Biometric Citizens: Adapting Our Selfies To Machine Vision

Jill Walker Rettberg

Pre-print of chapter forthcoming in Adi Kuntsman (ed.) *Selfie Citizenship*. Palgrave 2016.

Machines don't see as humans do. Machines convert that which humans can see into data that can be analysed by algorithms. Light, dark and contrast are analysed and computed. Eyes, nose and mouth are identified, converted, stored. The curves and features of a face are condensed into 3D grids, allowing the machine to recognise another picture of that same face even if it is tilted to the side or shown from another angle.

Machines are good at this kind of vision. iPhoto uses facial recognition to sort our personal photos so we can easily find all our photos of each of our friends. Facebook's Deepface algorithm can recognise individual faces in a crowd almost as well as a human can (Taigman, Yang, Ranzato, & Wolf, 2014). But the way in which the machine identifies that face has little to do with our human way of recognising a familiar face.

Machine vision is about data, not about the visual or optical. Carolyn L. Kane sees this as such a fundamental aspect of today's society that she, building upon Kittler (2009), calls our time *post-optical,* arguing that we no longer use color or sight itself as an end in itself but as a means to another end (Kane, 2014). The military paints pixellated patterns on the tops of drones, not because humans will see the patterns but to fool the machines watching from satellites above us. Makeup like CV Dazzle is specifically designed not to make humans appear more beautiful to each other but to fool surveillance cameras trying to identify human faces in a crowd. With chromakey video, producers use a coloured background to let a machine know to replace that colour with something else. "Color is not exclusively about vision," Kane writes (211). "Rather, it is a system of control used to manage and discipline perception and thus reality." We could say much the same thing of selfies.

Much has been written about the ways in which selfies can be seen as a form of discipline, where we learn to shape ourselves in certain ways that our visual culture allow (Burns, 2015; Rettberg, 2014; Senft & Baym, 2015). But we can also see selfies as post-optical, post-visual and posthuman. When we post selfies online, our audience will not only be other humans. Our audience is also, and perhaps primarily, machines.

In an excellent article discussing selfies in the context of photography, Paul Frosch argues that it is the *nonrepresentational* aspects that are new in selfies: the "distinctiveness from older forms of self-depiction seems to derive from nonrepresentational changes: innovations in distribution, storage, and metadata that are not directly concerned with the production or aesthetic design of images" (Frosh, 2015). This is true as far as it goes, but it neglects the ways in which machines analyse our selfies. In this chapter, my aim is to show you how the machines that will read our selfies actually affect and control the way that we create them. Machine vision is changing the way we see ourselves.

The development of photography as a technology was in many ways decided by our society's desire to control citizens by identifying them. Photographs in identification documents and police mugshots are among the earliest uses of photographic portraits. Today it is impossible or at least extremely difficult to cross a border, buy alcohol, vote or in some cases pick up medication or a parcel addressed to you without being able to prove your identity with a

driver's license, passport or other photographic identification. For many decades these photographs were intended to identify us to other human beings, but now they are designed to be machine readable. We can no longer smile for a passport photograph, because the smile makes it harder for a machine to match the image to a fixed identity. This changes the way we think about faces. Paul Frosh argues that the selfie's indexicality is more deixic, pointing towards the social gesture of communication, and in fact, machines can also read this aspect by counting which people who like or comment on each others selfies. Yet it is the fixing of the face to a unique body that is the main goal of facial recognition, whether it is used by the government to control the movements and actions of citizens or by Facebook to provide relevant content to users and relevant users to advertisers.

New security systems are skipping the intermediary of the ID card and are instead running facial recognition directly on the image of our face. Refugees in Jordan are given cash assistance from the UN which they withdraw from an ATM without a bankcard or a PIN: instead the ATM scans their iris to identify them (Dunmore, 2015). Soon MasterCard customers will be able to prove their identity when making purchases online by blinking at the camera in their smartphones. In reading press reports of MasterCard's payment-by-selfie system, it is interesting to note the emphasis on the face's conversion to data:

> The facial recognition scan will map out your face, convert it to 1s and 0s and transmit that over the Internet to MasterCard. Bhalla [the spokesperson for MasterCard] promised that MasterCard won't be able to reconstruct your face -- and that the information would transmit securely and remain safe on the company's computer servers. (Pagliery, 2015)

There is an underlying assumption here that the customer would not want her selfie to be uploaded to MasterCard. This discomfort points to the difference between seeing a selfie as an end in itself and as a means to an end. Technically, looking at a camera that records an image of my face is the same as taking a selfie, but it is not a selfie that is intended to be published or to be seen by any other human, or even by myself. This is a selfie that is being used as something else.

