Eurographics Conference on Visualization (EuroVis) 2015
H. Carr, K.-L. Ma, and G. Santucci
(Guest Editors)

Volume 34 (2015), Number 3

Adaptive Recommendations for Enhanced Non-linear Exploration of Annotated 3D Objects

M. Balsa Rodriguez, M. Agus, F. Marton, and E. Gobbetti

Visual Computing Group, CRS4, Italy – http://www.crs4.it/vic/

Abstract

We introduce a novel approach for letting casual viewers explore detailed 3D models integrated with structured spatially associated descriptive information organized in a graph. Each node associates a subset of the 3D surface 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. At run-time, users navigate inside the 3D scene using a camera controller, while adaptively receiving unobtrusive guidance towards interesting viewpoints and history- and location-dependent suggestions on important information, which is adaptively presented using 2D overlays displayed over the 3D scene. The capabilities of our approach are demonstrated in a real-world cultural heritage application involving the public presentation of sculptural complex on a large projection-based display. A user study has been performed in order to validate our approach.

Categories and Subject Descriptors (according to ACM CCS): I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction techniques I.5.2 [Information Interfaces And Presentation (HCI)]: User Interfaces—Interaction styles, Input devices and strategies

1. Introduction

Digital multimedia content and presentations means are rapidly increasing their sophistication and are now capable in many application domains of describing detailed representations of the physical world. Providing effective 3D exploration experiences is particularly relevant when the goal is to allow people to appreciate, understand and interact with intrinsically 3D virtual objects. Cultural Heritage (CH) valorization and Cultural Tourism are among the sectors that benefit most from this evolution, as multimedia technologies provide effective means to cover the pre-visit (documentation), visit (immersion) and post-visit (emotional possession) phases [EM11]. In order to effectively support a rich, informative, and engaging experience for the general public, 3D representations should, however, go 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. Much of this information requires spatial association, as it describes, or can be related to, different spatial contexts. For instance, cultural artifacts are often very complex 3D objects, with subtle material and shape details, presenting information at multiple scales and levels of abstractions (e.g., global shape and carvings). Even the finest material micro-structure carries valuable information (e.g., on the carving process or on the conservation status).

Until recently, the most widespread ways to present information around 3D reconstructions have been through mostly passive visual presentation modalities, such as videos or computer-generated animations. Interest is, however, now shifting towards more flexible active modalities, which let users directly drive exploration of 3D digital artifacts. These active approaches are known to engage museum visitors and enhance the overall visit experience, which tends to be personal, self-motivated, self-paced, and exploratory [FD00]. In general, visitors do not want to be overloaded with instructional material, but to receive the relevant information, learn, and have an overall interesting experience. To serve this goal, user-friendly and flexible systems are needed, and many challenges need to be addressed in parallel [KSZ*11]. Our work, motivated by a project for the museum presentation of Cultural Heritage objects (see Sec. 3) deals with the particular problem of letting casual viewers explore detailed 3D models integrated with structured spatially associated descriptive information in form of overlaid text and images.

Approach. Our goal is to let users explore spatially anno-

tated 3D models using a walk-up-and-use user-interface that emphasizes the focus on the work of art. Content preparation, done offline with a authoring tool, organizes descriptive information in an information graph. It should be noted that this graph does not provide a 3D scene description, as the usual scene graph, but is used to structure annotations and relate them to the 3D scene. Each node associates a subset of the 3D surface 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. 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 information, which is adaptively presented using 2D overlays displayed over the 3D scene. The approach is implemented within a scalable system, supporting exploration of information graphs of hundreds of viewpoints associated to massive 3D models, using a variety of GUI setups, from large projection displays to smartphones.

Contribution. Our approach, motivated by a real-world visual presentation project in the CH domain, combines and extends state-of-the-art results in several areas. Our main contribution is the flexible integration of stochastic adaptive recommendation system based on a structured spatial information representation, centered around annotated viewpoints, with a walk-up-and-use user interface that provides guidance while being minimally intrusive.

Advantages. Using a graph of views as a basis for information structuring has a number of practical advantages for authors and viewers. In particular, authors can use 2D tools for content preparation, leading to a simple but effective procedure to create spatially relevant rich visual information in forms of overlays, and can use weak dependencies to smoothly transition from constrained sequential presentations (stories) to more flexible independent annotations. The user interface, which automatically selects annotated views, suggesting them and smoothly guiding users towards them, is engaging as it gives users full control on navigation, while being unobtrusive and avoiding the requirements of precise picking, as opposed to the more common hot-spot techniques.

Limitations. This work only targets the problem of 3D model exploration with image/text overlays. Using other associated multimedia information (e.g., video) and/or supporting very complex narratives are orthogonal problems not treated in this work. Moreover, while the proposed information presentation system is of general use, the proposed camera navigation technique is tuned for object inspection rather than environment walkthroughs. Finally, the current evaluation focuses mostly on user satisfaction. More work is required to objectively assess the effectiveness of the user interface in a variety of settings. Addressing this would require cognitive measures that are beyond the scope of the paper, and are an important avenue for future work.

2. Related Work

We briefly discuss the methods that are most closely related to ours. We refer the reader to well-established surveys on 3D interaction techniques [JH13, CO09], CH visualization strategies [FPMT10], and cognitive aspects of visual-spatial displays [Heg11] for a wider coverage.

