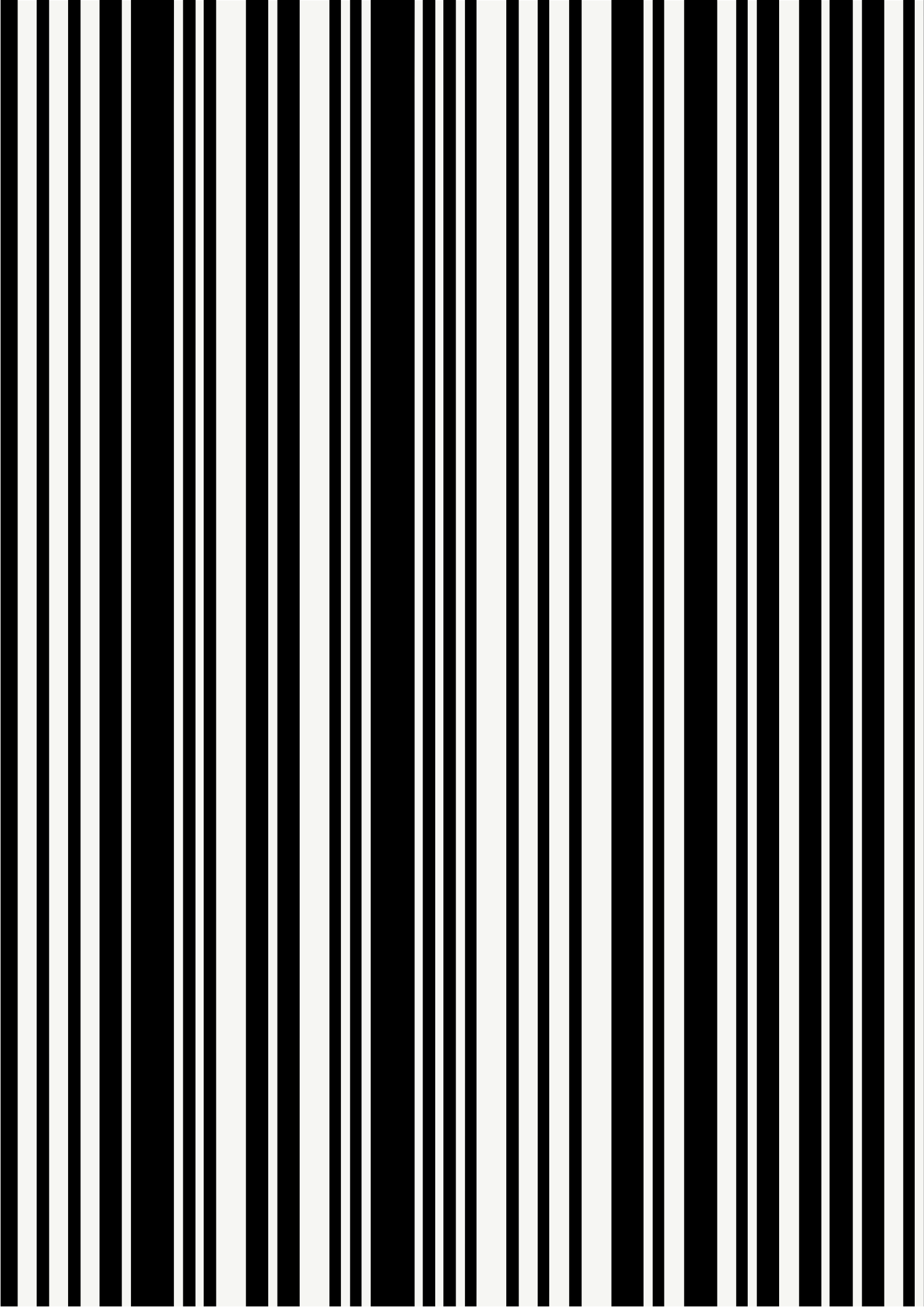




André Góes Mintz

**Visão computacional e visualidades contemporâneas:**  
composições do ver e do visível entre a técnica, a ciência e a arte

Programa de Pós-Graduação em Comunicação Social | Fafich | UFMG | 2015



André Góes Mintz

VISÃO COMPUTACIONAL E VISUALIDADES CONTEMPORÂNEAS:  
COMPOSIÇÕES DO VER E DO VISÍVEL ENTRE A TÉCNICA, A CIÊNCIA E A ARTE

Dissertação apresentada ao Programa de Pós-Graduação em Comunicação Social da Faculdade de Filosofia e Ciências Humanas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Comunicação Social.

Área de concentração: Comunicação e Sociabilidade Contemporânea

Linha de pesquisa: Pragmáticas da Imagem

Orientador: Prof. Dr. André Guimarães Brasil

Belo Horizonte  
Faculdade de Filosofia e Ciências Humanas da UFMG  
2015

301.16  
M667v  
2015

Mintz, André Góes  
Visão computacional e visualidades contemporâneas [manuscrito]  
: composições do ver e do visível entre a técnica, a ciência e a arte /  
André Góes Mintz. - 2015.  
215 f. : il.  
Orientador: André Guimarães Brasil.

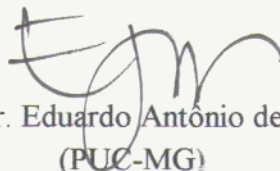
Dissertação (mestrado) - Universidade Federal de Minas Gerais,  
Faculdade de Filosofia e Ciências Humanas.  
Inclui bibliografia

1. Comunicação – Teses. 2. Visão por computador - Teses. I.  
Brasil, André Guimarães. II. Universidade Federal de Minas Gerais.  
Faculdade de Filosofia e Ciências Humanas. III. Título.

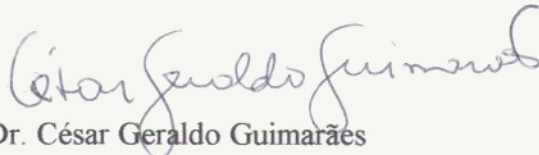
*“Visão computacional e visualidades contemporâneas: composições do ver e do visível entre a técnica, a ciência e a arte”*

André Góes Mintz

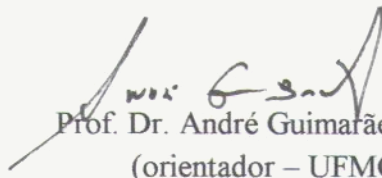
Dissertação defendida e aprovada pela banca examinadora constituída pelos professores:



Prof. Dr. Eduardo Antônio de Jesus  
(PUC-MG)



Prof. Dr. César Geraldo Guimarães  
(UFMG)



Prof. Dr. André Guimarães Brasil  
(orientador – UFMG)

Programa de Pós-graduação em Comunicação Social  
Faculdade de Filosofia e Ciências Humanas  
Universidade Federal de Minas Gerais  
Belo Horizonte, 23 de março de 2015.



## Agradecimentos

Gostaria, primeiramente, de agradecer ao meu orientador, André, não apenas pelo acompanhamento nestes dois anos, com uma leitura atenta e apoio próximo em todas as etapas, mas também pelo acolhimento generoso desta proposta de pesquisa, compartilhando de seus riscos e, ao mesmo tempo, concedendo-me espaço para experimentar.

À Geane Alzamora e ao Eduardo de Jesus, agradeço pela minuciosa leitura do texto na fase de qualificação e pelo apontamento de várias questões que tiveram participação crucial na elaboração do texto final. Ao Eduardo agradeço, ainda, junto ao César Guimarães, por toparem vir à banca final, bem como pelos incentivos, desde antes do mestrado, sempre estimulantes.

Aos professores com quem tive aulas e conversas neste período, agradeço pela carinhosa acolhida, especialmente: Ângela Marques, Luciana Oliveira, Vera França, Bruno Leal, Joana Ziller, Beatriz Bretas, Cláudia Mesquita, Roberta Veiga e Carlos Falci. Ao Carlos Mendonça agradeço, ainda, pelos comentários tecidos em seu parecer ao projeto, que tiveram fundamental importância no delineamento de seu percurso.

A meus professores noutros momentos, com os quais tive a oportunidade de cruzar em outras oportunidades e que, pela amizade e incentivos, não me deixaram perder o mestrado de vista, agradeço especialmente: à Laura Guimarães, ao Paulo B., à Patrícia Moran e, saudosamente, ao Rodrigo Minelli.

Aos colegas do curso, pelas conversas, encontros e botecos – partes integrantes desta formação, ainda não reconhecidas oficialmente pela CAPES. Às secretárias do programa, Elaine Martins e Tatiane Oliveira, pelo apoio e aconselhamento em meio aos trâmites – estes, sim, oficiais.

À Fernanda Bruno, agradeço pelo interesse no projeto e pelas referências compartilhadas em nosso breve encontro. Neste mesmo ensejo, agradeço ao Marcus Bastos, ao Gabriel Menotti, ao Sean Cubitt e à Priscila Arantes, bem como aos participantes e organizadores dos eventos em que pude apresentar etapas desta pesquisa e que renderam boas interlocuções, integrantes do processo de investigação.

Aos amigos (colegas e estudantes) da AIC/Oi Kabum!, pela partilha. Sinto-me privilegiado por integrar esta escola e devo destacar que considero-a, tanto quanto o mestrado, uma instância fundamental de minha formação nesses últimos dois anos. Agradeço-lhes, *todos*, pelo convívio, pela compreensão e valorização das eventuais ausências necessárias, ajudando-me a conciliar estes dois percursos.

Aos amigos do Espaço do Conhecimento UFMG, também, pelo estímulo, apoio e compreensão, inclusive do meu derradeiro afastamento, pela urgência da escrita. Devo à experiência em meio a este grupo, certamente, muitos dos interesses e referências que desembocam neste trabalho.

Aos demais amigos e amigas que não cheguei ainda a mencionar, com muitos dos quais fico praticamente dois anos em débito, espero que eu logo possa começar a sanar minha dívida, começando por este agradecimento, em especial a: Clá, Lu, Horts, Marina, X, Dé, Dou, Ricardo, Aline, Sara e Gabi.

À minha mãe, pai, queridas irmãs e à Tau, bem como ao restante da família, agradeço enormemente pelo incentivo e pelo apoio, em todos os momentos, e pela compreensão de minhas ausências e divagações.

Por fim, à Naína, nem sei por onde começar: agradeço por tudo, e mais um pouco. A companhia, o apoio, a ajuda, o cuidado, o carinho, a escuta. Por me lembrar de distrair e descansar. Por me lembrar do resto e me fazer mudar de assunto.



Este ensaio é um argumento em favor do *prazer*  
da confusão de fronteiras, bem como em favor  
da *responsabilidade* em sua construção.

