UNICA

UNIVERSITÀ
DEGLI STUDI
DI CAGLIARI

Ph.D. DEGREE IN
MATHEMATICS AND COMPUTER SCIENCE

Cycle XXXV

TITLE OF THE Ph.D. THESIS

Scalable Exploration of Complex Objects and Environments Beyond Plain Visual Replication

Scientific Disciplinary Sector(s)
INF/01 INFORMATICA

Ph.D. Student : Moonisa Ahsan
Supervisor : Prof. Riccardo Scateni (UniCa)
Co-Supervisor : Dr. Enrico Gobbetti (CRS4)

Final Exam Academic Year (2021-2022)
Thesis Defence: February 2023 Session

Dedicated to my dearest ***mother*** (Sumbla Amber), ***father*** (Ahsan Badr) and ***grandmother*** (Hameeda Akhtar Ghani).

Abstract

Digital multimedia content and presentation means are rapidly increasing their sophistication and are now capable of describing detailed representations of the physical world. 3D exploration experiences allow people to appreciate, understand and interact with intrinsically virtual objects.

Communicating information on objects requires the ability to explore them under different angles, as well as to mix highly photorealistic or illustrative presentations of the object themselves with additional data that provides additional insights on these objects, typically represented in the form of annotations. Effectively providing these capabilities requires the solution of important problems in visualization and user interaction.

In this thesis, I studied these problems in the cultural heritage-computing-domain, focusing on the very common and important special case of mostly planar, but visually, geometrically, and semantically rich objects. These could be generally roughly flat objects with a standard frontal viewing direction (e.g., paintings, inscriptions, bas-reliefs), as well as visualizations of fully 3D objects from a particular point of views (e.g., canonical views of buildings or statues). Selecting a precise application domain and a specific presentation mode allowed me to concentrate on the well defined use-case of the exploration of annotated relightable stratigraphic models (in particular, for local and remote museum presentation).

My main results and contributions to the state of the art have been a novel technique for interactively controlling visualization lenses while automatically maintaining good focus-and-context parameters, a novel approach for avoiding clutter in an annotated model and for guiding users towards interesting areas, and a method for structuring audio-visual object annotations into a graph and for using that graph to improve guidance and support storytelling and automated tours.

We demonstrated the effectiveness and potential of our techniques by performing interactive exploration sessions on various screen sizes and types ranging from desktop devices to large-screen displays for a walk-up-and-use museum installation.

Keywords: Computer Graphics, Human-Computer Interaction, Interactive Lenses, Focus-and-Context, Annotated Models, Cultural Heritage Computing.

ہر چند کہ ایجادِ معانی ہے خداداد
کوشش سے کہاں مردِ ہنرمند ہے آزاد!

Har Chand Ke Aeejad-e-Maani Hai Khudad
Koshish Se Kahan Mard-e-Hunarmand Hai Azad !

Indeed it is a gift by divine to [be able to]
coin fresh words with new meanings;
Yet skillful artist must work hard [to redeem this gift],
as it cannot be achieved without effort.

È davvero un dono che sa di divino [poter]
Dare vita a parole nuove con significati nuovi;
Ma l'artista per eccellenza è chi lavora sodo [per fare suo questo dono],
perché non si può ottenere senza sforzo.

—Dr. Allama Iqbal
Poem 141 : Ejad-e-Ma'ani | Book: Zarb-e-Kaleem

Acknowledgments

I'd like to thank a number of people whose support was vital in this professional and personal journey. First and foremost, the most important figure in my PhD path, my project supervisor, guide and mentor — *Dr. Enrico Gobbetti* (Director of Visual and Data-intensive Computing at CRS4); thank you for being generously supportive in every possible way. Your polite suggestions, clarity and constructive feedback allowed me to learn from my shortcomings, discover areas of improvement and shape my potential scientific interests in the long-term. I owe you my scientific accomplishments. Thanks to my immediate colleagues in the Visual Computing Group (ViC, CRS4) — Fabio Bettio, Fabio Marton and Ruggero Pintus; your kindness, expert knowledge, team-spirit, helpful attitude and humbleness is a constant source of inspiration for me.

Thanks to *Prof. Riccardo Scateni* — my academic supervisor at UniCa, for his supportive and cheerful nature, keeping this doctoral journey easy going for myself. Also thanks to Dr. Gianmarco Cherchi, Dr. Luca Pitzalis, Valentino Artizzu and all my departmental colleagues, for constant help regarding university matters. Thanks to Dr. Federico Ponchio (CNR, Pisa.) — for being my secondment host-and-supervisor and a wonderful mentor, all in the quickest time possible.

Wholehearted thanks to Katia Brigaglia, Antonio Zorcolo, Francesco Versaci, Francesca Frexia, Giovanni Pintore, Eva Almansa, and Alberto Jaspe for their kindness throughout. Thanks to the whole EVOCATION team for cultivating a friendly and healthy research environment. A huge thanks to dearest Raffaella Chierici, Loredana Coni and Luisa Giugnini for their extensive initial support in helping me settle here. I'd also like to extend my thanks to Rashid Ali — for his continuous and generous support in the politest manner possible. Thanks to Ayesha Sayed — for her earliest support to help me embark on this journey.

Big thanks to my first two teaching inspirations, my dearest mother — Mrs. Sumbla Amber and my school teacher — Madam Shazia Akbar, for polishing my childhood with utmost attention, active concern and vigilant teaching. I am grateful to all my teachers, professors and mentors for their precious contributions throughout school, college, university and workplace.

I'd like to thank all my friends, particularly Dr. Rabia Fatima and Shanshan Wang, for their love and encouragement; thank you for making me smile and clearing my selfdoubts when it was hard for me to do it on my own. Thanks to Lizeth Fuentes, Chiara Cidu, Adam Celarek, Daud Kiani, Samaneh Zolfaghari, Tarek A. Haila, Irfan Khalid and Alina Pranovich for the empathetic support, appreciation and pep-talks. Thanks to Bilal for the periodic reality-checks, it helped me with my academic priorities, limitations, actions and eventually life-goals.

Thanks to my fellows, colleagues and particularly dearest students(betu) from Pakistan for continuously sending good wishes, actively appreciating our scientific accomplishments and generously supporting the scientific surveys and questionnaires. Thanks to the vibrant people of Cagliari, for showing unconditional love, warmth, and openness. I will forever be grateful for it. (Cagliaritani — Your hearts are welcoming and your smiles shine like the sun). A part of my heart will always belong to Sardinia.

A virtual thanks to Dearest Dr. Jordan Peterson, Dr. Andrew Huberman, Joey Schweitzer, Ryan Holiday, Teal Swan, Robert Greene, Mel Robbins, James Clear, Ali Abdaal and Anna Akana, for creating valuable digital content that helped me develop protocols to move forward in times of darkness, doubt, loneliness, and failure with the best of my abilities. Thank you very much!

Big thanks to my backbone — my dearest *family*. I dedicate my PhD journey to my mom and dad for being my biggest support pillar. Thank you for believing in me, and actively prioritizing my personal goals over social norms. Your support is the root of my confidence. I love you immensely. Thanks to my grandmother — Respected Hamida Akhtar Ghani for having a vision for my bright future. My brothers — Dearest Talal and Abdul Moid *(meri jan)* for always cheering out the loudest for me. I am blessed to be your sister. I cannot imagine this without your love and support all throughout the years. In the end, a humble thanks to the part of ”me”, who did not give up, when I thought I might give up.

Moonisa Ahsan
Cagliari, Italy
7th November — 2022.

Contents

List of Figures

List of Tables

Preface

This thesis represents a summary of the work done from 2019 to 2022 at the Visual and Data-intensive Computing **(ViDiC)** group of **CRS4** (Center for Advanced Studies, Research and Development in Sardinia) under the direction and supervision of **Dr. Enrico Gobbetti**, whom I really want to thank for opportunity to be part of his team. Also, my gratitude to **Prof. Riccardo Scateni** for his constant motivation and academic support from University of Cagliari. This was, indeed, a rewarding experience both scientifically and personally.

The scientific work in this thesis has been performed within the international framework of **EVOCATION** (Advanced Visual and Geometric Computing for 3D Capture, Display, and Fabrication) project, which is a leading European-wide doctoral Collegium for research in Advanced Visual and Geometric Computing for 3D Capture, Display, and Fabrication supported by European Union's H2020 research and innovation program grant 813170. The consortium participants are the University of Rostock (UNIRO), the Center for Research, Development and Advanced Studies in Sardinia (CRS4), the University of Zurich (UZH), the Italian National Research Council (CNR), the Technical University of Vienna (TUW), Fraunhofer IGD (FHG-IGD), and the two companies Holografika (HOLO) and GEXEL. The goal of this ITN is to address the current and future major challenges in scalable and high-fidelity shape and appearance acquisition, extraction of structure and semantic information, processing, visualization, 3D display and 3D fabrication in professional and consumer applications. As an ESR and Marie-Curie Fellow in the project, my research trajectory focused mostly on user-interaction methodology applied to 3D graphics.

European Union Horizon 2020 | EVOCATION ITN-MSCA | Marie Skłodowska Curie Actions | POR FESR Sardegna 2014-2020 | Sardegna Ricerche | Museo Archeologico Nazionale Cagliari | Università degli Studi di Cagliari (Italy). | Center for Adv. Studies, Research & Development in Sardinia (Italy).

With this fellowship, I was also enrolled as PhD Student in Computer Science Program at the Department of Mathematics and Computer Science in the University of Cagliari under the kind tutoring of Prof. Riccardo Scateni.

My topic is ***"Scalable exploration of complex objects & environments beyond plain visual replication"*** – where I studied modeling methods and interaction techniques for permitting exploration of high fidelity 3D models annotated with semantic information". Researched topics included information modeling, rendering methods, and user interfaces. My particular case study was in the cultural heritage domain. The outcomes resulted in scientific publications and prototype implementation of exploration systems. All my activities performed supported the goals, milestones and deliverables as *ESR04* within the project.

I'd like to thank the National Archaeological Museum of Cagliari and its National Gallery, for access to the artworks for the purpose of digitization and for collaborating on our research work. This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sk lodowska-Curie Actions Innovative Training Network (MCSA-ITN) grant agreement *No 813170* and my co-authors also benefited by the support of Sardinian Regional Authorities for projects connected to CRS4 Visual Computing activities.

Moonisa Ahsan
Cagliari, Italy
7th November — 2022.

Chapter 1

Introduction

Digital multimedia content and presentation means are rapidly increasing their sophistication and are now capable of describing detailed representations of the physical world. 3D exploration experiences allow people to appreciate, understand and interact with intrinsically virtual objects. Communicating information on objects requires the ability to explore them under different angles, as well as to mix highly photorealistic or illustrative presentations of the object themselves, with additional data that provides additional insights on these objects, typically represented in the form of annotations. The main research challenges are how to effectively present a model under different angles, and how to communicate the auxiliary (geometric, conceptual and semantic) information. In this thesis, I concentrated on cultural-heritage-computing use cases, tackling these problems for the very common and important special case of mostly planar, but visually, geometrically, and semantically rich objects or visual representations. This chapter outlines the scientific motivation behind this research, a brief summary of research achievements, and enlists the structure of organization of this thesis.

1.1 Background and motivation

The virtual inspection of digital scenes, including simulation results or digital replicas of physical objects, is of fundamental importance for many use cases in disparate application fields. A typical example occurs in the Cultural Tourism and Cultural Heritage (CH) domains, where the virtual inspection of cultural objects is recognized as a precious means to support the three main stages related to the enjoyment of the artworks, i.e., the pre-visit (documentation and planning), visit (immersion and enhancement) and post-visit (emotional possession and linking) phases [2], [3].

While early approaches of 3D inspection mainly focused on passive visual presentations,

Figure 1.1: **Stratigraphic models**. Three of the many layers relevant for inspection of a painting. The main color is the albedo (diffuse reflectance) of the painting. The bottom circle displays the specularity layer, which emphasizes changes of the materials. The top circle displays a monochromatic shaded representation of the geometry, where surface details and cracks are visible. The painting has been acquired at the National Archaeological Museum of Cagliari (see chapter 6)

e.g. image galleries or authored videos in museums, the interest has now shifted to letting users directly drive exploration. This is because interactive methods are known to better support experts in inspection tasks, as well as to improve engagement of casual users in museums [4], [5].

In this context, in parallel to generic interactive viewers displaying fully-3D virtual replicas (e.g., [6], [7]), relighting interfaces, popularized by Reflectance Transformation Imaging (RTI) viewers [8], have emerged as one of the most successful exploration modes. This success is due to several practical reasons that facilitate the deployment of such viewers.

First of all, the continuous improvement of controllable lighting and digital photography has made the acquisition of high-resolution multi-light image collections practical and affordable using many different and affordable physical setups [9]. Such a collection of samples, typically arranged in image stacks, provides massive amounts of visual data that can be analyzed to extract information and knowledge on shape and appearance. In particular, from this data, it is possible to create stratigraphic representations, i.e., multi-layered images, that may contain maps computed from feature detection and

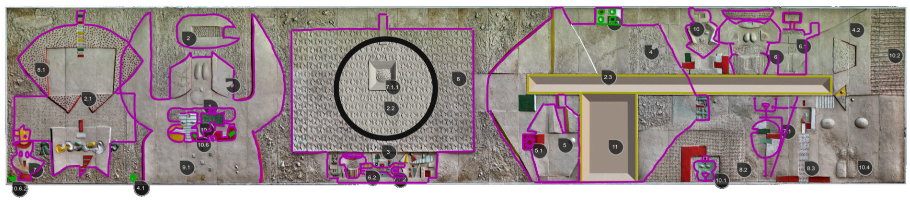


Figure 1.2: **Visual annotations of a cultural heritage object**. Interesting areas of the model have been annotated with overlay drawings. Each annotation also points to external data providing information, and may be connected also to other annotations. Hundreds of these markups may be associated with a model. Finding the relevant ones, and displaying them is a real challenge. The model is a reconstruction of the Olivetti Sandcast by Costantino Nivola done at CRS4 and ISTI-CNR and annotated by the Nivola Museum curators [1].

enhancement [10], shape information in the form of normal map [11], [12], appearance map [13], [14], and relightable models that allow the interactive simulation of different illuminations [15].

Such representations naturally support a type of visualization very appropriate to inspect fine surface details and resembling the classical physical inspection raking light sources to reveal surface detail of actual objects under study. Moreover, the restriction of camera motion to panning and zooming is very appropriate to a variety of cultural objects and, at the same time, removes one of the main difficulties of 3D exploration applications, reducing learning curves [16]. For this reason, interactive inspection of relightable images has been applied to a wide range of items [17], and are very appropriate both for expert and casual users.

Most often, several facets of the same models, have to be presented to fully understand them (see Figure 1.1). Examples include showing the appearance of an object before and after restoration interventions, displaying details only visible in particular spectral bands from a multi-spectral capture, showing underpainting and overpainting layers, or enhancing geometric information while removing or toning down material color [9]. For this purpose, a wide variety of tools have been proposed for targeting either static exploration of multi-faceted image data (e.g., multi-spectral or stratigraphic data [18], [19] or multi-light image collections [20], [21]), or dynamic exploration through relighting [9]. In all these cases, the selection, visualization, and comparison of different ways to look at an object proves challenging [22].

Moreover, creating an informative and engaging experience requires, however, to go beyond the pure visual presentation of one or more color or geometric layers. In particular, annotations linked to the digital model are often used to provide better insights to the user [23]. Traditionally, such annotations let authors identify specific regions, visually mark them with overlay text or drawing, and link them to metadata or other information that characterizes the significance of those regions [24]. However, finding relevant annotations, and presenting them in a comprehensible way without

Figure 1.3: **Result examples**. Two examples of application of our techniques to the interactive inspections of annotated relightable stratigraphic models on a large touch screen. On top, a multi-layered relightable model is explored with a visualization lens. During exploration, the camera responds to lens motion to always ensure a good focus-and-context situation. At the bottom, an annotated model is explored using the lens. The models have been acquired at the National Archaeological Museum of Cagliari (see chapter 6)

cluttering the display, and in a coherent order while conveying a context- and user-dependent narrative, is very challenging [22], [25] (see Figure 1.2 for a typical example).

1.2 Objectives

Based on aforementioned considerations, further expanded in chapter 2, I set as a goal of this theses to advance the state-of-the-art in the exploration of annotated stratigraphic 2D models by focusing on solving the following research problems:

1. **How to better display multiple layers and annotations to avoid clutter.**
   This problem is related to two sub-problems. *First*, users should focus on the object themselves and use annotations to understand them, so the display of the annotations should not mask relevant parts of the object under inspection.

Second, many of the objects' areas contain multiple annotations, and their mutual display creates overload and visual conflicts.

2. **How to guide the user within an annotated world with a suitable interface.** This is related to the fact that in order to understand objects, users must discover information which is often not immediately visible, and must be guided towards areas of interest that contain annotations. In order to deliver a coherent flow of information, moreover, this guidance should take into account not only relations between objects and annotations, but also the relation among the different annotations and the history of navigation.

1.3 Achievements

The solutions proposed in this thesis introduce a novel way to explore annotated models by building on the concepts of visualization lenses [26] and user guidance [27]. Lenses are the means to provide alternative visual representations for selected regions of interest of a display, while guidance techniques applied to lenses and structured annotation databases let the system provide assistance in discovering and selecting the most relevant annotations in response user needs (see Figure 1.3).

My main results and contributions to the state of the art are the following:

- **A novel technique for interactively controlling visualization lenses while automatically maintaining a good focus-and-context visualization** (chapter 3). The method, introduced at EUROVIS 2021 and published in the Computer Graphics Forum journal [22] automatically couples lens and camera control and is applicable to all sorts of multi-faceted visualizations. My prime contribution was on the conceptualization, methodology, and validation of the developed method.

- **A novel approach for avoiding clutter in an annotated model and guiding users towards interesting areas** (chapter 3). In addition to traditional visual markups and information links, we associate to each annotation a lens configuration that highlights the region of interest. During interaction, an assisting controller determines the next best lens in the database based on the current view and lens parameters and the navigation history. Then, the controller interactively guides the user's lens towards the selected target and displays its annotation markup. As only one annotation markup is displayed at a time, clutter is reduced. The method is also discussed in our EUROVIS 2021/Computer Graphics Forum contribution [22]. My prime contribution was on the conceptualization, methodology, and validation of the developed method.

- **A method for structuring audio-visual object annotations into a graph and for using the graph to improve guidance and support automated tours** (chapter 5). The approach, originally presented at STAG2021 [28] (honorable mention award) and later extended for the Computers and Graphics journal [29] makes it possible to support both autonomous and fully-guided visits. I have significantly contributed to the conceptualization, methodology, and validation of the method.

A later simplified refinement has been applied to the exploration of a very large annotated artwork for an exhibition [1]. For this latter use case, my contribution is only in the application of the previously designed approach based on lenses to this particular use case.

1.4 Organization

This thesis is based on the results that I have published in project deliverables [30]–[32], articles [22], [29], and conference proceedings [28]. I have organized them in order to show in a natural and coherent order all the outcomes obtained. Following is a brief overview of each chapter:

- **chapter 1** (this chapter) introduces the topic and motivation for this Ph.D. dissertation, described my objectives, and summarized my results.
- **chapter 2** provides a general background for the thesis, providing a wider view of previous approaches.
- **chapter 3** describes the technique I have introduced for controlling visualization lenses while maintaining focus-and-context, compares it with the relevant literature, and evaluates it through a user study;
- **chapter 4** describes how I exploit the concept of guidance to drive the lens towards interesting areas, letting user discover annotations; the chapter includes a user study in the cultural heritage domain that also evaluates the focus-and-context controller;
- **chapter 5** extends the two above techniques by organizing annotations into a graph to express desired presentation order and evaluates the new approach on a user study in the cultural heritage domain;
- **chapter 6** illustrates two examples of usage of the presented techniques in the context of cultural heritage pilots carried out within the EVOCATION project.
- **chapter 7** provides a conclusion and short summary of the achievements, and a critical discussion of the results obtained and of how they advance the state-of-the-art, as well as some reflections on future lines of work.

Chapter 2

General background

Before presenting the thesis contribution, I provide relevant background information on annotations linked to object, covering applications, creation and structuring, and annotation presentation. I will then focus on presentation and display issues, which are the main topics of the thesis, identifying open problems and introducing the solutions that will be detailed in the forthcoming chapters.

2.1 Contextual information representation and presentation

In this thesis, I tackle the problem of improving the exploration of objects associated with additional data that provides insights on these objects, typically represented in the form of *annotation* (see Figure 2.1 for an example).

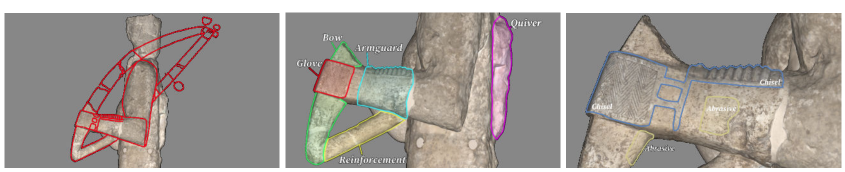


Figure 2.1: **Example of annotation on a 3D scan model**. Three different levels of annotation are associated to the 3D scan of a statue of an archer: *(left)* a hypothesis of the reconstruction of the missing archer's bow; *(center)* five archer's parts highlighted with contours and text annotation, i.e., the quiver, the armguard, the remaining part of the bow, the archer's glove, and a reinforcement part; *(right)* motifs of the decorated glove and armguard. (image from our JOCCH 2021 contribution [33])

The technical term *annotation*, in this context, refers to a mechanism that links a subportion of a geometrical representation of an object to some related information not present in the object itself. As noted elsewhere [24], 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. Many application domains, see subsection 2.1.1, are routinely creating and exploiting annotated 3D objects.

The use of auxiliary information to integrate and enhance the presentation of complex data within interactive digital applications has a long history, and 3D environments where textual information is attached to the models in a spatially coherent way have been thoroughly studied [34]. Moreover, a number of authors have proposed techniques for presenting 3D models and textual information in an integrated fashion [16], [35]–[37]. In this thesis, we aim to build on this previous research, focusing mainly on the problem of stratigraphic model exploration with annotation overlays.

2.1.1 Applications

Annotations are pivotal components for applications requiring the visual examination of (virtual) objects of focus.

In *3D modeling*, for instance, tools such as Sketchfab [7] or the ShapeNet system [38] allow users to add comments and metadata for better understanding of the object and also supporting quick search and retrieval process [39]. Moreover, the gaming industry uses annotated indicators to set-up game narrative for players, whereas developers use meta-data and annotated labels for tagging and organizing UI elements. In addition, *CAD Models* are prime sources to use annotated geometry, as they utilize both generic and numerical metadata in mechanical structures and drawings. Moreover, many web-based 3D applications are increasingly offering annotation features on 3D objects [38], [40]–[42].

Many areas including Medicine, Biological Sciences, Cultural Heritage, Digital Humanities, 3D Modeling, and Computer Aided Design are having a wide range of applications for Annotated Models [24].

For instance, in *medicine* and biological sciences, scanning equipment such as *CT (computerized tomography)* and *MRI (magnetic resonance imaging)* processes a stack of images for inspection of internal body parts, such as bones, joints, and organs. Data-sets are typically annotated with regions of interest over the 2D or 3D digital results, which must be preserved among the generated 2D/3D imagery data-set, as it is important in the diagnosis process [43].

Cultural Heritage and *Digital Humanities*, which are the target domain for the research performed in this thesis, routinely work on images or 3D models, and pioneered the usage of digital clones of artifacts for study [44]. In this context, annotations are a common feature of many systems [41], [42], [45]–[48]. These examples show evidence of the extended use of annotations in 2D and 3D systems.

2.1.2 Annotation creation and structuring

The majority of data representations for context-aware systems focus on general representations for data interconnections, rather than on interconnections between structured information and associated objects [49].

Since annotations are external information spatially associated to 3D models, they require both the identification of a location/region over a model and the creation of an explicit link between that spatial element and structured, semi-structured or unstructured data that provides a characterization of the selected location/region [24]. While regions and associated annotations can be sometimes defined by (semi-)automatic techniques, as in semantic segmentation approaches, this use case is typically limited, and in most cases each annotated region is selected accordingly by a user that links it to metadata or additional information.

In the recent survey of Ponchio et al. [24], annotations are characterized in four classes by considering the dimensionality of the reference on the 3D object: *point*, where a single point-wise position in the object 3D space locates and links the annotation; *linear*, where a polyline on the surface identifies linear structures such as fracture lines or discontinuity lines; *region*, where a subset of the surface specified by an irregular polygon is associated with a given annotation; *volume*, where sub-volumes define spatial regions to which annotations are linked.

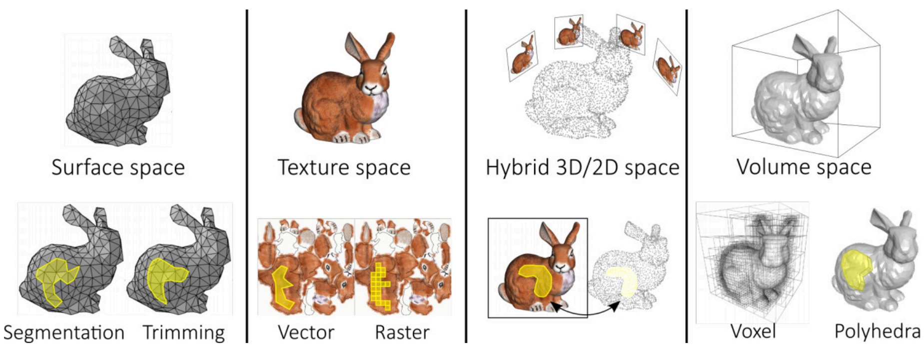


Figure 2.2: **Different region-selection approaches for 3D surfaces.** Selecting the region to be annotated strongly depends on the working space and on the representation scheme adopted for encoding the object of interest (image courtesy of Federico Ponchio [24])

Surface-based approaches (i.e., linear and region) are the most commonly used ones. Selecting the region to be annotated strongly depends on the working geometric space and on the representation scheme adopted for encoding the object of interest (see Figure 2.2). When working in *surface space*, region selection is done by triangle of point selection, either by *segmentation* through direct picking of surface primitives [40], [43], [50] or by *trimming* through clipping over the profile of the annotation. When working in *texture space*, a UV parameterization is defined over the 3D surface creating a link between the 3D space and a 2D texture space. In this case *raster* selection defines areas as subsets of texture pixels, while *vectorial* selection defines 2D polylines in texture space. *Hybrid 2D-3D projected image space* approaches define annotations by drawing on aligned images, and then propagates them onto the 3D surface or point-cloud by projection [48], [51], [52]. Finally, *Volumetric* selection defines the requested portion of the surface is defined by intersecting the model with a volume, using either *Voxels enumeration* on a volumetric structure [53], or by intersecting the surface with *Simple volumetric primitives* [45], [54], or finally by *Clipping polyhedra*, i.e., intersecting the surface with a a clipping volume defined by a closed manifold mesh [42], [55].

Balsa et al. [47] introduced a different sort of approach, which, instead of having annotations explicitly linked to a particular object region, considers them linked to a particular viewpoint. The approach is similar to *Hybrid 2D-3D projected image space*, but without explicitly transferring the annotation to the surface. Moreover, in their approach, not only the annotations are linked to the object, but they are also semantically linked between them, to guide their display (see subsection 2.1.3). In particular, they use a graph of 3D views to represent the various relations between annotations and their spatial position with respect to the 3D model. Each node associates a subset of the 3D surface (ROI) seen from a particular viewpoint to the related descriptive annotation, together with its author-defined importance. Graph edges describe, instead, the strength of the dependency relation between information nodes, allowing content authors to describe the preferred order of presentation of information. The information graph is used to guide the information display (see subsection 2.1.3)

2.1.3 Annotation presentation

While in previous work much of the focus has been on how to create, represent, and maintain annotations [24], [37], the work within this thesis to study more how, given a set of annotations, these can be effectively presented to effectively support a rich, informative, and engaging experience, with the goal of going beyond simple visual replication, supporting information integration/linking, allow shape-related analysis, and providing the necessary semantic information, be it textual or visual, abstract or tangible.

As an object or scene is typically associated with many spatially-associated annotations, displaying all of them at the same time is infeasible, and even if possible, would generate cluttering and cognitive overload. Special care should thus be taken to decide when and how an information is presented. In the context of information visualization, this is typically tackled by following guidelines inspired by the Visual Information-Seeking Mantra: "Overview first, zoom and filter, then details-on-demand" [56]. However, most works on information visualization [57] concentrate on data analysis, extrapolating results and presenting them using graphical representations tailored for better human comprehension, while in the context of annotation visualization the focus is on techniques for enhancing 3D object exploration.

Annotation presentation requirements

For the specific case of annotations on 3D objects, Balsa et al. [47] defined a specific set of requirements related to annotation display. While the original work was motivated by a cultural heritage application, they can be generalized to the general case of annotated data display:

1. **Information spatially connected with 3D models.** Most of the information, textual and visual, is spatially connected to a region of a 3D model. This implies that descriptive information should not only be associated to parts of the associated objects, but also in relation with a viewpoint that permits the visualization of the part of the object containing annotations. Moreover, different macro-structural and micro-structural views should be associated with different kinds of information, showing a dependence on scale.

2. **Annotation presentation order.** Relations exist among the different information to be presented, which leads to the importance of deciding the order in which the information is presented. An inspection can thus be seen as a storytelling instance, in which the subsequent presentation of several annotations delivers information by "telling a story". The presentation order is often not strict, and different storytelling paths are possible.

3. **Annotation importance.** Not all the information has the same importance. While some descriptions are mandatory, others are more anecdotal and can be skipped in some presentations.

4. **Annotation authoring.** Textual and visual information (drawings, images) should be supported. Editing should be made possible without particular training, and adding annotations and linking them should not require intervention of specialized personnel. While any of the techniques introduced in subsection 2.1.2) may be used, for effective visualization purposes they should be combined with methods defining importance and ordering.

5. **Focus on analyzed object (avoid occlusion from interaction widgets).** The important information is the visualized object itself, which should not be obstructed by general clutter (e.g., interaction widgets or annotation display).

6. **Fast learning curve and assisted navigation.** In most applications for the general public, where *walk-up-and-use* interfaces are expected, the visitor experience could be easily frustrated if the proposed interaction paradigm does not allow them to immediately explore the content, through a natural user interface with an extremely short learning curve. Moreover, in cases that must manage large amounts of visitors (e.g., museums), long training times and/or guided training are not affordable. The user interface should thus be perceived as simple, immediately usable. These considerations are also applicable to applications designed for experts, since they are particularly the sophisticated / advanced users.

