HAL open science

Impact of View-Dependent Image-Based Effects on Perception of Visual Realism and Presence in Virtual Reality Environments Created Using Multi-Camera Systems

Grégoire Dupont de Dinechin, Alexis Paljic, Jonathan Tanant

To cite this version:

Grégoire Dupont de Dinechin, Alexis Paljic, Jonathan Tanant. Impact of View-Dependent Image-Based Effects on Perception of Visual Realism and Presence in Virtual Reality Environments Created Using Multi-Camera Systems. Applied Sciences, 2021, 11 (13), pp.6173. 10.3390/app11136173 . hal-03283305

HAL Id: hal-03283305

https://hal.science/hal-03283305v1

Submitted on 9 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

applied sciences

MDPI

Article

Impact of View-Dependent Image-Based Effects on Perception of Visual Realism and Presence in Virtual Reality Environments Created Using Multi-Camera Systems

Grégoire Dupont de Dinechin [1,*], Alexis Paljic [1] and Jonathan Tanant [2]

[1] Center for Robotics, MINES ParisTech, PSL University, 75006 Paris, France; alexis.paljic@mines-paristech.fr
[2] JON LAB, 78790 Tilly, France; jonathan.tanant@gmail.com
* Correspondence: gregoire.dupont_de_dinechin@mines-paristech.fr

Featured Application: As a backdrop for our user study, we recreated a museum gallery in virtual reality. More generally, the results presented here may be interesting for all applications based on recreating real-world places as immersive virtual environments.

Abstract: Several recent works have presented image-based methods for creating high-fidelity immersive virtual environments from photographs of real-world scenes. In this paper, we provide a user-centered evaluation of such methods by way of a user study investigating their impact on viewers' perception of visual realism and sense of presence. In particular, we focus on two specific elements commonly introduced by image-based approaches. First, we investigate the extent to which using dedicated image-based rendering algorithms to render the scene with view-dependent effects (such as specular highlights) causes users to perceive it as being more realistic. Second, we study whether making the scene fade out beyond a fixed volume in 3D space significantly reduces participants' feeling of being there, examining different sizes for this viewing volume. To provide details on the virtual environment used in the study, we also describe how we recreated a museum gallery for room-scale virtual reality using a custom-built multi-camera rig. The results of our study show that using image-based rendering to render view-dependent effects can effectively enhance the perception of visual realism and elicit a stronger sense of presence, even when it implies constraining the viewing volume to a small range of motion.

Keywords: image-based rendering; view-dependent effects; multi-camera; virtual reality; user study; visual realism; presence

check for updates

Citation: de Dinechin, G.D.; Paljic, A.; Tanant, J. Impact of View-Dependent Image-Based Effects on Perception of Visual Realism and Presence in Virtual Reality Environments Created Using Multi-Camera Systems. *Appl. Sci.* **2021**, *11*, 6173. https://doi.org/10.3390/app11136173

Academic Editors: Diego Gutierrez, Ana Serrano and Qi Sun

Received: 2 June 2021
Accepted: 28 June 2021
Published: 2 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Several recent publications have demonstrated the use of inside-out multi-view capture to generate photorealistic 360° scenes for virtual reality (VR). In particular, such capture setups can be used to effectively recreate the captured scene in all directions, even as users move their head within a viewing volume of several tens of centimeters [1,2]. Additionally, various authors have shown how photographs taken from multiple different viewpoints can be used to accurately recover complex view-dependent effects such as specular highlights, in order to replicate the way light interacts with reflective and semi-transparent objects in real life [1,3–6]. In this way, such methods aim to make scenes rendered from photographs more realistic. However, these approaches also have drawbacks: most notably, they often generate visual inaccuracies such as flickering or ghosting artifacts (visual data from one object appearing on top of the surrounding objects during blending due to errors in the geometry), and may thus be distracting or disturbing for immersed viewers [1]. With this in mind, we believe that it is important to investigate how immersed viewers effectively respond to effects such as the display of view-dependent highlights and the pres-

ence of distracting visual artifacts, and in particular how this may impact their appreciation of the scene's realism. Few existing user studies have examined this question.

Additionally, because the input photographs are captured around a given viewpoint, the accuracy of the rendered views diminishes the more users move away from the viewpoint. Consequently, several approaches make the scene fade out beyond a fixed range of head motion [1,7]. However, it seems reasonable to assume that making the scene disappear in this way may cause users to feel like they are not actually there, visiting the recreated place, as the scene is thereby likely to feel intangible, fleeting, ephemeral. Therefore, we hypothesize that such limited viewing volumes are likely to cause a significant decrease in what is often referred to in the literature as the sense of presence [8–10] or place illusion [9]—that is, the sensation of being in a real place, of "being there" [8–10]. If this is indeed the case, it may thus be preferable instead to remove this fade effect, in spite of the decrease in visual accuracy as one moves away from the center. Similarly, few existing studies seem to address this question of how having to restrict the viewing volume (due to the virtual environment being created from photographs centered around a given viewpoint) may potentially impact users' sense of presence.

Therefore, in this paper we propose a user-centered evaluation of virtual reality environments created from inside-out multi-view capture systems. Specifically, we lead a user study to evaluate two hypotheses. First, we test the assumption that immersive environments capable of replicating view-dependent highlights are likely to be perceived by users as being more visually realistic than standard alternatives that blur out these highlights, despite the likely creation of visual artifacts. Second, we test the hypothesis that making the scene fade out beyond a small distance (several tens of centimeters) away from the central viewpoint is likely to cause a significant decrease in users' sense of presence. We believe that both of these questions are important, as they may provide valuable insight on how the size of the capture volume and the choice of a rendering method may make the immersive experience seem more realistic and provide a stronger sense of being transported in the recreated location.

Consequently, we make the following contributions in this paper:

- We provide an overview of the challenges and opportunities linked to capturing and rendering real-world scenes using multi-view setups. We notably discuss how we applied this analysis to recreate a museum gallery in VR (see Figure 1 for an illustration of the results) using a custom outward-facing multi-camera rig. These points are notably discussed in Sections 2 and 3.
- We present a user study investigating (1) whether rendering captured view-dependent effects significantly enhances perceived visual realism, by comparing 3D reconstructions rendered with and without these effects; and (2) the extent to which making the scene fade out beyond a fixed radius significantly reduces the sense of place illusion, considering different viewing radii ranging from tens of centimeters to over a meter. This is discussed in Sections 3–5.

We structure the paper as follows. In Section 2, we analyze the literature related to recreating immersive virtual environments from photographs of real-world scenes and evaluating viewers' perception of presence and visual realism. In Section 3, we then describe our study design. We notably provide details on how we developed a multi-camera rig with 40 outward-facing cameras arranged around a central position, and applied it to create a VR museum visit by capturing viewpoints in the main gallery. We also discuss how we designed our investigation into viewers' perception of realism and presence in the resulting virtual environments, outlining the key hypotheses of our study and detailing our experimental protocol. In Section 4, we then present the results of our investigation. Finally, in Section 5 we analyze these results and discuss their implications for future work.

Figure 1. Rendered views from our recreated viewpoint in the MINES ParisTech Mineralogy Museum, with view-dependent reflections visible on the windows and minerals in real-time VR.

2. Related Work

2.1. Recreating Real-World Scenes from Photographs

2.1.1. Issues with Single-Viewpoint Capture

To recreate real-world scenes as photorealistic background environments for VR, many authors today rely on the use of 360° capture devices [7,11–13]. Indeed, consumer-grade 360° cameras are light, easy to use, cheap, and capable of acquiring visual data coming from all directions in a single shot. However, 360° visuals alone can only render the scene as perceived from a single viewpoint: moving one's head in any direction will immediately reveal the absence of geometry through the lack of motion parallax. By comparing users' perception of 360° videos rendered with or without depth data, Serrano et al. [7] thus show that users are effectively more likely to feel discomfort and VR-induced sickness when there is no recovered geometry, thereby justifying the need for adding depth when recreating scenes for VR based on input photographs. Several recent works have thus been dedicated to finding efficient ways of recovering this geometry from 360° photographs to provide motion parallax in VR [7,12].

2.1.2. Strengths of Multi-View Capture

To recover the geometric information required for comfortable motion parallax from a set of input photographs, one very practical solution is to rely on multi-view capture, rather than capturing a single 360° image. Indeed, acquiring the scene from multiple different viewpoints enables the recovery of dense depth data through standard 3D reconstruction techniques [1,7,14]. For instance, Hedman et al. [14] thus demonstrate a method for transforming sets of casually captured photographs into complete and realistic 3D scenes that can be viewed with motion parallax in VR. Furthermore, this strength of multi-view capture has encouraged multiple research efforts towards developing custom inside-out multi-camera rigs designed specifically for recreating 3D scenes for VR. Such systems are typically composed of several tens of outward-facing cameras arranged on a sphere [15] or rotating arc [1], and are all the more practical that they enable capturing the entire set of photographs at the click of a button (rather than having to move a single camera to multiple viewpoints).