Donna Haraway



## Resumo

Este trabalho aborda, por uma perspectiva dos estudos da imagem, o domínio tecnológico da Visão Computacional, tipicamente compreendido pelo amplo conjunto de operações algorítmicas pelas quais programa-se uma máquina para a interpretação automatizada de informações visuais. Busca-se descrever o fenômeno de emergência e disseminação de suas máquinas derivadas, em um plano não estritamente técnico, atentando-se, em particular, aos modos com que agenciam, em seu funcionamento, a *imagem* e o *observador*. Amparado pelo referencial teórico no entorno do conceito foucaultiano do *dispositivo*, o trabalho compreende o domínio estudado por suas instanciações concretas, em formas de regulação e governo do *ver* e do *visível*, portanto profundamente atravessado por formações do saber, demandas institucionais, práticas de observação e discursos técnicos, estéticos e científicos – dentre outros elementos. A fim de tornar visíveis tais articulações, e de passar de uma constituição técnica a agenciamentos estéticos, sociais e políticos, o percurso teórico-analítico é elaborado pela reunião de múltiplos objetos, selecionados dos domínios da técnica, da ciência e da arte, segundo seu potencial de problematização de aspectos relacionados aos modos de agenciamento deste dispositivo. Tal esforço é animado pelo gesto metodológico da *composição*, proposto por Bruno Latour, pelo qual busca-se construir, por conexões e aproximações, o plano comum de nosso fenômeno de interesse a partir da multiplicidade de suas manifestações heterogêneas. Acerca do modo de agenciamento da imagem e do visível por este domínio, desenha-se um movimento alternante entre, de um lado, a singularização, a localização e a identificação e, de outro, a generalização, a categorização e a idealização. Quanto ao seu modo de agenciamento do *observador* – noção tomada de Jonathan Crary – sugere-se uma mutualidade constitutiva das visualidades humana e da máquina, inclusive por processos de retroalimentação aos modos de ver contemporâneos de formas de codificação do visível elaboradas pela Visão Computacional. Sugere-se, por tal caminho, o valor paradigmático deste dispositivo para a compreensão das dinâmicas contemporâneas de produção e circulação do visível, bem como de constituição de novas concepções e práticas observacionais, mesmo aquelas apenas indiretamente relacionadas com seu funcionamento.

**Palavras-chave:** visão computacional; dispositivo; imagem; observador.



## Abstract

This work approaches, through an image studies perspective, the technological domain of Computer Vision, typically understood by the broad spectrum of algorithmic operations by which a machine is programmed for the automatic interpretation of visual information. It aims at describing the phenomenon of emergence and dissemination of its derived machines, but not from a purely technical point of view, paying particular attention to the modes by which they agency, through their operation, the *image* and the *observer*. Supported by the theoretic framework surrounding foucaultian concept of the *dispositif*, the research understands the studied domain by its concrete instantiations, in forms of regulation and government of the act of *seeing* and of the *visible*, therefore deeply traversed by formations of knowledge, institutional demands, observational practices and technical, esthetical and scientific discourses – among other elements. In order to make such articulations visible, going from a technical constitution to esthetic, social and political agencies, this work's theoretic-analytical course is elaborated through the reunion of multiple objects, selected from technical, scientific and artistic domains, according to their potential for problematizing aspects of this *dispositif's* *agencement*. This effort is animated by the methodological gesture of the *composition*, proposed by Bruno Latour, by which it is sought to build, through connections and approximations, the common ground of the approached phenomenon from the multiplicity of its heterogeneous manifestations. Concerning the modes of *agencement* of the image and the visible, an alternating movement is described between operations of singling, locating and identifying, in one hand; and generalizing, categorizing and idealizing, on the other. Concerning the *agencement* of the *observer* – notion taken from Jonathan Crary – it is suggested that the visualities of the human and of the machine are mutually constituted through its operation, including processes of taking back to the contemporary modes of seeing the codification of the visible developed within Computer Vision. Through this course it is suggested the paradigmatic value of this *dispositif* for the comprehension of contemporary dynamics of the production and circulation of the visible, as well as the constitution of new conceptions and practices of observation, even those only indirectly related to its functioning.

**Keywords:** computer vision; *dispositif*; image; observer.



## Lista de ilustrações

Figura 1 – Fotograma de <i>O homem com a câmera</i> (1929), de Dziga Vertov. Fonte: <a href="http://www.moma.org/images/dynamic_content/ipad/62227.jpg">http://www.moma.org/images/dynamic_content/ipad/62227.jpg</a> . ....	19
Figura 2 – Fotograma de <i>Um cão andaluz</i> (1929), de Luis Buñuel e Salvador Dalí. Fonte: UN CHIEN, 2014. ....	23
Figura 3 – Montagem comparativa: Leitor de código de barras e fotograma de <i>Robocop</i> (1987), de Paul Verhoeven. Fontes: <a href="http://www.modber.ru/images/catalog/medium/c7c3365c902a7c255f4a8905eb71cb38.jpg">http://www.modber.ru/images/catalog/medium/c7c3365c902a7c255f4a8905eb71cb38.jpg</a> e <a href="http://i.ytimg.com/vi/5q10EXI5hgk/maxresdefault.jpg">http://i.ytimg.com/vi/5q10EXI5hgk/maxresdefault.jpg</a> . ....	24
Figura 4 – “Identity intelligence: image is everything”: documento da Agência de Segurança Nacional dos EUA vazado por Edward Snowden e divulgado em reportagem de James Risen e Laura Poitras (2014). Fonte: RISEN e POITRAS, 2014. ....	27
Figura 5 – Exemplo de CAPTCHA. Fonte: <a href="http://static.guim.co.uk/sys-images/Guardian/Pix/pictures/2008/08/27/captcha460.jpg">http://static.guim.co.uk/sys-images/Guardian/Pix/pictures/2008/08/27/captcha460.jpg</a> . ....	28
Figura 6 – Ilustração do processo de extração de silhuetas realizado pelo programa de Visão Computacional de <i>Videoplace</i> (1969). Fonte: KRUEGER, 2003, p. 385. ....	39
Figura 7 – Tela que apresenta os dados resultantes do algoritmo de reconhecimento de rostos do serviço <i>online</i> Rekognition. Fonte: impressão da tela a partir dos resultados da análise de fotografia de nossa autoria no site <a href="http://rekognition.com">http://rekognition.com</a> . ....	45
Figura 8 – “Galáxias formando-se ao longo dos fios, como gotículas ao longo das costas de uma teia de aranha”, de Tomas Saraceno, em exposição na Bienal de Veneza de 2009. Foto: Eduardo de Jesus. Fonte: LATOUR, 2012. ....	46
Figura 9 – Ilustração do relatório de Roberts (1963) explicitando os passos do processo algorítmico desenvolvido. Fonte: ROBERTS, 1963. ....	53
Figura 10 – Módulo interativo de <i>Videoplace</i> (1969), de Myron Krueger. Fonte: LEVIN, G., 2006a. ....	54
Figura 11 – <i>Text rain</i> (1999), de Camille Utterback e Romy Achituv. Fonte: <a href="http://camlleutterback.com/projects/text-rain/">http://camlleutterback.com/projects/text-rain/</a> . ....	57
Figura 12 – <i>Hand from above</i> (2009), de Chris O’Shea. Fonte: <a href="http://www.chrisoshea.org/hand-from-above">http://www.chrisoshea.org/hand-from-above</a> . ....	58

Figura 13 – <i>The giver of names</i> , de David Rokeby, em montagem de 2008. Fonte: <a href="http://www.davidrokeby.com/gon.html">http://www.davidrokeby.com/gon.html</a> .....	65
Figura 14 – Ilustração da sequência de algumas respostas visuais e textuais do programa de <i>The giver of names</i> , de David Rokeby. Fonte: <a href="http://www.davidrokeby.com/gon.html">http://www.davidrokeby.com/gon.html</a> .....	66
Figura 15 – Sinal perdido das “câmeras suicidas” em <i>frame</i> de um dos canais da videoinstalação <i>Eye/Machine</i> (2001), de Harun Farocki. Fonte: EYE/MACHINE, 2001.....	79
Figura 16 – Imagens de câmeras acopladas a mísseis em <i>frame</i> de um dos canais da videoinstalação <i>Eye/Machine</i> (2001), de Harun Farocki. Fonte: EYE/MACHINE, 2001.....	89
Figura 17 – <i>Frame</i> da versão monocanal consultada, da instalação <i>Eye/Machine I</i> (2001), de Harun Farocki, exibindo, simultaneamente, os dois canais da obra. Fonte: EYE/MACHINE I, 2001....	90
Figura 18 – <i>Frame</i> da versão monocanal consultada, da instalação <i>Eye/Machine I</i> (2001), de Harun Farocki, exibindo, simultaneamente, os dois canais da obra. Fonte: EYE/MACHINE I, 2001....	97
Figura 19 – <i>Frame</i> da versão monocanal consultada, da instalação <i>Eye/Machine III</i> (2003), de Harun Farocki, exibindo, simultaneamente, os dois canais da obra. Fonte: EYE/MACHINE III, 2003.	98
Figura 20 – Plano detalhe do <i>olho</i> de HAL 9000 em <i>2001: uma odisseia no espaço</i> (1968), de Stanley Kubrick. Fonte: 2001, 2014. ....	107
Figura 21 – Ponto de vista de HAL em <i>2001: uma odisseia no espaço</i> (1968). Fonte: 2001, 2014. ....	107
Figura 22 – Ponto de vista de HAL em <i>2001: uma odisseia no espaço</i> (1968). Fonte: 2001, 2014. ....	107
Figura 23 – <i>Frame</i> de <i>O exterminador do futuro</i> (1984), de James Cameron. Texto e elementos gráficos sobrepostos sugerem o reconhecimento da escrita manual pelo robô androide. Fonte: TERMINATOR, 2014. ....	108
Figura 24 – <i>Frame</i> de <i>Robocop</i> (1987), de Paul Verhoeven. Elementos gráficos sobrepostos indicam a operação de alinhamento da mira ao alvo realizada pelo policial ciborgue. Fonte: ROBOCOP, 2014.....	108
Figura 25 – Modelo visual de um gato gerado pelo programa de reconhecimento da <i>Google</i> , disponível em tamanho menor na matéria jornalística (MARKOFF, 2012), mas retirada de <a href="http://www.embedded-vision.com/sites/default/files/news/CatDetection_resized.jpg?1349732566">http://www.embedded-vision.com/sites/default/files/news/CatDetection_resized.jpg?1349732566</a> . ...	117
Figura 26 – Visualização dos 48 estímulos do grupo teste com melhores respostas do programa da <i>Google</i> para reconhecimento de rostos, que se valeu de imagens encontradas no <i>YouTube</i> como base de treinamento. Fonte: LE <i>et al.</i> , 2012. ....	118



