Aspects and Criticalities of the Fruition in Subjective of the Digital Space: the 'First Person View'

Graziano Mario Valenti, Alessandro Martinelli

Abstract

The research illustrated here stems from a critical reflection on the theme of 'first person view', as defined in the context of interactive virtual digital three-dimensional representations. A reflection arising from the perceptive, communicative, illusory and narrative wonder typical of seventeenth-century architectural perspectives, in contrast with the more slender –compared to the forces at play– suggestive efficacy of the typical scenarios of today's digital immersive perspective enjoyment. Through an analytical phase of observation of the natural experience of seeing and moving in the space of the real world, an attempt was made to identify some macroscopic critical issues and functional gaps, now deliberately or individually neglected in the implementation of the use of digitally simulated three-dimensional space. In this way, three macro areas of potential intervention have been identified and described in their salient features and, for each of them, some emerging qualitative factors have been focused, the care of which can perfect the experience of three-dimensional fruition and, consequently, the cognitive activity of the digital virtual space, as well as of the information contained therein. The final goal of the research is in fact to perfect and personalize the virtual fruition of digital museum spaces.

Keywords: first person view, perspective, perception, real-time 3D, virtual museums

Introduction

The museum installations and the cognitive paths designed within the cultural heritage present, with increasing frequency, spaces intended for the fruition of digital models by means of headsets, which drop the viewer in virtual, augmented or mixed reality.
Despite the extraordinary innovation and technological evolution that can be found both in the production of headsets and in the representation of digital models, the direct experience of real space is so complex, and sometimes even personal, that each digital simulation that aims to effectively reproduce the sensorial impact and the innate cognitive goal, results in an 'other', different experience. For this reason, it is frequent to resort to alternative solutions, which avoid facing the simulation of reality in its total complexity. In the context of videogames, for example, it is usual to drop the visitor into a virtual space with an unreal character, at times imaginative if not even 'cartoon', where the experience itself is worth: an immersion in a large digital playground, whose nature is never to be associated, much less to be compared, with the experience in the real world [Schwartz 2006] (fig. 1).
This experiential-communicative paradigm is the foundation of the information aids produced in the field of gamification which, in general, seem to prefer motor perceptual learning, rather than a symbolic reconstructive one [Antinucci 1998].
The narrative emphasis that derives from this approach is certainly consistent with the development of a video

game. However, it leaves some perplexity when it is extended as an optimal and universal solution for any other information activity. The digital use of museums, for example, carried out mostly in an undifferentiated form for the target user, is among the activities most at risk of this applicative singularity [Modena 2019].
Since the early 1990s, with the birth of the World Wide Web and the subsequent opening of the Internet to the vast public, the methods of sharing digital information relating to cultural heritage have taken on different forms, naturally of increasing complexity: from the most simple hypertextual/hypermedial product, to the 360-degree photographic sequence of the exhibition space, up to the most advanced interactive navigation experiences in virtual three-dimensional spaces, where the user is equipped with haptic and stereoscopic interfaces.
However, by observing and above all experiencing the state of the art directly, the reduced pervasiveness of information communication in a three-dimensional context appears to remain, a mode frequented more out of curiosity about technology than for actual communicative surplus value. There may be many interpretative keys aimed at identifying the reasons for the recurrent oscillation, between enthusiasm and disappointment, which characterizes the interest in three-dimensional digital operations. Reasons that in substance are all attributable to the consideration that in digital simulations it is possible to focus and 'increase' only some aspect of experience and communication, which however manifests itself in a sensorial context that is still too poor compared to that determined by real experience and with which it inevitably ends up confronting itself. The digital experience therefore remains, even today, an integrative activity and is far from being a replacement.
In analyzing what were the sensory gaps that still make the two experiences incomparable [Paes 2017] and seeing if there was space to provide a useful research contribution to fill them, the areas of study appeared numerous and mostly affordable only in an interdisciplinary way, requiring to put in system skills in geometry, optics, computer science, anatomy, neurology, psychology, and so on. Within this scenario, our attention was therefore focused on a central aspect of the problem and precisely of research in the field of drawing. A qualitative factor of three-dimensional fruition that, compared to the role it covers, appeared all too neglected: the knowledge of the digital space and the information contained in it in the mode defined as 'first-person view' (fig. 2). A first substantial survey was therefore initiated on this qualitative factor, followed by an analytical reflection, aimed at identifying criticalities and potentialities: operational, perceptive and cognitive.