Additionally, beyond the practical aspect of being able to more easily reconstruct 3D geometry from the captured photographs, multi-view acquisition systems are also interesting in that they enable the recovery of complex visual effects such as reflections or transparency [1–4,6]. Indeed, a number of view-dependent effects become particularly visible when capturing a scene from multiple viewpoints, such as specular highlights on a shiny

object or reflections on a mirror. Because multi-view setups enable capturing these effects, they can thus be coupled with dedicated image-based rendering algorithms to effectively replicate these captured view-dependent effects in VR. For instance, Overbeck et al. [1] present a system for capturing and rendering light field image datasets at VR-compatible speeds, demonstrating its view-dependent capabilities on several sample scenes and sharing the results by way of a complementary VR application. Similarly, de Dinechin and Paljic [6] present an open-source toolkit and interface for transforming photographs into virtual environments that can be rendered with view-dependent highlights in VR. As another example, Broxton et al. [5] also demonstrate that these approaches can also be extended to dynamic scenes, using a hemispherical setup to capture light field videos that they then render using a deep learning network.

2.1.3. Novel Issues and Need for User Evaluation

However, both processes of estimating 3D geometry and recovering view-dependent effects often also result in the creation of undesirable visual artifacts that are likely to be distracting or uncomfortable for immersed viewers [1,4,5]. Moreover, because the accuracy of the rendered views degrades as one moves further from the center of capture, several authors intentionally restrict the viewing volume by making the scene fade out beyond a given range. Serrano et al. [7] thus apply a fade to gray then to black after respectively 20 cm and 35 cm. Similarly, Overbeck et al. [1] make the scene smoothly fade out beyond a viewing volume 60 cm wide in diameter. Both these constraints and the undesirable presence of artifacts are thus likely to degrade viewers' appreciation of the scene.

Therefore, while these novel solutions are expected to enhance the rendered scenes such that viewers perceive them as being more comfortable and photorealistic, it may be important to lead user studies to evaluate the extent to which this is effectively the case. In this way, our approach significantly differs from previous work. Indeed, by proposing a user-centered investigation into the perception of view-dependent effects and constrained viewing volumes, our investigation is thus complementary to existing evaluations of multi-view rendering solutions that are often instead based on qualitative comparison with previous work and quantitative evaluation via image-quality metrics [1,2,5]. Additionally, by examining view-dependent image-based rendering based on multi-view capture, our work complements investigations into users' perception of comfort or realism in image-based VR scenes [7,12], that most often instead examine single-viewpoint 360° capture and are focused on finding efficient ways of recovering the scene's geometry.

2.2. Visual Realism and Presence

2.2.1. Factors for Achieving the Application's Goals

Because we study immersive virtual environments created from photographs of real-world scenes, it is important for the scene both to appear photorealistic and to elicit a sense of presence.

Indeed, one is most likely to rely on photographs when the goal is to generate an accurate digital replica, typically to create an immersive virtual visit of a specific place for training or educational purposes [11]. In this context, Guttentag [16] explains that it is particularly problematic if the reconstructed environment is not objectively accurate, as viewers may be misled by visual inaccuracies and thus retain incomplete or incorrect information. Furthermore, it is also critical for the reconstructed scene to subjectively feel visually realistic, as virtual tours are expected to provide a visual experience close to what would have been seen had one been physically there, that is, are expected to be virtual substitutes that elicit a sense of visual "authenticity" [16]: viewers may thus be disappointed if it does not seem real.

Additionally, if the reconstructed scenes are to be displayed in VR, then it is critical that they foster users' sense of presence. Indeed, this feeling of "being there" is often described as being the key characteristic of virtual reality [8,10] that motivates rendering the scene in an immersive system rather than on a standard desktop display [9]. Weech et al. [10]

thus explain that unlocking the potential of VR largely depends on our ability to detect and reduce factors that hinder presence, such as sources of cybersickness. Our study of the impact of viewing volume size on place illusion thus falls into the same category of approaches that aim to provide insight on possible barriers to presence in order to ensure that the best solution is adopted to make users feel like they are there.

2.2.2. Evaluating Visual Realism

In the literature, authors often evaluate the realism provided by a virtual scene either by directly recording users' perception that the environment is photorealistic, or by computing specific metrics of visual accuracy, in the latter case based on the assumption that users are likely to better appreciate the scene's realism if it scores higher in terms of objective visual accuracy.

In this way, several researchers rely on single-item 7-point Likert-type questions to assess viewers' perception of a scene's level of realism [7,12], using questions such as "How realistic was the experience of the scene?" [12] and "Which of the two methods offers a more realistic experience?" [7]. If it is expected for the rendering method to generate noticeable visual artifacts, that are thus likely to degrade the scene's realism, it is also possible to use questionnaires to explicitly record the extent to which viewers perceive or are disturbed by these visual inaccuracies [7].

Complementarily, to measure the extent to which a rendered environment accurately replicates the original scene as it was at the moment of capture, authors typically evaluate their proposed solutions by comparing the ground truth photographs with the corresponding output views captured from the same viewpoint, using per-pixel metrics that measure the difference between two images [2,5,14,17]. The work of Waechter et al. [17] is notably interesting in this regard: the authors present an analysis of existing image comparison metrics and geometry-based benchmarks, based on which they propose a novel evaluation methodology, virtual rephotography, aimed at evaluating both the accuracy and the visual completeness of the views rendered by a given solution. This methodology has inspired several ensuing research works as a means to evaluate proposed solutions and compare them with previous approaches [7,14,15]. In particular, because the absence of view-dependent surface reflections would be likely to lead to significant differences when comparing input photographs and rendered views, such image comparison metrics can also be used as an indicator of a rendering method's ability to convincingly render view-dependent effects [5].

2.2.3. Evaluating Presence

How to measure presence, and which factors foster or hinder this sense, are questions that have been thoroughly examined in the literature for several decades [8,10].

Evaluating presence is most often done by way of questionnaires: Schwind et al. [8] thus recently performed a comparative analysis that showed that multiple-item forms such as the Witmer and Singer (WS) [18] and Slater–Usoh–Steed (SUS) [19] questionnaires are commonly used to provide standardized self-reported measures of presence. In this way, when evaluating virtual environments created from photographs, many researchers rely on questionnaire data to evaluate presence, using either standard forms from the literature [13] or simple single-item assessments [7,12].

Through the evaluation of presence, user studies have also helped to better understand the different factors that may hinder it. For instance, Weech et al. [10] provide a detailed analysis of the association between presence and cybersickness in VR based on a review of existing work, in which they outline the different factors that seem to contribute to one or the other, and conclude that existing evidence points to a negative correlation between the two. Additionally, several authors discuss ways to reduce the occurrence of "breaks in presence" [8,9], which are notably expected to occur when users notice inconsistencies in the perceived stimuli due to reaching the boundaries of the immersive system's capabilities [9]. In order to better understand the influence of a proposed method

on viewers' sense of presence, it is thus important to anticipate these different barriers and evaluate the significance of their impact.

3. Materials and Methods

3.1. Multi-View Rendering for a Mineralogy Museum

3.1.1. Relevance for Minerals and the Museum Context

Using multi-view capture to render view-dependent effects is most interesting when visual realism is key to the application's goal. It thus seems particularly relevant for educational experiences, such as virtual museum tours. Indeed, not only are museum collections likely to comprise objects that display complex visual effects, but the context of a museum's educational mission may also specifically encourage a focus on recreating scenes in a realistic way.

Additionally, minerals are a particularly interesting object to capture from multiple viewpoints. Indeed, they often display various forms of highlights, reflections, and translucency that become visible when looking at the items from different sides. Consequently, minerals are difficult to recreate accurately using standard photogrammetry pipelines, and instead can be expected to strongly benefit from being rendered using view-dependent image-based techniques.

Therefore, we selected as a practical backdrop for our study the recreation of a mineralogy collection in VR. We thus partnered with the MINES ParisTech Mineralogy Museum in order to capture the data required for our study, and were thereby kindly given access to the museum to take photographs of the main gallery using our capture rig. The museum staff also provided us with photographs of individual minerals in the collection, based on which we created 3D replicas of these items.

3.1.2. Recreating Individual Minerals

In a mineralogy museum, visitors often only see the collection behind glass display cases: minerals can be rare or precious items, and precautions thus have to be taken to ensure that they are not damaged. In the same way, because we were not trained museum staff, we could only take photographs of the minerals inside their protective cases. However, the photographs we captured in this way proved to be unpractical for rendering with view-dependent effects. Indeed, doing so noticeably generated artifacts by which reflections on the display case windows were projected onto the reconstructed minerals, thereby degrading the quality of the virtual replica.

Therefore, we needed to find a way to obtain higher-quality photographs of the minerals, captured individually and on a uniform background. The museum's curators kindly agreed to help us in this task: the minerals were thus moved into a dedicated light box by museum staff, captured from multiple angles using a high-resolution hand-held camera, and returned into their display cases. These photographs were effectively a much better fit for our reconstruction and rendering pipeline: each mineral's 3D shape could be recovered with high accuracy, and could be rendered with photorealistic highlights in virtual reality. We provide an illustration of a mineral recreated in VR using this methodology in Figure 2.

