

Machine Vision as Viewed Through Art: Hostile Other or Part of Ourselves?

By Jill Walker Rettberg

Paper presented at Post-Screen Festival: PSF2016, Lisbon, 17-18 November 2016.

Abstract: This presentation analyses three recent works of art that interrogate the relationship between human perception and machine vision: Nadav Assor's art-documentary *Lessons on Leaving Your Body* (2014), Muse's VR music video *Revolt* (2015), and Erica Scourti's *Body Scan* (2014). How do these works present the relationship between human and machine vision? When machines can see us, do we see them as subjects in their own right, or as expansions of our human selves? The paper shows how the three works discussed portray machine vision in three different ways: as an expansion of human perception, as a hostile, controlling force that should be destroyed, and as a commercialised force altering or co-constructing the way we view our own humanity.

Key Words: Machine vision; drones; Google Cardboard; VR; posthumanism.

'Now objects perceive me,' the painter Paul Klee once wrote in his notebook, according to Paul Virilio in *The Vision Machine* (1994, orig. 1988). With the Internet of Things, objects that perceive us have become reality. Cameras watch us from satellites and drones, sharing information and using facial recognition algorithms to track individuals. Home surveillance systems measure air quality and send messages to parents when their facial recognition algorithms identify that a child has returned home from school. Alexa and Siri listen for our voices and answer our questions with information from the cloud.

Up until the last few decades, human vision was the only visual perspective available to us. Visual technologies such as drawing, photography and video give us an indirect access to other humans' visual experience of the world, but still controlled by humans. We know other animals have different senses than us, but do not have direct access to their visual experience. We have only had direct access to our own visual experience, and indirectly, through the technologies of cameras and drawings, to that of other humans' visual experience.

Now we also have access to a nonhuman form of vision: the machine vision of drones and satellites and surveillance cameras. Machine vision is the registration, analysis and representation of visual information by machines. Pre-digital technologies such as cameras gave us a taste of how machines perceive the world, but today's algorithms go even further in pushing us to see machines as perceiving beings with some kind of agency.

Of course visual machines and the software they run are built by humans, but we have designed our machines so that they are increasingly autonomous (Flusser 2000). We have also designed our laws and regulations to increase the autonomy of machines. For example, international maritime regulations require ships to use electronic chart display and information systems (ECDIS), which leave human captains and navigators acting as supervisors watching screens rather than being active navigators. Surveillance systems in cities or homes automatically track individuals over time and space, only alerting human operators if something unusual occurs (Andrejevic & Burdon, 2015). Self-driving cars similarly rely on visual sensors and other positioning systems to navigate through traffic, and humans need only take action if something goes wrong. Google Photos sorts through our personal photographs and generates videos and animations from them based on image recognition algorithms, and then asks us which of its creations we wish to keep. In these ways machine vision becomes primary, with humans only being notified when the machine identifies something it has perceived as being valuable or dangerous.

Sometimes, machine perception is integrated with human perception but designed to bypass conscious human decision-making. For instance, researchers claim they can show humans images ‘more rapidly than [humans] can consciously process in order to use their brains, hooked up to EEG monitors, as a detection device’ (Andrejevic, 2015 paragraph 4-7). Here human vision, or rather our brains’ processing of visual stimuli, simply provides an extra level of analysis to aid the machine in its work. Rather than us using the machine as a tool, we have designed a machine that uses us. In this posthumanist world, the human is no longer automatically the subject who perceives the world (Hayles, 1999). Objects may just as well perceive us, as Paul Klee wrote, or perhaps more accurately, objects may perceive *with* us.

Art is an area where machine vision is being examined and critiqued in very creative and provocative ways (Greene, 2015; Zylinska, 2016). This paper analyses three recent works of art that interrogate the relationship between human perception and machine vision: Nadav Assor’s art-documentary *Lessons on Leaving Your Body* (2014), Muse’s VR music video *Revolt* (2015), and Erica Scourti’s *Body Scan* (2014), with the goal of understanding how these works present the relationship between human and machine vision.