Fig. 1. The 'cartoon' representation of reality in Fortnite and Minecraft.

The 'first person view'

The 'first person view' is a conceptual model of representation: a relationship between observed reality and observer, where the latter has an active participation, as the protagonist of the scene and not as a mere spectator. It is also a spatial relationship understood in a geometric sense, proportional and kinematic between the elements making up the representation, since, in becoming the protagonist of the scene, the observer must assume the same presence and behavior of the subject he identifies. Not surprisingly, the first person view is otherwise known in cinematographic language with the name of 'subjective'.
By drawing on the extensive discussion that history and film criticism has dedicated to this particular way of shooting the scene, two families of influencing factors can be deduced that perfect the narrative effectiveness of the first person view. Representative families of factors, one exogenous and the other endogenous compared to the observer. We can consider exogenous factors those expedients external to the visual phenomenon adopted to communicate and reinforce the idea that the image being enjoyed is the product of a first person view. At the origins of cinema, when there was still no technology and in part even the theoretical knowledge to faithfully reproduce a subjective shot, the illusion of the viewer was emphasized through the visual narration associated with the use of mattes [Eugeni 2020].
For example, initially the protagonist of the scene was filmed in the act of positioning a telescope on the eye and in the following images what he saw in 'subjective' through it was simulated, cutting out the frame with a circular matte (fig. 3). Over time, the visual narrative has been very refined, emphasizing the illusion of subjective sight through the careful framing of parts of the subject's body, in which the viewer had to identify himself (fig. 3). These exogenous factors, to which the appropriate use of sound must certainly also be added –e.g. heartbeat and footsteps during a labored run– are of considerable interest also in the digital virtual environment, but, in the part of the research described here, the attention was based on endogenous factors, which strictly depend on the physiology and behavior of the human body, albeit investigated and reproduced with a reasonable margin of approximation. Within this family of factors, we distinguish three prevalent areas of possible intervention: geometric-projective, physiological-perceptive, kinematic.

Fig. 2. Rigging of the camera to obtain the first person view (FPV). Unreal 5.

Geometric-projective field

The representation of digital three-dimensional space, like the analog one, originated from the geometric simplification of the visual phenomenon. A simplification that is stronger the more the elements making up the scene are numerous and complex and the more the image must be formed rapidly in real time. The starting point of this simplification are the principles of *perspectiva artificialis*, which describe with sufficient approximation the projective phenomenon characteristic of a photographic apparatus and which the human being, with a consistent operation of mental interpretation, is now accustomed to associate identical to his own vision.
Of this simplified model we cite, by way of example, some aspects of the approximation: the eye is considered in the abstract form of a point; it is often unique and has a homogeneous sensitivity; therefore, there are neither perspective aberrations, nor peripheral areas of lesser detail; the surface on which the light rays are projected is flat. It should also be remembered that the image that is formed in the mind as a consequence of direct observation of the real world is constructed with the contribution of the complex phenomena of visual perception [Amoruso 2020], while, in the use of a digital image, perception takes over from the constructed image: it therefore collaborates in a minor form, to its definition and to those processes of attenuation-emphasis produced by the cultural and subjective interpretation of the observed space. Processes that could significantly change the color, size, orientation, geometry of what we observe.
We do not want to refer here specifically to the explored question of cultural interpretation of perspective, raised by Panofsky and Gioseffi, which would open a field of

