



# Seeing the World Through Machinic Eyes: Reflections on Computer Vision in the Arts

Marijke Goeting<sup>1,2</sup>(✉) 

<sup>1</sup> Radboud University, Nijmegen, The Netherlands  
m.goeting@artez.nl

<sup>2</sup> ArtEZ Institute of the Arts, Arnhem, The Netherlands

**Abstract.** Today, computer vision is broadly implemented and operates in the background of many systems. For users of these technologies, there is often no visual feedback, making it hard to understand the mechanisms that drive it. When computer vision is used to generate visual representations like Google Earth, it remains difficult to perceive the particular process and principles that went into its creation. This text examines computer vision as a medium and a system of representation by analyzing the work of design studio Onformative, designer Bernhard Hopfengärtner and artist Clement Valla. By using technical failures and employing computer vision in unforeseen ways, these artists and designers expose the differences between computer vision and human perception. Since computer vision is increasingly used to facilitate (visual) communication, artistic reflections like these help us understand the nature of computer vision and how it shapes our perception of the world.

**Keywords:** Art · Design · Perception · Computer vision · Google Earth · Media · Representation · Digital Image

*Once we are definitively removed from the realm of direct or indirect observation of synthetic images created by the machine for the machine, instrumental virtual images will be for us the equivalent of what a foreigner's mental pictures already represent: an enigma.*

—Paul Virilio

## 1 Introduction

In his 1994 book *The Vision Machine*, French cultural theorist Paul Virilio worried about the way automated artificial perception might come to influence our perception of the world. His idea of *vision machines* “that would be capable not only of recognising the contours of shapes, but also of completely interpreting the visual field” and that could “analyse the ambient environment and automatically interpret the meaning of events” [1] seems to have become reality. Today,

© Springer Nature Switzerland AG 2019

L. Leal-Taixé and S. Roth (Eds.): ECCV 2018 Workshops, LNCS 11130, pp. 653–670, 2019.

[https://doi.org/10.1007/978-3-030-11012-3\\_50](https://doi.org/10.1007/978-3-030-11012-3_50)

computer vision is broadly implemented – from automatic passport control to self-driving cars and interactive video games – and used for a variety of tasks: from collecting, processing and analyzing images to even understanding them.

It is significant that Virilio uses the word “interpretation” – a term normally applied to human understanding – to describe the ability of machinic vision. And he is not alone in this. Descriptions of advanced technologies such as computer vision often involve words that are usually associated with human experience. Many writers whose imagination has been captured by recent developments in the field of artificial intelligence, for example, have described intelligent systems like Google’s AlphaGo and DeepDream and Apple’s Siri as “intuitive”, “creative”, and even “funny”.<sup>[2]</sup> But applying human characteristics to computers is misleading: it blurs the distinction between the two and creates the illusion that man could be replaced, or at least rivaled, by machines.

For Virilio the threat of computer-based, artificial vision lies in its ability to create mechanized imagery from which we are often excluded. Virilio was concerned at the prospect of automatic perception needing no graphic or videographic output, therewith totally excluding us.<sup>[3]</sup> And indeed, there are many systems that use computer vision without producing any visual output (think of machine vision used in factories and self-driving cars), and it is true that visual output is not a necessity. Other systems do, however, create visual output that is meant to be seen by both humans and computers (for example QR codes), while yet other systems are designed especially to generate visual output to be seen by us (for example Google Earth). Regardless, with the advent of computer vision, we find ourselves in a new situation. For the first time in history, we are dealing with images that are not only *created* by machines, but that are also meant to be *seen* by machines. We are now in a situation in which we share the perception of our environment with our machinic other. This has given rise to the philosophical problem that Virilio called the “splitting of viewpoint.”<sup>[1]</sup> How can we understand a world seen by a synthetic, sightless vision? What modes of representation are created by it? And how does this affect the way we see the world?

This paper examines computer vision as a medium; as an extension of our sense of sight, as well as of our ability to analyze and recognize what we see, in other words our visual perception as a whole. As media theorist Marshall McLuhan already explained in *Understanding Media: The Extensions of Man* (1964), any extension of our body or mind can be regarded as a medium.<sup>[4]</sup> Computer vision is such an extension: it is an externalization and automation of visual perception by technological means. Like McLuhan and Virilio, I believe it is important to examine the specific characteristics of computer vision in order to understand its effect on our senses, on our perception of the world, and how it influences a society that increasingly relies on computers to do the looking for us.

For artists and designers who are in the business of not only producing images, but also (and perhaps more importantly) of looking at and understanding images and their effect on us, computer vision has become an important

medium – to use *and* to understand. For that reason, I will analyze the work of contemporary artists and designers that are already exploring the possibilities and effects of computer vision. In the past few years, designers and artists like Bernhard Hopfengärtner, studio Onformative, and Clement Valla have been experimenting with the specific characteristics of automated artificial perception and have even viewed computer vision as a particular system of representation. Although the works I have selected are not the most recent works in the field of design and media art, I have chosen them for their capacity to reflect on a diverse range of applications of computer vision (mapping, object recognition, Semacode/QR code and face recognition) while still connected by their use of Google Earth as an artistic medium, and for their ability to shed light on the differences between human and artificial perception. While the advent of convolutional neural networks entails a paradigm shift in the field of computer vision, this technique does not change the basic insights into the nature of perception and computer vision that these artworks provide.

