Int. J. Human–Computer Studies 180 (2023) 103135

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Increasing visitors attention with introductory portal technology to complex cultural sites

Antonio Origlia [a,b,*], Maria Laura Chiacchio [b], Marco Grazioso [a,b], Francesco Cutugno [a,b]

[a] *Department of Electrical Engineering and Information Technology, University of Naples Federico II, Italy*
[b] *URBAN/ECO Research Center, University of Naples Federico II, Italy*

ARTICLE INFO

Keywords:
Cultural heritage technology
Visitor Employed Photography
Semantically annotated 3D models

ABSTRACT

We investigate the effectiveness of visually impacting interfaces which were located next to the entrance of the San Martino Charterhouse in Naples (Italy), by using high quality 3D reconstructions annotated with semantic information. Semantic data were used to develop a gesture based orienting interface, for the generation of camera movements and pointing gestures for a 3D avatar. Using the Visitor Employed Photography protocol, we collected data about visitors noticing details and discovered that those exposed to the informative systems were able to detect more details than visitors who did not experience it.

1. Introduction

It is generally agreed that the design of technological solutions for Cultural Heritage should take into account several aspects concerning the nature of the museum, the type of visitors–users addressed, if the technological intervention is to be used before, during or after the visit. Also, complex museum environment pose a challenge per se, as for the case of museums that are historical monuments, rich of works of art stratified over time and also hosting collections not necessarily connected to the monument. Many museums already provide introductory content at the beginning of the visiting path to orient the visitors. These typically consist of simple videos with little or no possibility for interaction. In this study we investigate if in a complex museum environment the use of interactive technological installations designed to deliver introductory information before the visit, contributes to improving visitors attention. We consider the specific case of visually impacting technology, intended as all those technological interventions that primarily rely on the visual channel to convey information. Those include techniques like Virtual Reality, Augmented Reality etc. Specifically, we investigate the effectiveness of visually impacting communication devices using semantically annotated 3D models, which will refer to as *portals*, as they are located next to the entrance. In other words, we investigate the possibility of using visually impacting technology before cultural contents are actually present to avoid interfering with them, while enabling visitors to orient themselves independently, potentially increasing the quality of the experience. The main research question concerns the way semantically annotated 3D models impact the museum visit experiences and it can be split in two sub-questions as:

- Q1: How can semantically annotated 3D models be deployed in *portal* museum installations?
- Q2: Is there any different level of attention in people exposed to *portal* installations?

The paper is organised as follows: Section 2 presents an overview of related work; Section 3 lays the theoretical background supporting the design of the technological installations; Section 4 describes the 3D data and their semantic annotations together with the experimental chosen settings; Section 5 presents the evaluation protocol and the details of the experiments involving two different applications designed to provide introductory contents. The first consists of a CAVE installation designed to query semantically annotated 3D models to obtain information about the important items in the Charterhouse. The second consists of a totem kiosk equipped with a 3D avatar designed to present the museum, using spoken texts produced *ad hoc* and semantic annotated 3D models. These are used to inform the automatic system generating both pointing gestures and camera movements, according to the concepts explained by texts. Both experiments use the Visitor Employed Photography (VEP) protocol (Cherem and Driver, 1983) to measure attention changes occurred.

2. Related work

The recently updated (August 2022) definition provided by the International Council of Museums states that “A museum is a not-for-profit, permanent institution in the service of society that researches,

* Corresponding author at: Department of Electrical Engineering and Information Technology, University of Naples Federico II, Italy.
E-mail address: antonio.origlia@unina.it (A. Origlia).

https://doi.org/10.1016/j.ijhcs.2023.103135
Received 10 July 2022; Received in revised form 27 July 2023; Accepted 17 August 2023
Available online 23 August 2023
1071-5819/

collects, conserves, interprets and exhibits tangible and intangible heritage. Open to the public, accessible and inclusive, museums foster diversity, sustainability and citizen curation. They operate and communicate ethically, professionally and with the participation of communities, offering varied experiences for education, enjoyment, reflection and knowledge sharing".. Furthermore, if we refer to the wider idea of heritage (Babelon and Chastel, 2012), the concept of "education, enjoyment, reflection and knowledge sharing" remains fundamental. Ideally, technology should be employed to support cultural environments in achieving the mission stated before.

There is a wide spectrum of technological interventions that have been proposed to support museum visits. The most relevant aspects considered cover personalisation (Ardissono et al., 2012; Kuflik et al., 2015; Michalakis and Caridakis, 2022), virtual guides (Swartout et al., 2010), storytelling (Carrozzino et al., 2018), mixed/natural interaction (Brondi et al., 2016) and Augmented Reality (Zhou et al., 2022). All these research areas show a tendency to concentrate on testing the possibilities offered by new technologies using museums as case studies or test environment. Recent developments in 3D modelling of cultural sites led to represent or reconstruct large environments. This kind of experience has often been proposed, in museum settings, as virtual and augmented reality. The viability of deploying such approaches in cultural heritage settings has been repeatedly demonstrated in a number of cases (Chrysanthi et al., 2012; Kang, 2013; Gonizzi Barsanti et al., 2015). Also, the usage of 3D models has been considered (Apollonio et al., 2018).

We highlighted, in the past (Cera et al., 2018; Origlia et al., 2022) that the annotation of digital models lets scholars associate spatial shapes with the heterogeneous data outlining them through semantic descriptors and the most relevant approach is presented in De Luca (2006). This is based on the geometrical segmentation of architectural digital artefacts which become collections of separate elements, organised using part-whole relationships. More recently, the original methodology has been updated (Messaoudi et al., 2018) and implemented as a cloud-based service called *Aioli*.[1] This knowledge representation approach can be used for multiple applications, among which degradation monitoring (Veron et al., 2014). Another approach to the management of semantic annotations over 3D models is found in Ponchio et al. (2020).