Contextual information representation and presentation. 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 [Heg11]. Our work falls in the category of visual-spatial displays that dynamically mix 3D representation with associated overlays. We do not focus on designing navigation aids or displays for specific tasks (e.g., location awareness), but, rather, simply on providing flexible means to unobtrusively guide the user towards "interesting" nearby locations and to present contextual information. 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 [BCQ*07]. Most works on information visualization [LCWL14] concentrate on data analysis, extrapolating results and presenting them using graphical representations tailored for better human comprehension, while we focus on techniques for enhancing 3D object exploration. Riedl et al. [RY06] propose narrative mediation as alternative to branching stories in order to provide non-linear narrative generation, but there is no mention to handling spatial relations. We propose, instead, a graph-based representation that contains spatial and hierarchical dependencies between nodes. Using linked multimedia information to enhance the presentation of complex data has been long studied, mostly focusing on guided tours [FS97], text disposition and readability [SCS05, JSI*10], usability of interaction paradigms [PBN11], and the integration of interconnected text and 3D model information with bidirectional navigation [GVH*07b, JD12, CLDS13]. All these methods require precise 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. We propose, instead, a method to present contextual information associated to regions of interest in selected viewpoints, without requiring precise picking. 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 [IH12]. These are discussed below in the context of view-based navigation.

View-based navigation. A number of authors have proposed to automatically compute interesting viewpoints in order to guide the viewer [BDP00, SPT06, GVH*07a]. We focus, instead, on the orthogonal problem of proposing views that have been previously annotated. Using views to help navigate within a 3D dataset is often implemented with thumbnail-bars. At any given time, one image of the dataset can be selected by the user as current focus, moving the camera to the associated

viewpoint [Lip80, SSS06, DBGB*14]. Often, these images are also linked to additional information, which is displayed when the user reaches or selects them, as an alternative to the usage of hot-spots [ACB12]. The organization of the images in this kind of tools can be considered a challenging problem in many scenarios, since simple grid layout approaches do not scale up well enough with the number of images. A trend consists in clustering images hierarchically, according to some kind of image semantic [GSW*09, RCC10]. Most of these works strive to identify good clustering for images, rather than a good way to dynamically present and explore the clustered dataset. Our approach instead is navigation-oriented and it is organized in a way that, in any moment, users can decide to change the point of view and trigger display of overlaid information. Similar concepts can be found in the bi-directional hyperlink system of Goetzelmann et al. [GVH*07b] and in the exploration system of Marton et al. [MBB*14], which use gestures to provide contextual information. We significantly extend prior work by using a structured information representation and introducing an adaptive stochastic recommendation system.

Motion control. In the context of visualization of complex scenes, the user requires interactive control to effectively explore the data. Many solutions have been proposed for camera/object motion control [CO09], and our method can adapt to many of them. In this work, we use a Virtual Trackball with auto-centering pivot [BAMG14] since it has a familiar user-interface mapping and does not require precise picking to define the rotation pivot.

3. Overview

While our approach is of general use, our work has been motivated by the Mont'e Prama project, a collaborative effort between our center and the *Soprintendenza per i Beni Archeologici per le Province di Cagliari ed Oristano* (ArcheoCAOR, the government department responsible for the archeological heritage in South 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 buildings. The project covers aspects ranging from 3D digitization to visual exploration.

In order to design our model exploration technique, we embarked in a participatory design process involving domain experts with the goal of collecting the detailed requirements of the application domain; the experts included two archaeologists, two restoration experts, and one museum curator from *Museo Archeologico Nazionale di Cagliari*. Additional requirements stem from our analysis of related work (see Sec. 2) and our own past experience developing exploration systems in museum settings [BAMG14, MBB*14]. We now describe the main requirements, briefly summarizing how they were derived:

R1 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 be associated to parts of the cultural objects, typically seen from a canonical point of view, or at least close to it. Examples are descriptions of carvings and decorations, reconstruction hypotheses, comparisons with other objects. Different macro-structural and micro-structural views should be associated with different kinds of information.

R2 Information presentation order. Relations exist among the different information to be presented, and experts emphasize that a predefined order should be defined. This order is often not strict, and different storytelling paths are possible.

R3 Information 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.

R4 Information authoring. Textual and visual information (drawings, images) should be supported. Editing should be made possible for museum curators and archaeologists without particular training. Adding annotations and linking them should not require intervention of specialized personnel.

R5 Focus on cultural object (avoid occlusion from interaction widgets). The important information is the visualized object itself, which, as in a real exhibition should not be obstructed by general clutter (e.g., interaction widgets). Visitor focus should thus be guided to the presentation medium.

R6 Fast learning curve and assisted navigation. In a museum installation, 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, since museums must manage large amounts of visitors, long training times and/or guided training are not affordable. The user interface should thus be perceived as simple, immediately usable.

R7 Engaging experience. In general, 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 be personal, self-paced, and exploratory. The user interface should provide guidance, while not being perceived as overly obtrusive.

R8 User interface and display flexibility. In order to cover the whole pre-visit, visit, and post-visit phases, one should support a wide range of setups, including museum setups, as well as smartphone and tablet applications. The specific application case for our design is the presentation of larger-than-life human statues, reaching over 2.5m of height. The statues were constructed at this imposing scale on purpose, and this macro-structure information should be immediately conveyed to the visitor through a real-scale (or larger-than-real) presentation.

This means, in particular, that in museum setups we need to support large (wall-sized) displays.

R9 **Seamless interactive exploration.** Control modes should be active with real-time feedback, in order to engage visitors providing them the sense of control, and support smooth and seamless object inspection, going back and forth from shape inspection to detail inspection.