Figura 27 – Visualização do estímulo ótimo do programa da <i>Google</i> para o reconhecimento de rostos. Fonte: LE <i>et al.</i> , 2012.....	118
Figura 28 – Retratos compostos por Francis Galton exibindo características prevaletentes em condenados por furto sem uso de violência. Fonte: <a href="http://galton.org/">http://galton.org/</a> .....	121
Figura 29 – Retrato composto de um “criminoso médio”, elaborado ao modo de Francis Galton. Fonte: ELLIS, Havelock. <i>The criminal</i> . London: Walter Scott, 1890. <i>apud</i> SEKULA, 1986, p. 42. ....	122
Figura 30 – <i>Frame</i> de <i>Blade Runner</i> , de Ridley Scott: o teste <i>Voight-Kampff</i> . Fonte: BLADE, 2014. ....	127
Figura 31 – Exemplos de reconstrução dinâmica da experiência visual na pesquisa realizada por cientistas da Universidade da Califórnia em Berkeley. A linha superior – “Presented movies” – apresenta <i>frames</i> do vídeos apresentados como estímulo; a linha inferior – “Reconstructed movies (AHP)” – , <i>frames</i> dos estímulos reconstruídos pelo programa; e as linhas intermediárias, indicam as imagens da base de treinamento que seriam as melhores estimativas do programa, dentre as imagens já conhecidas. Fonte: NISHIMOTO <i>et al.</i> , 2011, p. 1644.....	133
Figura 32 – <i>Frame</i> do vídeo de demonstração da pesquisa dos neurocientistas de Berkeley (NISHIMOTO <i>et al.</i> , 2011) exemplificando a reconstrução realizada. A imagem à esquerda apresenta o estímulo original. À direita é apresentado o estímulo reconstruído a partir da atividade cerebral. Fonte: <a href="https://www.youtube.com/watch?v=nsjDnYxJ0bo">https://www.youtube.com/watch?v=nsjDnYxJ0bo</a> . ....	133
Figura 33 – Optograma do Dr. Vernois, revelado a partir da retina de uma mulher assassinada em 14 de junho de 1868. Fonte: DUBOIS, 2012, p. 230. ....	134
Figura 34 – <i>Frame</i> de <i>Blade Runner</i> , de Ridley Scott. Fonte: BLADE, 2014. ....	149
Figura 35 – Captura de tela do aplicativo <i>Bablrr</i> (2012), de Thiago Hersan.....	150
Figura 36 – <i>CVDazzle</i> (2010), de Adam Harvey. <i>Look 4</i> . Fonte: <a href="http://cvdazzle.com/">http://cvdazzle.com/</a> . ....	153
Figura 37 – <i>Frame</i> de <i>How not to be seen: a fucking didactic educational .mov file</i> (2013), de Hito Steyerl. Fonte: CONNOR, 2013. ....	154
Figura 38 – <i>reCAPTCHA</i> em sua versão clássica. Fonte: SHET, 2014.....	157
Figura 39 – <i>reCAPTCHA</i> com imagem do projeto <i>Street View</i> . Fonte: SHET, 2014. ....	157
Figura 40 – <i>reCAPTCHA</i> com tarefa de classificação de imagens. Fonte: SHET, 2014.....	158
Figura 41 – Excerto da sequência de <i>I’m Google</i> (201-), de Dina Kelberman. Fonte: KELBERMAN, 201-. ....	163
Figura 42 – Excerto da sequência de <i>I’m Google</i> (201-), de Dina Kelberman. Fonte: KELBERMAN, 201-. ....	164

Figura 43 – Excerto da sequência de <i>I'm Google</i> (201-), de Dina Kelberman. Fonte: KELBERMAN, 201-.	164
Figura 44 – Excerto dos resultados de busca realizada na <i>Google</i> a partir de imagem da sequência de <i>I'm Google</i> . Fonte: <a href="http://www.google.com">http://www.google.com</a> .	167
Figura 45 – Seção da visualização da rede composta a partir da base de imagens utilizadas no desafio ILSVRC 2012. Fonte: KARPATY, [2012?].	168
Figura 46 – Tela do <i>site</i> do projeto <i>Google Glass</i> , fora do ar desde janeiro de 2015. Fonte: <a href="http://www.google.com/glass/start/">http://www.google.com/glass/start/</a> .	175
Figura 47 – Imagem ilustrativa da função de navegação do <i>Google Glass</i> , retirada do <i>site</i> do projeto, fora do ar desde janeiro de 2015. Fonte: <a href="http://www.google.com/glass/start/">http://www.google.com/glass/start/</a> .	175
Figura 48 – Imagem de divulgação do projeto <i>HoloLens</i> , da Microsoft, apresenta uma representação daquilo que veria uma usuária do produto. Fonte: HEMPEL, 2015.	176

## Sumário

Introdução: o olho e a câmera .....	21
1 Máquinas que veem.....	41
1.1 Máquinas visuais e de visão .....	41
1.2 Visão–algoritmo .....	49
1.3 Dispositivo, máquina, paradigma .....	67
2 O que veem as máquinas .....	81
2.1 Câmeras suicidas, imagens evanescentes .....	81
2.2 Imagem–instrumento .....	88
2.3 <i>A visão subjetiva</i> das máquinas .....	104
3 Nós, que vemos .....	129
3.1 Quem vê? .....	129
3.2 Nós e elas .....	145
3.3 Nós e a rede.....	162
Considerações finais .....	183
Referências .....	193
Anexo I – Listagem ilustrativa de operações características da Visão Computacional.....	207
Anexo II – Entrevista com a artista Dina Kelberman .....	211





Figura 1 – Fotograma de *O homem com a câmera* (1929), de Dziga Vertov. Fonte: [http://www.moma.org/images/dynamic\\_content/ipad/62227.jpg](http://www.moma.org/images/dynamic_content/ipad/62227.jpg).



## Introdução: o olho e a câmera

A sobreposição de objetiva e olho (Figura 1), de *O homem com a câmera* (1929), de Dziga Vertov, talvez sirva como um prelúdio à discussão que se segue. Tradução visual da metáfora do *cine-olho* – *leitmotiv* da proposição de um cinema revolucionário pelo cineasta soviético – a imagem do olho arregalado que se abre através das lentes da câmera evidencia o acoplamento almejado para a transposição dos limites da visão humana. Alcançando pelo cinema, sugere o cineasta (VERTOV, 1983b, p. 261-262), “o que o olho não vê”, revelando a verdade em uma pedagogia comunista da visão. Para além dos aspectos mais particulares desta formulação de Vertov, é mais relevante para esta pesquisa o modo como esta imagem representa não um encontro historicamente isolado entre *câmera e olho*, mas, antes, uma analogia recorrente. Como indica William C. Wees (2003), embora tal correlação tenha nesse cineasta seu principal expoente, ela possui relevância mais geral para o cinema de vanguarda. Nesse contexto, ressurgue com frequência, seja explicitamente, como neste caso, ou indiretamente, como na dilaceração do globo ocular que antecipa a narrativa surrealista de *Um cão andaluz* (1929), de Luis Buñuel e Salvador Dalí (Figura 2).

Na esteira do projeto estético modernista, câmera e olho surgem como componentes fundamentais por cuja articulação o cinema seria capaz de produzir “novas formas de ver”. Não deixa de ser este, em larga medida, o discurso encampado por Vertov, que em um de seus manifestos chega, inclusive, a assumir a voz da máquina: “Eu sou o cine-olho. Eu sou o olho mecânico. Eu, máquina, vos mostro o mundo como só eu posso vê-lo” (VERTOV, 1983a, p. 256). Eis que o olho mecânico do cinema liberta-nos da cegueira: o olho arregalado que vemos então se abrir, através da objetiva, não seria apenas o da câmera, personificada, mas também o dos nascentes espectadores, assombrados, diante do mundo que lhes revelava o cinema, à sua frente. Este momento, que em *O homem com a câmera* vemos se realizar, parece ser uma constante da modernidade (ao menos almejada), que se estende até nós – uma constante pela variação, cabe destacar – em que o campo visual é seguidamente reordenado, reconfigurado, ampliado, entre outros fatores, através de acoplamentos atualizados de máquinas e olhos.

Mais além do cinema, a metáfora do *cine-olho* – ou, de modo mais abrangente, da *câmera-olho*, seguindo Wees (2003) – inscreve-se na própria câmera. Trata-se, afinal, de artefato que compreende, desde o início, uma tentativa de reprodução técnica da visão, compartilhando com o olho humano inclusive os nomes de alguns de seus componentes, como a *íris* que indica, em ambos os casos, o “mecanismo” que regula a abertura para a entrada da luz. Jonathan Crary (1992), abordando desenvolvimentos anteriores à câmera fotográfica, como a câmara escura e certos dispositivos ópticos do começo do século XIX – como o estereoscópio e o praxinoscópio – indica a força de tais metáforas na medida em que informariam não apenas a compreensão dos processos da percepção visual como também, de modo mais abrangente, concepções filosóficas de ramos como a fenomenologia e a

epistemologia, em que a visão é tomada (em algumas formulações) como um sentido privilegiado para conhecer ou atestar a existência das coisas do mundo. O autor recupera, por exemplo, o papel desempenhado pela câmara escura, como metáfora filosófica, nos desenvolvimentos de Descartes acerca da possibilidade de uma percepção objetiva do mundo ou da própria separação entre sujeito e objeto. Segundo argumenta Crary (1992, p. 25-66), seria, entre outras razões, pelo próprio modo de operação deste artefato, com a formação de uma imagem externa ao corpo e a configuração espacial que depende do isolamento de seu habitante ao observar o mundo, que subsidiariam uma tal compreensão. Vê-se, assim, que mais do que tratarmos dos aspectos técnicos ou oftalmológicos e cognitivos envolvidos em uma tal correlação entre a câmera e o olho – como guia, até certo ponto, o cotejamento feito por Wees (2003) – faz-se importante reconhecer que, ainda que metafórica, há a vinculação não somente a uma compreensão particular da percepção visual, mas também (e, talvez, principalmente) a uma determinada *visão de mundo*. Pois, enquanto expressão e instrumento<sup>1</sup> de transformação das compreensões e dos modos de *ver* em determinado contexto, dispositivos visuais não participam de forma neutra destes acoplamentos concretos ou assimilações metafóricas.

\*

Assistimos, hoje, à emergência de novos aspectos e configurações desta metáfora. Experimentando graus mais elevados de autonomização dos dispositivos de registro e processamento de imagens, observamos que as operações desempenhadas por diversas das máquinas contemporâneas direcionadas ao propósito de reprodução técnica da visão já não se restringem à produção de imagens, mas engajam-se também em sua *interpretação*<sup>2</sup>. Programas computacionais aplicados às imagens produzida pelas câmeras (muitas vezes inclusive integrados a elas) dedicam-se à tarefa de reconhecimento de formas, de recomposição do espaço tridimensional retratado ou mesmo de atribuição de conceitos à cena. Já há algum tempo lidamos cotidianamente com tais máquinas, das quais os leitores de códigos de barra talvez tenham sido os primeiros exemplos. O olho vermelho do escâner dos supermercados – que talvez não por acaso se assemelha tanto ao do *Robocop* (Figura 3)<sup>3</sup> – e a versão atualizada dos códigos de barra nos chamados *códigos QR*<sup>4</sup>, seguem como resquícios desta forma “primitiva” da visão das máquinas. Contudo, em seu estágio atual e em seus exemplos mais prementes, elas já não dependem de tais grafismos desenhados específica e exclusivamente para a leitura pela máquina – sensível a formas geométricas em alto contraste. Em operações como o reconhecimento de rostos – implementado em câmeras digitais, em redes sociais e em sistemas de videovigilância, dentre outros contextos – máquinas de visão contemporâneas dedicam-

---

1 Recorro à formulação de Maria Cristina Franco Ferraz (2013, p. 168) acerca do papel desempenhado pelos dispositivos nos processos de transformação social.

2 Evidentemente, não sugerimos que máquinas *interpretariam* imagens sem algum tensionamento. Trata-se, contudo, de uma abordagem recorrente que, como ficará mais claro ao longo da dissertação, incorporamos criticamente em nosso texto. Pontuaremos as implicações deste tipo de assimilação metafórica.

3 Curiosamente, o *remake* de José Padilha, de 2014, ainda parece reforçar esta associação, com a ênfase na linha vermelha delineando o olho robótico do personagem.

4 No inglês chamados de *QR Codes* são versões atualizadas de códigos de barras, compostos por padrões mais complexos, bidimensionais. Cf. [http://pt.wikipedia.org/wiki/Código\\_QR](http://pt.wikipedia.org/wiki/Código_QR).





Figura 2 – Fotograma de *Um cão andaluz* (1929), de Luis Buñuel e Salvador Dalí. Fonte: UN CHIEN, 2014.