3. Theoretical background

According to the museum studies literature, the definition of museum as a *Communications System* (Cameron, 1968; Davallon, 1999) refers to every kind of museum and includes every aspect: exhibits, their selection, the set-up, the lighting, labels, educational programs, videos, and whatever contributes to communicate a message related to the museum itself. Even considering such an articulate system, some museums show peculiar aspects that make them *complex*, i.e. when beyond the museum function of form and content, they are monuments in themselves. The nature of monument implies, on one side, historical aspects, a notable architecture, with decoration and often even remarkable works of art stratified over times. Usually, all these aspects are kept together by centuries of history that contributed to merge all of them by multiplying and enhancing the sense. From this point of view, *complex museum environments* can be defined as museum-monuments where several aspects (art, architecture, ornaments, history, landscape, etc…) contribute to the *Communication System*, and beside they host collections not necessarily connected to the monument itself and play all the modern function of museums. In other words, the whole environment, in its complexity, plays a role in the process of *creation of sense*.

[1] www.aioli.cloud

Once defined what we intend for complex museum environment, to describe a satisfying cultural experience, we refer to the Self-Determination Theory (Deci and Ryan, 1985; Ryan and Deci, 2000) as a general framework to develop technological museum installations. This states that the personal value assigned by an individual to an experience depends on the effect it has in terms of three basic *mental needs*: (a) Autonomy, a sense of choice and volition in the regulation of behaviour; (b) Competence, the sense of efficacy one has with respect to both internal and external environments; Relatedness: feeling connected to and cared about by others.

Applying the Self-Determination Theory in a museum experience setting, we interpret the role of technology, through interaction, as a provider of non-linear access to cultural information. This satisfies *Autonomy* by offering a kind of experience that is shaped by visitors. Satisfying *Competence* means enabling visitors to independently detect relevant aspects in complex environments, even beyond the specific subject technological experience concentrates on. Technologies designed to help people make sense of a complex environment without *spoon-feeding* have the potential to make visitors feel comfortable in relating themselves with the place they are visiting. Satisfying *Relatedness* means stimulating dialogue and information exchange among visitors.

Furthermore it should be taken into account that different kinds of visitors make sense of the environment in different ways, depending on their background and fundamental motivations for visiting cultural sites. A categorisation of museum visitors is found in Falk and Dierking (2016). In our work, we concentrate on the *Explorers* category, defined as Falk and Dierking (2016, p. 62)"[…] curiosity-driven visitors with a generic interest in the contents of the museum. They expect to find something that will grab their attention and fuel their curiosity and learning". This is one of the most relevant profiles in the categorisation: according to Christidou (2010, p.114), on a sample of visitors to the California Science Center in Los Angeles, the most common identity was the Facilitator with 41%, followed by the Explorer with 34%.

For the specific case of *Explorers*, the role of curiosity in a visit to the museum is fundamental. Interestingly, psychological literature confirms that a number of behaviours shown by curious people are of interest for cultural heritage. When curious, people ask questions (Peters, 1978), manipulate interesting objects (Reeve and Nix, 1997), read deeply (Schiefele, 1999); persist on challenging tasks (Sansone and Smith, 2000); examine interesting images (Silvia, 2005). From a formal point of view, curiosity, according to Loewenstein (1994), reflects the desire to close inherently unpleasant information gaps. This desire, however, depends on the perceived likelihood that the gap will be closed by accessing information (Loewenstein et al., 1992). Moreover, it has also been suggested that "[…] the amount of pre-existing knowledge in a particular domain may impact on the perceived likelihood of closure" (Miceli and Castelfranchi, 2014, p. 57). In this sense, interventions designed to create the conditions for people to autonomously approach and interact with cultural heritage are compatible with museum goals of stimulating reflection.

Summarising, the goal of technology should be to support the museum Communications System in providing cultural experiences satisfying the visitors' basic mental needs. Our view is that communication channels used by artworks should not be contended by technology. Works of art are *designed* to convey the messages intended by the artists. Our approach, rather than directly decoding this message for the viewer while attending at the work of art, aims at providing the necessary decoding keys before actual exposure. We expect that, this way, when only the interpretation mechanisms are provided to the viewer, the actual interpretation remains personal and has the potential to stimulate a feeling of competence. From a technological design point of view, we adopt the principle that cultural environments do not need to be *augmented*: it is the visitors' personal way to access cultural experiences that needs to be *enabled*.

For the case considered in this work, technology is deployed to enable *Explorers* to autonomously decode complex cultural environments

(a) A view of the San Martino Charterhouse exteriors.

(b) A view of the church.

Fig. 1. Different views of the San Martino Charterhouse in Naples. 1(a) shows an aerial view of the site. 1(b) shows one of the richest interior environments.

without overlapping or substituting artworks. Providing a way for people to feel more competent has the potential to stimulate curiosity and increase attention activating a virtuous cycle in which visitors will *spontaneously* look for more information.

4. Materials and methods

4.1. The San Martino Charterhouse

The cultural site considered in this paper is the San Martino Charterhouse in Naples (Italy). The Charterhouse perfectly matches the definition of *complex museum environment* given above. A monumental monastery, built to meet the specific requirements of the carthusian monastic rule, based on the benedictine motto *ora et labora* (pray and work). Built at the beginning of the 14th century but undergone several decoration campaigns, especially from the second half of the 16th century to the first half of the 18th century, this impressive and stunning monastic complex hosted the works of generations of craftsmen and artists from all over Europe. Since the religious function was abandoned, it has been converted into a museum in 1867 because it could in itself well represent the art, craft and architecture in Naples through the centuries. Since then, becoming a sort of museum of the city of Naples, it started to acquire several collections of very different items (nativity scenes, coaches, pottery, paintings, drawings, coats of arms, etc.), all of them connected in different ways to the history of the city, so creating a sort of meta-museum, made of the Charterhouse plus all the other collections. Views of the Charterhouse are shown in Fig. 1.