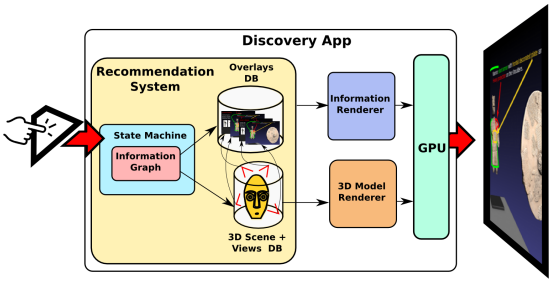

Figure 1: ***System overview.*** *At run-time, users navigate inside the 3D scene, while adaptively receiving unobtrusive guidance towards interesting viewpoints and history- and location-dependent suggestions on important information, which is adaptively presented using 2D overlays displayed over the 3D scene.*

Requirements **R1**–**R9**, as well as our analysis of related work presented in Sec. 2, were used as to drive our design process, which resulted in the definition of an approach based on the following concepts:

- **Information graph and authoring.** We 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 (**R1**), together with its author-defined importance (**R3**). 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 (**R2**). The information graph can be created off-line by selecting a reference view for each presented information, drawing an overlay image using standard 2D tools (**R4**), and indicating dependencies by selecting strong and weak predecessors for each view. This leads to a simple but effective procedure to create spatially relevant rich visual information in forms of linked overlays. Authoring details are orthogonal to the proposed method and are not detailed in this paper. In practice, in this paper, we used the exploration system to select the views that are to be annotated, and store snapshots as PNG images. Annotations are then created using a drawing system (*libreoffice draw*), and exporting the overlays as PNG images. The graph is then created with a simple image browser, that shows images+overlays and defines dependencies by referencing other images, saving the result as an XML file.
- **Exploration.** In order to provide an engaging self-paced experience (**R7**), we let users freely explore 3D models using an interactive camera controller (**R9**), with a user interface that presents in the main view only the 3D scene of interest (**R5**). 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 (**R1**, **R2**, **R3**), see Fig. 1. A suggestion is then stochastically identified among these candidate nodes, with a probability proportional to the score (**R2**, **R7**). The non-deterministic choice respects mandatory presentation orders, supporting classic authored storytelling, while introducing variations in the exploration experience (**R7**). If the selected information node's view parameters are close enough to the current view, the user is unobtrusively guided towards it by smoothly interpolating camera parameters during interaction towards the best view (see below). Otherwise, the proposal is visually presented for a limited time to the user in a small inset viewport (**R5**). If the user accepts it, a small animation is activated to bring the user to the selected target viewpoint. When the user is aligned with the target view, the corresponding textual and visual overlay is displayed on top of the 3D view (**R1**, **R5**). This approach avoids the use of a series of hot-spots over the model, which require pointing methods and/or produce clutter (**R5**). 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, which is able to provide content in a consistent manner, but with different flavors depending on the user attitude to follow the proposed indications (**R7**). This approach mimics the experience of a tour with an expert which describes and highlights the parts of the model on which the user is mostly interested.
- **User-interface and device mapping.** The proposed approach poses little constraints on the GUI, as it requires only means for controlling the camera and accepting a suggestion (**R8**). In particular, we do not employ hot-spots (**R5**) and can rely on incremental controls for camera navigation, as, in particular, we do not require 2D or 3D picking. This makes it possible to implement the method in a variety of settings. In this work, we employ an approach that decouples the devices for interaction and for rendering as to allow for large projection surfaces and enable multiple users to watch the whole screen without occlusion problems and staying at a suitable distance from it when viewing large objects of imposing scale (**R5**,**R8**). The widespread diffusion of touch devices, such as tablets or smartphones, has made people used to touch-based user interfaces. While no real standard for 3D touch-based interaction exists [KI13], touch surfaces are now so common that people are encouraged to immediately start interacting with them, which is an important aspect of *walk-up-and-use* interfaces. Moreover, even if the mapping between 2D and 3D motions is non-trivial and varies for a user interface to the next, users are encouraged to learn by trial and error while interacting. In this work, we use a 3D variation of the well-known 2D multi-touch RST technique,

that allows the simultaneous control of Rotations, Scaling, and Translations from two finger inputs to control a modification of a virtual trackball with auto centering capabilities [BAMG14]), which provides automatic pivot without requiring precise picking. Accepting a suggestion is mapped to a long press, while rejection automatically occurs upon time-out. The motion of the trackball, in addition, is modified so as to attract the view towards the currently selected best view by applying a small nudge force in the direction of the currently selected best view, but only use the component which is orthogonal to the current direction of motion (see Sec. 5). This helps gently guiding the user towards good viewpoints with associated information.

4. The recommendation engine

At the core of our approach is a recommendation engine running in parallel with user navigation. It is based on a state machine (see Sec. 4.2) that evaluates the node contributions and stochastically selects one node with a probability proportional to a context-dependent score depending on the current spatial position and the navigation history (see Sec. 4.3).

4.1. Data representation

The information exploited by the recommendation engine are the *3D model*, an *annotated view graph*, the *current viewpoint*, and the *interaction history*. The first two elements are static and provide the scene description, while the two latter ones are dynamic and evolve during navigation.

The 3D model can be any kind of surface model with a renderable representation. In this work, we use multiresolution triangulated surfaces (see Sec. 6 for scalability issues).

The view graph describes annotations in a structured form, as described in Sec. 3). We denote as $\gamma \in [0..1]$ the author-defined *importance* of each node and as $\omega \in [0..1]$ the dependency weight. Strict dependencies ($\omega = 1$) are useful to model cases where prior information is mandatory (e.g., global introduction is required before presenting some particular detail), while weak dependencies ($\omega < 1$) enable a more adaptive navigation, and if $\omega = 0$ no dependency exists. The descriptive information associated to each node is a 2D overlay image (a semitransparent bitmap or scalable vector graphics with the same aspect ratio of the rendered 3D view). The 2D overlay contains drawings, images or even text which is tightly attached to the object from the node's viewpoint (e.g., imagine a statue with a missing arm and a drawing proposing what could be the missing part, see Fig.2). The 2D ROI consists of a bitmask denoting the relevant part of the view that contains the information referenced in the textual information. We use this ROI in the ranking process, to identify the 3D region that participates in view similarity computation. All this information is connected to a viewpoint that is also stored in the node in the form of a view matrix. In order to speed-up view-similarity computation (see Sec. 6), we also maintain with each node the bounding box of the 3D points contained in the ROI.