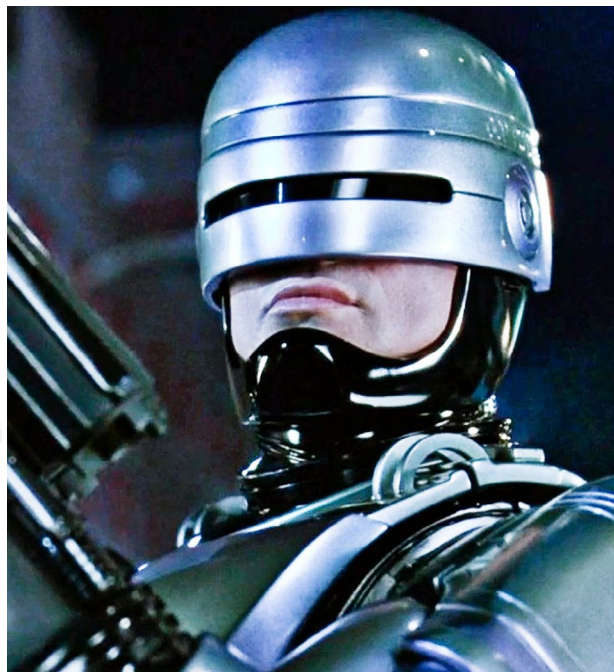


Figura 3 – Montagem comparativa: Leitor de código de barras e fotograma de *Robocop* (1987), de Paul Verhoeven. Fontes: <http://www.modber.ru/images/catalog/medium/c7c3365c902a7c255f4a8905eb71cb38.jpg> e <http://i.ytimg.com/vi/5q10EX15hgk/maxresdefault.jpg>.

se frequentemente à interpretação de imagens de base fotográfica<sup>5</sup> (estáticas ou em movimento). Assistimos à proliferação de câmeras miniaturizadas e integradas aos mais diversos artefatos, como celulares, computadores, televisores, óculos<sup>6</sup>, capacetes<sup>7</sup> e aeronaves não tripuladas (*drones*). Somam-se às próprias câmeras o crescente recurso à computação embarcada nestes dispositivos (com microprocessadores integrados) ou sua conexão constante com a internet e com serviços de *computação em nuvem*<sup>8</sup> – permitindo o processamento instantâneo das imagens capturadas, seja local ou remotamente. Já além das máquinas produtoras de imagens, chegamos ao domínio do que, certa vez, Paul Virilio (1994) denominou *máquinas de visão*. Potencialmente – e provavelmente – toda e qualquer imagem hoje produzida ou digitalizada é passível de ser interpretada por seus programas.

Exemplos são cada vez mais frequentes. A empresa responsável pela rede social *Facebook* vale-se, desde poucos anos, de algoritmos sofisticados de reconhecimento facial para a geração automatizada de metadados sobre as imagens postadas por seus usuários. Também, a *Google* incorporou às suas ferramentas de pesquisa a opção de valer-se de uma imagem (ao invés de um texto) como chave da requisição, além de oferecer, por meio do aplicativo para dispositivos móveis *Google Goggles*, diferentes modalidades de extração de dados de uma fotografia, como o reconhecimento óptico de caracteres, leitura de códigos de barra e códigos QR, reconhecimento de logomarcas, busca por imagens visualmente semelhantes, ou mesmo a identificação (literalmente) de um livro pela sua capa. No âmbito da vigilância, tais algoritmos, presentes em geral de forma oculta, atuam sem que tenhamos real dimensão da amplitude ou profundidade de sua aplicação. “Imagem é tudo”<sup>9</sup> (Figura 4), diz o título de um dos documentos vazados recentemente por Edward Snowden, em que tomamos conhecimento de que a Agência de Segurança Nacional dos Estados Unidos vem se valendo de algoritmos de reconhecimento de rostos na varredura de imagens postadas na *web*, anexadas a *emails* e em videoconferências (RISEN e POITRAS, 2014<sup>10</sup>). Estamos tratando, portanto, de um campo extremamente estratégico e cada vez mais presente em nosso cotidiano. Suplantando, em alguma medida, as limitações da atenção ou da resistência humana à amplitude da tarefa requisitada diante da produção massiva de imagens (que demandam organização, classificação e indexação constantes), as máquinas de visão contemporâneas parecem visar à visibilidade irrestrita.

---

5 Entendemos por imagens de base fotográfica não apenas fotografias, propriamente ditas, mas quaisquer representações figurativas produzidas pelo recurso a câmeras (tendo como base o modelo fotográfico) sejam elas estáticas ou em movimento – como cinema e vídeo. Excluem-se, assim, por exemplo, imagens produzidas a partir de diferentes formas de inscrição manual (desenho, gravura, pintura) ou de síntese numérica (computação gráfica).

6 Como exemplo mais evidente, temos o *Google Glass* (<http://www.google.com/glass/start/>).

7 Câmeras voltadas para esportes de aventura como a *GoPro* possuem acessórios para acoplamento a capacetes e são frequentemente utilizadas por ciclistas como forma de registro de eventuais acidentes e foram recentemente utilizadas por policiais militares em Minas Gerais como recurso de vigilância durante a repressão de manifestações populares.

8 Chama-se computação em nuvem o conceito de arquitetura de sistemas de informação que se baseia no recurso às capacidades de armazenamento e processamento de computadores remotos, através da internet, como forma de potencializar as capacidades de dispositivos locais, geralmente mais simples e menos potentes.

9 No original: “*Image is everything*”. Tradução nossa.

10 A reportagem, publicada em *The New York Times*, trouxe uma reprodução parcial dos *slides* de uma apresentação confidencial da Agência de Segurança Nacional, intitulada “*Identification intelligence: image is everything*”.

Numa espécie de contraponto a tal onividência, contudo, um dos traços mais evidentes dos agentes computacionais de visão talvez sejam justamente os dispositivos desenvolvidos visando seus limites: os *CAPTCHA*. Diante de uma figura ruidosa de letras e números distorcidos, somos solicitados a identificar os caracteres para seguir adiante na navegação (Figura 5). Enquanto um teste cujo objetivo é distinguir, entre aqueles que acessam determinada página na *web*, os *humanos* dos *bots* (literalmente, robôs), os *CAPTCHA* jogam com uma espécie de fronteira da visibilidade computacional: visíveis para humanos, mas não para máquinas. A necessidade atual de distinguirmos de agentes computacionais através de um tal dispositivo não apenas revela a expectativa de certo nível de desempenho destes programas e de sua presença e atuação ubíqua na *web* como também parece expressar uma ambiguidade do desenvolvimento de tal tecnologia, dado que em determinadas circunstâncias julgamos necessário recorrer à possibilidade de ainda demarcar a diferença entre visão humana e visão da máquina. A falha dos dispositivos de visão contemporâneos, portanto, longe de um não dito desta tecnologia, integra a dinâmica de seu desenvolvimento, apontando para a complexidade do trânsito de fronteiras entre o humano e a máquina que ela parece inicialmente sugerir<sup>11</sup>.

\*

O desenvolvimento contemporâneo desta tecnologia talvez se configure como a mais recente manifestação da metáfora câmera-olho, aprofundada pelo recurso à tecnologia digital e, ainda, à correspondência corrente entre *computação* e *cognição*. Dentre as denominações comuns deste conjunto de objetos, encontramos *câmeras inteligentes*, *visão robótica* ou *percepção da máquina*. Optamos, em vista da analogia de base que identificamos, por tratar este conjunto heterogêneo por *Visão Computacional*, como se denomina a disciplina das Ciências da Computação dedicada ao estudo e ao desenvolvimento de tais processos, e reunindo, assim, toda a ampla classe de algoritmos e aplicações derivadas. Trata-se, também, de denominação que já traz inscrita a analogia entre humano e máquina ao articular a *visão*, cujo referencial é, evidentemente, o humano ou o animal, ao *computacional*, cujo referencial contemporâneo é a máquina, o computador.

Pelos exemplos já pontuados, nota-se tanto a relevância contemporânea dessa tecnologia quanto o caráter problemático de sua presença. De um lado, recebemos a Visão Computacional com certo tom de celebração e simpatia como, por exemplo, no caso das câmeras fotográficas que têm, integrados, algoritmos de reconhecimento de sorrisos, efetuando disparos automáticos quando rostos sorridentes são detectados<sup>12</sup>. Destaca-se, também, sua aplicação como interface sensível ao

---

11 Shoshana Amielle Magnet (2011), em outro registro, abordando o tema da biometria, sugere que a falha, longe de uma aberração ou exceção, constitui uma dimensão endêmica da tecnologia que ela aborda (p. 6). A autora sugere ainda, em um plano geral, que a falha tecnológica seria algo como uma dimensão recalçada destes desenvolvimentos, tão produto de matrizes culturais e ideologias quanto os seus sucessos (p. 3).

12 Hoje presente em aplicativos de celular, a tecnologia de detecção de sorrisos em câmeras fotográficas, foi introduzida no mercado pela Sony, em 2007, sob o nome *Smile Shutter*. Cf.: <https://www.sony-mea.com/article/237954/section/events>.

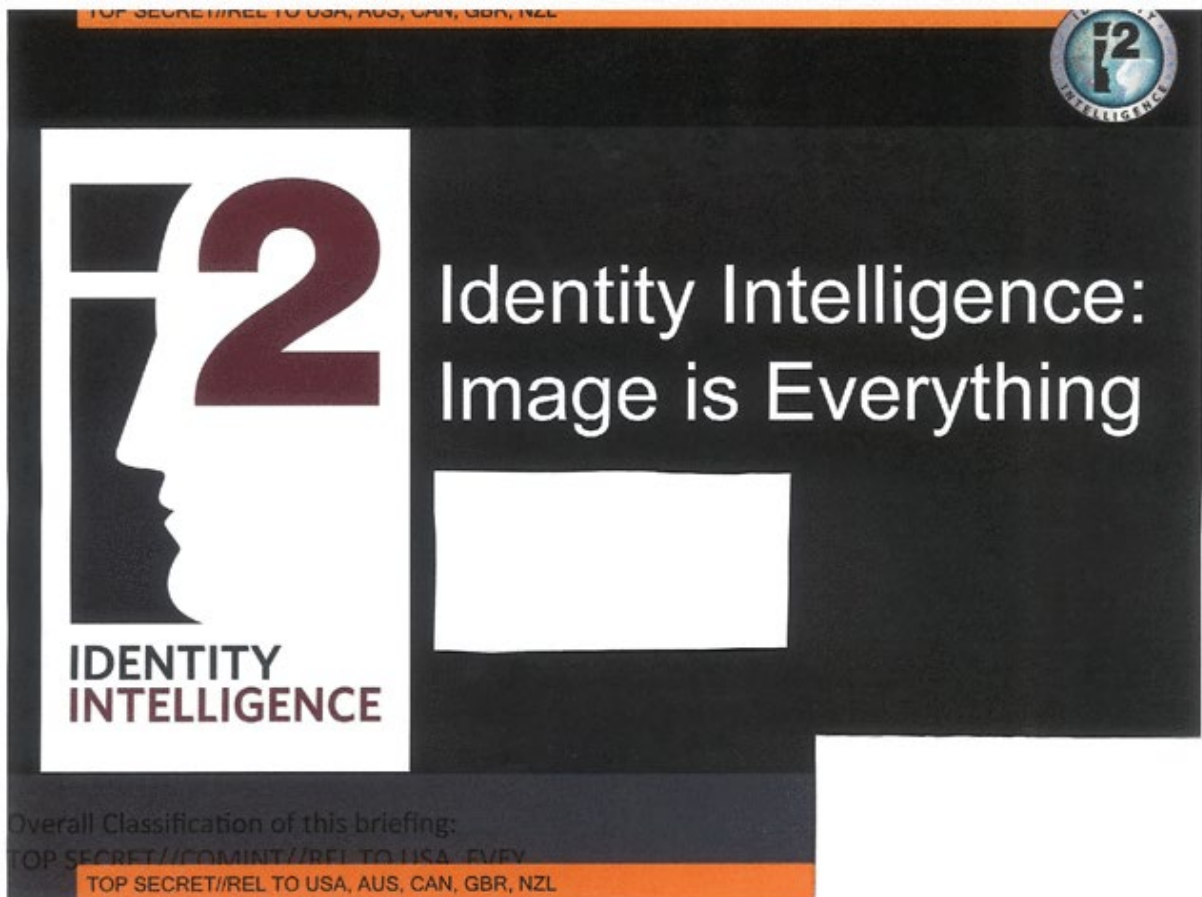


Figura 4 – “Identity intelligence: image is everything”: documento da Agência de Segurança Nacional dos EUA vazado por Edward Snowden e divulgado em reportagem de James Risen e Laura Poitras (2014). Fonte: RISEN e POITRAS, 2014.



