



# Natural Interaction in Virtual Reality for Cultural Heritage

Riccardo Galdieri<sup>(✉)</sup> and Marcello Carrozzino

Institute of Communication, Information and Perception Technologies,  
Perceptual Robotics Laboratory, Scuola Superiore Sant'Anna, Pisa, Italy  
{r.galdieri,m.carrozzino}@santannapisa.it

**Abstract.** Now that virtual reality has finally become a customer ready product, museums can use this new mean to enhance their exhibitions. The main problem however is that such a tool was not thought for casual users, and to adapt this new technology to short experiences such as the ones museums could provide, it is necessary to reduce the adaptation time to the new mean. In this paper, we discuss how removing physical controllers in favour of visually-tracked virtual hands could significantly reduce the time needed by casual users to adapt to new experiences, underlying the current technological limitations both in terms of technology and design.

**Keywords:** Human Computer Interaction · Virtual reality · Cultural heritage  
Interaction metaphors · Natural interaction

## 1 Introduction

After many decades of incubation, fully immersive virtual reality (VR) has finally become a customer-ready technology. It is not so hard to imagine how such a new way of living surrounding spaces could be used to enhance interaction and fruition of virtual worlds, and many different fields, such as Industrial manufacturing, medicine and entertainment, are adopting these new technologies to improve their products. Despite some initial hesitation, museology and humanities disciplines in general are catching up with this major technological breakthrough, developing dedicated soft-ware to enhance the way in which the public interacts with cultural heritage.

As it often happens with new technologies, in these early stages VR is still far from expressing its full potential. Amongst the remaining problems, the lack of natural interaction within the simulated environments is one of the hardest to solve. Major selling companies ship their head mounted displays (HMDs) with fully tracked controllers, but gameplay interaction is still based on button clicking. This situation is not ideal for casual users such as the ones that museums have, and the time needed by these people to learn new interaction metaphors with controllers could significantly affect their overall enjoyment. Different contexts have different needs, and the inter-action metaphors must be designed in order to produce the best compromise between inter-action, presence, enjoyment, learning and fatigue.

Building full hands tracking in VR would be an important breakthrough: natural interaction would speed up the adaptation process for casual users while increasing the

overall perceived immersion. Unfortunately there is still a conceptual, rather than technological, problem we need to solve. What is keeping real hands out of VR, regardless of the technical implementation, is that virtual and real hands belong to different systems that have different constraints, and an action can be both possible and impossible at the same time when translated from a system to the other. For instance, the surrounding space can be perceived as empty in one system but can also be blocked in the other, and when an action performed in the free space is translated to the other world it creates a logical conflict to the scene where the action was not allowed in the first place, resulting in a loss of presence. When the empty space is the simulation, the risk is to hit objects in the real world, and when the empty space is reality, simulated hands can interpenetrate objects in the simulated world, causing non-realistic behaviours.

In this paper we will discuss how to build natural interaction in mono-user immersive controller-free experiences for cultural heritage applications, introducing a test case scenario currently under development. After a summary of the theoretical background in Sect. 2, in Sect. 3 the current state of the art technologies for natural interaction in VR will be explored and current limitations will be exposed. In Sect. 4 an experiment currently under development to test hands free interaction will be presented together with some expected results, before to draw conclusions in Sect. 5.

## 2 Theoretical Background

### 2.1 Human Computer Interaction

As human beings, the decisions we take are based on what our senses perceive from the environment. It is therefore important to find a way to feed our sensory apparatus as much as possible in VR, so that our actions can still be based on our perceptions. This is why, when the first home computers came out decades ago, it was important to study users' abilities to interact with these new machines in the smoothest possible way.

The first studies in the so called human-computer interaction (HCI) field, a name that was popularized by Stuart Card in 1983 [1], are dated back to 1976 [2]. During its infancy, HCI research focused on simple interactions such as moving the cursor around the screen: early studies used Fitts' law to measure accuracy with different hardware such as the mouse, trackball, joystick, touchpad, helmet-mounted sight, and eye tracker [3]. With time, HCI evolved from being an engineering problem to an interdisciplinary field [4], benefitting from studies in Psychology [5], cognitive studies [6], and even memory studies [7].