In the course of the navigation, we collect data on the user interaction with the system in order to extract aggregated information. This information is held in nodes attributes that provide aggregated information on user preferences like most visited nodes and the amount of time spent per node, providing new relevance weights for each node which can provide better suggestions to the user during navigation. This information is then exploited by the selection algorithm in order to improve future suggestions (see below).

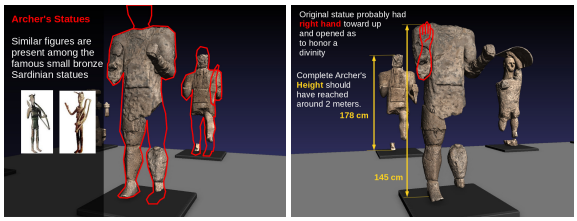

Figure 2: ***Overlaid information.*** *Left: Drawing showing a possible reconstruction of the missing parts of the object; Right: Textual information is presented without cluttering the region of interest.*

4.2. The recommendation state machine

The state machine (*SM*), see Fig. 3) runs in background while the user can freely move around the scene. The SM proposes specific views depending on the user behavior. On user acceptance the corresponding information is visualized. The SM, using the node graph, is able to produce a sequence of contents which tells a structured consistent story, according to user preferences. The SM states (*discover, attract, propose, goto, show*) are here described in detail:

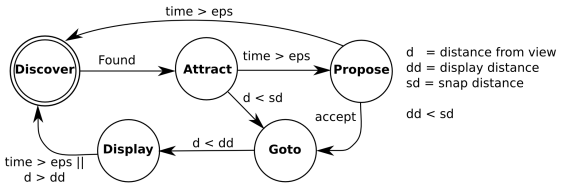

Figure 3: ***State Machine.*** *State machine of the recommendation system.*

- **Discover:** it is the start state. Here the SM lets elapse a few seconds to avoid a continuous flow of suggestions, then it looks for a new node to propose (see Sec. 4.3). Once the node is selected the state passes to *attract*.
- **Attract:** a hidden attraction force is active while the user explores the scene, trying to drive her toward the active view (see Sec. 5). The machine can exit this state after a timeout, and in this case the state changes to *propose*, or if user gets close to the node, and in this case the state changes to *goto*.
- **Propose:** a thumbnail with a snapshot of the selected view is proposed to the user. If the user accepts the proposal, the state changes to *goto*. Otherwise, if the proposed node

is not accepted in a given time, the SM gets back to the *discover* state.

- **Goto:** a small animation is computed and the user is moved to the point of interest associated to the node. After reaching the target point the state changes to *display*.
- **Display:** the information related to the node is displayed. After a time proportional to the length of the textual content is elapsed, or if the user moves away from this position, the state returns to *discover*.

4.3. Next best view selection

The adaptive recommendation system aims to guide the user to a structured exploration, taking into account either user movements and author preferences. At given times, the system selects the next best view to be proposed, possibly in the neighborhood of the area currently explored.

Selection algorithm. The selection is performed according to a ranking of all the visitable views in the graph. First of all, the views are partitioned into two sets (visible and invisible), based on view culling with respect to the current view position. At this point each visible view i is compared to the current view, according to the similarity measure described in next paragraph. If this measure is below a given threshold, the view is added to the set of invisible views, otherwise a score is computed according to the following equation:

$$S_i = \gamma_i \times D_i \times R_i \times \sigma_i \tag{1}$$

where γ_i is the author defined view relevance, D_i the dependency weight, R_i the recent navigation weight, and σ_i is the view similarity weight. All the weights appearing in equation 1 are inside the range $[0..1]$. The dependency weight D_i is a product among all the view dependencies, and it is computed as follows:

$$D_i = \prod_{j=1}^{N_i} (1 - \omega_j \times (1 - vis_j)) \tag{2}$$

where N_i the number of dependencies of view i, ω_j is the dependency weight of view j with respect to view i and vis_j is 1 if the view j has been already visited and 0 otherwise. The weight R_i takes into account the user recent navigation: giving lower priority to the views which have been recently displayed, or presented but not accepted. This value is 1 for all not visited and not proposed nodes, otherwise its value is computed by $R_i = min((\frac{\Delta T}{T_{max}})^2, 1)$, where ΔT is the time elapsed from the last event (propose or visualization), and T_{max} is a time threshold. If at least a view in the visible set has a positive score, the next best view is selected randomly with a probability proportional to the score, otherwise the views inside the invisible set need to be considered. In this case, the scores are computed according to equation 1, but the view similarity weights employ a metric which is robust to distance, which will be detailed in next paragraphs. At this point, the next best view is selected among the ones with positive score, with a probability proportional to the latter.

View similarity metric for close views. For comparing the current view with respect to *visible* annotated views, we derived a metric based on the fact that two similar views would approximately project the same 3D points to the same image pixels. Therefore, the normalized squared sum of the distances of projected points provides an adequate distance metric for deriving the similarity measure used in Eq. 1:

$$\sigma_i = 1 - \frac{\xi^2(K-R) + \sum_{j=1}^{R} min\left(\left(PV_i\mathbf{s_j} - PV_{cur}\mathbf{s_j}\right)^2, \xi^2\right)}{K\xi^2} \tag{3}$$

where, in the current stochastic sampling composed by K samples, $\mathbf{s_j} = \{\mathbf{s_1}, ..., \mathbf{s_R}\}$ is the set of R points inside the region of interest of view i, P is the current projection matrix, V_i is the view matrix of the view i, V_{cur} is the current observer view matrix, and ξ^2 is the maximum squared distance between two visible points in the normalized clipping cube.