Figura 5 – Exemplo de CAPTCHA. Fonte: <http://static.guim.co.uk/sys-images/Guardian/Pix/pictures/2008/08/27/captcha460.jpg>.

gesto – em modalidades de interação geralmente indicadas como *intuitivas* ou, ainda, *amigáveis* – como no sensor *Kinect*, da Microsoft<sup>13</sup>, que é utilizado no aparelho de jogos eletrônicos da empresa e em inúmeras instalações interativas, obras de artemídia, dispositivos museais e estandes publicitários. De outro lado, os desenvolvimentos neste campo despertam preocupações com os riscos da visibilidade irrestrita e da biometria, com os temores relacionados à vigilância e ao controle. Provocam, assim, abordagens mais cautelosas ou francamente críticas e combativas com relação ao papel assumido por tais desenvolvimentos em sua inserção nos coletivos sociotécnicos contemporâneos – encampadas principalmente por artistas, acadêmicos e ativistas.

A discussão destes posicionamentos diversos permearão inevitavelmente o percurso desta dissertação e inclusive colocam-se como motivações iniciais da pesquisa, mas não é a eles que direcionaremos nossa análise. Situamo-nos, particularmente, em uma abordagem do *dispositivo* da Visão Computacional que busca compreender de perto os aspectos de sua operação que atravessariam diferentes configurações e contextos de aplicação e que, articulados à dimensão do *visual*, provocariam deslocamentos substanciais em nossa relação com a *imagem* e com o próprio ato de *ver*. Longe de recusarmos os aspectos políticos da questão, como brevemente pontuamos, buscamos compreendê-los na forma como seriam em alguma medida constituídos na própria configuração dos objetos analisados e pelos demais elementos que os sustentam, além é claro, enquanto desdobramentos importantes das reconfigurações de que participam. Dizemos, assim, que não concebemos o debate político implicado por tal tecnologia como um elemento exógeno que pudéssemos analisar em separado ou mesmo como uma consequência posterior aos desenvolvimentos que estudamos – como se a ciência e a tecnologia fossem *objeto* do debate político quando, na verdade, são já expressões e instrumentos deste debate. Compreendendo, portanto, uma articulação mais intrínseca entre ciência, tecnologia, estética e política, enfocaremos o modo como o dispositivo que abordamos já traz, como parte intrínseca de sua constituição, tensionamentos e respostas a questões destas distintas ordens, buscando, assim, retrazar as conexões entre tais domínios. Em suma, se abordamos um desenvolvimento tecnológico que visa automatizar processos de tradução do visível em informação, não há nem mesmo como iniciarmos sua descrição sem observarmos suas aplicações presumidas – que não apenas justificam o investimento realizado em sua pesquisa quanto informam seu funcionamento. Por mais que, no contexto da prática científica e tecnológica o problema venha a ser fragmentado e reduzido ao desenvolvimento de uma operação lógica e à racionalização de um determinado processo, devemos ainda nos atentar ao quadro geral, até porque, mesmo nestas subdivisões analíticas do problema, provavelmente seguirão presentes rastros do problema mais amplo que ajudam a compor e das transformações que ajudam a realizar.

---

13 O sensor *Kinect*, com primeira versão lançada em 2010, consiste em um dispositivo dedicado de Visão Computacional, composto, entre outros elementos, por duas câmeras, sendo uma de captação em cores e outra em preto e branco, e sensível à luz infravermelha. Esta última capta no ambiente um padrão luminoso projetado por um *laser* infravermelho integrado ao dispositivo, cuja distorção ao incidir sobre o espaço é utilizada pelo sensor para inferir as distâncias de cada ponto visível na imagem e assim recompor o espaço representado em uma simulação tridimensional.

Ao nos atentarmos a tais questões, reivindicamos uma compreensão da própria visão – mais além das próprias técnicas – como uma prática construída e historicamente situada. Este é o sentido que atribuímos à noção de *visualidade*. Hal Foster (1988, p. ix) nos oferece uma boa formulação da questão ao sugerir que *visão* e *visualidade* seriam dimensões complementares da constituição de um campo visual, no qual a *visão* pareceria dizer respeito a um componente *natural* – a uma fundamentação fisiológica e cognitiva da visão (em alguma medida compreendida como *universal*) – enquanto a *visualidade* apontaria para uma dimensão social e histórica. Seria pelo jogo entre estas dimensões que se constituiriam *regimes escópicos* de determinados períodos:

Com sua própria retórica e representações, cada regime escópico busca desfazer essas diferenças: fazer de suas muitas visualidades sociais uma visão essencial, ou ordená-las numa hierarquia natural da visão. É importante, então, deslocar estas superposições para fora de foco, perturbar o arranjo dado de fatos visuais (talvez seja, absolutamente, a única forma de vê-los)[...] (FOSTER, 1988, p. ix)<sup>14</sup>.

Portanto, mais do que reforçar uma distinção essencial entre *visão* e *visualidade*, Foster sugere que, para reconhecer esta última, precisamos desnaturalizar a primeira: abalar determinada compreensão da visão como dada ou transparente, independente de sua inscrição histórica. Trata-se de abordagem que analistas contemporâneos como o próprio Hal Foster, Jonathan Crary (2014; 2001; 1992) e Martin Jay (1988) herdaram de Walter Benjamin, entre outras razões, por sua defesa das transformações da percepção como o real sentido do estudo da estética (BENJAMIN, 1994, p. 194)<sup>15</sup>.

No caso estudado, as dimensões da visão e da visualidade se articulam pela via da naturalização de uma compreensão particular da metáfora câmera-olho prevista já na denominação deste campo de pesquisa. Afinal, ao dizermos de uma *Visão Computacional* parecemos sugerir que a visão seria uma competência destacável do corpo e reprodutível na máquina, o que já traduz uma compreensão cuja construção precisaria ser explicitada. Não raramente, pesquisadores da área irão omitir o adjetivo, dizendo apenas: “Visão”<sup>16</sup>. Mais do que uma abreviação tácita, própria de um jargão, tal formulação se insere em uma concepção bem estabelecida que compreende a possibilidade de correlação direta entre visão humana e visão da máquina que, embora não seja um ponto de pleno acordo, parece ser uma crença remanescente e frequentemente recuperada – como se observa na tomada do desempenho humano como índice de referência para a aferição do desempenho da máquina<sup>17</sup>.

14 No original: “*With its own rhetoric and representations, each scopic regime seeks to close out these differences: to make of its many social visualities one essential vision, or to order them in a natural hierarchy of sight. It is important, then, to slip these superpositions out of focus, to disturb the given array of visual facts (it may be the only way to see them at all)...*”. Tradução nossa.

15 Este tratamento, que parece atravessar muitos de seus ensaios, aparece nesta formulação na 1ª versão de seu conhecido ensaio “A obra de arte na era de sua reprodutibilidade técnica”, tendo sido excluída de versões posteriores (cf. BENJAMIN, 2014).

16 Forsyth e Ponce (2012, p. xvii), na introdução de um livro técnico da área, chegam a se desculpar àqueles que estudam a visão humana ou animal logo após introduzir tal denominação genérica a seu campo de estudos.

17 Por exemplo, vemos pesquisa recente empreendida pelo *Facebook* na área de reconhecimento de rostos este tipo de comparação – o algoritmo conseguiu um índice de acerto de 97,25% contra 97,53% do índice humano. Cf. TAIGMAN *et al.*, 2013.



David Marr (2010), um dos pesquisadores pioneiros da Visão Computacional, introduz seu livro *Vision*, publicado originalmente em 1982, sugerindo a compreensão da visão (humana, inclusive) enquanto tarefa de processamento e representação da informação. Em concepção herdada dos campos das Ciências Cognitivas e da Inteligência Artificial, dos quais a Visão Computacional é frequentemente compreendida como um subdomínio, a informação é tomada, assim, como ponto de passagem entre cérebro e máquina. A analogia, porém, não é simplista: como o autor (MARR, 2010, p. 5) chega a mencionar, ao se comparar cérebro e computador, toma-se como ponto de passagem a tarefa de processamento e representação da informação, mas não se diz que um computador qualquer – como o que utilizamos na escrita deste texto – seria diretamente equiparável ao cérebro humano. Enquanto nosso órgão estaria mais apto ao desenvolvimento de atividades específicas por suas partes e regiões, os computadores, de um modo geral, foram projetados para atividades padronizadas e genéricas, e por isto não seriam tão eficientes na reprodução técnica de processos da mente humana. Isto não desfaz, contudo, a base da comparação, que diz da vinculação direta entre as noções de computação e cognição e, portanto, da possibilidade, ainda que por enquanto apenas teórica, de equiparação entre cérebros e máquinas, uma vez que estas tenham sido adequadamente projetadas. Seria assim aplicável também à formulação de David Marr o que indica a antropóloga Lucy Suchman (2007, loc. 835<sup>18</sup>), em sua discussão das relações humano–máquina:

O acordo entre todos os participantes da ciência cognitiva e disciplinas afiliadas [...] é que a cognição não é apenas potencialmente *como* computação; ela é literalmente computacional. Não há razão, em princípio, por que não poderia haver uma explicação computacional da mente, portanto, e não há uma razão *a priori* para traçar uma fronteira fundamentada entre pessoas, tomadas como processadoras de informação ou manipuladoras de símbolos, [...] e certas máquinas de computação<sup>19</sup>.

Neste sentido, da mesma forma que a cognição é compreendida nestes domínios como literalmente computacional, também parece ser o caso da visão, de modo que trata-se mais do que como uma metáfora a denominação “Visão Computacional”. Mas, se assumimos o que propõe Foster acerca da tarefa do analista, deslocar as superposições de *visão* e *visualidade* demanda, no caso abordado, desfazermos precisamente a sobreposição naturalizada entre uma Visão Computacional e um sentido geral da visão para nos atentarmos aos elementos da construção de tal assimilação. Em alguma medida, a própria existência dos *CAPTCHA* aponta para a persistência desta diferenciação e, talvez, até mesmo para a necessidade prática de sua manutenção. Contudo, a possibilidade de demarcar um limite não significa uma distensão da relação entre o olhar humano e o da máquina, pois a metáfora constituída por sua sobreposição segue a mediar a projeção de um sobre a compreensão do

---

18 O livro foi consultado em sua edição digital no formato *Kindle*, da *Amazon*, e, portanto, não possui indicação de número de página, mas um índice de localização (*location*). Utilizarei esta referência para as passagens citadas.

19 No original: “*The agreement among all participants in cognitive science and its affiliated disciplines [...] is that cognition is not just potentially like computation; it literally is computational. There is no reason, in principle, why there should not be a computational account of mind, therefore, and there is no a priori reason to draw a principled boundary between people, taken as information-processors or symbol manipulators, [...] and certain computing machines*”. Tradução nossa.

outro. Tendo em vista explorar esta relação, será o caso, grosso modo, de desfazer a analogia inscrita na própria denominação do campo, tomando-a na forma de uma *visualidade computacional*.