Drones as Alternative Bodies

In artist-filmmaker Nadav Assor’s documentary *Lessons on Leaving Your Body* (2014), we see an example of a human using machine vision to see the world and himself from another perspective. The human user, Jake Wells, sees himself as a ‘fleshpilot’ who happens to be temporarily in control of his physical body. The film begins with footage shot by one of Wells’s drones, descending slowly between trees until we see Wells himself standing on the forest floor, holding a controller and looking up at the drone. The drone crashes among the trees and even after it is still we see and hear its perspective on the world: moss and grass, very close to the camera. Part of the screen is black. A soft, regular beeping, apparently emitted from the crashed drone, reminds me of the beeping machinery keeping humans alive in hospital beds. The film cuts to a static, wide shot showing Wells sitting at a table on his deck, drone parts spread out before him. He speaks to the camera, explaining his ideas about the drones. He sees the drone as being a body he can inhabit, just as he inhabits his body of flesh, and for him, this comparison is religious: ‘When Jesus comes back, he’s going to give us a new body,’ he explains. His body is broken, just as his drone is broken, he explains, but both his flesh body and his drone can be repaired or replaced.



Figure 1: This still image from Nadar Assav's film *Lessons on Leaving Your Body* (2014) shows the protagonist, Jake Wells, with his head down as he views the world through the camera of the drone he is flying overhead. Reprinted with the artist's permission.

Next, the camera slowly pans up from Wells' hands holding the drone's controller to his bowed head. We hear the whirr of the drone, the cawing of a bird. Cut to a painterly landscape shot, the camera static as Wells sits, motionless, on a rock, his head between his knees as though in prayer. The drone flies into the frame of the image, white against the light grey sky (see fig. 1). Cut to a closeup of Wells' bent head, his breathing heavy and his eyes staring into a black visor to see the camera footage from the drone, then we see the drone's camera view of the same scene, and Wells speaks in voiceover: 'Your flesh is not who you are. You are flying that flesh (..) My body is just completely a rock, it's not even moving, it's just sitting there, you know.'

While Wells seems to see his human body and its sensory systems as equivalent to the machine body of the drone, the visual and auditive aesthetics of the film show stark contrasts between machine and human. The human-controlled camera footage is calm, carefully composed and almost motionless, whereas the drone footage is blurry and full of chaotic movement and crashes. And yet this apparent contrast is sometimes broken, as in the opening shot which begins with meditatively beautiful images of a forest seen from above, until the drone crashes in the undergrowth. We never hear the filmmaker speak. Only Wells, the 'flesh-pilot,' is given a voice in the film. Perhaps we should also include the drone as a speaker: we hear its whirr, its beeps when it has crashed, and we see the world through its perspective.

FPV (First Person View) videos taken from the perspective of drones are a common genre on YouTube, and while Wells' religious take on them may be unusual, experiencing the drone as inhabiting a body is a fairly well-established theme. 'I put on the specs and on a sudden, I am able to move about in a space where I normally can't be,' another FPV-pilot explains in an academic article about drones (Jablonowski, 2015). This sounds like something that goes beyond McLuhan's idea of media as extensions of ourselves.

When drones are used as military and surveillance tools, images captured by them are of course combined with algorithms that recognize targets and that can take action. Wells doesn't use algorithms to analyse the images captured by his drones. He simply wants to see the world from another perspective, from the point of view of a different kind of body. In a sense Wells

does not really crave machine vision, he craves human vision from a machine body. Or perhaps it is more correct to say he sees human and machine vision as interchangeable, as perfect equivalents. The film itself shows how different the machine vision of the drone is from the selective, aestheticizing vision of a human. For Wells, though, machine vision is not dangerous or threatening, it is simply an expansion of our perceptual possibilities.

Seeing as a Drone

Even without owning a drone or an Oculus Rift, we can experience new ways of seeing by slipping a smartphone into a cheap or homemade Google Cardboard kit and using a VR app like VRSE to view stereoscopic, 360° videos. One of the initial works released for VRSE in 2015 was a VR music video for ‘Revolt’ by Muse. The video is focalized through the camera in a security drone. As we can view the scene in all directions by turning around, we also feel some agency, almost as though we *are* the drone, or at least its operator. The non-VR version of the video uses fisheye lens distortion to give the impression it was recorded by a drone, but the video cuts between these ‘drone’ images and traditional cinematic footage, so the focalization is not consistent.