Given the complexity of the environment, it makes sense to verify the impact of the proposed technologies in this particular site. We developed two different technological installations to be positioned at the beginning of the visiting paths, to provide introductory contents and, thus, act as *portals*. This way, visual contents provided by technology are present only when the contents provided by the museum are not, so that they do not compete on the visual channel. Our hypothesis is that high fidelity reconstructions can be used to *prime* visitors in autonomously recognising important details during the visit.

4.2. Semantically annotated 3D data

The vast amount of textual knowledge concerning cultural heritage can be linked to 3D data to support queries coming both from the users and from automated systems. In this work, we deploy semantic annotations for 3D models in both ways, testing two different systems designed to provide introductory information to the visit.

The 3D models collected for these experiments were obtained using aerial photogrammetry, for the external part, and laser scanning for the internal environments. In order to use them in the two experiments presented here, they were annotated using *semantic maps*, following the method presented in Cera et al. (2018). Each model has a corresponding UV map,[2] consisting of a projection of the vertices in the model on information represented on a 2D surface. In general, rendering engines will use information from multiple textures (e.g. colorimetric, normal, ambient occlusion) to produce the final output. The same mechanism can be used to store semantic information. Our approach uses UV mapping to associate a greyscale texture to the models representing, for each vertex, a relevance level for a specific semantic label. Specifically, the scale goes from black, indicating, *not relevant*, to white, being *totally relevant*. Greyscale values can either be obtained by averaging the annotations of multiple experts, as in the reference work, or to represent concepts *blending* into one another when clear boundaries cannot be found, even by an expert architect. In Figs. 2 and 3 an example colour map and a semantic map of the *painting* concept are shown.

Fig. 2. Colour map relative to a painting found on a 3D model. This is the map used by rendering engines to assign colour to the polygons of the represented 3D models. The position of the colour information is generated according to the UV map associated to the model.

[2] UV maps refer to the coordinate space of an image texture, where the point of origin is located in the upper left corner and the axes are conventionally named U and V

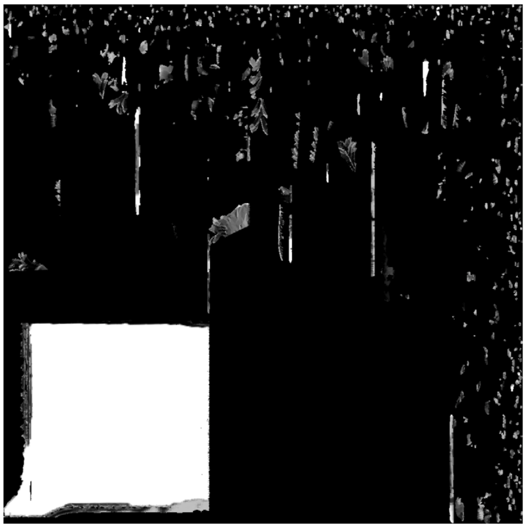

Fig. 3. Semantic map of the *painting* concept found on a 3D model. White areas show which polygons are relevant for the considered concept. The position of the semantic information is generated according to the UV map associated to the model.

In the presented experiments, two different 3D models were considered to design the applications and evaluate their effect on the visitors. Specific details were annotated by an expert architect using a texture painting tool to obtain the semantic maps. The models were used, in one case, to design a model query system and, in the other case, to support the generation of camera movements and pointing gestures for a 3D avatar. The specific details of the considered models and the corresponding semantic annotations are reported in the sections dedicated to each installation, together with the way they were used in the presented applications.

Semantic annotations, in general, have the potential to support the development of interactive technologies based on 3D contents. With the recently increasing presence of Real Time Interactive 3D approaches, this technology becomes relevant to support, for example, question answering systems, interactive totem kiosks and innovative query systems. The tests presented in this work are also intended to validate this technology so it can be integrated in future installations based, for example, on embodied conversational agents.

4.3. CAVE-based installation

The first experiment was conducted using gesture-based interaction with a *portal* CAVE environment positioned at the beginning of the visiting path, in the museum, just after the tickets booth. The system was designed to act as an orienting interface, providing a better understanding of the Charterhouse size and organisation.

The whole setup of the interaction environment created an immersive experience for the users and it consisted of a curved screen 2,5 m high and 4,4 m long. A semantically annotated 3D model of the San Martino Charterhouse was slowly rotating on the screen and users movements were tracked in real time using a Microsoft Kinect 2 (Zhang, 2012). By pointing towards the screen, users were able to move a pointer over the 3D model. As soon as a semantically annotated area, marked by particle emitters, was reached, the model stopped rotating and the name of the area was shown. By keeping the pointer on the semantically annotated area for at least 3 s, users were able to start a video presentation about the pointed area. When more users were together in front of the screen, the system would track the one closer to the centre. Fig. 4 shows the deployed setting. Five different interactive areas were labelled using semantic maps: the great cloister, the procurators cloister, the church, the women's church and the pharmacy. The application was implemented using the Unreal Engine 4.[3] The narrated text for the video was written by an expert art historian to highlight the following targets: monks' cells, cemetery, great cloister well, church vault, church floor, main altar, pharmacy fresco, procurators cloister well, cells' service hatches, arches on columns, angular doors, side chapel flowers, papier-mache angels, sound box grate, mercy, procurators cloister well masks, arches on pillars, coats of arms.

4.4. 3D avatar installation

In the San Martino Charterhouse, the church and the environments immediately connected to it are of utmost importance from the artistic value point of view. They are rich in decorations and full of details that can easily overwhelm visitors so that particularly relevant works of art may be lost in the plethora of visual stimuli that the baroque decorations present. The second *portal* explicitly covers this sub-section of the visiting path, to concentrate on an environment that is particularly difficult to decode by visitors. In this experiment, the interior environments (example shown in Fig. 5) were semantically annotated to guide both non-verbal behaviour in 3D avatars and camera movements. The goal, in this case, is to support people in recognising important elements in a potentially confusing setting using pointing gestures from the avatar and camera movements to anticipate navigation in the church environment.