Poderíamos formular, então, a questão que nos mobiliza – nosso problema – como: quais as implicações da Visão Computacional, enquanto tecnologia e ciência da visão e do visual, para as práticas e o estudo das imagens? Com tal pergunta lançamo-nos a uma investigação conduzida na forma de um mapeamento de seu *dispositivo*, desdobrando a diversidade de elementos que o compõem, a economia interna destes elementos, bem como o modo como agencia e incorpora, em seu funcionamento, a *imagem* e o *observador*. Trata-se, evidentemente, de problema bastante extenso e que não pretendemos resolver plenamente, circunscrevendo-o tanto a partir do percurso teórico sugerido, em torno daqueles que identificamos como conceitos-chave na abordagem deste objeto – o *dispositivo*, a *imagem*, e o *observador* –, quanto por meio dos objetos particulares que tomamos enquanto manifestações da Visão Computacional. Compreendendo-a enquanto um fenômeno disperso que não poderíamos resumir seja aos programas que a realizam, seja às suas manifestações singulares, esforçamo-nos por recompô-la a partir destas partes – que tampouco poderíamos compreender como fenômenos isolados.

Ultrapassando os limites que talvez pudéssemos traçar, circunscrevendo a Visão Computacional a um âmbito técnico ou científico, compreendemo-la enquanto uma questão de maior amplitude que deveríamos tomar não apenas por seus algoritmos – ainda que também sejam objeto de nossa análise – mas também: pelas suas aplicações no cotidiano; pelo modo como a concebem os cientistas engajados em seu desenvolvimento; pelos subprodutos de seu desenvolvimento e operação; pelo modo como figura ou é antecipada na ficção; pelos seus antecedentes históricos; ou, ainda, pela sua apropriação na produção artística contemporânea. Pelo próprio caráter disperso deste território, bem como por sua amplitude e limites incertos, tomamos a pergunta enunciada mais como motivação do que como questão por se responder em definitivo – levá-la a cabo talvez seria algo como a anedota borgiana, do mapa igualado em tamanho ao território que descreve<sup>20</sup>. De outro modo, longe de abarcarmos a totalidade dos aspectos que poderíamos considerar constituintes da Visão Computacional, enfocamos alguns objetos selecionados – seja propositalmente por sua relevância e caráter problemático para nosso enfoque particular (a partir das várias remissões encontradas na investigação); seja, em alguma medida, pelo acaso, que teria nos levado a encontrar este e não aquele objeto em nosso percurso. Em todo caso, assumindo o caráter tanto circunscrito quanto contingente que inevitavelmente permeia nossa abordagem, compreendemos que a composição que elaboramos configura uma dentre muitas compreensões possíveis do fenômeno estudado que, embora passível de justificativa e fundamentação, certamente não seria capaz de excluir outras leituras – ainda que com elas entre em debate. Trata-se de um esforço por conferir certo sentido ao conjunto de elementos que agregamos em nossa abordagem, algo que certamente poderíamos fazer

---

20 Neste conhecido microconto, Borges imagina um império em que a cartografia era realizada com tal rigor que o seu mapa, inútil, coincidia em tamanho e forma com o próprio território (BORGES, 1998).

de outra forma, ainda que, talvez, sem o mesmo posicionamento, ou abarcando outros objetos, além dos que aqui reunimos.

Já, mais de uma vez, fizemos menção à noção de composição (inclusive em nosso título). Trata-se, como voltaremos a discutir adiante, de uma noção desenvolvida por Bruno Latour (2012 e 2010) pela qual sugere-se um exercício de reunião e conexão de elementos heterogêneos e dispersos sem, contudo, homogeneizá-los, encontrando um modo de passarmos de um a outro deles, compondo uma mesma rede, ainda que sejam elementos contraditórios ou incongruentes. Assim compreendemos nosso desafio neste estudo, no qual esforçamo-nos por encontrar nosso fenômeno enquanto uma instância emergente deste conjunto de objetos que mobilizamos para descrevê-lo – mais do que tomando-o como uma categoria previamente definida que apenas projetaríamos sobre os casos concretos abordados. Quanto a eles, vale destacar a remissão a objetos de natureza diversa – oriundos da técnica, da ciência e da arte –, o que não apenas desdobra-se desta noção da composição, como busca encontrar algo similar ao que Latour indica, em outro momento (2005, p. 74-78), como os momentos intermitentes em que a agência dos objetos se faz visível. Neste sentido, assim como ele sugere ao indicar situações propícias para flagrarmos os objetos em ação (LATOURE, 2005, p. 79-82), também selecionamos nossos exemplos em função da visibilidade e inteligibilidade que são capazes de dar a aspectos do fenômeno estudado. Alguns casos demonstram-se especialmente adequados para esta tarefa, como textos técnicos e artigos científicos, que enunciam explicitamente alguns dos pressupostos e modos de operação das máquinas estudadas além de, implicitamente, expressarem crenças e fundamentos envolvidos em seu desenvolvimento. Cabe destacar, contudo, o recurso que fazemos, com perceptível frequência, a exemplos da apropriação e figuração da Visão Computacional na arte e no imaginário da ficção, os quais desempenham um papel importante neste estudo. De um lado, enquanto instâncias de visibilidade e apropriação crítica dos modos de operação da Visão Computacional (como é o caso, por exemplo, de instalações da artemídia abordadas); de outro, enquanto instâncias de formulação de certo imaginário destas máquinas que informam o modo como nos relacionamos com elas ou mesmo certo caminho de nossa reflexão sobre elas – reforçando e atualizando metáforas que são inclusive recuperadas em seu desenvolvimento científico. Também, evidentemente, os casos tomados da arte têm particular importância para a discussão da incidência da Visão Computacional na configuração de modos de ver, de visualidades contemporâneas, embora não tomemos este domínio enquanto instância exclusiva ou mesmo privilegiada de elaboração e manifestação desta questão. Como sugere Crary<sup>21</sup>, haveria uma grande diferença entre práticas representativas (em especial restritas ao contexto da arte) e os modos de percepção próprios de determinado período, os quais são resultado de uma rede bem mais ampla de elementos. Em suma, a arte participa, em nossa abordagem, como um dos

---

21 Na elaboração do problema de pesquisa a que se dedica em *Técnicas do observador*, Crary critica abordagens que lidariam com a visualidade de determinados períodos apenas pela aparência das imagens e trabalhos de arte, indicando que o processo de transformação experimentado no período foi bem além de convenções da prática representativa, estendendo-se a formas de produção de conhecimento e outras práticas sociais que reconfiguraram as capacidades observacionais do sujeito (CRARY, 1992, p. 3).

terrenos em que se manifestam e se fazem conhecer os aspectos deste domínio técnico e científico que compreendemos em torno da Visão Computacional, tensionando referências mais reconhecidas deste domínio, assim como por elas seriam tensionadas.

Cabe destacar, ainda, que ao descrevermos o resultado desta investigação como composições assumimos, também, algumas das implicações do que Latour sugere em sua proposição desta noção (LATOURE, 2010). Por ele assumimos a posição intermediária, sugerida pelo autor, entre universalismo e relativismo: “Do universalismo ele [, o composicionismo,] toma a tarefa de construção de um mundo comum; do relativismo, a certeza de que este mundo comum precisa ser construído de partes completamente heterogêneas que nunca vão compor um todo, mas no máximo um material composto, frágil, revisável e diverso”<sup>22</sup> (LATOURE, 2010, p. 474). Se buscamos, aqui, desenvolver uma compreensão daquilo que seria a Visão Computacional, portanto, dedicamo-nos a uma tal tarefa de construção. Adotamos uma postura que compreende tal fenômeno – diferentemente da aceção que o tomaria como uma ideia que se manifesta concretamente em cada um dos casos trazidos à discussão – como um material composto, frágil e heterogêneo que não é formado senão por estas manifestações, que delas emerge. E, se a partir delas uma idealização qualquer poderia, talvez, ser indutivamente elaborada, buscamos, de outro modo, não transcender as manifestações concretas de tal fenômeno, mas assumi-las como sua real instância de existência, por mais diversas, ou mesmo discordantes, que elas possam ser.

\*

A respeito da tarefa proposta, portanto, deve-se destacar que não haveria como tomarmos a Visão Computacional como um objeto dado e, como discutiremos no Capítulo 1, há mesmo questões por se discutir para que possamos compreendê-la como um dispositivo. Tomando este conceito no entorno das formulações de Foucault (1999 e 1979) e nas suas leituras por Deleuze (2005 e 1999) e Agamben (2009), encontramos, é claro, ampla margem para abarcarmos sob esta concepção configurações as mais diversas. Contudo, em especial no âmbito da operacionalização do conceito para os estudos da imagem – tais como em Jacques Aumont (1993), Jonathan Crary (1992), Anne-Marie Duguet (2012) ou Cezar Migliorin (2005), dentre outros – apesar da diversidade de construções metodológicas, dificilmente poderíamos propor uma transposição simples destas experiências ao nosso estudo. Como já parece se indicar nesta caracterização introdutória, os processos de que participa tal tecnologia encontram-se enredados a tal nível que já não temos clareza, frequentemente, de quando efetivamente ela age (e com quem, ou por meio de quem) ou mesmo de quando começa e quando acaba uma circunstância particular de sua aplicação. Dificulta-se, portanto, a discretização destas máquinas, já que suas configurações complexas e difusas envolvem não apenas equipamentos locais como também remotos, com frequência operados por agentes

---

22 No original: “From universalism, it takes up the task of building a common world; from relativism, the certainty that this common world has to be built from utterly heterogeneous parts that will never make a whole, but at best a fragile, revisable, and diverse composite material”. Tradução nossa.

computacionais ocultos, constantemente interligados. Diferentemente, no caso do cinema, por exemplo, por mais que busquemos compreender o seu dispositivo de forma a englobar diferentes aspectos e momentos de sua constituição, há ainda objetos emblemáticos que em alguma medida, paradigmaticamente, se prestariam a uma abordagem particularizada, como a câmera ou a sala de exibição – que inclusive permanecem como casos exemplares da genealogia do conceito, como na abordagem de Jean-Louis Baudry (1983). Faz-se importante, também, termos clareza quanto a que se deve o foco sobre o objeto técnico, bem como a respeito de seus limites, algo que fazemos numa compreensão que tomamos, ainda que por vieses distintos, tanto da referência em Jonathan Crary (1992, p. 8), para quem os dispositivos ópticos devem ser tomados como pontos de interseção entre técnicas, forças socioeconômicas e discursos filosóficos, científicos e estéticos, entre outros elementos; quanto em Bruno Latour (2005, p. 46), para quem, longe de qualquer determinismo (seja técnico ou humano), a responsabilidade por uma ação não pode, nunca, ser resumida a um ator isolado, mas deve sempre se referir a uma rede mais ampla e heterogênea de elementos. Não apenas existe certa continuidade a conectar estas compreensões como também, na forma que discutiremos adiante, elas se articulam intimamente com aquela do dispositivo.