In order to analyze and expand on the ideas that these designers and artists have explored in their work, I will use concepts from the field of media theory and philosophy as developed by Jay David Bolter and Richard Grusin, Beatrice Colomina and Mark Wigley, Anke Coumans, Vilém Flusser, Marc Hansen, Marshall McLuhan, Anna Munster, and Paul Virilio. By looking critically at computer vision as a medium and a system of representation, I hope to advance our understanding of the nature of artificial perception as well as its effect on image culture and visual communication.

## 2 Gaps in the Landscape: The Paradox of Perspective Imagery as Data

One artwork that investigates the nature of computer vision and its system of representation, is the project *Postcards from Google Earth* (2010-present) by French, Brooklyn-based artist Clement Valla. The work consists of screenshots of strange images found on Google Earth (Fig. 1). While navigating the virtual globe, Valla discovered landscapes that did not meet his expectations, such as bridges that appeared to droop down into valleys they were supposed to cross – like Salvador Dali’s watches melting over tables and trees. To Valla, the images he collected felt alien, because they seem to be incorrect representations of the earth’s surface.[5] At first, he thought they were glitches or errors, but later he realized that they were not: they were actually the logical results of the system. “These jarring moments,” Valla writes, “expose how Google Earth works, focusing our attention on the software.”[6]

The way Google Earth’s imagery is created is a rather complex process. Google uses a variety of sources (from space shuttle shots, satellite imagery and airplane photography to GPS data) and a range of techniques (from digital imaging, image stitching, image rendering, and 3D modeling to texture mapping) supported by computer vision.[7] As Valla makes clear, the images produced by Google Earth “are hybrid images, a patchwork of two-dimensional photographic

data and three-dimensional topographic data extracted from a slew of sources, data-mined, pre-processed, blended and merged in real-time.” [5] With the help of computer vision, Google is able to automatically locate features within overlapping photos of the earth’s surface that are the same. By connecting these features to GPS data, it becomes possible to know where photos were taken and from which angle. This allows Google to generate depth maps from different cameras and then automatically stitch these together into one big 3D reconstruction, which can subsequently be textured: photographs can now be applied to the 3D model. Together, tens of millions of images make up Google Earth, whose structure resembles a series of interconnected Russian dolls all made up out of puzzle pieces.[8] Consequently, the imagery found on Google Earth can no longer be regarded as an index of the world (a physical trace of light), but is instead a calculated rendering of data. Thus, even though it seems counter-intuitive, we should not think of Google Earth’s imagery (or in effect any kind of digital image) as a photograph that is simply digital, for it is something else entirely.[6] It is the result of the way the computer is programmed to “see” and how it represents this information visually. According to Valla, Google Earth is a database *disguised* as a photographic representation.[6]



**Fig. 1.** Clement Valla, *Postcards from Google Earth*, 2010-present.

The illusion of reality created by Google Earth is based on its aspiration to be seamless, continuous, complete and up-to-date. In order to achieve this, the computer vision software, which Google uses to automatically generate its virtual globe, selectively chooses its data and creates a very specific representation of

the earth. It does this by training on a few basic traits – it learns to recognize and select images that contain no clouds, high contrast, shallow depth and daylight – to give us a smooth and continuous 24-hour, cloudless, day-lit world.[5] The inclusiveness and presentness of this idealized representation is the result of the speed by which it comes into being. Accordingly, the imagery of Google Earth cannot be defined in relation to a particular time or place, but rather in relation to the speed of calculation.

The effect of this is that Google Earth does not privilege a particular viewpoint, but aims at a *universal perspective* – which is exactly why it needs to be collected, processed, analyzed and rendered by a computer, instead of seen from the limited viewpoint of an embodied observer. Not only is perspective arbitrary when it comes to 3D modeled imagery, it is actually considered an obstacle to the total automation of sight by vision researchers and programmers. As media theorist Marc Hansen points out in his article “Seeing with the Body: The Digital Image in Postphotography” (2001), it is only by deprivileging the particular perspectival image that a totally and fully manipulable grasp of the entire data space becomes possible. Hence, “[w]ith this deterritorialization of reference,” Hansen writes, “we reach [...] the moment when a computer can ‘see’ in a way profoundly liberated from the optical, perspectival, and temporal conditions of human vision.”[10]

As art historian Jonathan Crary has remarked in the introduction of his book *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century* (1990), with the advent of techniques like computer-aided design and robotic image recognition, “[m]ost of the historically important functions of the human eye are being supplanted by practices in which visual images no longer have any reference to the position of an observer in a ‘real,’ optically perceived world.”[11] Consequently, according to Crary, these techniques “are relocating vision to a plane severed from a human observer”.[11] I, however, would argue that this does not happen entirely or definitively. Although Google Earth’s imagery is indeed both “seen” and represented by technology, its *raison d’être* is that it will be perceived by a human observer. Thus, while computer vision operates independently from human perception (in terms of its opticality, perspective and speed), it is paradoxically also (still) bound to it: the human observer remains the prime focus of these images; s/he forms both its starting and end point.

This paradox is exactly what Valla’s *Postcards from Google Earth* reveal. They show how the imagery of Google Earth is the result of a double recoding – from image to data, and from data to image – by using photographs as textures to decorate the surface of a 3D model. The problem, however, is that we – as humans – see *through* a photograph, and unconsciously look *at* a surface. “Most of the time this doubling of spaces in Google Earth goes unnoticed,” Valla explains.[5] But when the photographs are taken from a particular angle and contain depth and shadows, suddenly the two spaces do not align. At that moment, “we are both looking *at* the distorted picture plane,” the artist writes, “and *through* the same picture plane at the space depicted in the texture. In other words, we are looking at two spaces simultaneously.”[5] This clash of embodied

perception (with its sense of perspective and experience of space) and computer vision (with its mathematical calculation of data), reveals the friction inherent in a human-computer connected perception of the world.