3.1.3. Recreating Viewpoints in the Museum Gallery

We then aimed to recreate the sense of being inside the museum itself. To do so, we needed to be able to capture large amounts of photographs around a given viewpoint. This led to our development of a custom camera rig, which we describe in Section 3.2.

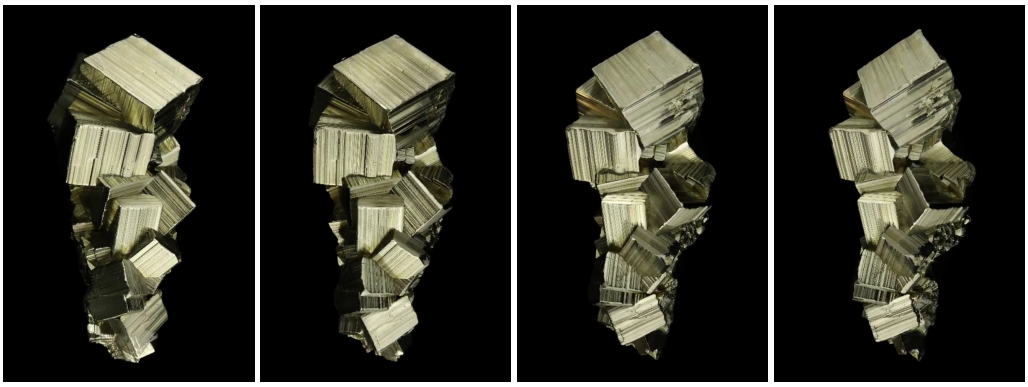


Figure 2. Rendered views of a pyrite mineral we recreated in VR (from original photographs of the mineral by: MINES ParisTech Mineralogy Museum/E. Gaillou). The virtual object displays view-dependent highlights where light was captured bouncing off the mineral's facets.

We thus applied our custom-built multi-camera rig to recreate a central viewpoint within the museum's main gallery. The scene contained multiple reflective surfaces, most notably the glass display cases and the minerals themselves. Therefore, by capturing large amounts of densely packed images, we were able to capture a dense sample of view-dependent highlights. However, while capturing more photographs enabled us to enhance the final rendering quality, this also increased processing time and rendering latency. Consequently, we led several reconstructions with different image counts to find the trade-off we were most satisfied with. Ultimately, for our user study, we decided to capture a museum viewpoint using 320 photographs of the scene. Using our multi-camera rig, we were able to lead this capture process in about 15 min. Views rendered from this reconstruction are illustrated in Figure 1.

3.2. Designing and Prototyping the Capture Rig

The specifications for our inside-out camera system were as follows. We required the system to be capable of capturing large numbers of photographs at the click of a button, such that the photographs could then be processed by existing 3D reconstruction tools to generate a complete 3D environment fit for use with view-dependent image-based rendering methods. The system also had to be relatively lightweight and portable, so as to enable capturing viewpoints in different locations. Finally, we required the camera layout to enable recovering visual data in all directions, with the potential exception of small areas at the zenith and the nadir deemed less essential.

3.2.1. Hardware Setup

For reasons of availability, size, weight, cost, and ease of use, we chose to build the system using 40 Raspberry Pi portable computers connected to the corresponding 8 Megapixel v2 camera modules. To simulate different possible camera layouts using these modules, we then virtually captured sample 3D scenes in the Unity game engine using various outside-in and inside-out setups, taking into account the fields of view of our selected cameras. We then led alignment and reconstruction tests in Reality Capture, a commercial photogrammetry software, to rapidly evaluate each design based on the quality and completeness of the reconstructed 3D models.

As a result of this virtual prototyping phase, we ultimately selected a layout with 5 circular levels of 8 outward-facing cameras arranged in a uniform manner. Two successive cameras on a given level were thus separated by a 45° angle. To obtain a better coverage, we also set each level to be offset by 22.5° from the next. We chose the height of the center level to make it correspond to the average height of view of an adult human in a standing position (about 165 cm), to be able to capture and replicate a scene as if it was naturally experienced by a standing person in reality. To achieve this height, we initially planned to

mount the rig on a tripod; however, we also needed to be able to move it easily and we required space for gear such as network switches, routers, cables, and a power supply. We thus decided to mount the upper part of the rig on a supporting cart, which allowed us to store the required equipment for easy handling. The cart was built around the rack that encased all the gear we needed, notably a power supply, network switches, and a surface to hold an open laptop. Adding 9 mm plywood for the structure gave us a square section of 51 × 51 cm. We also added wheels (two of them with brakes), and set additional space for an optional counterweight (or battery), to prevent the height and weight distribution of the rig and cart from causing tilting. A 5 V/120 A power supply was finally installed to distribute the voltage that each Raspberry Pi computer required.

Additionally, in order to increase sampling density and camera overlap, we mounted the upper part of the rig on a lazy Susan (turntable). Indeed, this enabled us to rotate the camera levels around the vertical axis in a continuous fashion, to obtain as many intermediate angular positions as needed. Given that humans also essentially look around by rotating their heads around a vertical axis, we deemed it less critical to also increase sampling density along the vertical arc. The system thus only rotates in the horizontal plane.

Finally, we designed and 3D-printed camera mounts to attach the modules, using Blender and a stereolithography resin printer. Indeed, we had initially planned to use camera mounts from the market, but they proved to be very fragile and not flexible enough to configure the tilt angle. Our 3D-printed mounts thus allowed us to tilt each camera to 3 different tilt angles, and could be set in reverse on the plate, providing 2 additional angles. Specifically, each camera could thus be set to a tilt angle of −40°, −20°, 0°, 20°, or 40°. We could thus implement complete cylindrical capture, as well as partial spheric or quasi-spheric capture (without the zenith and nadir areas).

Photographs of the camera rig, with close-ups on a camera with a 3D-printed mount and on the set of rotating wooden panels, are shown in Figure 3.

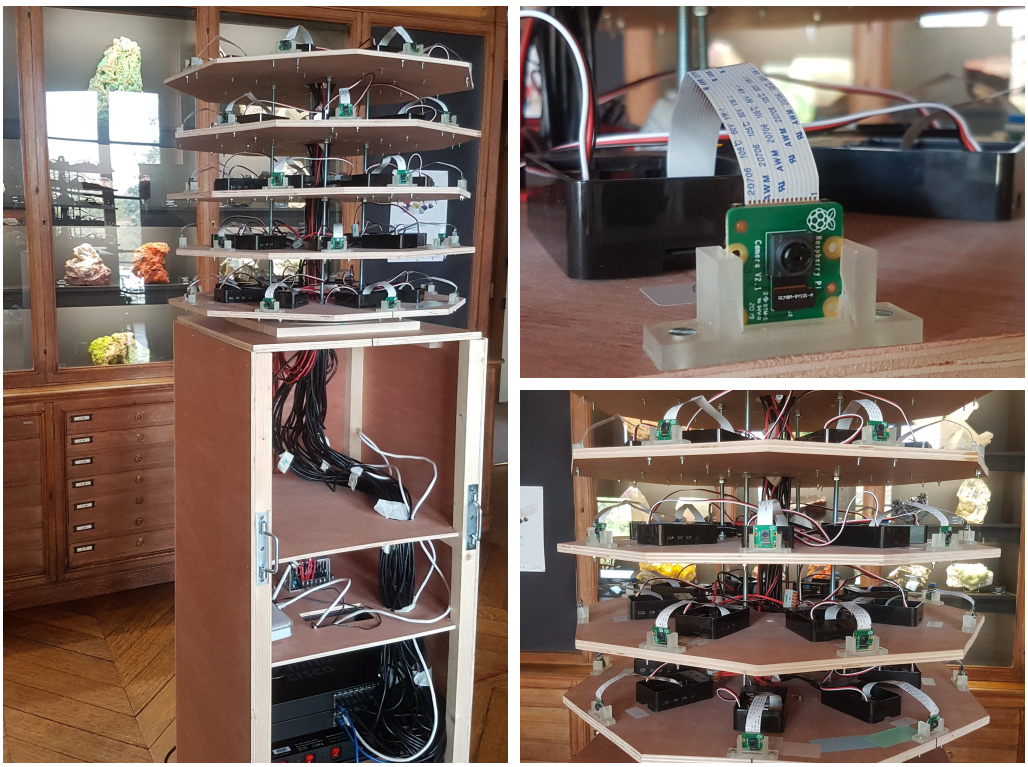


Figure 3. Our rotating 40-camera system for multi-view capture.

3.2.2. Software Implementation

We decided to use a fixed addressing of cameras to get information on the location of each camera taking the photographs. We thus defined an address nomenclature and reserved the corresponding IP addresses in the router: level 1 was 192.168.0.11x, level 2 was 192.168.0.12x, and so on, the last digit being the camera index on the corresponding level. Each Raspberry Pi ran a Python backend, including a TCP/IP server that waits for incoming commands to set the white balance and shutter time. This server also enables launching the capture, with the image data then being transferred over the same network TCP/IP connection. A semi-automatic procedure was developed that allowed the easy installation of a fresh Raspberry Pi running Raspbian, in command line and over the network (SSH), with minimum manual intervention.