7. **Engaging experience.** For public installations, visitors do not want to be overloaded with instructional material, but to receive the relevant information, learn,

and have an overall interesting experience, which should be personal, self-paced, and exploratory. The user interface should provide guidance, while not being perceived as overly obtrusive.

8. **User interface and display flexibility.** Since a given annotated object is typically explored in different settings, one should support a wide range of setups. Information display should in particular be easily retargetable to different display sizes and aspect ratios.

9. **Seamless interactive exploration.** Following the Visual Information-Seeking Mantra, most control modes should be active with real-time feedback, in order to provide a sense of control, and support smooth and seamless object inspection, going back and forth from shape inspection to detail inspection, and from overview to zooming and filtering operations.

Annotation presentation techniques

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[58]. Annotation display falls in the category of visual-spatial displays that dynamically mix 3D representation with associated overlays.

Using linked multimedia information to enhance the presentation of complex data has been long studied, mostly focusing on guided tours [59], text disposition and readability [60], [61], usability of interaction paradigms [62], and the integration of interconnected text and 3D model information with bidirectional navigation [35], [36], [63]. Most of the CH data presentation tools also support the integration of interconnected text and model information with bidirectional navigation [6], [35], [36], [63], and several tools are emerging that also offer interfaces for multi-user annotation creation (e.g., Aioli [37]).

In most of these cases, pickable regions (points or surface areas) are displayed above the object and trigger an annotation display when selected, much as when clicking on hyperlinks in a web browser. Sometimes, e.g. in the bi-directional hyperlink system of Goetzelmann et al. [35], links present in the textual annotation can also refer to the object, and their selection triggers camera motion.

All these methods, however, require precise 3D picking to navigate through the information, thus presenting problems when targeting non co-located interaction setups (e.g., large projection displays), and often introduce clutter in the 3D view to display the pickable regions, especially when many annotations are present or they are extremely large. An alternative to picking are methods that use postures or gestures to trigger visualization of contextual information, e.g., in the form of contextual menus [64]. By using contextual selection, the active annotations below the cursor are only displayed on demand in the context menu, reducing clutter but losing the direct visual connection between annotation and 3D object.

Another solution to reduce clutter in annotation display has been recently introduced by Jaspe et al. [48] by adopting an interactive lens approach [26], whose main purpose

Figure 2.3: **Interactive lens.** At run-time, the user moves a lens that defines a local area in which annotations are displayed. In this case, in addition to linear annotations highlighting cracks, an alternate achromatic geometry view is displayed inside the lens. Left: without lens; Middle: lens with achromatic view; Right: lens with achromatic view and linear annotation layer (image from our JOCCH 2021 contribution [33])

is to support multi-faceted data exploration. In their approach, on top of a defined visualization mode for the 3D object, an alternative visual representation of a local area of the data including annotations is displayed. This method has proved successful for multi-field visualization, and has been so far applied for annotation display only in the context of relightable image visualization.

All the above techniques, however, 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. This is typically handled by defining annotation categories and explicitly enabling/disabling the visible categories in the interface [6]. In this thesis, I will significantly expand upon this research, by defining novel techniques to move lenses, display annotations, discover annotated areas, and receive guidance.

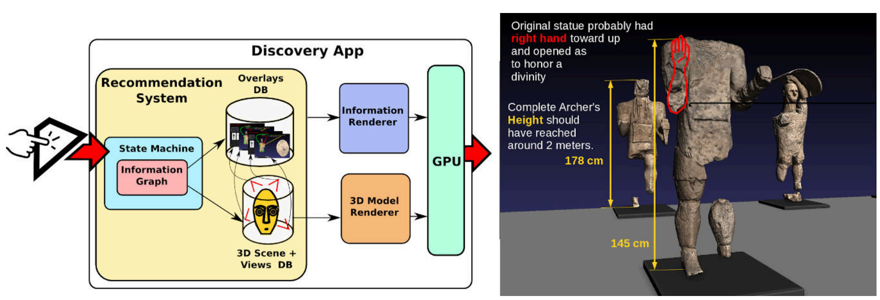


Figure 2.4: **Automatic recommendation approach.** At run-time, users navigate inside the 3D scene, while adaptively receiving unobtrusive guidance towards interesting viewpoints and history- and location-dependent suggestions on important annotations, which is adaptively presented using 2D overlays displayed over the 3D scene. Image reprinted from Balsa et al. [47])

An alternative solution has been introduced by Balsa et al. [47], which automatically selects only a single annotation at a time by exploiting the organization of annotation into an information graph with importance and precedence relations (see subsection 2.1.2). In order to provide an engaging self-paced experience, they let users freely explore 3D models using an interactive camera controller, while an adaptive recommendation engine based on a state machine runs in parallel with user interaction, and identifies which are the current most interesting information nodes, using a scoring system based on the previous history of visited nodes, the dependency graph and the current user viewpoint (see Figure 2.4). A suggestion is then stochastically identified among these candidate nodes, with a probability proportional to the score. The non-deterministic choice respects mandatory presentation orders, supporting classic authored story-telling, while introducing variations in the exploration experience. After a suggestion is taken or ignored. the information graph is updated, and a new suggestion is selected based on the new state. The so created story telling path is a non-linear dynamic exploration of the information graph.

Such an approach is very promising, but has only been applied to annotations linked to view-points, forcing the camera to snap to fixed positions for annotation display. In this thesis, as we will see, I significantly expanded on this research by switching from cameras to lenses.

2.2 Discussion and workplan

A wide variety of use cases and application domains require the ability to combine 3D objects with annotations linking regions of them to additional information. Many solutions have been proposed for the creation, organization, and display of annotations. For this thesis, within the domain of EVOCATION, I have looked, in particular, at the exploration of models with annotations. Open problems that I have identified within this area of research include

- How to ***define relations*** not only between objects and annotations, but also among annotations themselves in order to guide their presentation;
- How to ***define recommendation systems*** that exploit this graph representation to determine which annotation to display based on the current interaction context;
- How to ***better display annotations*** to avoid clutter;
- How to ***guide the user*** within an annotated world with a suitable interface.

In the following chapters, I will discuss the techniques that I have introduced to tackle this problem. First, I introduce the usage of visualization lenses to provide selective display in a focus-and-context setting (chapter 3), and a will define solutions that allow users to solely interact with the lens, without modal switches, to obtain a full multi-scale visualization of models. On top of this technique, I will then define a method for associating specific lens configurations to annotations in the database, and I will exploit this information to display one annotation at a time, to avoid clutter, and to guide the user towards the areas with the most-relevant, context-dependent annotations (chapter 4). Finally, I will organize annotations in a graph that defines not only the relations between the annotations and regions of the object of interest, but the precedence relations among annotations, allowing authors to define meaningful information paths, and the system to provide improved guidance within an interface that can mix-and-match guided tours with free exploratory navigation (chapter 4). Finally, I will show two simple examples in which I have applied these techniques in the cultural heritage pilot of the EVOCATION project (chapter 6).

2.3 Bibliographic notes

The major part of this chapter has been taken from my contribution to the Chapter 4 of EVOCATION Deliverable D3.1 titled as "*Scalable visualization techniques for large captured datasets beyond simple visual replication*" [30].

Chapter 3

Supporting focus-and-context exploration with visualization lenses

As a first contribution towards better interaction with complex multi-faceted objects, I introduce in this chapter a novel approach for controlling interactive lenses. Focus-and-context exploration is supported by translating user actions to the joint adjustments in camera and lens parameters that ensure a good placement and sizing of the lens within the view. This general approach, implemented using standard device mappings, overcomes the limitations of current solutions, which force users to continuously switch from lens positioning and scaling to view panning and zooming. While the method is generally applicable to general 2D visualization, it is presented and evaluated for the exploration of stratigraphic relightable models, which are extremely common use cases in the target domain of cultural heritage. A user study has been performed in order to validate our approach.

3.1 Introduction

Interactive visualization lenses are movable tools that provide alternative visual representations for selected regions of interest of a display. Due to their flexibility, they are among the most widely used techniques in scientific and information visualization [26]. In particular, they offer support to *overview+detail* (through a spatial separation in depth between the detail view in the lens and the overview outside it), *focus+context* (through the minimization of the seam between views), as well as *cue-based techniques* (thanks to the selective alteration of the visual representations) [65].

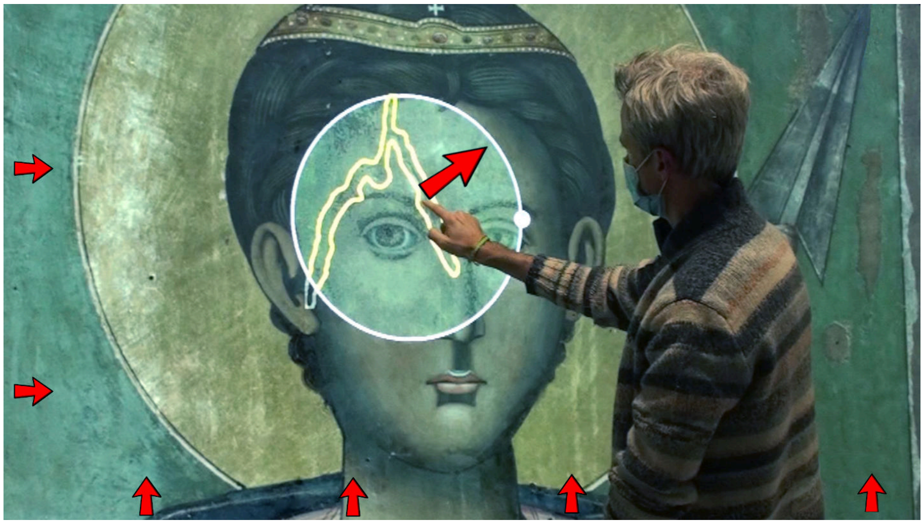


Figure 3.1: **Exploration of annotated models.** We introduce an approach for improving navigation with interactive lenses. The general control scheme simplifies focus-and-context exploration by jointly adjusting camera and lens parameters in response to user actions.

Research on lenses is extremely wide. Tens of different techniques have been presented for visualization, and far more in related fields, the vast majority targeting the design of the intended lens effect for solving specific visualization problems [26]. In this chapter, we seek, instead, to define user-interface mechanisms to support effective navigation strategies based on lenses.

Most real-world datasets typically have spatially-spread information that appears at different scales and can be presented in various ways. While camera and lenses are typically handled separately (see section 3.2), an effective multi-scale focus-and-context visualization imposes stringent constraints, which forces users to repeatedly perform complex combinations of control actions. The lens must not only be maintained visible within the current view, but it must also have a reasonable size in screen space, and should be surrounded by enough context [65]. Current user-interface solutions either assume that the view remains static during lens-based exploration, limiting the size and scale of the exploration area, or force users to find reasonable exploration conditions by continuously switching from lens positioning and scaling to view panning and zooming, thus increasing cognitive load. To overcome these limitations, we introduce a novel user-interface controller that maps user actions to the joint adjustments in camera and lens parameters that ensure a good placement and sizing of the lens within the view (section 3.3). This general approach, implemented using standard device mappings, is seamlessly integrated within a classic panning and zooming user interface. It makes it possible to perform detail analysis with a lens without distraction, as well as to use the lens for wide-area exploration.

While the introduced methods are generally applicable to any 2D visualization, our motivating application is in the cultural heritage domain, where it is essential to deliver informative and engaging real-time experiences to the general public, that cannot be overloaded with instructional material given within a very limited time span for using the system. In particular, we have implemented them for the exploration of stratigraphic relightable models. These models are very common in cultural heritage use cases. A user study has been performed in order to validate the basic principles of our approach (section 3.4).

3.2 Related work

Lens-based visualization has many aspects, and we refer the reader to the recent survey by Tominski et al. [26] for an extensive coverage of the domain. While most of the work on lenses focuses on the definition of particular lens functions, several authors have studied the problem of interacting with lenses, which is the focus of this work. Solutions, especially developed in the context of multi-touch interfaces include methods to create and delete lenses (e.g., with five-finger picks [66]), to manipulate the lens geometry (e.g., with pinch gestures [67]), or to parameterize the lens operations (e.g., by controlling zoom levels with pinch [68]). The manipulation of lens position and scale has been treated, so far, especially related to lens magnification, by introducing high-precision control [69] or hierarchies of focus regions [70]. We introduce, instead, new ways to jointly control the relative positioning and scaling of the focus and context areas.

While the techniques presented in this work are of general usage, and can be applied to various multi-faceted 2D exploration tasks, we have focused our implementation on the special case of stratigraphic relightable models. In the last decade, a wide variety of tools have been proposed for targeting either static exploration of multi-faceted or multi-light image data [18]–[21], or dynamic exploration through relighting [9]. In this context, lens-based interaction with such models has been used previously by Jaspe et al. [33], [48] solely with the purpose of letting users see different layers inside or outside the lens. By contrast, this chapter proposes novel techniques to move the lens for free navigation. We will show in the next chapter how the method can be extended for assisted or automated motion, and can be combined with annotation display.

3.3 Focus-and-context lens and camera control

Interactive lenses maintain visual attention in the interior of the lens, emphasizing the data analyzed by the user. The surrounding base visualization serves as context: it helps users understanding relations between the altered and the base visualization, and provides spatial information to support location awareness while navigating. To be effective, such a visualization must thus respect several constraints. In particular, the lens must be large enough to show a good amount of data inside it, but at the same time small enough to allocate screen space for displaying enough surrounding context to interpret the lens content and avoid getting lost in the dataset. Having surround space also permits lens motion in the neighborhood, to profit from spatio-temporal cue changes. In order to keep the lens relatively well centered in the view, and not too big or too small in screen space, users must repeatedly switch between camera control and lens control, thus increasing cognitive load. In the following, we introduce a joint controller that allows users to better concentrate on their analysis task by automatically adjusting camera and lens parameters in response to user actions in real-time.

3.3.1 Control scheme

Our joint camera and lens controller, see Figure 3.2, evolves as a state machine responding to user events, using the mapping described in subsection 3.3.3. Its behavior is as following:

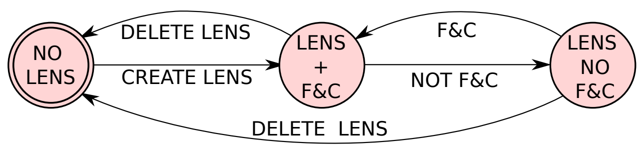


Figure 3.2: **State machine for joint camera and lens control.**

(S0): No lens At the beginning, the application starts without a lens, and all the user actions are enacted on the camera, letting users pan and zoom across the model. When the user activates a lens, the controller automatically ensures that the focus-and-context condition is met (subsection 3.3.2), and changes its state to *(S1)*.

(S1): Focus-and-context condition verified When the lens is moved or scaled, the parameters of both the camera and the lens are adjusted to ensure that we remain in the focus-and-context condition (subsection 3.3.2). If, instead, the camera is moved, the controller checks if the focus-and-context condition is violated after the motion, and, if so, changes the state to *(S2)*. Lens deletion simply removes the current lens and changes the state to *(S0)*.

(S2): Focus-and-context condition not verified The motion of the camera and the lens are mostly decoupled, as in common user interfaces mappings. So, panning and zooming with the camera simply updates the view without changing the object-space position and scale of the lens; moving the lens changes its object-space position, and the camera is only adjusted when it is needed to keep the lens in view. After updating the camera or the lens, the controller checks whether the focus-and-context condition is now met and, if so, changes the state to *(S1)*. Lens deletion simply removes the current lens and changes the state to *(S0)*.

3.3.2 Joint camera- and lens-parameters adjustment

At the core of our technique is the detection and enforcement of a focus-and-context condition. Given a lens of radius r placed at a position (x, y), a change in the relative

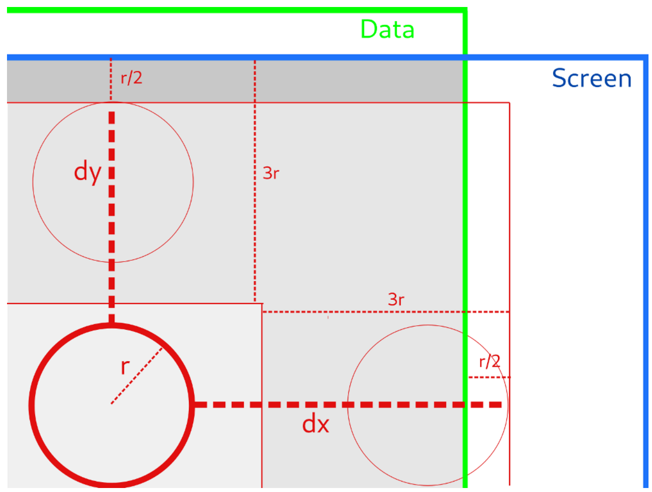


Figure 3.3: **Joint camera and lens parameter adjustment.** The motion of the lens is subdivided between motion of lens and motion of camera based on the amount of context available, as indicated by dx and dy, and the direction of motion.

positioning or scaling of the lens with respect to the view can be obtained either by directly moving the lens, or by applying the inverse of the same change to the camera. In our controller, we smoothly transition from camera control to motion control based on the available amount of context.

First of all, we seek to have a lens which is not too small or too big with respect to the current view, as measured by the size in pixels of the smallest length between viewport-width, viewport-height, viewed-dataset-width, and viewed-dataset-height. Therefore, we adjust the camera and not the lens if the scaling causes the lens radius to be smaller than 10% or larger 20% of that size.

We then take into account the distance from the boundary to verify whether we need to adjust and compensate for a missing context. We start by measuring the horizontal and vertical distance to the visible context boundary resulting from just moving the lens (see Figure 3.3). For each of the directions, this distance is the smallest between the distance to the viewport boundary and the distance to the dataset border expanded by an amount $\frac{r}{2}$ in screen coordinates. This expansion takes into account that users might want to explore up to the boundary of a dataset even though there is no visible context across the boundary. We then consider, independently for each direction, how to subdivide the requested change in parameters between camera and lens. If the change is in the direction of increasing the context, i.e., away from the boundary, all the changes are applied to the lens. If, instead, the motion is towards the boundary, we consider that, after the requested translation or scaling, at least a context of dimension of half the radius of the lens should be preferably maintained to provide the user with enough information around the lens to help with data interpretation. Thus, if the distance to boundary falls below that value, all the change requested is applied to the camera. If, instead, there is a large amount of context available ($d > 3r$ in this thesis chapter), all the change for that direction is applied to the lens. For the in between values, i.e., $d = r..3r$, we proportionally apply the change to both the camera and the lens. To apply uniform scaling, we average the independent solutions and clamp the result to guarantee that we do not exceed the allowable distance to the boundary. We then apply the same scaling to both dimensions.

With this approach, the same input has a result that smoothly varies from lens control to camera control and, if the user starts in a good focus-and-context condition (i.e., $d \geq r$), it is guaranteed that the focus-and-context condition is also valid after motion.

3.3.3 User interface and device mapping

Our user interface for joint lens and camera control requires minimal user input, and can be mapped to input devices in a variety of ways (see Figure 3.1). In our current implementation, we realized both a multi-touch solution and a mouse-controlled version. Lens creation is triggered by a long press (or a center mouse button click) at the point in which the lens must be initialized. Lens deletion is activated by long press (or a center mouse button click) inside the current lens. Panning the camera or moving the lens is achieved by a one-finger pan gesture (or by dragging with the left mouse button pressed), differentiating whether we intend to control the lens or the camera by the position of the cursor at the beginning of the gesture. Scaling the lens or zooming the camera works similarly, using the pinch-to-zoom gesture for the multi-touch interface, and the mouse wheel or a up/down right button drag for the mouse control version.

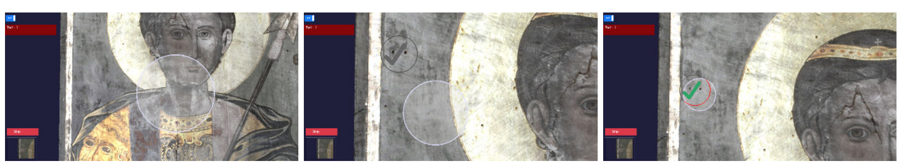


Figure 3.4: **Lens control user interface evaluation.** Participants were asked to find, as quickly as possible, small annotations made on the model, using a small image of the surroundings of the target annotation as the only guidance *(left)*. When the user-controlled lens is in the neighborhood of the annotation, a target lens is displayed over the annotation *(middle)*. The task is accomplished when the users places its lens over the target *(right)*.

3.4 Implementation and results

A reference system integrating all techniques described in this chapter has been implemented on a web-based platform. Stratigraphic relightable image preparation is done off-line and results in a repository containing a set of image layers and a configuration file that describes the arrangement of layers. The data is made available by a standard web server to a web client running in a browser on top of WebGL2, a JavaScript API that closely conforms to OpenGL ES 3.0 and can be used in HTML5 <canvas> elements without requiring plugins. The viewer is used both for annotating models and for exploring them.

We have extensively tested our system with a number of complex heterogeneous datasets. In this thesis, without loss of generality, we demonstrate its usage on a simple painting use case stemming from the cultural heritage domain. The videos are available together with the original publication [22].

The painting use-case concerns the exploration of relightable stratigraphic model of the *Icon of St. Demetrios* (17^{th} - 18^{th} century), see Figure 3.1 left, containing a normal map and six color layers (visible, 2xIR, 2xUV, FC) generated from a multispectral RTI acquisition. A total of *33* annotations describe various damages (in particular cracks, woodworms, paint defects) and artistic/decoration details. The model was acquired by the CRS4 team at Ormylia Foundation during a previous European Project (Scan4Reco - EU H2020 grant 665091). We thank Ormylia Foundation for the access to the artworks for the purpose of digitization and for annotation information.

In order to provide a preliminary assessment of the effectiveness of our approach, we designed and carried out a user study focused on the proposed novel interaction capabilities.

For the user analysis, *25 participants* (*14 males* and *11 females*, with ages ranging from *11* to *69*, median *41 years*) were recruited among students, families and friends of researchers working at our center. All subjects had normal or corrected to normal vision and, as now extremely common, had basic computer or smartphone literacy.

Our first user evaluation focuses on the control scheme for jointly interacting with lenses and cameras, without any reference to an underlying annotation database.

The main goal of the evaluation was to assess whether the proposed joint camera controller provides advantages with respect to the classic controller in which the lens and the camera are separately controlled, in which actions outside the lens move the camera, and actions within the lens move the lens. In the following, our control scheme is identified with *LC* (for lens+camera), while the standard scheme is identified by *STD*.

3.4.1 Setup

The experimental setup considered the reference system implementation described above. In order to reduce variability of results, we limited the comparison to the interface operated with mouse control, using the web-based implementation on desktop or laptop platforms. The testing model was the painting dataset, which has a lot of visual and geometric details spread over the entire image.

3.4.2 Tasks

The experiments consisted in letting users try the two different manipulation controllers in the context of a target-oriented user interaction task [71]. We designed our task to measure performance for the macro-structure and micro-structure inspections tasks typical of cultural heritage model explorations (see Figure 3.4). Participants were asked to find, as quickly as possible, small annotations made on the model, with the help of only an image of the surrounding of the target annotation. When the user-controlled lens is in the neighborhood of the annotation, a target lens is displayed over the annotation. The task is accomplished when the user places its lens over the target lens.

3.4.3 Design

Users were first allowed to become familiar with the two controllers by watching a brief video showing how they work. Then, each participant used the two interfaces in randomized order. The test started with a short training session, in which the user could familiarize with the interface and performed one task freely without it being scored. After the training session, the measured tests consisted of five trials, where targets were randomly selected from a list of 20 potential candidates, so as to avoid any bias due to a-priori knowledge of target positions. Including training, users dedicated less than 5 minutes to complete the evaluation.

In order to measure and quantify the perception of usability, the participants were also asked to fill a System Usability Scale (SUS) questionnaire [72], a simple ten-item Likert scale form with five response options for respondents (from *Strongly agree* to *Strongly disagree*). The questions are related to (Q1) desired frequency of use; (Q2) perceived complexity; (Q3) perceived ease of use; (Q4) perceived need for support; (Q5) integration of functions; (Q6) perception of inconsistency; (Q7) possibility of using it without training; (Q8) perceived interface complexity; (Q9) confidence in using it; (Q10) and perceived quantity of information needed. As identified by Lewis and Sauro [73], Q4 and Q10 provide indications on learnability. while the other questions provide indication on usability.

All the tasks and filling of questionnaires were autonomously performed by the users, without supervision, by accessing web forms.

3.4.4 Performance evaluation

For completing their trials, users needed times ranging from 28s to 4min46s (median 1min16s). Before collecting the results, we expected our controller to be faster, due to the joint control of camera and lens.

Figure 3.5 shows the boxplots of the task completion times. The bottom and top of each box are the first and third quartiles, the band inside the box is the second quartile (the median), and the ends of the whiskers extending vertically from the boxes represent the lowest datum still within 1.5 IQR (inter-quartile range) of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. Outliers are indicated as small circles. The analysis of results reveals that, independently from the expertise, the LC controller appears significantly faster and more stable than the standard approach of alternatively moving camera and lens. The median completion time for all users

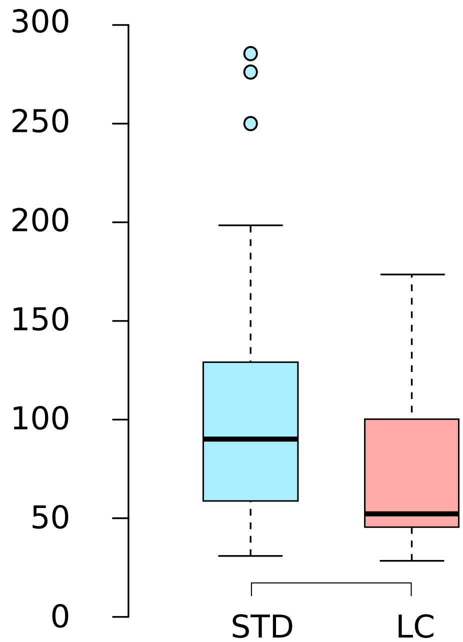


Figure 3.5: **Performance evaluation.** Our controller (LC) was compared to the standard separate controller for camera and the lens (STD). The graphs show the time in seconds used to complete the task consisting of 5 target-positioning trials. A total of 25 users were evaluated. In the boxplots, center lines show the medians, box limits indicate the 25th and 75th percentiles as determined by R software, and whiskers extend 1.5 times the inter-quartile range from the 25th and 75th percentiles, while outliers are represented by dots.

using the standard interface is 90.08s, against 52.34s for LC (42% improvement). The analysis of the IQR range and outliers also reveals that LC provides a more homogeneous performance (see Figure 3.5). A one-way analysis of variance (ANOVA) further confirms that there was a significant effect on completion time at the $p < 0.05$ level for the two interfaces [$F(1,48) = 4.047$, $p = 0.0499$].

Direct observation of user behavior indicates that in several cases, when using the standard interface, the lens had to be picked and re-centered manually multiple times, as it tended to leave the field of view. The fastest users, when searching for targets far from the current location tend to quickly zoom out to see a larger area of the object, and then zoom in to reach the target, both with the modal interface and our controller, while most users tend to analyze the object at a smaller scale using longer panning motion. A possible improvement in our interface might thus be to incorporate speed-dependent zooming.

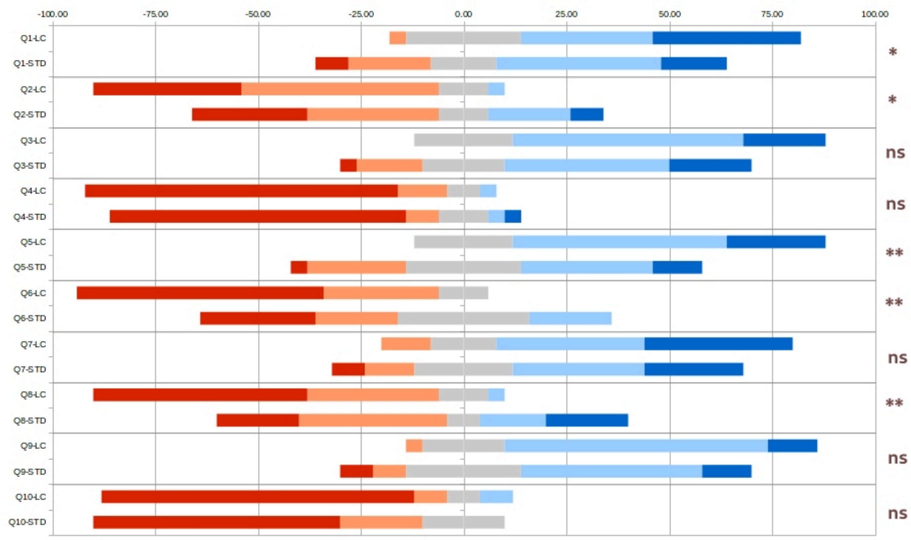


Figure 3.6: **Usability evaluation of lens control.** Diverging stacked bar charts of SUS questionnaire responses concerning our controller (LC) and the standard controller (STD). The color scale goes from red (strongly disagree) to blue (strongly agree). The labels near the right axis summarize the per-question statistical significance resulting from ANOVA ($ns \rightarrow p > 0.05$; $\star \rightarrow p \leq 0.05$; $\star\star \rightarrow p \leq 0.01$).

3.4.5 Usability evaluation

By analyzing the responses of the SUS questionnaires, summarized in the bar charts of Figure 3.6, we obtain for our joint controller a SUS score of 79.6, which, according to standard practices [74], rank the results as good. By contrast, for the standard controller splitting camera and lens motion, we obtain a significantly lower SUS score of 65.7. The ANOVA results are reported in Table 3.1. They confirm that there was a significant effect on SUS score at the $p < 0.05$ level for the two interfaces [$F(1,48) = 7.035$, $p = 0.011$]. ANOVA on answers to individual questions revealed that there was a very significant effect on the perception of integration and consistency (Q5 and Q6 with $p \ll 0.01$). Moreover, users perceived the standard method much more cumbersome than our joint controller (very significant effect on Q8 with $p \ll 0.01$) and also more complex

LC vs. STD	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	SUS
F(1,48)	4.407	4.267	2.335	0.333	8.544	11.977	1.898	10.839	2.429	0.226	7.035
p	0.041	0.044	0.133	0.567	0.005	0.001	0.175	0.002	0.126	0.637	0.011
Significance	$\star$	$\star$	ns	ns	$\star\star$	$\star\star$	ns	$\star\star$	ns	ns	$\star$

Table 3.1: **Usability evaluation of lens control.** Comparison of our method (LC) with the standard disjoint controller (STD) using a one-way ANOVA on responses to SUS questionnaires. The last row summarizes the per-question statistical significance resulting from ANOVA ($ns \rightarrow p > 0.05$; $\star \rightarrow p \leq 0.05$; $\star\star \rightarrow p \leq 0.01$).