In their book *Remediation: Understanding New Media* (2001), media theorists Jay David Bolter and Richard Grusin state that (new) media often strive for what they call “immediacy”: to ignore, deny or erase the presence of the medium and the act of mediation, so that the viewer or user appears to stand in an immediate relationship with what is represented. According to Bolter and Grusin, digital technologies such as virtual reality and three-dimensional computer graphics are seeking to make themselves “transparent”. However, often, the desire to achieve immediacy involves a large amount of (re)mediation, of combining many different media – what Bolter and Grusin call “hypermediacy” – to produce this effect.[12] This can also be seen in the example of Google Earth, which uses a combination of many different media and techniques (satellite images, airplane photography, GPS data, 3D modeling, image stitching, image rendering and texture mapping) in order to create one single, apparently seamless, “transparent” and immersive visual space. According to Bolter and Grusin, such immediacy is therefore paradoxical by nature: to achieve it, especially in new digital media, hypermediacy is required. In some cases, the experience of immediacy can even flip to an experience of hypermediacy when the viewer suddenly becomes aware of the medium and the act of mediation.

An example that Bolter and Grusin use to explain the coexistence and interdependency of immediacy and hypermediacy is the photomontage, or collage. This medium incorporates a tension between looking *at* a mediated surface and looking *through* to a “real” space beyond the surface.[13] “We become hyperconscious of the medium in photomontage,” Bolter and Grusin write, “precisely because conventional photography is a medium with such loud historical claims to transparency.”[14] As explained earlier, Google Earth is also a kind of photomontage or collage. However, in Google Earth the user is not aware that s/he is actually looking at a photomontage – which is, of course, the intention. The experience of transparent immediacy dominated Google’s hypermediated virtual space,[15] until Valla exposed the unintentional moments of obvious mediation in an otherwise seemingly transparent, unmediated Google Earth. By selecting moments when the technology “fails,” the illusion of immediacy is immediately breached. In so doing, Valla makes viewers aware of the nature of the medium, while simultaneously reminding them of their desire for – and habitual reliance on – “transparent” photorealistic imagery.

In her text “De stem van de grafisch ontwerper” (The voice of the graphic designer), film theorist Anke Coumans explains how “bad” images – meaning images that do not successfully use or fully employ the possibilities of the technical apparatus by which they are made – reveal the apparatus itself. In other words, a bad image reveals how the image was programmed, but also how it programs us. For instance, bad lighting in a photo, she writes, draws the viewers’ attention to the lighting, making them aware of the distinction between a photo and reality, as well as of the reality that existed before the camera.[16] With

Valla's work, the distorted images reveal how Google Earth was programmed (literally) and how it programs us to see it as an indexical photographic representation of the world. However, in the case of Valla's *Postcards*, one could wonder if there ever was a reality before the camera. Although Google uses satellite images, Google Earth's "universal texture", as Valla calls it, is in fact a computational model.

Coumans' analysis builds on the philosophy of Vilém Flusser, who has examined the nature of "Technobilder". Technical images like photographs, film, video and computer graphics are not surfaces that represent objects or scenes, Flusser argues, but rather mosaics, visualizations of numerical code.[17] Likewise, the highly technical image of Google Earth does not represent the surface of the earth, but a programming language that constructs a computational model. As Flusser writes, "technical images [...] are produced, reproduced and distributed by apparatuses, and technicians design these apparatuses." [18] Addressees of technical images are therefore often unaware of the specific program, of the level of consciousness that went into the creation of these apparatuses, Flusser argues.[19]

Through specific design strategies, artists and designers are able to "break" the program, Coumans argues, allowing space for the viewer to enter into dialog with the otherwise predetermined visual communication.[16] Instead of images that affirm, they create "transapparatic images," [20] which transcend the technical program and evoke contemplation. The difference between the graphic design strategies that Coumans discusses and the work of Valla is that Valla simply selected the right "wrong" images. Consequently, these images do not point to the artist, but rather to the technician. In this sense, discovering the technical "failures" or glitches[21] of the apparatus can provide rare moments in which the viewer can suddenly see through the image and glimpse its origin, i.e. its technical production and the programmer's intention.



**Fig. 2.** Bernhard Hopfengärtner, *Hello, World!*, 2006.

### 3 Perception in an Expanded Field: Between Technology and the Body

Another “transapparatic” image that transcends its technical program and reflects on the collision of human perception and computer vision, is the design project *Hello, world!* (2006) by German designer Bernhard Hopfengärtner. The work consists of a graphical pattern that the designer mowed into a wheat field near the town of Ilmenau in Thuringia, Germany (Fig. 2). The 324 bright and dark squares together form a  $160 \times 160$  meter-wide Semacode, a machine-readable pattern, similar to a QR-code, which is used to connect online information to objects and locations in the real world.[22] The code translates into the phrase “Hello, world!”.[23] Without the aid of computer vision, however, it is very difficult, if not impossible, to read and understand this abstract code.