As the VR video begins, you see code all around you, like a startup screen, followed by white text against a black background as you hear a crowd and sounds of machines: ‘The year 2025. Government drones fill the sky. Freedom has been banished. Until now.’ You quickly realize that you are positioned as a moving drone, sometimes at ground level, sometimes right next to the musicians in the band, and sometimes above the women who storm the police. Sometimes it seems you (that is, the drone) fire upon the women. As police officers die, their bodies are labeled ‘asset casualty’, and the rebel women’s faces have numbers attached to them, emphasising the way that human identity is flattened into data by machine vision and its algorithms (Gates, 2011 loc. 262).

Even the simple VR experience offered by Google Cardboard is dizzyingly visceral. There is a strong sensation of actually being there, although with somewhat blinkered vision. As the drone swoops down, my stomach feels the movement, despite my knowledge that I am standing on the floor of my home. Watching from the point of view of the drone, with no power to intervene, the viewer is situated as part of the regime that the heros are revolting against. At the same time, the lyrics speak directly to the viewer/listener as though it is they who must revolt. The overall story of *Drones*, the album the song ‘Revolt’ is taken from, tells of a person who becomes a drone-like soldier, a ‘killing machine,’ but who in ‘Revolt’ breaks free and becomes an autonomous human subject again. The lyrics speak to this protagonist, and the second person address of the ‘you’ places the listener in the position of the protagonist:

I can see you're trapped in a maze
Let's find a way to escape
(..)
You're not afraid
You're not a drone
(..)
You can revolt

Visually, however, the viewer remains locked in the position of a drone, although the mechanical body of the drone is replaceable. Towards the end of the video, the lead singer kicks the drone we inhabit, and for an instant, we see the world through cracked glass, before our point of view suddenly shifts to a different position, above the rebels. The video ends when this second drone is shot down and the video ends. We remain complicit in the oppressor’s regime

until our systems fail and we have no choice but to accept that the fight is over—or to become fully human again, released from this machine body as we take off the cardboard headset.

Positioning the viewer or player as a drone or a machine is a fairly common technique in current digital culture. Games like *République* (2013-16) and *Watch Dogs* (2014) position the player as watching the action through surveillance systems, which is a similar position to that of an omniscient narrator in a novel—except that the machine vision of the surveillance cameras, while shown as vastly superior to the singular human point of view, is still presented as partial and corruptible.

While the player of *République* or *Watch Dogs* has some agency, and, at least in *République*, has the sense of using the oppressors' tools against them, the viewer of *Revolt* has no agency. *Revolt* presents the machine vision of the drone as though it is the enemy of humanity. All we can do is watch, and turn around to look at the video from other angles. Yet we must revolt against this oppressive system that we are trapped inside and complicit in, the lyrics and video insist that we must, and the words tell us that we can do so, even if we are only armed with flares against police officers in full body armour with tanks and drones. The relationship between human and machine is presented as a battlefield where only one side can win, and where the machines will relentlessly attempt to destroy our humanity. And yet, as humans, we are trapped, simultaneously fighting the machines yet complicit in their oppressiveness.

Searching your body

Erica Scourti's *Body Scan* uses a more intimate form of technology that offers another form of machine vision: image recognition software for smartphones. CamFind is an app that allows users to visually search the internet for items captured with the phone's camera. The app is aimed at shoppers, so if you point the camera at a watch the app will identify the brand of watch and point the user to a range of websites where they can purchase similar items. CamFind can identify a surprisingly broad range of objects, and artist and poet Erica Scourti has made use of this feature to create *Body Scan*, telling a story from a series of images of her body.

Body Scan is a 5:02 minute long video in portrait layout, like the screen of an iPhone, and most of the video is edited from recorded sessions with the CamFind app. An excerpt of the work is available online, but the full work can only be seen in the art exhibitions where it is displayed, where it is shown on a large screen, turned to portrait view, which makes the screen look a lot like a giant iPhone.