View similarity metric for distant views. This above similarity measure is reliable when an adequate number of points are visible in the region of interest of view i, but it is not applicable to views outside the view frustum or with only few sample points in it. In these cases, similarity should not be computed in the image plane. Just computing the distance between view matrics, e.g., using L1 or Frobenius norms, is an applicable solution, but would not take into account the distance from the camera to the (average) lookat point. Thus, small variations in camera orientation, that could lead to large variations in image space, would not be captured. This is why we combine in our metric the motion of the viewpoint with motion of the lookat point, considering the eye-target-twist parameterization of viewing transformation, as a quick way to estimate the length of the path needed to reach the view V_i from the current view position V_{cur}. Specifically, the similarity metric is computed in this way:

$$\sigma_i = 1 - \frac{\|e_i - e_{cur}\| + \|t_i - t_{cur}\|}{2\delta} \tag{4}$$

where δ is the diagonal of the scene bounding box, e_i, e_{cur} are the eye positions associated to view i and the current view, t_{cur} is the center of visible points from current view, and t_i is computed as $t_i = e_i + \|t_{cur} - e_{cur}\|v_i$, with v_i the viewing direction of view i.

5. User interface

The recommendation engine can be integrated in a variety of settings, as, in terms of input, it requires only means for controlling the camera and accepting a suggestion, while, in terms of output, it requires only real-time 3D navigation, suggestion display, and overlay display. In this work, we focus on a museum setting that decouples input and output devices.

Setup and assisted camera control. 3D models and associated information are presented on a large display (a back-projection screen in this work), controlled by a touch-enabled surface placed at a suitable distance in front of it. Note that we avoid using the touch-screen to display content-related information, in order to encourage the user to focus on the

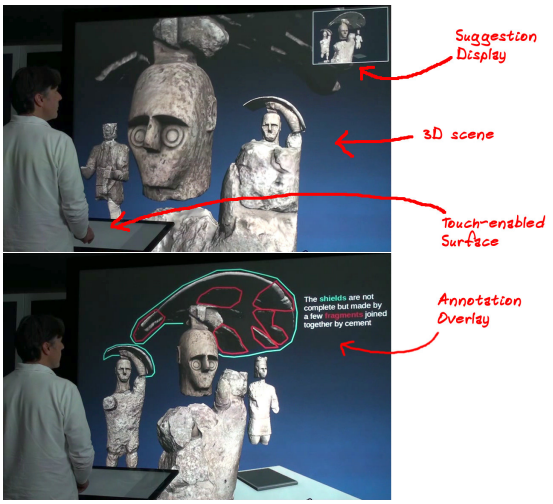

Figure 4: ***Suggestions and overlays.*** *Top: Suggestions are presented in a small inset, using animations to relate them to the spatial context. These suggestions appear only when the attraction forces do not drive the user close enough to the current view. Bottom: when moving close to the currently selected view or accepting a suggestion, annotations are overlaid to 3D view.*

visualization screen instead of concentrating on the user interface, see Fig. 4. An alternative to this setup would be to use 3D devices, e.g., Kinect or Leap Motion, and gestures/posture recognition. Such an implementation, however, is less practical to deploy in crowded museum settings. Camera control is implemented through a multi-touch interface controlling the auto-centering virtual trackball. In order to reduce training times, we considered RST multi-touch gestures, as used for 2D actions in commodity products such as smartphones or tablets, and mapped them to analogue 3D behaviors in our constrained controllers. As we deal with statues, we use a fixed up-vector, and map two-finger pinch to dolly-in and dolly-out, two-finger pan to camera panning, and one-finger horizontal motion to orbiting. It should be noted that, similarly to Secord et al. [SLF*11], we deform the motion of the trackball in order to be attracted towards the currently selected view both during pan and rotate and during throwing (i.e., the small period of time after a release). We also add in a small friction force in the neighborhood of the selected view, so as to slow down near good viewpoints, and, when the view is sufficiently close, we snap it to the best view. A long press, instead, is used to accept the displayed suggestion. Suggestion is also accepted when the view similarity d is below a user-defined threshold. This means that when the view is almost similar, the user is automatically moved to the selected node's position, and the related overlay is accepted.

Displaying suggestions. Each time a new view is selected by the state machine, and the user has not moved close enough to trigger automatic acceptance within a small amount of time, a suggestion is displayed to the user. It should be noted that this situation does not occur very often, since the attraction force automatically drives the user towards the currently selected view during interaction. In our current implementation, suggestions are presented in small inset images using animations (see Fig. 4 top). First, the inset image fades in in a corner of the main 3D view. The initial image presented is a clone of the target view. Then, an animation starts, showing a path from the current view to the target view. This animation is employed to inform the user on the location of the target without cluttering the main 3D view. The target image then remains fixed in the inset for a predetermined amount of time. If, within this time, the user does not accept it (by moving close to the target or performing a long-press to trigger an automatic go-to), the suggestion is considered ignored, the inset image fades out, and the state machines starts looking for alternatives.

Visualizing overlays. When a target node is reached, the associated information is displayed in overlay (see Fig. 4 bottom). The information remains visible until the user decides to move to another position. In order to reduce distraction, the suggestions appears smoothly, combining fade-in/fade-out animation with incremental zooming. We plan in the future to investigate less obtrusive methods, e.g., by using mechanisms for better exploiting change blindness events [Int02].

6. Scalability

Both the interactive inspection and the recommendation system require specialized spatial indexing and multiresolution structures and adaptive algorithms to ensure real-time performance on large datasets (billions of triangles and hundreds of points of interest per scene). The most costly operations are 3D rendering and recommendation computation.

When computing recommendations, the graph is first partitioned in a set of feasible nodes, which are the ones for which predecessors are satisfied. Only these nodes, typically a small subset of the total graph, are checked for similarity. We associate to each node a bounding box, which contains all the points in its ROI, and keep the bounding boxes of potentially visible nodes in a bounding volume hierarchy (BVH). When ranking starts, the BVH is traversed, and nodes are compared with the current view frustum, classifying potentially visible and invisible ones. Potentially visible nodes are pushed in a priority queue, ordered by inverse difference in projected ROI area between target view and current view. Nodes are extracted from this queue one by one, starting with the nodes with most similar projected ROI area, view similarity is computed, and nodes are pushed in a queue sorted by recommendation score, until a small predefined number of nodes is found or there are no more nodes to check. Any node for which similarity is zero is pushed to the invisible set. If the set of nodes for which a score has been computed is non-empty, we stochastically select the suggestion by randomly picking from it with a selection probability proportional to the score. Otherwise, a score is computed for the invisible node, using a linear scan and a fast method that does not require view similarity computation. This approximate tech-

nique keeps the number of view similarity computations low in order to maintain interactivity.