By enlarging the pattern to a size beyond human scale, Hopfengärtner highlighted the fact that these visual codes are strictly speaking not meant to be seen and read by humans, but by computers. The work can really only be seen from an airplane, drone, or satellite – as a photograph. Its perception, seeing the work visually and as a whole, is therefore dependent on technological mediation. Its interpretation in turn is dependent on computer vision: algorithms are needed to extract the meaning of the pattern. As such, this work calls into question the role of the viewer, or even the viewer itself, which, arguably, need not be human at all. On the other hand, when seen from ground level, from within the field, the work had a specific tactile, olfactory, auditory, spatial and temporal nature. When walking through the extensive field, people were able to feel and smell the wheat, hear the wind whistling through it. They could get a different view of the field from different angles and under different weather conditions. Over a period of time, the wheat grew and changed color, and eventually the pattern disappeared. This particular perception of the work, this bodily experience, can really only be had by a human viewer.

Hansen argues how, in contrast to computer vision, human perception is always contaminated with affection. Hansen writes: “human perception takes place in a rich and evolving field to which bodily modalities of tactility, proprioception, memory, and duration – what I am calling affectivity – make an irreducible and constitutive contribution.”[24] In other words, duration (how long something takes, but also intervals in occurrences), the position from which we view something and the movement of the body through space, the way things feel, the memories of previous encounters and the associations these memories trigger: all these aspects are meaningful to an embodied viewer. They are inextricably linked to, and therefore greatly influence not only *what* we see, but also *how* we interpret it.

Computer vision of course lacks this bodily perception and consequently it sees and interprets the world quite differently. As Virilio already observed, computers do not so much look at images (in the sense that they see and perceive images as humans do), as process data. According to Virilio, the word “image” in this context is empty, and perhaps even deceptive, since in reality the computer is rapidly decoding information and analyzing it statistically.[25] So not only



does computer vision lack the position of an embodied viewer in a real, multi-sensorial perceivable world, it is also looking at a remediation of the world that consists exclusively of data.[26]

Hopfengärtner's work points to the differences between computer vision and human perception – between a mediated, disembodied, data-driven gaze and bodily experience. Whether the perception of this work can be best arrived at through computer vision or human experience remains a question, although I would argue that it lies precisely in between. This is explained in part by the fact that the “real” space where this work exists – or at least where it was meant to be seen is online, in the virtual world of Google Earth.

The project's aim was to send a message – “Hello, world!” – to the world via the digital globe of Google Earth. This message – a reference to mastering a programming language[27] – can be said to be directed both at the technology itself and the users of that technology, the designer included. In this sense, the project can be seen as a way to gain access to, control, or at least engage in a dialog with the complex and omnipresent digital realm of technology giant Google. Hopfengärtner did this not by working directly in the program of Google Earth, but by altering a part of the physical landscape, assuming that it would be automatically integrated in the satellite images Google uses to construct its virtual globe, because we do not only watch Google Earth, it also watches us.[28] Whether users of Google Earth were/are actually able to see Hopfengärtner's message is unlikely, since Google updates its aerial views regularly.[29] In this way, his message is more a statement about “self-determination and possessions in the digital world,” as the designer describes it.[30]

Since the industrial revolution, writers Beatrice Colomina and Mark Wigley argue in *Are We Human? Notes on an Archeology of Design*, the debate about design has centered on the complex relationship between humans and technology. Designers and thinkers like William Morris, for example, suggested that the machine was no longer a human tool, but had become a new life-form that was turning humans into its tools.[31] According to Colomina and Wigley, “[d]esign was framed as a way to deal with the increasingly dominant logic of the industrialized and globalized world while resisting the perceived dehumanizing impact of that world. [...] The word *design* was called on in the 1830s to explicitly negotiate between human and machine”.[32] Today, a large and important part of design still focuses on precisely this balancing act. With Hopfengärtner's outcry “Hello, world!” (perhaps a question mark – “Hello, world?” – would have struck a more fitting tone), the designer lays bare some essential questions about the relationship between humans and their technology.

Today we literally live inside design, Colomina and Wigley make clear. For them, design includes everything, from the materials and objects that we use to network systems and the process of bioengineering. In this sense, Colomina and Wigley's definition of design resembles McLuhan's definition of media as any extension of our body and mind that “gradually creates a totally new human environment” which in turn “shapes and controls the scale and form of human association and action.”[33] As a result of the continuous process of redesign-

ing the human by design, Colomina and Wigley argue, there is no longer an outside to design: the whole planet is covered in countless overlapping and interacting webs, from underground transit systems and submarine communications cables to buildings, cities, transportation infrastructure and cell phone towers to satellites and space stations that circle the earth.[34]

The parallel world of Google Earth reflects this, with its interwoven layers of satellite images, 3D textures, and GEO-specific information. But what kind of planet is Google Earth exactly? In her book *An Aesthesis of Networks: Conjunctive Experience in Art and Technology* (2013), media artist and theorist Anna Munster describes the particular nature of Google's virtual globe. While at first it may appear to be the ultimate simulation of the world, she observes, one crucial aspect is missing: collective human sociality. The experience of Google Earth is solitary: you "fly" around from location to location without ever encountering others.[35] "Instead of producing a heterogeneously populated world," Munster writes, "Google Earth produces a world and its peoples as a loose database of individual users initiating and retrieving their individual queries bereft of any sociality".[36]