We developed the remote/client software in Unity, and tested it on both Windows and Mac OS X. The remote software implements 40 TCP/IP clients and displays a user interface allowing the operator to set capture parameters and remotely trigger cameras in synchronization. After the capture process is launched, the images are gradually displayed side-by-side on the screen for easy monitoring. Each image is then saved on disk with a file naming nomenclature that includes a timestamp and a camera index. Launching this process via WiFi allows the user to carry the laptop during capture, so as to be able to move outside of the rig's field of view.

3.2.3. On-Site Tests

We tested the multi-camera rig to perform several captures on site. We thus validated that the rig can be moved easily to different locations, its overall size ($190 \times 51 \times 51$ cm) enabling it to pass through doors and elevators with no difficulty. We also tested several intermediate angle captures, with 8 and 16 rotations total (thereby yielding respectively 320 and 640 images). In the end, we selected for our study a capture process with 8 intermediate rotations. To subdivide the 45° angle between consecutive cameras into equal rotation angles, we used a pen to draw small markings in-between two cameras on one of the circular levels, corresponding to the fixed angles we needed. The tilt angles we selected for the cameras at each level were, from top to bottom, 40°/20°/0°/−20°/−40°.

3.3. Creating and Rendering a 3D Environment

3.3.1. Generating 3D Models from the Captured Photographs

To reconstruct a 3D model of our captured items and scenes, we decided to rely on Schönberger and Frahm's [20] COLMAP (see Figure 4), which had consistently achieved satisfying results during testing. We were thus able to recover intrinsic parameters, positions, and orientations for every camera. The software's dense reconstruction step also provided us with a plausible 3D model of the minerals and environments, which we simplified and cleaned up using Blender and Jakob et al.'s [21] Instant Meshes.

From the 41 high-resolution photographs of the pyrite mineral, we thus obtained a final 3D mesh composed of 122 k vertices and 242 k triangles. Similarly, we simplified and cleaned up the 3D model reconstructed from the 320 photographs of the museum gallery to finally obtain an output mesh composed of about 80 k vertices and 160 k triangles. Because part of our investigation consisted of comparing the scene's appearance with and without view-dependent highlights, and since COLMAP does not texture the output model, we also applied the global texture mapping method implemented in de Dinechin and Paljic's [6] COLIBRI VR to obtain textures for these models.

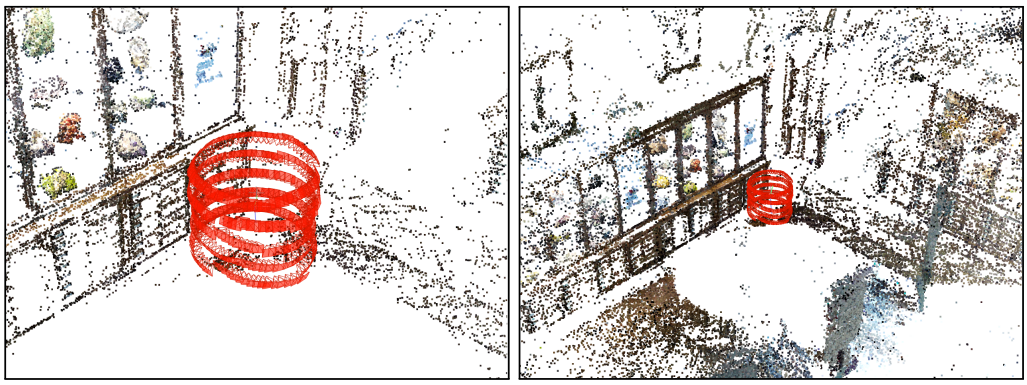


Figure 4. Reconstructing the museum scene in COLMAP.

3.3.2. Applying a View-Dependent Rendering Method

We then rendered the obtained models with view-dependent highlights in VR in Unity, using the implementation of unstructured lumigraph rendering [3] made available in the open-source COLIBRI VR toolkit [6]. We simply applied the toolkit's unstructured lumigraph rendering method using a global 3D mesh as proxy, without any modification to the algorithm. Using this method, we were able to render the scene at framerates of more than 60 frames per second on our setup, with specular highlights and reflections effectively becoming visible on the facets of the minerals and on the glass display cases. With the obtained VR scene, we could now lead our user study.

3.4. Study Protocol

3.4.1. Hypotheses

We designed our study to investigate two hypotheses:

Hypothesis 1 (H1). *Larger viewing volumes provide participants with a stronger sense of presence, due to the virtual location not fading out as the viewer moves away.*

Hypothesis 2 (H2). *Scenes rendered with view-dependent effects provide participants with a stronger impression of visual realism, due to the corresponding increase in fidelity with regard to the real-world scene.*

3.4.2. Scenes and Overall Procedure

To test these hypotheses, we relied on two sample scenes. In one scene, viewers were thus shown a virtual mineral we had recreated from the photographs captured by the museum's curators (see Section 3.1.2), specifically a pyrite mineral. This was used to study H2, by rendering the mineral with or without view-dependent highlights. In the other scene, users were shown the recreated viewpoint captured in the museum's gallery using our multi-camera rig (see Section 3.1.3). This second scene was used to study both H1 and H2, by varying the size of the viewing volume between participants, and showing the scene successively with and without view-dependent effects. To give a better sense of what the scenes looked like, we provide a panoramic view of the museum scene and its underlying geometry in Figure 5.

Each participant thus observed both scenes, each rendered successively both with and without view-dependent highlights, in a randomized order. Specifically, participants were tasked with observing each scene for a few seconds, before answering a series of questions related to perceived realism, comfort, and presence. After seeing the two ways of rendering a given scene, participants were also asked to state which rendering solution they had preferred, for both scenes.

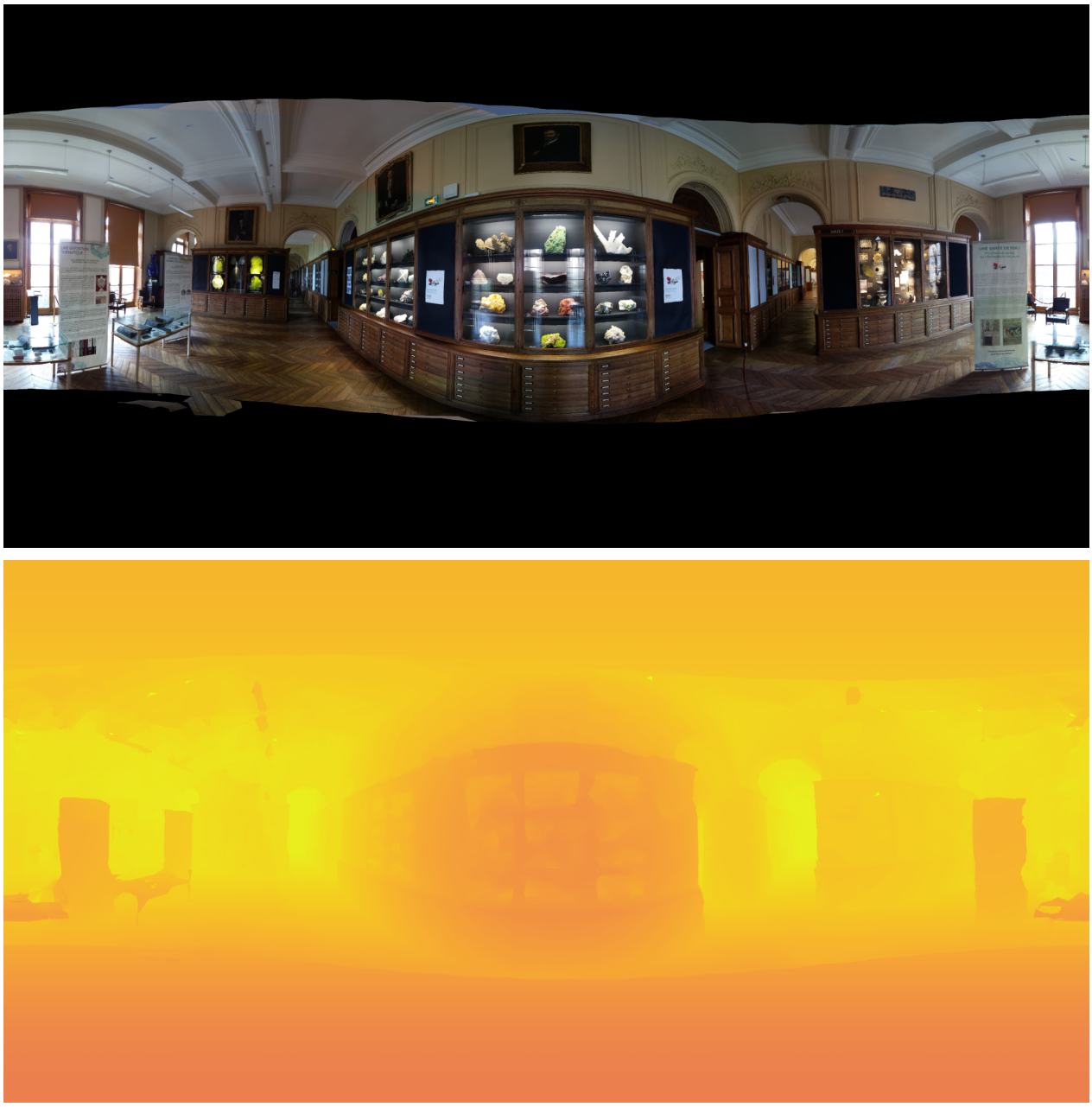


Figure 5. The images show 360° color and depth panoramas obtained by placing a virtual camera in the center of the "Museum" scene. Note that we generated these panoramas from the scene simply for illustration: no 360° photographs were captured or rendered during the study.