As pointed out by many researches, HCI benefits by a nature-driven approach [8, 9]. Being these interactions always artificial to a certain degree, it was necessary to create some metaphors to mimic a real behaviour in a three-dimensional space [10], the so called interaction metaphors. Through these interactions, it is easier for the public to interact with new environments without any domain specific knowledge or acclimatization programme, by translating their previous knowledge to the new situation.

## 2.2 Virtual Reality and Hand-Pose Recognition

Historically speaking, in the early stages of virtual reality definitions tended to be strictly related to hardware constraints, categorizing VR based on the different hardware types in use [11]. What those definitions lacked, according to Steuer, was a more human-focussed approach, he therefore proposed a new definition based on the key concepts of presence and telepresence [12], allowing desktop applications to be considered virtual reality even without dedicated hardware. According to Slater, the definition of presence was still too broad and somehow confusing, proposing to categorize VR based on immersion, meant as objective level of sensory fidelity, and presence, which refers to a subjective psychological response [13–15].

With the exponential growth of desktop VR, a wide range of hardware technologies has been released to support and enhance virtual experiences. Among these, head mounted displays (HMD) and non-invasive cameras have attracted a lot of attention, especially in the academic field. In regards of HMD, they have been used for a wide range of topics, including phobias treatment [16], anxiety [17], and education [18], while controller free interaction has been used in scenarios such as Stroke rehab [18], Sign Language recognition [20, 21], surgery [22] and data visualization [23]. Even though these two technologies are widely used in research, only a few experiments have been carried out with active combinations of them [24], and even less seem to address the problem of physically accurate interaction [25]. In one case, given the high efficiency of native controllers shipped with VR, natural interaction has even been defined as “obsolete” [26].

## 2.3 On Gesture Recognition and Interaction

When discussing hands interaction in virtual worlds, there are two different topics that must be taken into account: pose recognition and interaction. Despite being not mutually exclusive, it is important not to consider them as synonyms, as the first topic studies how to identify the current hands’ position in real world and the second topic is interested in understanding how acquired hands can be used to interact with a digital scene [27].

As regards hands position recognition in a three-dimensional space, the two main devices that can perform reliable recognition without haptic interfaces are the Leap Motion and Microsoft Kinect. Leap Motion software return a pre-rigged fully animated mesh of both hands, with advanced API to use the acquired information in custom environments applications. Despite being tested periodically [28, 29], its tracking software is updated almost on monthly basis, and accuracy tests are outperformed most of the times. Also, as proved by Marin [31, 32], Leap Motion results can be further improved by using machine learning algorithms. On the other side, Microsoft Kinect is way more extensible and programmable but it does not provide any hands identification tool. Nevertheless, it has successfully been used to do perform hand gesture recognition [33, 34].

## 2.4 Museums and Technology

While it is commonly believed that museums are still reticent when it comes to apply technology to exhibitions [30], this tendency has been proven false in recent years [35]. The first milestone in this direction was the creation of the International conference on hypermedia and Interactivity in Museums in 1991 (ICHIM), followed by Museums and the Web established in 1997.

In that period the idea of museums as static exhibitions of art and history was drifting towards the idea of interactive places where people were not passive to their surroundings, but could enhance their experience through new interactive tools [36]. The role of the museum itself was questioned, arguing that museums should not be passive to information, but have an active role in promoting culture and research like other media [37, 38].

As regards user experiences in the so called “Virtual Museums”, defined by the International Council of Museums (ICOM) as “A non-profit, permanent institution in the service of society and its development, open to the public, which acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment.” (ICOM, 2007), it has been proven that the usage of virtual tools to enhance exhibitions does not affect users’ enjoyment nor the learning experience in any way [39]. As a matter of fact, it is quite the opposite. Studies have shown that using technology to customize the way guests explore a museum could improve the overall level of satisfaction [40, 41].

## 3 Background Material

When designing a virtual application for cultural heritage, it is important to keep two elements in mind: the maximum number of simultaneous users and their technological background.

Talking about big audiences, museums want to have as many people as possible to try to enjoy the virtual experience. This leads to an important consequence: unless the application allows many users to control the application simultaneously, all the interaction will be performed by one user at a time with all the other being spectators.

