Shaping Virtual Image †

Daniele Rossi

School of Architecture and Design, University of Camerino, 62032 Camerino, Italy; daniele.rossi@unicam.it

† Presented at the International and Interdisciplinary Conference IMMAGINI? Image and Imagination between Representation, Communication, Education and Psychology, Brixen, Italy, 27–28 November 2017.

Abstract: This paper presents a reflection on themes related to the virtual image. In addition, it tries to define a grammar of the digital virtual image, desuming it from other languages such as those related to photography, cinema or video games. After analyzing in detail the set of information and organizational and strategic activities that govern the shaping of the virtual image, the paper outlines the history of innovations that have technically supported the possibilities of the virtual image over the last decade.

Keywords: virtual reality; artificial reality; first person shot; point of view; framing; HMD; panorama; Camcorder

1. The Virtual Image as Language

The visual imagery of virtuality exists entirely in the paradoxical condition between the implicit intangibility of the digital image, which anticipates something that is only imagined or prefigured, and its three-dimensional placement composed of geometric entities orderly composed in a Cartesian space.

The three-dimensional nature is then no longer a simulated condition but an intrinsic attribute of the design. Drawing for virtual reality means developing images that can be explored effectively. The metaphor of navigation has been abused in informatics, referring, for example, to the possibility of conducting an Internet search. Likewise, one can refer to the virtual imagination through actions that roughly imply seeing, but which instead allude to an experience of multi-sensory exploration. We are therefore used to using verbs that, to indicate the analysis or consultation of ordered data, imply a pilgrimage into an unknown space. However, now more than ever, this space corresponds to reality.

The present article therefore presents a reflection on themes related to the virtual image. It aims to define a logistical dimension intended as the set of information and organizational/strategic activities that govern the formation of the image.

Delineating the basic characteristics of the virtual image means composing the grammar of a young language that is still being defined, and for this reason we will borrow from other languages: photography, cinematography, and video games.

The grammar is composed of compositional rules based on some "morphemes" such as: framing, the form of framing, i.e., the format, and the point of view, and is completed and concretized with a series of innovative technological vision devices that in recent years have allowed the imagery of virtual reality to unfold in all its brilliant and psychedelic potential.

Separating and analyzing these morphemes and examining their qualifying technologies not only has an analytical goal; it is also useful to work backwards to discover the mechanisms of constructing and organizing the virtual image.

2. Virtual Reality: Some Definitions

When Myron Krueger coined the term artificial reality [1] in the mid-1970s, his objective was to define a type of digital experience that was so immersive it was perceived as real. He used the concept of artificial reality as a tool to examine human-machine relationships, both analyzing their possible interfaces of exchange and examining their related social/cultural relationships.

At the beginning of the 1990s, the idea of artificial reality was overtaken by the concept of the reality-virtuality continuum synthesized graphically by Paul Milgram [2]. He defined real and virtual as the two ends of a horizontal line and between the two, a type of mixed reality that blends them in augmented reality and augmented virtuality. These intermediate levels therefore pertain to a mixed reality in which the relationship between the observer (the user that experiences reality) and background (the environment in which the user is immersed) determines the point of the continuum.

If the user experiences a reality in which the real surroundings are enhanced with structured digital information, we find ourselves in the field of augmented reality, that is, the field of computer graphics that studies the possibility of overlaying perceived reality with digital objects. Vice versa, if the user experiences a completely artificial digital reality in which the digital information is structured to conform to the perceived world, we are found in what Jaron Lanier at the beginning of the 1980s called virtual reality, thereby concretizing some lines of technical/cultural research that prototypically pervaded cyberpunk literature and the movie industry up to the end of the last century (In virtual reality, the user is called to move and interact with the surrounding environment by means of gestures useful only for probing the environment with means that have nothing to do with the real world. There are also no constraints with respect to the laws of physics regarding gravity, collisions, or interference among objects, forcing users into a complex phase of behavioural training. In augmented reality instead, the user's behaviours correspond to those in reality and the interaction with the surroundings is not constrained to particular preventive instructions; approaching augmented reality therefore is much more intuitive and friendly).