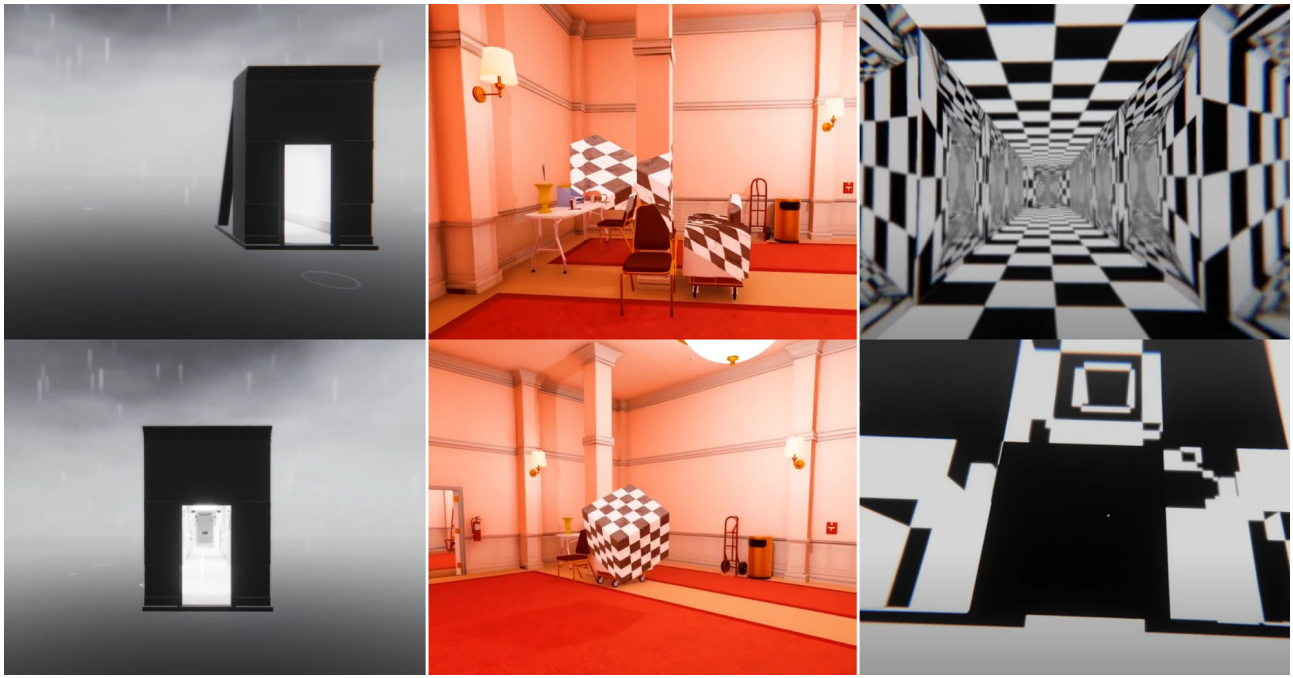
Fig. 3. Use of the circular matte in Ce que l'on voit de mon sixième (F. Zecca 1901); blurred vision of hair in the wind in Notorius (A. Hitchcock 1946) ; from murder to suicide in Spellbound (A. Hitchcock 1945).

conceptual treatment closer to aspects related to visual perception than to geometry; instead, we allude to the ability and habit of the mind to normalize the observed space, as happens for example in the face of small fluctuations in the point of view. Attenuations of the projective alterations of space, which are typical of the natural vision process and which are generally not present in the digital dynamic perspective image, because there are no devices suitable for their control or simply because it has been chosen to neglect them. Consider, for example, how little evident it is, without tracing the thought back to the principles of perspective, to perceive live and on a small scale the possible convergence in vanishing line of the vertical lines, consequent to a consistent zenith rotation of the head upward or downward, and how vice versa this convergence appears instead evident in a two-dimensional image, a digital projection of the same space. Think again of the stability-straightening of the vertical and horizontal orientation of the image perceived in the face of small rotations of the head, compared to what would happen by rotating the camera of a digital representation or even more simply by observing a perspective drawn on a sheet arranged with orientation random in front of us.

The more we manage to immerse the viewer in virtual space, the more the mind is deceived and stimulated to behave as in real space.

Binocular vision, obtainable by wearing modern stereoscopic headsets, attenuates the presence of some of the phenomena described. However, the attention dedicated to the development of this technology and the care in its rigorous use do not seem to be proportional to the significant contribution they could offer in favor of the implementation quality of virtual three-dimensional fruition systems.

