A novel approach for exploring annotated data with interactive lenses

F. Bettio, M. Ahsan, F. Marton, and E. Gobbetti

CRS4, Italy
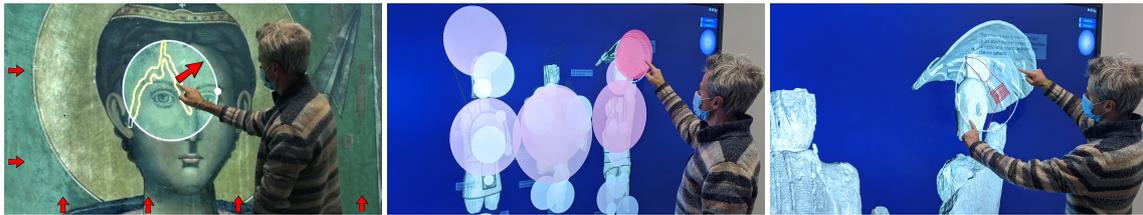
Figure 1: *Exploration of annotated models. We introduce two approaches for improving navigation with interactive lenses. Left: a general control scheme simplifies focus-and-context exploration by jointly adjusting camera and lens parameters in response to user actions. Right: knowledge of an authored annotation database with pre-computed lenses guides users towards interesting regions through an unobtrusive interface. Middle: guidance is provided by selecting target lenses based on a relevance score computed from the current lens position, camera parameters, and navigation history.*

Abstract
We introduce a novel approach for assisting users in exploring 2D data representations with an interactive lens. 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. Navigation is further assisted by exploiting data annotations. 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. While our methods are generally applicable to general 2D visualization, we have implemented them for the exploration of stratigraphic relightable models. The capabilities of our approach are demonstrated in cultural heritage use cases. A user study has been performed in order to validate our approach.

CCS Concepts
• Computing methodologies → Computer graphics; Graphics systems and interfaces;

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 [TGK*17]. 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) [CKB09].

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 [TGK*17]. In this paper, 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 Sec. 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 [CKB09]. 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 (Sec. 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. We further improve navigation with lenses by exploiting and extending the concept of *data annotations* to provide *guidance* (Sec. 4). Guidance is a process where the system provides assistance in response to information on user needs [CGM*16]. While existing approaches mostly guide the interpretation of visualizations [CGM19], 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 [PCDS20]. Such visual cues, which can come from automated analysis or manual mark-up processes, are known to make data understanding easier for the viewer [VETL18]. Finding relevant annotations, and presenting them in a comprehensible way without cluttering the display, however, is very challenging [CCJ14].

In our approach, we associate to each annotation a lens configuration that highlights the region of interest. The stored information include 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.

While our methods are generally applicable to any 2D visualization, our motivating application is in 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 very limited time span for using the system. In particular, we have implemented them for the exploration of stratigraphic relightable models, for which the tunable parameters include the visualized layer, its rendering mode, and the illumination environment. 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 (Sec. 5).

2 Related work

Interactive lenses and annotated-model visualization are vast and well-researched subjects, and a complete review of the literature is out of the scope of this paper. We discuss here only the approaches most closely related to ours.

Guiding or assisting users 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 [Sil91, Hor99], guidance has more recently targeted the support to users during interactive visual analytics work [CGM19]. Ceneda et al. [CGM*16] 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 [CGM*16]. 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 [SPT06, GVH*07, DBGBR*14], others focus on creating animated paths, by arranging viewpoints into graphs [DH02] or letting users define video-tours [WSA*18]. None of the previous approaches target lenses. The camera-control work of Balsa et al. [BAB*16] 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 [Fur86] for trees and extended by van Ham and Perer [VHP09] to graphs. Similarly to Gladisch et al. [GST13], 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.

Interactive control of lenses and surrounding regions Lens-based visualization has many aspects, and we refer the reader to the recent survey by Tominski et al. [TGK*17] 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 multitouch interfaces include methods to create and delete lenses (e.g., with five-finger picks [CBF14]), to manipulate the lens geometry (e.g., with pinch gestures [KRD14]), or to parameterize the lens operations (e.g., by controlling zoom levels with pinch [SS14]). The manipulation of lens position and scale has been treated, so far, especially related to lens magnification, by introducing high-precision control [ACP10] or hierarchies of focus regions [JGE12]. We introduce, instead, new ways to jointly control the relative positioning and scaling of the focus and context areas.