Inside the 3D reconstructed internal environments, five virtual points of interest (POIs) were identified, one for each environment, thus covering the parlour, the capitol room, the choir, the sacristy and the treasure chapel. For each of these environments, an expert art historian produced illustrative texts that contained, as in the previous experiment, a set of details that were considered relatively hard to spot given the richness of the environment. Semantic maps were produced for the 3D model to label the areas corresponding to the items named in the text. To support pointing gestures and camera movements, texts were produced using the Speech Synthesis Markup Language (SSML). Labelled items were marked accordingly to the SSML syntax so that a speech synthesiser would be able to provide the time offset at which each labelled item in the text was actually pronounced by the synthetic voice. The 3D avatar speech and animation was managed inside the Unreal Engine 4 using the FANTASIA plugin (Origlia et al., 2019, 2022), which is freely available on Github.[4] When labelled items are pronounced by the avatar, the system queries the environment for the centroid of vertex points that are relevant for the pronounced concept. A pointing gesture and a camera movement are, then, dynamically generated to accompany the presentation from the avatar.

The user interface, for this installation, consisted of a touch interface that allowed visitors to navigate the environment from one POI to the other. Visitors could, therefore, navigate the virtual reconstruction by moving from one presentation point to the following one, given the environments sequence in the church, and listen to the presentations accompanied by the automatically generated pointing gestures from the avatar and camera movements. Camera movements also accompanied transitions among the POIs to anticipate the visiting path to the visitors, so that they could more easily identify items of interest during the actual path. Fig. 6 shows the touch based interface deployed on a totem kiosk structure.

5. Evaluation

In this Section, we describe how we applied the Visitor Employed Photography protocol, using images taken by visitors to the site of interest, to estimate their capability of detecting *important* details, from an art history point of view.

[3] www.unrealengine.com

[4] https://github.com/antori82/FANTASIA

Fig. 4. The CAVE installation at the San Martino Charterhouse.

Fig. 5. An example of the 3D reconstruction of the internal part of the San Martino Charterhouse.

5.1. The visitor employed photography protocol

In this work, we propose the use of VEP as a way to investigate if visitors were able to detect and recognise, as important, specific details in a complex museum environment. Textual investigation methods indeed retain their value especially concerning quality of learning, which is a common goal in the field of technologies for cultural heritage. In our case, to collect data about the visitors' attention to important details is the main interest: taking a picture is interpreted as a testimony of having *noticed* something and declaring its importance from a personal point of view.

Our focus is to measure how visually impacting *portals* influence the way visitors perceive the Charterhouse. In particular, we measure how much the developed applications supported people, in the following visit, to autonomously recognise characteristics that would otherwise be missed. Both experiments involved visitors to the San Martino Charterhouse during separate exhibits, each lasting 15 days, and they were divided in two parts:

- Participants were briefly instructed on how to use the active installation and they were left free to use them for as long as they wanted;
- The possibility of participating to the second part of the experiment was offered to the visitors. The experimenters provided a digital camera to the participants who accepted the offer and instructed them to take pictures, during their visit, *as if it was their own camera*.

The second part of the experimental procedure implements the VEP protocol. Using pictures produced by participants as evaluation data has been questioned, in the past, as a research method. This was because of potential difficulties in interpretation and because of the impact of participants' subjectivity on the data with respect to *normalised* approaches like questionnaires and interviews. Modern views on the topic, however, reclaim the value of photography as a research method because of its characteristic to "[...] provide tourism researchers with a different kind of information that is able to embrace the embodiment

Fig. 6. The 3D avatar touch-based interface.

of experiences" (Bell and Davison, 2013). The significance attributed to pictures and competence about photography have also changed substantially because of multiple factors: the possibility to immediately check pictures and retake them to obtain the desired effect; the practical absence of limits in the number of pictures to take; the availability of devices allowing picture taking; the influence of social media among others. These changes both reduce the perceived *cost* of taking pictures and increase the value of pictures as research data. A foundational work strongly advocating for the use of photography as data in tourism research is Urry and Larsen (2011). While there is still an open discussion about its use in the social sciences community, the VEP protocol has been repeatedly used to conduct research about landscape and urban landscape scenarios (Traweek, 1977; Chenoweth and Niemann, 1981; Cherem and Driver, 1983) and, more recently, in Balomenou and Garrod (2014), Prestholdt and Nordbø (2015), Hansen (2016). In Matteucci (2013), the authors explicitly states that the VEP has much to offer to researchers in tourism, while Balomenou et al. (2017) underlines that they remain underused as a research tool. A complete overview on this topic is found in Balomenou and Garrod (2019).

Before using VEP, people who asked for more information were explained that it could be provided only after the visit in order to avoid biasing. All participating groups agreed with this and were informed about the goals of the experiment after they brought back the camera. No personal data were collected and the camera's memory was erased after downloading the pictures on a PC to avoid influencing other visitors. At each time, only one group participating to the second part of the experiment was active to avoid meeting during the visits and influence each other, as the Charterhouse does not have a single visiting path and can be visited in a non-linear way.

At the end of the 15 days in which the installations were active (on selected days in that time interval), 19 groups per installation, were recruited for the VEP and represent the experimental groups. For each experiment, moreover, another 19 groups who were not exposed to the installation were recruited to represent the control group. The difference between the experimental and control groups lies in the introductory experience they had with the proposed installations. Our goal is to investigate if there is an increased level of attention, in visitors exposed to the *portal* installations towards details that are typically hard to notice/isolate from the context. In all cases, people were free to use any additional material they had, like paper guides. Due to the architectural characteristics of the Charterhouse, moreover, Internet access was slow if not at all available during the visit.

For the analysis of the collected data, we concentrated on pictures taken in the environments covered by the active installations, checking if the details included in the provided contents were noticed by the visitors. Three of the authors independently annotated the pictures taken by the experimental and control groups, indicating whether a picture represented one of the target details. Also, three independent judges were provided with reference pictures of the targets to perform the same task. In general, subjects centrality and frame occupation were considered during the labelling phase but judges were free to express their own opinion about the considered targets being a relevant part of the image. As an example, the image shown in Fig. 7 clearly has the main altar as primary subject and was labelled as such. More general images of the church containing the main altar were not assigned to the target. Also, the analysis concentrated on whether a target was detected or not: a target was counted only once per group even if there were multiple pictures representing it.