Fig. 4. Perspective illusions and multidimensional spaces in Superliminal. The solution of the alterations and perspective, illusions guide the player through the virtual space.



The attention and commitment of the creative operators in the sector of communication of cultural heritage is mostly biased towards the project of information content, while with regard to the tools –specifically in the choice of the methods of use– are often used standard procedures which, except for a few cases of technological experimentation, they are often very simplified simulations of the real phenomenon they want to allude to.
With Vitruvian etymological inspiration from the term 'scenography', but with a precise reference to the geometries that make up the space observed, to this geometric-projective field belong the scenic factors too, which can affect and in particular direct the exploratory path of digital space. In the cognitive appropriation of space achieved through the vision of a film product, the viewer is bound to the movements of the camera, along paths that have been designed by the director and specially characterized by the set designer. In the theatrical field, even if there is no interposition of the shooting equipment between the observer and the observed space, the position of the spectator is generally fixed and, once again, the director and set designer determine what must and can be seen. In interactive digital three-dimensional fruition, on the other hand, the observer can move freely and is able to potentially explore every corner of the scene. In the construction of a virtual museum experience, therefore, the design of the environment cannot be limited to the definition of the exhibition equipment useful for emphasizing the knowledge of the objects on display. Rather, it will be necessary to pay particular attention to the introduction of scenic precautions, which have both limiting and attractive value, in order to contain and direct the user's path in the virtual space [Nielsen 2016] (fig. 4).
With specific reference to the seventeenth-century architectural perspectives, which initiated the reflection on the theme of first-person view, think, for example, of the mastery in the use of anamorphosis, visible in the work attributed to Jean François Niceron (1613-1646), present in the convent of Trinità dei Monti in Rome and concerning Saint John the Evangelist on the island of Pathmos while writing the Apocalypse. A work where the user is encouraged to move from perspective expedients to reach knowledge and make the narrated message their own [Trevisan 2015].
To prevent the user from getting lost or lingering in the digital space, some videogames have used so-called cutscenes, during which the user loses or reduces the autonomy of control of the virtual space navigation and proceeds 'automatically' to a new place of the scene. While helping to keep the pace and dynamics of the game high, the judgments on their use are very discordant: since the more these tricks are present, the more we move away from the quality and benefits of personal experience.

Physiological-perceptive field

The eye is a complex sensory organ with heterogeneous performance in humans. Its digital form, we have said to be mostly simplified, but various physiological factors, should instead be considered with great attention, since they significantly contribute to modifying the projective geometric rigor and communicative pervasiveness of the image. Sev-

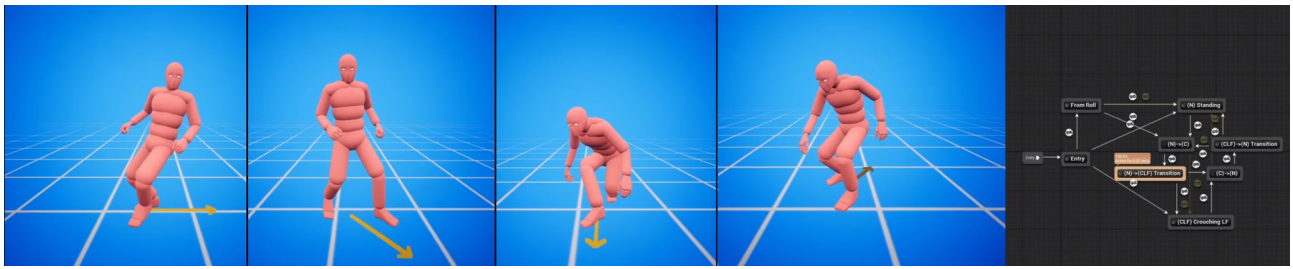
Fig. 5. Examples of images captured in photomode. Polygon: <https://www.youtube.com/watch?v=PZ4jY0LOPVg> (consultato il 12 maggio 2021).