Exploration of models with visual annotation overlays 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 [PCDS20]. While annotation markers can be placed on surfaces in many ways, including labels [BNC*03, SCS05, JSI*10] or hot-spots [CLDS13, PCD*15, PCDS20], our focus is on visual image overlays, which augment the annotated regions with text or drawings draped over the surface. This 2D representation of-

fers 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 [BGMT13, LBM17, DPM*19, CCDL*20]. 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 [PCD*15], 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) [ED07]. 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 [EBD05]), but not for overlay images draped over surfaces. Jaspe et al. [JPGG19, JAP*21] also used lenses, but assumed non-overlapping annotations. In our context, the automatic selection of annotations also provides navigation assistance.

Stratigraphic and relightable models 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 model. In the last decade, 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 [MAD*18, PSK09] or multi-light image collections [VHW*18, Mac15]), or dynamic exploration through relighting [PDC*19]. In this context, lens-based interaction with such models has been used previously by Jaspe et al. [JPGG19, JAP*21] solely with the purpose of letting users see different layers inside or outside the lens. By contrast, this work proposes novel techniques to move the lens for free, assisted, and fully guided navigation.

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 better concentrate on their analysis task by automatically adjusting camera and lens parameters in response to user actions.

3.1 Control scheme

Our joint camera and lens controller, see Fig. 2, evolves as a state machine responding to user events, using the mapping described in Sec. 3.3. Its behavior is as following:
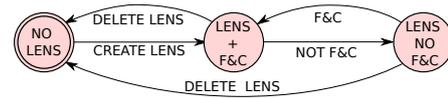

Figure 2: *State machine for joint camera and lens control.*

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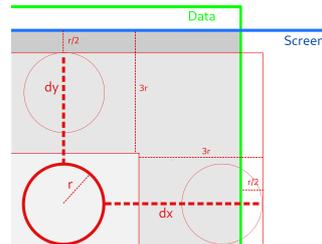

Figure 3: *Joint camera and lens parameter adjustment. The motion of the lens is subdivided between motion of lens and motion of camera based on the amount of context available, as indicated by dx and dy, and the direction of motion.*

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 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 pixel 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 Fig. 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 paper), 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 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 Fig. 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 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.

4 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 [PCDS20]. In this paper, 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 Fig. 4).

4.1 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 article, 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 Sec. 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.

4.2 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 Sec. 4.2.1, and H_i is the history score depending upon the activity log of the active user (see Sec. 4.2.2). The next best lens is the one with the largest S_i.

4.2.1 Spatial and semantic similarity score

Navigating through visually annotated details in multiscale visualizations requires a tradeoff between several conflicting criteria.
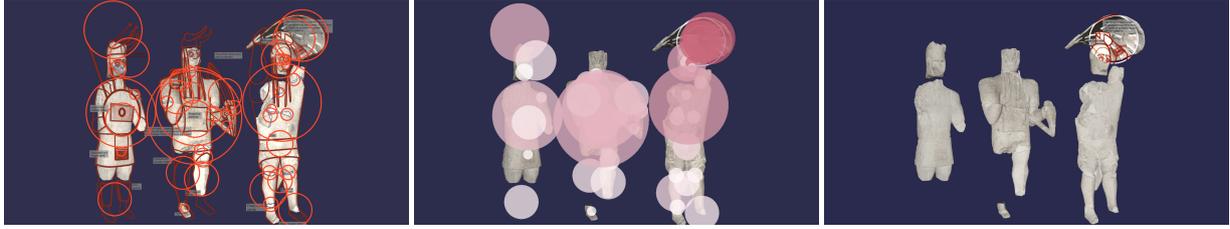
Figure 4: *Annotation selection. Annotations with annotated lenses cover the dataset with a lot of overlap (left). At run-time we rank the annotations based on a similarity computation with the current lens and view (middle), and select the best annotations based on the assigned score. If the selected 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 annotation are color-coded white to red based on the score computed for the lens in the right image.*

