nicola Desogus, Paolo Cignoni, and Roberto Scopigno. 2013. Surface Light Field from Video Acquired in Uncontrolled Settings. In *Proc. Digital Heritage*, Vol. 1. 31–38.

Anthony Pamart, Odile Guillon, Stéphan Faraci, Eloi Gattet, Michel Genevois, Jean-Marc Vallet, and Livio de Luca. 2017. Multispectral photogrammetric data acquisition and processing for wall paintings studies. In *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 42. 559–566.

Anthony Pamart, Federico Ponchio, Violette Abergel, A Alaoui M'Darhri, Massimiliano Corsini, Matteo Dellepiane, François Morlet, Roberto Scopigno, and Livio De Luca. 2019. A complete framework operating spatiall-oriented RTI in a 3D/2D cultural heritage documentation and analysis tool. In *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, Vol. 44. 573–580.

Ruggero Pintus, Tinsae Dulache, Irina Ciortan, Enrico Gobbetti, and Andrea Giachetti. 2019. State-of-the-art in Multi-Light Image Collections for Surface Visualization and Analysis. *Computer Graphics Forum* 38, 3 (2019), 909–934.

Gilles Pitard, Gaëtan Le Goïc, Alamin Mansouri, Hugues Favrelière, Maurice Pillet, Sony George, and Jon Yngve Hardeberg. 2017. Robust Anomaly Detection Using Reflectance Transformation Imaging for Surface Quality Inspection. In *Scandinavian Conference on Image Analysis*. Springer, 550–561.

Federico Ponchio and others. 2019. Relight website. (2019). http://vcg.isti.cnr.it/relight/ [Online; accessed-March-2019].

Federico Ponchio, Marco Callieri, Matteo Dellepiane, and Roberto Scopigno. 2020. Effective Annotations Over 3D Models. *Computer Graphics Forum* 39, 1 (2020), 89–105.

Federico Ponchio, Massimiliano Corsini, and Roberto Scopigno. 2018. A compact representation of relightable images for the web. In *Proc. ACM Web3D*. 1:1–1:10.

K. Ponto, M. Seracini, and F. Kuester. 2009. Wipe-Off: An Intuitive Interface for Exploring Ultra-Large Multi-Spectral Data Sets for Cultural Heritage Diagnostics. *Computer Graphics Forum* 28, 8 (2009), 2291–2301.

Marco Potenziani, Marco Callieri, Matteo Dellepiane, Massimiliano Corsini, Federico Ponchio, and Roberto Scopigno. 2015. 3DHOP: 3D heritage online presenter. *Computers & Graphics* 52 (2015), 129–141.

Mattias Roupé, Mikael Johansson, Mikael Viklund Tallgren, Fredrik Jörnebrant, and Petru Andrei Tomsa. 2016. Immersive visualization of Building Information Models. In *Living Systems and Micro-Utopias: Towards Continuous Designing, Proceedings of the 21st International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA 2016)*. 673–682.

Emanuele Salerno, Anna Tonazzini, Emanuela Grifoni, Giulia Lorenzetti, Stefano Legnaioli, Marco Lezzerini, Luciano Marras, Stefano Pagnotta, and Vincenzo Palleschi. 2014. Analysis of multispectral images in cultural heritage and archaeology. *Journal of Applied Laser Spectroscopy* 1 (2014), 22–27.

Roberto Scopigno, Marco Callieri, Paolo Cignoni, Massimiliano Corsini, Matteo Dellepiane, Federico Ponchio, and Guido Ranzuglia. 2011. 3D models for cultural heritage: Beyond plain visualization. *Computer* 44, 7 (2011), 48–55.

Xiaoyong Shen, Li Xu, Qi Zhang, and Jiaya Jia. 2014. Multi-modal and multi-spectral registration for natural images. In *Proc. ECCV*. 309–324.

Sketchfab. 2019. Sketchfab - Publish and find 3D models online. (2019). https://sketchfab.com/ [Online; accessed 13-June-2019].