(significant effect on Q2 with $p < 0.05$). This is very likely due to the frequent modal switches imposed by the decoupled controller, which lead to confusion. The preference for our controller is also reflected by the significant effect on the desired frequency of use (significant effect on Q1 with $p < 0.05$).

3.5 Discussion

We have presented an enhanced interaction controller that helps interactive exploration of a model with a lens by providing a mapping, mediated by an interaction metaphor, that meaningfully links user actions on the inside or outside of the lens to coordinated camera and lens motions that support focus-and-context exploration. Our evaluation of this aspect of the interface shows that the method appears to be well received and intuitive for casual users, making exploration times shorter, especially when inspecting an object at multiple scales, independently from the presence of annotations.

While our evaluation targeted a particular data kind (stratigraphic relightable models) coming from a single domain (cultural heritage), our methods are general enough to be readily applied to other information visualization using lenses on a variety of 2D datasets. An important avenue of future work will be to extend them also to more general 3D visualization. A particularly promising solution would be to explore their combination with decal lenses [75], which act on patches of 2D manifolds built to attach smoothly to non-flat surfaces. A promising solution would be to extend our approach to 3D by sliding and scaling these patches around the surface while maintaining enough context visible.

3.6 Bibliographic notes

Most of the content of this chapter was presented in our EUROVIS 2022 contribution and published in the Computer Graphics Forum journal [22], which also includes the follow-up work on annotations presented in the next chapter. Our EUROVIS talk is publicly available here, and includes demonstration video and further related content. I have significantly contributed to the conceptualization, methodology, and validation of the method and was one of the primary authors of the paper. An early approach for using lenses together with stratigraphic models was presented in our JOCCH 2021 contribution [33]. That work, to which I contributed for the evaluation and data preparation, allowed users to visualize multiple layers, but used only standard means for lens control.

Chapter 4

Assisted exploration of annotated models using interactive lenses

The previous chapter has focused on methods for translating user actions to joint camera and lens adjustments, with the goal of simplifying focus-and-context exploration of general stratigraphic models. Here, we assume that the explored models are annotated, and tackle the problem of annotation discovery and display. In order to provide guidance, in addition to traditional visual markups and information links, we associate to each annotation a lens configuration that highlights the region of interest. During interaction, an assisting controller determines the next best lens in the database based on the current view and lens parameters and the navigation history. Then, the controller interactively guides the user's lens towards the selected target and displays its annotation markup. As only one annotation markup is displayed at a time, clutter is reduced. Moreover, in addition to guidance, the navigation can also be automated to create a tour through the data. The capabilities of our approach are demonstrated through a user study in a cultural heritage use case.

4.1 Introduction

In chapter 3, we have shown how users can explore a stratigraphic dataset at multiple scales by mapping user actions to the joint adjustments in camera and lens parameters that ensure a good placement and sizing of the lens within the view (Figure 3.1). Such an approach, implemented using standard device mappings, can be seamlessly integrated within a classic panning and zooming user interface, and makes it possible to perform detail analysis with a lens without distraction, as well as to use the lens for wide-area exploration.

Figure 4.1: **Exploration of annotated models.** We introduce an approach for improving navigation with interactive lenses. *Right*: Knowledge of an authored annotation database with pre-computed lenses guides users towards interesting regions through an unobtrusive interface. *Left*: guidance is provided by selecting target lenses based on a relevance score computed from the current lens position, camera parameters, and navigation history.

We further improve navigation with lenses by exploiting and extending the concept of *data annotations* to provide *guidance* (section 4.3). Guidance is a process where the system provides assistance in response to information on user needs [76]. While existing approaches mostly guide the interpretation of visualizations [27], we focus here on assisting users in discovering interesting areas while navigating with the lens. In this context, we assume that the data under inspection has been enriched with visual annotations that mark and describe the areas of interest in the dataset [24]. Such visual cues, which can come from automated analysis or manual mark-up processes, are known to make data understanding easier for the viewer [23]. Finding relevant annotations, and presenting them in a comprehensible way without cluttering the display, however, is very challenging [25].

In our approach, we associate to each annotation a lens configuration that highlights the region of interest. The stored information includes the lens location and rendering parameters that were used to inspect the region while creating the annotation. During the interaction, a recommendation system determines in background the next best recorded annotation as a function of the current camera position, lens parameters, and navigation history. The user is then interactively guided towards that annotation in different ways, depending on the situation. Only a single context-dependent annotation is selected at a time in order to reduce the amount of clutter. Moreover, at annotation display, the current lens parameters, and the dependent viewing context, can be smoothly adjusted towards the pre-recorded ones, leading to the automatic selection of the best visualization mode. Finally, in addition to assistance, the navigation can also be automated to create a tour through the data.

As for the methods presented in chapter 3, the techniques introduced in this chapter are, in principle, generally applicable to any 2D visualization, but have been implemented and tested for the typical stratigraphic relightable models common in cultural heritage use cases, for which the tunable parameters include the visualized layer, its rendering mode, and the illumination environment. A user study has been performed in order to validate the basic principles of our approach (section 4.4).

4.2 Related work

Annotations are mechanisms that link a sub-portion of a geometrical representation of an object to some related information not present in the object itself [24]. While annotation markers can be placed on surfaces in many ways, including labels [34], [60], [61] or hot-spots [6], [24], [36], our focus is on visual image overlays, which augment the annotated regions with text or drawings draped over the surface. This 2D representation offers direct spatial association with the annotated region, and is very common even for 3D models, since it is much easier to select the annotation on projected 2D media than on 3D objects themselves [37], [77]–[79]. As objects have typically many different spatially-associated annotations, special care should be taken to decide when and how the information is presented, in order to avoid clutter and cognitive overload. In addition to letting users explicitly enabling/disabling categories in the interface [6], the techniques proposed in the literature deal with overcrowded displays by modifying the appearance (e.g., filtering data or using variable opacity), distorting the image (e.g., zooming), or using space-time trade-offs (e.g., using serial temporal presentation) [80]. In our approach, we use both temporal and appearance modification techniques, by selecting one annotation at a time and exploring it with a lens. Lenses have also been classically used to reduce congestion (e.g., by using sampling inside the lens to reduce clutter in a local area[81]), but not for overlay images draped over surfaces. Jaspe et al. [33], [48] also used lenses, but assumed non-overlapping annotations. In our context, the automatic selection of annotations also provides navigation assistance through user guidance.

Guidance approaches are based on the assumption that intelligent services and users may often collaborate efficiently to achieve the user's goals. Starting from research in human–computer interaction [82], [83], guidance has more recently targeted the support to users during interactive visual analytics work [27]. Ceneda et al. [76] provide a full characterization of the domain and highlight how existing approaches mostly support the interpretation of visualization. Our technique, instead, aims to assist direct interaction during an analysis task. We do so by combining camera and lens motion to support focus-and-context exploration, and by suggesting or directing users towards previously annotated areas, thus providing both prescribing and directing guidance [76].

A number of authors have proposed to manually or automatically compute interesting viewpoints in order to guide users towards areas of interest within their data. While some solutions use these viewpoints to aid camera control [84]–[86], others focus on creating animated paths, by arranging viewpoints into graphs [87] or letting users define video-tours [88]. None of the previous approaches target lenses. The camera-control work of Balsa et al. [47] is the most similar to ours, as it selects only a single item at a time from a viewpoint graph. Selection is based on a score that extends to viewpoints the Degree-of-Interest (DOI) concept introduced by Furnas [89] for trees and extended by van Ham and Perer [90] to graphs. Similarly to Gladisch et al. [91], DOI computation also takes into account past behavior. In a different context, we also use a scoring system with a history term to help navigation. Our work, however, does not use a graph of views, augments the annotation database with lenses and rendering attributes, and introduces specialized scoring functions targeting lens navigation.

Figure 4.2: **Annotation selection.** Annotations with annotated lenses cover the dataset with a lot of overlap (left). At run-time we rank the annotations based on a similarity computation with the current lens and view (middle), and select the best annotations based on the assigned score. If the selected annotation is close enough to the current lens, it is immediately displayed (right), otherwise it is suggested to the user, who can accept or reject the suggestion. In the middle image, lenses associated to individual annotations are color-coded white to red based on the score computed for the lens in the right image.

4.3 Assisted and automatic navigation in an annotated model

In addition linking camera and lens motion, we further improve navigation by exploiting and extending the concept of *data annotations* to support assisted and automatic navigation.

Traditionally, annotations let users identify specific regions, visually mark them with overlay text or drawing, and link them to metadata or other information that characterizes those regions [24]. In the work presented in this chapter, we exploit annotations coming out of a user-driven analysis for guidance and data presentation. Our aim is to let users explore an annotated scene by just controlling the lens at their own pace, while the system supports them in finding annotated areas and in presenting annotations without cluttering the scene. This is achieved by running, in background, an assisted navigation system that selects the single next best annotation in the database, based on the current viewing parameters and the navigation history, and presents it in context-dependent ways (see Figure 4.2).

In the following, after summarizing the structure of our annotation database (subsection 4.3.1), we illustrate how we select at run-time the next best annotation and lens based on a similarity computation (subsection 4.3.2). Then, we discuss how this similarity is used to drive the guiding controller for assisted or automatic navigation (subsection 4.3.3). Finally, we discuss the user interface and device mapping realization (subsection 4.3.4).

4.3.1 The annotation database

In order to support navigation, each annotation stores, in addition to the *visual overlay* and the *external annotation description*, also the parameters that should be used for an effective lens-based exploration of the annotated area. This information consists in an *annotation importance*, a *lens and context area description*, and a set of *rendering parameters*.

The *importance* is a user-determined scalar weight. Annotations with larger importance values are more likely to be displayed. The *lens and context area description* geometrically determines the initial viewing setup for exploring the annotation area. It consists in the position and size of a lens and of its context area, i.e., the viewing rectangle used at annotation creation. *Rendering parameters* describe, instead, the visualization inside and outside the lens. For this thesis, targeting relightable stratigraphic models, these parameters include the light configuration and the layers that should be activated inside and outside the lens to ensure the visual representation most suited for understanding the annotation. All the indications are optional. Omitting one of them means that understanding does not depend on the recorded lighting or displayed layers. Note that this aspect is the only one strictly targeting stratigraphic images, and, in a more general context, might be replaced and extended by a wider definition of the data facet that must be displayed (e.g., defining parameters for data extraction in a multi-field dataset).

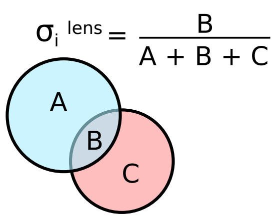
Authoring details are orthogonal to our method. For the sake of completeness, we mention here that we annotate our models by using the viewer itself, controlling the lens using the methods in section 3.3 to identify the interesting area, and drawing the annotation with a simple image editor. The lens and context area description and rendering parameters stored with the annotations are extracted from the viewer's state at annotation time.

4.3.2 Finding the next best annotation and lens

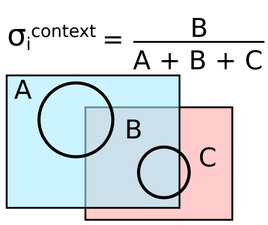
The selection of the next best annotation to display has to take into account three different concepts. First of all, the algorithm should be favoring annotations that are close to the current lens, not only in terms of position and scale but also of presented content, in order to permit the seamless presentation of annotations under the lens during user controlled motion and limit the amount of visual and semantic changes that would be caused by changes in presented layer as well as by large modification of overall position and scale. Second, we should take into account authoring information, by favoring annotations marked more important by the user with a higher priority. Finally, the algorithm should take into account the navigation history, in order to avoid repeatedly presenting the same information over and over again if other information is available. This is particularly important for the target application in which user engagement is paramount. We achieve these goals by assigning to each recorded annotation i a score $S_i = \gamma_i \sigma_i H_i$, where γ_i is the author-defined annotation importance, σ_i is the similarity score depending on spatial and semantic distance (see section 4.3.2, and H_i is the history score depending upon the activity log of the active user (see section 4.3.2). The next best lens is the one with the largest S_i.

Spatial and semantic similarity score

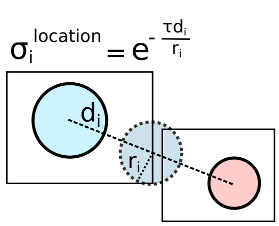
Navigating through visually annotated details in multiscale visualizations requires a trade-off among several conflicting criteria. In order to reduce travel times and foster continuity of exploration, we should prefer annotations that are present in the surrounding of the current lens over annotations that are far in terms of position and scale. At the same time, we should favor annotations that are similar in content or presented data facet over annotations that force a semantic change. We tackle the problem by defining a similarity score γ_i that compares the current lens with a lens i in the data, considering both purely geometric factors and semantic criteria.



Lens overlap If during exploration the current lens hovers over a lens i in the database having the same scale, we should favor the selection of the associated annotation, as the user is already optimally placed to explore it. We thus set the lens similarity term σ_i^{lens} to the Jaccard Similarity (a.k.a Intersection over Union (IoU) metric) between the current lens and the stored lens for annotation i. This value will be non-zero only in case of overlap, and will take its maximum for matching lens size and position.

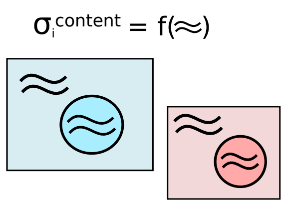


Context overlap Intuitively, selecting a lens that requires small changes in the camera position or scale to preserve good focus-and-context conditions should be favored. Such a choice would preserve locality even when lenses are not overlapping. To take into account this fact, we compute the context area determined by our focus-and-context approach when moving the current lens to the position and scale of lens i, using the constraints described in subsection 3.3.2. The current and target contexts are two rectangles in world space coordinates, determining the currently displayed area and the area that will be imaged when moving to position i. We then set the context similarity $\sigma_i^{context}$ to the Jaccard Similarity between these two rectangles. This measure is 1 for totally matching rectangles (i.e., the camera won't move if we select lens i), 0 if the two rectangles do not overlap (i.e., the camera will view a totally different area of the dataset when selecting lens i), and grows from 0 to 1 proportionally to the amount of overlap normalized by the union of current and target context pixels. Such a measure provides thus an indication of visual change.



Location similarity The context similarity measure $\sigma_i^{context}$ returns, by design, a constant score for all lenses i very close or very far to the current lens, since the context will either remain unchanged or will be without overlap. In both extreme cases, however, it is reasonable to favor close lenses to far ones, since moving to a closer lens favors locality and reduces travel time and visual change. Thus, we introduce a location similarity score, $\sigma_i^{location}$, which provides a smoother variation of score as a function of distance between the current lens and the target lens i. Moreover, such a measure should be scale-dependent, since traveling long distances with small lenses requires more time and produces more discomfort than with large lenses due to loss of context.

Thus, we define $\sigma_i^{location} = e^{-\tau \frac{d_i}{r_i}}$ where d_i is the world-space distance from the current lens to lens i and r_i is the average between the current lens radius and the radius of lens i, and τ is a scaling constant ($\frac{1}{10}$ in the work presented in this chapter). Intuitively, this measure takes the maximum at 1 when the lens does not move, and decreases as a function of the traveled distance in terms of lens radii, which is a measure of visual change during animation.



Content similarity While the three above measures concern geometric changes, the content similarity measure $\sigma_i^{content}$ indicates the change that will occur due to semantic changes in the areas inside the lens and outside the lens. This measure is application-dependent. Since in this thesis we target annotated relightable stratigraphic models, we consider that there is a significant change if, when moving to target annotation i we must change the layer or the annotation class. We compute weights for affected areas as $w = \frac{area_{currentlens}}{area_{currentcontext}}$, and set $\sigma_i^{content} = (1-w)s_{in} + ws_{out}$, where s_{in} and s_{out} are zero if a change inside their affected area occurs and one otherwise.

We finally compute the total similarity score σ_i as a normalized weighted sum of the individual similarity components. Currently we use unit weights for each component.

History score

The recommendation system should favor the selection of annotations that have not recently been proposed to the user to avoid repetitions, but should still consider them as an option in case no more information is present, or local information is exhausted and a very large travel is needed to move to other annotated areas. We implement this concept by introducing a history score H_i, which smoothly varies over time as a function of past user behavior.

In order to define a smooth variation of scoring factors, we employ the smoothstep function $S_1(x, x_0, x_1)$, which returns 0 if $x \leq x_0$, 1 if $x > x_1$, and performs smooth Hermite interpolation between 0 and 1 when $x_0 < x < x_1$. For shaping the temporal behavior of the system, we also define the fading function $F_1(x, x_0, x_1, x_2)$ which returns $1 - S_1(x, x_0, x_1)$ if $x \leq x_1$ and $S_1(x, x_1, x_2)$ if $x > x_1$. The function has a value that starts at 1, smoothly decreases to 0 when $x > x_0$, and then raises again to 1 for $x > x_1$.

In particular, we define $\Delta t_i^{presented}$ as the time that has passed since the last time the annotation i has been displayed, $\Delta t_i^{rejected}$ the last time it has been presented but not accepted. We then define $w_i^{presented} = F_1(t_i^{selected}, t_0, t_1, t_2)$ to control the priority for selecting the annotation i based on when it has been last displayed. If it has never been displayed, or has been displayed extremely recently ($t < t_0 = 5s$), the priority is maximum, while it smoothly lowers until some time has passed ($t < t_1 = 30s$), after which we consider that the user might have forgotten it and the priority starts to rise again, reaching maximum value to ($t < t_2 = 1m$). We also define $w_i^{rejected} = F_1(t_i^{selected}, \varepsilon, \varepsilon, t_2)$ to control the priority for selecting a rejected annotation. In this case, the priority instantly goes to zero, since we don't want to re-propose immediately a rejected annotation. The history score thus becomes $H_i = w_i^{presented} w_i^{rejected}$.

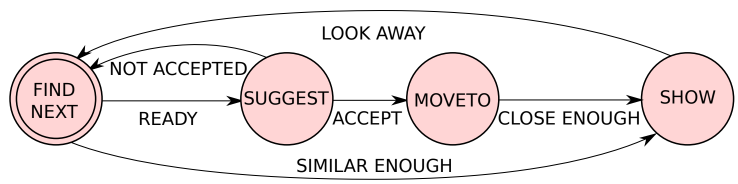


Figure 4.3: **State machine for assisted navigation in an annotated model.**

4.3.3 Assisting navigation

Our assisted navigation system, see Figure 4.3, is activated whenever a lens is active. When no annotation is currently displayed or when the user explicitly asks for suggestions, the system applies the method of subsection 4.3.2 to find the next best annotation. If such an annotation exists, it marks it as the next possibly displayable annotation. If the lens stored with the annotation is sufficiently similar to the current lens, it is immediately presented to the user by activating its display. We consider the lens sufficiently similar for immediate display if $\sigma^{context} > 0.9$ and $\sigma^{content} = 1$. This approach allows the system to seamlessly activate the display of the annotations under the lens while the user is moving. Otherwise, it is considered as a suggestion, i.e., a signal to the user that he could control the lens to find something potentially interesting in the suggested direction of change of the lens parameters. The suggestion is presented to the user only if the user has requested it or sufficient time has passed since the last time a suggestion was made. Such an automatically generated suggestion can be accepted by the user or rejected/ignored (see subsection 4.3.4). The time between successive automatically generated suggestions is controlled by the user behavior. Every time the user accepts a suggestion, we consider it helpful, and, thus, reduce the time without suggestions. Conversely, every time the user rejects a suggestion, the time to wait for the next suggestion to be presented is increased, as the user is considered less interested in receiving suggestions. This is achieved by setting the time between suggestions to $t_{wait} = median(t_{min}, t_{max}, t_{wait} * \alpha)$ where $t_{min} = 10s$, $t_{max} = 60s$, and α is 1.2 for rejection and $\frac{1}{1.2}$ for acceptance.

4.3.4 User interface and device mapping

Assisted navigation based on annotations must augment the user interface and device mapping of subsection 3.3.2 to handle information coming from the recommendation system that runs in background.

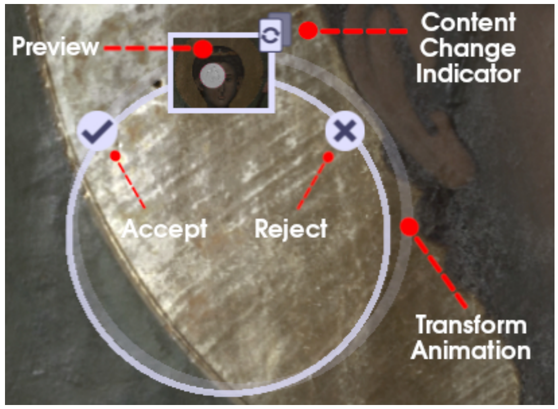


Figure 4.4: **Lens with suggestions.** During suggestion presentation, accept/reject buttons and indications of content and direction of changes for target lens are presented.

When the next best annotation is judged to be sufficiently similar (see subsection 4.3.3), and thus also close in position and scale, the rendering parameters of the current lens are, if needed, changed to the target ones, and the recorded overlay is displayed, with an additional transparency outside the lens.

Handling suggestions requires supporting the display of hints and the expression of acceptance or rejection (see Figure 4.4). A suggestion must indicate that some important information can be found by scaling and/or moving the lens in a particular direction, as well as eventually changing layer or rendering parameters. The target position, scale, or other parameters are those of the lens stored with the selected annotation. In order to guide towards them, we simply display a small semitransparent animation that shows the current lens boundary starting to move towards the target. The animation area is kept small (10% of the radius of the lens), so as to provide a hint without being too intrusive if the user wants to ignore it. In addition, a small icon on the lens boundary shows the target area of the annotation. If a significant change in rendering properties is required (i.e. $\sigma_i^{content} < 1$, a small glyph is also displayed. Moreover, two small accept/reject buttons are also displayed on the lens area. Such a suggestion indication stays visible until it is accepted, rejected, or ignored for a given amount of time.

We also offer users a gestural interface for accepting suggestions by launching the lens towards the target, in addition to clicking on the visible accept/reject buttons. If the user quickly moves or scales the lens in the direction indicated by the suggestion, the suggestion is considered accepted. If the total duration of interaction of the pan/zoom gesture is low (less than 1s), while the final velocity is high and in the right direction in terms of translation and scaling, the acceptance condition is verified. In all other situations, the suggestion is rejected, and the controller proceeds as usual.

Every time a suggestion is accepted, the lens is moved to the target by smoothly changing all the continuous parameters during the animation, and using cross-blending to implement the smooth changes of discrete parameters (e.g., displayed layer).

The process can be fully automated by telling the system to accept suggestions without manual intervention, so as to produce a guided tour of the data that successively shows the selected annotations.

4.4 Implementation and results

For the evaluation, we expanded the same reference system discussed in chapter 3, and tested it on a number of complex heterogeneous datasets. In this chapter, without loss of generality, we demonstrate its usage on a use case stemming from the cultural heritage domain: an annotated 2D projection of fragmented sculptures. The videos are available together with the original publication [22].

The sculpture use-case concerns the exploration of a multi-layered rendered image of three representative models from the Mont'e Prama collection of prehistoric stone sculptures [47], [92]: Archer $n.5$, Boxer $n.15$, and Warrior $n.3$ (see Figure 4.1 right). The original 3D models are one of the outcomes of the Digital Mont'e Prama, a multi-year project carried out by CRS4 in collaboration with The Digital Mont'e Prama project is a collaborative effort between CRS4 (Visual Computing Group) and the Soprintendenza per i Beni Archeologici Sardegna (ArcheoSAR, the government department responsible for the archaeological heritage of Sardinia), which aims to digitally document, archive, and present to the public the large and unique collection of pre-historic statues from the Mont'e Prama complex, including larger-than-life human figures and small models of prehistoric nuraghe (cone-shaped stone towers).

The relightable stratigraphic models, created from the original scans, contains two layers: a normal map with diffuse color, and an unsharp-masked normal map with monochromatic color. 44 annotations at multiple scales and with lots of overlap (see Figure 4.2) concern reconstruction hypotheses, artistic details and part descriptions. I created these annotations based on primary literature on the statues [93]. Raffaella Chierici also assisted by drawing all the overlays concerning the reconstruction hypotheses.

In order to provide a preliminary assessment of the effectiveness of our approach, we designed a user study focused on the proposed novel interaction capabilities for the assisted exploration of annotated scenes.

The user analysis was done jointly with the evaluation of the lens controller with the same 25 participants (14 males and 11 females, with ages ranging from 11 to 69, median 41 years) were recruited among students, families and friends of researchers working at our center. All subjects had normal or corrected to normal vision and, as now extremely common, had basic computer or smartphone literacy.

While our navigation assistance approach should generally be applicable to support a human analyst in understanding complex data, we focused here on our motivating domain-specific application: the provision of effective exploration experiences in cultural heritage settings. In this context, physical installations in museums, as well as virtual exhibits over the web, have to deliver educational and pleasant experiences in a very limited amount of time [94]. Since museums must manage large amounts of visitors, long training times and/or guided explorations with the support of personnel are hardly affordable. The user interface should, thus, be perceived as simple, immediately usable, and provide guidance in complex operations (e.g., to avoid lost-in-space situations during navigation), while not being perceived as overly obtrusive [95].

In order to support self-paced exploration, our approach mixes a free navigation component, which lets users freely explore data by directly manipulating a lens, with guidance components, which use authored information to drive the users towards interesting annotated regions. Quantifying the effectiveness of user learning from data using various interfaces is difficult, if only because of the lack of consensus on metrics and methods, and because information learning has to be balanced with user engagement [96]. Thus, similarly to previous work on evaluating camera control in museum settings [97], we set as a goal of our preliminary user study only to have an indication of interface usability, user satisfaction and user performance in a context in which users are asked to freely explore a cultural heritage item, much as in a museum.

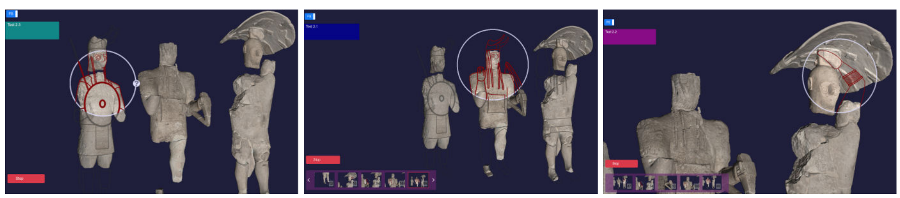


Figure 4.5: **Assisted navigation user interface evaluation.** Left: our controller; Middle: static thumbnail bar; Right: Adaptive thumbnail bar

4.4.1 Setup

We used the same web-based setup of section 3.4, applying it to the annotated sculptures dataset, which contains a database with 44 annotations pertaining to decoration descriptions and reconstruction hypotheses. Three alternatives were considered for the experiments (Figure 4.5): our fully assisted navigation system described in section 4.3 (LC), and two versions in which recommendations of far annotations are replaced by user selection in two kinds of thumbnail bars. The thumbnail bars are activated on demand by the user by clicking on a button and automatically disappear when the user selects the target annotation, triggering lens and camera motion towards that target. The first version of the thumbnail bar (FIX) is static and always presents all the annotations ordered according to authoring importance. Instead, the second version (DYN) is dynamic and presents the current top five annotation targets according to our similarity score.

4.4.2 Tasks

The experiments consisted in letting users to freely explore the annotated sculptures, with little or no training and no external direction. Users were told that their goal was simply to enjoy the experience and acquire information at their own pace in a prescribed short limited amount of time. This is expected to be a typical situation for walk-up-and-use user-interfaces in museum settings, where installations must engage museum visitors and enhance the overall visit experience in short times, if only because of the need to have many visitors use the installation. Moreover, it can also be considered a typical situation in an online museum with many datasets available, each one competing for the user's attention span.

4.4.3 Design

Similarly to what presented in section 3.4, each participant tested the three exploration systems in randomized order after seeing all of them in action in a short video, to understand the goal of the evaluation. Before each test, users familiarized with the interface by using it for less than 2 minutes on a different scene. The evaluation was performed by simply letting users try the three different interfaces for 3 minutes each one, for a total of less than 20 minutes per user testing session, including introduction and training. The remaining time of exploration was made visible to the user. User actions and system behaviors were monitored and stored in a log for further analysis. At the end of each experiment, participants were asked to evaluate the interface using the same SUS questionnaire of subsection 3.4.3 and to optionally provide free-form comments.

4.4.4 Performance evaluation

In order to assess the amount of information presented, we recorded for all the interfaces the number of annotations presented. For the assisted navigation interface (LC), we subdivided the number of annotations presented into annotation displayed directly because considered close to the current lens, suggestions presented but not accepted, and suggestions presented and accepted, as defined in subsection 4.3.3. We also recorded the number of annotations proposed but ignored (i.e., annotations that were indicated as "next best annotations" by our system but were not reached by a lens). For the two non-assisted versions (FIX and DYN), we recorded, instead, the time spent browsing the list of annotations in the thumbnail bar, measured as the interval from thumbnail bar activation to annotation selection. This time is an indication of the amount of time a viewer loses the main focus on scene exploration to decide where to look next. Using our assisted navigation approach, the participants visualized an average of 25.2 annotations (median 25, minimum 14, max 39). Of the visualized annotations, an average of 50.7% (median 50%, minimum 14.3%, maximum 87.5%) were directly displayed when the lens was judged close, while the remaining ones were displayed as a result of accepting a guidance suggestion. On average, 82% of the suggestions were accepted, while the remaining were rejected. These figures indicate that in over half of the cases annotations appeared transparently during the navigation, without the need of additional inputs which could distract users from interaction. Moreover, when suggestions were proposed without directly displaying the target annotations, the high acceptance rate of suggestions proposed without directly displaying the target annotations indicates their relevance for the user.

A comparison with the results obtained with the thumbnail bar versions also offer some interesting insights. First of all, the number of viewed annotations is lower, dropping to an average of 17.8 (median 17, minimum 8, maximum 30) for the fixed version (FIX) and 18.2 (median 17, minimum 6, maximum 31) for the dynamic version (DYN). ANOVA further confirms that there was a significant effect on number of viewed annotations at the $p < 0.05$ level for the three interfaces [$F(1,48) = 19.038$, $p = 0.00007$ when comparing our method with the FIX and $F(1,48) = 11.20$, $p = 0.00006$ when comparing it with DYN].