For each participating group, a target was considered as noticed (hit) if at least four judges identified a picture from the considered group as representative of the target. For each experiment, we consider, at different levels, the probability of both pictures from the experimental group and from the control group to hit multiple targets. Specifically, the target hit probability was computed for each number of detected targets (level) so that, for each group, the probability of people belonging to it to detect at least a threshold value of i targets was computed as

$$P(t_i) = \frac{N_t}{N} \tag{1}$$

where N_t is the number of groups that detected at least i targets and N is the total number of groups. Expectations are that people from the experimental group have a higher probability of detecting multiple targets. We also report the pictures distributions over the targets to provide further details about the groups *focusing* behaviour.

The two installations covered different aspects of the Charterhouse so the target items are different between the two experiments. This avoids the risk posed by using one set of target items, which may contain biases, and allows us to investigate situations with different densities in visual stimuli, as explained in the next Section.

5.1.1. CAVE installation — results

After applying the annotation procedure to the collected pictures, as previously explained, the obtained target hit probability levels for the two groups were compared to check if there was a difference in the attention shown by the two groups. The Shapiro test confirmed the normality of the distributions so a paired t-test was used to check if the number of groups exposed to the system that detected targets at each threshold was different, on average, than the ones in the control group. The test indicated that the difference was significant ($p < 0.01$), so we can conclude that visitors in the experimental group have a higher chance to detect multiple target details than the ones in the control group at each threshold level. An overview of the number of groups hitting at least i targets at different thresholds, is shown in Fig. 8.

Concerning pictures distributions over the considered targets, a paired t-test over the pictures distribution over the considered targets showed a statistically significant difference ($p < 0.01$). From this, we conclude that people from the experimental group also had an advantage in detecting targets that were not detected from the control group. A more detailed view of the pictures distributions over the targets is shown in Fig. 9. In most of the cases, a higher number of visitors from the experimental group detected the selected targets. In those cases where more visitors from the control group detected a specific target, we note that the difference is negligible and that the targets are among the most easily detectable ones because of position and size. On the contrary, in cases where more visitors in the experimental group detected a specific target, the difference is clear in most of the cases.

Summarising, the CAVE installation appeared to influence the capability of the visitors to detect more details, during their tour.

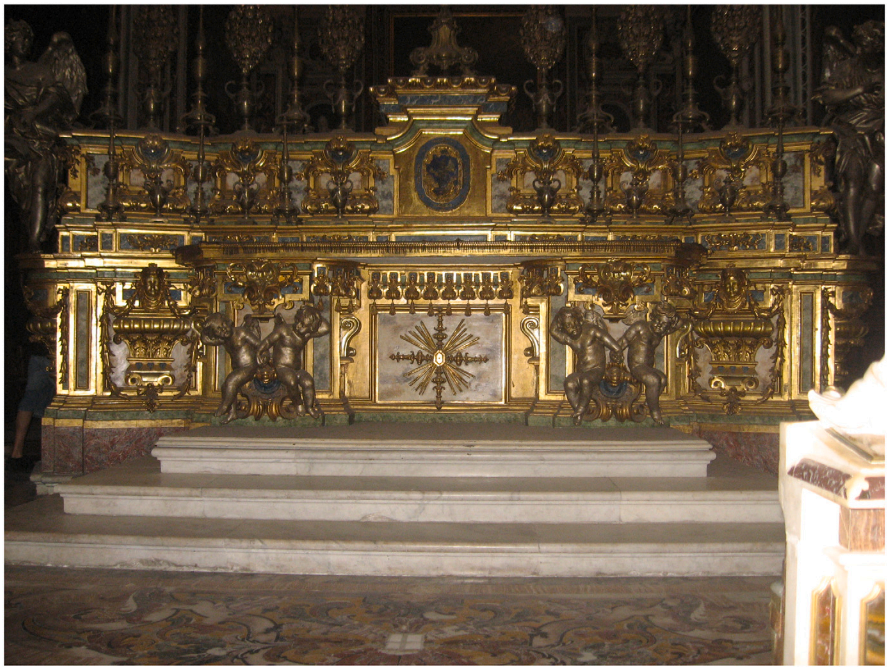

Fig. 7. A picture of the main altar in the San Martino church.

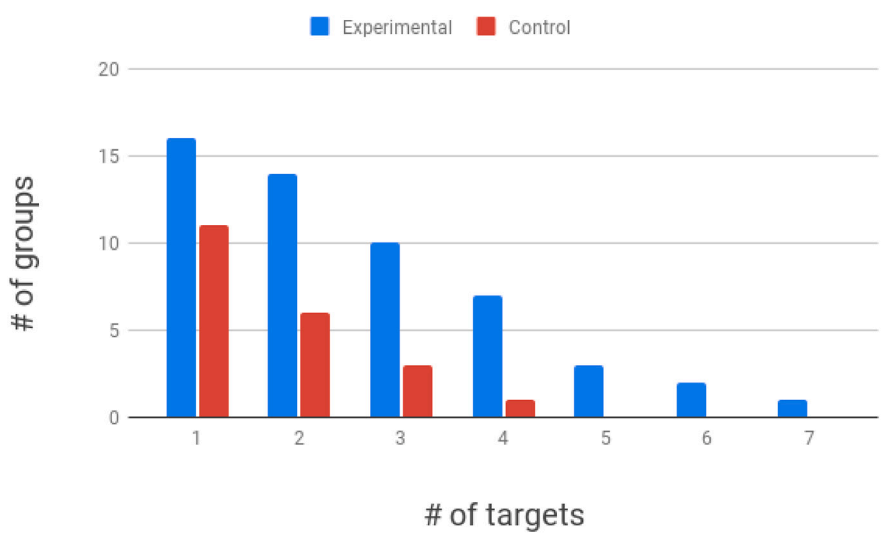

Fig. 8. Targets detected by the groups included in the experimental and control groups for the CAVE installation.