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 Sec. 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.
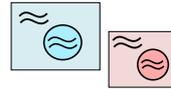

Location similarity The context similarity measure $\sigma_i^{context}$ returns, by design, a constant score for all lenses i very close or very far to the current lens, since the context will either remain unchanged or will be without overlap. In both extreme cases, however, it is reasonable to favor close lenses to far ones, since moving to a closer lens favors locality and reduce travel time and visual change. Thus, we introduce a loca-

tion 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 this paper). 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 paper 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 $\sigma_i^{content} = w s_{in} + (1-w) s_{out}$, 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.

4.2.2 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 need 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 $x \le x_0$, 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 $x \le x_1$ 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 raise 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 $H_i = w_i^{presented} w_i^{rejected}$.
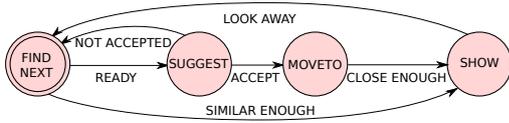
Figure 5: *State machine for assisted navigation in an annotated model.*

4.3 Assisting navigation

Our assisted navigation system, see Fig. 5, 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 Sec. 4.2 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 Sec. 4.4). 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.

4.4 User interface and device mapping

Assisted navigation based on annotations must augment the user interface and device mapping of Sec. 3.3 to handle information coming from the recommendation system that runs in background.

When the next best annotation is judged to be sufficiently similar (see Sec. 4.3), and thus also close in position and scale, the rendering parameters of the current lens are, if needed, changed to the target
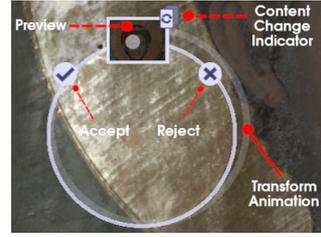

Figure 6: *Lens with suggestions. During suggestion presentation, accept/reject buttons and indications of content and direction of changes for target lens are presented.*

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 Fig. 6). 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.

5 Implementation and results

A reference system integrating all techniques described in this paper 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.

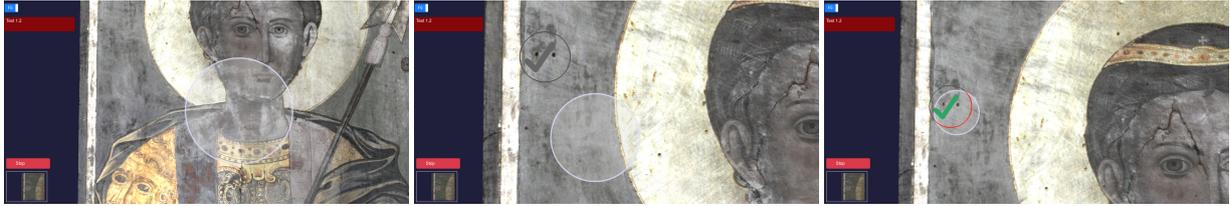We have extensively tested our system with a number of complex

Figure 7: *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 surrounding 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).*

heterogeneous datasets. In this paper, without loss of generality, we demonstrate its usage on two use cases stemming from the cultural heritage domain: a painting and a 2D projection of fragmented sculptures (see accompanying videos). The painting use-case concerns the exploration of relightable stratigraphic model of the Icon of St. Demetrios (17th - 18th century), see Fig. 1 left, containing a normal map and six color layers (visible, 2xIR, 2xUV, FC) generated from a multispectral RTI acquistion. 33 annotations describe various damages (in particular cracks, woodworms, paint defects) and artistic/decoration details. The sculpture use-case concerns, instead, the exploration of a multi-layered rendered image of three representative models from the Mont'e Prama collection of prehistoric stone sculptures [BJM*15]: Archer *n*.5, Boxer *n*.15, and Warrior *n*.3 (see Fig. 1 right). The relightable stratigraphic model 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 Fig. 4) concern reconstruction hypotheses, artistic details and part descriptions.

