3D Avatars and Semantic Models Annotations for Introductory Cultural Heritage Presentations

Antonio Origlia[1], Marco Grazioso[1], Maria Laura Chiacchio[1] and Francesco Cutugno[1]

[1]*University of Naples "Federico II", Naples, Italy*

Abstract

In this paper, we investigate the effectiveness of visually impacting interfaces located at the beginning of visiting paths at the San Martino Charterhouse in Naples (Italy), using high quality 3D reconstructions annotated with semantic information. Semantic data were used to develop an application generating camera movements and pointing gestures for a 3D avatar to accompany introductory contents. Observed behaviours, collected using the Visitor Employed Photography protocol, of visitors exposed to the informative systems show that the groups who interacted with the installations were able to detect more details than visitors who did not experience it.

Keywords

cultural heritage, visitor employed photography, 3D semantic annotation

1. Introduction

The design of technological solutions for Cultural Heritage should take into account several aspects concerning the nature of the museum, the kind of visitors-users who it is addressed to, if the technological intervention is supposed to be used before, during or after the visit. In this study we investigate if in a complex museum environment, where exhibits are not *atomised* (i.e. organised in a series of clearly recognisable exhibits), the use of technological installations at the beginning of the visiting path and designed to deliver introductory information, which we will refer to as *portals*, contributes in enhancing visitors experience. Specifically, we investigate the effectiveness of visually impacting communication devices located at the beginning of a visiting path in a complex museum environment using semantically annotated 3D models. The main research question posed in this work investigates whether behavioural changes can be observed in people who were exposed to this kind of *portal* installations before the visit. This paper presents one part of a larger experiment described in [1].

There is a wide spectrum of technological interventions that have been proposed to support museum visits. The most relevant aspects that have been considered cover personalisation [2], virtual guides [3], storytelling [4] and mixed/natural interaction [5]. 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. While this is, obviously, important to advance knowledge in the technological field, considering the need museums have to *push*

AVI-CH 2022 Workshop on Advanced Visual Interfaces and Interactions in Cultural Heritage. June 06, 2022. Rome, Italy

antonio.origlia@unina.it (A. Origlia); marco.grazioso@unina.it (M. Grazioso); marialaura.chiacchio@gmail.com (M. Chiacchio); cutugno@unina.it (F. Cutugno)

people towards deeper reflection on cultural contents, it is also important in order to identify how technology can support museums on different aspects. The profound sense in visiting cultural sites lies in accessing information feeding mental needs [6] to support personal growth.

Recent developments in 3D modelling of cultural environments have led to the possibility to represent or reconstruct large environments. This kind of experience has been proposed, in museum settings, in the form of virtual and augmented reality. The viability of deploying such approaches in cultural heritage settings has been repeatedly demonstrated in a number of cases [7, 8, 9]. Detailed 3D reconstructions of architectural heritage are of interest in this work. We concentrate on exploring how technological installations developed on the basis of semantic annotation approaches can be used in a museum setting to improve the way in which visitors autonomously access complex museum environments [10]. The annotation of digital models lets scholars associate spatial shapes with the heterogeneous data describing them through the use of semantic descriptors. The most relevant approach to this kind of semantic annotation is presented in [11] and it is based on the geometrical segmentation of architectural digital artefacts. More recently, the original methodology has been updated [12] and implemented as a cloud-based service called *Aioli*[1]. This kind of knowledge representation approach can be used for multiple applications, among which degradation monitoring [13]. In this work, we explore its use to support interactive applications, possibly integrated with Artificial Intelligence.

The paper is organised as follows: Section 2 presents the 3D data and their semantic annotations together with the evaluation protocol adopted to test the impact of the installations on the visit; Section 4 presents two experiments deploying semantically annotated 3D models in a technological application designed to provide introductory presentations.

2. Materials and methods

In this Section, we summarise how 3D data were collected and semantically enriched to support the development of the applications described in Section 4. We also describe here the experimental procedure adopted to investigate the research questions.

2.1. Semantically annotated 3D models

Visually impacting technology can significantly benefit from semantic annotations. In particular, 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 testing a system designed to provide introductory information to the visit.