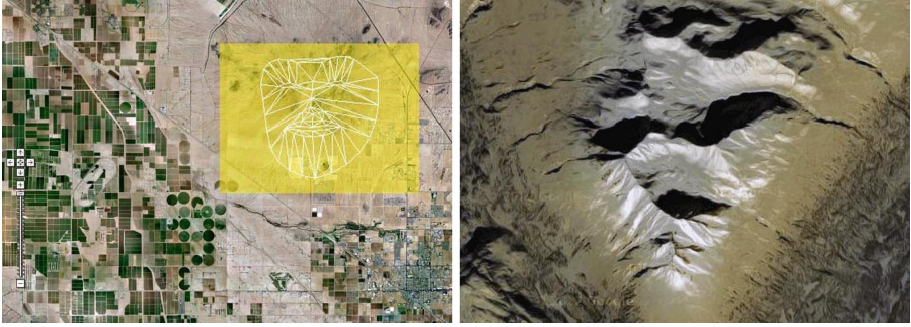
Even though Google Earth is made by/of us (because we both contribute content to it and produce the world that is its subject), the experience of this parallel world remains a solitary one. In Google Earth, the liveliness and messiness of social relations seems to be forever just out of reach. It does not seem to matter how detailed Google Earth becomes, it never really turns into a social place. Instead, it is a beautiful clean image of the globe, which can be consumed by nomadic, solitary individuals who live a database-mode of existence and for whom the Google Earth experience has become an end in itself.[37]

Hopfengärtner's message tries to cut through this and reach others: "Hello, world!" But in the solitary world of Google Earth, it becomes little more than a faint echo of the designer's presence. Yet at the same time its encoding as abstract visual pattern suggests that its receiver is not human, but machine – or some kind of mix. In this sense, his work is perhaps a message to the humans we have become. "What makes the human human is not inside the body or brain, or even inside the collective social body, but in our interdependency with artifacts", Colomina and Wigley state.[38] "Artifacts are interfaces," they write, "enabling different forms of human engagement with the world but equally enabling the world to engage with the human differently." [39] The increasing use of – and dependency on – computer vision technology, however, begs the question of how the world is increasingly engaging with us and how our engagement with others is being shaped by it.

## 4 Looking for Faces: Statistics and the Imagination

The conflation of human perception and extended technological perception is at the heart of another work, entitled *Google Faces* (2013). This digital design project consists of a computer vision program that autonomously searches for faces hidden in the surface of the earth (Fig. 3). Its makers, German designers

Julia Laub and Cedric Kiefer, who founded design studio Onformative, developed an application that automatically analyzes satellite images from Google Maps by using a face-detection algorithm.[40] As the designers explain on their website, their aim was “to explore how the cognitive experience of pareidolia can be generated by a machine”.[41]



**Fig. 3.** Studio Onformative, *Google Faces*, 2013.

Pareidolia is a psychological phenomenon used to describe the human tendency to detect meaning in vague visual (or auditory) stimuli. Whenever we navigate our surroundings, it is of vital importance that we identify and recognize visual patterns, whether it is the face of a friend in a crowd or the speed of an approaching car. Sometimes this mechanism continues to work in situations where it should not, and we recognize a face in the shape of a mountain. Onformative’s fascination is with such erroneous and seemingly useless pattern recognition. According to the designers, “we also tend to use this ability [pattern detection] to enrich our imagination. Hence we recognize meaningful shapes in clouds or detect a great bear upon astrological observations.”[42] Consequently, pattern recognition can be considered an important faculty when it comes to looking at art, and is responsible for our ability to infuse certain landscapes and objects with symbolic meaning.

Pattern recognition is something computers share with us, even, as it turns out, the ability to see faces when there are none. In fact, it was the high rate of false positives (the detection of a face when there is none) that Laub and Kiefer noticed when they worked with face-tracking technology for an earlier project, which led them to further investigate this phenomenon with *Google Faces*.[43] But the similarity between computer vision and human perception is only superficial. In looking for faces in landscapes, Onformative’s application *simulates* our tendency to see meaningful patterns, while in fact it is calculating the number of light and dark spots which together pass for a face-like configuration in the form of two eyes, a nose and a mouth. As Virilio already observed, “blindness is thus very much at the heart of the coming ‘vision machine’. The production of *sightless vision* is itself merely the reproduction of an intense blindness that will

become the latest and last form of industrialization: *the industrialization of the non-gaze*" [44] (original italics).

An adaptation of René Magritte's painting *The Treachery of Images*, posted by the Computer Vision Group of the University of California Berkeley, also – perhaps unintentionally – points to a fundamental blind spot in the way computers see the world. [45] In this image, the painting of Magritte is overlaid with a pink rectangle, referring to the field recognized by computer vision, along with its estimation of the subject – “pipe” – and its correspondence rate: 94 percent (Fig. 4). Amusingly, in this image the computer itself is not completely sure whether this is a pipe or not, providing a 94 percent accuracy. Arguably, this has little to do with the fact that it understands the difference between object and representation, since it knows only representation and not the world.



**Fig. 4.** Contemporary adaptation of René Magritte's painting *The Treachery of Images* (1928–1929). (Color figure online)

Thus, while Magritte's painting makes the viewer aware of the difference between object and representation – emphasized by the caption *Ceci n'est pas une pipe* (This is not a pipe) – this image makes us aware of how far a computer-interpreted version of the world is removed from the rich field of human perception and interpretation. While an image of something is not the thing itself, but a representation, what the computer “sees” is a pre-established category, based on a large number of representations applied to yet another representation. It compares a representation of a pipe with a database of representations of pipes. It therefore knows not the object nor reality, but only the pattern “pipe” – an abstracted and reduced version of reality. Paradoxically, a man-made artistic interpretation of the world, such as Magritte's painting, is therefore closer to reality than a computer's calculated account of it.