The piece begins with soft images in the rosy colours of dawn fading one into the next, so unfocused it is not possible to tell what they represent. Scourti's voice narrates what appears to be a moment between lovers, perhaps one waking before the other and gazing at the other's body: 'Lying comfortably in your bed, while you were still asleep. Feel your body as if for the first time. Gently place your attention on your foot. A human foot.' Already we are not quite sure whether 'your body' means the lover's body or the narrator's body, or whether the narrator is perhaps a third entity speaking to a human or humans who are lying in bed. A third confusing element appears as the meditative visuals are replaced by a screenshot of the CamFind app with a photo of a foot (see fig. 2). It is quickly replaced by search results: images of human feet fill the screen (fig. 3).



Figure 2: Still image from Erica Scourti's BodyScan (2014). Reprinted with the artist's permission.

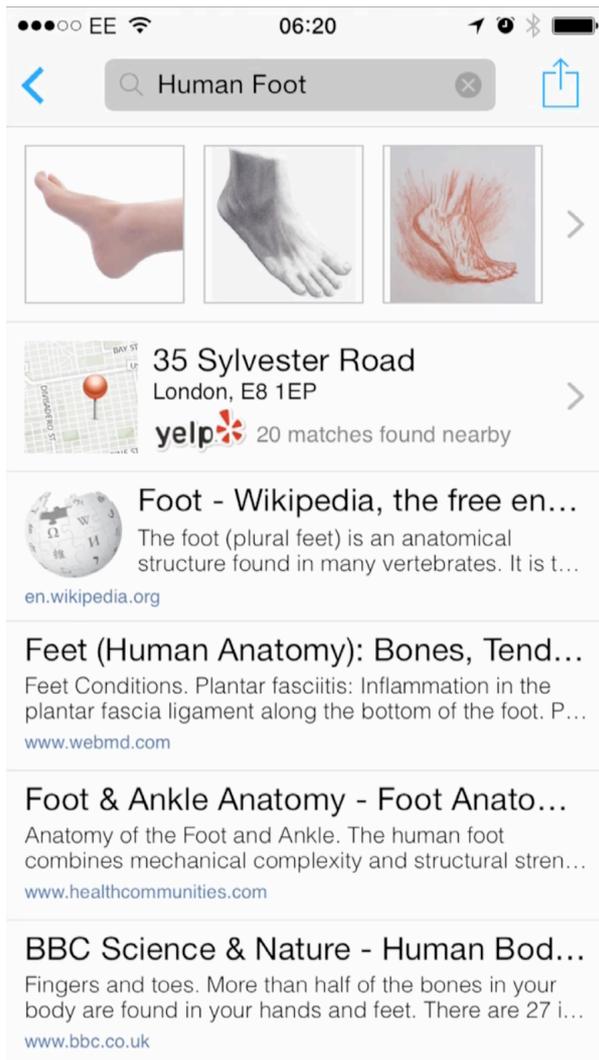


Figure 3: Still image from Erica Scourti's *BodyScan* (2014). Reprinted with the artist's permission.

A photo of breasts fills the mobile-phone shaped screen for a moment, quickly followed by CamFind's result: 'woman breast', then moving on to the search results: breast enlargement, fast enlargement. A quick montage of various male and female body parts follows. 'Identifying human, human armpit, human feet.' CamFind's merging of the personal and the consumerist machine algorithm is skilfully mocked. A picture of a foot returns the search result 'Baby' and the voiceover reads 'Baby, I can't wait forever 21', shifting from the human emotion of 'I can't wait forever' to shopping as 'forever' is coupled to '21', thus becoming the women's clothing store chain Forever 21.