3. Virtual Reality as the Place of the Image

With reference to the cinema and what it means to go to the cinema as a physical place, Francesco Casetti writes, "it means therefore and especially leaving a habitual territory and appearing in another world. In this sense, it means confronting a heterotopia, that is, a special place that, located 'here', opens 'elsewhere'" [3] (pp. 223–224).

This statement is as appropriate for the cinema as it is adaptable, with some effort, to the conditions imposed by virtual reality, especially in relation to the definition given by Foucault who, speaking about heterotopias, says they are, "… a kind of effectively enacted utopia in which all the other real sites that can be found within the culture, are simultaneously represented, contested. And inverted. Places of this kind are outside of all places, even though it may be possible to indicate their location in reality" [4].

These heterotopias share recurring characteristics that were specified by Foucault himself. In particular:

- "Heterotopias always presuppose a system of opening and closing that both isolates them and makes them penetrable" [4]. By analogy, one can think of the need to wear more or less bulky visors or to enter appropriate recesses to access content in virtual reality and immerse oneself in a totally synthetic environment.
- "The heterotopia has the power to juxtapose in a single real place several spaces that are in themselves incompatible" and again, "the heterotopia begins to function at full capacity when men arrive at a sort of absolute break with their traditional time" [4]. In this case, it is even too easy to reconnect the infinite possible combinations through the whirling temporal jumps allowed by virtual reality in its video-entertainment applications.
- "… they create a space that is other, another real space, as perfect, as meticulous, as well arranged as ours is messy, ill constructed, and jumbled" [4]. One can consider, for example, the use of virtual reality in the area of architectural prefiguration. Real-estate agencies are beginning

to use apps dedicated exclusively to viewing their homes for sale, in which the potential buyer hyper-realistically has the possibility to observe in first person the layout of the rooms visually decorated with luxury furniture, coloured rugs, and copies of famous paintings, all accompanied by a relaxing soundtrack of jazz music and the chirping of springtime birds.

Moving within immersive virtuality means walking, running, driving, flying, looking. These are actions that always imply measurement of the space. The three-dimensional image no longer represents space, no longer deceives the user, but rather creates a state of complete cognitive immersion.

What changes, therefore, is the way of seeing things: the period eye, to use Michael Baxandall's fortunate definition [5]. From a static vision tied to the canons of the Renaissance perspective, we reach a panoptic vision that allows us to simultaneously share in all of what happens around us. In this sense, the virtual image, finally overcoming the limits of the frame, the screen, is provided as an aggregate of provisional, moving data.

This image, however, does not move in front of us; rather, we move within an image that no longer needs a protective surface in that it anatomically coincides with our retina. The screen through which we look, intended as a barrier, a window, is gone, collapsing into the latest-generation wearable visors, thereby increasing the field of view both in depth and expanse.

4. Framing and Format

"The visual world is endless. It surrounds us as an unbroken space, richly subdivided but without limits. When we isolate a portion of the world for a photograph or a realistic painting, for example, it is always with the understanding that the world continues beyond the segment's borders" [6]. With these words, R. Arnheim recalls how extracting a single portion from the infinity that can be represented means making a synthesis that is not only visual, but cultural in particular.

Framing is a term commonly used today in the language of photography and cinematography to isolate the subject one wants to represent. Framing implies a formal and narrative choice that orients the observer's view, influencing it. In Renaissance perspective, the format/framing issue was resolved by the subject represented. Instead, during the modernity era the idea of framing without a point of view is affirmed (An example is the photographs of Alexandr Rodčenko), in the same way that it is an axonometric drawing.