In order to provide a preliminary assessment of the effectiveness of our approach, we designed and carried out two user studies focused on the proposed novel interaction capabilities. The first one targets on lens control for general scenes (Sec. 5.1), while the second one focuses on assisted exploration of annotated scenes (Sec. 5.2). 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.

5.1 Evaluation of lens control

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 controls scheme is identified with *LC* (for lens+camera), while the standard scheme is identified by *STD*.

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.

Tasks The experiments consisted in letting users try the two different manipulation controllers in the context of a target-oriented

user interaction task [MCG09]. We designed our task to measure performance for the macro-structure and micro-structure inspections tasks typical of cultural heritage model explorations (see Fig. 7). 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 users places its lens over the target lens.

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 [Bro96], 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 [LS09], Q4 and Q10 provide indication 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.
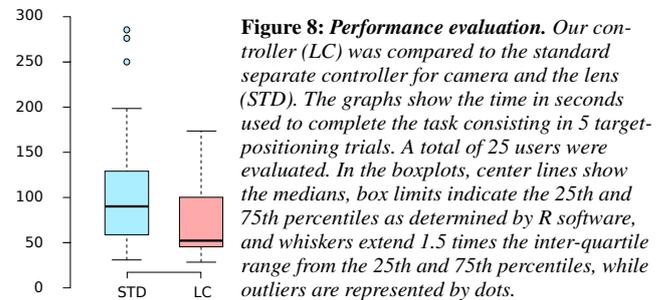

Figure 8: *Performance evaluation. Our controller (LC) was compared to the standard separate controller for camera and the lens (STD). The graphs show the time in seconds used to complete the task consisting in 5 target-positioning trials. A total of 25 users were evaluated. In the boxplots, center lines 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.*

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. Fig. 8 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 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 Fig. 8). 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.
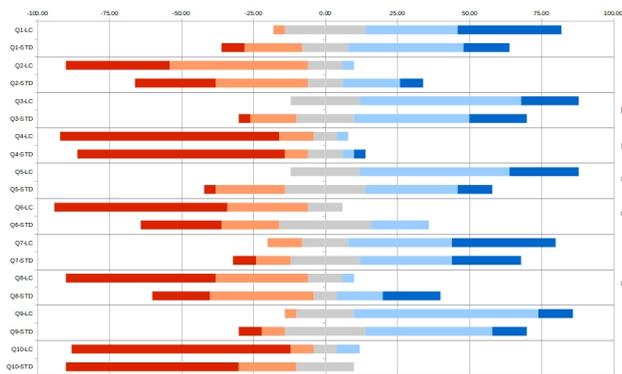

Figure 9: *Usability evaluation of lens control. Diverging stacked bar charts of SUS questionnaire responses concerning our controller (LC) and the standard controller (STD). The color scale goes from red (strongly disagree) to blue (strongly agree). The labels near the right axis summarize the per-question statistical significance resulting from ANOVA (ns → $p > 0.05$; ⋆ → $p \leq 0.05$; ⋆⋆ → $p \leq 0.01$).*

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	⋆	⋆	ns	ns	⋆⋆	⋆⋆	ns	⋆⋆	ns	ns	⋆

Table 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 → $p > 0.05$; ⋆ → $p \leq 0.05$; ⋆⋆ → $p \leq 0.01$).*

Usability evaluation By analyzing the responses of the SUS questionnaires, summarized in the bar charts of Fig. 9, we obtain for our joint controller a SUS score of 79.6, which, according to standard practices [Bro13], 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 Tab. 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 (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 $p < 0.05$).

5.2 Evaluation of assisted exploration of annotated models

The previous user study concerned lens control in isolation. Our second user study was focused, instead, on a preliminary assessment of the effectiveness of our assisted navigation techniques for an annotated model. While our 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 [BBM17]. 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 [HSC08].

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 [ZMVPL18]. Thus, similarly to previous work on evaluating camera control in museum settings [BAMG15], 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.