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*. The 3D model collected for this experiment was obtained using laser scanning: an example of the result is shown in Figure 1. In order to use it in the experiment presented here, it was annotated using *semantic maps*, following the method

[1] www.aioli.cloud/

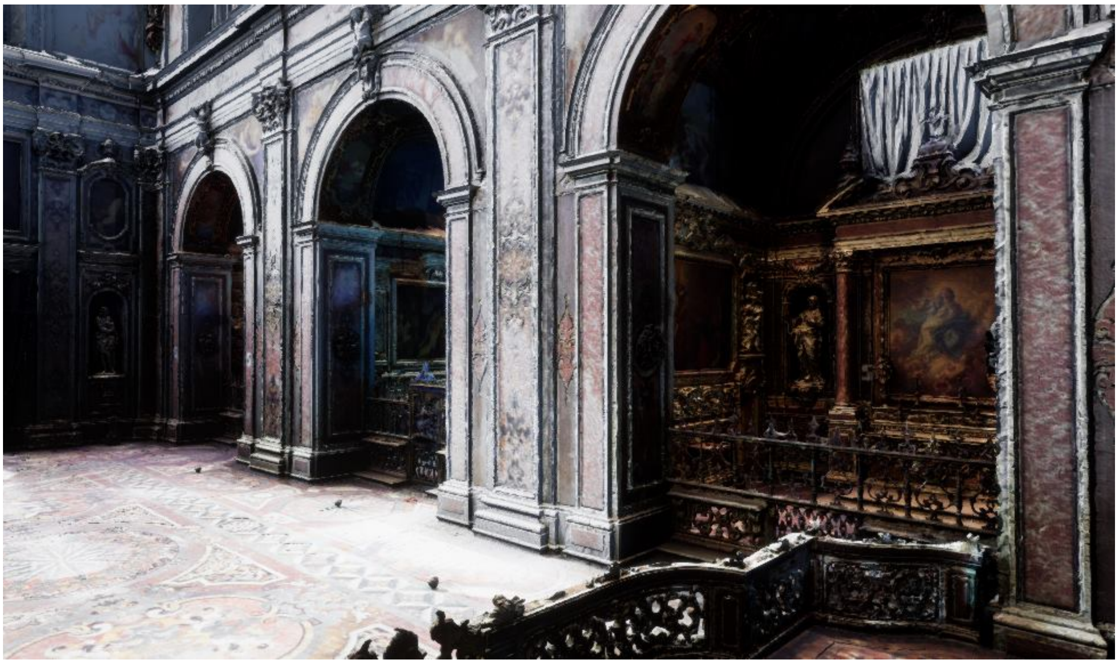

Figure 1: An example of the 3D reconstruction of the internal part of the San Martino Charterhouse.

presented in [14]. This 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.

2.2. Experimental procedure

The introductory contents provided by the installation covered specific aspects of the Charterhouse, both architectural and decorative, and were selected by an expert art historian. Being a baroque cultural site, the Charterhouse is very rich in visual stimuli. Not all of them, however, have the same importance so it is not always easy, for the visitors, to separate important details from the general view of the majestic environments. In this paper, our focus is to measure how visually impacting *portals* influence the way visitors perceive the Charterhouse. In particular, we measure how much the developed application supported people, in the following visit, to autonomously recognise characteristics that would otherwise be missed. The experiment involved visitors to the San Martino Charterhouse during separate exhibits, each lasting 15 days, and it was divided in two parts:

- Participants were briefly instructed on how to use the 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 Visitor Employed Photography (VEP) protocol [15]. 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 of experiences" [16]. 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 complete overview on this topic is found in [17]. The VEP technique has been repeatedly used in studies concerning landscape in urban landscape studies. 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. While textual investigation methods indeed retain their value especially to evaluate quality of learning, which is a common goal in the field of technologies for cultural heritage, in our case collecting data about the visitors' experience 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.

Before the 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 (samples) 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 samples. At each time, only one sample participating to the second part of the experiment was active. This is because the Charterhouse does not have a single visiting path and can be visited in a non-linear way so, to avoid different samples meeting during the visits and influence each other, the offer to participate to the VEP was presented only if there was not another sample already participating.

At the end of the 15 days in which the installation was active, 19 samples were recruited for the VEP experiment and represent the experimental group. Other 19 samples who were not exposed to the installation were recruited to perform comparisons and represent the control groups. In all cases, people were free to use any additional material they had, like paper guides, but only a very limited number of samples had one. Due to the architectural characteristics of the Charterhouse, moreover, Internet access was slow if not at all available during the visit. Although the museum does provide audioguides, none of the participating samples had one.