With photography first and then cinema, the shooting format became a basic attribute of the image. The form of framing obviously depends on the content of the representation, establishing at the same time a sort of unique relationship between format and content. The form of a depiction is normally based on the rectangle because it can vary significantly according to the proportion between the sides, easily adaptable to the representative needs. The rectangle is usually oriented in landscape mode, i.e., horizontally, thereby enhancing the scene, confirming the supremacy of the field of view over the framing. Only recently, due to the use of smartphones as a tool for photography has the vertical format in portrait mode become a type of shooting capable of enhancing entire figures to the detriment of the context, thereby favoring the use of framing over the field of view.

In the pre-photography era, the frame symbolically represented the earth with its stable form and central symmetry; the circle symbolized the sky and helped to hierarchize the content of the composition from the center towards the edge. In the Baroque era, the ellipse then introduced some dynamics with the two poles of attraction which could organize the composition of the painting. Other forms were used less frequently: the oval, lunette, the Gothic window, up to the polyptychs spread over the fourteenth and fifteenth centuries. This latter form in particular—in which a work is organized on separate painted panels, each of which autonomously depicted a character or an episode—was re-evaluated in the modern era precisely due to its possibility of modulating the depiction, hypothetically making it infinite. Numerous video-artists have used composite screens or large video walls that, with their fragments, are able to construct a large image or sequence of images.

Returning to the realm of cinema, the form of framing has always been rectangular, given that spectators were already used to the typical forms of most photography and paintings. The light that enters through the objective lens would in reality form a circular image on the focal plane if there

were no masks or sensors to "cut" away a considerable portion to obtain a rectangle (Figure 1). As is obvious, the rectangle is not a natural form, given that binocular vision creates an oval shape on the retina with a field of view of about 120° both vertically and horizontally, which extends to about 135° if the far peripheral vision is also considered (Figure 2).
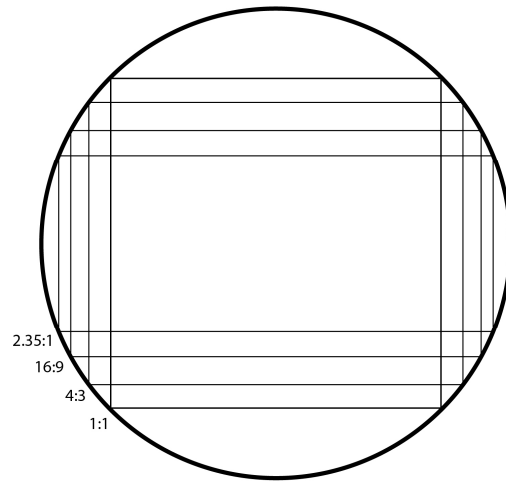

Figure 1. Relationship between the aspect ratios: TV (4:3), widescreen (16:9) and CinemaScope (2.35:1).
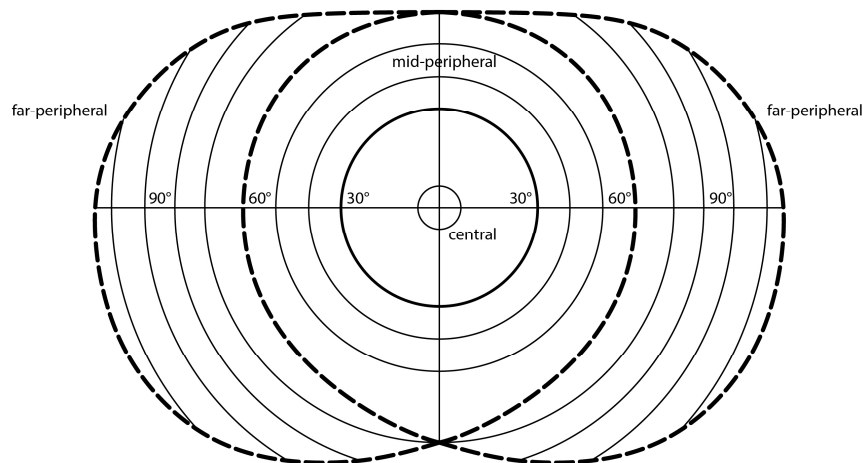

Figure 2. Field of view of the human eye.