No caso da Visão Computacional precisaremos ir ao nível de seus algoritmos, de modo a descrevermos seu funcionamento e algumas de suas operações sobre o espaço e os corpos retratados, mas também precisaremos ter sempre em vista aquilo em que se apoiam e com o que se articulam tais formulações – tanto material quanto discursivamente. Uma fonte fundamental para este esforço são as referências primárias do próprio campo – textos técnicos e teóricos – mas também nos auxiliarão os rastros deixados por seu funcionamento em manifestações particulares desta tecnologia. Visamos, em todo caso, compreender mais claramente os modos pelos quais a Visão Computacional se inscreveria em diferentes compostos actantes, em sua diversidade. Tal abordagem, desenvolvida no âmbito da Teoria Ator-Rede (TAR), em particular de Bruno Latour (2005; 2001; 1994), desempenhará papel fundamental em nosso percurso, em que nos permitirá ir ao nível das agências que participam da composição dos dispositivos abordados, de modo a buscarmos compreender a economia interna de forças e tensionamentos que os constituem. Assim poderemos reconhecer, também, a participação dos objetos técnicos na constituição de determinada visualidade, o que implica reconhecermos que quando *nós* vemos, não o fazemos sozinhos, senão em uma associação com uma variedade de outros atores, dentre os quais objetos como a câmera fotográfica, a tela de televisão, ou os algoritmos de Visão Computacional. O que não significa, contudo, tomar os objetos como determinantes deste processo, mas como participantes efetivos que, além de agir, prestam-se como expressões particulares, ou como materializações de algumas das formulações discursivas, inclusive, em que se apoiam. Compreendendo o caráter problemático da própria definição ou circunscrição dos processos da Visão Computacional como um tal *objeto* ou *dispositivo*, será justamente a esta tarefa que se dedicará o primeiro capítulo desta dissertação.

Iniciando-se por esta discussão, o texto se desenvolverá em três capítulos que, como já indicado para o primeiro, se dedicarão a categorias analíticas que em alguma medida conduzirão a discussão proposta. Após o *dispositivo*, os dois capítulos subsequentes se dedicarão, respectivamente, à *imagem* e ao *observador*. Para além de seu caráter fundamental para o estudo das imagens técnicas, constantemente mobilizadas no entorno da fotografia, do cinema e do vídeo, por exemplo, convocamos tais categorias ao texto justamente pelo caráter problemático de sua vinculação ao fenômeno abordado. Diferentemente de uma matriz que se aplicaria de modo inalterado a diferentes objetos, visamos compreender os deslocamentos ou as distorções que sobre elas seriam aplicadas em cada caso. Jonathan Crary (1992, p. 8) sugere abordagem similar no tratamento dado à relação entre técnica e história ao dizer que não haveria como tomar diferentes momentos históricos como *grids* análogos nos quais diferentes objetos poderiam ocupar posições correspondentes. Este parece ser também o caso da relação entre objetos e categorias de análise e que se tornaria ainda mais forte no caso da tentativa de estudar a Visão Computacional pelo viés da imagem ou das visualidades contemporâneas: como compreender o *ver* se este passa a ser atribuído como uma competência, também, das máquinas? Podemos chamar, o que veriam, de *imagens*? Elas transformariam em alguma medida nosso próprio modo de *ver*?

No segundo capítulo, portanto, a discussão se voltará para as questões que nosso objeto de estudo coloca para a discussão da *imagem*. Embora se trate de um desenvolvimento bastante vinculado à imagem digital, percebemos que não são suficientes para a Visão Computacional os aspectos em geral mais destacados em relação a este estágio da história das imagens – como a simulação ou as imagens de síntese, mais relacionadas à Computação Gráfica, no âmbito, por exemplo, do que Santaella e Nöth (2001) elaboram como o “terceiro paradigma da imagem”. Embora aplicações dessa tecnologia articulem-se com frequência à Computação Gráfica, estamos diante de processos que não têm, necessariamente, o *visual* como ponta de saída: diferentemente de uma tecnologia produtora de imagens, neste caso as imagens são tomadas como entrada ao sistema, em geral no registro de base fotográfica, estático ou em movimento. Neste sentido, a Visão Computacional parece trazer questões particularmente importantes para a passagem da fotografia ao regime do digital, ou, se quisermos, informacional: parece operar em seu funcionamento uma tradução de um regime indicial – em geral apontado como aspecto fundamental da fotografia (DUBOIS, 2012; SANTAELLA e NÖTH, 2001) – para um regime mais marcadamente simbólico, com uma forte ênfase no numérico e mesmo no estatístico. Dentre as raras abordagens em torno do tema, Lev Manovich (1993a) propõe uma noção que nos será importante desenvolver que é aquela do *nominalismo visual*, pela qual a Visão Computacional se situaria – juntamente com outras tecnologias, como o radar – no âmbito de uma abordagem profundamente pragmática da imagem enquanto instância de controle e atuação sobre o mundo. Um desenvolvimento que nos indicará um segundo caminho a seguir nesta discussão é o de Paul Virilio (1994 e 1993), em sua indagação acerca das *imagens virtuais instrumentais*, que seriam, para as máquinas, o correspondente de nossas imagens mentais. De um lado, Manovich aponta para uma reflexão acerca de como as imagens são vistas – ou, como sugere

Stella Senra (2013, p. 18), *lidas* pelas máquinas –; de outro, Virilio aprofunda a indagação quanto a que tipo de representação interna às máquinas este processo produziria. Tomaremos estas duas frentes, no segundo capítulo, para discutirmos a imagem neste contexto.

Por fim, voltamos-nos para a questão do *observador*, categoria que tomamos de Jonathan Crary (1992), que sugere tal nomeação para dizer daquele que, em cada momento histórico, seria engendrado pelo dispositivo através do qual engaja-se em práticas de observação do mundo. Embora guarde similaridades fundamentais com a noção de *espectador*, geralmente preferida no âmbito dos estudos em cinema<sup>23</sup>, Crary (1992, p. 5-6) defende o termo *observador* tanto para diferenciá-lo da figura do espectador supostamente passivo circunscrito à situação do *espetáculo* quanto para valer-se de um aspecto presente na ressonância etimológica da palavra, pela raiz latina *observare*. O autor define sua compreensão do observador numa exploração do duplo sentido da palavra, pela qual o observador não apenas vê mas *observa* (submete-se, atende) um arranjo de possibilidades e prescrições que informariam seu modo de ver (CRARY, 1992, p. 6). Em nosso caso, a questão ganha contornos particulares a partir da sugestão, prevista na própria proposição da Visão Computacional, de que as máquinas poderiam ser, a seu próprio modo, *observadores* contemporâneos. Evidentemente, trata-se de suposição que precisará ser discutida, mas, de todo modo, a figura da máquina de visão, como veremos, parece integrar já nossa compreensão do mundo e inclusive motivar determinadas posturas e reações diante do seu modo de *ver* particular.

Caberá, portanto, compreender não apenas em que medida a Visão Computacional se integraria ao arranjo de convenções no qual nós mesmos nos inserimos, quanto, também, em que medida tais máquinas poderiam adquirir, ou não, a posição de *observadores*. Trata-se de uma recolocação particularizada para o campo visual da discussão das fronteiras e das passagens entre humanos e máquinas, tópico que – em particular em meio à discussão da Inteligência Artificial – ganha hoje importância especial para o campo de Estudos da Ciência e da Tecnologia (*Science and Technology Studies*), não apenas em Latour (2005 e 2001) como também, com matizes distintos, no ciborgue de Donna Haraway (2013) e nas reconfigurações entre humanos e máquinas apontadas por Lucy Suchman (2007). Será, ao final, com base neste autor e nestas autoras que perseguiremos o jogo que parece se desenhar a partir das questões interpostas por nosso objeto: entre as formulações da distinção ou da indistinção entre visão humana e visão da máquina, talvez seja o caso de nos determos, de outro modo, mais à construção destes limites e à forma com que – esquivando-nos da equiparação – as máquinas informariam nosso modo de ver e, nós, o delas. Seria neste ponto, afinal, que talvez seja mais produtivo indagarmo-nos, também, quanto aos limites que nós gostaríamos de traçar e às dinâmicas que gostaríamos de estimular entre estas visualidades – em suma, posicionando-nos quanto a como gostaríamos de ver, quanto a que modos gostaríamos de construir e praticar. Será, ao final, sobre este aspecto que retomaremos a questão das implicações políticas do dispositivo que abordamos.

---

23 Cf. COMOLLI, 2008; e SCHWARTZ, 2004







Figura 6 – Ilustração do processo de extração de silhuetas realizado pelo programa de Visão Computacional de *Videoplace* (1969).  
Fonte: KRUEGER, 2003, p. 385.



# 1 Máquinas que veem

## 1.1 Máquinas visuais e de visão

Embora máquinas visuais tenham uma presença histórica reconhecida desde pelo menos o Renascimento<sup>24</sup>, seria apenas mais recentemente que, para além da produção de imagens, teríamos elaborado máquinas capazes de tomar o visual como ponto de partida para a produção de outras formas de representação, como viemos indicando ser o caso daquelas baseadas na Visão Computacional. Se comumente indicamos pelas máquinas visuais artefatos construídos com o objetivo de produzir imagens, indo da câmara escura ao vídeo ou, mais recentemente, à Computação Gráfica ou à fotografia e ao vídeo digitais, tratamos aqui de tecnologia que, ainda que sempre a partir de imagens produzidas por câmeras ou escâneres (portanto delas dependentes), dedica-se a outra tarefa. Lev Manovich (1997) explicita tal especificidade ao propor duas operações que poderíamos identificar no desenvolvimento das máquinas visuais: ele sugere haver, complementarmente à busca da *automação da produção de imagens* (“*automation of imaging*”), um esforço pela *automação da visão* (“*automation of sight*”). Ainda que se trate de uma distinção operatória, que não se propõe exaustiva ou exclusiva, a nomeação destas duas frentes identifica um traço fundamental da caracterização de nosso objeto.

Diferentemente das diversas tecnologias que visam oferecer imagens como saída (*output*) para serem vistas por nós, as máquinas baseadas nos desenvolvimentos da Visão Computacional são desenvolvidas com o propósito de elaborar informações numéricas ou simbólicas acerca da imagem de que dispõem como entrada (*input*). O que se oferece como produto nesta operação geralmente não será composto por imagens (ainda que possa ser elaborado visualmente em estágio posterior), mas frequentemente se dará na forma de dados e informações que não se encontram inscritos, senão visualmente, nas imagens tomadas de início, sendo necessário o trabalho de mecanismos computacionais de inferência e interpretação para descrever seus aspectos aparentes. Tal é o caso, por exemplo, de algoritmos de reconhecimento de rostos que, por mais que possam demarcar com pontos e linhas os aspectos identificados, oferecem, como produtos principais, dados numéricos como: a posição relativa do rosto em relação ao plano da imagem; as distâncias entre traços característicos (olhos, nariz, boca, sobrancelhas); a expressão facial reconhecida; ou, ainda,

---

<sup>24</sup> Evidentemente, seria arriscado postular um início. Em todo caso, parece ser um recorte razoável tomar o Renascimento como um começo operatório desta história, ainda que reconheçamos – como indica Dubois (2004, p. 33) – que mesmo antes da câmara escura, da perspectiva linear e do maquinário desenvolvido como auxiliares à representação pictórica racionalizada, toda e qualquer imagem seria produzida por alguma forma de tecnologia. Não apenas parece ser necessário delimitar um período de maior interesse como as técnicas renascentistas parecem constituir o principal substrato sobre o qual se elaboram as máquinas contemporâneas.

estimativas acerca da idade e mesmo de quem poderia ser a pessoa retratada (Figura 7). Neste sentido, os elementos gráficos apresentados, em geral sobrepostos às imagens tomadas de início, seriam, antes, representações gráficas destes dados numéricos e, portanto, um produto secundário do processo de interpretação. Por este processo reverso que tem a imagem à montante de sua operação, talvez poderíamos dizer que as máquinas de que tratamos ocupam um terreno singular na história da imagem técnica e das máquinas visuais.