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. In general, subjects centrality and frame occupation were considered during the labelling phase. The analysis concentrated on whether a target was detected or not: a target was counted only once per sample even if there were multiple pictures representing it.

For each participating sample, a target was considered as noticed (hit) if at least two judges identified a picture from the considered group as representative of the target. For each experiment, we report, at different levels, the probability of both samples 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 a sample belonging to it to detect at least i targets was computed as

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

where N_t is the number of samples that detected at least i targets and N is the total number of samples. Expectations are that samples 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 samples behaviour.

3. System architecture

Inside the 3D reconstructed internal environments, five virtual points of interest (POIs) were identified, one for each environment, thus covering the parlor, 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 synthesizer 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 [18].

A specific plugin was developed to manage semantic annotations and support environment queries from the AI controlling the avatar. The plugin included an interface towards the UE4 visual scripting language *Blueprints* to be easily redeployed in other research scenarios. The system connects the main Blueprint managing the 3D avatar with a specialised Animation Blueprint controlling the arms movements. When a semantically labelled term is pronounced by the avatar, accordingly to a specific event produced by FANTASIA also containing the concept's ID, the avatar's Event Graph, controlling its general logic, queries the semantically annotated 3D model. The enriched model internally queries the available semantic maps and the geometric data to compute the relevant centroids for the given concept and returns them to the 3D avatar. At this time, the Event Graph selects the most *reachable* centroid, as the closest one to the front of the avatar among the ones that are found in a range of 120 degrees. This is to ensure that the pointing gesture can be produced with natural movements. If such a centroid exists, the

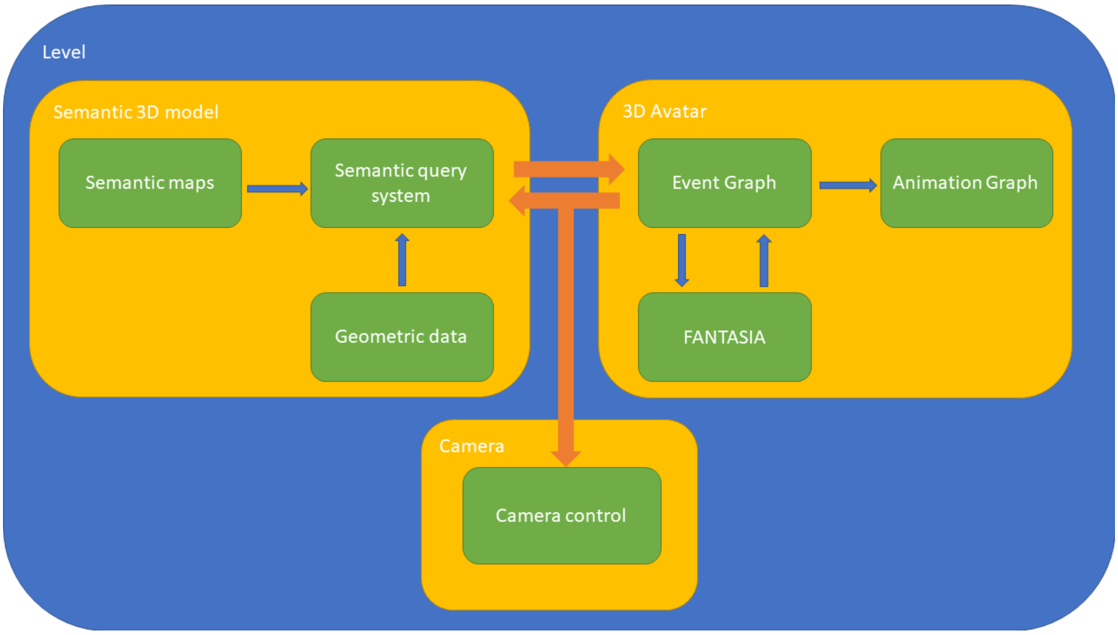

Figure 2: Message exchange system among 3D actors in UE4 to generate camera and arm movements when semantically relevant concepts are pronounced by the avatar.

avatar's Animation Blueprint is passed the location of the centroid and the appropriate arm for the pointing gesture is selected. Then, the target position of the arm, following the vector connecting the shoulder to the centroid is computed and the animation to reach it is generated. The pointing gesture is sustained for 2 seconds before returning to the rest position. At the same time of the arm animation being generated, the 3D camera is also informed about the position of the target centroid by the 3D avatar's Event Graph. Following the same procedure, the camera is animated in order to look at the same position pointed by the avatar. In this case, too, after 2 seconds the camera goes back looking at the avatar. Figure 2 shows the message passing organisation among the involved 3D actors.