View similarity computation is computed on top of the same adaptive multiresolution triangulation [CGG*04] used for rendering. A small random set of 3D points is extracted from the view-adapted tetrahedron graph. This is done using a simple traversal of the graph leafs, selecting a few points per node. These points are then projected using the view parameters both from the candidate view and the user viewpoint, in order to calculate the average screen space distance between the two point sets and compare it according to Eq. 3. Note that only the points falling within the ROI of the target view participate in the score.

7. Implementation and User Study

A reference system integrating all techniques described in this paper has been implemented on Linux using OpenGL and Qt 4.7. The hardware setup for the interactive stations was composed of a 2.5m-diagonal back-projection screen, a 3000 ANSI Lumen *projectiondesign F20* SXGA+ projector, and a 27" Dell P2714T multi-touch screen, both controlled by a PC with with Ubuntu Linux 14.10, a Intel Core i7-3820 @ 3.6Ghz, with 8GB of RAM and a NVIDIA GTX 680 GPU.

The system, illustrated in the accompanying video, has been tested in a variety of settings. In this paper, we report on tests made using 8 representative models from the Mont'e Prama collection [BJM*14], for a total of 390M triangles, which have been documented using a graph of 109 information nodes linked by 132 edges. Of these, 12 describe mandatory dependencies ($\omega = 1$), 36 strong dependencies ($\omega = 0.8$), and the remaining weak dependencies ($\omega = 0.2$). The graph is a hierarchical DAG with 5 levels, loosely ordered from general collection-level information to micro-structure description. In all tests, the models were adaptively rendered using a target resolution of 0.5 triangles/pixel, leading to an average 2.5M triangles/frame and maintaining frame rates never going below 30Hz. The time required in the recommendation engine to generate a new recommendation has been generally less than 5 ms, leading to minimal interaction delays.

In order to provide a preliminary assessment of the effectiveness of our approach, we designed and carried out a simple user study for evaluating its effectiveness and comparing the proposed method for unobtrusively guiding the user through the recommendation engine with two alternative techniques based on thumbnail-bars [DBGB*14]. While our approach might be applicable to support a human analyst in understanding complex 3D objects and presumably scenes, an important and complex problem that elicits much debate in cognitive science community [Heg11], the proposed tests are tuned for our particular domain-specific application of casual museum visits.

Goal. Our system is composed by a combination of narrative components [SH10] together with a free customized 3D user interface, which makes it difficult to evaluate from a user perspective, if only because of the lack of consensus on metrics and methods for assessing user understanding. In theory, for an adequate system evaluation, the various parts should be considered separately in order to quantify their effects over users [Sch06]. In our case, we opted to design our user study with the target to try to quantify the user satisfaction, in terms of fun and attractiveness, and the user performance, in terms of effort, learning curve, and information gathering [Dia11]. To this end, we measured user performance during free explorations together with a simplified version of NASA task load index questionnaire [LBF*13]. Furthermore, we gathered information from think-a-loud comments.

Configurations. Various alternatives of the usage of the system were considered for the experiments: a free exploration interface with adaptive recommendations, and two versions in which the exploration is decorated with a thumbnail-based bar exploration interface [MBB*14, DBGB*14]. In one of them the views are ordered according to the authoring importance (weights and dependencies), while in the other one the views are ordered according to the ranking of the recommendation system (views similarity and authoring criteria). In any moment, users could scroll the thumbnail-bar and decide to explore a specific view of the scene. The experimental setup considered the reference system implementation described in section 7. All exploration alternatives were operated using the same precise optical touch screen device using a multi-touch device mapping.

Tasks. The experiments consisted in letting users try and enjoy the system using the three different exploration strategies (with adaptive suggestions, with importance sorted thumbnails, and with rank sorted thumbnails) in the context of a free interaction task. We designed our task to measure learning and satisfaction performance in inspections tasks typical of cultural heritage model explorations. Participants were asked to freely explore the model and follow the narrative visualization with the goal of enjoying and acquiring as more useful information as possible.

Participants. For the user analysis, 15 participants (11 males and 4 females, with ages ranging from 31 to 65, mean 42.1 ± 8.9 years) were recruited among students and employees of our center. The user group was composed of 4 members of the administrative staff, 2 members of the technical staff (1 technician, 1 janitor), 6 system administrators, 1 researchers in computational science, and 2 PhD students (1 computer science, 1 bioengineering).

Design. Each participant tested the three exploration systems in randomized order. Users were first allowed to become familiar with the explorative systems by watching a brief video showing how it works (part of the help system of the museum installation). After the training session, the measured tests consisted of trying the 3 different narrative exploration interfaces for 5 minutes each one. For a complete testing session, users needed 15 minutes. In summary, the complete test design consisted of 15 participants, each one testing the 3 exploration interfaces for a total of 45 complete measurements. At the end of the experiments, participants were also asked to fill a questionnaire comparing the performance of the

three systems by indicating a score in a 7-point Likert scale with respect to six factors: mental demand, learning time, physical demand, performance, effort, and frustration level. Since the objective of the tasks was to enjoy the models as to acquire interesting information as much as possible, we asked subjects to quantify as performance level their perception of satisfaction (how much they enjoyed the scene exploration).