The interaction mean has therefore to be designed to be interactive for one user only, while it has to display data to many. While this is the common case for tools such as CAVE and interactive kiosks, fully immersive VR represents a harder challenge for museums. Given the more immersive nature of the technology, headset users expect a higher degree of interaction with the environment. By default, this interaction is performed through standard controllers in two ways: they can have either have one single action to be performed with a button, which is easy to understand and perform, or a rather complex system of interaction that would require users to learn in advance. For this reason, building a controller free interaction could benefit both immersion and presence, increasing the degree of interactivity while removing the needs of previous knowledge, and speed up the usage time by a significant factor.

While Microsoft Kinect is a valid option for hands tracking acquisition in controlled environments, in a more unsupervised space it could be better to use a shorter-range tool like the Leap Motion. Given the high accuracy that can be reached with it, the consequent step is to blend its data with a fully immersive world. Leap Motion pose data has been used to perform gesture recognition – meant as the interpretation of human gesture – but this data has rarely been used to perform real time interaction with a fully immersive virtual reality system. The main reason for this is realism. Both worlds have physical constraints, but while real world laws cannot be changed, virtual environments' simplified physics interactions are not capable of handling each possible scenario, and when real actions are translated it often happens that the result falls outside the simulated physical model. Something simple like grabbing a glass bottle proves to be a challenge in virtual reality, as physical engines are extremely sensible to mesh interpenetration and are not capable of handling events that, in their own environment would not be allowed, such as having a hand narrow a rigid body.

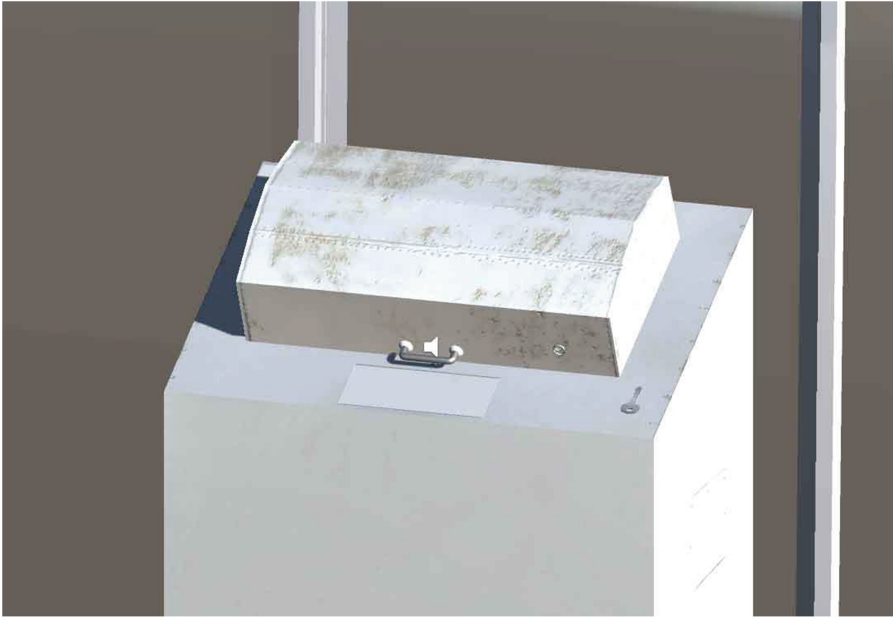
In June 2017, Leap Motion released an API to tackle this problem. This new software puts himself between the hand poses obtained by the Leap Motion and the 3D engine physics simulation, disabling any collision calculation when the hands are performing a physically inaccurate action. While this approach works from a physical point of view – by preventing the engine from carrying out wrong calculations – it still breaks the perception of reality within the simulation, as it allows the hands to interpenetrate the scene objects without any response. Some applications prefer to limit the degree of visual feedback in the simulation by always showing a physical response to the users, but this creates a mismatch between the perceived hand position and the visual hand. Given the purpose of this project to investigate real hands interaction in VR, the idea of having a mismatch between perception and visualization was discarded, and the compromise offered by Leap Motion accepted and noted.