Participants conducted the experiment in a standing position, and could move around the virtual scenes by walking normally within a tracking space about 2 m wide and 2 m long. The head-mounted display used for the study was a HTC Vive Pro. The scenes were displayed using Unity version 2019.2.15f, using a computer with a NVIDIA GeForce GTX 1070 GPU. The experiment lasted between 15 and 20 min for each user.

3.4.3. Participants, Ethics, and Safety

We conducted the study on N = 18 participants—12 male and 6 female. Users were aged between 19 and 39 years old (Mdn = 21, SD = 1.08). Many were university students enrolled in an ongoing course on virtual reality, and most thus had prior experience with VR head-mounted displays. Participation in the study was not mandatory for the course

on VR, and users took part in the study anonymously, with no record of their name in the collected data.

We recruited participants on the university campus by way of an online form. In this form, we provided details on the study protocol and on the way we would use the recorded data. Interested users then replied to the form by selecting an option stating that they wished to take part and indicating their preferred time slot. We obtained approval to lead our user study after submitting our protocol and registration form to the relevant ethics and safety committee of our institution—specifically, the Health, Safety, and Environment (HSE) officer.

We also adapted our protocol to prevent risks related to the ongoing pandemic. Participants were thus required to wear a face mask to take part in the experiment, and were provided with hand sanitizer as they entered the study room. A safe distance was also consistently kept between the experimenter and the participant. Additionally, to ensure as little interaction with the experimenter as possible, participants were asked the study questions while still in VR, instead of being given a post-test questionnaire to fill out by hand. We do not believe that this change hinders the results in a meaningful way: in particular, previous work in the literature has shown that completing questionnaires while in VR does not significantly change measures of presence [8]. Finally, we sanitized the VR head-mounted displays between each participant using CleanBox UVC decontamination systems, and ventilated the room regularly by opening the windows.

3.5. Study Conditions and Measures

3.5.1. Independent Variable: ViewRadius

To study H1, we considered as a first independent variable the fade radius of the viewing volume. This variable was used in the "Museum" scene captured using our camera rig, thereby constraining the viewing volume as users moved away from the center of capture. It was chosen as a between-subjects variable, in order to be able to ask participants which rendering method they preferred without the answer being biased due to different viewing volume sizes.

Specifically, we gave this variable one of three levels in the study:

- 20 cm: In our scene, this was the empirical limit until which the rendering method was capable of producing noticeable view-dependent effects, given our 30 cm capture radius. This value is also relevant because it is close to the radii used in recent works [1,7].
- 50 cm: This was an intermediate radius, which we expected would not be perceived as being extremely constraining. However, within the second half of the corresponding viewing volume in our sample scene, view-dependent effects were no longer accurately rendered.
- 100 cm: This was the volume which we considered to be a reasonable upper bound, given the size of our tracking space.

Whichever its value, the scene linearly faded to black when moving beyond this radius, over a fixed 5 cm distance. This is illustrated in Figure 6.

Figure 6. User's view in the "Museum" scene rendered with view-dependent effects and a view radius of 20 cm. From left to right, the user's head has moved to $x = 0$ cm, $x = 10$ cm, $x = 20$ cm, and $x = 22.5$ cm. Notice how the view fades out (at $x = 25$ cm, it would be entirely black), as well as how the reflection on the window moves in front of the items inside the glass case.

3.5.2. Independent Variable: RenderType

To study H2, we considered as a second independent variable the rendering condition. It was chosen as a within-subjects variable, in order to ask participants their preferred solution. This variable was used in both the "Museum" and "Pyrite" scenes, which displayed different types of view-dependent effects: the pyrite exhibited mirror-like reflections on its different facets, while the museum scene presented more subtle effects such as transparency and reflectivity on the glass display cases. Each scene could thus be rendered either with view-dependent highlights (using unstructured lumigraph rendering) or without (using only a global texture map).

3.5.3. Dependent Variables: Questionnaire

In the course of the study, we essentially aimed to measure users' perception of visual realism and presence. To do so, we built upon several relevant questionnaires from the literature to create a study form with multiple Likert-type items. We provide an overview of this questionnaire in Table 1. We notably included questions from previous works investigating visual realism, comfort, and general preference [7,12] (questions 1–3, 9, and 10). We also included questions related to our more specific question of viewing volume size (questions 4 and 5). Finally, to measure presence, we used the three-item SUS questionnaire from the literature [19] (questions 6–8), which was seen as more practical than longer forms given that participants answered from within VR. All questions were on 7-point scales, with short descriptions provided for the scales' extremes.

Table 1. Study questionnaire, with "MP" if asked in both the "Museum" and "Pyrite" scenes, and "M" if only asked in the former.

1. How visually realistic was the scene?	MP
2. If you perceived visual artifacts in the reconstruction, to what extent did you find them disturbing?	MP
3. How visually comfortable was the scene?	MP
4. How comfortable were you moving around in the scene?	M
5. Did you feel constrained in your movements?	M
6. To what extent did you feel like you were there, in the scene?	M
7. To what extent did the scene become the reality for you?	M
8. Did you feel more like you were seeing images of a place, or visiting a place?	M
9. Did you feel nausea, vertigo, or headache?	M
10. Overall, which method did you prefer? (after seeing the scene rendered with and without view-dependent effects)	MP

4. Results

We analyzed the recorded observations by conducting the relevant analysis of variance (ANOVA) tests, followed by pairwise t-tests to refine the analysis when appropriate. We used a significance level of 5%, and applied the Bonferroni correction to adjust p-values. Statistical testing was computed in R.

4.1. Experiment 1: "Pyrite" Scene

To analyze the observations in the "Pyrite" scene, we conducted a repeated-measures ANOVA test with RenderType as a within-subjects factor. This analysis validated a statistically significant main effect of RenderType on VisualComfort (question 3): $F_{(1,17)} = 8.774$, $p.adj < 0.01$.

Users thus found the scene to be more visually comfortable when there were no rendered view-dependent highlights ($M = 6.00$, $SD = 1.08$) than when these were rendered ($M = 4.67$, $SD = 1.57$), as illustrated in Figure 7.

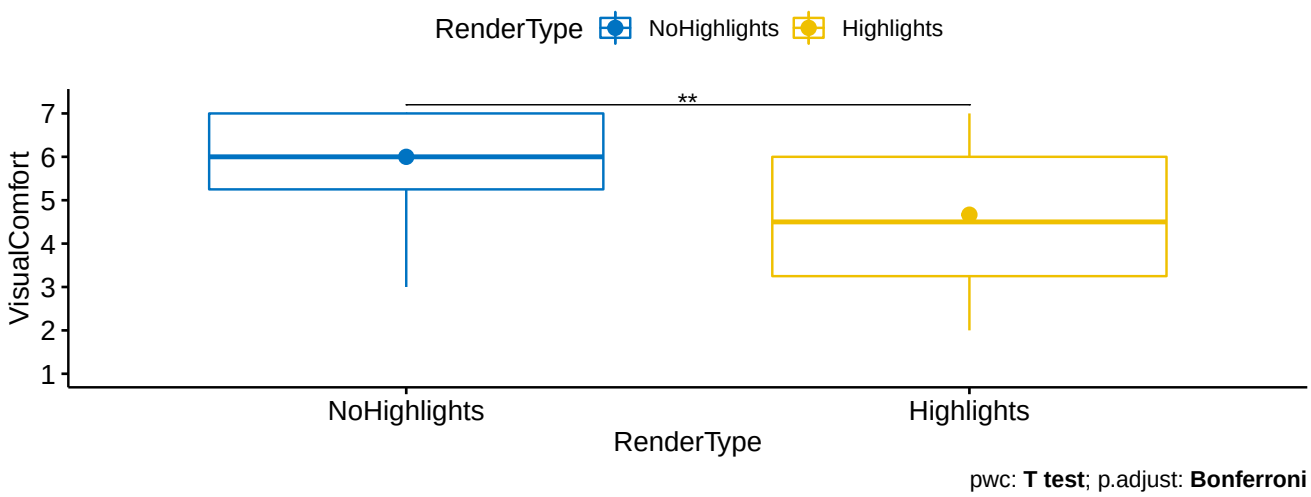


Figure 7. In the "Pyrite" scene, visual comfort (higher is more comfortable) was shown to significantly differ based on rendering type.