Current visors for virtual reality tend to approximate this form, and using OLED technology are able to guarantee a field of view of 110° on a display of 5.7" This means that almost the entire visible area is captured and, precisely for this reason, peripheral vision, which blurs the outer edges, becomes an integral part of the representation delineating the borders. The edges no longer sharp as with the usual rectangular depictions, but are blurred, thereby forcing the use of compositional rules that are no longer based on an orthogonal two-dimensional grid from the canons of painting and photography, but rather on parallax and a three-dimensional changeable dynamic grid.

In the field of computer graphics, the term used to define an image—either photorealistic or not—of a three-dimensional environment constructed digitally is called rendering. In the area of digital graphics, rendering is equivalent to the photographic snapshot. From this, however, they are step back operationally because in the term rendering there is no allusion to the point of view from which the image is "framed", even less are there references to the form of printing. Rendering indicates only the algorithmic process of collapsing a three-dimensional environment onto a two-dimensional surface.

Therefore, while taking a photo implies a series of wasteful operations that tend to isolate, cut, exclude, and separate, rendering implies only a two-dimensional collapsing operation in order to obtain a clipped image for "comfort" according to some mostly rectangular photographic standards (4/3, 16/9, etc.).

5. Panorama and Diorama

Traditional images in art history are representations contained within the populated environment of the spectator. With the digital paradigm, from representation one moves on to simulation [7] and as mentioned above, to rendering, that is, a potentially infinite, panoptic image capable of completely enveloping the spectator.

From the "framed" space of painting, photography, and cinema, one therefore moves to a simulated space that envelops the person experiencing it. Well before the development of virtual-reality visors, the scenic artifice capable of representing and deluding the spectator found within a simulated environment was the panorama and then the diorama. Technically these were large canvases painted and mounted on a circular wall in order to completely surround the spectator, who was then found at the center of a cylinder whose internal face depicted mostly landscapes or war scenes (Figure 3). The resulting effect was a complete, enwrapping view. Viewing a panorama meant immersing oneself within the painting, in isolation from the external world; it was no longer possible to find the edges, or better, the frame of the representation. Analogously, the first dioramas of the 1800s also introduced sophisticated movements and illusions in order to also add dynamics and change within the scene.
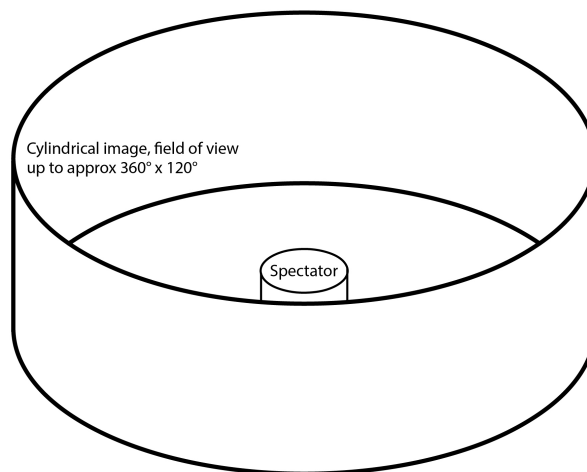
Cylindrical image, field of view up to approx 360° x 120°

Spectator

Figure 3. Panorama set-up.

It is not by chance that the decline of panoramas and dioramas as a spectacle for entertainment coincided with the rise of cinema, which even from the outset promised a realism usually approached only by painted depictions. What was lost was the illusion of really being within the scene, which instead the panoramic, immersive effect of the exhibit guaranteed.

6. The Point of View

Continuing with the analogies with cinematographic language and drawing on its grammar, it is also necessary to touch on the exclusive discipline of cinema and what distinguishes it from other expressive forms: editing.