Performance evaluation. The following measures were recorded during explorations using the adaptive recommendation system interface (ASI): number of nodes displayed, subdivided in nodes reached through attraction (overlays appearing during exploration), nodes reached through goto animations, and nodes proposed and ignored during exploration. The subjects were proposed an average of 24.6 ± 3.4 nodes, of which 18.7 ± 2.8 were accepted and displayed, 13 ± 1.8 were reached during the attract state, and just 5.9 ± 3.5 were reached by explicit accept through goto animation. This means that the adaptive recommendation system appeared to generally show appropriate contents with respect to subjects curiosity, and that in many cases this content appeared transparently during the navigation, without the need of additional inputs which could distract users from interaction. In order to compare the adaptive recommendation system with respect to the thumbnail-bar systems, we also measured for all the interfaces the total number of nodes displayed, and the time that subjects employed for observing overlay information (we assume it to be proportional to the interest to the content displayed), the time that they employed for 3D exploration of the scene, and the time that they employed for scrolling the thumbnail-bars. The number of nodes visited for thumbnail-based interfaces was 17.4 ± 2.4 for the interface with the authoring importance based thumbnail (ITI), while it was 18 ± 2.5 for the interface with the ranking based thumbnail (RTI). With respect of measured times, scrolling times were 84.7 ± 25.1 sec. for ITI, and 58.3 ± 15.4 sec. for RTI, while overlay display times were 81.7 ± 24.8 sec. for ITI, 81.8 ± 26 sec. for RTI, and 94.9 ± 14.1 sec. for the adaptive suggestion interface (ASI), and finally 3D exploration times were 132.8 ± 34.7 sec. for ITI, 160.1 ± 27.8 sec. for RTI, and 205.7 ± 13.8 sec. for ASI. It appears evident that, even if with all interfaces subjects were able to visit a similar number of annotated views, when using scroll-based interfaces users employed at least 20% of interaction time for scrolling operations, thus losing the main focus of the 3D scene exploration. Moreover, the reported scrolling time measures are underestimated, since they are based only on touch screen interactions without gaze tracking.

Work-load evaluation. All factors of the NASA task load index questionnaire were individually analyzed in order to find differences between the three proposed interface. The average values of Likert-scores for the factors are presented in Table 1. We noticed an effect with respect to performance ($p < 0.001$ and $F(2,42) = 15.8$), and a slight effect with respect to mental demand ($p = 0.04$ and $F(2,42) = 3.37$). We think, from think-a-loud comments, that a significant part of subjects considered distracting and demanding the scrolling operation on thumbnail bars, especially in the case of the importance based ordering. No significant effects were found with respect to the other factors, namely physical demand, effort, learning time, and frustration, meaning that in general subjects considered all three interfaces easy to learn and use.

	ASI	RTI	ITI
Mental demand	1.8 ± 0.77	2.67 ± 1.35	2.93 ± 1.44
Physical demand	1.93 ± 0.96	2.33 ± 1.11	2.73 ± 1.67
Learning time	2.27 ± 1.33	2.27 ± 1.39	2.87 ± 2.06
Performance	5.87 ± 0.64	5 ± 1	3.67 ± 1.35
Effort	2.8 ± 1.37	2.93 ± 1.22	3 ± 1.77
Frustration	2.53 ± 1.55	2.33 ± 1.34	3.27 ± 2.05

Table 1: *Results of NASA task load index questionnaire.*

Qualitative evaluation. We also gathered useful hints and suggestions from think-a-loud comments made by subjects during the tests. In general, users perceived as appealing the overlays decorating the 3D models, and appreciated the transparent attraction force driving them to interesting views, while giving them the chance to freely explore the 3D scene. On the other side, few subjects considered intrusive the attractive force, while others considered the animation inset distracting with respect to 3D exploration. Finally, most users appreciated the adaptive suggestion system, and we noticed that the non-linear graph lead to a significant variability in node exploration (all subjects carried out different paths and enjoyed different versions of the informative content). We plan to further explore this aspect in future.

8. Conclusions

We have presented a new method and a scalable representation for letting casual users explore, at their own pace, spatially annotated 3D models. Our evaluation shows that the method appears to be well received and intuitive enough for casual users who quickly understand how to browse statue models in a short trial period. The resulting virtual environment, which combines structured information with a simple interface that does not require precise picking, appears to be well suited both for installations at museums and for interaction on mobile devices. We are currently focusing on improving the proof-of-concept prototype, and planning to perform large-scale tests in museum setting. So far, we mostly focused on the recommendation system, in order to provide meaningful navigation. Our future work will concentrate on improving the assisted navigation subsystem, in order to improve guidance towards interesting viewpoints during free navigation. Since the current evaluation focuses mostly on user satisfaction, more work is required to objectively assess the effectiveness of our user interface. Addressing this would require cognitive measures that are beyond the scope of the paper, and are an important avenue for future work. It will be also interesting to evaluate whether the proposed approach, currently tuned to museum applications, can be extended to more complex situation requiring specific tasks to be solved.

Acknowledgments. This work is partially supported by the EU FP7 Program under the DIVA project (290277) and by the HELIOS projet (RAS L7). We thank the personnel of *ArcheoCAOR* for their participation in the design process.