The lower number of annotations displayed by the competing interfaces is generated from the fact that interaction with the thumbnail bar takes time, reducing the time dedicated to exploring the scene. In fact, we measured that users interact with the scrolling widgets for large amounts of time. On average, for FIX, on average 20.1% of the time is spent interacting with the thumbnail bar (median 17%, minimum 3.7%, maximum 75.1%). Numbers are also important for DYN, where on average 14.6% of the time is spent interacting with the thumbnail bar (median 13%, minimum 0.0%, maximum 41.5%). It is interesting to note, here, the two extreme behaviors on these interfaces. One user of DYN decided to completely ignore the bar, and explore the scene solely by moving the lens, waiting for suggestions to appear when hovering over them, reducing to almost zero the time interacting with the bar, but reducing the number of annotations viewed (9). By contrast, a user of FIX decided to explore the scene almost solely with the thumbnail bar, jumping from one precomputed view to the next without moving the lens or the camera, therefore using the system more as a slide show than as an interactive exploration tool. This other extreme behaviour also led to the same small number of viewed annotations (9).

LC vs. FIX	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**SUS**
F(1,48)	6.683	5.233	1.589	3.647	2.602	1.321	0.716	4.200	4.184	1.485	6.054
p	0.013	0.027	0.214	0.062	0.113	0.256	0.402	0.046	0.046	0.229	0.018
Significance	⋆	⋆	ns	ns	ns	ns	ns	⋆	⋆	ns	⋆
LC vs. DYN	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**	**Q7**	**Q8**	**Q9**	**Q10**	**SUS**
F(1,48)	7.124	5.038	2.934	1.755	2.584	0.387	3.008	4.545	8.397	4.190	6.824
p	0.010	0.029	0.093	0.191	0.115	0.537	0.089	0.038	0.006	0.046	0.012
Significance	⋆	⋆	ns	ns	ns	ns	ns	⋆	⋆⋆	⋆	⋆

Table 4.1: **Usability evaluation of assisted exploration of annotated models.** Comparison of our method with the static (FIX) and dynamic (DYN) thumbnail bars using two one-way ANOVA on responses to SUS questionnaires. The last row of each comparison summarizes the per-question statistical significance resulting from ANOVA ($ns \rightarrow p > 0.05$; $\star \rightarrow p \leq 0.05$; $\star\star \rightarrow p \leq 0.01$).

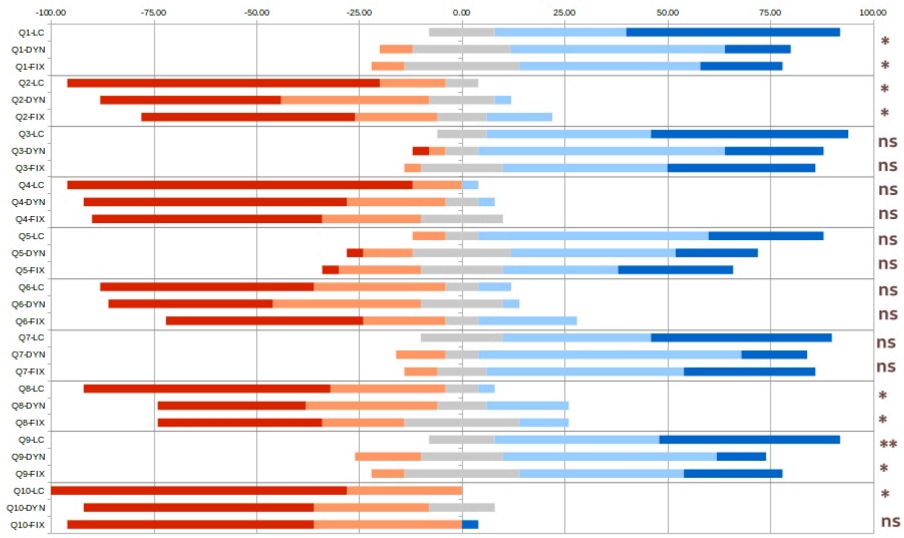


Figure 4.6: **Usability evaluation of assisted exploration of annotated models.** Diverging stacked bar charts of SUS questionnaire responses concerning our controller (LC), static thumbnail bars (FIX), and dynamic thumbnail bars (DYN). The color scale goes from red (strongly disagree) to blue (strongly agree). The labels near the right axis summarize the per-question statistical significance resulting from ANOVA ($ns \rightarrow p > 0.05$; $\star \rightarrow p \leq 0.05$; $\star\star \rightarrow p \leq 0.01$).

4.4.5 Usability evaluation

By analyzing the responses of the SUS questionnaires, summarized in the bar charts of Figure 4.6, we obtain for our guided interface a SUS score of 85.4, which, according to standard practices [74], rank the results as excellent. By contrast, the versions using the thumbnail bars obtain much lower scores, i.e., 74.5 for DYN and 74.9 for FIX. The results of ANOVA comparing LC to FIX and LC to DYN are reported in Table 4.1. They confirm that there was a significant effect on SUS score at the $p < 0.05$ level for the three interfaces [$F(1,48) = 6.824$, $p = 0.012$ when comparing FC to DYN and $F(1,48) = 6.054$, $p = 0.018$ when comparing it to FIX]. ANOVA on answers to individual questions revealed that there was a significant effect at the $p < 0.05$ level on the perception of complexity (Q2) and awkwardness (Q8), as well as on the confidence in using the method (Q9). Of particular importance for museum applications, in which walk-up-and-use interfaces are paramount, is the fact that there was a significant effect also on the desire to use the method frequently (Q1) and on the amount of training required (Q10).

We also gathered useful hints and suggestions from comments recorded by subjects in the final form. In general, most users appreciated the idea to use a lens for navigation in an annotated database. Some users mentioned that they liked the idea of actively requesting suggestions, in order to jump to another location when the local interactive exploration is considered complete. One user found the idea of suggestion interesting, but considered the animated glyph showing the direction not clear, as it did not show the actual annotation target. To solve this problem, scalable insets [98] could be explored as a way to complement lenses for providing guidance towards far or off-screen locations. Other users, by contrast, liked the fact that the suggestion has little intrusiveness. We believe that we can further explore these aspects, in particular, by expanding guided tour features and combining more intrusive suggestions with a ”snooze” option for users that do not want to get distracted too much.

4.5 Discussion

We have presented a novel approach for exploring visually annotated models using an interactive lens. By mixing and matching the concept of interactive lenses with that of annotations, we introduced a new method for guiding users in the self-paced exploration of annotated 2D models. The presented results on a use-case stemming from the cultural heritage domain demonstrate how this technique leads to a new way of mixing casual interaction with storytelling from data. One important result in this area is that our approach of selecting the next best annotation to display and of differentiating between immediately displayable annotations and possible future annotation to display makes it possible to support a variety of use cases. In particular, we can effectively support the usual way to display relevant data under the lens during fully free user-controlled exploration, always selecting scale-specific data and avoiding clutter while displaying the single selected annotation also in the context area. Moreover, by the introduction of suggestions, we can assist navigation to direct users towards interesting areas. Finally, we can provide fully guided tours, that can be started at any time by accepting all suggestions in a sequence.

In the next chapter, we will show how this approach can be further extended by arranging lenses in a graph, in order to take into account also precedence relations among annotations. Since the current evaluation focuses mostly on extracting basic performance measures and getting data on user satisfaction, more work is required to objectively assess the effectiveness of our user interface for specific tasks other than casual inspection. Addressing this would require cognitive measures that are beyond the scope of the thesis, and are an important avenue for future work.

4.6 Bibliographic notes

Most of the content of this chapter was presented in our EUROVIS 2022 contribution and published in the Computer Graphics Forum journal [22], which also includes the follow-up work on annotations presented in the next chapter. Our EUROVIS talk is publicly available here, and includes demonstration video and further related content. I have significantly contributed to the conceptualization, methodology, and validation of the method and was one of the primary authors of the paper.

Chapter 5

Annotation graphs for guiding lens-based scene exploration

We extend the techniques introduced in the previous chapters by considering not only the relations of annotations with objects, but also the relation of annotations with themselves, in order to provide authors with storytelling features. In this new approach, we do not use a flat annotation database, but record information on the interesting areas of the model is encoded in an annotation graph generated at authoring time. Each graph node contains an annotation, in the form of a visual and audio markup of the area of interest, as well as the optimal lens parameters that should be used to explore the annotated area and a scalar representing the annotation importance. Directed graph edges are used, instead, to represent preferred ordering relations in the presentation of annotations, by having each node point to the set of nodes that should be seen before presenting its associated annotation. A scalar associated to each edge determines the strength of this constraint. At run-time, users explore the scene with the lens, and the graph is exploited to select the annotations that have to be presented at a given time. The selection is based on the current view and lens parameters, the graph content and structure, and the navigation history. The best annotation under the lens is presented by playing the associated audio clip and showing the visual markup in overlay. When the user releases control, requests guidance, opts for automatic touring, or when no available annotations are under the lens, the system guides the user towards the next best annotation using glyphs, and potentially moves the lens towards it if the user remains inactive. This approach supports the seamless blending of an automatic tour of the data with interactive lens-based exploration. The approach is tested and discussed in the context of the exploration of multi-layer relightable models.

Figure 5.1: **Overview.** ***Left***: The user explores the scene using an interactive lens, and the best annotation under the lens is presented by playing the associated audio clip and showing the visual markup in overlay. ***Middle***: when the user releases control, requests guidance, opts for automatic touring, or when no available annotations are under the lens, the system indicates the next best annotation using glyphs. ***Right***: if the user remains inactive, the lens is moved towards the selected target. This approach can be used to generate intuitive tours through the data that dynamically respond to user actions, seamlessly transitioning from full user control to automatic navigation.

5.1 Introduction

In the previous chapters, I have shown how a context-dependent clutter-free display can be achieved by displaying a single annotation at a time under a movable lens, selecting it using a recommendation system that takes into account the current camera position, current interactive lens parameters, and navigation history [22]. This approach, however, does not consider the relation among annotations themselves, and has thus limitations in the ability to prescribe presentation orders to define meaningful tours through the data [99], [100].

In this chapter, we introduce a novel approach for guiding users in the exploration of annotated 2D models by exploiting an annotation graph generated at authoring time (Figure 5.1). Each graph node contains an annotation, in the form of a visual and audio markup of the area of interest, as well as the optimal lens parameters that should be used to explore the annotated area and a scalar representing the annotation importance. Directed graph edges are used, instead, to represent preferred ordering relations in the presentation of annotations. A scalar associated to each edge determines the strength of this constraint (section 5.3). Such edges let us introduce storytelling features by letting each node point to the set of nodes that should be seen before presenting its associated annotation.

At run-time, a user explores the scene with the lens, and the graph is exploited to select the annotation that has to be presented at a given time (section 5.4). We call it the *best annotation*, to reflect it is the particular one which optimizes a set of selection criteria, that considers the current view and lens parameters, the graph content and structure, and the navigation history, through a novel technique that also takes into account topological distance among subsequently presented nodes in the annotation graphs (section 5.5). The best annotation under the lens is presented by playing the associated audio clip and showing the visual markup in overlay. The use of audio clips to audibly present the additional information lets users focus on the visual content lens, without further clutter. When the user releases control, requests guidance, opts for automatic touring, or when no more annotations are available under the lens, the system points towards the next best annotation using glyphs, and potentially moves the lens towards it if the user remains inactive (section 5.6). This approach can be used to automatically generate intuitive tours through the data that dynamically responds to user actions in real-time.

As for the previous solutions, we present an objective and subjective evaluation of our method for the exploration of stratigraphic relightable models (section 5.7).

5.2 Related work

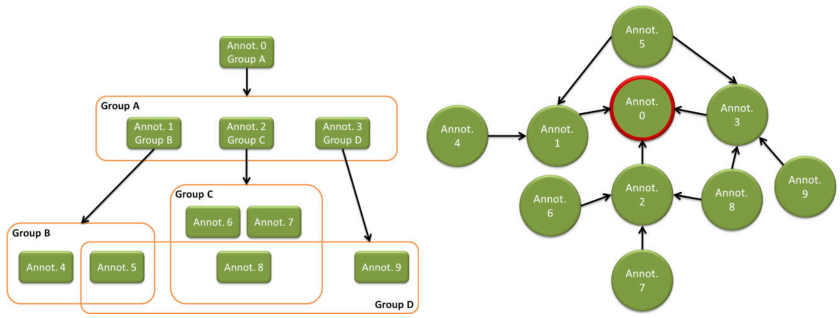
Exploration of annotated models and interactive lenses are broadly studied topics within the visualization community, and most of the related work has been covered in the previous chapters. I, therefore, summarize here only the most relevant works related to the novel content presented here.

Serial temporal presentation to enhance the view [80] is one of the approaches that has been used to deal with overcrowded display, and, in conjunction with authoring or automatic determination of temporal precedence, it provides a way to deliver a narrative meaningful tours through the data [97]. Manually writing or defining fixed key-frames and forcing a single path is one of the most adopted solutions [59], which has also been used by touring through annotations [88]. This approach, however, leads to the generation of static videos rather than interactive experiences.

In this chapter, we build on our prior approach [22] by significantly extending the annotation representation, moving from a simple flat list of annotations to an annotation graph, in which the edges express semantic relationships among nodes, exploiting these relations for automatic data touring and generating guided suggestions. Annotation selection is based on a score that extends to annotated lens graphs, with the Degree-of-Interest (DOI) concept introduced by Furnas [89] for trees and extended by Van Ham and Perer [90] to graphs. Similarly to Gladisch et al. [91], DOI computation also takes into account the past behavior of the system. The camera-control work of Balsa et al. [97] is the most similar to ours, as it selects only a single item at a time from a viewpoint graph. Our annotation graph and scoring system is, however, targeted to support lens-based navigation of an annotated model, and has a different structure. In particular, we expand our previous work [22] by introducing a dependency score to support hierarchical grouping, and a topology score to drive the system towards an orderly visit of the graph by penalizing changes in levels of abstraction of topic switches. Moreover, we introduce a new state machine design to seamlessly combine automatic touring with self-guided visits.

5.3 The audio-visual annotation graph

Traditionally, annotations are used to identify specific regions, linking them to metadata or other characteristic information [24]. In this thesis, we want to exploit annotations for



(a) Hierarchical grouping (b) Graph representation

Figure 5.2: **The annotation graph for hierarchical grouping.** Edges in the graph point to enabling nodes.

guidance and data presentation. Similarly to what presented in chapter 4, we associate to each annotation a *visual overlay* and an *external annotation description*, together with the parameters that should be used for an effective lens-based exploration of the annotated area. In this chapter, the visual overlay is simply an image or text that can be drawn over the model, while the external annotation description is a link to a hypertext with additional information. The exploration parameters are used for navigation control, and consists an *annotation importance* (i.e., a user-defined weight to associate a higher (or lower) likelihood that an annotation will be displayed before (or after) the others), a *lens and context area description* (i.e., the position and size of the best lens and camera-angle for viewing the annotated area), and a set of *rendering parameters*. For this work, focusing on stratigraphic relightable models, the rendering parameters include the layers that are displayed inside and outside the lens and the light configuration in terms of both direction and type (e.g., collinear or spot light); for the spot light we also specify the light beam aperture. Other rendering parameters can be defined (e.g., brightness, gamma value), also related to different rendering strategies (e.g., shape/color enhancement operators).

In addition, we also include with each annotation an *audio description*, which is an explanatory text that describes the annotated area, and is intended to be played when the annotation is visited. The audio clip can be generated by synthesizing the textual description, or be a pre-taped recording. In both cases, the audio clip duration defines the minimum time that the system considers should be spent for considering an annotation seen (section 5.5). Using audio to describe the annotated area is particularly appropriate for our use case and interface design. In particular, using the audio clip rather than a displayed text to convey non-visual information allows us to let users concentrate on the model, and to produce a lean visual overlay when exploring the scene with the lens.

Using audio for enhancing a museum visit is very common, and it is employed in a range of solutions, from conventional audio guides presenting short bursts of audio information at each stop [101] to virtual audio-visual visits [88]. Supplementing visual information with audio has also been shown to improve memorability [102]. However, usage of audio may not always be appropriate. For instance, handling multiple co-located users performing separate visits requires special care. The typical solutions in museum settings are the set-up of an isolated display area, with space management complications and limitations in the number of active interactive displays, or the usage of headsets by individual visitors, locking them into isolated experiential bubbles with the risk of reducing inter-personal interaction [101]. For handling those cases, we can provide a purely visual experience, in which the annotation description is displayed on screen. Note that, since we use a lens, we cannot simply display the text in a separated area, as in classic annotated model presenters [6], since this solution would force users to lose their focus. Our current solution is to display the annotation in a small area under the lens. We plan to improve this approach by considering an optimal shape, placement, and scaling of the text box attached to the lens border, so as to reduce the masking of the annotated area, as done in external labeling techniques [103].

In order to specify a preferred presentation order, we introduce dependency links, transforming the annotation database into an *annotation graph*. In this representation, each node is an annotation, and directed edges point to a set of enabling nodes (one or multiple parents) that should be seen before visiting it. The presence of edges allow authors to define a preferred global order, that can be used to create a story-like structure between annotations, or, e.g., to go from coarse to finer details as prescribed by the visualization mantra [56]. A weight associated with the edge, ranging from zero to one, defines the strength of the dependency between the nodes (see section 5.5).

By using this graph representation, for instance, it is possible to structure an annotation database into hierarchical groups of nodes, to represent information at various levels of abstraction, as done for complex graph exploration [91]. Semantically, each node in the graph can be seen as a coarser representation of its children, and this translates into the fact that the annotation associated with a leaf node is best presented to the user only after its parents. This particular view of the dependencies helps guide authors in the definition of links, as they can proceed to structure annotations coarse to fine, or inversely grouping them from fine to course during their editing process.

Figure 5.2 shows an example of hierarchical organization of a set of ten annotations. In particular, Figure 5.2a depicts the idea of *hierarchical grouping*, where each annotation can represent a set (or group) of other annotations, which we can consider its children. As previously explained, visualization order depends on high level annotations that, once viewed, enable the visualization of the groups they represent. On the right (Figure 5.2b), we present how the grouping and the corresponding visualization order is implemented through dependency edges; note that while all the authored hierarchical groupings can be expressed/transformed in a graph, not all the graphs can be transformed in a hierarchy of groups, e.g., some cyclic graphs. If an arrow points from a node *B* to a node *A*, it means that the visualization of *B* depends (with a certain level of dependency) on the fact that the node *A* has already been visualized. The annotation at the highest level is the first displayed annotation (we can consider it the root of the navigation), and it is depicted with a bold red contour in Figure 5.2b.

The annotation attributes and their organization into a graph is exploited by our system to present information in a context-dependent and graph-dependent order during navigation (section 5.4).

Note that authoring details are orthogonal to the subject of this thesis. For the sake of completeness, we mention here that we annotate the models during the lens navigation as described in the previous chapter, with slight modifications to support the inclusion of dependency information. The system allows users to move the lens to interesting areas, draw annotations with a simple image editor, and store them in an annotation database containing the lens and context area description, as well as the rendering parameters. The node table is then edited off-line by adding dependencies to nodes, and enriching the description of each annotation with an audio recording.

5.4 Interactive and guided lens-based exploration

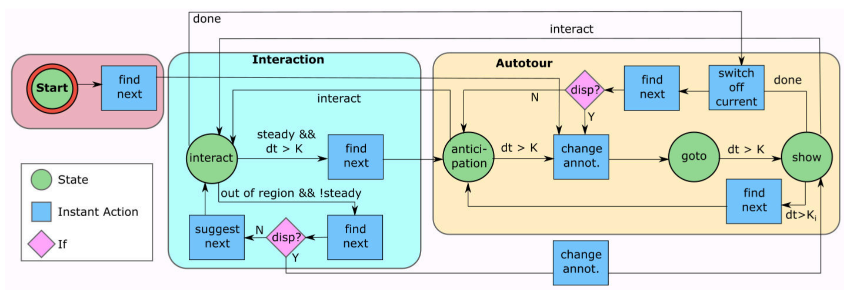


Figure 5.3: **Annotation Navigation State Machine**. Two main navigation modalities have been implemented, i.e., the manual interaction (cyan box) and the auto-tour (yellow box). In the first mode the users freely move the lens, while in the latter they are guided through annotations that are automatically selected. To enter the auto-tour mode the users just stop the interaction with the lens interface; re-touching the interface will bring it to the manual mode.

At run-time, the user explores the annotated scene using a visualization lens that interacts with the scene by moving and scaling the focus area and activating relevant annotations. Since only a single context-dependent annotation is selected at a time, clutter is reduced. The sequence of selected annotations must be relevant to the current spatial context and maintain a flow, so that more general information is presented before dependent details. We do that by running a state machine that exploits the annotation graph and responds to user actions. Our goal is to support the seamless transitioning between two behaviors. On one extreme, we would like the system to be capable of producing automatic tours of the data, by presenting annotations in a sequence, as for a video tour. On the other extreme, users should be allowed to explore the scene at their own pace, with relevant annotations appearing in sequence as the user moves to the annotated areas. In the common intermediate situation, we would like to support users that start with automatic touring, then explore the scene for a while, then restart auto-touring, possibly in other areas depending on their interest.

The state machine is made basically of two intertwined parts, one devoted to react to interaction, and one devoted to perform automatic tours. The user can interact with the state machine in three ways: by doing nothing (the user accepts what is proposed by the state machine), moving the lens (the state machine accepts what the user proposes), or sending a `Done` signal to communicate that the exploration of the current annotation is completed.

The machine is composed of 5 states: *Start*, *Anticipation*, *Goto*, *Show*, *Interact* (Figure 5.3). The loop *Anticipation*, *Goto*, *Show*, constitutes the auto-tour part. From this loop the user can exit only by starting to interact with the lens.

The state machine employs a function *Find Next* that, given the current situation, identifies the next best annotation. From the *Start* state, the first annotation is selected and the lens is moved over it through the *Goto* state. During *Goto*, the lens position and the rendering parameters are interpolated from the currently displayed situation toward the target one, encoded in the node database together with the annotation. Note that, for the particular case of the relightable stratigraphic models used in this thesis, adjusting the rendering parameters includes the selection of inner and outer layers and the update of the illumination settings in terms of light intensity and direction. This means that, when interpolation is complete, the model is displayed with the full visualization settings that annotation-author has stored in the database.

When interpolation finishes, the state changes to *Show*. The *Show* state displays the annotation for a content-dependent time K_i. In particular, if not specified by the user in the annotation database, this time is equal to a small setup time of half a second plus the duration of the audio clip associated with the annotation.

It is possible to exit from this state in three ways: by interacting with the lens, changing the state to *Interact*; when the allocated time elapses, entering the *Anticipation* state; or finally when the user signals that he has `Done` with the current annotation. This last operation is used to speed-up exploration in case the user is not interested in the current content anymore. To decide what to do next, the system evaluates if the next selected annotation is directly *displayable*.

An annotation is considered displayable if, by drawing it as an overlay, it can be reasonably well perceived by the viewer, without the need to resort to directing or leading cues to indicate where the annotation is located. The displayable condition, thus, requires checking whether the view is approximately at the same scale of the annotation (in this thesis, within a factor of two larger or smaller in zoom factor), and at least some portion of the annotation is within the focus area of the lens.

If the next best annotation is displayable, the machine changes the state to *Goto* and the lens is moved to properly center the new annotation while remaining in the auto-tour loop. If the next annotation is not displayable, the state is changed to *Anticipation*.

The *Anticipation* state has a twofold purpose: it alerts the user with visual cues that the lens is going to move towards the next annotation and provides the users with information on where the next annotation is placed using visual hints (see section 5.6 for details on visual signals and direction hints). From the *Anticipation* state, there are two possible existing transitions: if a certain amount of time with no user action elapses, the system considers the next annotation accepted, and the auto-tour continues by changing to the *Goto* state; otherwise, upon user interaction, the system enters the *Interact* state.

During the *Interact* state, the user keeps visiting the current displayed annotation, possibly moving the lens. From this state it is possible to exit in three ways: being steady after the whole annotation time has elapsed, going outside the annotation area, or through the `Done` signal. Steadiness is considered as having completed the inspection, thus the system alerts the user by entering the *Anticipation* state. Instead, when going outside the current annotation area, an indication of the next best annotation is presented. Then, two events can produce the state change: if the user keeps moving but passes over an annotation considered displayable, the state changes to *Show* and the annotation is made immediately visible, or if the user stops interacting, after a small amount of time the state goes to *Anticipation* to present the new annotation. Finally, sending `Done`, produces a situation similar to auto-touring: if the next annotation is displayable the state changes to *Goto*, otherwise to *Anticipation*.

Note that, with this approach, we cannot distinguish if a user remains inactive after having inspected an annotation because the inspection has been completed or because the user is closely inspecting/pondering on the current view. In the latter case, repeatedly receiving a suggestion, through entering the *Anticipation* phase, might be considered annoying. To reduce this effect, we increase the time for receiving a suggestion by doubling the inactive timeout each time the user does not accept the suggestion by interacting during the *Anticipation* phase, up to a maximum timeout. By contrast, the timeout is halved each time a suggestion is accepted or a suggestion is requested, until we reach the default timeout. In the work presented in this chapter, the minimum and default timeout is 5s, and the maximum is 40s. An alternative solution would have been to remove the timeout, and explicitly ask the user to signal the completion of interaction viewing, which is now optional. This is still possible by configuring the timeout to (much) larger values. However, we consider in this work the small-timeout version, to test the typical setting of cultural heritage visits, which take into account the limited span of attention of visitors and the need to streamline visits in order to increase the visitor throughput while delivering enjoyable experiences.

During the visualization experience, thus, the system continuously performs two main tasks. The first is the selection (when required) of the next best annotation to display (section 5.5). The second is the management of the user activity through several device mapped interactions (section 5.6). Those two elements drive the annotated model visualization and allow the user to seamlessly switch between interactive and automatic navigation.

5.5 Best annotation selection

During the navigation the system selects the next best annotation for the automatic tour using a scoring system. Following Bettio et al. [22], we assign to each recorded annotation node i a score $N_i = \gamma_i \sigma_i H_i$, where γ_i is the author-defined annotation importance, σ_i is the similarity score depending on spatial and semantic distance, and H_i is the history score depending upon the activity log of the active user, that equals to 1 when the node has not been visited in a recent time, and 0 when it has just been visited. H_i is thus initialized to 1 for all the not visited nodes. When a node is visited it goes immediately to 0, in order to avoid presenting again the same node, then smoothly gets back to 1 over a certain time, meaning that after a certain elapsed time the user tends to forget the content of a node, and could be presented again; this time can be set proportional to the amount of the duration of the visual-audio annotations, in order to avoid disturbing repetition of annotations before having seen the vast majority (or all) of them. More details of each of these individual scores are presented in the original publication [22].

In order to consider dependencies, we extend this formulation by multiplying the node score N_i by a dependency score δ_i, which takes into account node precedence relations and their weights, and by a topology score τ_i, which depends from level of abstraction distances, to obtain a final annotation score

$$S_i = \delta_i \tau_i N_i = \delta_i \tau_i \gamma_i \sigma_i H_i \tag{5.1}$$

5.5.1 The dependency score

The dependency score δ_i takes into account node precedence relations and the corresponding weights. It expresses the fact that the author would prefer a given node to be presented after its enabling nodes, with a weight depending on the edge strength. This is achieved by taking the fuzzy logic AND (i.e., min operator) of a per-edge quantity that expresses if the node has already been presented, and which strength should have this information. The dependency score of node i is thus given by

$$\delta_i = \min_j (1 - e_{ij} H_j) \tag{5.2}$$

where j loops over all enabling nodes, e_{ij} is the author-selected edge weight linking node i to node j, and H_j is the history weight of the node j. For a strong dependency with weight $e_{ij} = 1$, when the parent node is not visited ($H_j = 1$), the dependency weight δ_i is 0, thus blocking the presentation of the node. When, instead, the parent is visited ($H_j = 0$), the node has $\delta_i = 1$, so the node is completely enabled, and is thus included in the potential candidates for selection. When dependencies are, instead, weak, i.e. $e_{ij} < 1$, the dependency score will not reach 0, permitting, with low probability, the visit of a node even if all the enabling nodes have not yet been visited.

5.5.2 The topology score

The aim of the topology score is to provide a configurable orderly visit of the annotation database, without large semantic jumps among proposed content. Since the graph structure encodes relations among annotations, e.g., by grouping and definition of levels of abstraction, we define a weight that favors proximity relations in the graph.

For instance, it is preferred to visit children and siblings of the last active node, as the content of their annotation is likely more strictly related to what just presented than the content of other annotations in the graph.

For that purpose, we can define the semantic distance among two annotations as the topological distance between the two nodes containing them, which is the length of the shortest path among the two nodes. In order to give the same proximity value to siblings and children, we virtually insert edges among siblings. Note that this can be done directly in the distance computation method, without any structure modification, by setting the distance among siblings to be one rather than two, as it would be required if we were forced to go up to the parent and then down again.

Given the current candidate node i and the last visited node j, the topology score τ_i is then defined as:

$$\tau_i = 1 - \beta \frac{\min(d_{ij}, d_{MAX})}{d_{MAX}} \tag{5.3}$$

where d_{ij} is the shortest path in the graph between node i and node j, while d_{MAX} is a normalization factor used to define the maximum allowed topology distance, which is independent of global graph size (i.e., adding or removing distant nodes), and is defined at authoring time (Figure 5.4), and β is a scalar that is equal to zero if the user is interactively moving the lens and one otherwise.

In order to speed-up run-time evaluation of the topology score, we precompute in advance all the mutual distances between graph nodes with the Floyd Warshall Algorithm [104], which provides the shortest distances between every pair of vertices in a graph.

The scalar β allows us to tune the behavior of the system depending on the current situation. Taking into account the topological distance in the annotation graph is of primary importance during automatic touring or when the user explicitly asks for suggestions, as it is very reasonable to strongly favor semantic proximity over spatial proximity. On the other hand, when the user is freely moving, especially far from the last displayed annotation, switching subjects in order to show locally relevant information is often the expected behavior, as users typically move for new knowledge discovery, also signaling with their motion that the current story flow can be modified.