Kelly Gates writes that facial recognition technology "treats the face as an index of identity, disregarding its expressive capacity and communicative role in social interaction" (2011, p. loc. 251). There are other modes of machine vision that can interpret selfies more broadly though, such as emotion analysis software.

Companies offer systems with names like Emotient, Real Eyes and Affectiva where the webcam on a test subject's computer captures video of their face as they watch an advertisement or other material on their computer. Their face is instantly analysed for emotions and the marketer or content creator who is paying for the service receives a report pinpointing the exact moments of their video that elicit joy, tears, boredom or frustration. Expression analysis is used in other situations, too. The marketing video for Affectiva shows a young boy using expression analysis with his Google glasses to interpret his older sister's responses to what he is saying (Affectiva, 2014). A video made for Emotient shows a smart car with a camera trained on the driver's face. When she frowns the screen on the dashboard is shown to read "Driver frustrated. Suggest alternate route" (Voltage Advertising + Design for Emotient, 2014).

Some of the early versions of expression analysis software were designed to help people with autism to understand emotions, but current marketing assumes that all humans would find this useful, especially if they are trying to sell something. Machines constantly watching us

and our homes is also posited as helpful in the marketing of home surveillance and communication devices like the Withings Home, a camera that alerts absent parents when a person comes near the device. In the marketing video, a child arrives home after school, and the mother, who is at work, receives a message on her phone and is able to start a video conversation with her son (Withings, 2014).

Most consumer-oriented machine vision systems, such as Facebook's facial recognition algorithms, Affectiva's emotion-sensing car and the Withings Home, are intended to feel helpful and to improve our lives. However, they are not neutral. In her study of users' experiences with heart-rate variability monitors, Minna Ruckenstein (2014) found that users told the story of their day differently before and after seeing the data from the heart-rate variability monitor, which indicates stress levels. The users negotiated with the data, adjusting their lived experience to match the story the data was telling. The more we outsource our emotions and social communication to machines, the less likely we are to trust ourselves.

Most machine vision strategies rely on the assumption that selfies can be broken down into individual parts. Faces become 3D grids, emotions are reduced to certain muscular movements, and a child arriving home is simply motion in a quiet living room. These machine vision techniques are also easily used by humans, especially when we try to use machines to understand more about a phenomenon such as selfies. Elizabeth Losh discusses the ways in which Lev Manovich's *Selfiecity* project "depicts human individuals as discrete elemental particles":

> Yet applying an atomistic social physics to selfie culture perpetuates the stereotype of the independent and autonomous self as isolated media creator and media subject in the cultural imaginary of personal consumer electronics and ignores how people are embedded in complex rhetorical situations. (Losh, 2015)

For digital machines to "see" selfies, this atomism is necessary. Digital algorithms rely on 0s and 1s. But what does this do to the ways we see ourselves? How does machine vision influence human vision?

One genre of selfies that illustrates such influence is the "time-lapse selfie" (Rettberg, 2014, pp. 36–40), where people take a photo of themselves every day and generate a time-lapse video showing their face changing day by day and year by year. The app Everyday automates this process in a way that makes the algorithm very visible. When you take a photo, you are asked to align your face on the screen with the selfie you took the day before. By aligning your eyes and mouth to the grid the app provides, you ensure that the video produced will be a smooth as possible.

Snapchat's popular selfie lenses similarly mark our faces as machine-readable. To activate the lenses, you use the front facing camera so you see your own face, and then touch the image of your face on the screen until a biometric grid pops up on your face. You can now swipe between different lenses that apply various special effects to your face, mapping them to the grid so that the effects follow the movements of your face. Some distort your face, removing your nose or twisting your chin, while others add makeup and accessories or make rainbows gush out of your mouth. Like funhouse mirrors that show us distorted versions of ourselves, Snapchat's lenses allow us to play with our visual identity, and of course to send silly selfies to our friends. They also normalize biometrics and automated image

manipulation. They make us more used to having our faces read by machines (Rettberg, 2016).