Consequently, we should be careful not to confuse human interpretation with statistical calculation or to contribute human abilities to the computer. Some of

the less recognizable results of *Google Faces*, for example, have been described by writer Margaret Rhodes as “subjective,” and the machine’s eye as “more conceptually artistic” than ours.[46] In so doing, we are not only anthropomorphizing the landscape, we are also anthropomorphizing the computer. While we might feel we have found “faces staring back,”[46] it is in fact *we* that *project* our gaze: not only on the landscape that appears to be looking back at us, but also on the computer that is doing the looking for us.

If we consider computer vision a medium – an extension of visual perception – we realize that what Onformative is doing with *Google Faces* is examining what happens when we delegate our perception *along with our imagination* to a computer. The result is, as Kiefer and Laub describe it, an inseparable process in which objective investigations (computers) and subjective imagination (humans) collide.[41] It is important to remember, however, that in the end, it is the human observer that contributes meaning to these images, not the computer.[47] While a computer might be able to detect a face in a landscape, it is (as of yet) not able to detect this recognition itself as an instance of pareidolia or an act of the imagination.[48]

## 5 Conclusion

When Virilio wrote *The Vision Machine*, he worried that artificial vision would leave humans out of the perceptual loop altogether, i.e. that we would share the perception and even the interpretation of our environment with machines without any need for visual feedback. While this is still cause for concern, artists like Clement Valla, studio Onformative, and Bernhard Hopfengärtner, who explore the nature of Google Earth, show that computer vision is also used to construct images that are especially meant to be seen by us, humans. Instead of solving the problem of exclusion, however, this application of computer vision creates problems of its own – particularly when it comes to generating a visual system of representation that is used to understand the world and even more so when that system of representation simulates another. The resulting trouble is that we fail to perceive, and therefore understand, the technical program that shapes our communication.

As Valla’s *Postcards from Google Earth* show, Google Earth is neither a photographic, indexical representation of the world, nor connected to the position of a real embodied observer. Google Earth does not reflect a particular perspective, but instead aims at an idealized and universal depiction of the earth’s surface. Its aim, however, is strongly contrasted by its method. As Bolter and Grusin make clear, the desire for immediacy is often approached through hypermediacy. Similarly, Google Earth’s smooth, continuous space is actually a patchwork of tens of millions – very selective! – images all stitched together through processes of automated visual analysis. This disguise, this illusion of reality, serves to counter or transcend our own “limited” (perhaps “undesired” is the better word) perception of the world by filtering out clouds, depth, strange angles, darkness, or any kind of obscurity or ambiguity.

But it doesn't stop here. Its speed of calculation, combined with automation, is one of the reasons why computer vision is called upon to do the looking for us. This means that the convoluted and slow process of human perception, which involves duration, tactility, movement, changing perspectives, memories and associations, is left out of the loop. As Hopfengärtner's work *Hello, world!* shows, in machine vision, our rich field of interpretation is reduced to a coded pattern that contains only a limited amount of information. In addition, Hopfengärtner's failure to get his message into Google Earth reflects how the speed of artificial perception can work against an often less efficient or predictable human communication. Perhaps this is why human sociality is lacking in a place like Google Earth. As Munster argues, the messiness of human social relations becomes an obstacle to the consumption of Google's nice, clean image of the globe.

The risk is that, despite all these deficiencies and due to complex and opaque technical processes, we start to project our own gaze onto the technology. As studio Onformative has shown with *Google Faces*, it is not difficult to anthropomorphize computer vision, to regard its outcome as subjective or to add human characteristics to it. However, as Virilio rightfully pointed out, blindness is at the heart of any vision machine – not only because it statistically calculates data, but also because this calculation is based on representation and not the world. In this sense, any artificial visual analysis is at least twice removed from reality: first, by looking only at representations, second, by comparing those representations to other representations, but also to predetermined patterns and concepts. For that reason, to describe the process of computer vision as a process of interpretation is anthropomorphic, since there really is no understanding involved, only (an intricate process of) selection. By examining what happens when we use computer vision to simulate our tendency to see meaningful patterns in random data, studio Onformative demonstrates that we are all too willing to conflate human perception and computer vision.

As McLuhan, Colomina and Wigley make clear, any technology gradually creates a totally new human environment that shapes and controls human thought and action. This is as true for computer vision, as it is for any other technology – perhaps even more so, since it externalizes and automates one of our most dominant faculties: visual perception. It is therefore important to fully understand the nature of computer vision and the way it analyzes and represents our world. By exposing unintentional moments of mediation, by highlighting the frictions that are part of a human-computer connected perception and by discovering technical failures, artists and designers allow us to peak through the cracks of otherwise often hermetically sealed technical processes. For them, technical failures and “bad” images do not need to be eradicated or quickly fixed. Instead, these instances are valuable. They not only reveal the particular nature of the apparatus itself, but also how it programs us to see the technology and the world it creates.

