

Edge Detection

Shane Denson

There is a remarkable (and often remarked) scene a little over halfway into Denis Villeneuve's *Blade Runner 2049* (US/UK/Hungary/Canada/Spain, 2017) that speaks powerfully to our changing media landscape and the uncertain relations that it generates between human life and perception on the one hand and the spatiotemporal situations created by digital images on the other. The scene in question is a sex scene involving three characters: the Blade Runner known as K (Ryan Gosling), his AI-powered and holographically projected girlfriend Joi (Ana de Armas), and a sex worker named Mariette (Mackenzie Davis) who has been hired to lend Joi a tangible body in order to make love to K. This unusual ménage à trois requires Mariette to stand in corporeally for the immaterial Joi, whose likeness is projected onto the sex worker's body, thereby forming a composite of the two characters—and indeed, *compositing* is the technique used to merge the two actors' bodies in digital screen space.¹ But rather than blending together into a perfect, seamless union (an effect certainly within the power of the VFX team), Joi's and Mariette's images remain distinct from one another, slipping in and out of phase in imperfect alignment.

The emphatic *seamfulness* of this union is foregrounded as a visual spectacle in its own right. As a spectacle, it is staged not as something to be taken for granted but as something that has to be achieved. Thus, before the composited women can jointly make love to K, they must, with some effort, be synced up with one another: Mariette stands facing K while Joi slowly approaches; the two women regard each other briefly, and Joi tries to match Mariette's position in the room; the holographic woman flickers slightly

while she gets into position; the two roughly aligned women shift back and forth and look in opposite directions; then they raise their right hands and look down to inspect them; the camera cuts to a subjective shot, either from one or both of the women's perspectives; the hands are waved back and forth several times, one of them lagging noticeably behind the motions of the other, until they suddenly snap into place and line up visually with each other. Cutting back to the women's faces, still visibly out of phase, Mariette smiles and says with some amazement: "Look at you!" (thus giving voice to the movie's self-reflexive display of its effects, which demand the viewer to *look* at them). Joi, more seriously, responds: "Quiet! Now I have to sync!" And now, indeed, the women's faces and bodies line up, but they continue to alternate between the one and the other as more focally dominant—just as if someone were playing with the transparency settings of each layer in Photoshop or AfterEffects.

Because the synchronization remains incomplete, it retains an obstinate visibility. The scene is therefore something like those VFX reels that digital effects studios upload to YouTube or Vimeo to show off their work: videos that peel back the multiple composited layers of CGI textures, particle physics, and simulated lighting to reveal the technical complexity behind the finished images in blockbuster movies, a complexity that is otherwise invisible in its seamless integration.² But here, through intentionally seamful compositing, the execution of VFX is put on display in the movie itself. Paradoxically, the very failure of integration is situated as a visual spectacle.

The images' imperfect temporal synchronization and seamful spatial alignment serve to focus the movie's thematic interrogation of boundaries between human and artificial being, between life and nonlife, and between real and fake (just prior to syncing with Mariette, Joi assures K: "It's OK, she's real. I want to be real for you."). But staged in this way, the images also layer atop this questioning a further interest in imaging processes themselves—an interest that is motivated narratively by the presence of the holographic Joi, of course, but that goes beyond that level and aims squarely at a problematization of the images on the spectator's screen. The scene therefore synthesizes—or composites—these levels materially: the question of the digitally mediated image becomes transposable with the question of the definition of human life and perception itself. So, like the characters on screen, the spectator too is enlisted in a questioning that might be termed the problem of *edge detection*: Joi's struggle to stay in sync with Mariette translates a computationally demanding task of matching a moving object's

visible outline in real time; K's job (as Blade Runner) of discerning between human and nonhuman entities is twisted into an effort to *see* (both visually and conceptually) the two women as one without thereby confusing them; and the viewer is tasked with mediating between diegetic and nondiegetic complications of the image, along with the resulting thematic and material confusions of human and technological agencies.

Technically speaking, *edge detection* refers to a set of computational processes that are fundamental to machine vision, computer vision, and automated image processing; it encompasses algorithms that identify discontinuities in brightness within digital images, extracting line segments called *edges* that correspond (ideally) to the actual edges of physical objects or symbols.³ Edge detection is implemented in applications ranging from optical character recognition (OCR), automatic license plate readers (ALPR), industrial robotics, drones and self-driving cars, and other areas where a computational system is responsible for detecting, recognizing, classifying, or processing visual phenomena.