When building the narrative setting of the virtual image, editing should necessarily be rejected as an artifice capable of directing the spectator's interpretation. The French critic André Bazin [8], precisely to remedy the use of montage, suggested the extreme use of another cinematographic technique: the long take, that is, a single shot that alone carries out the functions of a sequence or scene. In other words, this is a long shot that exhausts an entire narrative sequence without cuts. The

long take is perfectly combined with a filming technique that has become a pervasive stylistic figure of new media: the first-person shot. Ruggero Eugeni defines it as a symbolic form, and more precisely, "a ubiquitous and almost omnipresent figure within the intermediate and post-cinematographic galaxy that characterizes the current era" [9].

The first-person shot, which in the grammar of films represents a well-defined filming technique, is intersected by areas and experience deriving from the world of first-person video games and symbolizes a visual culture characterized by perceptual habits dominated by first-person experiences. The reawakening of commercial interest in really wearable HMD, the miniaturization of filming technologies and action cams, the worldwide success of some first-person shooter (FPS) video games on domestic gaming consoles and the presence of full-length films visibly based on first-person filming and interminable sequence plans, are the main factors constituting the new experiential means that find in virtual reality a fruitful place for experimentation.

These experiments would have therefore been impossible if the technological innovation had not simultaneously accompanied this process.

7. Viewing Devices

Even just an outline of the history of innovations that have technologically supported the possibilities of the virtual image would also merit separate treatment. I limit myself here to highlighting those technological innovations that only in the last decade have allowed sophisticated hardware devices for imagery linked to mixed-reality experiences to be launched on a large scale and at low cost.

With the launch of Google Cardboards (Google Cardboard. https://vr.google.com/cardboard (online on 5 September 2017)) and the crowdfunding campaign to finance the Oculus Rift project (Oculus Rift. https://www.oculus.com (online on 5 September 2017)), commercial interest reawakened unexpectedly at the beginning of the 2010s in a sector that for too long was the object of interest for only a few research centres or those working in sci-fi cinema.

First, Google, with its Glass, tried out new pioneering roads in the field of augmented reality then in rapid succession the first visors for virtual reality were released by the Oculus Rift team, followed by HTC with Vive (HTC Vive. https://www.vive.com (online on 5 September 2017)), and Playstation VR (Playstation VR. www.playstation.com/PSVR (online on 5 September 2017)), which definitively attacked the market of domestic gaming consoles. In addition to these complete solutions for HMD displays, some less expensive alternatives imply the use of smartphones and tablets to integrate mobile apps and solutions thanks to "adaptors" such as Samsung Gear VR (Samsung Gear VR. http://www.samsung.com/global/galaxy/gear-vr (online on 5 September 2017)) or Google Daydream (Google Daydream. https://vr.google.com/daydream (online on 5 September 2017)).

Finally, returning to the field of augmented reality, Microsoft is ready for the world launch of its Hololens (Hololens. https://www.microsoft.com/en-us/hololens (online on 5 September 2017)), which will realize the promises made and then temporarily abandoned by the technology on which Google Glass is based.

8. Camcorder

Of course these technological innovations are accompanied by those characterizing some communicational means that then had significant effects in the languages of cinema, TV, and video games.For example, we cannot avoid mentioning the evolution and process of miniaturizing video cameras. In 1975, the first Steadicam was introduced on the market, that is, a machine to film videos capable by means of inertial and gyroscopic shocks, to move in symbiosis with the operator. The cinematographic eye, which Robert Montgomery made coincide with the human eye in 1947 with Lady in the Lake, could now finally simulate visual trajectories in a real, fluid manner, something unthinkable with traditional filming cameras heavily anchored to tripods or dollies.

In the 1990s, portable digital video cameras introduced new expressive means in the field of documentaries and journalism, introducing the figure of the citizen journalist. In cinema, instead, they have led to the extreme possibilities of the steadicam, innovating filming methods in every

genre. One thinks, for example, of the success of The Blair Witch Project, a low-budget film from 1999 filmed entirely in first person, which was capable of transmitting to viewers all the tension of the actor/operator committed to filming his friends in what was supposed to be a calm adventure in the forest.