While, in the current version, texts are static and manually labelled, the system also supports dynamically generated texts and labels using, for example, entity linking approaches. For the purposes of this paper, it was not necessary to generate texts automatically and this is left for future work. The user interface, for this installation, consisted of a touch interface that allowed visitors to navigate the environment from one POI to the other. Camera movements accompanied transitions among the environment to anticipate the visiting path to the visitors, so that they could more easily identify items of interest during the actual path. Figure 3 shows the touch based interface deployed on a totem device.

4. Results

After applying the annotation procedure to the collected pictures, as explained in Section 2, the obtained target hit probability levels for the two groups were compared to check if there was a

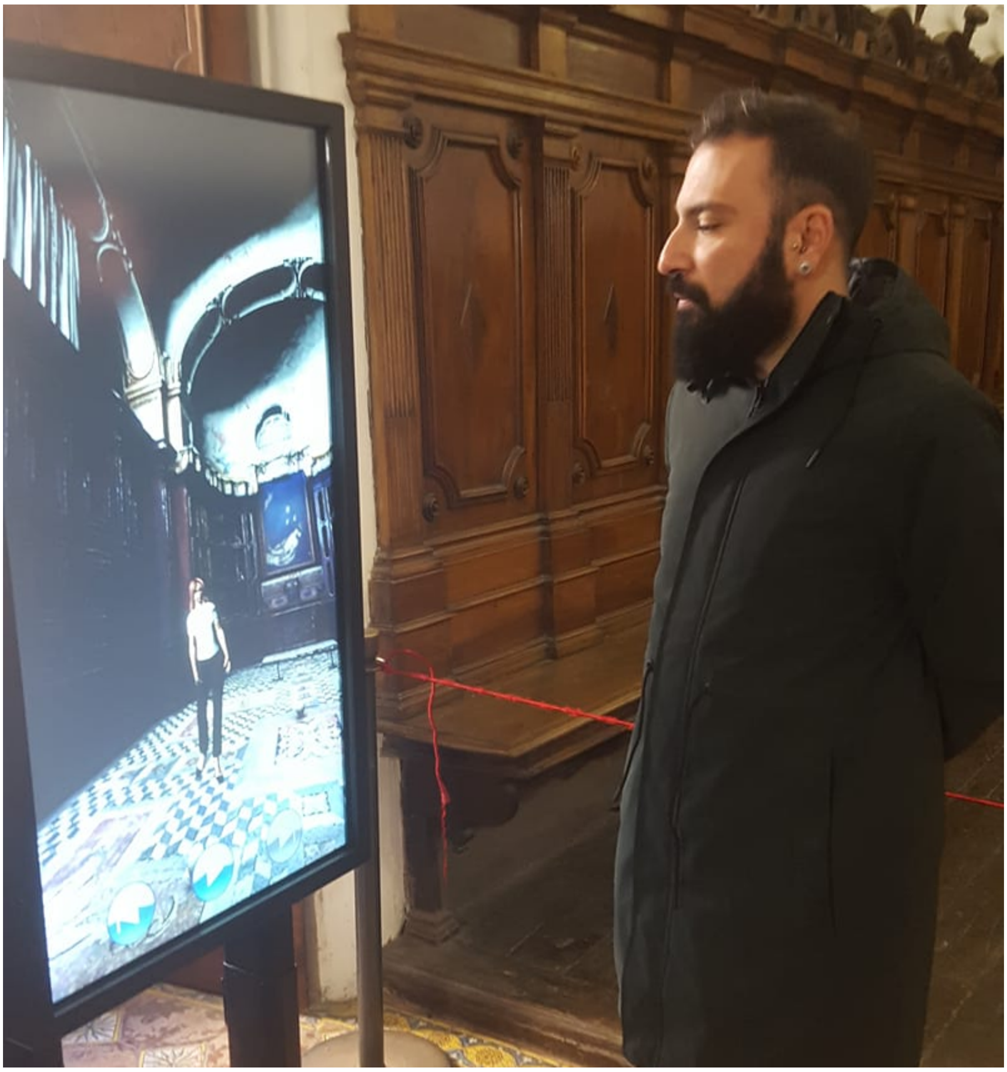

Figure 3: The 3D avatar touch-based interface

difference in the behaviour of 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 target details than the control group at each threshold level. An overview of the probability for a sample from each group to detect the selected targets at different thresholds is shown in Figure 4.