eral of these physiological factors, characteristic of the human eye, also belong to inorganic photographic equipment, perhaps also for this reason they have found greater attention in digital implementation: focus, width and depth of field, blur, sharpness, brightness, contrast. In recent years, the most advanced videogames have begun to offer a particular mode of use called *photomode* (fig. 5). When the player enters this mode, he turns into a virtual photographer and the scene is arrested in an instant. It is not a simple capture of what is projected on the screen, but a real exploration of the immobilized virtual space, with the possibility of controlling the virtual photographic set as if it were a real set. More than a game within a game, this operational opportunity is assuming the important role of vector and training aid in the dissemination of the culture of the image and its communicative power. Thus, a new profession of digital scenery photographer has emerged, which creates another bridge, perhaps closing the circle, between three worlds of digital photography, characterized by shots produced in real life, in a deferred rendering environment and finally in real-time virtual rendering scenarios. If this is the trend that is emerging –as a need– in the field of videogames, it is natural to think that even in the virtual museum fruition we will soon proceed in this direction. The photographic shot of the cultural asset is now already widely desired in the real world and recently mostly granted, given the benefit it brings in the role of promotional image when shared on social networks. It is therefore a question of making the project of fruition of the virtual museum and the possibility of seeing of the user even more complex, studying in detail also this emerging mode of interaction.
Referring to the qualitative factors deriving from considerations on visual perception, the research has identified several of them, during the recognition phase. The theme, as expected, turned out to be very wide [Casale 2016] and untraceable in this context. However, there is an aspect that cannot be avoided and that concerns inclusiveness: one of the key challenges of the current research European 'Culture, creativity and inclusive society'. The role of visual perception, understood as a physiological and cultural media, is the tool to be explored, to create personalized digital environments on the heterogeneity of users. Inclusive virtual spaces, where different skills, different cultures, different generations can also access and understand information.

Kinematic field

The third macro area identified concerns kinematics, that is the geometric description of the motion of the organs of vision. The solutions prevalent today use excessive simplifications or complex descriptions of the movement which, however, are customized on particular dynamic activities of the game. In still other cases, there is a tendency to reproduce the camera movements which, although familiar, belong to the expressive repertoire of the cinematographic image and are therefore not characteristic of the personal visual experience. In this type of digital experience, camera movements such as panning, tracking shots, zooming are frequently mixed continuously, undoubtedly making the aesthetic experience of digital use rich but, at the same time, distancing it from reality.
A correct kinematics of the first person view should be based on the study of a motion that necessarily passes through the analysis of the chain of independent move-

Fig. 6. Unreal Advanced Locomotion System V. 4: visual programming of the avatar's movements.



ments that defines the absolute orientation of the visual organs compared to the observed space [Boletsis 2019] (Fig. 6). With reference to the natural movement of the human being, who proceeds by walking in real space observing what surrounds him, the corresponding avatar designed for the virtual space should have a pair of points of view –perspective projection centers– placed in correspondence with the eyes. Their orientation will be determined respectively: by the movement of the eyes, by the orientation of the head and, finally, by the posture of the body that supports the head.

Eyes

The eyes orient themselves in space through movements of various kinds, both voluntary and involuntary. Some are almost imperceptible and have little influence on the representation, others, conversely, are more consistent and not negligible. In the context of involuntary movements, the vestibular-ocular and optokinetic reflexes must certainly be included, both of which are useful for keeping the point collimated by compensating for any movements of the head and body. Instead, belong to voluntary movements those called vergence, which allow the two eyes to collimate a single point, even if moving away (divergence) and approaching (convergence) to our face. For our evaluations, the voluntary saccadic movement plays a fundamental role, which aims to rotate the eyeball, bringing the observed objects to collimate with the central region of the retina, of maximum visual acuity [Sun 2018]. This movement, which would allow an excursion of about 90 degrees, is actually normally used within 20 degrees –for example, about two degrees when walking through the lines of a text–, beyond which one instinctively tends to bring into play the head rotation. On a perceptual level, as can be learned by rotating the eyeballs in a constrained view situation, the changes to the projective product are minimal. It is therefore difficult to perceive projective differences between the saccadic movement carried out in front of a real space compared to the same movement obtained by observing a two-dimensional projection of the same space made on a screen/headset.