Setup We used the same web-based setup of Sec. 5.1, 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 (Fig. 10): our fully assisted navigation system described in Sec. 4 (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 users 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
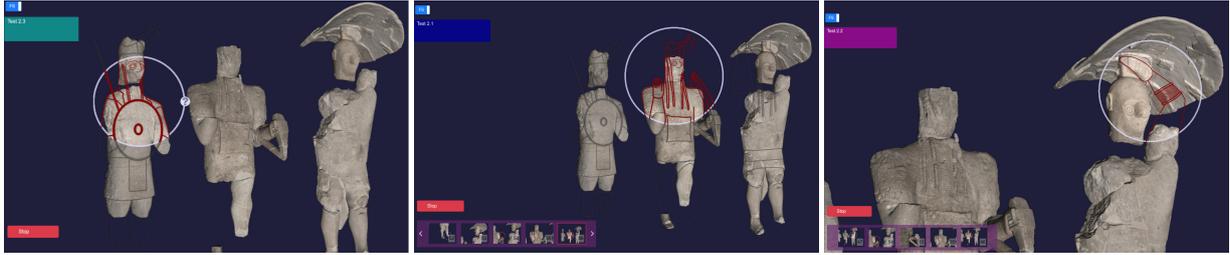
Figure 10: *Assisted navigation user interface evaluation. Left: our controller; Middle: static thumbnail bar; Right: Adaptive thumbnail bar*

importance. Instead, the second version (DYN) is dynamic and presents the current top five annotation targets according to our similarity score.

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 user's attention span.

Design Similarly to what presented in Sec. 5.1, 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 Sec. 5.1 and to optionally provide free-form comments.

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 Sec. 4.3. 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).

Usability evaluation By analyzing the responses of the SUS questionnaires, summarized in the bar charts of Fig. 11, we obtain for our guided interface a SUS score of 85.4, which, according to standard practices [Bro13], 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 Tab. 2. 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 compar-
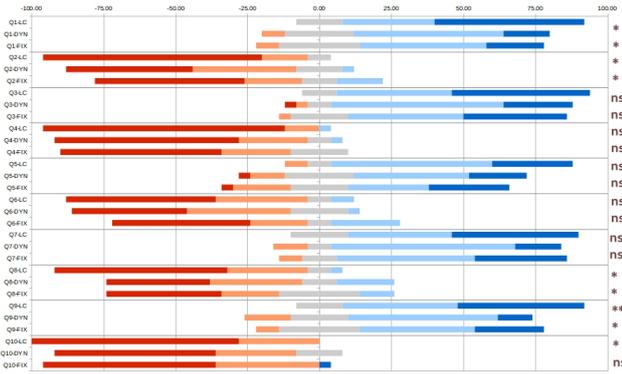
Figure 11: *Usability evaluation of assisted exploration of annotated models. Diverging stacked bar charts of SUS questionnaire responses concerning our controller (LC), static thumbnail bars (FIX), and dynamic thumbnail bars (DYN). The color scale goes from red (strongly disagree) to blue (strongly agree). The labels near the right axis summarize the per-question statistical significance resulting from ANOVA (ns → $p > 0.05$; ⋆ → $p \leq 0.05$; ⋆⋆ → $p \leq 0.01$).*

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 2: *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 → $p > 0.05$; ⋆ → $p \leq 0.05$; ⋆⋆ → $p \leq 0.01$).*

ing 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 [LBB*19] 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.

6 Conclusions

We have presented a novel approach for exploring visually annotated models using an interactive lens.

First of all, we have introduced 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 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.

Moreover, 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 use-cases 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. We plan to further extend this approach 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 different than casual inspection. Addressing this would require cognitive measures that are beyond the scope of the paper, and are an important avenue for future work.

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 [RSA*18], which act on patches of 2D manifolds built to attach smoothly to non-flat surfaces. We plan to look at extending our approach to 3D by sliding and scaling these patches around the surface while maintaining enough context visible, and to extend the guidance to surface-based annotations by suitably modifying the distance functions.