## References

1. Virilio, P.: *The Vision Machine*, p. 59. Indiana University Press, Indianapolis (1994)
2. Shubber, K.: Artificial artists: when computers become creative. *Wired*. 7 Aug 2013. <http://www.wired.co.uk/article/can-computers-be-creative>. Naughton, J.: Can Google's AlphaGo really feel it in its algorithms? *The Guardian*. 31 Jan 2016. <http://www.theguardian.com/commentisfree/2016/jan/31/google-alphago-deepmind-artificial-intelligence-intuitive>. Titcomb, J.: The best of Siri: 11 funny responses from the iPhone's virtual assistant. *The Telegraph*. 1 Jul 2015. <http://www.telegraph.co.uk/technology/apple/11709991/The-best-of-Siri-11-funny-responses-from-the-iPhones-virtual-assistant.html>
3. Virilio: *The Vision Machine*. 60
4. McLuhan, M.: *Understanding Media: The Extensions of Man*. Ginko Press, Berkeley (1964). 5, 19, 34
5. Valla, C.: *The Universal Texture*. *Rhizome*, 31 July 2012. <http://rhizome.org/editorial/2012/jul/31/universal-texture/>
6. Postcards from Google Earth. <http://www.postcards-from-google-earth.com/info/>. Accessed 1 Mar 2016
7. In 2012, Google switched from using its geo-modeling community (which consists of many volunteers that manually created detailed 3D models of buildings) to automatic image rendering and computer vision techniques to create a 3D representation of entire metropolitan areas. *The Never-Ending Quest for the Perfect Map*. <https://googleblog.blogspot.com/2012/06/never-ending-quest-for-perfect-map.html>. Accessed 15 Mar 2016
8. Google Earth's Incredible 3D Imagery, Explained (Youtube). <https://bit.ly/2pnyZsG>. Accessed 26 Jun 2018
9. See also Goeting, M.: Digital fluidity: the performative and reconfigurable nature of the digital image in contemporary art and design. *Int. J. New Media Technol. Arts* **11**(4), 27–46 (2016)
10. Hansen, M.: Seeing with the body: the digital image in postphotography. *Diacritics* **31**(4), 54–84 (2001)
11. Crary, J.: *Techniques of the Observer: On Vision and Modernity in the Nineteenth Century*, p. 2. MIT Press, Cambridge (1990)
12. Bolter, J.D., Grusin, R.: *Remediation: Understanding New Media*. MIT Press, Cambridge (2000). 5–6, 22–23
13. Bolter, Grusin: *Remediation*. 38–41
14. Bolter, Grusin: *Remediation*. 38
15. In this regard, Google Earth departs from the tradition of cartography. Cartography's aim to map places always involves a form of reduction; relevant geographical information is collected and transformed into a schematic representation of an area. Traditionally, what gets on the map are distinctive elements, not uniformity, because including sameness hinders the functionality of a map. The aim of Google Earth, however, is to "build the most photorealistic version of our planet," as one of Google's developers makes clear, and this involves including as much (overlapping) information as possible. See Adams, C.: *Imagery update: Explore your favorite places in Google Earth*. Medium. <https://medium.com/google-earth/imagery-update-explore-your-favorite-places-in-google-earth-5da3b28e4807>. Accessed 19 Sep 2018. Consequently, Google Earth can no longer be considered a map. Instead, it functions as an all-encompassing virtual image space that builds on the photographic paradigm and owes more to the history of immersive (virtual) spaces than cartography

16. Coumans, A.: De stem van de grafisch ontwerper, drie vormen van dialogische verbeelding in het publieke domein. *Esthetica*. <http://estheticatijdschrift.nl/wp-content/uploads/sites/175/2014/09/5-Esthetica-Destemvandegrafischontwerper-Drievormenvandialogischeverbeeldinginhetspubliekedomein-2010-12-20.pdf>. Accessed 29 May 2018
17. For this reason, the distinction between “real” and “virtual” had lost its meaning to Flusser. He preferred to talk about the distinction between gestures of “abstraction” and “concretion”
18. Flusser, V.: *Writings*, p. 128. University Minnesota Press, Minneapolis (2002)
19. Flusser: *Writings*. 129–130
20. The term “transapparatische Bilder” was originally conceived by Flusser. Flusser, V.: *Medienkultur*. Fischer Taschenbuche Verlag, Frankfurt am Main (2005). 75, 77
21. According to Dutch glitch artist Rosa Menkman, “Glitch, an unexpected occurrence, unintended result, or break or disruption in a system, cannot be singularly codified, which is precisely its conceptual strength and dynamical contribution to media theory.” Menkman, R.: *The Glitch Moment(um)*. Institute of Network Cultures, Amsterdam (2011) 26. “It [the glitch] is the moment at which this flow [of technology] is interrupted that a counter-experience becomes possible. [...] the possibility for an alternative message unfolds. Through the distorted lens of the glitch, a viewer can perceive images of machinic inputs and outputs. The interface no longer behaves the way it is programmed to; the uncanny encounter with a glitch produces a new mode that confounds an otherwise predictable masquerade of human-computer relations.” Skyers, E. I.: *Vanishing Acts*. Link Editions, Brescia (2015). 48–49
22. The Semacode was originally designed to encode Internet URLs, but it is also used by postal services to automate the distribution of parcels and letters, and by the railway and concert venues to sell tickets online
23. Hello, world! <http://hello.w0r1d.net/description.html>. Accessed 16 May 2018
24. Hansen: *Seeing with the Body*. 61–62
25. Virilio: *The Vision Machine*. 75 However, one could question whether the computational algorithms employed in computer vision are really that different from how we as humans perceive and analyze the world. After all, many of these technologies are modeled after us. Like us, they work with pattern recognition, although of course not as advanced as humans. However, although many scientists now agree that the human brain works similar to computers/algorithms, there is a risk in using this analogy. It may cause us to overlook the differences. As historian Yuval Noah Harari explains, during the Industrial Revolution scientists described the human body and mind as a steam engine because it was the dominant technology at the time. While this sounded logical in the nineteenth century, it seems naïve today and the same thing most likely applies to the human-computer analogy. It explains only a very small part of the human, and even less about what it means to be human. See Harari, Y. N.: *Homo Deus: A Brief History of Tomorrow*. Random House, New York (2016)
26. The digital photograph taken of the wheat field is actually a list of data that describes values of color, contrast, size, etc. It is only after using software to convert this data into an image that we can talk about an image. See also Goeting: *Digital Fluidity*
27. The phrase “hello, world” was first used by Canadian computer scientist Brian Kernighan in 1978 in his instruction manual *The C Programming Language* as a small test program to display a basic understanding of the programming language.