5.2. Avatar installation — results

The second experiment considers a situation where important details may be lost due to the complexity, in terms of richness in visual stimuli. The same analysis performed for the previous experiment was applied to this one. The Shapiro test confirmed the normality of the distributions so a paired t-test was used to check if the number of groups exposed to the system that detected targets at each threshold was different, on average, than the ones in the control group. The test indicated that the difference was significant ($p < 0.01$), so we can conclude that visitors in the experimental group have a higher chance to detect target details than the control group at each threshold level. An overview of the number of groups hitting at least i targets at different thresholds, is shown in Fig. 10.

Concerning pictures distributions over the considered targets, a paired t-test over the pictures distribution over the considered targets found a statistically significant difference between the two groups ($p < 0.01$). From this, we conclude that people by the experimental group were able to detect targets that were not detected by the control group, thus being able to focus their attention on important details that would otherwise be lost. A detailed view of the pictures distributions over the targets is shown in Fig. 11. For this distribution, the same observations made for the first experiment hold.

6. Conclusions

We have presented an investigation on visually impacting technologies in the case of complex museum environments relying on the visual communication channel to provide cultural contents. Our design approach deploys the technological intervention at the beginning of the visiting path, as a *portal*, to avoid overlapping with works of art and to enable visitors in moving more confidently in a complex environment.

From the technological point of view, we used semantically annotated 3D models to investigate the use of a data representation

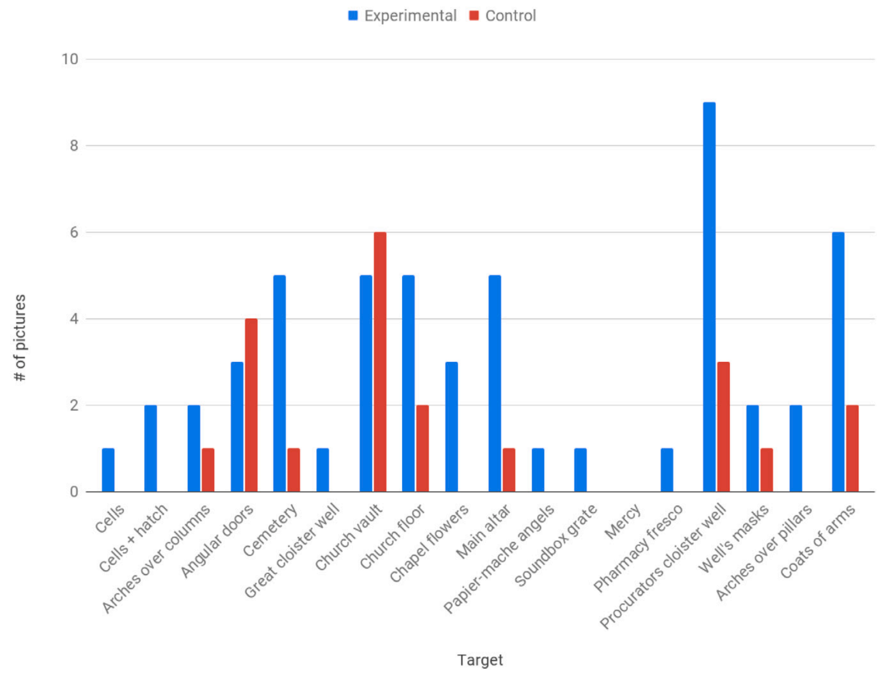

Fig. 9. Pictures distribution over the targets considered in the CAVE experiment.

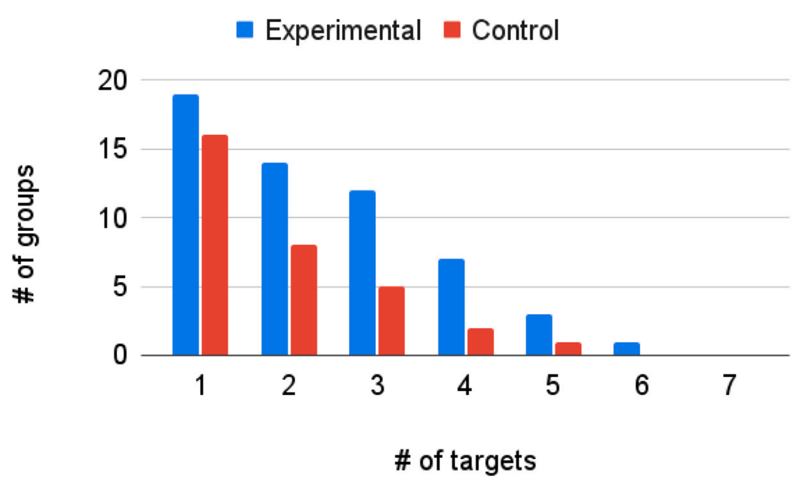

Fig. 10. Targets detected by the groups included in the experimental and control groups for the avatar installation.

technique to develop interactive applications. Our experiments motivate the research effort in this direction, showing the application of semantic 3D annotations in this setting. Also, we have shown that the example applications were able to influence the attentive behaviour of the attending visitors. People who were exposed to the installations showed a higher capability of autonomously detect relevant details among the multiple visual stimuli offered by the case of the San Martino Charterhouse. To verify this, we proposed the use of the VEP technique, which is usually adopted for landscape investigations. This has proven useful to evaluate what people noticed and considered important without relying on more invasive methods that could interfere with the visit. The increased attention observed in the Experimental groups is indeed more *competent* than the one observed in the Control groups for both experiments. We cannot, however, confirm that participating people *felt* more competent as the setting we chose, prioritising ecology, did not allow us to ask people to compile questionnaires. Given our data, no difference appears to be present in the effect provided by the two installations. However, user experience may substantially differ so we will investigate this aspect more deeply in the future.