## 4 The Experiment

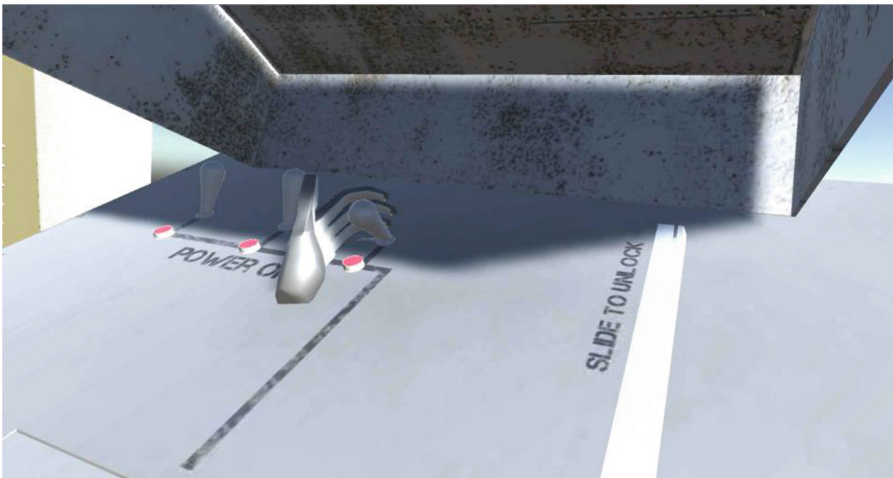
As already discussed, hands free interaction in VR is a rather unexplored field. We designed an experiment to understand how different interactions can be perceived as natural by a variegated audience, hoping to find a preliminary way to categorize single-handed actions. The ideal outcome would be to find common features among gestures that could potentially be used in future natural interaction metaphors design.

To test the previous assertions, we're currently developing a game-like test case scenario application in immersive VR, where users will be required to perform a series of actions on a console in order to unlock a new room with a piece of art in it. The sequence of actions, at the current state of the application, is as follows: in the first instance, users must grab and hold a key, which they must put in a lock. Once inserted, the key must be rotated in order to unlock the case. When unlocked, users will be required to pull it up to access a control panel hidden below. At this point, once a series of switches is activated and the panel powered, a secret box opens and a card is found. The card must be grabbed like the key and slid through a rail. Once the card arrives at the end, the door unlocks and the prize is revealed. Before the simulation starts, the

operator will be able to choose whether if he wants to activate a pre-recorded speech that guides the users over the different challenges, or to keep it quiet and leave them to the task (Figs. 1 and 2).



**Fig. 1.** The closed hatch.



**Fig. 2.** The virtual hand interacting with the environment. Materials are temporary.

There will be two evaluation metrics for this challenge: time and accuracy. The demo will be monitoring both the overall time needed to access the room and the time needed to complete each single task. If a user takes a significantly longer time but just one attempt to perform a subtask, it means that he was not able to understand what he was required to do in the first place, and the metaphor was not clear. On the other side, if he attempts many times and fail, it could mean that the manipulations were not easy enough to be executed in VR rather than in reality, bringing up further discussion on both technology and design.

A control group has also been created in order to compare how the usage of controllers instead of hands could affect performances. While receiving the same instructions and the same support throughout the tests, the control group will use a single button to interact with the scene instead of touching, grabbing and pulling with his hands.

#### 4.1 Expectations

There are some results that we are expecting, given the discussion above. First and most important, interaction metaphors deriving from different physical interactions will have different degrees of success. In real life, it is almost impossible to insert a key without scratching around the hole, and even though the application gives users some margin, by allowing the key to fit even if not perfectly positioned, they won't be aware of this facilitation and will try to achieve a perfect result.

In addition, the overall time needed to complete each single subtask must be crosschecked with the number of attempts to perform an action. For instance, we might have a small number of users who try to turn the switches on and off in order to repeat the animation. If that is the case, the overall completion time data will be less relevant than in other cases. This behaviour must be noted during the data analysis phase, and data-wise, noisy experiences must be ignored if possible.

Another crucial factor to consider is the size of the objects people will interact with. Every object should have a significant size in order to be physical accurate, and while there is no precise measurement on what the minimum suggested size could be, it has been noticed that small objects such as a key could be subject to problems if too small. For this reason, all graspable objects in the scene are bigger than their real life matches. While it may not seem a significant factor in achieving the desired interaction, as the scale is not so significantly different, further investigations should be made in order to exclude possible score contamination by the scale difference.

Generally speaking, we expect the overall interaction time not to be significantly different among participants. We do however expect some people to take a longer period to adapt, meaning that they will spend more time than others completing the first challenge. As regards the control group, we expect them to score less mistakes in grabbing challenges, while we expect them to take longer rotating the key and clicking the switches. Moreover, current state of the art applications for VR provide vibration as force feedback during interactions. We decided not to provide any, to keep the two interaction means as equal as possible.