5.5.3 Choosing the best annotation

The next best annotation is then selected by taking into account the scores S_i, which determine the suitability of each particular annotation for a given context. When scores are widely different, the annotation with the highest score must definitely be preferred, as lower-scored annotations would seem out-of-context. However, when scores are very similar, several different annotations might be considered suitable. This is not an unlikely situation, especially when no annotation is overlapping with the current lens. To take into account this situation, rather than just selecting the annotation with the highest score, we perform a stochastic selection among a small set of nodes that have a similar high score. In particular, we select a cutoff score S_c equal to a fraction C of the

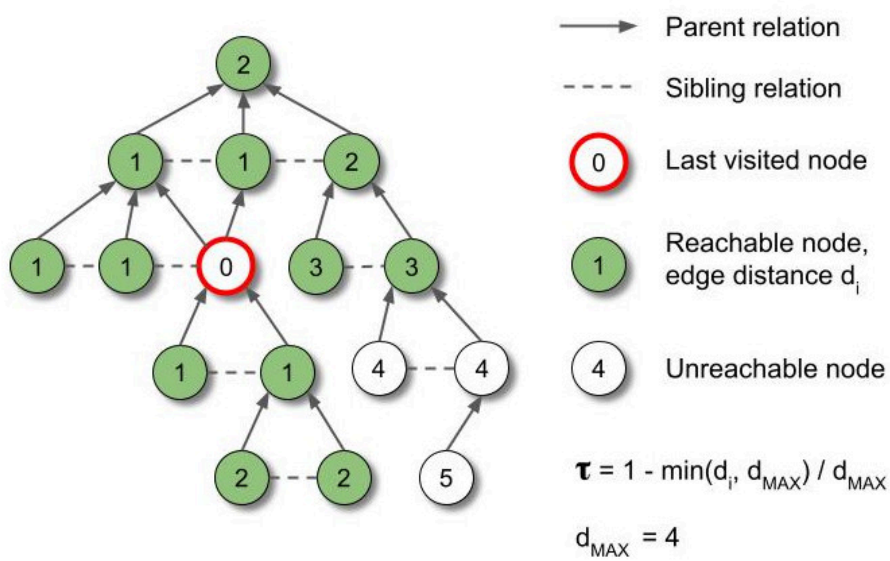


Figure 5.4: **Topology distance.** Annotation graph with parent and sibling relations. Topology distances computed with respect to the red node. Topology score is derived by the depicted formula with $d_{MAX} = 4$.

maximum achieved score, and extract the subset of K nodes which have a score higher than this threshold. We then assign to each node in this subset a picking probability $p_k = \frac{S_k}{\sum_i S_i}$, and select the next best annotation according to this probability. In such a way, the exploration is open to a wider range of possible paths, while maintaining the author dependency requirements.

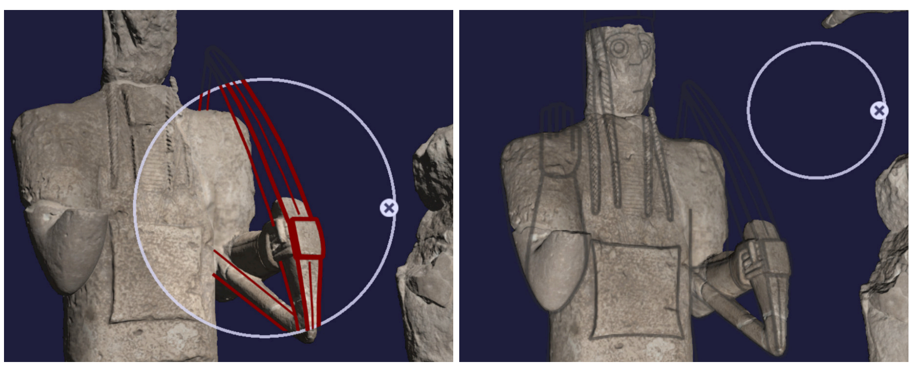
When the cutoff C is set to 100%, there is no stochastic selection, and the system, let alone, always repeats the same tour. In the work presented in this chapter, the cutoff C is, instead, tuned depending on the current situation. In particular, it is equal to 95% when searching for annotations while the user moves the lens, and to 60% otherwise. This makes it possible to choose among a large number of likely paths when performing automatic tours or the user is requesting suggestions, increasing the variability of the exploration experience, while avoiding the selection of incoherent solutions. This variability is important for casual visitors, as it makes the visit more engaging and less repetitive (see subsection 5.7.2 for an evaluation of the effect).

5.6 User interface and device mapping

Our user interface for lens-based exploration requires minimal user input, and can be mapped to input devices in a variety of ways. For lens and camera movement, we employ the recently-introduced approach of Bettio et al. [22], that couples lens and camera motion to always ensure a good focus-and-context placement of the lens within the view. Using this approach, the user manipulates only the lens, changing its position and radius, and the system automatically computes the camera translation and scale updates in order to always maintain a good focus-and-context situation. In our current implementation, we realized both a multi-touch solution and a mouse-controlled version.

At the start of the navigation, the lens is moved to the best position for the first annotation selected (a selected root node in the annotation graph). Then, the user can pan the lens by a one-finger pan gesture (or by using the left mouse button), and use pinch-to-zoom (or the wheel or up/down movement holding the right mouse button) to modify lens scale. In both cases, the controller adjusts camera position and zoom to maintain the focus-and-context condition. The state machine, running in background, reacts to lens and camera motions to change interaction modes and update the display, as detailed in section 5.4.

The user interface also includes additional features that implement all the characteristics of the controller. In particular, during the *Show* and *Interact* state, the lens always has a small button with a cross that, when clicked (with a touch or a left mouse click), triggers the `Done` signal (Figure 5.6). That signal communicates to the system the fact that the user has finished inspecting the current annotation, and asks to visualize the next best annotation. The `Done` button is not available during the *Anticipation* and the transition state *Goto*.



(a) (b)

Figure 5.5: **Annotation Rendering.** Rendering within the lens shows the original annotation colors, instead for content outside of the lens the colors are transformed into grayscale.

Visual signals also enrich the interface during transitions. The *Anticipation* state, in particular alerts the user that the current annotation is going to be replaced by the next one. To convey this message, we progressively change the color of the lens boundary from white to red, to alert the user that the system is about to go to the next annotation if the user does not restart to interact with the lens.

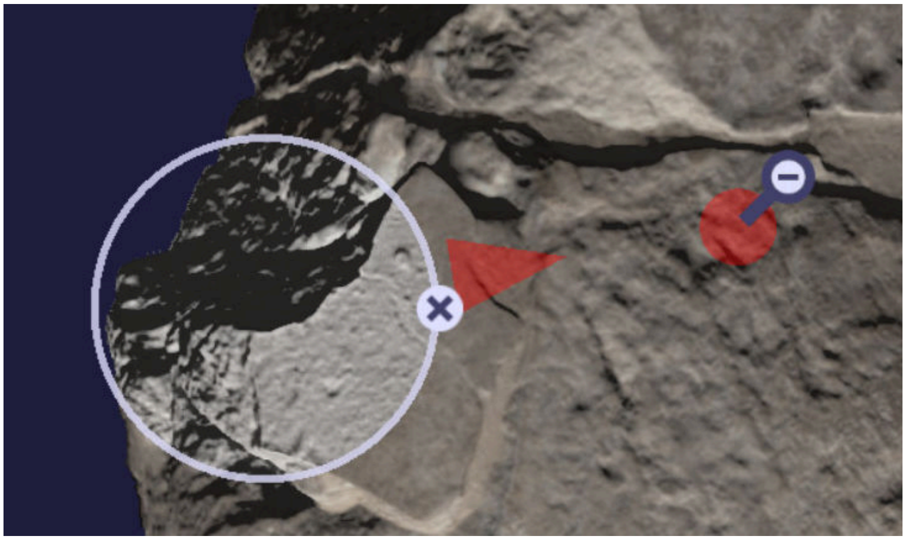


Figure 5.6: **Interface glyph**. Glyphs rendered during interaction with the lens outside the current annotation area. A cross button, placed over the lens border, can be used to communicate `Done` signal. Red arrow and spot indicate the direction and position of the next annotation center. Hand lens with minus sign, indicates necessity of zoom out (plus sign would be used for zoom in).

The other important visual signals concern annotation display. The representation used for the annotation depends on whether it is the currently active annotation or the next proposed one, as well as on whether the annotation is within its display range or outside it.

The current annotation, when considered displayable, is rendered with full color within the lens and dimmed outside the lens (Figure 5.5a). In the *Anticipation* state, as well as in the *Interact* states when the lens moves out of the current annotation, both the current annotation and the next one are displayed. When the next annotation becomes current, the previous one disappears.

One of the main problems to tackle is the display of annotations that are not currently visible (e.g., far from the lens, outside of the view frustum, or outside of the zoom range). This occurs, in particular, when the user must suggest the next annotation and provide directions towards it.

The problem of displaying out-of-view objects is subject of much research, and the main techniques are distinguished among the use of *leading cues* attached to the target, meaning that some part of the cue is always spatially connected to the target, and *directing cues*, mostly fixed in the user's view and giving the user a general direction to the target instead of providing a direct path to follow [105]. In this work, we use a combination of both.

In particular, to indicate to the user an annotation that cannot be displayed, we use a combination of three indication glyphs: arrow, spot and zoom (Figure 5.6). A dynamically oriented arrow placed around the lens, similar to a compass needle, points in the direction toward the next annotation center. It acts as a directing cue, but is not fixed in the view but attached to the lens, since it is, at the same time, the object that controls navigation and the area where the user is focusing. Moreover, if the annotation cannot be displayed but is within the viewport (e.g., because out of zoom range), we use as leading cue a red filled dot placed at the center of the new annotation. A zoom indicator, a small hand lens with plus or minus sign, shows if a change of zoom is required to properly see the annotation.

In addition to annotation overlays and leading and directing cues, we also employ audio to convey semantic information without overloading the visual channel. This is particularly important for our lens-based interface, since fixed text areas would require users to move their focus out of the lens, while moveable text areas attached to the lens would provoke considerable masking in the lens context.

5.7 Implementation and results

We implemented the proposed approach by extending the web-based platform presented in the previous chapters that targets multi-layered relightable models. The client can run in regular web browsers (we tested, in particular, Firefox, Chromium, and Chrome on both Windows and Linux platforms, and Edge on Windows), supporting both mouse or multi-touch input using the TouchEvent API. Figure 5.7 shows how we can adapt to multiple use cases, including full-screen display on large multitouch installations, and desktop or tablet visualization for web distribution.

The preparation of the relightable images and their layers, all the annotations and the authored annotation grouping (node and edge attributes) in the relation-dependent, hierarchical graph, and all the associated audio clips are done off-line. They are stored in a repository that contains the set of image layers, the audio clips, a configuration file that manages the arrangement of those layers, and a file that includes both the text annotations with all the graph structure. At run-time, the viewer loads a scene description that includes the annotation database, and starts navigation by placing a lens at the root position.

We have tested our system on a variety of models. In this chapter, we provide an objective and subjective evaluation centered around the exploration of a cultural heritage scene, with the aim of analyzing and assessing the suitability of our navigation system for casual users, as typical on museum web sites or walk-up-and-use installations. The accompanying videos provide an illustration of the behavior of our method, as well as sample footage from our user tests.

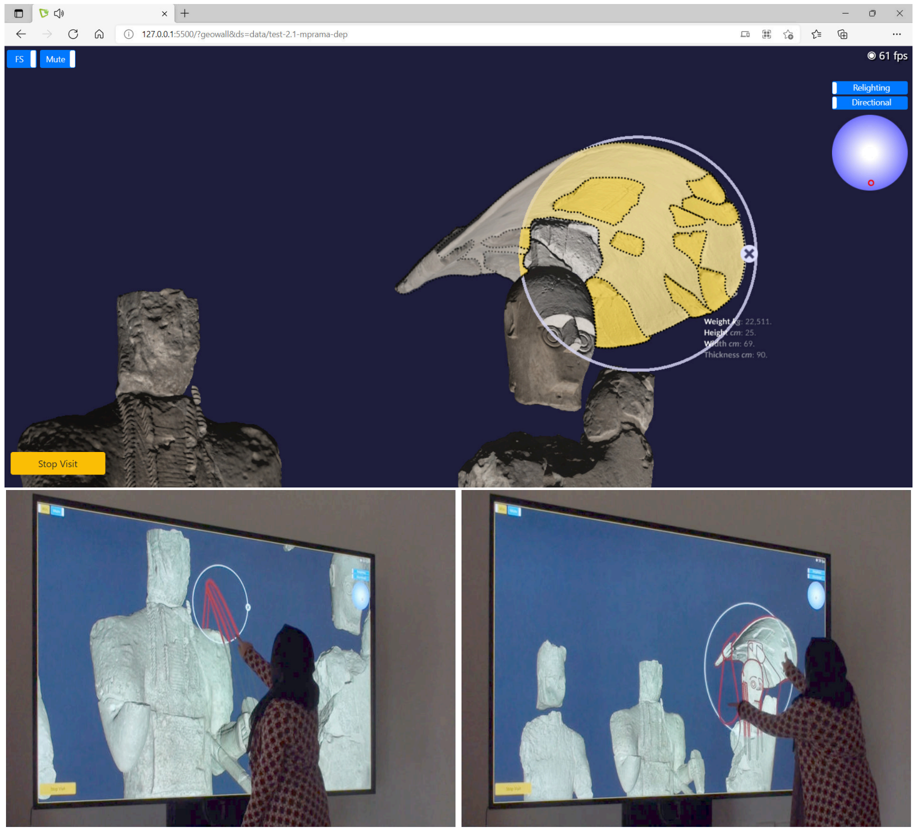


Figure 5.7: **Multiplatform application.** The same web-based implementation is used for multiple use cases. The top image shows the application running inside a web browser on a desktop platform. The bottom images show two frames from the recorded video of an interactive session on a large touch screen for a walk-up-and-use museum installation.

5.7.1 Annotations Creation and Dataset Preparation

The test dataset is a relightable multi-layered rendered image of three representative models from the Mont'e Prama collection of prehistoric stone sculptures [47], [92]: Archer n.5, Boxer n.15, and Warrior n.3 (Figure 5.8), already used for the evaluation of the technique presented in chapter 4. The annotation database concerns reconstruction hypotheses, artistic details and part descriptions. It contains 108 annotations at multiple scales that form 25 annotation groups; in total we have 107 edges that express groups and nodes dependencies. An illustration of all annotations in one single frame and the density of all lenses is shown in Figure 5.8. All the annotations were given the same authored importance.

In creating our annotation database, our goal was to enrich the plain visual represen-

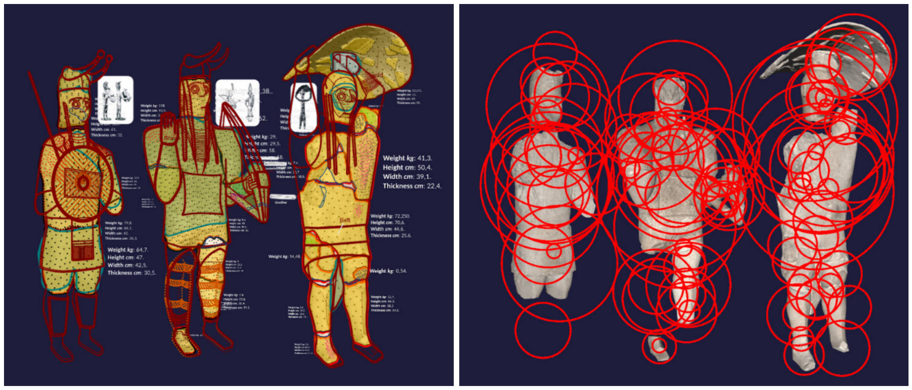


Figure 5.8: **Mont'e Prama Dataset**. Three statues from the Mont'e Prama collection of prehistoric stone sculptures (from left to right): Warrior n.3, Archer n.5, and Boxer n.15. The left image shows the content of all the annotations of the database, while the right image shows the corresponding lenses.

tation with pieces of interesting information taken from the historical and semantic knowledge in the Mont'e Prama related literature [93]. The intention was to ensure that the annotations, in terms of visual and audio content, are easy to interpret for a common user, without any prior knowledge of paleo history. We started from the work described in the previous chapter (chapter 4), adding more annotations (from 44 to 108), including more annotation kindes, adding audiovisual information, and structuring them in a graph.

Starting from the source information contained in a dedicated archaeological books series [93], we created a variety of new annotations, that can be concisely classified into (A) graphical extensions of missing parts, limbs and accessories of the statues *(total n. 22)* (Figure 5.9a), (B) prominent regions of peculiar patterns and designs *(total n. 11)* (Figure 5.9b), (C) highlighted areas with particular conservation states (e.g., showing well preserved parts for virtual reconstruction or their dimensions) *(total n. 54)* (Figure 5.9c), (D) visual pointers of biological phenomena unseen to the naked eyes *(total n. 13)* (Figure 5.9d), (E) frames marked with exclusive segments for additional historic and sculpting details (e.g. highlighting fine carving details over the surface together with information about sculpture techniques and tools) supported by pictorial references or images (e.g., comparison with small bronze statues) *(total n. 08)* (Figure 5.9e). Each annotation contains both a visual markup, intended lens position and rendering parameters, and an explanatory audio clip.

5.7.2 Scoring system analysis

The proposed framework allows one to mix purely automatic navigation with interaction, since the user may take control of the lens during any path, and auto navigation restarts from the new user-updated lens and view configuration. We show this behaviour in Figure 5.11. The transitions marked with red arrows depict lens movements/positions

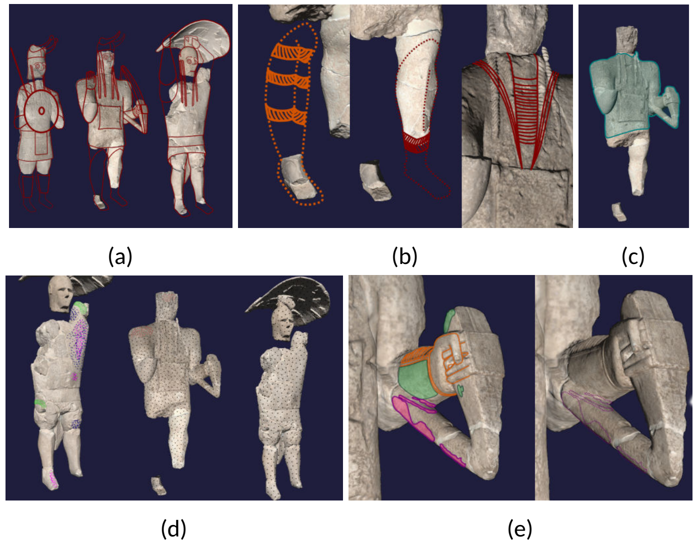


Figure 5.9: **Annotation Classes**. Derived from the literature [93], we create a variety of annotation classes i.e., (a) graphical extensions of missing parts, (b) regions of peculiar patterns and designs, (c) highlighted areas with particular conservation states, (d) visual pointers to regions interested by biological phenomena hardly visible to the naked-eye, and (e) historic and sculpting details.

decided by the user (not by the automatic generated path); the next annotation selected by the automatic algorithm (transitions marked with green arrows) takes into account the dependency graph and the history, while being consistent with the lens positioning provided by the user. In Figure 5.11, after the first three automatic frames, the user interrupts the automatic navigation three times, in order to move and inspect all the three statues. The accompanying video shows additional examples of this behavior, in which we seamlessly move from automatic touring to interactive exploration, and, each time, the tour restarts taking into account the possibly largely modified local context.

In order to evaluate the behavior of our scoring system, we tested the automatic navigation without the free movements introduced by the manual exploration (see subsection 5.7.3 for a detailed analysis of users' interaction and their subjective validation).

In this setup, since the graph was authored with a single root node (the statue overview) required for all further inspection (dependency weight=1), we expect that the navigation always starts from the root and, from there, a relevance-based order would be followed by navigation, taking into account graph hierarchy and node/edge priorities. In addition, we expect our navigation to enable a good level of variety, due to the stochastic aspects of our annotation selection (subsection 5.5.3). We performed 20 automatic tours, each of them visiting 20 nodes, always starting from an initial position at the center of the screen and viewing all the visible scene in the viewport. Despite the same initial conditions, the 20 tours visited a total of 69 nodes with respect to the 108 contained in

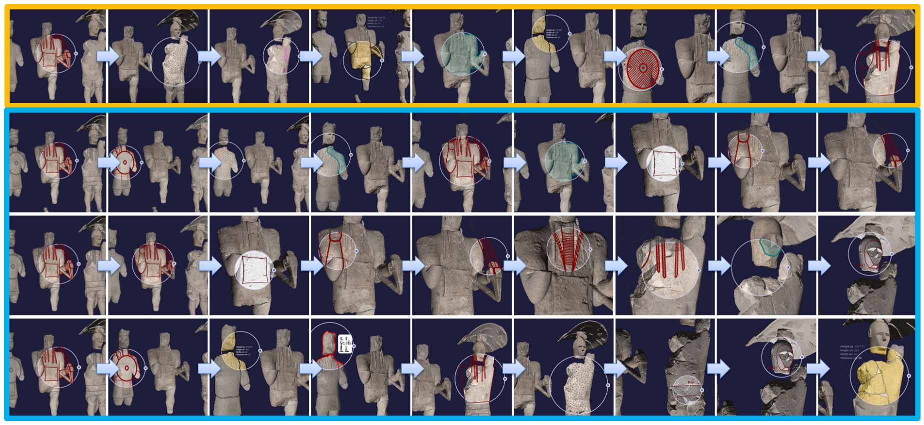


Figure 5.10: **Automatic navigation**. Top row (yellow outline):an example of automatic navigation without using the dependency graph. The path proceeds by going from an annotation to the most similar one, without taking into account semantic aspects (e.g., same statue, from more general to specific annotation). Other rows (blue outline): several examples of automatic navigation with the dependency graph. All exploration paths start from the same annotation, and all tours share a similar flow, dictated by authored graph dependencies. Nonetheless, they introduce variations due to our stochastic next-best annotation selection process. The dependencies introduce semantic aspects, in this example favoring the presentation of a statue's detail after presenting its overview.

the graph. This fact shows how a stochastic component in path selection avoids full repetitiveness, providing different exploration experiences to the users, also in fully automated mode.

Figure 5.10 (bottom three rows) shows three runs of the auto-navigation, with time going from left to right. It is clear how the first view is always the same, i.e., the graph root presenting the annotation related to the whole set of statues. Even if the lens is centered, providing a higher node priority to the center statue, there are several situations in which one of the other statues is selected first, due to our selection strategy with picking probability proportional to weight and our loose β value in this mode of operation. From there, the navigation continues with a spatial and semantic consistency,

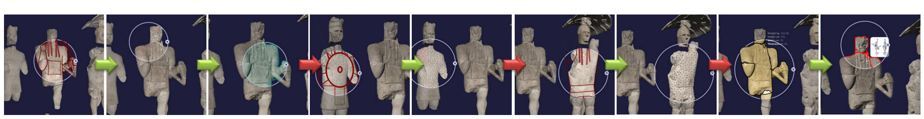


Figure 5.11: **Mixing automatic and free exploration**. Our framework enables both automatic and free navigation. As soon as the user moves the lens (transitions marked with red arrow), the automatic navigation stops. When it restarts (transitions marked in green), the next frame is selected by taking into account both the dependency graph, the navigation history, and the user-updated lens and view configuration.

e.g., if a high-level node of a statue has been visited, the navigation continues with higher probability in the leaf nodes of that same statue. The semantic choice cooperates with the spatial one; if the visualization of a statue's annotation enables the visiting of the detail nodes of that statue, the visiting of nearby details of another statue still remain blocked until their enabling nodes have been enabled. Thus, the authored hierarchical grouping of the annotations enables the introduction of semantic constraints.

Conversely, when edges are not present, as in previous work on lens navigation [22], such constraints are not possible, and the next best annotation in a navigation path may be selected from a nearby statue based on pure proximity consideration. We repeated the 20 runs with 20 annotations each, using the same database, but with edge dependencies disabled. In such a situation, we explored a total of 72 nodes. Since more degrees of freedom are available due to the lack of edges linking to enabling nodes, the number of visited nodes is slightly larger. However, the paths are less structured, as they jump more frequently, for instance, from one statue to another. The first row of Figure 5.10 shows an example of that kind of navigation, where edges are removed, and navigation proceeds purely by selecting the most similar annotations. Without taking into account a hierarchy of nodes, the storytelling aspect of the automatic navigation might get lost, as also demonstrated by our dedicated user study (section 5.7.3).

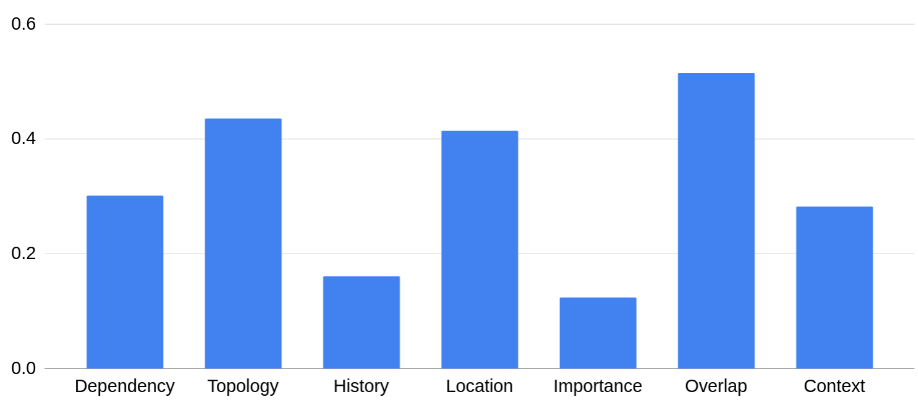


Figure 5.12: **Score vs Weights Correlation**. We show the Pearson correlation coefficient between the final annotation score and each factor that contributes to that score. We can see that the three most important factors are the *Overlap*, *Topology*, and *Location* weights.

In order to better understand how the next annotation is selected, and to have a more clear idea about the contributions of the single weights to the final annotation ranking score, we also launched several autotours, and we collected a series of data to compute the Pearson correlation coefficient $\rho_{S,x} = {cov(S,x)}/{\sigma_S \sigma_x}$ between the final score S and the individual components x of the scoring system of Equation 5.1. In particular, we consider the author-defined annotation importance γ, the history score H, the dependency δ, the topology weight τ, and the three weights that produce the similarity value σ, i.e., the lens overlap σ_{lens}, the context overlap σ_{cont}, and the location similarity σ_{loc}.

The results are reported in Figure 5.12. As we can see, there is a good balance in the different terms, but the three most important factors are the *Overlap*, *Topology*, and *Location* weights. It is important to note that overlap and location are closely related to selection by spatial proximity, while topology relates to semantic continuity.

5.7.3 User study

The proposed navigation framework has been obtained by combining two main elements. From the narrative side, we have the audio-visual annotations and their structured organization, while from the purely visual data we have the multi-layered 2D model. From the user point of view, it is extremely challenging to assess and validate the combination of a narrative element and the interface that drives the communication between that and the user. This is mainly due to the lack of reliable and standard metrics or practices that have enough consensus when assessing interface usability together with content user understanding. Considering and evaluating each part separately might be a good way to quantify their contribution to the user experience [106]. However, this approach does not take into account the effects that only arise because of the combination of several elements. If those components increase in number, the combinatorial nature of the problem makes the evaluation even more complex, unreliable, and non-practical.

For this reason, we concentrate here on answering two main research questions that are connected to the main differences between this work and previous ones.

The first question is whether the introduction of an annotation graph, with edges connecting annotations to their enabling nodes, leads to explorations that are perceived as an important improvement over presentations using methods that only considered an individual list of annotations (e.g., [22]). This question is explored through our *Autotour* test (section 5.7.3).

The second question is, instead, related to the overall user experience. In particular, we want to investigate whether casual users remain active or passive in presence of system that provide both automatic touring and interactive exploration, and if they prefer a system that actively follows their actions or prefer to limit their interaction to local investigations inside an inflexible authored story. This question is explored through our *Interaction* test (section 5.7.3).

Autotour Test

Goal The purpose of this test is to understand, from the user perspective, if the introduction of the structured annotation graph increases the user experience during the *Autotour* mode compared to an automatic navigation that is free and ignores the semantic relationships between annotations. In this setup, in order to have a more controlled experiment, user interaction is not considered.

Configurations We produce two types of videos of automatic exploration of the digital model. One is obtained by launching the *Autotour* mode that uses the structured

S1-G	GRAPH Visit is more engaging than FREE Visit.
S2-G	GRAPH Visit held my attention more than FREE Visit.
S3-F	GRAPH Visit is more boring than FREE Visit.
S4-G	Information presentation is more clear in GRAPH Visit than in FREE Visit.
S5-G	GRAPH Visit presents the information more organically than FREE Visit.
S6-G	GRAPH Visit provided me with more intellectual stimulation than FREE Visit.
S7-G	GRAPH Visit, more than FREE Visit, motivated me to learn more about the Mont'e Prama collection
S8-F	GRAPH Visit, more than FREE Visit, presents the cultural heritage content in a more scattered way.
S9-F	GRAPH Visit is more distracting than FREE Visit.
S10-G	The story told by GRAPH Visit is better structured than that told by FREE Visit.
S11-G	In GRAPH Visit I gained more knowledge than in FREE Visit.
S12-G	I enjoyed GRAPH Visit more than the FREE Visit.

Table 5.1: **Autotour Test - Statements.** List of statements in the *Autotour* evaluation Likert-scale questionnaire. In order to avoid agreement bias, half of the participants were presented the questions in their reverse form, i.e., swapping GRAPH and FREE as the preferred method.

annotation graph, while the other produces an automatic exploration by completely ignoring annotation semantic relationships encoded in edges, therefore approximating the method of Bettio et al. [22] that works on a flat annotation database. The content presented in the two videos starts exactly from the same database in terms of visual representations (shape, color, illumination, etc.), texts, drawings, and audio clips. The main difference between the two videos is the way the information has been selected and organized for presentation. Each visit has an equal length of two minutes. We have produced many different *Autotour* navigations with the two modalities, in order not to have biases produced by a particular exploration run. We call the two modalities *GRAPH* visit and *FREE* visit.

Tasks We ask participants to take a look at the two virtual visits of a set of three statues from the Mont'e Prama collection, and to build an opinion on which of them they prefer. We will ask them several questions to understand that opinion.