In terms of preference (question 10), 8 users reported preferring the replica in the condition with view-dependent highlights, and 10 users in the condition without. A χ^2 test on these scores did not show a significant preference for one method over the other.

4.2. Experiment 2: "Museum" Scene

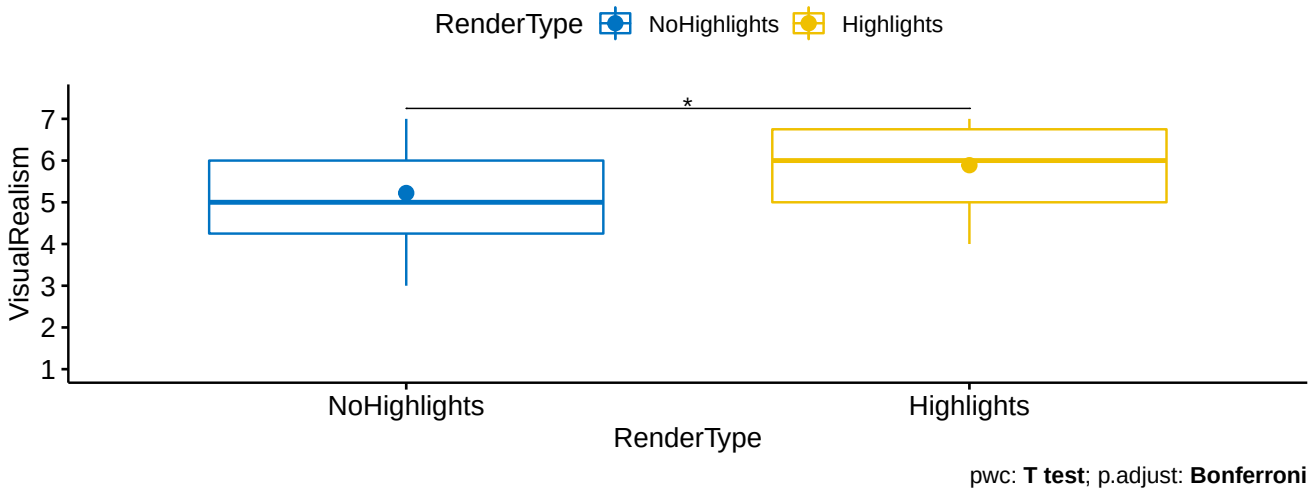
To analyze the observations in the "Pyrite" scene, we conducted a two-way mixed ANOVA test with ViewRadius as a between-subjects factor and RenderType as a within-subjects factor.

As illustrated in Figure 8, this analysis validated a statistically significant main effect of RenderType on:

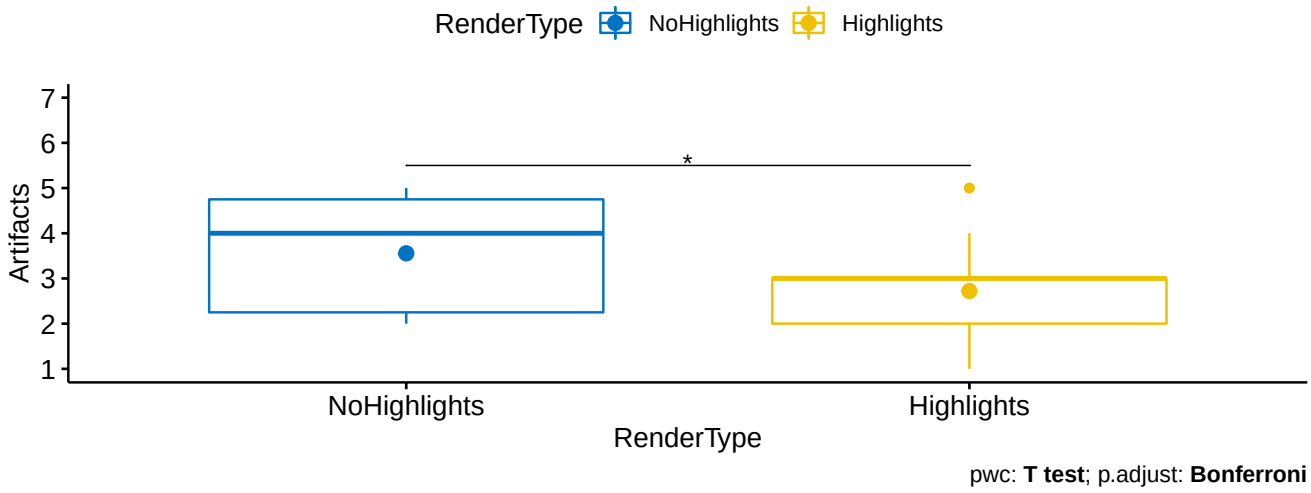
- VisualRealism (question 1): $F_{(1,15)} = 4.661$, $p.adj < 0.05$. Users thus found the scene to be more visually realistic when there were rendered highlights ($M = 5.89$, $SD = 0.90$) than when these were not rendered ($M = 5.22$, $SD = 1.40$).
- Artifacts (question 2, higher is more comfortable): $F_{(1,15)} = 5.046$, $p.adj < 0.05$. Users were thus more disturbed by artifacts when there were rendered view-dependent highlights ($M = 2.72$, $SD = 1.23$) than when these were not rendered ($M = 3.56$, $SD = 1.20$).
- Presence (questions 6–8): $F_{(1,15)} = 9.023$, $p.adj < 0.01$. Users thus felt a stronger sense of presence when there were rendered view-dependent highlights ($M = 5.39$, $SD = 1.04$) than when these were not rendered ($M = 4.76$, $SD = 1.36$). Leading the test on each individual component showed a significant impact on the response to question 7 in particular.

Additionally, as illustrated in Figure 9, this analysis validated a statistically significant main effect of ViewRadius on the sense of being Constrained (question 5, higher is more free): $F(2,15) = 5.242$, $p.adj < 0.05$. Leading a pairwise t-test to refine the results confirmed a significant difference both between the 20 cm and 100 cm conditions ($p.adj < 0.001$), and between the 50 cm and 100 cm conditions ($p.adj < 0.01$). Users thus felt more constrained in their motion when viewing the scene with a 20 cm fade radius ($M = 3.14$, $SD = 1.61$) than with a 100 cm fade radius ($M = 5.7$, $SD = 1.16$), and similarly a 50 cm fade radius ($M = 3.50$, $SD = 1.51$) was found to be more constraining than a 100 cm one.

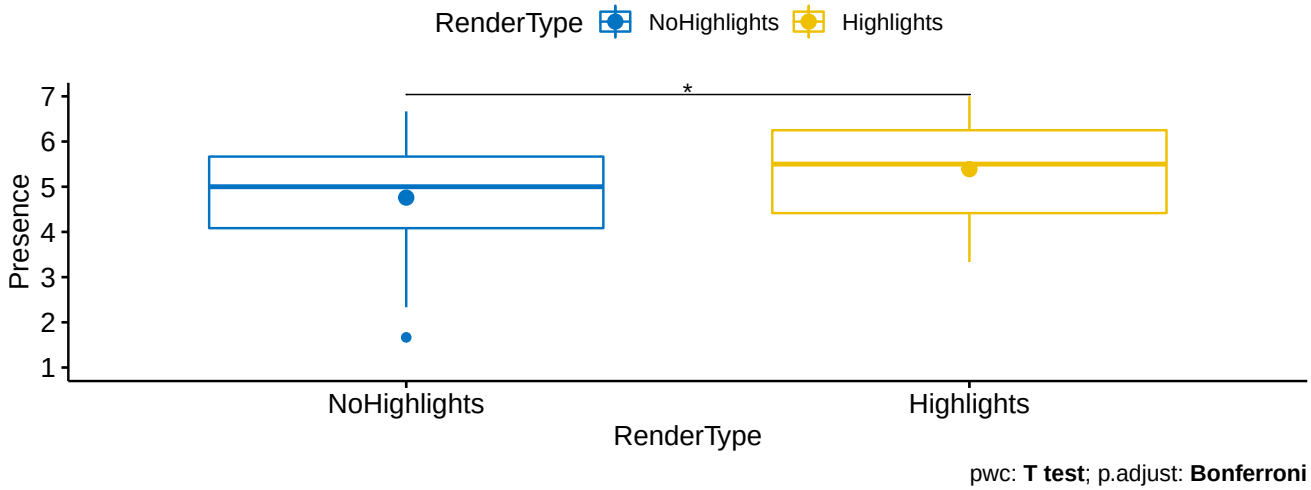
Furthermore, as illustrated in Figure 10, our analysis validated a statistically significant interaction between RenderType and ViewRadius on perceived Sickness (question 9): $F_{(2,15)} = 4.341$, $p.adj < 0.05$. Leading a pairwise t-test to refine the results confirmed a significant difference both between the 20 cm and 100 cm conditions ($p.adj < 0.05$), and between the 50 cm and 100 cm conditions ($p.adj < 0.05$) when considering scenes rendered without view-dependent highlights. Specifically, in this rendering condition, users reported feeling a significantly stronger sense of sickness when they had the larger 100 cm viewing volume to explore ($M = 4.60$, $SD = 1.82$) than in the restricted 20 cm ($M = 6.71$, $SD = 0.488$) and 50 cm ($M = 6.50$, $SD = 0.837$) conditions.