## 5 Conclusions

The experiment we are currently setting up only concerns simple interactions, and purposely avoid complex gestures like throwing, pulling, squeezing or any two hands interaction. While the problem of hands interaction is easy to define, we are far from even scratching the surface of how to handle such complexity.

Now that the quality of virtual reality has reached such a high level of interactivity, it is time to start thinking about immersive virtual experiences as a whole and not as a cluster of problems that can be solved individually. The collision of real and simulated worlds is far too complex, and without an accurate evaluation of colliding aspects, it will be impossible to reach the level of interaction that is expected in a realistic simulation.

Museums could and should be part of this challenge. Given their extremely wide audience, specific interactions must be designed to create immersive controller-free in VR, and general guidelines will not be exhaustive enough to be borrowed and applied to cultural heritage application. Hands interaction among exhibitions could make the difference between being passive to history and actively be part of it.

**Acknowledgements.** This paper is supported by EU Horizon 2020 research and innovation programme under grant agreement No 692103, project eHERITAGE (Expanding the Research and innovation capacity in Cultural Heritage Virtual Reality Applications).

## References

1. Card, S.K., Newell, A., Moran, T.P.: *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc., Hillsdale (1983)
2. Carlisle, J.H.: Evaluating the impact of office automation on top management communication. In: *Proceedings of the June 7–10, 1976, National Computer Conference and Exposition*, pp. 611–616. ACM, June 1976
3. MacKenzie, I.S.: Fitts' law as a research and design tool in human-computer interaction. *Hum. Comput. Interact.* **7**(1), 91–139 (1992)
4. Jaimes, A., Sebe, N.: Multimodal human-computer interaction: a survey. *Comput. Vis. Image Underst.* **108**(1–2), 116–134 (2007)
5. Newell, A., Card, S.K.: The prospects for psychological science in human-computer interaction. *Hum. Comput. Interact.* **1**(3), 209–242 (1985)
6. Hollan, J., Hutchins, E., Kirsh, D.: Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **7**(2), 174–196 (2000)
7. Nass, C., Brave, S.: Emotion in human-computer interaction. In: *The Human-Computer Interaction Handbook*, pp. 94–109. CRC Press (2007)
8. Villaroman, N., Rowe, D., Swan, B.: Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor. In: *Proceedings of the 2011 Conference on Information Technology Education*, pp. 227–232. ACM, October 2011
9. Francese, R., Passero, I., Tortora, G.: Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pp. 116–123. ACM, May 2012



10. Mackay, W.E., Fayard, A.L.: HCI, natural science and design: a framework for triangulation across disciplines. In: Proceedings of the 2nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, pp. 223–234. ACM, August 1997
11. Coates, G.: Program from Invisible Site—a virtual sho, a multimedia performance work presented by George Coates Performance Works. San Francisco, CA (1992)
12. Steuer, J.: Defining virtual reality: dimensions determining telepresence. *J. Commun.* **42**(4), 73–93 (1992)
13. Slater, M.: Measuring presence: a response to the Witmer and Singer presence questionnaire. *Presence* **8**(5), 560–565 (1999)
14. Slater, M.: A note on presence terminology. *Presence Connect* **3**(3), 1–5 (2003)
15. Bowman, D.A., McMahan, R.P.: Virtual reality: how much immersion is enough?. *Computer* **40**(7) (2007)
16. Carlin, A.S., Hoffman, H.G., Weghorst, S.: Virtual reality and tactile augmentation in the treatment of spider phobia: a case report. *Behav. Res. Ther.* **35**(2), 153–158 (1997)
17. Wiederhold, B.K., Wiederhold, M.D.: Virtual Reality Therapy for Anxiety Disorders: Advances in Evaluation and Treatment. American Psychological Association, Washington DC (2005)
18. Freina, L., Ott, M.: A literature review on immersive virtual reality in education: state of the art and perspectives. In: The International Scientific Conference eLearning and Software for Education, vol. 1, p. 133. “Carol I” National Defence University, January 2015
19. Bassily, D., Georgoulas, C., Guettler, J., Linner, T., Bock, T.: Intuitive and adaptive robotic arm manipulation using the leap motion controller. In: Proceedings of ISR/robotik 2014; 41st International Symposium on Robotics, pp. 1–7. VDE, June 2014
20. Chuan, C.H., Regina, E., Guardino, C.: American sign language recognition using leap motion sensor. In: 2014 13th International Conference on Machine Learning and Applications (ICMLA), pp. 541–544. IEEE, December 2014
21. Potter, L.E., Araullo, J., Carter, L.: The leap motion controller: a view on sign language. In: Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration, pp. 175–178. ACM, November 2013
22. Harrison, B., et al.: Through the eye of the master: the use of Virtual Reality in the teaching of surgical hand preparation. In: 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), pp. 1–6. IEEE, April 2017
23. Donalek, C., et al.: Immersive and collaborative data visualization using virtual reality platforms. In: 2014 IEEE International Conference on Big Data (Big Data), pp. 609–614. IEEE, October 2014
24. Blaha, J., Gupta, M.: Diplopia: a virtual reality game designed to help amblyopics. In: 2014 IEEE Virtual Reality (VR), pp. 163–164. IEEE, March 2014
25. Lee, P.W., Wang, H.Y., Tung, Y.C., Lin, J.W., Valstar, A.: TranSection: hand-based interaction for playing a game within a virtual reality game. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 73–76. ACM, April 2015
26. Hilfert, T., König, M.: Low-cost virtual reality environment for engineering and construction. *Vis. Eng.* **4**(1), 2 (2016)
27. Mitra, S., Acharya, T.: Gesture recognition: a survey. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **37**(3), 311–324 (2007)
28. Weichert, F., Bachmann, D., Rudak, B., Fisseler, D.: Analysis of the accuracy and robustness of the leap motion controller. *Sensors* **13**(5), 6380–6393 (2013)
29. Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., Sodnik, J.: An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. *Sensors* **14**(2), 3702–3720 (2014)