Design The test is subdivided in three phases. We first ask general questions to the users, in order to understand the type and distribution of the participants. In the second part, we blindly show participants two videos of a navigation of a digital model. The participants don't know which video is the *GRAPH* or *FREE* mode; the users don't even

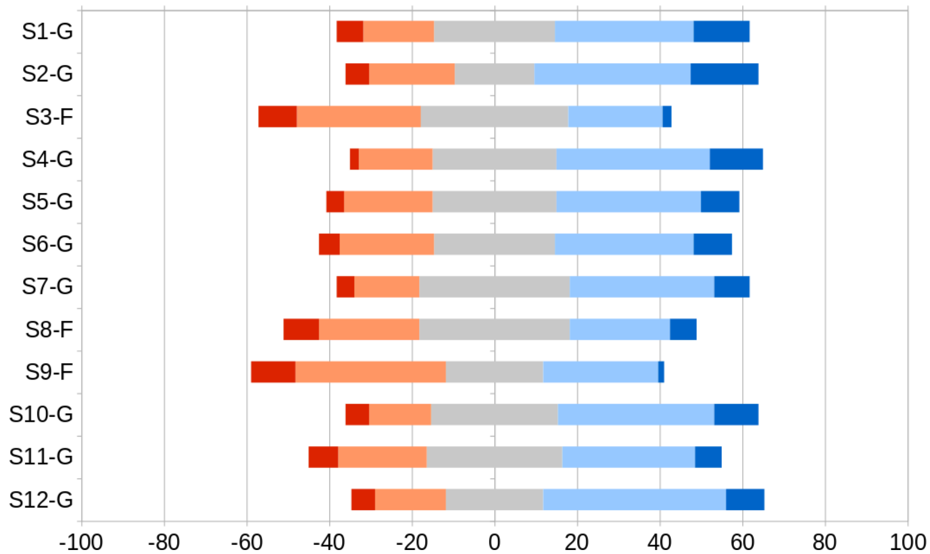


Figure 5.13: **Autotour Test - Evaluation**. Histograms of responses for the statements in Table 5.1. Responses are color mapped from left (dark red, *Strongly Disagree*) to right (dark blue, *Strongly Agree*).

know the details of the two modalities, they only know that the two videos present different model explorations. From the produced videos in the two modalities (with or without structured annotation graph), for each user we randomly pick one example from each modality, and we randomized the video presentation order. Finally, we ask several questions to understand which videos/exploration they have preferred. We design the questionnaire as a Likert Scale [107] form with a series of twelve statements (Table 5.1), with five possible choices, i.e., *Strongly Disagree*, *Disagree*, *Neutral*, *Agree*, *Strongly Agree*. The statements are marked as $SX-G$ or $SX-F$ depending on the fact that a positive feedback is respectively given to the proposed solution (*GRAPH*) or the reference navigation strategy (*FREE*). The questions are inspired by Othman's work [108] about measuring visitors' experiences and engagement in museum visits. To avoid the agreement bias, i.e., the tendency of a respondent to agree with a statement when in doubt, half the respondents were presented with the questions reversed, i.e., swapping *GRAPH* and *FREE* references in their formulation. To simplify the presentation of the results, we have transformed back the order for the half cases in which we inverted the G/F order. In addition, some questions are slightly similar or opposite to each other to create a redundancy that is useful to test if the user has given consistent responses. We take this into account in the computation of the questionnaire consistency score.

Participants The group of participants consists in 140 users (65% female and 35% male). The 3.6% are high school graduates, 5% with an associate's degree, 29.3% with a Bachelor's degree, 41.4% with a Graduate or professional degree, and 20% have a PhD. About 84% have a STEM background, while 10% of them come from the Humani-

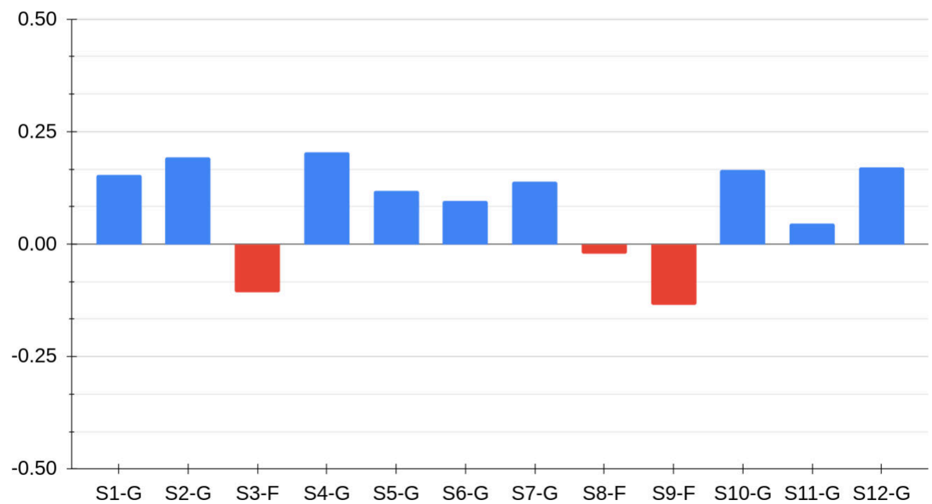


Figure 5.14: **Autotour Test - Statements Score**. Scores obtained by each statement in Table 5.1. Positive scores mean agreement, while negative scores mean disagreement. In blue are statements that favor the *GRAPH* visit, while in red are those that favorite the *FREE* visit. In all statements, users agree that *GRAPH* visit is better than *FREE* one.

ties field. They were recruited using a mailing lists across various leading institutions involved in both Computer Science (specifically Computer Graphics and Visualization), CH studies, and applications. Through direct mailing, we have also tried to include participants representative of the general public, with a more heterogeneous background. They are researchers (15%), students (22.1%), teachers/professors (22.1%), IT professionals (2.1%), developers (5.0%), house wives (4.3%), and others (29.3%), which include freelancers, technologists, managers, and unemployed people. The age is ranging from 18-25 (41.4%), to 26-35 (29.3%), 36-50 (18.6%), 51-64 (10%), and over 65 (0.7%). We also have a heterogeneous set of people in terms of familiarity with museums/exhibitions and virtual museum presentations. About 60% of them have visited a museum last year, but 45% of them have no familiarity with virtual museum presentations; for the 60% of them, this is the first time they try an interactive setup similar to that proposed in this thesis. Finally, half of them did not have any knowledge of the Mont'e Prama collection presented in the test.

User evaluation We evaluate the *Autotour* test from three points of view, i.e., graphically, by a scoring system, and by computing the Cronbach's alpha reliability of the questionnaire. We choose the Cronbach's alpha since it is the most commonly used metrics to assess the internal consistency of a questionnaire made up of multiple Likert-type scales [109], [110]. First, we plot the histogram of responses for each statement (Figure 5.13). The responses are color mapped from left (dark red is *Strongly Disagree*) to right (dark blue is *Strongly Agree*. It is clear how the $SX-G$ statements are more towards the *Agree* and *Strongly Agree* part, while the $SX-F$ statements contain more disagreement form the user. These results confirm that the user prefers more the *GRAPH* than the *FREE* navigation; this can be seen in the last statement, which explicitly ask the user the preference between the two exploration strategy. Here, 53.6% of the participants prefer the *GRAPH* Visit, 23.6% have a neutral opinion, while only 22.8% prefer the *FREE* Visit. In order to assign a numerical score to each single statement and a global score to the entire test, we linearly map each of the five responses to scoring value, as typical for Likert scales [107]. In our case, respectively, from *Strongly Disagree* to *Strongly Agree*, we assign -1, $-1/2$, 0, $1/2$, 1 values. The statement score is the average of the responses received by participants. As illustrated in Figure 5.14, the statements marked as $SX-G$ obtain a positive score, while the statements that judge positively the *FREE* Visit (marked as $SX-F$), received a negative score. This, again, confirms that in each statement the users prefer the *GRAPH* Visit. In order to compute the final global score, we take the average of statement scores, after negating those marked as $SX-F$, obtaining a value between -1 and 1. A positive global score would mean that the users prefer our proposed automatic exploration system, a negative score that they prefer the other one, while a close to zero score would mean no preference. The final global score is 0.55, showing a very marked preference for the *GRAPH* version. We found that the reliability of the questionnaire is very high, with a Cronbach's alpha equal to 0.91. Since some questions are by design redundant, and since this can cause a bias in the Cronbach's alpha computation, we have also estimated the reliability by removing statements $2, 5, 10$, and 12; the Cronbach's alpha becomes 0.81, which is still very high. The user test thus confirms that the more coherent order induced by the graph, as evaluated in subsection 5.7.2, leads to a perceivably improved experience.

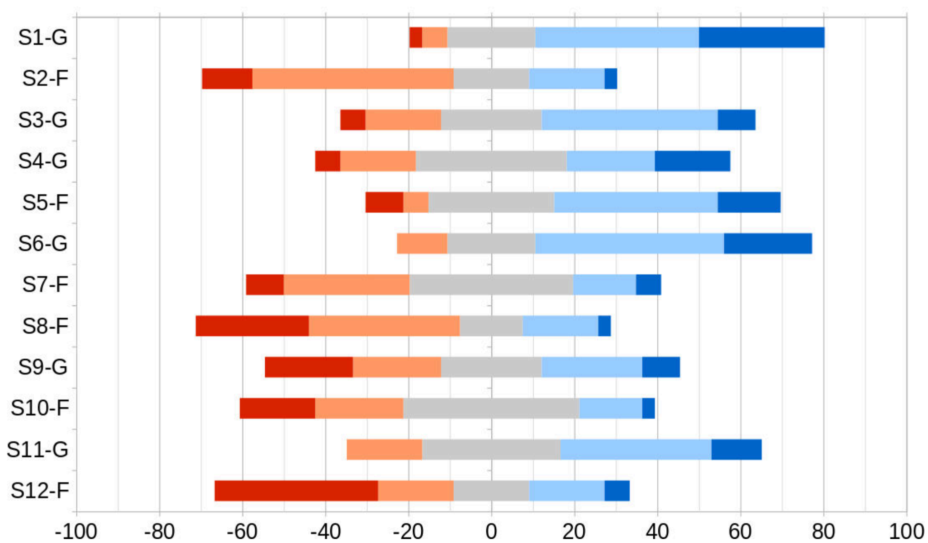


Figure 5.15: **Interaction Test - Evaluation**. Histograms of responses for the statements in Table 5.2. Responses are color mapped from left (dark red, *Strongly Disagree*) to right (dark blue, *Strongly Agree*).

S1-A	ADAPTIVE exploration is more engaging than FIXED exploration.
S2-F	FIXED exploration held my attention more than ADAPTIVE exploration.
S3-A	FIXED exploration is more boring than ADAPTIVE exploration.
S4-A	Information presentation is more clear with ADAPTIVE exploration than with FIXED exploration.
S5-F	ADAPTIVE exploration, more than FIXED exploration, guides you toward new annotations far from the region you want to explore.
S6-A	ADAPTIVE exploration provided me with more intellectual stimulation than FIXED exploration.
S7-F	FIXED exploration, more than ADAPTIVE exploration, motivated me to learn more about the Mont'e Prama collection.
S8-F	FIXED exploration, more than ADAPTIVE exploration, follows better your exploration intention.
S9-A	FIXED exploration is more distracting than ADAPTIVE exploration.
S10-F	They story told by FIXED exploration satisfies you more than that told by ADAPTIVE exploration.
S11-A	With ADAPTIVE exploration I gained more knowledge than with FIXED exploration.
S12-F	I enjoyed more FIXED exploration than ADAPTIVE exploration.

Table 5.2: **Interaction Test - Statements.** List of statements in the *Interaction* evaluation Likert-scale questionnaire. In order to avoid the agreement bias, half of the participants were presented the questions in their reverse form, i.e., swapping FIXED and ADAPTIVE as the preferred method.

Interactive navigation test

Goal We aim to compare a classical exploration based on fixed authored tours with the new proposed solution where the tour is adaptively adjusted in response to user actions. For more generality, rather than restricting our comparison to the fully static video presentations proposed by systems such as CHER-Ob [88], we chose as a term of comparison the slightly more flexible interruptible video navigation method popularized by ArtMyn [111], which allows users to pause the video presentation to perform local exploration. In addition to analyzing user preferences, we also want to investigate whether casual users remain active or passive in front of these presentation systems.

Configurations We configure two types of interaction experiences. In the first one, called *FIXED*, we show participants an image of an artwork, and we guide them through a pre-established and pre-recorded (but interruptible) navigation of a sequence of annotations attached to it. At any moment, the user can interrupt the navigation, and interact with the virtual environment to inspect the database, e.g., to look in more detail at some areas discussed in the pre-recorded story. After interaction, the automatic navigation continues from the point where it was stopped (as in the navigation method

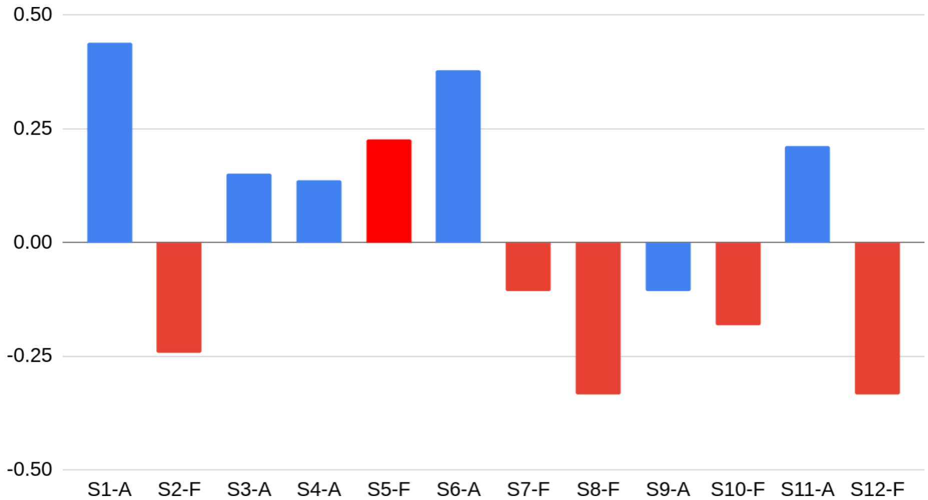


Figure 5.16: **Interaction Test - Statements Score**. Scores obtained by each statement in Table 5.2. Positive scores mean agreement, while negative scores mean disagreement. In blue are statements that favorite the *ADAPTIVE* exploration, while in red are those that favorite the *FIXED* navigation. Apart from statement 5 and 9, users agree that *ADAPTIVE* visit is better than *FIXED* one.

popularized by ArtMyn [111]), following the fixed pre-recorded annotation path. In the second interaction test, called *ADAPTIVE*, we adaptively select and present annotations with the methods presented in this chapter, which allow users to freely mix interaction with guided touring.

Tasks The experiments consisted in letting users to freely explore the annotated sculptures, after a minimal training and without external direction. Users were told that their goal was simply to enjoy the experience and acquire information at their own pace in a prescribed short limited amount of time, exploiting the audio-visual annotations provided by the system, and using the interaction capabilities of the lens-based interface. This reflects well the scenario of a walk-up-and-use experience in a museum setup, as well as the situation encountered in museum web sites.

Design The test is subdivided in two phases, focused on the interface usability and the presentation rationale/order. In the first phase, each participant actively tests the two exploration modalities, i.e. *FIXED* and *ADAPTIVE*. The modalities are presented to the participant in a random order. First, we make users familiar with the interface and the navigation task; so, before the actual test, users receive a one-page instruction describing the overall test, the interface, and the user-interface mapping; they are also allowed to test the tool without performing the task. After that, the exploration task is performed with the two configurations, and a series of variables are recorded to measure the user experience, i.e., number of annotation visited (manually or during an autotour), autotour or interaction time, etc. Each test has a fixed duration of 3 minutes. At the end of both the two interaction experiments, the participants were asked to fill a Likert scale questionnaire with five options for each question, i.e., *Strongly Disagree*, *Disagree*, *Neutral*, *Agree*, *Strongly Agree*. The statements are marked as $SX-A$ or $SX-F$ depending on the fact that a positive feedback is respectively given to the proposed solution (*ADAPTIVE*) or the reference navigation strategy (*FIXED*). The twelve statements of the questionnaire are shown in Table 5.2. The type and order of the statements are designed and presented to the user with the same rationale of the previous test (see section 5.7.3), including the strategy to avoid the agreement bias. To simplify result presentation, questions are presented here in their canonical form.

Participants The group of participants consists in 33 users (21.2% female and 78.8% male) recruited among students, families and friends of researchers working at our center. All subjects had normal or corrected to normal vision and, as now extremely common, had basic computer or smartphone literacy. The 6.1% are high school graduates, 24.2% with a Bachelor's degree, 33.3% with a Graduate or professional degree, 33.3% have a PhD, and 3.0% prefer not to answer. About 94% have a STEM background, while 3% of them come from the Humanities field. They are researchers (54.5%), students (18.2%), teachers/professors (6.1%), developers (12.1%), and others (9.1%), which include home workers, designers, and administrators. The age are ranging from less than 18 (3.0%), 18-25 (9.1%), 26-35 (24.2%), 36-50 (48.5%), and 51-64 (15.2%). We also have an heterogeneous set of people in terms of familiarity with museums/exhibitions and virtual museum presentations. About 80% of them have visited a museum last year, and 9.1% of them have no familiarity with virtual museum presentations; for 15.2% of them, this is the first time they try an interactive setup. Finally, only 6.1% of them did not have any knowledge of the Mont'e Prama collection presented in the test.

User evaluation As for the previous test, we evaluate the *Interaction* test from three points of view, i.e., graphically, by a scoring system, and by computing the Cronbach's alpha reliability of the questionnaire. First, we plot the histogram of responses for each statement (Figure 5.15). The responses are color mapped from left (dark red, *Strongly Disagree*) to right (dark blue, *Strongly Agree*). It is clear how the majority of $SX-A$ statements are more towards the *Agree* and *Strongly Agree* part, while most of the $SX-F$ statements express a disagreement form the user. While responses are generally consistent, we discovered that some questions are not clear to some of the users. $S5-F$ asks the users if the *ADAPTIVE* exploration leads farther away from the region they want to explore than the *FIXED* path navigation; *ADAPTIVE* exploration typically remains close to the position of the user, while *FIXED* mode will continue to the next pre-defined annotation, completely ignoring user will. In fact, when the statement expresses the same concept, but in a different way, such as the statement $S8-F$, some users recognize that the *FIXED* exploration does not follow the participants exploration intention better than the *ADAPTIVE* exploration. Although users gained more knowledge from the *ADAPTIVE* exploration ($S11-A$), and found the proposed solution more satisfactory ($S12-F$,$S6-A$) and engaging ($S1-A$). Nonetheless they found the *ADAPTIVE* slightly more distracting than the *FIXED* one ($S9-A$). The effect is very small, and it is not evident to judge the reason as no specific comments were made. Our hypothesis is that, especially for naive users, the reduced set of possibilities offered by the *FIXED* exploration requires less learning and mental effort to manage navigation decisions, reducing the mental mode switches from fully guided to interactive. In any case, the unambiguous final statement demonstrates that the users prefer more the *ADAPTIVE* modalities than the *FIXED* one. In particular, 63.6% of the participants prefer the *ADAPTIVE* navigation, 9.1% does not have a preference, while 27.3% prefer the *FIXED* alternative. We assign a numerical score to both each single statement and a global score to the entire test similarly to the previous test. As done in the *User evaluation* of section 5.7.3, we convert qualitative responses to numerical values, thus obtaining the final scores per statement in Figure 5.16. Even with the presence of the inconsistently interpreted statements, most of the statements marked as $SX-A$ obtain a positive score, while most of those that judge positively the *FIXED* exploration (marked as $SX-F$), received a negative score. This, again, confirms the strong preference towards the *ADAPTIVE* solution. In order to compute the final global score, we sum all the single statement scores, by multiplying by -1 those marked as $SX-F$, and we remap between 0 and 1. A positive global score means that the users prefer our proposed automatic exploration system, otherwise they prefer the other one. The final global score is 0.58, showing a large preference. We found that the reliability of the questionnaire is very high, with a Cronbach's alpha equal to 0.91 for all questions, and 0.87 after the removal of redundant statements, i.e., $2, 5, 10$, and 12.

Usage statistics The different perception of the two modalities is also reflected in a different usage pattern, despite the very similar control interface. On average, users spend considerably more time interacting using the *ADAPTIVE* solution. On average, 41.6s (median 40.5, min 0, max 108) are spent actively moving the lens, against the 29.2 for the *FIXED* solution (median 28, min 0, max 77.3). Interestingly, there has been a user that in both cases remained completely passive, just listening to the story without ever attempting to move the lens. The interactive exploration in the *ADAPTIVE* solution also leads to a slightly larger number of annotations visited. On average, 16.6 annotations (median 17, min 9. max 27) are presented (with overlay and audio explanation) against the 15.2 (median 14, min 8, max 51) for the *FIXED* version. The higher number of annotations is due to the dynamic activation of new annotations when the users explore new areas. Again, here, it is interesting to note the singular behavior of a user (one of the two that never attempted to move the lens) that continuously skipped to the next annotation at maximum speed (reaching the max of 51 annotations displayed and played in 3 minutes).

Free comments After the experiments, we collected in the web forms a series of comments about both the *FIXED* and *ADAPTIVE* explorations. Several users explicitly stated that the *FIXED* modality is a little confusing, since they ”have no control over the system. The system just explains things and the only thing the user can do is skip the explanations”. They find it annoying that, when they move around, they can only explore the model without having any explanation of what they are looking at, so they find it of little use to let the user move around interrupting the guided tour.

Similarly, other users complain about the fact that they don't find too intuitive how to guide the interaction; they could easily find regions of interest in the three statues, but couldn't find a way to visualize their details by themselves without waiting for the auto-tour to show them (if the *Autotour* decides to show them). Moreover, when they focus on a detail they continue to hear another audio explanation from the pre-defined series of visual/audio annotation. From this perspective they much preferred the *ADAPTIVE* configuration, finding it pretty nice and flexible. They can enjoy jumping from an annotation to another without waiting for the predefined path tour. So the *ADAPTIVE* exploration allows them to inspect much more details. We can also conclude that the *FIXED* exploration is more geared towards a mostly passive experience, with little differences than watching a video. Finally, several comments suggested possible changes in the interface implementation. For the *FIXED* modality they ask to provide more visual cues and colors. Conversely, for both modalities, they find that would be useful to speed up the annotation time (it takes too long to change the lens color), to allow users to move not only the lens but also the background scene, and to change the glyph for the `Done` button (they think that an X isn't the perfect button for this action). While in preparing the annotations we favored the audio and visual overlays to avoid clutter, a part of the users suggested us to add some more text to the annotated regions, since they think it could make the CH content easier to understand. Concerning the audio annotation, users suggest to fade that out smoothly when changing the annotation, rather than stopping it abruptly. Moreover, one user suggested to include a list of available annotations displayed somewhere in the screen (e.g. a thumbnail bar) to complement the current presentation display. Most of these suggestions point to aspects orthogonal to this work, and might be integrated in future versions of the system.

5.8 Discussion

We have proposed a framework that aims at presenting annotations in a structured way. The approach is meant to support casual users to explore, at their own pace, spatially annotated 2D models using an interactive lens that moves from an interesting area to the next, while also responding to user inputs, following shifts in interest and attention. The presentation order is dynamically dependent on lens position, navigation history and authoring information encoded in an annotation graph. The integration of a stochastic recommendation system that interprets context-dependent scores as transition probabilities makes it possible to increase the variability of exploration paths. Moreover, the user can freely mix personal/free exploration with automatic touring.

Our very preliminary evaluation has shown the potential interest of the approach, but also highlighted areas for future research. First of all, the current approach is targeted towards the exploration of areas that fit well on a circular lens, but should be refined when pointing at areas where linear or extended features should be explored. We plan to address this problem by storing at each node not only a single lens position, but a lens path for the exploration of the annotated area. Second, the dependencies presented here currently target the definition of simple precedence relations expressed by taking the fuzzy AND of values coming from enabling nodes. It is worth exploring whether fully supporting other logical operators (i.e., at least OR, XOR, and NOT) would be beneficial for improving the authoring expressiveness. Edges, in addition, might also benefit from being augmented with audio information, which could be played when a particular transition is activating, extending the current experience that limits audio clips to individual annotations. This latter feature, while interesting, is feasible within the current system for fully automatic transitions that move from one node to the next, but might require special care to be integrated with free-form lens motion.

The proposed annotation graph, state machine, and navigation interface have been applied in this work to interactions on an image plane. Such a 2D interactive exploration is natural for 2D objects, and is often applied also to fixed views of general 3D objects. The relightable 2.5D dataset used in this work is a typical example. A particularly interesting extension would be to apply this work to full 3D models using a less constrained interface. While, from the annotation point of view, our proposed concepts should already support a direct 3D extension, the interactive control and guiding components would need to be significantly extended. First of all, interactively manipulating lenses on 3D models require special care. Several solutions have been proposed to control lenses in screen-space (e.g., [112]–[114] or object-space (e.g., [75], [115]), but none of these techniques seamlessly supports navigation on multiple models with coupled lens and camera control. How to control a lens while keeping an effective focus and context situation is an open problem in 3D. In terms of guidance, moreover, the various terms used for determining the next best lens would need to be adapted to 3D, in particular taking into account 3D visibility. A starting point could be the work done by Balsa et al. [97] for camera navigation.

Moreover, authoring, orthogonal to this work, also deserves attention, in particular in case of extension of the dependency logic. Finally, our current evaluation was very preliminary, and focused mostly on responding to our the main research questions, i.e., whether the presence of dependency among annotations perceivably improves the experience, and whether users enjoy our flexible interactive or mostly interactive tours better than the more standard fixed auto-touring features. More work is required to objectively assess the effectiveness of our user interface. It will be also interesting to evaluate whether the proposed approach, currently tuned to museum applications, can be extended to more complex situations requiring specific visualization tasks to be solved or particular user needs to be addressed.

5.9 Bibliographic notes

The research presented in this chapter was originally presented in our STAG2021 contribution [28], that received the Honorable mention in best paper award category. The approach was later very significantly extended and published in the Computer & Graphics Journal [29]. The content of this chapter is based on the journal version. Video demonstrations are included as additional multimedia together with the original publication [29]. I have significantly contributed to the conceptualization, methodology, and validation of the method and was one of the primary authors of these papers. A later simplified refinement has been applied to the exploration of a very large annotated artwork for an exhibition [1]. For this latter paper, my contribution was only in the application of the previously designed approach based on lenses and graph-based annotation databases to this particular use case.

Chapter 6

Cultural heritage pilot

This thesis has been carried out in the context of the European Union's H2020 research and innovation program grant 813170 (EVOCATION). Within this project, we have carried out a pilot to evaluate the application of our research to two use cases in the cultural heritage domain. I have, in particular, participated to the acquisition of cultural heritage items, to the creation and annotation models, and to creation of interactive application for their exploration using the lens-based techniques introduced in this thesis. This chapter briefly summarizes the work performed and the results achieved.

6.1 The EVOCATION pilots

The EVOCATION project, within which this thesis was carried out, targeted the creation of shareable representations of shape and material models from high-fidelity capture, as well as their annotation and virtual presentation in web clients.

In this chapter, I discuss two use cases performed at the National Archaeological Museum of Cagliari (Italy), an associated project partner, that illustrate the entire pipeline up to virtual presentation, and use, for that purpose, the methods and techniques that were developed within this thesis. We are grateful to the staff and curators of the National Archaeological Museum, Cagliari (Museo Archeologico Nazionale di Cagliari) and its National Gallery (Pinacoteca Nazionale di Cagliari) for their support in access to the artworks for the purpose of digitization and for annotation information.

6.2 Retablo S. Bernardino

The first pilot has been devoted to the capture and visualization of paintings held at the National Gallery (Pinacoteca Nazionale) of Cagliari (Italy), one of the expository sites of the National Archaeological Museum of Cagliari. The museum is well known

Figure 6.1: **Capture and relighting of a painting (panel of the polyptych retable of Saint Bernardino (1455), Cagliari, Italy)**. The optical response of the painting surface to variable illumination is measured by taking a few tens of photos using a fixed reflex camera and a hand-held LED (left). The Multi-Light Image Collection is then transformed to a shape and material representation, which is used for interactive relighting (right)

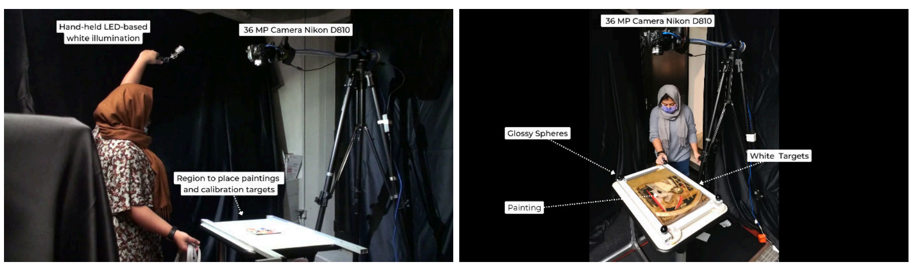


Figure 6.2: **Capture and relighting of a painting (panel of the polyptych retable of Saint Bernardino (1455), Cagliari, Italy)**. Details of the components of the capture setup.

for the display of several paintings of Sardinian artists dated from 16^{th} to 20^{th} century, pictorial works (15^{th} to 18^{th} century) of the Genoese, Neapolitan, and Roman school, and some Sardinian and Catalan retabli of the 15^{th} to 16^{th} centuries.

Given the current research and restoration activity performed by scholars on the set of retabli, we have applied the techniques developed in the project to the on-site capture, reconstruction, and web-based multi-layer visualization of the shape and material (BRDFs) of two panel parts of the *Retablo S. Bernardino* (1455). This retablo is a polyptych originally from the chapel of St. Bernardino in the St. Francesco church in Cagliari, Italy, and currently preserved and displayed at the Pinacotheca. The first analyzed panel, measuring 34x25cm, depicts the prophet Daniel, while the second, measuring about 53x35cm, shows the piety of Christ. Both are painted in oil on a wooden support.

Both paintings have been acquired in a completely dark room, by performing a free-form RTI setup, with a 36.3 Mpixels DSLR FX Nikon D810 Camera with a $50mm$ AF Nikkor Lens and a handheld white LED (5500K) covering the entire visible spectrum (Figure 6.1

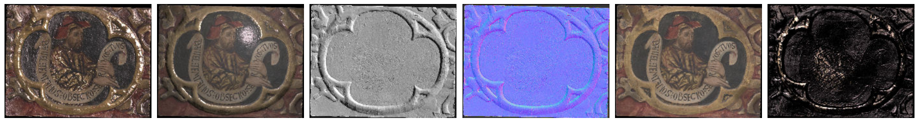


Figure 6.3: **Retablo S. Bernardino - Prophet Daniel**. From left to right, top to bottom: relighting of the computed SV-BRDF by using a directional light; relighting of the computed SV-BRDF by using a spot light; monochromatic rendering; normal map layer; diffuse/albedo map; relighting of the specular component by using a directional light.