(**a**)



(**b**)



(**c**)

Figure 8. In the "Museum" scene, the following were shown to significantly differ based on rendering type: (**a**) visual realism (higher is more realistic); (**b**) perception of artifacts (higher is less disturbing); (**c**) presence (higher is stronger).

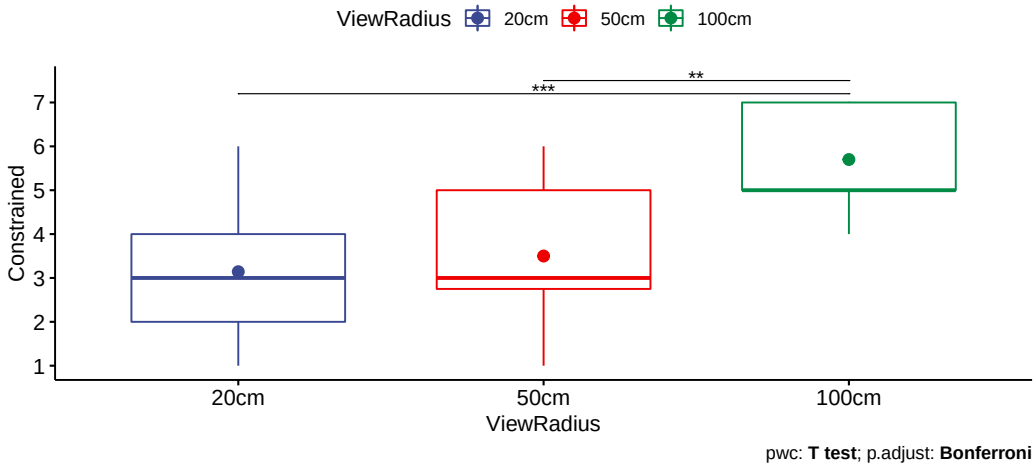


Figure 9. Perceived constraint during motion (higher is more free to move around) was shown to significantly differ based on viewing radius in the "Museum" scene.

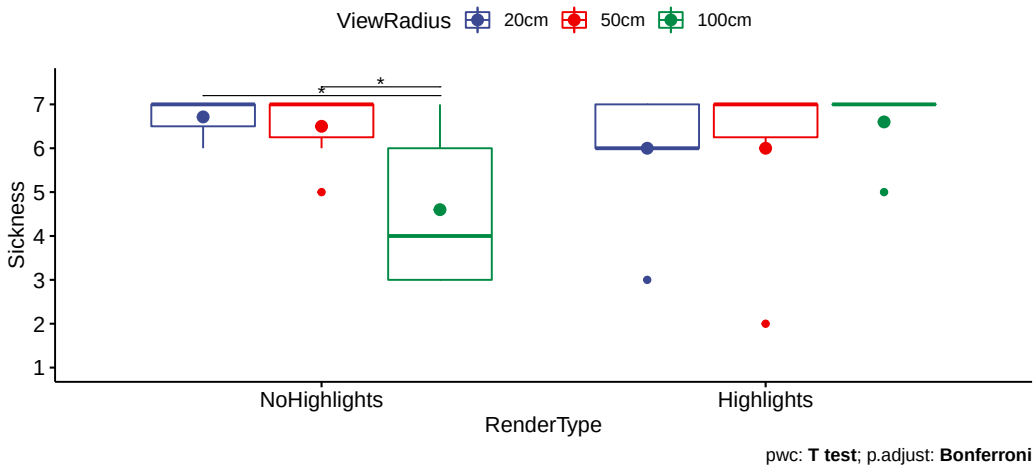


Figure 10. Perceived cybersickness (higher is less sick) was shown to significantly differ based on the interaction between rendering condition and viewing radius in the "Museum" scene.

Finally, in terms of preference (question 10), 13 users reported preferring the replica in the condition with view-dependent highlights, and 5 users in the condition without. A χ^2 test did not show a significant preference for one method over the other.

5. Discussion

5.1. Results Analysis

In the "Pyrite" scene, our analysis shows that the mineral was perceived as being less comfortable to look at when rendered with view-dependent highlights. This result was also consistent with additional comments made by many participants, with several noting that the reflections made the object flicker even when performing little motion, thereby causing discomfort. Overall, most participants thus found the mineral with highlights to be more visually realistic, but less comfortable to look at, which likely explains the lack of consensus in terms of preference. Interestingly, some participants even found the mineral to be less realistic when rendered with highlights, as the transitions between reflections from different images were deemed too rough and exaggerated.

In contrast, the more subtle view-dependent effects in the "Museum" scene seem to have elicited more consistently positive responses from users, and the scene rendered with view-dependent effects seems to have been preferred overall by the participants in our sample group. In particular, the environment was found to be significantly more

visually realistic when rendered with view-dependent reflections: several participants thus commented that details seemed crisper in this condition, and the reflections on the glass windows and on the minerals were mostly found to be convincing. Additionally, this condition elicited higher presence scores, which we believe is likely also linked to the increase in perceived resolution and visual realism. On the other hand, rendering the scene with unstructured lumigraph rendering instead of a fixed texture map caused viewers to perceive more artifacts: this is consistent with observations from previous works that underline the importance of mitigating such artifacts [1,5], as these visual inaccuracies are bound to appear when applying image-based rendering methods.

Finally, the results of our analysis underline that making the scene fade out at closer ranges does indeed make viewers feel more constrained in their movements. Fade radii of 20 cm and 50 cm were thus perceived as similarly quite constraining, while the radius of 100 cm—that roughly corresponded to the physical bounds of our tracking space—was perceived as less of a barrier. This is consistent with user comments, which underlined eagerness to move around to explore the scene, and frustration when virtually stuck within a small radius. More surprisingly, our analysis also seems to show that rendering the scene without highlights and with a large viewing volume caused several users to perceive stronger symptoms of cybersickness, which can perhaps be interpreted as a result of viewers moving further away from the central viewpoint to explore the scene. In any case, the size of the viewing volume did not seem to have had an impact on viewers' perception of presence, which was instead more impacted by the choice of a rendering method.

5.2. Hypothesis Validation

Based on the results of our analysis, we thus cannot validate H1. Indeed, the fade radius did not appear to have an impact on perceived presence: even when viewers were constrained to stay within a small viewing volume, they reported feeling a sense of "being there", with levels similar to those reported in larger viewing volumes. In this way, it seems that viewers' mental construct of being in a place does not require the place to be visible at all times. Perhaps this result can be interpreted by analogy with the fact that the real world around us "disappears" regularly when we close our eyes, yet does not feel less real because of it, because we know by experience that opening our eyes again will make the world appear once more. In the same way, perhaps viewers feel that the virtual world still exists around them even though it disappears beyond a given range of motion, simply because the place consistently re-appears when they move back to the central viewpoint: they can regularly check that the world exists by moving back to this viewpoint, confirming that it is not an intangible, fleeting world that stops existing beyond a given range, but rather a world upon which a dark veil is placed beyond this range, the world itself still existing in spite of it.

On the other hand, our analysis is quite consistent with H2, and the results of our experiment in the "Museum" scene confirm that rendering view-dependent highlights can significantly enhance viewers' perception of visual realism. However, our results on the "Pyrite" scene also lead us to exert caution: the added visual accuracy does not appear to always be perceived as enhancing the scene's realism, as some reflections may be perceived as being too exaggerated or visually uncomfortable to seem realistic. Consequently, additional research should lead to a better understanding of the conditions under which view-dependent effects are perceived as realistic when rendered with image-based techniques, i.e., whether this depends on the type of effect, sampling density, or other aspects of the rendering method that still have to be investigated.

5.3. Paths for Future Work

Our results encourage several paths for future work. In particular, the observation that viewing volume size does not appear to impact presence is interesting, as it may encourage focusing on rendering higher-quality views (fewer artifacts, more realistic effects) within small viewing volumes rather than necessarily augmenting the size of the volume itself.

Additional studies could also be led to evaluate image-based rendering methods on a wider array of scenes and with larger sample sizes, in order to better understand the conditions under which view-dependent effects are appreciated, preferred over global texture maps, and perceived as being realistically rendered.

As for the capture process, further testing could be led to evaluate the extent to which supplementing the rotations with additional captures in translation could help 3D reconstruction by consolidating parallax. Several different angle layouts (both cylindrical and spherical) could also be tested, and the resulting reconstructions compared. Furthermore, coverage at the zenith could be improved by adding cameras facing upwards. Other paths for improvement also include using a battery instead of the current power supply, and switching to higher-quality cameras.