Head

The movements of the head, in the digital environment, are generally translated as rotation of the main direction of the perspective view, applied precisely in correspondence with the center of projection. This completely eliminates the phenomenon of parallax, which instead is perceived

Fig. 7. Death Stranding (2019): one of the most advanced digital simulations of the anatomical and motor dynamics of the avatar.

in natural rotation, as it is impossible for us to rotate our head while keeping still what we perceive to be our center of projection. It should also be said that the movements in the digital environment are mostly linear, as are the panoramas, but in reality, when we move the head, the saccadic movement takes its liberties: it pauses and accelerates according to need, transforming the perception of the space in a decidedly different experience than the one tried digitally.

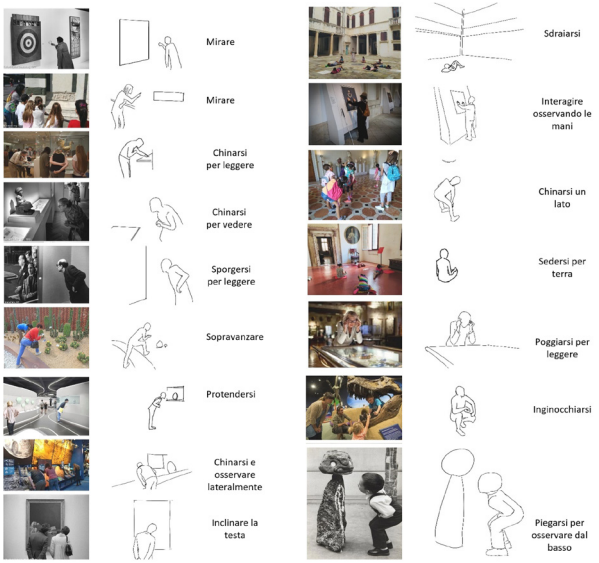
Body

Over time, from the origins of digital three-dimensional use until today, there have been significant improvements in the simulation of postures that have described the movements of the avatar in the virtual environment. If originally the movement was translated as a simple translation from the point of origin to that of destination, today –in the most refined models– it is instead the product of a complex kinematic sequence that affects all the organs of movement in which the body of the avatar is discretized. End of this chain, in the case of binocular vision, is a pair of vision centers located in correspondence with the eyes. The cinematic sequences are identified and reproduced thanks to the study and normalization of data collected through motion capture techniques. Recently, automatic recognition technologies have also spread, conducted with the aid of machine learning procedures, today widely used for face swapping [Nirkin 2019] or physic character control activities [Bergamin 2019]. If, on the one hand, in the panorama of Virtual Reality applications, there are vast motion-capture databases dedicated to the most extreme and demanding physical activities that can be assigned to the avatar (sports, fights, dance etc.), (fig. 7), much less material is available to fully describe and reproduce the slower, more reflective movements characteristic of normality. Only in the field of animation cinema, where the need to best characterize the characters with expressions and emotions of considerable realism has arisen, a greater attention was paid to the detail of normality.
A Part of the research was therefore dedicated to detecting a first group (fig. 8) of movements and postures that people assume when visiting an exhibition space. Movements that, in a subsequent phase, will be digitally implemented and tested.

Conclusions

The analysis of the three thematic macro areas and the related qualitative factors through which it is possible to perfection the experience of use and knowledge of the digital three-dimensional space, clearly shows the need to pursue the search for a new goal of equilibrium, where technological advancement on the one hand and the narrative project on the other find integration at a higher level of complexity. From the analysis of the state of the art, these operational scenarios today appear weakly linked: on the one hand, the technology useful for the use of three-dimensional digital space appears too influenced by a development oriented –obviously– to respond to the demands of the great demand, now from the world of videogames; on the other hand, the designers of the narrative paths useful for accessing the knowledge of three-dimensional space and the information it contains end up welcoming technological aids as a 'dogma' and focusing all their critical attention and innovation only on the subject

Fig. 8. Exemplary extract, in the form of an abacus, of the postures and movements to be digitized, identified by observing the activity of use of real exhibition spaces.