The choice to prefer a more ecological setting is motivated by the observation that technological interventions in museum environment can be considered well-established, so we decided to prioritise experimentation with *actual* users, who independently chose to visit the museum and had no previous idea they would have been presented with the technologies of interest for this work. The drawback, of course, is that it was not possible to collect personal data, as involved subjects were investing their free time in the visit and could not be enrolled in a full study. Also the agreement with the hosting venue specified that the experimental protocol should be as less imposing as possible for the visitors. The high ecological value of our data, therefore, implies less control over the participants background. Deeper investigations about the visitors personal experience will be the subject of future, targeted experiments.

The validation of the technology using semantically annotated 3D models and the observed effect on visitors motivate the development of more advanced applications, in the future, following the same approach. These will make use of the semantically annotated models to guide interaction with virtual agents and will be designed to prepare

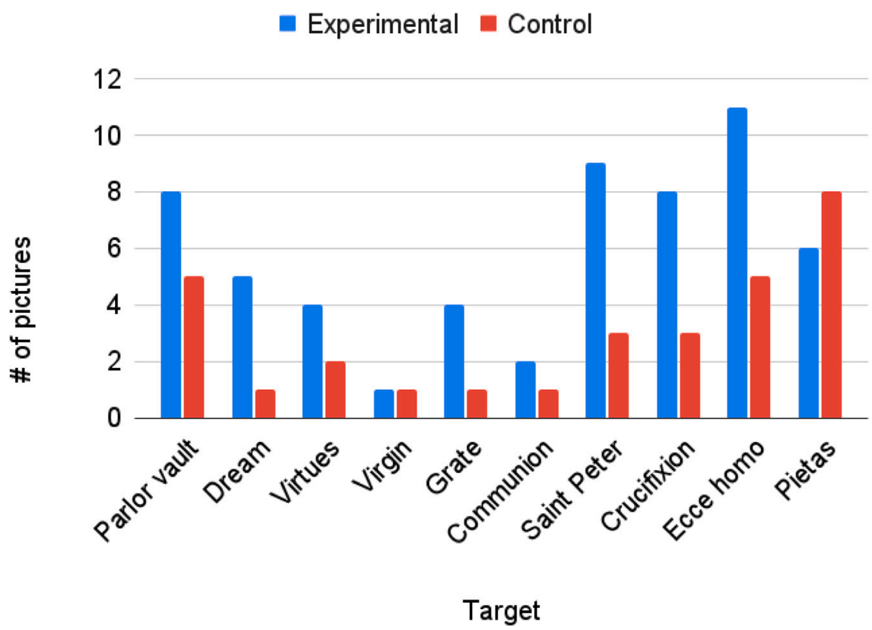

Fig. 11. Pictures distribution over the targets considered in the avatar experiment.

visitors in autonomously orient themselves and make sense of complex museum environments. As the VEP protocol has provided interesting insight in documenting the way people relate themselves with the cultural site, it will be adopted again in future test.

In the current version of the avatar installation, texts are static and manually labelled, but the system also supports dynamically generated texts and labels using, for example, entity linking approaches. The validated technology for movements generation based on semantic annotations will be integrated in future applications involving dialogue systems to further exploit the possibility to deal with dynamically generated content, as per in the case of questions answering systems.

CRediT authorship contribution statement

Antonio Origlia: Conceptualization, Methodology, Software, Formal analysis, Investigation, Funding acquisition, Writing – original draft, Writing – review & editing, Visualization. **Maria Laura Chiacchio:** Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Marco Grazioso:** Software, Investigation, Data Curation, Writing – original draft. **Francesco Cutugno:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code of the FANTASIA plugin is freely available on Github. We may provide the Unreal project although they run on an outdated version of the engine.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijhcs.2023.103135.

References

Apollonio, F.I., Basilissi, V., Callieri, M., Dellepiane, M., Gaiani, M., Ponchio, F., Rizzo, F., Rubino, A.R., Scopigno, R., et al., 2018. A 3D-centered information system for the documentation of a complex restoration intervention. J. Cult. Herit. 29, 89–99.
Ardissono, L., Kuflik, T., Petrelli, D., 2012. Personalization in cultural heritage: the road travelled and the one ahead. User Model. User-Adapt. Interact. 22, 73–99.
Babelon, J.-P., Chastel, A., 2012. La notion de patrimoine. Liana Levi.
Balomenou, N., Garrod, B., 2014. Using volunteer-employed photography to inform tourism planning decisions: A study of St David's Peninsula, Wales. Tour. Manag. 44, 126–139.
Balomenou, N., Garrod, B., 2019. Photographs in tourism research: Prejudice, power, performance and participant-generated images. Tour. Manag. 70, 201–217.
Balomenou, N., Garrod, B., Georgiadou, A., 2017. Making sense of tourists' photographs using canonical variate analysis. Tour. Manag. 61, 173–179.
Bell, E., Davison, J., 2013. Visual management studies: Empirical and theoretical approaches. Int. J. Manag. Rev. 15 (2), 167–184.
Brondi, R., Carrozzino, M., Lorenzini, C., Tecchia, F., 2016. Using mixed reality and natural interaction in cultural heritage applications. Informatica 40 (3).
Cameron, D.F., 1968. A viewpoint: The museum as a communications system and implications for museum education 1. Curator Museum J. 11 (1), 33–40.
Carrozzino, M., Colombo, M., Tecchia, F., Evangelista, C., Bergamasco, M., 2018. Comparing different storytelling approaches for virtual guides in digital immersive museums. In: International Conference on Augmented Reality, Virtual Reality and Computer Graphics. Springer, pp. 292–302.
Cera, V., Origlia, A., Cutugno, F., Campi, M., 2018. Semantically annotated 3D material supporting the design of natural user interfaces for architectural heritage. In: Proc of the AVI-CH Workshop.
Chenoweth, R., Niemann, B., 1981. Scenic Beauty of the Alpine Lakes Wilderness Area: A Scenic Assessment Survey of On-Site Users. Landmark Research Inc, Madison.
Cherem, G.J., Driver, B., 1983. Visitor employed photography: A technique to measure common perceptions of natural environments. J. Leis. Res. 15 (1), 65–83.
Christidou, D., 2010. Re-introducing visitors: Thoughts and discussion on John Falk's notion of visitors' identity-related visit motivations. Pap. Inst. Archaeol. 20, 111–122.
Chrysanthi, A., Papadopoulos, C., Frankland, T., Earl, G., 2012. 'Tangible Pasts': User-centred design of a mixed reality application for cultural heritage. Archaeol. Digit. Era 31.
Davallon, J., 1999. L'exposition à l'oeuvre: stratégies de communication et médiation symbolique. L'harmattan.
De Luca, L., 2006. Relevé et multi-représentations du patrimoine architectural Définition d'une approche hybride pour la reconstruction 3D d'édifices (Ph.D. thesis). Sciences de l'Homme et Société. Arts et Métiers ParisTech.
Deci, E.L., Ryan, R.M., 1985. The general causality orientations scale: Self-determination in personality. J. Res. Pers. 19 (2), 109–134.
Falk, J.H., Dierking, L.D., 2016. The Museum Experience Revisited. Routledge.