Additionally, future works could reproduce the study with high-dynamic-range (HDR) data, by capturing multiple exposures and rendering the data on an adapted display or using a tone-mapping operator. Indeed, using HDR data may be significant, because photographs with a limited dynamic range may cause areas with highlights to be oversaturated and may thus be the source of noticeable artifacts during rendering. In this way, it could be interesting to lead a user study investigating the extent to which viewers perceive fewer artifacts when using HDR data as input. Future research could also focus on the choice of a tone-mapping operator, which could be selected either using objective criteria or based on user study results (to obtain a perceptive-driven metric).

6. Conclusions

In this paper we presented our development of a custom solution for multi-view capture, which we applied in a museum context to investigate the impact of viewing volume size and view-dependent effects on participants' perception of presence and visual realism. Our analysis confirms that view-dependent image-based rendering methods can be applied to significantly enhance a scene's visual realism and elicit a stronger sense of presence, although some drawbacks remain that should encourage future work on this subject.

Author Contributions: Conceptualization, G.D.d.D., A.P., and J.T.; methodology, G.D.d.D. and A.P.; software, G.D.d.D. and J.T.; validation, G.D.d.D. and A.P.; formal analysis, G.D.d.D.; investigation, G.D.d.D. and A.P.; resources, A.P.; data curation, G.D.d.D. and A.P.; writing—original draft preparation, G.D.d.D.; writing—review and editing, G.D.d.D., A.P., and J.T.; visualization, G.D.d.D.; supervision, A.P.; project administration, A.P.; funding acquisition, A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a PhD Grant from MINES ParisTech, PSL University for the time period 2017–2020, as well as a grant from French National Research Agency CARNOT ("Lightfield" project).

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the HSE officer of MINES ParisTech, PSL University on 5 November 2020.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The study data is openly available and can be accessed from the "data" folder of the repository at the following address: https://github.com/caor-mines-paristech/colibri-vr (accessed on 23 June 2021).

Acknowledgments: We express our warmest thanks to the team of the MINES ParisTech Mineralogy Museum, and in particular to Didier Nectoux, Eloïse Gaillou, and Michel Lechartier, for giving us access to the museum and helping to capture the photographs used for this study. We would also like to thank the reviewers for helping us to improve the final version of the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Overbeck, R.S.; Erickson, D.; Evangelakos, D.; Pharr, M.; Debevec, P. A System for Acquiring, Processing, and Rendering Panoramic Light Field Stills for Virtual Reality. *ACM Trans. Graph.* **2018**, *37*, 197:1–197:15. [CrossRef]
2. Bertel, T.; Yuan, M.; Lindroos, R.; Richardt, C. OmniPhotos: Casual 360° VR Photography. *ACM Trans. Graph.* **2020**, *39*, 266:1–266:12. [CrossRef]
3. Buehler, C.; Bosse, M.; McMillan, L.; Gortler, S.; Cohen, M. Unstructured Lumigraph Rendering. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, Los Angeles, CA, USA, 12–17 August 2001; ACM: New York, NY, USA, 2001; pp. 425–432. [CrossRef]
4. Hedman, P.; Ritschel, T.; Drettakis, G.; Brostow, G. Scalable Inside-out Image-based Rendering. *ACM Trans. Graph.* **2016**, *35*, 231:1–231:11. [CrossRef]
5. Broxton, M.; Flynn, J.; Overbeck, R.; Erickson, D.; Hedman, P.; DuVall, M.; Dourgarian, J.; Busch, J.; Whalen, M.; Debevec, P. Immersive Light Field Video with a Layered Mesh Representation. *ACM Trans. Graph.* **2020**, *39*, 86:1–86:15. [CrossRef]
6. de Dinechin, G.D.; Paljic, A. From Real to Virtual: An Image-Based Rendering Toolkit to Help Bring the World Around Us Into Virtual Reality. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020. [CrossRef]
7. Serrano, A.; Kim, I.; Chen, Z.; Di Verdi, S.; Gutierrez, D.; Hertzmann, A.; Masia, B. Motion parallax for 360° RGBD video. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 1817–1827. [CrossRef]
8. Schwind, V.; Knierim, P.; Haas, N.; Henze, N. Using Presence Questionnaires in Virtual Reality. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; ACM: New York, NY, USA, 2019; pp. 1–12. [CrossRef]
9. Slater, M. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 3549–3557. [CrossRef] [PubMed]
10. Weech, S.; Kenny, S.; Barnett-Cowan, M. Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Front. Psychol.* **2019**, *10*. [CrossRef] [PubMed]
11. Zhao, J.; LaFemina, P.; Carr, J.; Sajjadi, P.; Wallgrün, J.O.; Klippel, A. Learning in the Field: Comparison of Desktop, Immersive Virtual Reality, and Actual Field Trips for Place-Based STEM Education. In Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Atlanta, GA, USA, 22–26 March 2020; pp. 893–902. [CrossRef]
12. Sayyad, E.; Sen, P.; Höllerer, T. PanoTrace: Interactive 3D Modeling of Surround-view Panoramic Images in Virtual Reality. In Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden, 8–10 November 2017; ACM: New York, NY, USA, 2017; pp. 32:1–32:10. [CrossRef]
13. He, L.; Li, H.; Xue, T.; Sun, D.; Zhu, S.; Ding, G. Am I in the Theater?: Usability Study of Live Performance Based Virtual Reality. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology, Tokyo, Japan, 28 November–1 December 2018; ACM: New York, NY, USA, 2018; pp. 28:1–28:11. [CrossRef]
14. Hedman, P.; Alsisan, S.; Szeliski, R.; Kopf, J. Casual 3D Photography. *ACM Trans. Graph.* **2017**, *36*, 234:1–234:15. [CrossRef]
15. Pozo, A.P.; Toksvig, M.; Schrager, T.F.; Hsu, J.; Mathur, U.; Sorkine-Hornung, A.; Szeliski, R.; Cabral, B. An integrated 6DoF video camera and system design. *ACM Trans. Graph.* **2019**, *38*, 216:1–216:16. [CrossRef]
16. Guttentag, D.A. Virtual reality: Applications and implications for tourism. *Tour. Manag.* **2010**, *31*, 637–651. [CrossRef]
17. Waechter, M.; Beljan, M.; Fuhrmann, S.; Moehrle, N.; Kopf, J.; Goesele, M. Virtual Rephotography: Novel View Prediction Error for 3D Reconstruction. *ACM Trans. Graph.* **2017**, *36*, 8:1–8:11. [CrossRef]
18. Witmer, B.G.; Singer, M.J. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence Teleoperators Virtual Environ.* **1998**, *7*, 225–240. [CrossRef]
19. Slater, M.; Usoh, M.; Steed, A. Depth of Presence in Virtual Environments. *Presence Teleoperators Virtual Environ.* **1994**, *3*, 130–144. [CrossRef]
20. Schönberger, J.L.; Frahm, J.M. Structure-from-Motion Revisited. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4104–4113. [CrossRef]
21. Jakob, W.; Tarini, M.; Panozzo, D.; Sorkine-Hornung, O. Instant Field-aligned Meshes. *ACM Trans. Graph.* **2015**, *34*, 189:1–189:15. [CrossRef]

Short Biography of Authors

Grégoire Dupont de Dinechin Grégoire Dupont de Dinechin recently completed his PhD work under the supervision of Alexis Paljic at the Centre for Robotics of MINES ParisTech, PSL University (Paris, France). He holds a PhD in Computer Science (2020) and an Engineer's degree (2017). His research interests include immersion and interaction using consumer-grade virtual reality (VR) and augmented reality (AR) devices, realistic 3D reconstruction of real-world locations using image-based modeling and rendering, and real-time computer graphics.

Alexis Paljic Alexis Paljic is an associate professor at the Center for Robotics of MINES ParisTech, PSL University (Paris, France). He holds an accreditation to supervise research (2018), a PhD in Computer Science (2004), a Master's degree in Physics (1998) and Robotics (1999) from Université Pierre et Marie Curie, Paris Sorbonne Université. He was formerly a researcher in Virtual Reality for Airbus and Stellantis. He now leads the VR and AR group at MINES ParisTech that he joined in 2008. His core research interests are 1)

the conditions of realistic user response and behavior in VR and 2) immersion and interaction for wide public uses of VR in cultural heritage, cinema, museums and performing arts. He also does research on context dependent adaptive AR using machine learning. He is a member of the IEEE Computer Society.

Jonathan Tanant Jonathan Tanant is a French software engineer and researcher with a Master's degree in Music from the University of Paris 8, followed by a Master's degree (2002) in Video Games and Interactive Medias at CNAM / IRCAM in Paris. His areas of expertise include virtual reality and 3D, imaging systems, video mapping, DSP, hardware, interactive installations and simulation. He works with museums, agencies, art schools, artists and research labs.