of the narrative. Already the first results of this research show, instead, how vast is the space of intervention between these two operational scenarios and how much it is necessary to strengthen interdisciplinary collaborations in order to increase and consolidate their links. In the geometric-projective field, for example, the relationship between derogation and rule in perspective representation should be further investigated, fueling the experimentation with knowledge from studies on visual perception. In the physiological-perceptive field, the coherence and narrative efficacy of the physiological aspects of human vision should be clarified, codified and tested, and on a strictly perceptive level, investigated the attractive and dissuasive potential of the space useful for directing the visitor's path. Finally, in the cinematic field, if on one front it is necessary to continue to develop the immersive technologies necessary to access the digital space, so that they can offer the recording and control of multiple parameters of human behavior, on the other hand it is appropriate that these parameters are adequately regulated by the designers of the narration, so that digital information can be accessed in ways similar to natural ones. Parameters that more than enumerated, valued and counted, must be mapped as Laura Marcolini suggests [Amoruso 2020].

Authors

Graziano Mario Valenti, Department of History, Representation and Restoration of Architecture, Sapienza University of Roma, grazianomario.valenti@uniroma1.it
Alessandro Martinelli, Department of History, Representation and Restoration of Achitecture, Sapienza University of Roma, alessandro.martinelli@uniroma1.it

Reference List

Amoruso, G. (2020). disegnare con… Laura Marcolini. In *DisegnareCon*, vol. 13, n. 25.

Antinucci, F. (1998). Musei e nuove tecnologie: dov'è il problema? In *Sistemi intelligenti*, vol. X, n. 2, pp. 281-306.

Bergamin, K., Clavet, S., Holden, D., Forbes, J. R. (2019). DReCon: data-driven responsive control of physics-based characters. In *ACM Trans. Graph.*, vol. 38, n. 6, Article 206.

Boletsis, C., Cedergren, J. E. (2019). VR locomotion in the new era of virtual reality: an empirical comparison of prevalent techniques. In *Advances in Human-Computer Interaction*, vol. 2019. <https://doi.org/10.1155/2019/7420781> (accessed 2021, May 12).

Casale, A. (2018). *Forme della percezione: dal pensiero all'immagine*. Roma: Franco Angeli.

Eugeni, R., Guerra, M. (2020). Far sentire la macchina. Appunti sulla soggettiva cinematografica e la teoria dell'enunciazione. In *E/C*, anno XIV, n. 29, pp. 134-144.

Modena, E. (2019). Musei nei videogiochi| Videogiochi nei musei. In *Piano B Arti e culture visive*, n. 4, pp. 83-105.

Nielsen, L. T., et al. (2016). Missing the point: an exploration of how to guide users' attention during cinematic virtual reality, In S. N. Spencer (ed.). *VRST 2016. Atti del 22nd ACM Conference on Virtual Reality Software and Technology*, Monaco, Germania, 2-4 novembre, pp. 229-232. New York: Association for Computing Machinery Inc.

Nirkin, Y., Keller, Y., Hassner, T. (2019). FSGAN: Subject Agnostic Face Swapping and Reenactment, In *IEEE/CVF 2019*. Proceedings International Conference on Computer Vision, Seul, Corea, 27 ottobre-2 novembre, pp. 7184-7193: <https://openaccess.thecvf.com/content_ICCV_2019/papers/Nirkin_FSGAN_Subject_Agnostic_Face_Swapping_and_Reenactment_ICCV_2019_paper.pdf> (accessed 2021, May 12).

Paes, D., Arantes, E., Irizarry, J. (2017). Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems. In *Automation in Construction*, n. 84, pp. 292-303.

Schwartz, L. (2006). Fantasy, realism, and the other in recent video games, In *Space and culture*, n. 9, pp. 313-325.

Sun, Q., et al. (2018). Towards Virtual Reality Infinite Walking: Dynamic Saccadic Redirection. In *ACM Trans. Graph.*, vol. 37, n. 4. <https://doi.org/10.1145/3197517.3201294> accessed 2021, May 12).

Trevisan, E. (2015). Il san Giovanni Evangelista di Jean François Niceron: la scoperta di un'apocalisse dell'Ottica. In G. M. Valenti (a cura di), *Prospettive Architettoniche*, pp. 365-374 Roma: Sapienza Edizioni.