References

[ACP10] APPERT C., CHAPUIS O., PIETRIGA E.: High-precision magnification lenses. In *Proc. SIGCHI* (2010), pp. 273–282. doi:10.1145/1753326.1753366. 2

[BAB*16] BALSA RODRIGUEZ M., AGUS M., BETTIO F., MARTON F., GOBBETTI E.: Digital Mont'e Prama: Exploring large collections of detailed 3D models of sculptures. *ACM JOCCH 9*, 4 (September 2016), 18:1–18:23. doi:10.1145/2915919. 2

[BAMG15] BALSA RODRIGUEZ M., AGUS M., MARTON F., GOBBETTI E.: Adaptive recommendations for enhanced non-linear exploration of annotated 3D objects. *Computer Graphics Forum 34*, 3 (2015), 41–50. doi:10.1111/cgf.12616. 8

[BBM17] BARBIERI L., BRUNO F., MUZZUPAPPA M.: Virtual museum system evaluation through user studies. *Journal of Cultural Heritage 26* (2017), 101–108. doi:10.1016/j.culher.2017.02.005. 8

[BGMT13] BALSA RODRIGUEZ M., GOBBETTI E., MARTON F., TINTI A.: 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. 3

[BJM*15] BETTIO F., JASPE VILLANUEVA A., MERELLA E., MARTON F., GOBBETTI E., PINTUS R.: Mont'e Scan: Effective shape and color digitization of cluttered 3D artworks. *ACM JOCCH 8*, 1 (2015), 4:1–4:23. doi:10.1145/2644823. 7

[BNC*03] BOWMAN D. A., NORTH C., CHEN J., POLYS N. F., PYLA P. S., YILMAZ U.: Information-rich virtual environments: theory, tools, and research agenda. In *Proc. ACM VRST* (2003), pp. 81–90. doi:10.1145/1008653.1008669. 2

[Bro96] BROOKE J.: SUS: a quick and dirty usability scale. In *Usability Evaluation In Industry*, Jordan P., Thomas B., McClelland I., Weerdmeester B., (Eds.). CRC Press, 1996, ch. 21, pp. 189–195. doi:10.1201/9781498710411-35. 7

[Bro13] BROOKE J.: SUS: a retrospective. *Journal of usability studies 8*, 2 (2013), 29–40. URL: 10.5555/2817912.2817913. 8, 9

[CBF14] CHAPUIS O., BEZERIANOS A., FRANTZESKAKIS S.: Smarties: An input system for wall display development. In *Proc. SIGCHI* (2014), pp. 2763–2772. doi:10.1145/2556288.2556956. 2

[CCDL*20] CROCE V., CAROTI G., DE LUCA L., PIEMONTE A., VÉRON P.: Semantic annotations on heritage models: 2D/3D approaches and future research challenges. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 43* (2020), 829–836. doi:10.5194/isprs-archives-XLIII-B2-2020-829-2020. 3

[CCJ14] CAMBA J., CONTERO M., JOHNSON M.: Management of visual clutter in annotated 3D CAD models: a comparative study. In *International Conference of Design, User Experience, and Usability* (2014), Springer, pp. 405–416. doi:10.1007/978-3-319-07626-3_37. 2

[CGM*16] CENEDA D., GSCHWANDTNER T., MAY T., MIKSCH S., SCHULZ H.-J., STREIT M., TOMINSKI C.: Characterizing guidance in visual analytics. *IEEE TVCG 23*, 1 (2016), 111–120. doi:10.1109/TVCG.2016.2598468. 2

[CGM19] CENEDA D., GSCHWANDTNER T., MIKSCH S.: A review of guidance approaches in visual data analysis: A multifocal perspective. *Computer Graphics Forum 38*, 3 (2019), 861–879. doi:10.1111/cgf.13730. 2

[CKB09] COCKBURN A., KARLSON A., BEDERSON B. B.: A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys 41*, 1 (2009), 1–31. doi:10.1145/1456650.1456652. 1, 2

[CLDS13] CALLIERI M., LEONI C., DELLEPIANE M., SCOPIGNO R.: 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. doi:10.1145/2466533.2466548. 2

[DBGBR*14] DI BENEDETTO M., GANOVELLI F., BALSA RODRIGUEZ M., JASPE VILLANUEVA A., SCOPIGNO R., GOBBETTI E.: Exploremaps: Efficient construction and ubiquitous exploration of panoramic view graphs of complex 3D environments. *Computer Graphics Forum 33*, 2 (2014), 459–468. doi:10.1111/cgf.12334. 2