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 samples from the experimental group were able to detect targets that were not detected from the control group, thus being able to focus

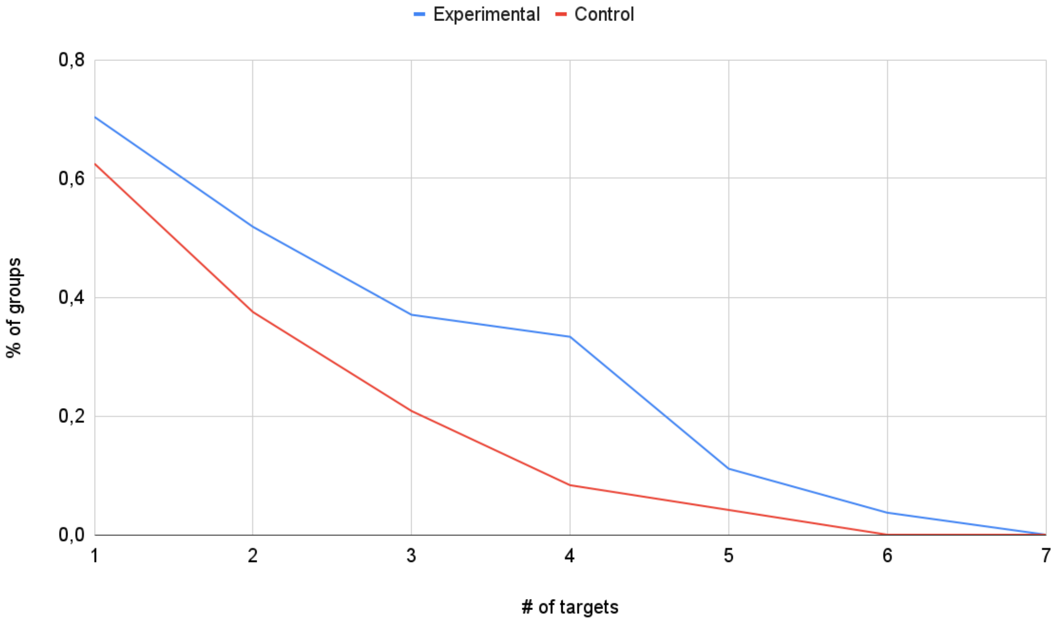

Figure 4: Target detection probability for the experimental and control groups at different thresholds. At each level *i*, the percentage of samples that detected at least *i* targets is plotted.

their attention on important details that would otherwise be lost. A detailed view of the pictures distributions over the targets is shown in Figure 5.

5. Conclusions and future work

We have presented an investigation on visually impacting technology in the case of complex museum environments relying on the visual communication channel to provide cultural contents. Our design approach deploys the technological intervention, based on 3D semantic annotations, 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. To verify the approach, we used of the VEP technique, which is usually adopted for landscape investigations, in the case of complex museum environments. 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 Experimental group was found to be able to detect more target items, during the visit, than the Control group, indicating a successful application of the design concepts to the technological installation we designed.