Artists have of course experimented with the many ways in which we use machines to see ourselves and our surroundings. CamFind is an app that will search the internet to find images the user captures with the phone's camera. The app is aimed at shoppers, so for instance if you point the camera at a crayon the app will identify the object as a crayon and point the user to a range of websites where they can purchase crayons. CamFind can identify a surprisingly broad range of objects, including parts of the human body. Artist Erica Scourti has made use of this feature to create *Bodyscan,* a short video composed of images she has taken of her body using CamFind and the results it finds in response to her images. I saw the full video at the Transmediale exhibition in Berlin in January 2015, and an excerpt is also available online (2015). The video is in portrait layout, like the screen of an iPhone, and is an edited recording of a session using the CamFind app, where Scourti reads the descriptions of items identified by the app as a voiceover. Sometimes she also narrates beyond the terms provided by the app, blending her human voice with the words of the machine. The video moves faster than the app, cutting away all the lag and cutting between search results at a frenetic pace while reading aloud the identification text the app produces.

A photo of breasts fills the mobile-phone shaped screen for a moment, quickly followed by CamFind's result: "woman breast", then quickly 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." The merging of the personal, the human, and the consumerist machine algorithm in CamFind 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.

Selfies are of course not wholly defined by the machines that read them. There is resistance against the machines, for instance in nude or sexually provocative selfies that are deliberately taken in ways that make any kind of automated facial recognition or other identity-fixing strategies difficult or impossible (van der Nagel, 2013). I remain optimistic, and believe selfies serve an important communicative purpose for humans. But it is important to also be aware of the ways in which our machines are steering and controlling the ways in which we can see ourselves. In many ways, we are becoming biometric citizens, identified and shaped by the digital images of our faces.

References

Affectiva. (2014). "Affectiva Overview" *YouTube*. Available from https://www.youtube.com/watch?v=mFrSFMnskI4&feature=youtu.be [Accessed 30 March 2016].

Burns, A. (2015). Self(ie)-Discipline: Social Regulation as Enacted Through the Discussion of Photographic Practice. *International Journal of Communication*, *9*, 1716–1733.

Dunmore, C. (2015, March 23). Iris scan system provides cash lifeline to Syrian refugees in Jordan. Retrieved October 7, 2015, from http://www.unhcr.org/550fe6ab9.html

Frosh, P. (2015). The Gestural Image: The Selfie, Photography Theory, and Kinesthetic Sociability. *International Journal of Communication*, *9*(0), 22.

Gates, K. (2011). *Our Biometric Future: Facial Recognition Technology and Our Culture of Surveillance*. New York: New York UP.

Kane, C. L. (2014). *Chromatic Algorithms: Synthetic Color, Computer Art, and Aesthetics After Code*. Univ. of Chicago Press.

Kittler, F. (2009). *Optical Media*. Cambridge, UK ; Malden, MA: Polity.

Losh, E. (2015). Feminism Reads Big Data: "Social Physics," Atomism, and Selfiecity. *International Journal of Communication*, (9), 1647–1659.

Pagliery, J. (2015, July 1). MasterCard will approve purchases by scanning your face. Retrieved October 7, 2015, from http://money.cnn.com/2015/07/01/technology/mastercard-facial-scan/index.html

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

Rettberg, J. W. (2016) "How Snapchat Uses Your Face (A Snapchat Research Story)." *YouTube*. 6 April 2016. Available from: https://www.youtube.com/watch?v=2XOd-rc7r98&list=PL46Xs2itPIMlDBL0tPfg-2WzXwZrnTePh&index=1 [Accessed: 19 May 2016].

Ruckenstein, M. (2014). Visualized and Interacted Life: Personal Analytics and Engagements with Data Doubles. *Societies*, *4*(1), 68–84. http://doi.org/10.3390/soc4010068

Scourti, E. (2015). *Body Scan (excerpt)*. Retrieved from https://vimeo.com/111503640

Senft, T. M., & Baym, N. K. (2015). Selfies Introduction ~ What Does the Selfie Say? Investigating a Global Phenomenon. *International Journal of Communication*, *9*(0), 19.

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1701–1708). http://doi.org/10.1109/CVPR.2014.220

van der Nagel, E. (2013). Faceless Bodies: Negotiating Technological and Cultural Codes on reddit gonewild. *Scan: Journal of Media Arts Culture*, *10*(2). Retrieved from http://scan.net.au/scn/journal/vol10number2/Emily-van-der-Nagel.html

Voltage Advertising + Design for Emotient. (2014). *Emotient - Emotional Awareness for Devices*. Retrieved from https://vimeo.com/70048902

Withings. (2014). *Meet Withings Home. Connect. Sense Protect.* [YouTube video]. Retrieved from https://www.youtube.com/watch?v=qUydD9ycSWQ