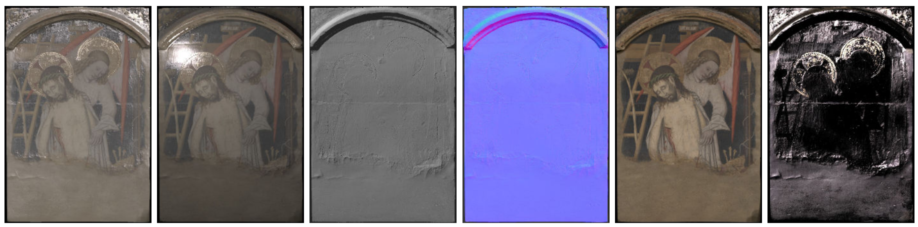


Figure 6.4: **Retablo S. Bernardino - Piety of Christ**. From left to right, top to bottom: relighting of the computed SV-BRDF by using a directional light; relighting of the computed SV-BRDF by using a spot light; monochromatic rendering; normal map layer; diffuse/albedo map; relighting of the specular component by using a directional light.

left). I have performed the capture at the museum together with my colleagues Ruggero Pintus and Antonio Zorcolo. We have acquired about 60 images for each MLIC. The acquired data has been calibrated with four glossy spheres (for light direction), and with a gray frame positioned around the object (see Figure 6.1 left), using the camera and light calibration method recently presented by Pintus et al. [116]. All the components of the capture setup are depicted in Figure 6.2.

The extraction of the shape, in terms of normal maps, and the optical material behaviour, represented by a spatially-varying BRDF, has been performed by employing the algorithms presented in our GCH 2021 contribution [117]. For that work, I did not contribute to the design of the technique, but I supported the generation of results.

The final processed data is represented by a multi-layer data structure, and we used the tools and techniques discussed in chapter 3 to allow users to explore the model at multiple scales, performing relighting, layer selection and combination, and focus-and-context examination with visualization lenses.

In our implementation, used for the retablo inspection, we enable the concurrent visualization of a base layer and a single lens layer, and we use visualization modes fully configured using presets. In this scenario, various rendering modes are carefully authored at data preparation time, and we finally rely on a multilayered dataset description and on two lists of preset rendering configurations, one for the base layer and one for the lens layer. Such simplified setup makes it possible to target a non-expert audience, with a wide range of possible users. In addition, this viewer is complemented

Figure 6.5: **Retablo S. Bernardino**. Multi-layer visualization through a lens tool in a focus-and-context setup. The context consists in the rendering of the computed SV-BRDF by using spot (left) or directional (right) light. Inside the lens we render a different layer or use a different rendering mode, in order to facilitate the inspection of some shape and material characteristics of the surface (e.g., normal map and surface roughness).

by two other interaction widgets: one for enabling/disabling the lens and another for switching from relighting mode to measurement mode. In relighting mode, one-finger dragging is mapped to light direction control using the standard trackball interface. In measurement mode, the single touch operation allows us to display $\Phi_d = 90$ slice of the BRDF under the cursor, which is a good indicator of the main reflectance features of the underlying material. For the general interaction with the object and the lens, and for the navigation across the dataset, we designed the viewer behaviour in order to reduce training time, by relying on analogs of common 2D Rotate-Scale-Translate (RST) gestures. For instance, 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.

Figure 6.3 and Figure 6.4 show some rendering modes used to visualize the computed shape and material information from the two panel parts of the *Retablo S. Bernardino*. We show the computed SV-BRDF relighted with two different kind of light sources (i.e., directional and spot light); we also visualize some shape and material based layers and rendering modes, such as the normal, albedo and specular map, and the monochromatic rendering. Figure 6.5 shows how we can exploit the multi-layered nature of the computed data in order to inspect shape and material characteristics of the surface (e.g., normal map and surface roughness).

We perform a multi-layer visualization through a lens tool in a focus-and-context setup. The context consists in the rendering of the computed SV-BRDF by using spot (left) or directional (right) light, while within the lens we render a different layer or use a different rendering modes. Finally, Figure 6.6 and Figure 6.7 illustrate the usage of the system on a large multitouch surface.

Figure 6.6: **Retablo S. Bernardino - Prophet Daniel - Interaction sequence on large touch screen**. Representative frames from an interactive session on a 98 inch touch screen.

A video demonstration of the entire pilot, from acquisition to exploration, is available on the project website at https://evocation.eu/videos/.

6.3 The Nora Stone

In the second pilot, we applied our acquisition, processing, annotation and visualization pipeline to a the Nora Stone (*Stele di Nora* in Italian), an ancient Phoenician inscribed stone found at Nora on the south coast of Sardinia in 1773. The object is housed in the National Archaeological Museum in Cagliari (Italy). One of the museum's major purposes is to hold and preserve findings from the pre-Nuragic and Nuragic age to the Byzantine age, including bronze statuettes from the Nuragic age, artworks and findings related to the Phoenician settlement (the Nora Stone is one of them), and other objects (e.g., ceramics and jewels) from the Carthaginian, Roman, Italic, and Byzantine culture. The Stone, whose dating varies from 850 to 725 BC, has a trapezoidal shape and is made of sandstone with a rough surface. The current, preserved visible area is 105cm tall. 49-59cm wide, and about 20cm thick. The stone is of large archaeological importance, since many scholars claim that it is the oldest written document in Sardinia and, possibly, in the entire Western Mediterranean region. Moreover, according to widespread interpretations, the stone contains the first attestation of the old name of Sardinia: "Shrdn".

I carried out the acquisition on-site, together with my colleagues Ruggero Pintus and Antonio Zorcolo. The acquisition has been performed without moving the Nora Stone from the exhibition area. We darkened the capture area using dark cloths. We per-

Figure 6.7: **Retablo S. Bernardino - Piety of Christ - Interaction sequence on large touch screen**. Representative frames from an interactive session on a 98 inch touch screen.

formed a free-form MLIC acquisition, with a fixed 36.3 Mpixels DSLR FX Nikon D810 Camera with a $50mm$ AF Nikkor Lens and a moved handheld white LED (5500K) covering the entire visible spectrum (Figure 6.8 left). We gathered more than 60 images for the MLIC. The acquired data has been calibrated with four glossy spheres (for light direction), and with a white frame positioned around the object (see Figure 6.8 right), using the camera and light calibration method recently presented by Pintus et al. [116]. The extraction of the shape and the optical material behaviour has been performed with the same protocol that we used for the paintings (section 6.2), even though, in this rather diffuse case, a simpler solution could have been applied to just extract the albedo. The surface properties come directly from the sandstone material, and the only added pigments are remains of old restoration attempts that tried to emphasize the shape of the carved letters.

With the help of scholars in the humanities, we produced a series of layers of interest from the captured data and documented the model using an annotation graph. As for layers, from the original data we produced the original albedo, a version of the albedo map without the color restoration intervention, and a version in which the inscription is digitally enhanced. We also produced monochrome maps to be used to emphasize the geometry. Figure 6.9 show several relighting results. First, we show the unmodified captured shape and material relighted with a directional light source. Then, we show a monochromatic rendering, together with the normal map and the surface albedo. Finally, we present relightings of the two edited layers with removed or enhanced coloring of the inscriptions. Figure 6.10 (left) shows two layers inspected at the same time with a visualization lens.

The model is also enriched with structured annotations exploited to guide the user in a journey through the model. For this use case, we used the simplified model presented

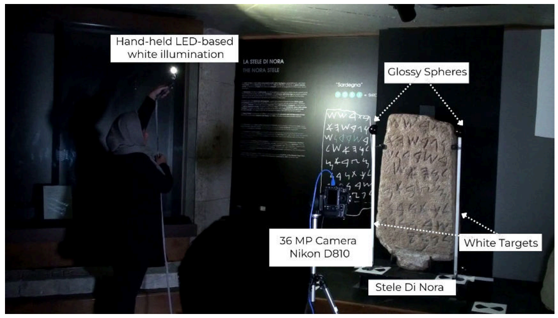


Figure 6.8: **Nora Stone acquisition**. We performed an on-site free-form MLIC acquisition, with a fixed 36.3 Mpixels DSLR FX Nikon D810 Camera with a $50mm$ AF Nikkor Lens and a moved handheld white LED (5500K) covering the entire visible spectrum with more than 60 images. The acquired data has been calibrated with four glossy spheres (for light direction), and with a white frame positioned around the object

Figure 6.9: **Stele di Nora - Multi-layered Representation**. From left to right, top to bottom: relighting of the computed model by using a directional light; monochromatic rendering; normal map layer; diffuse/albedo map; two layers created by editing the original layers, i.e., a standard rendering with a restored original rock without the letter colors, and a map with highlighted letters.

in chapter 5, that arranges annotations in a multi-level hierarchy. As for the follow-up work on Nivola [1], we have used the *openlime* framework, and with the support of my colleague Fabio Bettio, I have enriched the user interface to display a text box with visual and text information associated to each annotation in the area near the lens, and added buttons to navigate in the multi-resolution hierarchy. This approach offers the possibility to follow a multiresolution linear flow with details on demand. The interface of the openlime viewer and editor is depicted in Figure 6.11, while the annotation workflow is presented in Figure 6.12.

We simplify user interaction by using a single virtual object (the visualization lens) as the sole target for user manipulation. By moving or scaling the lens, the system jointly controls both a focus area and the camera of the surrounding view. Moreover, the lens has an attached dashboard to trigger all interactively controlled actions, in particular

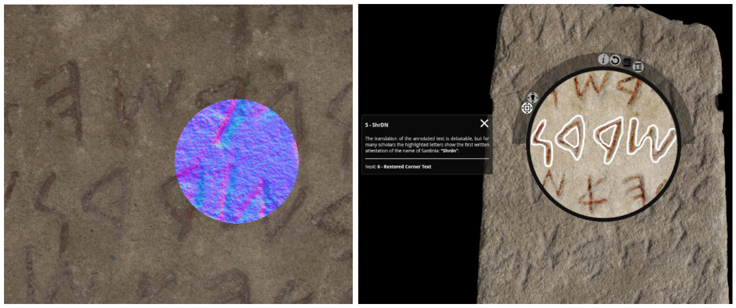


Figure 6.10: **Nora Stone - Lens-based visualization of the Annotated Multi-layered model** . Left: multi-layered visualization with standard rendering in the context vs normal map inside the lens. Right: decorated lens with overlay annotation and its description in a side box.

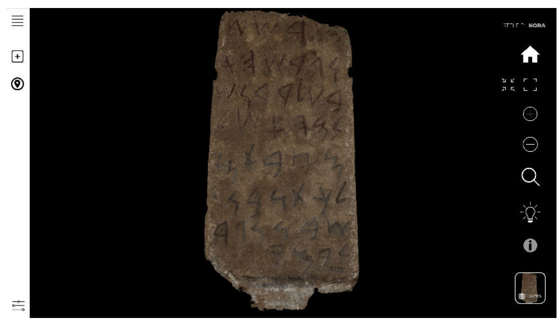


Figure 6.11: **OpenLime viewer interface.** The basic viewer is also used for editing annotations by placing the lens in the desired position, setting the viewing condition, and recording the annotation in a database.

for the navigation through annotations, or for switching from moving the camera or the light for navigation or relighting. Each annotation is shown in overlay with its corresponding text and/or image description. We always show the annotation overlay inside the lens, while the text/image description is visible only when the lens is steady and inactive; it is presented in a box placed near, but outside, the lens. Figure 6.10 (right) depicts a decorated lens showing the overlay annotation and enhanced layer in the focus area, and the base layer in the context area.

This strong focus+context design, that simplifies user interaction by using a single virtual object (a visualization lens) as the sole target for user manipulation, also enables the effective support of large touch screens, where users close to the screen for manipulation purposes naturally focus on a small moving display area, using the rest of the display as an immersive context in the visual periphery.

The overall interactive experience is also enriched by an automatic playback mode, which sequentially loops over the annotations one after the other. Figure 6.13) illustrates

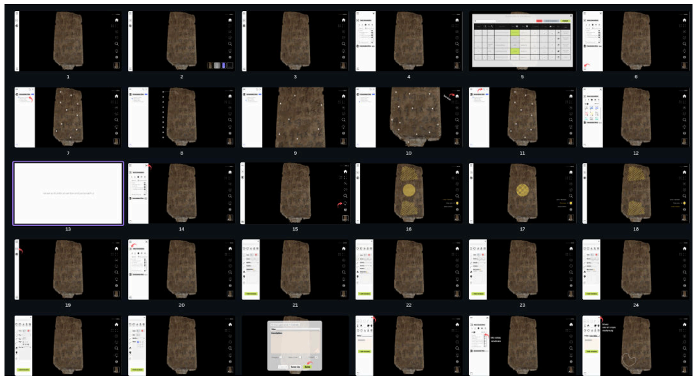


Figure 6.12: **OpenLime viewer interface.** Screenshots from the openlime annotation workflow.

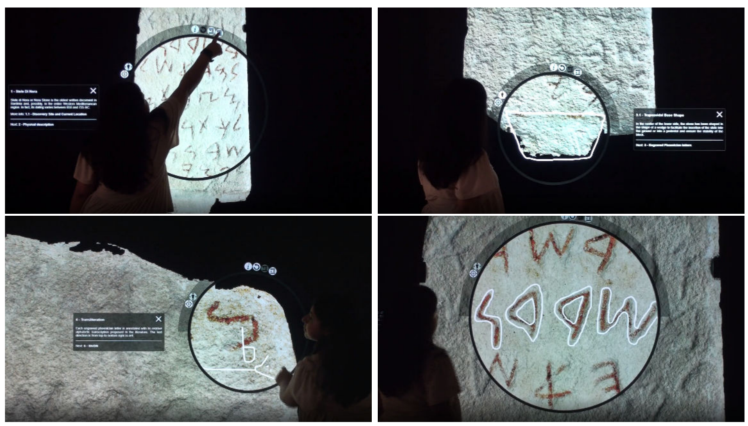


Figure 6.13: **Nora Stone - Interaction sequence on large touch screen**. Representative frames from an interactive session on a 98 inch touch screen.

an interaction sequence on a 98 inch touch display through four frames extracted from a live recording. A video demonstration of the entire pilot, from acquisition to exploration, is available on the project website at `https://evocation.eu/videos/`.

6.4 Bibliographic notes

Major part of this chapter has been taken from my contribution to the EVOCATION Deliverable D6.1 titled as ”*Pilot: Capture and Replication for Cultural Heritage*” [32].

Chapter 7

Conclusion

This thesis has introduced novel techniques based on visualization lenses and guidance to advance the state-of-the-art in interactive and scalable exploration beyond plain visual replication. This final chapter provides a concise summary of the achieved results and briefly discusses the potential directions for future work.

7.1 Overview of achievements

In this thesis, I tackled the problem of improving the exploration of objects associated with additional data that provides insights on these objects. In my analysis of related work (chapter 2), I have shown how that a very wide variety of use cases and application domains require both the ability to combine 2D or 3D objects with annotations linking regions of them to additional information, and to explore this resulting annotated objects for a variety of needs.

I studied these problems in the cultural-heritage-computing domain, focusing on the very common and important special case of mostly planar, but visually, geometrically, and semantically rich objects. These could be generally flat objects (e.g., paintings, bas-reliefs), as well as visualizations of fully 3D objects from a particular point of views (e.g., frontal or side views of buildings or statues). Selecting a precise application domain and a specific presentation mode allowed me to concentrate on the well defined use-case of the exploration of annotated relightable stratigraphic models for local and remote museum presentation.

The analysis of the state-of-the-art at the beginning of my PhD work (chapter 2) showed that, despite the fact that many solutions have been proposed for the creation, organization, and display of annotations, several important open problems remained, including:

- How to ***define relations*** not only between objects and annotations, but also among annotations themselves in order to guide their presentation;
- How to ***define recommendation systems*** that exploit this graph representation to determine which annotation to display based on the current interaction context;
- How to ***better display annotations*** to avoid clutter;
- How to ***guide the user*** within an annotated world with a suitable interface.

During the course of my thesis, with the invaluable and generous help of my mentors and colleagues Enrico Gobbetti *(Supervisor)*, Fabio Bettio, Fabio, Marton, and Ruggero Pintus, I strived to consistently solve these problems through novel solutions building on the concepts of visualization lenses [26] and guidance [27]. In particular, my main achievements have been:

- **A novel technique for interactively controlling visualization lenses while automatically maintaining a good focus-and-context visualization** (chapter 3).
- **A novel approach for avoiding clutter in the visualiation of an annotated model and for guiding users towards interesting areas** (chapter 4).
- **A method for structuring audio-visual object annotations into a graph and for using the graph to improve guidance and support automated tours** (chapter 5).

7.2 Characterization of proposed solutions

The introduced solutions offer practical methods for clutter-free exploration of annotated models. The first presented approach (chapter 3) consists in a general enhanced interaction controller that helps interactive exploration of a model with a lens by providing a mapping, mediated by an interaction metaphor, that meaningfully links user actions on the inside or outside of lens to coordinated camera and lens motions that support focus-and-context exploration. Our performance evaluation has shown that the method is intuitive, well received by users, and efficient for multi-scale exploration.

While the approach has been presented for exploration of annotated relightable stratigraphic models, the solution is general enough to be readily applied to other information visualization using lenses on a variety of 2D datasets. We expect, in particular, that it could replace standard camera and lens controllers in typical pan-and-zoom 2D interfaces.

The other two solutions build on this controller to provide an enhanced exploration of annotated models. The first approach expands the usual flat database of annotations attached to objects by associating to each annotation the best lens configuration for viewing it (chapter 4). This simple extension allows to exploit the database by looking for the next-"best" lens in the database based on contextual information, guiding users in the self-paced exploration of annotated 2D models. The presented results demonstrate how this technique leads to a new way of mixing casual interaction with storytelling from data. One important result in this area is that our approach of selecting the next best annotation to display and of differentiating between immediately displayable annotations and possible future annotation to display makes it possible to support a variety of use cases, including direct display of one annotation at a time without clutter, suggestions for directing users towards interesting areas, as well as guided tours enriched with story-telling aspects.

Since flat databases do not permit to define relations among annotations, for example for prescribing whether some information has to be presented before some other during storytelling, we have further enhanced this approach by structuring the annotation database in a graph (chapter 5) in which each graph node contains an annotation, in the form of a visual and audio markup of the area of interest, as well as the optimal lens parameters that should be used to explore the annotated area and a scalar representing the annotation importance, while directed graph edges are used, instead, to represent preferred ordering relations in the presentation of annotations, by having each node point to the set of nodes that should be seen before presenting its associated annotation. We have shown how this graph can be exploited by a recommendation system that strives to preserve precedence relations while presenting context-dependent annotations, using a presentation order which is dynamically dependent on lens position, navigation history and authoring information encoded in an annotation graph. The introduction of stochastic path selection that interprets context-dependent scores as transition probabilities makes it possible to increase the variability of exploration paths. Moreover, the user can freely mix personal/free exploration with automatic touring.

7.3 Future directions

Our very preliminary evaluation has shown the potential interest of the approach, but also highlighted important areas for future research.

Lens Database & Exploration Path Both the flat database solution and the graph approach are targeted towards the exploration of areas that fit well on a circular lens. They should, however, be refined when pointing at areas where linear or extended features should be explored. We plan to address this problem by storing at each node not only a single lens position, but a lens path for the exploration of the annotated area. This would require, however, also the redefinition of parts of the scoring system, to keep into account the distance of the current lens to the stored lens path, rather to the stored lens with the same circular shape.

Extension to dependencies and annotated graph edges For the graph solution, dependencies only express simple precedence relations from one node to a set of predecessors. It would be interesting to explore whether the concept of dependencies could be expanded into a full logic language to improve storytelling features. Moreover, all the annotation information is stored in the nodes, while the edges only express relations. It would be interesting, especially for automatic tours, to evaluate how to include annotations in the edges, for example to display information during the transition from one node to the next.

Extensions to 3D Datasets Our methods have been applied to 2D examples, but, in principle, they could be extended to 3D. An important avenue of future work is therefore the extension more general 3D visualization. For focus-and-context control, several solutions have been proposed to control lenses in screen-space (e.g., [112]–[114] or object-space (e.g., [75], [115]), but none of these techniques seamlessly supports navigation on multiple models with coupled lens and camera control. We have identified decal lenses [75], which act on patches of 2D manifolds built to attach smoothly to non-flat surfaces, as a one of the most promising extensions, since we could extend our approach by sliding and scaling decals around the surface while maintaining enough context visible. For guidance, the extension to 3D is also very interesting, but requires considerable research. In particular, the various terms used for determining the next best lens would need to be adapted to 3D, in particular taking into account 3D visibility. A starting point could be the work done by Balsa et al. [97] for camera navigation.

Usability Evaluation While our techniques have been evaluated with users, our current evaluation was very preliminary, and focused mostly on extracting basic performance measures and getting data on user satisfaction, more work is required to objectively assess the effectiveness of our user interface for specific tasks different than casual inspection. Addressing this would require cognitive measures that are beyond the scope of the thesis, and are an important avenue for future work.

Accessibility Features / UX For future applications, an overlooked area is the accessibility aspect of the use-cases, particularly considering users with special needs and mobility disabilities; user experience (UX) design considerations for various physical height ranges e.g., for short-heighted users like children (in case of tall-display-screens restriction) and tall users (in case of accessing the features located in the lower parts of the display) would be interesting.

Annotations Authoring Finally, authoring, orthogonal to this work, also definitely deserves attention. We currently just extended the viewer, that already allows users to move the lens to interesting areas and to modify rendering parameters, in order to draw overlay annotations with a simple image editor. We then store them in an annotation database containing the lens and context area description, as well as the rendering parameters. For the method presented in chapter 5, the node table is then edited offline by adding dependencies to nodes, and enriching the description of each annotation with an audio recording. We have later provides support, within the *OpenLime* editor, for direct editing of dependencies within the viewer [1], but the methods employed, while usable by end users in the cultural heritage domain, as demonstrated by our examples, are currently still not immediate to use, as they require, for instance, to search for target annotation labels by name in the annotation database in order to record dependencies. Such an approach puts a lot of burden in the user, and would become even more cumbersome in case of extension of the dependency logic. Several works have recently focused on the problem of (semantic) annotation and could serve as a basis for improving annotation creation [37], [39], [43], [50], [79], extending them, in particular, for annotation linking.

7.4 Publications

The scientific results obtained during this PhD work also appeared in related publications, for which I significantly contributed to the conceptualization, methodology, and validation of the developed method. These main publication, sorted by their introduction in this thesis, are the following:

- **A novel approach for exploring annotated data with interactive lenses.**
  Fabio Bettio, Moonisa Ahsan, Fabio Marton, and Enrico Gobbetti, Computer Graphics Forum, 40(3): 387-398, 2021. Proc. EUROVIS 2021.
  DOI: 10.1111/cgf.14315 — View PDF.
  — This is the original work forming the basis of my core concepts on interactively controlling visualization with focus-and-context (chapter 3) and avoiding clutter in an annotated model while interactively guiding users towards interesting areas in the object of interest (chapter 5).

- **Guiding Lens-based Exploration using Annotation Graphs.**
  Moonisa Ahsan, Fabio Marton, Ruggero Pintus, and Enrico Gobbetti. In Proc. Smart Tools and Applications in Graphics (STAG). Pages 85-90, October 2021.
  DOI: 10.2312/stag.20211477 — View PDF.
  **Honorable mention in best paper award category at STAG 2021.*
  — This work originally introduced the concept of graphs for structuring the annotation database in order to express preferences for lens-based navigation.

- **Audio-visual annotation graphs for guiding lens-based scene exploration.**
  Moonisa Ahsan, Fabio Marton, Ruggero Pintus, Enrico Gobbetti, In the Special Section on STAG21, Computers & Graphics, May 2022.
  DOI: j.cag.2022.05.003 — View PDF.
  — This article is a significant extension of our STAG2021 contribution that introduced a new representation of graph dependencies, that makes it possible to express hierarchical grouping and levels of abstraction, an improved scoring system that best preserves the ordering relations by exploiting topological distance in the annotation graph, an improved state machine for intuitive transition between interactive control and auto-touring features, and the seamless handling of audio markups (chapter 5).

In addition, during the course of my thesis, I have also contributed to the following related publications, which have not been included in this work:

- **Web-based Exploration of Annotated Multi-Layered Relightable Image Models**
  Alberto Jaspe Villanueva, Moonisa Ahsan, Ruggero Pintus, Andrea Giachetti, and Enrico Gobbetti, ACM Journal on Computing and Cultural Heritage, 14(2): 24:1-24:31, May 2021.
  DOI: 10.1145/3430846. — View PDF.
  — The work on this paper was the preliminary introduction to my research domain for this thesis. I contributed to the paper on the generation of data and evaluation.

- **Exploiting Neighboring Pixels Similarity for Effective SV-BRDF Reconstruction from Sparse MLICs**
  Ruggero Pintus, Moonisa Ahsan, Fabio Marton, and Enrico Gobbetti, In The 19th Eurographics Workshop on Graphics and Cultural Heritage, November 2021.
  DOI: 10.2312/gch.20211412. — View PDF.
  **Best paper award at GCH 2021.*
  — The work describes an efficient approach for creating relightable models from calibrated multi-light collections. Here, I mainly supported the capture of data and support in conducting tests. The technique has been used for generating all the input datasets that are then explored with the techniques presented in this thesis.

- **Ebb & Flow: Uncovering Costantino Nivola's Olivetti Sandcast through 3D Fabrication and Virtual Exploration.**
  Moonisa Ahsan, Giuliana Altea, Fabio Bettio, Marco Callieri, Antonella Camarda, Paolo Cignoni, Enrico Gobbetti, Paolo Ledda, Alessandro Lutzu, Fabio Marton, Giuseppe Mignemi, and Federico Ponchio, In The 20th Eurographics Workshop on Graphics and Cultural Heritage, November 2022.
  DOI: 10.2312/gch.20221230 — View PDF.
  — This paper is simplified refinement of [29] that has been applied to the exploration of a very large annotated artwork (Nivola's Olivetti Sandcast). In this use case, the graph is restricted to a tree. I contributed with the integration and tuning of the presented techniques within the *openlime* framework, under development at CRS4 and ISTI-CNR.

7.5 Demonstration videos

In the context of the EVOCATION project, I have also illustrated the outcomes of my research in the following demonstration videos that are available on the project web site at the URL evocation.eu/videos/:

- **Scalable exploration of complex objects and environments beyond plain visual replication: Joint Camera and Lens Control for Focus-and-Context Exploration** — Demo video.
  Demonstration of the technique presented in chapter 3

- **Scalable exploration of complex objects and environments beyond plain visual replication: Assisted and automatic navigation in an annotated model** — Demo video.
  Demonstration of the technique presented in chapter 4

- **Scalable exploration of complex objects and environments beyond plain visual replication: Pilot: Acquisition, reconstruction, and exploration of paintings from retable of San Bernardino.** — Demo video.
  Demonstration of the painting pilot discussed in chapter 6

- **Scalable exploration of complex objects and environments beyond plain visual replication: Pilot: Acquisition, reconstruction, and exploration of Stele di Nora.**
  — Demo video.
  Demonstration of the Nora Stone pilot discussed in chapter 6

Bibliography

[1] M. Ahsan, G. Altea, F. Bettio, M. Callieri, A. Camarda, P. Cignoni, E. Gobbetti, P. Ledda, A. Lutzu, F. Marton, G. Mignemi, and F. Ponchio, "Ebb & flow: Uncovering Costantino Nivola's Olivetti sandcast through 3D fabrication and virtual exploration," in *The 20th Eurographics Workshop on Graphics and Cultural Heritage*, Sep. 2022, pp. 85–94.

[2] M. Economou and E. Meintani, "Promising beginnings? evaluating museum mobile phone apps," in *Proc. Rethinking Technology in Museums Conference*, 2011, pp. 26–27.

[3] T. Kuflik, A. Wecker, J. Lanir, and O. Stock, "An integrative framework for extending the boundaries of the museum visit experience: Linking the pre, during and post visit phases," *Information Technology & Tourism*, vol. 15, pp. 17–047, 2014. doi: 10.1007/s40558-014-0018-4.

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

[5] T. Kuflik, O. Stock, M. Zancanaro, A. Gorfinkel, S. Jbara, S. Kats, J. Sheidin, and N. Kashtan, "A visitor's guide in an active museum: Presentations, communications, and reflection," *JOCCH*, vol. 3, no. 3, 11:1–11:25, 2011.

[6] M. Potenziani, M. Callieri, M. Dellepiane, M. Corsini, F. Ponchio, and R. Scopigno, "3DHOP: 3D heritage online presenter," *Computers & Graphics*, vol. 52, pp. 129–141, 2015.

[7] Sketchfab, *Sketchfab: Publish, share and discover 3d content over the web*, [Online; accessed 28-Sep-2022], 2022. [Online]. Available: http://www.sketchfab.com/.

[8] CHI, *Cultural heritage imaging website*, [Online; accessed-March-2019], 2019. [Online]. Available: http://culturalheritageimaging.org.

[9] R. Pintus, T. Dulache, I. Ciortan, E. Gobbetti, and A. Giachetti, "State-of-the-art in multi-light image collections for surface visualization and analysis," *Computer Graphics Forum*, vol. 38, no. 3, pp. 909–934, 2019. doi: 10.1111/cgf.13732.

[10] R. Fattal, M. Agrawala, and S. Rusinkiewicz, "Multiscale shape and detail enhancement from multi-light image collections," *ACM TOG*, vol. 26, no. 3, 51:1–51:9, 2007.

[11] R. J. Woodham, “Photometric method for determining surface orientation from multiple images,” *Optical engineering*, vol. 19, no. 1, pp. 513–531, 1980.

[12] J. Ackermann and M. Goesele, “A survey of photometric stereo techniques,” *Foundations and Trends in Computer Graphics and Vision*, vol. 9, no. 3-4, pp. 149–254, 2015.

[13] M. Weinmann and R. Klein, “Advances in geometry and reflectance acquisition (course notes),” in *SIGGRAPH Asia 2015 Courses*, ACM, 2015.

[14] J. Dorsey, H. Rushmeier, and F. Sillion, *Digital modeling of material appearance*. Elsevier, 2010.