Abstracting from this technical understanding, we might relate the concept of edge detection to broader issues of perception in an age of digital mediation. As the technical implementations of edge detection algorithms demonstrate, many contemporary images are radically *discorrelated* from human perception, and even the ones that we see with our own eyes are impacted by the transformative effects of computational processing.⁴ From popular worries over the veracity of images in an age of Photoshop and CGI to scholarly debates over the loss of indexicality and its meaning for our experience of moving images, the shift from a cinematic to a post-cinematic media regime has introduced all sorts of uncertainties into our relationships with images and the underlying technologies that produce and support them. Nearly all of the images that we see today—whether in a movie theater equipped with digital video projectors, on a smart TV connected to digital cable or a BluRay player, or on a handheld device or computer screen streaming video from the Internet—virtually all of these images have been processed by a computer prior to our seeing them, and much of this processing is done on the fly (for example, in the cases of video decompression, upscaling, or motion smoothing). At stake in these processes are, among other things, the “edges” that visually define the objects of our perception. Compression algorithms, which are responsible for making playback and streaming more efficient by reducing the amount of digital information that must be processed at any given time, work by eliminating

perceptually “extraneous” data—minor differences of intensity or hue between pixels that would be imperceptible under normal circumstances. For example, employing motion estimation algorithms to determine which information is redundant, codecs such as MPEG-2 and MPEG-4 are crucially concerned with determining and preserving the shapes of moving objects, defined in terms of edges moving across the two-dimensional plane of the screen.⁵ But if the compression settings are off, or if there are errors in the execution of the codec’s algorithms, perceptual edges are subject to effacement or exaggeration: the “false” edges of blocky compression artifacts might even be generated where no edge should be. Video compression therefore involves a precarious balancing act between computational and human perception, a seamful negotiation between human and nonhuman ways of seeing and processing visual information.

The seams, usually hidden from view, are perhaps nowhere more visible today than in the controversies surrounding “DeepFake” videos and other AI-generated imagery. Named after the Reddit user “deepfakes,” who in 2017 uploaded a series of face-swapped fake celebrity porn videos featuring the likenesses of Scarlett Johansson, Gal Gadot, Taylor Swift, and others, DeepFake videos use machine learning algorithms known as “generative adversarial networks” (GANs) to automate the morphing and superimposition of images (such as a celebrity’s face) onto unrelated video (such as pornographic clips or political speeches).⁶ The danger that such techniques will be used for disinformation campaigns or “fake news” has lately garnered increased attention, but it is telling that the most popular uses of DeepFakes so far have been in the realm of fake celebrity porn and revenge porn, almost exclusively targeting women.⁷ Taking notice of this trend, Google has recently added the category of “involuntary synthetic pornographic imagery” to its list of banned content types.⁸ Reporting about DeepFake videos emphasizes alternately that faces are “seamlessly grafted . . . onto someone else’s body,”⁹ or that the syncing of the images “isn’t perfect.”¹⁰ For example, the December 2017 article widely credited with making the DeepFake phenomenon known to a wider public notes with regard to the fake porn video featuring *Wonder Woman* star Gal Gadot: “In the Gadot video, a box occasionally appeared around her face where the original image peeks through, and her mouth and eyes don’t quite line up to the words the actress is saying—but if you squint a little and suspend your belief, it might as well be Gadot.”¹¹ Imperfections are evident here, but there is leeway “if you squint a little”—thus signaling a real perceptual indeterminacy.

Clearly, though, technical capabilities are evolving quickly, and more sophisticated algorithms may well erase these seams. What these discussions of the user's (in)ability to detect the edges in DeepFakes and GAN-generated faces point to, however, is the low-level imbrication taking place between technological imaging processes and human perception, whereby the borders between human and nonhuman, living and nonliving are called into question. The generative capabilities of artificial intelligences do not constitute life, to be sure, but the microtemporal operations involved in algorithmic imaging put them into close proximity with the subperceptual processes of human embodiment, potentially altering the metabolic pathways that mark our broadly ecological entanglements (all mediated today by technological apparatuses and systems) and that structure our own pre-conscious processing of time and space. Pornography presents itself as an obvious test case, as it aims to short-circuit subjective cognition, affecting viewers' bodies directly by "animating" sexual desire. In this way, the life-giving force of *animation* is plunged into a realm of indistinction, referring at once to technological processes like neural network-driven CGI, science fiction fantasies of artificial creation, and biological facts of life. In this context, edge detection reveals itself as a both technical and human-epistemological feat that leads straight back to the "edges" of human-technological interfaces as well as gender's blurred and contested lines, all of which are wrapped up in issues of generativity (with the Proto-Indo-European root **gene-* implicating gender, genre, genetics, and the generativity at the heart of CGI).