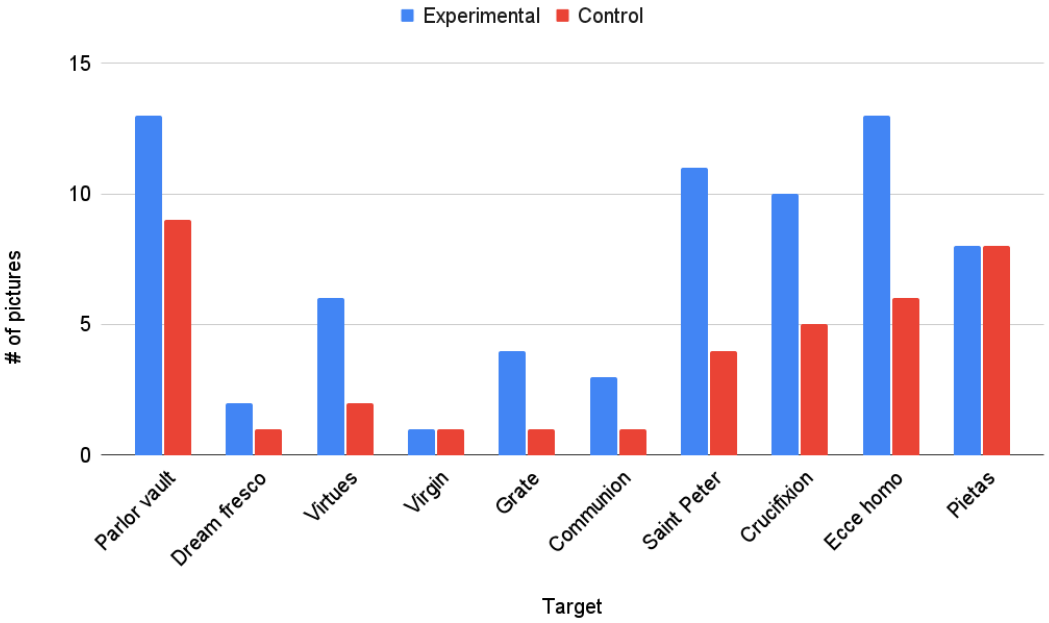

Figure 5: Pictures distribution over the targets.

References

[1] A. Origlia, M. Grazioso, M. L. Chiacchio, F. Cutugno, The role of visually impacting technology in introducing visits to complex cultural sites, in: Proc. of ACM Multimedia (submitted, 2022.

[2] T. Kuflik, A. J. Wecker, J. Lanir, O. Stock, An integrative framework for extending the boundaries of the museum visit experience: linking the pre, during and post visit phases, Information Technology & Tourism 15 (2015) 17–47.

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

[4] M. Carrozzino, M. Colombo, F. Tecchia, C. Evangelista, M. Bergamasco, Comparing different storytelling approaches for virtual guides in digital immersive museums, in: International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Springer, 2018, pp. 292–302.

[5] R. Brondi, M. Carrozzino, C. Lorenzini, F. Tecchia, Using mixed reality and natural interaction in cultural heritage applications, Informatica 40 (2016).

[6] E. L. Deci, R. M. Ryan, The general causality orientations scale: Self-determination in personality, Journal of research in personality 19 (1985) 109–134.

[7] A. Chrysanthi, C. Papadopoulos, T. Frankland, G. Earl, 'tangible pasts': User-centred

design of a mixed reality application for cultural heritage, Archaeology in the Digital Era (2012) 31.

[8] J. Kang, Ar teleport: digital reconstruction of historical and cultural-heritage sites for mobile phones via movement-based interactions, Wireless personal communications 70 (2013) 1443–1462.

[9] S. Gonizzi Barsanti, G. Caruso, L. Micoli, M. Covarrubias Rodriguez, G. Guidi, et al., 3d visualization of cultural heritage artefacts with virtual reality devices, in: 25th International CIPA Symposium 2015, volume 40, Copernicus Gesellschaft mbH, 2015, pp. 165–172.

[10] J.-P. Babelon, A. Chastel, La notion de patrimoine, Liana Levi, 2012.

[11] L. De Luca, 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, 2006.

[12] T. Messaoudi, P. Véron, G. Halin, L. De Luca, An ontological model for the reality-based 3D annotation of heritage building conservation state, Journal of Cultural Heritage 29 (2018) 100–112.

[13] P. Veron, T. Messaoudi, A. Manuel, E. Gattet, L. De Luca, 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, 2014.

[14] V. Cera, A. Origlia, F. Cutugno, M. Campi, Semantically annotated 3d material supporting the design of natural user interfaces for architectural heritage, in: Proc of the AVI-CH Workshop, 2018.

[15] G. J. Cherem, B. Driver, Visitor employed photography: A technique to measure common perceptions of natural environments, Journal of Leisure Research 15 (1983) 65–83.

[16] E. Bell, J. Davison, Visual management studies: Empirical and theoretical approaches, International Journal of Management Reviews 15 (2013) 167–184.

[17] N. Balomenou, B. Garrod, Photographs in tourism research: Prejudice, power, performance and participant-generated images, Tourism Management 70 (2019) 201–217.

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