30. Schweibenz, W.: The "virtual museum": new perspectives for museums to present objects and information using the internet as a knowledge base and communication system. *ISI* **34**, 185–200 (1998)
31. Marin, G., Dominio, F., Zanuttigh, P.: Hand gesture recognition with leap motion and kinect devices. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 1565–1569. IEEE, October 2014
32. Marin, G., Dominio, F., Zanuttigh, P.: Hand gesture recognition with jointly calibrated leap motion and depth sensor. *Multimedia Tools Appl.* **75**(22), 14991–15015 (2016)
33. Tang, M.: Recognizing hand gestures with microsoft's kinect. Department of Electrical Engineering of Stanford University, Palo Alto (2011)
34. Yao, Y., Fu, Y.: Contour model-based hand-gesture recognition using the Kinect sensor. *IEEE Trans. Circ. Syst. Video Technol.* **24**(11), 1935–1944 (2014)
35. Styliani, S., Fotis, L., Kostas, K., Petros, P.: Virtual museums, a survey and some issues for consideration. *J. Cult. Heritage* **10**(4), 520–528 (2009)
36. Pearce, S.M.: Thinking about things. In: Pearce, S.M. (ed.) *Interpreting Objects and Collections*, pp. 125–132. Routledge, London (1994)
37. Silverstone, R.: The medium is the museum: on objects and logics in times and spaces. In: Durant, J. (ed.) *Museums and the Public Understanding of Science*, pp. 34–42. The Science Museum, London (1992)
38. Macdonald, S.: Theorizing museums: an introduction in *Theorizing Museums* (Doctoral dissertation, ed. S. Macdonald and G. Fyfe, Oxford: Blackwell Publishers) (1996)
39. Pierdicca, R., Frontoni, E., Zingaretti, P., Sturari, M., Clini, P., Quattrini, R.: Advanced interaction with paintings by augmented reality and high resolution visualization: a real case exhibition. In: De Paolis, L.T., Mongelli, A. (eds.) *AVR 2015. LNCS*, vol. 9254, pp. 38–50. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-22888-4\\_4](https://doi.org/10.1007/978-3-319-22888-4_4)
40. Pagano, A., Armone, G., De Sanctis, E.: Virtual Museums and audience studies: the case of "Keys to Rome" exhibition. In: *Digital Heritage 2015*, vol. 1, pp. 373–376. IEEE, September 2015
41. Choi, H.S., Kim, S.H.: A content service deployment plan for metaverse museum exhibitions—centering on the combination of beacons and HMDs. *Int. J. Inf. Manag.* **37**(1), 1519–1527 (2017)

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