Obviously, such videos raise serious ethical and political questions, but the notion that they constitute the sort of ontological-epistemological nexus described above will likely strike many readers as hyperbolic at best.¹² And while it is beyond the scope of this short essay to fully substantiate these claims, a brief look at the phenomenology of DeepFake videos might be instructive.¹³ Such videos split viewers' attention between representational and presentational levels, asking the viewer to alternate between the objects depicted and the manner (or the quality of technical execution and plausibility) of their depiction. At stake in both forms of regard is the question of the visual edge—that which delineates the object itself and that which belies its fabrication. This split or oscillating form of regard decenters conventional modes of spectatorial engagement, deprivileging narrative interest and sutured engrossment in the diegesis. Instead, human perception is brought into closer contact with computer vision. "Object recognition" in

the latter involves scanning images for objects, conditions, or information. Looking at DeepFakes engages the viewer in a similar task, which demands a loosening of focal depth and a flattening of the distance between representation and mediation—the viewer essentially assumes a machinic vision not unlike the computer’s as it mechanically and non-subjectively scans pixels to recognize objects without regard for the difference between the image and that which it depicts. (It goes without saying that in the context of pornography, this de-subjectivizing of the spectator does nothing to mitigate the objectification of women’s bodies.)

Meanwhile, various attempts to recognize fake imagery seek to reassert the difference between human perception and computer vision: artists attempt to disentangle humans’ vision from that of AI by training our eyes to see telltale signs of computer-aided forgery,¹⁴ while DARPA and others approach the problem from the other side, turning computer vision back on itself in order to automate the process of debunking fake images.¹⁵ However, the fact that these efforts must be constantly updated points to the underlying fact that human vision and machinic visuality are fundamentally co-implicated today, not just in GANs and other AI-based visualizations, but in digital imaging processes more generally, including the much more mundane compression algorithms mentioned above. These algorithms subject nearly all the images we perceive to processes known as *perceptual coding*: the balancing act whereby imperfections in human vision are exploited to reduce file size and streaming performance.¹⁶ As we have seen, such processes turn crucially on the presence of edges and a means of apprehending them: differences in brightness trigger different degrees of granularity and subdivide the visible image into (ideally) invisible blocks, eliminating perceptually redundant information.¹⁷ But then the “edge” at stake in edge detection, both broadly and narrowly construed, is the would-be seamless integration or synchronization of human and nonhuman forms of vision, whereby perceptual coding blends imperceptibly into a *coding of perception*.¹⁸

Seen in this light, *Blade Runner 2049*’s uncanny sex scene—which joins movies like *Her* (dir. Spike Jonze, US, 2013) and *Ex Machina* (dir. Alex Garland, UK, 2014) in imagining erotic relations between humans and artificial beings created under conditions of the digital—can be seen as a concise emblem of the contingencies, exigencies, and balancing acts involved in contemporary mediated life. The scene is unusual for its seamful display and focus on the difficulties of synchronization, which it stages as a question

of *images* but also of *animation*—a life-giving generativity that implicates biological and artificial life both within the diegesis and without, at the interface where an embodied viewer, a holographic woman, and the images' digital infrastructures meet. All of the principal actors are engaged in operations of edge detection: Joi and Mariette share the task of matching movement based on visible outlines—a task usually carried out by video codecs—while K plays the role of the spectator who might need to “squint a little” to fulfill the (gendered) fantasy of seamless integration, but whose perception remains under constant threat of breakdown in the form of a visual glitch.

As I have argued, this seamful display stages a not only technological but also profoundly philosophical problematic and can therefore be read as a visual allegory for our post-cinematic situation more generally. It points to the uncertainties and seamful negotiations of human and non- or post-human spatiotemporal situations that take place in, but that far exceed, mundane acts of viewing digital video. The scene unearths the deep imbrication of perceptual and technological capacities as they are distributed across human and nonhuman agencies, thus problematizing relations between human perceivers and computational images that, because they operate in spatial and temporal registers that are significantly different from those of embodied humans, are in an important sense dis-correlated from that perception. Most importantly, by staging this encounter as an emphatically seamful spectacle, the scene adds an aesthetic dimension that helps us to make sense of subperceptual encounters that, because they take place in a microtemporal interval inaccessible to subjective perception, categorically elude higher-order sensory presentation. The scene therefore mediates between human perceivers and the invisible conditions of life today. This is an aesthetics of edge detection.

Notes

- 1 For details about the making of the scene, see the somewhat hyperbolically titled article: Joe Skrebels, “*Blade Runner 2049*: How Denis Villeneuve Created the Most Complicated Sex Scene of All Time,” *IGN*, 15 February 2018, www.ign.com/articles/2018/02/15/blade-runner-2049-how-denis-villeneuve-created-the-most-complicated-sex-scene-of-all-time.
- 2 For a compelling and relevant example, see the VFX reel documenting the layers of compositing involved in the (extra-diegetic) creation of the fembot at the heart of *Ex Machina*: www.youtube.com/watch?v=4sFD-YbeIX4.
- 3 For greater technical detail, see chapter 5, “Edge Detection,” in Ramesh Jain, Rangachar Kasturi, and Brian G. Schunck, *Machine Vision* (New York: McGraw-Hill, 1995).