[DH02] DENNIS B. M., HEALEY C. G.: Assisted navigation for large information spaces. In *Proc. IEEE Visualization* (2002), pp. 419–426. doi:10.5555/602099.602165. 2

[DPM*19] DE LUCA L., PIERROT-DESEILLIGNY M., MANUEL A., CHEVRIER C., BENJAMIN LOLLIER P. B., PAMART A., PETELER F., ABERGEL V., ALAOUI A.: Aioli – a reality-based 3D annotation platform for the collaborative documentation of heritage artefacts, 2019. [Online; accessed 03-Mar-2021]. URL: http://www.aioli.cloud/. 3

[EBD05] ELLIS G., BERTINI E., DIX A.: 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. 3

[ED07] ELLIS G., DIX A.: A taxonomy of clutter reduction for information visualisation. *IEEE TVCG 13*, 6 (2007), 1216–1223. doi:10.1109/TVCG.2007.70535. 3

[Fur86] FURNAS G. W.: Generalized fisheye views. *ACM SIGCHI Bulletin 17*, 4 (1986), 16–23. doi:10.1145/22627.22342. 2

[GST13] GLADISCH S., SCHUMANN H., TOMINSKI C.: 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. 2

[GVH*07] GÖTZELMANN T., VÁZQUEZ P.-P., HARTMANN K., GERMER T., NÜRNBERGER A., STROTHOTTE T.: Mutual text-image queries. In *Proc. Spring Conf. Comput. Graph.* (2007), ACM, pp. 139–146. doi:10.1145/2614348.2614368. 2

[Hor99] HORVITZ E.: Principles of mixed-initiative user interfaces. In *Proc. SIGCHI* (1999), pp. 159–166. doi:10.1145/302979.303030. 2

[HSC08] HINRICHS U., SCHMIDT H., CARPENDALE S.: EMDialog: Bringing information visualization into the museum. *IEEE TVCG 14*, 6 (2008), 1181–1188. doi:10.1109/TVCG.2008.127. 8

[JAP*21] JASPE VILLANUEVA A., AHSAN M., PINTUS R., GIACHETTI A., GOBBETTI E.: Web-based exploration of annotated multi-layered relightable image models. *ACM JOCCH* (2021). To appear. 3

[JGE12] JAVED W., GHANI S., ELMQVIST N.: 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. 2

[JPGG19] JASPE VILLANUEVA A., PINTUS R., GIACHETTI A., GOBBETTI E.: Web-based multi-layered exploration of annotated image-based shape and material models. In *Proc. GCH* (November 2019), pp. 33–42. doi:10.2312/gch.20191346. 3

[JSI*10] JANKOWSKI J., SAMP K., IRZYNSKA I., JOZWOWICZ M., DECKER S.: Integrating text with video and 3D graphics: The effects of text drawing styles on text readability. In *Proc. ACM SIGCHI* (2010), pp. 1321–1330. doi:10.1145/1753326.1753524. 2

[KRD14] KISTER U., REIPSCHLÄGER P., DACHSELT R.: 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. 2

[LBB*19] LEKSCHAS F., BEHRISCH M., BACH B., KERPEDJIEV P., GEHLENBORG N., PFISTER H.: Pattern-driven navigation in 2D multiscale visualizations with scalable insets. *IEEE TVCG 26*, 1 (2019), 611–621. doi:10.1109/TVCG.2019.2934555. 10

[LBM17] LAVRIČ P., BOHAK C., MAROLT M.: 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. 3

[LS09]   LEWIS J. R., SAURO J.: 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`. 7

[Mac15]   MACDONALD L. W.: *Realistic visualisation of cultural heritage objects*. PhD thesis, UCL (University College London), 2015. 3

[MAD*18]   MOUTAFIDOU A., ADAMOPOULOS G., DROSOU A., TZOVARAS D., FUDOS I.: Multiple material layer visualization for cultural heritage artifacts. In *Proc. GCH* (2018), pp. 155–159. `doi:10.2312/gch.20181353`. 3

[MCG09]   MARTINET A., CASIEZ G., GRISONI L.: 3D positioning techniques for multi-touch displays. In *Proc. ACM VRST* (2009), pp. 227–228. `doi:10.1145/1643928.1643978`. 7