- It was subsequently used by many others as a symbolic first step in mastering a programming language
28. Dambeck, H.: Code in Kornfeld: Gre an die Welt ber Google Earth. Spiegel Online, 9 May 2006. <http://www.spiegel.de/netzwelt/web/code-im-kornfeld-gruesse-an-die-welt-ueber-google-earth-a-415135.html>
  29. It is unclear whether Bernhard Hopfengärtner succeeded in getting his design recorded by the satellites Google uses and consequently on Google Earth, since the satellites scan the globe with intervals of approximately one year, by which time the pattern might have faded. Moreover, the satellite images are updated with the same interval, so if it did get on to Google Earth, it was only visible for a short period of time. That is, for those people who managed to find it among the vast imagery of Google's virtual globe
  30. Bauhaus-Universitt Weimar. <http://www.uni-weimar.de/projekte/iwantmymkg/en/hello-world>. Accessed 18 May 2018
  31. Colomina, B., Wigley, M.: Are We Human? Notes on an Archeology of Design, p. 80. Lars Mller Publishers, Zrich (2016)
  32. Colomina, Wigley: Are We Human? 76–77
  33. McLuhan: Understanding Media. 12, 20
  34. Colomina, Wigley: Are We Human? 9
  35. Munster, A.: An Aesthesia of Networks Conjunctive Experience in Art and Technology, pp. 45–55. MIT Press, Cambridge (2013)
  36. Munster: An Aesthesia of Networks. 53
  37. Munster: An Aesthesia of Networks. 51–61
  38. Colomina, Wigley: Are We Human? 23 See also Stiegler, B.: Time and Technics 1: The Fault of Epimetheus. Stanford University Press, Stanford (1998)
  39. Colomina, Wigley. Are We Human? 25 The following text by Bernhard Hopfengärtner can be connected to this: “The interaction with our environment is producing our mental representation of it. Designing objects of interaction, be it physical objects, services or cultural technologies, is also a method of generating new or alternative ways of thinking. As interaction design can be used as an approach to explore this field, it also allows the incorporation social [sic] or philosophical observations and ideas into a tangible form.” Royal College of Art. <https://www.rca.ac.uk/students/bernhard-hopfengartner/>. Accessed 22 May 2018
  40. Their “algorithmic robot,” as Kiefer and Laub call it, browses Google Earth day in, day out, continuously scanning the virtual globe by moving along the latitude and longitude of the earth. After it has circled the globe, it starts again, zooming in closer every time, which exponentially increases the number of images that need to be analyzed. After running the bot non-stop for several weeks, it only traveled a small part of the globe. Kiefer quoted in Solon, O.: Google Faces searches for Faces in Google Maps, and finds Forever Alone Guy. Wired. 23 May 2013. <http://www.wired.co.uk/news/archive/2013-05/23/google-faces>
  41. Google Faces. <http://onformative.com/work/google-faces>. Accessed 7 Mar 2016
  42. Onformative quoted in Garber, M.: These Artists are Mapping the Earth with Facial Recognition Software. The Atlantic, 21 May 2013. <http://www.theatlantic.com/technology/archive/2013/05/these-artists-are-mapping-the-earth-with-facial-recognition-software/276101/>
  43. Kiefer quoted in Solon, O.: Google Faces searches for Faces in Google Maps
  44. Virilio: The Vision Machine. 72–73
  45. Berkeley Computer Vision Group. <https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/shape/>. Accessed 2 Jun 2018

46. Rhodes, M.: Finding Hidden Faces in Google Earth's Landscapes. Fast Company, 6 October 2013. <http://www.fastcodesign.com/1672781/finding-hidden-faces-in-google-earths-landscapes>
47. Likewise we can delegate our wish to achieve chance or randomness in the production of artworks to computers, believing the computer to be the better instrument to achieve this goal. While computational algorithms can produce long sequences of apparently random results, they are in fact based on deterministic logic. Because these computational algorithms can never be regarded as a "true" random number source (as tossing a coin or rolling dice are), they are called pseudorandom number generators. Here again, it is our desire to perceive randomness and our inability to discern patterns and predict the outcome of a computational algorithm, that creates the impression or illusion of randomness
48. Moreover, as Stuart Hall makes clear, we humans are able to form concepts of rather obscure and abstract things, which we can't in any simple way see, have never seen and possibly can't or won't ever see. Examples are concepts like war, death, friendship or love. Hall, S.: Representation: Cultural Representations and Signifying Practices. Open University, London (1997) 17 For instance, how would computer vision be able to detect an image as a representation or expression of love or revenge?