- 4 For a provocative account of images that circulate between machines, wholly apart from human
- 5 For technical details, see A. Murat Tekalp, *Digital Video Processing*, 2nd ed. (New York: Prentice Hall, 2015), and Iain E. Richardson, *The H.264 Advanced Video Compression Standard*, 2nd ed. (West Sussex: John Wiley & Sons, 2010).
- 6 For an accessible introduction to DeepFake videos, see Jonathan Hui, “How Deep Learning Fakes Videos (Deepfakes) and How to Detect It?,” *Medium*, 28 April 2018, medium.com/@jonathan_hui/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9.
- 7 See, for example, Robert Chesney and Danielle Keats Citron, “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security,” *California Law Review* 107 (2019), papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954.
- 8 Google’s addition of the category is reported in Drew Harwell, “Fake-Porn Videos Are Being Weaponized to Harass and Humiliate Women: ‘Everybody is a potential target,’” *Washington Post*, 30 December 2018, www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target.
- 9 Harwell, “Fake-Porn Videos.”
- 10 Samantha Cole, “AI-Assisted Fake Porn Is Here and We’re All Fucked,” *Motherboard*, 11 December 2017 motherboard.vice.com/en_us/article/gdydm/gal-gadot-fake-ai-porn.
- 11 Cole, “AI-Assisted Fake Porn.”
- 12 Indeed, issues of politics, perception, and media ontology are deeply intertwined in the category of “involuntary synthetic pornographic imagery”—a concept that encompasses the media objects of images, relates them to the “body genre” of porn, implicates generativity with the qualifier “synthetic,” and connects it centrally with a politics of volition, which is deeply troubled by DeepFakes.
- 13 A more comprehensive attempt to substantiate the claims in question is the subject of my forthcoming book, *Discorrelated Images* (Durham, NC: Duke University Press).
- 14 See Kyle McDonald, “How to Recognize Fake AI-Generated Images,” *Medium*, 5 December 2018, medium.com/@kcimc/how-to-recognize-fake-ai-generated-images-4d1f6f9a2842.
- 15 DARPA’s efforts to battle DeepFakes are reported in Will Knight, “The Defense Department Has Produced the First Tools for Catching Deepfakes,” *MIT Technology Review*, 7 August 2018, www.technologyreview.com/s/611726/the-defense-department-has-produced-the-first-tools-for-catching-deepfakes/.
- 16 See, for example, Zhenzhong Chen, Weisi Lin, and King Ngi Ngan, “Perceptual Video Coding: Challenges and Approaches,” *Proceedings of IEEE International Conference on Multimedia and Expo (2010)*: 784–89
- 17 I am referring here to the division of the image into “macroblocks,” a process that is crucial for reliable motion estimation between frames while maintaining high compression rates (macroblocks with no edges are less likely to change and can therefore be considered redundant, while blocks with edges may include boundaries of objects that move between frames). For technical details, see Richardson, *The H.264 Advanced Video Compression Standard*.
- 18 At stake, in other words, is the question raised by N. Katherine Hayles in *How We Became Posthuman* of “the limit to how seamlessly humans can be articulated with intelligent machines.” Hayles has explored this question more recently in terms of the

concept of the “cognitive nonconscious”—an important site of subperceptual interactions that are crucial to the questions I am pursuing here. See Hayles’s *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics* (Chicago: University of Chicago Press, 1999), 284; and her *Unthought: The Power of the Cognitive Nonconscious* (Chicago: University of Chicago Press, 2017).

Shane Denson is Assistant Professor of Film & Media Studies in the Department of Art & Art History at Stanford University. His research and teaching interests span a variety of media and historical periods, including phenomenological and media-philosophical approaches to film, digital media, comics, games, videographic criticism, and serialized popular forms. He is the author of *Postnaturalism: Frankenstein, Film, and the Anthropotechnical Interface* (Transcript-Verlag/Columbia University Press, 2014) and co-editor of several collections: *Transnational Perspectives on Graphic Narratives* (Bloomsbury, 2013), *Digital Seriality* (special issue of *Eludamos: Journal for Computer Game Culture*, 2014), and the open-access book *Post-Cinema: Theorizing 21st-Century Film* (REFRAME Books, 2016). His next book, *Disrelated Images*, is forthcoming with Duke University Press. See also shanedenson.com for more information.