[PCD*15]   POTENZIANI M., CALLIERI M., DELLEPIANE M., CORSINI M., PONCHIO F., SCOPIGNO R.: 3DHOP: 3D heritage online presenter. *Computers & Graphics 52* (2015), 129–141. `doi:10.1016/j.cag.2015.07.001`. 2, 3

[PCDS20]   PONCHIO F., CALLIERI M., DELLEPIANE M., SCOPIGNO R.: Effective annotations over 3D models. *Computer Graphics Forum 39*, 1 (2020), 89–105. `doi:10.1111/cgf.13664`. 2, 4

[PDC*19]   PINTUS R., DULACHE T., CIORTAN I., GOBBETTI E., GIACHETTI A.: State-of-the-art in multi-light image collections for surface visualization and analysis. *Computer Graphics Forum 38*, 3 (2019), 909–934. `doi:10.1111/cgf.13732`. 3

[PSK09]   PONTO K., SERACINI M., KUESTER F.: Wipe-Off: An intuitive interface for exploring ultra-large multi-spectral data sets for cultural heritage diagnostics. *Computer Graphics Forum 28*, 8 (2009), 2291–2301. `doi:10.1111/j.1467-8659.2009.01532.x`. 3

[RSA*18]   ROCHA A., SILVA J. D., ALIM U. R., CARPENDALE S., SOUSA M. C.: Decal-lenses: Interactive lenses on surfaces for multivariate visualization. *IEEE TVCG 25*, 8 (2018), 2568–2582. `doi:10.1109/TVCG.2018.2850781`. 10

[SCS05]   SONNET H., CARPENDALE S., STROTHOTTE T.: Integration of 3D data and text: The effects of text positioning, connectivity, and visual hints on comprehension. In *Proc. Interact* (2005), vol. 3585 of *LNCS*, pp. 615–628. `doi:10.1007/11555261_50`. 2

[Sil91]   SILVER M. S.: Decisional guidance for computer-based decision support. *MIS quarterly* (1991), 105–122. `doi:10.2307/249441`. 2

[SPT06]   SOKOLOV D., PLEMENOS D., TAMINE K.: Methods and data structures for virtual world exploration. *The Visual Computer 22*, 7 (2006), 506–516. `doi:10.1007/s00371-006-0025-3`. 2

[SS14]   SADANA R., STASKO J.: 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`. 2

[TGK*17]   TOMINSKI C., GLADISCH S., KISTER U., DACHSELT R., SCHUMANN H.: Interactive lenses for visualization: An extended survey. *Computer Graphics Forum 36*, 6 (2017), 173–200. `doi:10.1111/cgf.12871`. 1, 2

[VETL18]   VANHULST P., EVEQUOZ F., TUOR R., LALANNE D.: A descriptive attribute-based framework for annotations in data visualization. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics* (2018), Springer, pp. 143–166. `doi:10.1007/978-3-030-26756-8_7`. 2

[VHP09]   VAN HAM F., PERER A.: "search, show context, expand on demand": Supporting large graph exploration with degree-of-interest. *IEEE TVCG 15*, 6 (2009), 953–960. `doi:10.1109/TVCG.2009.108`. 2

[VHW*18]   VANDERMEULEN B., HAMEEUW H., WATTEEUW L., VAN GOOL L., PROESMANS M.: Bridging multi-light & multi-spectral images to study, preserve and disseminate archival documents. *Archiving Conference 2018*, 1 (2018), 64–69. `doi:10.2352/issn.2168-3204.2018.1.0.15`. 3

[WSA*18]   WANG Z., SHI W., AKOGLU K., KOTOULA E., YANG Y., RUSHMEIER H.: Cher-ob: A tool for shared analysis and video dissemination. *ACM JOCCH 11*, 4 (2018), 1–22. `doi:10.1145/3230673`. 2

[ZMVPL18]   ZAMORA-MUSA R., VELEZ J., PAEZ-LOGREIRA H.: Evaluating learnability in a 3D heritage tour. *Presence: Teleoperators and Virtual Environments 26*, 4 (2018), 366–377. `doi:10.1162/PRES_a_00305`. 8