[15] T. Malzbender, D. Gelb, and H. Wolters, “Polynomial texture maps,” in *Proc. SIGGRAPH*, 2001, pp. 519–528.

[16] J. Jankowski and M. Hachet, “A survey of interaction techniques for interactive 3D environments,” in *Eurographics STAR*, 2013.

[17] P. Artal-Isbrand and P. Klausmeyer, “Evaluation of the relief line and the contour line on greek red-figure vases using reflectance transformation imaging and three-dimensional laser scanning confocal microscopy,” *Studies in Conservation*, vol. 58, no. 4, pp. 338–359, 2013.

[18] A. Moutafidou, G. Adamopoulos, A. Drosou, D. Tzovaras, and I. Fudos, “Multiple material layer visualization for cultural heritage artifacts,” in *Proc. GCH*, 2018, pp. 155–159. doi: 10.2312/gch.20181353.

[19] K. Ponto, M. Seracini, and F. Kuester, “Wipe-Off: An intuitive interface for exploring ultra-large multi-spectral data sets for cultural heritage diagnostics,” *Computer Graphics Forum*, vol. 28, no. 8, pp. 2291–2301, 2009. doi: 10.1111/j.1467-8659.2009.01532.x.

[20] B. Vandermeulen, H. Hameeuw, L. Watteeuw, L. Van Gool, and M. Proesmans, “Bridging multi-light & multi-spectral images to study, preserve and disseminate archival documents,” *Archiving Conference*, vol. 2018, no. 1, pp. 64–69, 2018. doi: 10.2352/issn.2168-3204.2018.1.0.15.

[21] L. W. Macdonald, “Realistic visualisation of cultural heritage objects,” Ph.D. dissertation, UCL (University College London), 2015.

[22] F. Bettio, M. Ahsan, F. Marton, and E. Gobbetti, “A novel approach for exploring annotated data with interactive lenses,” *Computer Graphics Forum*, vol. 40, no. 3, pp. 387–398, 2021. doi: 10.1111/cgf.14315.

[23] P. Vanhulst, F. Evequoz, R. Tuor, and D. Lalanne, “A descriptive attribute-based framework for annotations in data visualization,” in *Proc. International Joint Conference on Computer Vision, Imaging and Computer Graphics*, 2018, pp. 143–166. doi: 10.1007/978-3-030-26756-8_7.

[24] F. Ponchio, M. Callieri, M. Dellepiane, and R. Scopigno, “Effective annotations over 3D models,” *Computer Graphics Forum*, vol. 39, no. 1, pp. 89–105, 2020. doi: 10.1111/cgf.13664.

[25] J. Camba, M. Contero, and M. Johnson, "Management of visual clutter in annotated 3D CAD models: A comparative study," in *Proc. International Conference of Design, User Experience, and Usability*, 2014, pp. 405–416. doi: 10.1007/978-3-319-07626-3_37.

[26] C. Tominski, S. Gladisch, U. Kister, R. Dachselt, and H. Schumann, "Interactive lenses for visualization: An extended survey," *Computer Graphics Forum*, vol. 36, no. 6, pp. 173–200, 2017. doi: 10.1111/cgf.12871.

[27] D. Ceneda, T. Gschwandtner, and S. Miksch, "A review of guidance approaches in visual data analysis: A multifocal perspective," *Computer Graphics Forum*, vol. 38, no. 3, pp. 861–879, 2019. doi: 10.1111/cgf.13730.

[28] M. Ahsan, F. Marton, R. Pintus, and E. Gobbetti, "Guiding lens-based exploration using annotation graphs," in *Proc. Smart Tools and Applications in Graphics (STAG)*, 2021, pp. 85–90. doi: 10.2312/stag.20211477.

[29] M. Ahsan, F. Marton, R. Pintus, and E. Gobbetti, "Audio-visual annotation graphs for guiding lens-based scene exploration," *Computers & Graphics*, 2022. doi: 10.1016/j.cag.2022.05.003.

[30] A. Celarek, S. Fraiss, M. Ahsan, F. Ganovelli, E. Gobbetti, F. Marton, G. Pintore, J. Cardoso, S. Eilemann, L. Calla, R. Pajarola, M. Ritz, and M. Domajnko, "State-of-the-art report on interactive visualization," EVOCATION Project, H2020 MSCA 813170, Deliverable D3.1, Apr. 2020.

[31] M. Ahsan, E. Almansa Aranega, Fabio Bettio, L. Romero Calla, A. Celarek, J. Cardoso, E. Gobbetti, A. Islam, F. Marton, B. Mohanto, R. Pintus, O. Staadt, M. Wimmer, and A. Zorcolo, "Interim report on the development of new interactive visualization techniques," EVOCATION Project, H2020 MSCA 813170, Deliverable D3.2, Sep. 2022.

[32] M. Ahsan *et al.*, "Pilot: Capture and replication for cultural heritage," EVOCATION Project, H2020 MSCA 813170, Deliverable D6.1, May 2023, To appear.

[33] A. Jaspe Villanueva, M. Ahsan, R. Pintus, A. Giachetti, and E. Gobbetti, "Web-based exploration of annotated multi-layered relightable image models," *ACM JOCCH*, 2021, To appear.

[34] D. A. Bowman, C. North, J. Chen, N. F. Polys, P. S. Pyla, and U. Yilmaz, "Information-rich virtual environments: Theory, tools, and research agenda," in *Proc. ACM VRST*, 2003, pp. 81–90. doi: 10.1145/1008653.1008669.

[35] T. Götzelmann, P.-P. Vázquez, K. Hartmann, A. Nürnberger, and T. Strothotte, "Correlating text and images: Concept and evaluation," in *Proc. Smart Graphics*, Berlin, Heidelberg, 2007, pp. 97–109.

[36] M. Callieri, C. Leoni, M. Dellepiane, and R. Scopigno, "Artworks narrating a story: A modular framework for the integrated presentation of three-dimensional and textual contents," in *Proc. ACM Web3D*, 2013, pp. 167–175.

[37] L. De Luca, M. Pierrot-Deseilligny, A. Manuel, C. Chevrier, P. B. Benjamin Lollier, A. Pamart, F. Peteler, V. Abergel, and A. Alaoui, *Aioli – a reality-based 3D annotation platform for the collaborative documentation of heritage artefacts*, [Online; accessed 03-Mar-2021], 2019. [Online]. Available: http://www.aioli.cloud/.

[38] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, *et al.*, "Shapenet: An information-rich 3D model repository," *arXiv preprint arXiv:1512.03012*, 2015.

[39] M. Attene, F. Robbiano, M. Spagnuolo, and B. Falcidieno, "Semantic annotation of 3D surface meshes based on feature characterization," in *International Conference on Semantic and Digital Media Technologies*, 2007, pp. 126–139.

[40] C.-H. Yu and J. Hunter, "Documenting and sharing comparative analyses of 3D digital museum artifacts through semantic web annotations," *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 6, no. 4, pp. 1–20, 2013.

[41] M. Auer, N. Billen, L. Loos, A. Zipf, and G. Agugiaro, "Web-based visualization and query of semantically segmented multiresolution 3D models in the field of cultural heritage," *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 2, no. 5, 2014.

[42] F. I. Apollonio, V. Basilissi, M. Callieri, M. Dellepiane, M. Gaiani, F. Ponchio, F. Rizzo, A. R. Rubino, R. Scopigno, *et al.*, "A 3D-centered information system for the documentation of a complex restoration intervention," *Journal of Cultural Heritage*, vol. 29, pp. 89–99, 2018. doi: 10.1016/j.culher.2017.07.010.

[43] I. Banerjee, C. E. Catalano, G. Patané, and M. Spagnuolo, "Semantic annotation of 3D anatomical models to support diagnosis and follow-up analysis of musculoskeletal pathologies," *International journal of computer assisted radiology and surgery*, vol. 11, no. 5, pp. 707–720, 2016.

[44] R. Scopigno, M. Callieri, P. Cignoni, M. Corsini, M. Dellepiane, F. Ponchio, and G. Ranzuglia, "3d models for cultural heritage: Beyond plain visualization," *Computer*, no. 7, pp. 48–55, 2011.

[45] S. P. Serna, H. Schmedt, M. Ritz, and A. Stork, "Interactive semantic enrichment of 3d cultural heritage collections," in *Proc. VAST*, 2012, pp. 33–40.

[46] J. Hunter and A. Gerber, "Harvesting community annotations on 3D models of museum artefacts to enhance knowledge, discovery and re-use," *Journal of Cultural Heritage*, vol. 11, no. 1, pp. 81–90, 2010.

[47] M. Balsa Rodriguez, M. Agus, F. Bettio, F. Marton, and E. Gobbetti, "Digital Mont'e Prama: Exploring large collections of detailed 3D models of sculptures," *ACM Journal on Computing and Cultural Heritage*, vol. 9, no. 4, 18:1–18:23, Sep. 2016. doi: 10.1145/2915919.

[48] A. Jaspe Villanueva, R. Pintus, A. Giachetti, and E. Gobbetti, "Web-based multi-layered exploration of annotated image-based shape and material models," in *The 16th Eurographics Workshop on Graphics and Cultural Heritage*, 2019.

[49] C. Bolchini, C. A. Curino, E. Quintarelli, F. A. Schreiber, and L. Tanca, "A data-oriented survey of context models," *SIGMOD Rec.*, vol. 36, no. 4, pp. 19–26, 2007, issn: 0163-5808. doi: 10.1145/1361348.1361353. [Online]. Available: http://doi.acm.org/10.1145/1361348.1361353.

[50] S. P. Serna, R. Scopigno, M. Doerr, M. Theodoridou, C. Georgis, F. Ponchio, and A. Stork, "3d-centered media linking and semantic enrichment through integrated searching, browsing, viewing and annotating.," in *VAST*, 2011, pp. 89–96.

[51] A. Manuel, L. De Luca, and P. Véron, "A hybrid approach for the semantic annotation of spatially oriented images," *International Journal of Heritage in the Digital Era*, vol. 3, no. 2, pp. 305–320, 2014.

[52] T. Messaoudi, P. Véron, G. Halin, and L. De Luca, "An ontological model for the reality-based 3d annotation of heritage building conservation state," *Journal of Cultural Heritage*, vol. 29, pp. 100–112, 2018.

[53] F. Soler, J. C. Torres, A. J. León, and M. V. Luzón, "Design of cultural heritage information systems based on information layers," *Journal on Computing and Cultural Heritage (JOCCH)*, vol. 6, no. 4, pp. 1–17, 2013.

[54] K. Rodriguez-Echavarria, D. Morris, and D. Arnold, "Web based presentation of semantically tagged 3D content for public sculptures and monuments in the UK," in *Proceedings of the 14th International Conference on 3D Web Technology*, 2009, pp. 119–126.

[55] W. Shi, E. Kotoula, K. G. Akoglu, Y. Yang, and H. E. Rushmeier, "CHER-Ob: A tool for shared analysis in cultural heritage.," in *Proc. GCH*, 2016, pp. 187–190.

[56] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Proc. IEEE symposium on visual languages*, 1996, pp. 336–343.

[57] S. Liu, W. Cui, Y. Wu, and M. Liu, "A survey on information visualization: recent advances and challenges," *The Visual Computer*, 2014, issn: 0178-2789. doi: 10.1007/s00371-013-0892-3. [Online]. Available: http://link.springer.com/10.1007/s00371-013-0892-3.

[58] M. Hegarty, "The cognitive science of visual-spatial displays: Implications for design," *Topics in Cognitive Science*, vol. 3, no. 3, pp. 446–474, 2011, issn: 1756-8765.

[59] P. Faraday and A. Sutcliffe, "Designing effective multimedia presentations," in *Proc. ACM SIGCHI*, 1997, pp. 272–278.

[60] H. Sonnet, S. Carpendale, and T. Strothotte, "Integration of 3D data and text: The effects of text positioning, connectivity, and visual hints on comprehension," in *Proc. Interact*, ser. LNCS, vol. 3585, 2005, pp. 615–628.

[61] J. Jankowski, K. Samp, I. Irzynska, M. Jozwowicz, and S. Decker, "Integrating text with video and 3D graphics: The effects of text drawing styles on text readability," in *Proc. ACM SIGCHI*, 2010, pp. 1321–1330.

[62] N. F. Polys, D. A. Bowman, and C. North, "The role of depth and gestalt cues in information-rich virtual environments," *International Journal of Human-Computer Studies*, vol. 69, no. 1-2, pp. 30–51, 2011.

[63] J. Jankowski and S. Decker, "A dual-mode user interface for accessing 3D content on the world wide web," in *Proc. WWW*, 2012, pp. 1047–1056.

[64] T. Isenberg and M. Hancock, "Gestures vs. postures: Gestural touch interaction in 3D environments," in *Proc. 3DCHI*, 2012, pp. 53–61.

[65] A. Cockburn, A. Karlson, and B. B. Bederson, "A review of overview+detail, zooming, and focus+ context interfaces," *ACM Computing Surveys (CSUR)*, vol. 41, no. 1, pp. 1–31, 2009. doi: 10.1145/1456650.1456652.

[66] O. Chapuis, A. Bezerianos, and S. Frantzeskakis, "Smarties: An input system for wall display development," in *Proc. SIGCHI*, 2014, pp. 2763–2772. doi: 10.1145/2556288.2556956.

[67] U. Kister, P. Reipschläger, and R. Dachselt, "Multi-touch manipulation of magic lenses for information visualization," in *Proc. International Conference on Interactive Tabletops and Surfaces*, 2014, pp. 431–434. doi: 10.1145/2669485.2669528.

[68] R. Sadana and J. Stasko, "Designing and implementing an interactive scatterplot visualization for a tablet computer," in *Proc.International Working Conference on Advanced Visual Interfaces*, 2014, pp. 265–272. doi: 10.1145/2598153.2598163.

[69] C. Appert, O. Chapuis, and E. Pietriga, "High-precision magnification lenses," in *Proc. SIGCHI*, 2010, pp. 273–282. doi: 10.1145/1753326.1753366.

[70] W. Javed, S. Ghani, and N. Elmqvist, "Polyzoom: Multiscale and multifocus exploration in 2D visual spaces," in *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 287–296. doi: 10.1145/2207676.2207716.

[71] A. Martinet, G. Casiez, and L. Grisoni, "3D positioning techniques for multi-touch displays," in *Proc. ACM VRST*, 2009, pp. 227–228. doi: 10.1145/1643928.1643978.

[72] J. Brooke, "SUS: A quick and dirty usability scale," in *Usability Evaluation In Industry*, P. Jordan, B. Thomas, I. McClelland, and B. Weerdmeester, Eds., CRC Press, 1996, ch. 21, pp. 189–195. doi: 10.1201/9781498710411-35.

[73] J. R. Lewis and J. Sauro, "The factor structure of the system usability scale," in *Proc. International conference on human centered design*, 2009, pp. 94–103. doi: 10.1007/978-3-642-02806-9_12.

[74] J. Brooke, "SUS: A retrospective," *Journal of usability studies*, vol. 8, no. 2, pp. 29–40, 2013. [Online]. Available: 10.5555/2817912.2817913.

[75] A. Rocha, J. D. Silva, U. R. Alim, S. Carpendale, and M. C. Sousa, "Decal-lenses: Interactive lenses on surfaces for multivariate visualization," *IEEE TVCG*, vol. 25, no. 8, pp. 2568–2582, 2018. doi: 10.1109/TVCG.2018.2850781.

[76] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, and C. Tominski, "Characterizing guidance in visual analytics," *IEEE TVCG*, vol. 23, no. 1, pp. 111–120, 2016. doi: 10.1109/TVCG.2016.2598468.

[77] M. Balsa Rodriguez, E. Gobbetti, F. Marton, and A. Tinti, "Compression-domain seamless multiresolution visualization of gigantic meshes on mobile devices," in *Proc. ACM Web3D International Symposium*, 2013, pp. 99–107. doi: 10.1145/2466533.2466541.

[78] P. Lavrič, C. Bohak, and M. Marolt, "Collaborative view-aligned annotations in web-based 3D medical data visualization," in *Proc. IEEE MIPRO*, 2017, pp. 259–263. doi: 10.23919/MIPRO.2017.7973430.

[79] V. Croce, G. Caroti, L. De Luca, A. Piemonte, and P. Véron, "Semantic annotations on heritage models: 2D/3D approaches and future research challenges," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 43, pp. 829–836, 2020. doi: 10.5194/isprs-archives-XLIII-B2-2020-829-2020.

[80] G. Ellis and A. Dix, "A taxonomy of clutter reduction for information visualisation," *IEEE TVCG*, vol. 13, no. 6, pp. 1216–1223, 2007. doi: 10.1109/TVCG.2007.70535.

[81] G. Ellis, E. Bertini, and A. Dix, "The sampling lens: Making sense of saturated visualisations," in *Proc. CHI extended abstracts on Human Factors in Computing Systems*, 2005, pp. 1351–1354. doi: 10.1145/1056808.1056914.

[82] M. S. Silver, "Decisional guidance for computer-based decision support," *MIS quarterly*, pp. 105–122, 1991. doi: 10.2307/249441.

[83] E. Horvitz, "Principles of mixed-initiative user interfaces," in *Proc. SIGCHI*, 1999, pp. 159–166. doi: 10.1145/302979.303030.

[84] D. Sokolov, D. Plemenos, and K. Tamine, "Methods and data structures for virtual world exploration," *The Visual Computer*, vol. 22, no. 7, pp. 506–516, 2006. doi: 10.1007/s00371-006-0025-3.

[85] T. Götzelmann, P.-P. Vázquez, K. Hartmann, T. Germer, A. Nürnberger, and T. Strothotte, "Mutual text-image queries," in *Proc. Spring Conf. Comput. Graph.*, ACM, 2007, pp. 139–146. doi: 10.1145/2614348.2614368.

[86] M. Di Benedetto, F. Ganovelli, M. Balsa Rodriguez, A. Jaspe Villanueva, R. Scopigno, and E. Gobbetti, "Exploremaps: Efficient construction and ubiquitous exploration of panoramic view graphs of complex 3D environments," *Computer Graphics Forum*, vol. 33, no. 2, pp. 459–468, 2014. doi: 10.1111/cgf.12334.

[87] B. M. Dennis and C. G. Healey, "Assisted navigation for large information spaces," in *Proc. IEEE Visualization*, 2002, pp. 419–426. doi: 10.5555/602099.602165.

[88] Z. Wang, W. Shi, K. Akoglu, E. Kotoula, Y. Yang, and H. Rushmeier, "CHER-Ob: A tool for shared analysis and video dissemination," *J. Comput. Cult. Herit.*, vol. 11, no. 4, Nov. 2018. doi: 10.1145/3230673.

[89] G. W. Furnas, "Generalized fisheye views," *ACM SIGCHI Bulletin*, vol. 17, no. 4, pp. 16–23, 1986. doi: 10.1145/22627.22342.

[90] F. Van Ham and A. Perer, ""search, show context, expand on demand": Supporting large graph exploration with degree-of-interest," *IEEE TVCG*, vol. 15, no. 6, pp. 953–960, 2009. doi: 10.1109/TVCG.2009.108.

[91] S. Gladisch, H. Schumann, and C. Tominski, "Navigation recommendations for exploring hierarchical graphs," in *International Symposium on Visual Computing*, 2013, pp. 36–47. doi: 10.1007/978-3-642-41939-3_4.

[92] F. Bettio, A. Jaspe Villanueva, E. Merella, F. Marton, E. Gobbetti, and R. Pintus, "Mont'e Scan: Effective shape and color digitization of cluttered 3D artworks," *ACM Journal on Computing and Cultural Heritage (JOCCH)*, vol. 8, no. 1, 4:1–4:23, 2015. doi: 10.1145/2644823.

[93] A. Boninu, L. Usai, A. Costanzi Cobau, M. Minoja, and A. Usai, Eds., *Le sculture di Mont'e Prama – Conservazione e restauro – La Mostra – Contesto, scavi e materiali*. Gangemi, 2015, isbn: 9788849229844.

[94] L. Barbieri, F. Bruno, and M. Muzzupappa, "Virtual museum system evaluation through user studies," *Journal of Cultural Heritage*, vol. 26, pp. 101–108, 2017. doi: 10.1016/j.culher.2017.02.005.

[95] U. Hinrichs, H. Schmidt, and S. Carpendale, "EMDialog: Bringing information visualization into the museum," *IEEE TVCG*, vol. 14, no. 6, pp. 1181–1188, 2008. doi: 10.1109/TVCG.2008.127.

[96] R. Zamora-Musa, J. Velez, and H. Paez-Logreira, "Evaluating learnability in a 3D heritage tour," *Presence: Teleoperators and Virtual Environments*, vol. 26, no. 4, pp. 366–377, 2018. doi: 10.1162/PRES_a_00305.

[97] M. Balsa Rodriguez, M. Agus, F. Marton, and E. Gobbetti, "Adaptive recommendations for enhanced non-linear exploration of annotated 3D objects," *Computer Graphics Forum*, vol. 34, no. 3, pp. 41–50, 2015. doi: 10.1111/cgf.12616.

[98] F. Lekschas, M. Behrisch, B. Bach, P. Kerpedjiev, N. Gehlenborg, and H. Pfister, "Pattern-driven navigation in 2D multiscale visualizations with scalable insets," *IEEE TVCG*, vol. 26, no. 1, pp. 611–621, 2019. doi: 10.1109/TVCG.2019.2934555.

[99] E. Segel and J. Heer, "Narrative visualization: Telling stories with data," *IEEE TVCG*, vol. 16, no. 6, pp. 1139–1148, 2010. doi: 10.1109/TVCG.2010.179.

[100] C. G. Healey and B. M. Dennis, "Interest driven navigation in visualization," *IEEE TVCG*, vol. 18, no. 10, pp. 1744–1756, 2012. doi: 10.1109/TVCG.2012.23.

[101] P. M. Aoki, R. E. Grinter, A. Hurst, M. H. Szymanski, J. D. Thornton, and A. Woodruff, "Sotto voce: Exploring the interplay of conversation and mobile audio spaces," in *Proc. SIGCHI*, 2002, pp. 431–438. doi: 10.1145/503376.503454.

[102] R. Hutchinson and A. F. Eardley, "Inclusive museum audio guides: 'guided looking' through audio description enhances memorability of artworks for sighted audiences," *Museum Management and Curatorship*, vol. 36, no. 4, pp. 427–446, 2021.

[103] M. A. Bekos, B. Niedermann, and M. Nöllenburg, "External labeling techniques: A taxonomy and survey," *Computer Graphics Forum*, vol. 38, no. 3, pp. 833–860, 2019.

[104] R. W. Floyd, "Algorithm 97: Shortest path," *Commun. ACM*, vol. 5, no. 6, pp. 345–349, 1962. doi: 10.1145/367766.368168.

[105] J. W. Kristensen, A. Schjørring, A. Mikkelsen, D. A. Johansen, and H. O. Knoche, "Of leaders and directors: A visual model to describe and analyse persistent visual cues directing to single out-of view targets," in *Proc. VRST*, 2021, 88:1–88:3. doi: 10.1145/3489849.3489953.

[106] J. Scholtz, "Beyond usability: Evaluation aspects of visual analytic environments," in *Proc. IEEE VAST*, 2006, pp. 145–150. doi: 10.1109/VAST.2006.261416.

[107] R. Likert, "A technique for the measurement of attitudes," *Archives of psychology*, vol. 22, no. 140, p. 55, 1932.

[108] M. K. Othman, "Measuring visitors' experiences with mobile guide technology in cultural spaces," Ph.D. dissertation, University of York Toronto, Canada, 2012.

[109] S. Hajjar, "Statistical analysis: Internal-consistency reliability and construct validity," *International Journal of Quantitative and Qualitative Research Methods*, vol. 6, no. 1, pp. 46–57, 2018.

[110] J. Olhager and E. Selldin, "Manufacturing planning and control approaches: Market alignment and performance," *International Journal of Production Research*, vol. 45, no. 6, pp. 1469–1484, 2007.

[111] Artmyn, *Artmyn - new generation technological tools and services for the art ecosystem*, [Online; accessed 15-December-2021], 2021. [Online]. Available: https://artmyn.com/.

[112] R. Gasteiger, M. Neugebauer, O. Beuing, and B. Preim, "The FLOWLENS: A focus-and-context visualization approach for exploration of blood flow in cerebral aneurysms," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2183–2192, 2011.

[113] C. Pindat, E. Pietriga, O. Chapuis, and C. Puech, "JellyLens: Content-aware adaptive lenses," in *Proc. UIST*, 2012, pp. 261–270.

[114] S. Kluge, S. Gladisch, U. Freiherr von Lukas, O. Staadt, and C. Tominski, "Virtual lenses as embodied tools for immersive analytics," in *GI VR / AR Workshop*, 2020, pp. 1–12.

[115] R. C. R. Mota, A. Rocha, J. D. Silva, U. Alim, and E. Sharlin, "3De interactive lenses for visualization in virtual environments," in *Proc. IEEE Scientific Visualization Conference (SciVis)*, 2018, pp. 21–25.

[116] R. Pintus, A. Jaspe Villanueva, A. Zorcolo, M. Hadwiger, and E. Gobbetti, "A practical and efficient model for intensity calibration of multi-light image collections," *The Visual Computer*, vol. 37, no. 9, pp. 2755–2767, 2021.

[117] R. Pintus, M. Ahsan, F. Marton, and E. Gobbetti, "Exploiting neighboring pixels similarity for effective SV-BRDF reconstruction from sparse MLICs," in *The 19th Eurographics Workshop on Graphics and Cultural Heritage*, Best paper award at GCH 2021, Nov. 2021. doi: 10.2312/gch.20211412.

Appendix A

Curriculum Vitae

Moonisa Ahsan worked as an Early-stage Researcher (ESR) in the Visual and Data-intensive Computing (ViDiC) Group of CRS4, Italy from 2019-2022, supported by a Marie-Curie Fellowship in the EVOCATION MSCA-ITN (H2020 Research and Innovation Funding Programme grant 81370). Her scientific and research interests are Computer Graphics, Interface Design, Human-Computer-Interaction, User Experience (UX), Cultural Heritage and Digital Illustrations. Within the context of EVOCATION, she is pursuing her PhD from University of Cagliari (UniCa) from 2019-2023. She holds a Bachelors (Hons) Degree in Information Technology (IT) from the University of Gujrat (Pakistan), and a Masters Degree in Information Technology (IT) from the University of Lahore (Pakistan). During her ESR career, she was a part of research works mainly focusing on interactive exploration of annotated models.

Contact Information

Name	**Moonisa Ahsan**
E-Mail	moonisa.ahsan@gmail.com
Contact	+923066158053
Website	www.imoonisa.com
LinkedIn	Linkedin.com/moonisa
Google Scholar	scholar.google.com/moonisa
ORCID	0000-0003-0688-3919

Education

2019 - 2023	**Ph.D.** candidate in Computer Science. / **Marie-Curie** Fellowship Doctoral program in Department of Mathematics and Computer Science. University of Cagliari (UniCa), Italy.
2016 - 2018	**MSIT** - Masters in Information Technology. The University of Lahore (UOL), Pakistan.
2011 - 2015	**BSIT** - Bachelors(Hons) in Information Technology. University of Gujrat (UOG), Pakistan.

Work Experience

2019 - 2022	**Researcher / Marie-Curie Fellow** Visual and Data-Intensive Computing (ViDiC) Group, Center for Advanced Studies, Research and Development in Sardinia (CRS4), Italy.
2015 - 2019	**University Lecturer** — GC Women University Sialkot, Pakistan.
2014 - 2019	**Graphics Designer** — Plagiarism Checker X, LLC.

Scientific Publications

Journal Articles

1. **Audio-visual annotation graphs for guiding lens-based scene exploration.**
   Moonisa Ahsan, Fabio Marton, Ruggero Pintus, Enrico Gobbetti, In the Special Section on STAG21, Computers & Graphics, May 2022.
   DOI: j.cag.2022.05.003 — View PDF.
2. **A novel approach for exploring annotated data with interactive lenses.**
   Fabio Bettio, Moonisa Ahsan, Fabio Marton, and Enrico Gobbetti, Computer Graphics Forum, 40(3): 387-398, 2021. Proc. EUROVIS 2021.
   DOI: 10.1111/cgf.14315 — View PDF.
3. **Web-based Exploration of Annotated Multi-Layered Relightable Image Models**
   Alberto Jaspe Villanueva, Moonisa Ahsan, Ruggero Pintus, Andrea Giachetti, and Enrico Gobbetti,
   ACM Journal on Computing and Cultural Heritage, 14(2): 24:1-24:31, May 2021.
   DOI: 10.1145/3430846. — View PDF.

Conference Papers

1. **Ebb & Flow: Uncovering Costantino Nivola's Olivetti Sandcast through 3D Fabrication and Virtual Exploration.**
   Moonisa Ahsan, Giuliana Altea, Fabio Bettio, Marco Callieri, Antonella Camarda, Paolo Cignoni, Enrico Gobbetti, Paolo Ledda, Alessandro Lutzu, Fabio Marton, Giuseppe Mignemi, and Federico Ponchio, In The 20th Eurographics Workshop on Graphics and Cultural Heritage, November 2022.
   DOI: 10.2312/gch.20221230 — View PDF.
2. **Guiding Lens-based Exploration using Annotation Graphs.**
   Moonisa Ahsan, Fabio Marton, Ruggero Pintus, and Enrico Gobbetti, In Proc. Smart Tools and Applications in Graphics (STAG), Pages 85-90, October 2021.
   DOI: 10.2312/stag.20211477. — View PDF. | **Honorable mention in best paper awards*
3. **Exploiting Neighboring Pixels Similarity for Effective SV-BRDF Reconstruction from Sparse MLICs**
   Ruggero Pintus, Moonisa Ahsan, Fabio Marton, and Enrico Gobbetti, In The 19th Eurographics Workshop on Graphics and Cultural Heritage, November 2021.
   DOI: 10.2312/gch.20211412. — View PDF. | **Best paper award at GCH 2021.*

Personal Details

Languages Proficiency	Urdu (native), Punjabi(native), English (fluent), Italian (basic)
Place of Birth	Pakistan

Creative Works / Cultural Heritage Pilots

1. **Stele Di Nora** — Demo Video.
   Cultural Heritage Pilot: acquisition, reconstruction and exploration.
2. **Retablo di San Bernardino** — Demo Video.
   Cultural Heritage Pilot: acquisition, reconstruction and exploration (2 selected paintings).

Volunteer Work

Student Mentoring — Remote/Internationally, 2015 – till present
Online sessions with students regarding international scholarships, scientific research and STEM opportunities all over the world, particularly in EU.