References

[ACB12] ANDUJAR C., CHICA A., BRUNET P.: Cultural heritage: User-interface design for the Ripoll monastery exhibition at the National Art Museum of Catalonia. *Computers & Graphics 36*, 1 (2012), 28–37. 3

[BAMG14] BALSA M., AGUS M., MARTON F., GOBBETTI E.: HuMoRS: Huge models mobile rendering system. In *Proc. ACM Web3D* (2014), pp. 7–16. 3, 5

[BCQ*07] BOLCHINI C., CURINO C. A., QUINTARELLI E., SCHREIBER F. A., TANCA L.: A data-oriented survey of context models. *SIGMOD Rec. 36*, 4 (2007), 19–26. 2

[BDP00] BARRAL P., DORME G., PLEMENOS D.: Visual understanding of a scene by automatic movement of a camera. In *Proc. 3IA* (2000), pp. 3–4. 2

[BJM*14] BETTIO F., JASPE 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 (2014), Article 4. 8

[CGG*04] CIGNONI P., GANOVELLI F., GOBBETTI E., MARTON F., PONCHIO F., SCOPIGNO R.: Adaptive TetraPuzzles: efficient out-of-core construction and visualization of gigantic multiresolution polygonal models. *ACM TOG 23*, 3 (2004), 796–803. 8

[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. 2

[CO09] CHRISTIE M., OLIVIER P.: Camera control in computer graphics: models, techniques and applications. In *ACM SIGGRAPH ASIA Courses* (2009), pp. 3:1–3:197. 2, 3

[DBGB*14] DI BENEDETTO M., GANOVELLI F., BALSA M., JASPE 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. 3, 8

[Dia11] DIAKOPOULOS N.: Design challenges in playable data. In *CHI Workshop on Gamification* (2011). 8

[EM11] ECONOMOU M., MEINTANI E.: Promising beginnings? evaluating museum mobile phone apps. In *Proc. Rethinking Technology in Museums Conference* (2011), pp. 26–27. 1

[FD00] FALK H. J., DIERKING L. D.: *Learning from Museums: Visitor Experience and the Making of Meaning*. Rowman & Littlefield, 2000. 1

[FPMT10] FONI A. E., PAPAGIANNAKIS G., MAGNENAT-THALMANN N.: A taxonomy of visualization strategies for cultural heritage applications. *ACM JOCCH 3*, 1 (2010), 1:1–1:21. 2

[FS97] FARADAY P., SUTCLIFFE A.: Designing effective multimedia presentations. In *Proc. ACM SIGCHI* (1997), pp. 272–278. 2

[GSW*09] GIRGENSOHN A., SHIPMAN F., WILCOX L., TURNER T., COOPER M.: MediaGLOW: organizing photos in a graph-based workspace. In *Proc. ACM IUI* (2009), pp. 419–424. 3

[GVH*07a] 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. 2

[GVH*07b] GÖTZELMANN T., VÁZQUEZ P.-P., HARTMANN K., NÜRNBERGER A., STROTHOTTE T.: Correlating text and images: Concept and evaluation. In *Proc. Smart Graphics* (Berlin, Heidelberg, 2007), pp. 97–109. 2, 3

[Heg11] HEGARTY M.: The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science 3*, 3 (2011), 446–474. 2, 8

[IH12] ISENBERG T., HANCOCK M.: Gestures vs. postures: Gestural touch interaction in 3D environments. In *Proc. 3DCHI* (2012), pp. 53–61. 2

[Int02] INTILLE S. S.: Change blind information display for ubiquitous computing environments. In *Proc. UbiComp*. 2002, pp. 91–106. 7

[JD12] JANKOWSKI J., DECKER S.: A dual-mode user interface for accessing 3D content on the world wide web. In *Proc. WWW* (2012), pp. 1047–1056. 2

[JH13] JANKOWSKI J., HACHET M.: A survey of interaction techniques for interactive 3D environments. In *Eurographics STAR* (2013). 2

[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. 2

[KI13] KEEFE D., ISENBERG T.: Reimagining the scientific visualization interaction paradigm. *Computer 46*, 5 (2013), 51–57. 4

[KSZ*11] KUFLIK T., STOCK O., ZANCANARO M., GORFINKEL A., JBARA S., KATS S., SHEIDIN J., KASHTAN N.: A visitor's guide in an active museum: Presentations, communications, and reflection. *ACM JOCCH 3*, 3 (2011), 11:1–11:25. 1

[LBF*13] LIU Y., BARLOWE S., FENG Y., YANG J., JIANG M.: Evaluating exploratory visualization systems: A user study on how clustering-based visualization systems support information seeking from large document collections. *Information Visualization 12*, 1 (2013), 25–43. 8

[LCWL14] LIU S., CUI W., WU Y., LIU M.: A survey on information visualization: recent advances and challenges. *The Visual Computer* (2014). 2

[Lip80] LIPPMAN A.: Movie-maps: An application of the optical videodisc to computer graphics. *SIGGRAPH 14* (1980), 32–42. 3

[MBB*14] MARTON F., BALSA M., BETTIO F., AGUS M., JASPE A., GOBBETTI E.: IsoCam: Interactive visual exploration of massive cultural heritage models on large projection setups. *ACM JOCCH 7*, 2 (2014), Article 12. 3, 8

[PBN11] POLYS N. F., BOWMAN D. A., NORTH C.: The role of depth and gestalt cues in information-rich virtual environments. *International Journal of Human-Computer Studies 69*, 1-2 (2011), 30–51. 2

[RCC10] RYU D.-S., CHUNG W.-K., CHO H.-G.: PHOTOLAND: a new image layout system using spatio-temporal information in digital photos. In *Proc. ACM SAC* (2010), pp. 1884–1891. 3

[RY06] RIEDL M. O., YOUNG R. M.: From linear story generation to branching story graphs. *IEEE CG & A 26*, 3 (2006), 23–31. 2

[Sch06] SCHOLTZ J.: Beyond usability: Evaluation aspects of visual analytic environments. In *Proc. Visual Anal. Science and Techn.* (2006), pp. 145–150. 8

[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. 2

[SH10] SEGEL E., HEER J.: Narrative visualization: Telling stories with data. *IEEE TVCG 16*, 6 (2010), 1139–1148. 8

[SLF*11] SECORD A., LU J., FINKELSTEIN A., SINGH M., NEALEN A.: Perceptual models of viewpoint preference. *ACM TOG 30*, 5 (2011), 109:1–109:12. 7

[SPT06] SOKOLOV D., PLEMENOS D., TAMINE K.: Methods and data structures for virtual world exploration. *The Visual Computer 22*, 7 (2006), 506–516. 2

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3D. In *SIGGRAPH* (2006), pp. 835–846. 3