Continuing the process of miniaturizing film cameras, we finally arrive at the beginning of the twenty-first century with action cams, that is, video cameras that are so small and light that they can be easily fixed on a helmet and used to film extreme sports. Footage from an action cam, no longer recorded on magnetic tape, but on external supports (SD, micro SD, etc.), meaning it can be easily transferred to and edited on a computer, find a capillary, extremely quick transmission channel on the Internet. In addition to their reduced size, the other innovation of action cams is in the absence of zoom, and especially the use of a large-angle lens capable of capturing images with a field of view of about 170°. Viewed on a traditional monitor, the image has the characteristic form of a barrel, while they find their profitable viewing device precisely in visors for virtual reality (Figure 4).
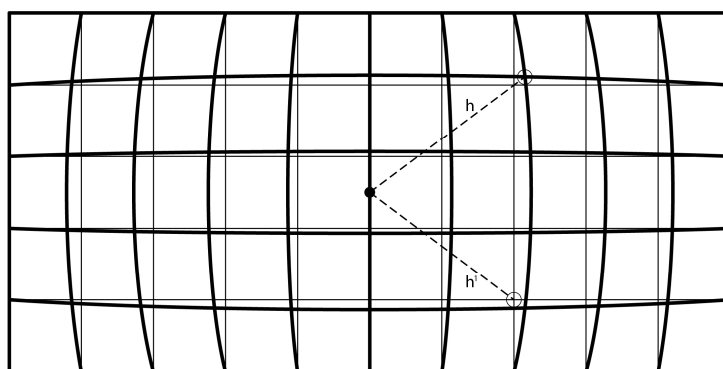

Figure 4. Barrel distortion. In barrel distortion, the image magnification decreases with distance from optical axis.

This quick summary on the evolution of filming devices must end by mentioning a type of video camera called omnidirectional, which is capable of filming immersive 360° video. These videos can be reproduced on both normal computer screens—in which case the interaction and possibility of exploring images is manipulated using click-and-drag—and smartphones—using the gyroscope—as well as on devices designed for virtual reality. These videos, based on the stylistic principle of the first-person shot are spreading rapidly, also thanks to the large digital players such as Google (YouTube support. How to load 360° videos. https://goo.gl/U1CXpv (online on 5 September 2017)) and Facebook (Facebook 360. https://facebook360.fb.com (online on 5 September 2017)), which, since 2015 have supported their use.

Some directors are also experimenting with creating short immersive films usable with virtual reality devices, probing their expressive potential. It is still a long time before a new grammar of film is defined, when the way of understanding scripts and directing will be completely revolutionized. The spectator will decide in real time what to follow in the scene, eliminating the need for editing, as mentioned above. Cinemas will also have to be updated since spectators will be free to move around them to look for the "proper" frame.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Krueger, M.W. *Artificial Reality*; Addison-Wesley: Boston, MA, USA, 1983.
2. Milgram, P.; Kishino, F. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.
3. Casetti, F. La Galassia Lumiere. In *Sette Parole Chiave per il Cinema che Viene*; Bompiani: Milan, Italy, 2015.
4. Foucault, M. Spazi Altri. In *I Luoghi Delle Eterotopie*; Mimesis: Milan, Italy, 2001.

5. Baxandall, M. *Painting and Experience in XVth Century Italy: A Primer in the Social History of Pictorial Style*; Oxford University Press: Oxford, UK, 1976.
6. Arnheim, R.; Pedio, R. *Il Potere del Centro: Psicologia Della Composizione Nelle Arti Visive*; Einaudi: Milan, Italy, 1994.
7. Manovich, L. *Il Linguaggio dei Nuovi Media*; Olivares: Milan, Italy, 2002.
8. Bazin, A. *Che Cos'è il Cinema*; Garzanti: Milan, Italy, 1967.
9. Eugeni, R. *La Condizione Postmediale*; La Scuola: Brescia, Italy 2015.