Gonizzi Barsanti, S., Caruso, G., Micoli, L., Covarrubias Rodriguez, M., Guidi, G., et al., 2015. 3D visualization of cultural heritage artefacts with virtual reality devices. In: 25th International CIPA Symposium 2015, Vol. 40, No. 5W7. Copernicus Gesellschaft mbH, pp. 165–172.

Hansen, A.S., 2016. Testing visitor produced pictures as a management strategy to study visitor experience qualities–A Swedish marine case study. J. Outdoor Recreat. Tour. 14, 52–64.

Kang, J., 2013. AR teleport: digital reconstruction of historical and cultural-heritage sites for mobile phones via movement-based interactions. Wirel. Pers. Commun. 70 (4), 1443–1462.

Kuflik, T., Wecker, A.J., Lanir, J., Stock, O., 2015. An integrative framework for extending the boundaries of the museum visit experience: linking the pre, during and post visit phases. Inf. Technol. Tour. 15 (1), 17–47.

Loewenstein, G., 1994. The psychology of curiosity: A review and reinterpretation. Psychol. Bull. 116 (1), 75.

Loewenstein, G., Adler, D., Behrens, D., Gillis, J., 1992. Why Pandora Opened the Box: Curiosity as a Desire for Missing Information. unpublished manuscript, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA.

Matteucci, X., 2013. Photo elicitation: Exploring tourist experiences with researcher-found images. Tour. Manag. 35, 190–197.

Messaoudi, T., Véron, P., Halin, G., De Luca, L., 2018. An ontological model for the reality-based 3D annotation of heritage building conservation state. J. Cult. Herit. 29, 100–112.

Miceli, M., Castelfranchi, C., 2014. Expectancy and Emotion. OUP Oxford.

Michalakis, K., Caridakis, G., 2022. Context awareness in cultural heritage applications: a survey. ACM J. Comput. Cult. Herit. (JOCCH) 15 (2), 1–31.

Origlia, A., Cutugno, F., Rodà, A., Cosi, P., Zmarich, C., 2019. FANTASIA: a framework for advanced natural tools and applications in social, interactive approaches. Multimedia Tools Appl. 78 (10), 13613–13648.

Origlia, A., Di Bratto, M., Di Maro, M., Mennella, S., 2022. Developing embodied conversational agents in the unreal engine: The FANTASIA plugin. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6950–6951.

Peters, R.A., 1978. Effects of anxiety, curiosity, and perceived instructor threat on student verbal behavior in the college classroom. J. Educ. Psychol. 70 (3), 388.

Ponchio, F., Callieri, M., Dellepiane, M., Scopigno, R., 2020. Effective annotations over 3D models. In: Computer Graphics Forum, Vol. 39, No. 1. Wiley Online Library, pp. 89–105.

Prestholdt, R., Nordbø, I., 2015. Norwegian landscapes: An assessment of the aesthetical visual dimensions of some rural destinations in Norway. Scand. J. Hosp. Tour. 15 (1–2), 202–222.

Reeve, J., Nix, G., 1997. Expressing intrinsic motivation through acts of exploration and facial displays of interest. Motiv. Emot. 21 (3), 237–250.

Ryan, R.M., Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am. Psychol. 55 (1), 68.

Sansone, C., Smith, J., 2000. In: Sansone, J. (Ed.), Interest and Self-Regulation: The Relation Between Having to and Wanting to. Intrinsic and Extrinsic Motivation: The Search for Optimal Motivation and Performance.

Schiefele, U., 1999. Interest and learning from text. Sci. Stud. Read. 3 (3), 257–279.

Silvia, P.J., 2005. What is interesting? Exploring the appraisal structure of interest. Emotion 5 (1), 89.

Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., Piepol, D., et al., 2010. Ada and grace: Toward realistic and engaging virtual museum guides. In: International Conference on Intelligent Virtual Agents. Springer, pp. 286–300.

Traweek, D.E., 1977. Visitor Employed Photography on the Huron River: A Tool for Interpretative Planning. The Ohio State University.

Urry, J., Larsen, J., 2011. The Tourist Gaze 3.0. Sage.

Veron, P., Messaoudi, T., Manuel, A., Gattet, E., De Luca, L., 2014. Laying the foundations for an information system dedicated to heritage building degradation monitoring based on the 2D/3D semantic annotation of photographs. In: Proc. of the Eurographics Workshop on Graphics and Cultural Heritage.

Zhang, Z., 2012. Microsoft kinect sensor and its effect. IEEE Multimedia 19 (2), 4–10.

Zhou, Y., Chen, J., Wang, M., 2022. A meta-analytic review on incorporating virtual and augmented reality in museum learning. Educ. Res. Rev. 100454.