Constantinos Sophocleous, Luigi Marini, Mohammed Elfarargy Ropertos Georgiou, and Kenton McHenry. 2017. Medici 2: A Scalable Content Management System for Cultural Heritage Datasets. *The Code4Lib Journal* 36 (2017).

Jiuai Sun, Melvyn Smith, Lyndon Smith, and Abdul Farooq. 2013. Sampling light field for photometric stereo. *International Journal of Computer Theory and Engineering* 5, 1 (2013), 14–18.

Christian Tominski, Stefan Gladisch, Ulrike Kister, Raimund Dachselt, and Heidrun Schumann. 2014. A Survey on Interactive Lenses in Visualization. In *EuroVis - STARs*, R. Borgo, R. Maciejewski, and I. Viola (Eds.). 43–62.

Bruno Vandermeulen, Hendrik Hameeuw, Lieve Watteeuw, Luc Van Gool, and Marc Proesmans. 2018. Bridging Multi-light & Multi-Spectral images to study, preserve and disseminate archival documents. In *Proc. Archiving Conference*, Vol. 2018. Society for Imaging Science and Technology, 64–69.

Pierre Vanhulst, Florian Evequoz, Raphael Tuor, and Denis Lalanne. 2018. A descriptive attribute-based framework for annotations in data visualization. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*. Springer, 143–166.

Vincent Vanweddingen, Chris Vastenhoud, Marc Proesmans, Hendrik Hameeuw, Bruno Vandermeulen, Athena Van der Perre, Frederic Lemmers, Lieve Watteeuw, and Luc Van Gool. 2018. A Status Quaestionis and Future Solutions for Using Multi-light Reflectance Imaging Approaches for Preserving Cultural Heritage Artifacts. In *Euro-Mediterranean Conference*. Springer, 204–211.

Dylan K Wainwright, George V Lauder, and James C Weaver. 2017. Imaging biological surface topography in situ and in vivo. *Methods in Ecology and Evolution* 8, 11 (2017), 1626–1638.

Zeyu Wang, Weiqi Shi, Kiraz Akoglu, Eleni Kotoula, Ying Yang, and Holly Rushmeier. 2018. Cher-ob: A tool for shared analysis and video dissemination. *Journal on Computing and Cultural Heritage (JOCCH)* 11, 4 (2018), 1–22.

Lieve Watteeuw, Hendrik Hameeuw, Bruno Vandermeulen, Athena Van der Perre, Vanessa Boschloos, Luc Delvaux, Marc Proesmans, Marina Van Bos, and Luc Van Gool. 2016. Light, shadows and surface characteristics: the multispectral Portable Light Dome. *Applied Physics A* 122, 11 (2016), 976.

Michael Weinmann and Reinhard Klein. 2015. Advances in geometry and reflectance acquisition (course notes). In *SIGGRAPH Asia 2015 Courses*. ACM.

Florian Windhager, Paolo Federico, Günther Schreder, Katrin Glinka, Marian Dörk, Silvia Miksch, and Eva Mayr. 2018. Visualization of cultural heritage collection data: State of the art and future challenges. *IEEE transactions on visualization and computer graphics* 25, 6 (2018), 2311–2330.

Robert J Woodham. 1980. Photometric method for determining surface orientation from multiple images. *Optical engineering* 19, 1 (1980), 513–531.

Limin Xie, Zhan Song, Guohua Jiao, Xinhan Huang, and Kui Jia. 2015. A practical means for calibrating an LED-based photometric stereo system. *Optics and Lasers in Engineering* 64 (2015), 42–50.

Jian Zhao, Michael Glueck, Simon Breslav, Fanny Chevalier, and Azam Khan. 2016. Annotation graphs: A graph-based visualization for meta-analysis of data based on user-authored annotations. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 261–270.

Yufeng Zheng, Erik Blasch, and Zheng Liu. 2018. *Multispectral Image Fusion and Colorization*. SPIE Press.