Using images of your body to tell a story is a familiar technique in electronic literature. Shelley Jackson's *My Body—A Wunderkammer* (2006, orig. 1997) is a well-known early example, where a fragmented, white-on-black print of a woman's body is the interface to a story. Each part of the body is hyperlinked to a text about that body part, and the texts are also densely interlinked. In both pieces the texts are associative, but in Jackson's piece, the associations are hypertextual and personal, while *Bodyscan* deliberately plays upon the commercialised search associations of the CamFind app. Each work relies on the logic of the visual: Scourti on the automated vision of the app, and Jackson on her descriptions of how she drew each part of her body as a child. Vilém Flusser (2000) writes of the apparatus (in this case CamFind or drawing) as containing a fixed set of possibilities, which the artist attempts to

subvert. Scourti's piece very much includes CamFind as an 'active cognizer,' to use N. Katherine Hayles' term (2004). The logic of the app determines the way Scourti can tell a story about herself and her own body, but she subverts the commercial filters of the app by insisting on showing it her own body instead of brandname items, and her voiceover goes well beyond the simple verbal responses of the app. While *Lessons on Leaving Your Body* explores contrasts and similarities between human and machine vision, and 'Revolt' shows machine vision as hostile to humans, *BodyScan* uses machine vision to explore and rewrite the interplay between humans and machines.

Human Vision, Machine Vision

Our perception of the world is becoming more and more entwined with machine perception, and this is something that art, games and popular culture are increasingly grappling with. As we try to understand machines as things-that-sense we rethink our own ways of sensing and seeing (Rettberg, 2014). Sometimes this means that we limit our ideas of what can be seen to that which our machines can see, which can lead to a world seen as a set of binaries or patterns that simplify reality (Losh, 2015) or to a biased or even racist version of the world (Roth, 2009). *BodyScan*, *Lessons* and 'Revolt' all show that although we often start by assuming machine and human vision are separate, these two modes of vision profoundly influence each other. 'Now objects perceive me,' Paul Klee wrote. This machine vision is developing in symbiosis with human vision, and human perception is changing too. We see differently as we begin to see as machines can see.

References

- Andrejevic, M. (2015). The Droning of Experience. *The Fibreculture Journal*, (25), 203–218.
<http://doi.org/10.15307/fcj.25.187.2015>
- Andrejevic, M., & Burdon, M. (2015). Defining the Sensor Society. *Television & New Media*, 16(1), 19–36.
<http://doi.org/10.1177/1527476414541552>
- Flusser, V. (2000). *Towards a Philosophy of Photography*. London: Reaktion Books.
- Gates, K. (2011). *Our Biometric Future: Facial Recognition Technology and the Culture of Surveillance*. New York: NYU Press.
- Greene, D. (2015). Drone Vision. *Surveillance & Society*, 13(2), 233–249.
- Hayles, N. K. (1999). *How We Became Posthuman*. Chicago: University of Chicago Press.
- Hayles, N. K. (2004). Print is Flat, Code is Deep: Rethinking Signification in New Media. *Poetics Today*, 25(1), 67–90.
- Jablonowski, M. (2015). Drone it Yourself! On the Decentering of "Drone Stories." *Culture Machine*, 16.
- Jackson, S. (2006). My Body--A Wunderkammer. In N. K. Hayles, N. Montfort, S. Rettberg, & S. Strickland (Eds.), *Electronic Literature Collection*. The Electronic Literature Organization. Retrieved from http://collection.eliterature.org/1/works/jackson__my_body_a_wunderkammer.html
- Losh, E. (2015). Feminism Reads Big Data: "Social Physics," Atomism, and Selfiecity. *International Journal of Communication*, (9), 1647–1659.

Rettberg, J. W. (2014). *Seeing Ourselves Through Technology: How We Use Selfies, Blogs and Wearable Devices to See and Shape Ourselves*. Basingbroke: Palgrave.

Roth, L. (2009). Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity. *Canadian Journal of Communication*, 34(1). Retrieved from <http://www.cjc-online.ca/index.php/journal/article/view/2196>

Virilio, P. (1994). *The Vision Machine*. (J. Rose, Trans.). University of Indiana Press.

Zylinska, J. (2016). The Creative Power of Nonhuman Photography. In Zylinska, Joanna, K. Kuc, J. Shaw, R. Varney, & M. Wamposzyc (Eds.), *Photomediations: An Open Book* (pp. 203–224). Open Humanities Press. Retrieved from <http://photomediationsopenbook.net>