Fundamentalmente, o processo empregado parte da matriz de valores numéricos que compõem uma imagem plana digitalizada, o *bitmap*, para dele extrair aspectos daquilo que se encontraria representado, o que envolve tanto operações aritméticas mais simples, quanto procedimentos mais complexos baseados em métodos da Inteligência Artificial. Cabe lembrarmos, neste sentido, que, se pudéssemos supor um ponto de vista *da* máquina, *para* um computador, a *imagem* nomearia apenas um modelo de entrada e saída de dados que orienta tanto o modo pelo qual determinados valores numéricos armazenados na memória seriam traduzidos a estes ou aqueles pulsos elétricos enviados a um equipamento de visualização (um monitor ou um projetor); quanto, vice-versa, o modo pelo qual os pulsos elétricos produzidos por um sensor de imagens em uma câmera ou escâner seriam traduzidos em arranjos de dados numéricos armazenados. Ou seja, dizemos que no nível de sua representação informacional uma imagem digitalizada traduz-se a uma sequência de valores numéricos em nada essencialmente distinta do que também é, neste nível, um som, por exemplo. Golan Levin (2006, p. 468) afirma, a esse respeito, que a Visão Computacional enfrenta o desafio fundamental de que o vídeo digital – mas poderíamos estender também às imagens estáticas – seria computacionalmente *opaco*:

Diferentemente de textos, os dados de vídeo digital, em sua forma básica, não contêm nenhuma informação intrínseca semântica ou simbólica. Como resultado, um computador, sem programação adicional, não é capaz de responder mesmo as questões mais elementares sobre se um clipe de vídeo contém uma pessoa ou objeto, ou se uma cena exterior de vídeo retrata o dia ou a noite etc. A disciplina da visão computacional foi desenvolvida para responder a esta necessidade (LEVIN, G., 2006, p. 468)<sup>25</sup>.

Os processos algorítmicos que abordamos, neste sentido, lidam fundamentalmente com o problema de tornar a imagem digital computável em um nível simbólico ou semântico. Dizermos isto após tal aprofundamento na natureza da representação informacional nas máquinas digitais poderia sugerir uma diferença fundamental entre a visão humana e uma visão da máquina mas, do ponto de vista dos discursos que inclusive sustentam o desenvolvimento da Visão Computacional e outras tecnologias seria, pelo contrário, justamente neste nível que haveria uma passagem possível entre computador e mente humana, ainda que por via da analogia, com a informação como termo comum. Com base

---

25 No original: “Unlike text, digital video data in its basic form [...] contains no intrinsic semantic or symbolic information. As a result, a computer, without additional programming, is unable to answer even the most elementary questions about whether a video stream contains a person or object, or whether an outdoor video scene shows daytime or nighttime, etc. The discipline of computer vision has developed to address this need”. Tradução nossa.

em tais discursos, o que dissemos com relação aos modos de leitura da imagem por tais algoritmos seria transponível aos mecanismos cognitivos empregados em processos similares<sup>26</sup>.

Não é nosso objetivo argumentar a favor ou contra a validade desta equiparação. Em todo caso, do mesmo modo como observamos criticamente a construção discursiva que a sustenta, parece-nos que o tensionamento efetuado na proposição de uma diferença fundamental entre humano e máquina também merece atenção quanto aos argumentos que a sustentam. Trata-se de um caminho que se desenha com frequência na discussão acerca da imagem, por via dicotômica. Embora chegue a relevar em certas passagens o peso da contraposição<sup>27</sup>, é perceptível o aceno a uma tal compreensão por Philippe Dubois (2004, p. 31-67) em seu tratamento do eixo “maquinismo–humanismo”, que elege como um dos três que guiam sua abordagem das *máquinas de imagem*<sup>28</sup>:

as máquinas, enquanto instrumentos (technè), são intermediários que vêm se inserir entre homem [sic] e mundo no sistema de construção simbólica que é o princípio mesmo da representação. Se a imagem é uma relação entre o Sujeito e o Real, o jogo das máquinas figurativas, e sobretudo seu progressivo incremento, virá cada vez mais distender e separar os dois pólos, como um jogo de filtros ou de telas se adicionando (DUBOIS, 2004, p. 38).

O autor indica, portanto, que enquanto elemento intermediário da nossa relação com o mundo – entre o Sujeito e o Real – o desenvolvimento das máquinas de imagem apontaria para um afastamento sempre crescente entre estes “polos”, camada sobre camada de mediação. Tomando como exemplos os desenvolvimentos do Renascimento, que efetuariam uma primeira operação de racionalização e geometrização do olhar, Dubois (2004, p. 36-37) sugere que técnicas como a câmara escura, a *tavoletta* de Brunelleschi ou as portinholas de Dürer já conteriam tal operação fundamental de separação, que seria aprofundada por tecnologias subsequentes. Assim se daria a passagem da perspectiva renascentista à fotografia, desta ao cinema e deste ao vídeo. A imagem informática, por sua vez, levaria ao extremo a operação constantemente renovada de afastamento. Em uma tal narrativa, este último estágio representa algo como uma culminância deste desenvolvimento progressivo, em que o próprio Real se ausentaria em favor da geração de imagens de pura forma, matemáticas. Nas palavras do autor: “A maquinaria que se introduz aqui é extrema. [...] De fato, com a imagem informática, pode-se dizer que é o próprio ‘Real’ (o referente originário) que se torna maquínico, pois é gerado por computador” (DUBOIS, 2004, p. 47). Na imagem de síntese, portanto, o lugar do Real seria ocupado pela máquina na produção de representações em que já não haveria qualquer inscrição indicial do referente, mas apenas simulações e modelizações<sup>29</sup>. Se

---

26 Cf. MARR, 1982.

27 Em certa passagem, o autor indica que não haveria correspondência direta entre o desenvolvimento tecnológico e a artisticidade, de modo que mais tecnologia não significaria menos estética (DUBOIS, 2004, p. 42-43). Com tal afirmação, ele relewa, embora não desfaça, algo da oposição que constrói em seu texto entre o maquinismo e o humanismo.

28 Os outros dois, que não são interessantes tanto a este momento da discussão, são os eixos “semelhança–dessemelhança” e “materialidade–imaterialidade” (DUBOIS, 2004).

29 Cabe ressaltarmos que esta compreensão das imagens informáticas não se restringe a Dubois e que tampouco a contestamos por completo. Outras abordagens que se aproximam ou insinuam tal compreensão incluem as

seguimos a mesma linha de raciocínio, poderíamos supor que, embora não conste da narrativa do autor, a Visão Computacional, faria, por sua vez, a substituição também do Sujeito pela máquina, em que ela assumiria o lugar daquele que vê. Embora esta seja uma abordagem intuitiva e que tem seu valor didático, logo percebemos algumas implicações de tomarmos a questão de tal maneira: pois, se tanto o Real quanto o Sujeito se ausentassem, em um encerramento da máquina sobre si mesma, o que nos restaria deste curto-circuito? Seriam ainda imagens? Seria ainda visão? Seriam ainda sujeitos? O caminho indicado por Dubois parece, dessa maneira, reforçar um enimesmamento que encerraria, sobre si, também a explicação: como se as máquinas pudessem dizer apenas de si quando, evidentemente, as imagens que produzem e as imagens que *veem* seguem tendo sentido para nós e seguem tendo alguma relação, ainda que de outra natureza, com o Real e os Sujeitos. A condição do Real nas imagens de síntese e a condição do Sujeito na Visão Computacional certamente se colocam enquanto questões fundamentais a se fazer diante de tais tecnologias, mas respondê-las pela postulação da ausência parece ser o mesmo que abandonarmos o problema proposto. Melhor seria buscarmos compreender como tais noções seriam reconfiguradas pela operação das máquinas em questão, descrevendo os deslocamentos efetuados, mais que excluindo os termos da equação.

Voltaremos, em capítulos adiante, à questão da imagem e do sujeito e à problemática de sua ausência ou substituição. Neste, nos dedicamos às próprias máquinas, buscando um modo de abordá-las que seja atento à sua constituição e às suas especificidades, mas que, ao mesmo tempo, não reforce dicotomias ou seu isolamento. Esperamos, assim, melhor compreender o modo como as máquinas de visão se inscrevem contemporaneamente na constituição de formas de ver, de *visualidades*, e que, portanto, não excluem nem se opõem a uma visão humana, mas com ela jogam em uma dinâmica de mútua afetação e de composição. Neste sentido, partimos de uma compreensão da máquina para a qual o próprio eixo “maquinismo–humanismo”, que sugere Dubois, precisaria ser repensado. Se, como indicamos, a imagem e o sujeito surgem enquanto termos problemáticos no tratamento destas máquinas, também a própria discussão da concepção de máquina precisa ser recolocada.

Mais do que o uso comum pelo qual circunscreveríamos o sentido de máquina ao de *máquinas tecnológicas*, nos aproximamos, em nossa abordagem, do sentido que propõe Félix Guattari (2003) em um texto em que enfrenta o desafio – cuja formulação toma de Pierre Lévy – de derrubar a “cortina de ferro ontológica que separa o sujeito de um lado e as coisas de outro” (GUATTARI, 2003, p. 51). O filósofo propõe superar a visão estrita que compreenderia a máquina como um problema específico do domínio da técnica, invertendo a hierarquia presumida e tomando, ao contrário, a técnica como um subdomínio de uma problemática maquínica que abarcaria, para

---

“imagens sem gravidade”, de Alain Renaud-Alain (1994), as “imagens utópicas” de Edmond Couchot (2003, p. 160-164) e a formulação por Santaella e Nöth (2001) de um “terceiro paradigma da imagem”. No caso de Dubois, como em vários outros, tal formulação parece se basear principalmente na ausência de registro indicial pelo qual, efetivamente, as imagens permitiriam a inscrição do Real. Na Computação Gráfica tratamos de imagens compostas principalmente a partir de processos de modelização e simulação, nos quais, embora o Real, em sua dimensão imediata, se faça mais distante, permanece enquanto referente à representação, ainda que o vínculo entre eles passe pelo nível do simbólico e, assim, com maior grau de mediação.

confidence : true ( value : 1 )  
 recognize : Mark\_Zuckerberg:43%  
 pose :roll(-0.41) ,yaw(-8.45) ,pitch(1.08)  
 race : white(0.98)  
 emotioin : confused:53%,happy:43%  
 age : 28.6 ( value : 28.6 )  
 smile : true ( value : 0.61 )  
 glasses : no glass ( value : 0.03 )  
 sunglasses : false ( value : 0 )  
 eye\_closed : open ( value : 0 )  
 mouth\_open\_wide : 0% ( value : 0 )

Drop directly in the above box  
 or click to upload your own images

API Key:   
 API Secret:   
 Name Space:   
 User ID :

```
{
  "url" : "base64",
  "face_detection" : [{
    "boundingbox" : {
      "tl" : {
        "x" : 412.8,
        "y" : 154.8
      },
      "size" : {
        "width" : 255.6,
        "height" : 255.6
      }
    },
    "confidence" : 1,
    "name" : "Mark_Zuckerberg:0.43,Barack_Obama:0.42,Jim_Parsons:0.34,Bill_Clinton:0.22"
  }]
```

Figura 7 – Tela que apresenta os dados resultantes do algoritmo de reconhecimento de rostos do serviço *online* Rekognition. Fonte: impressão da tela a partir dos resultados da análise de fotografia de nossa autoria no site <http://rekognition.com>.

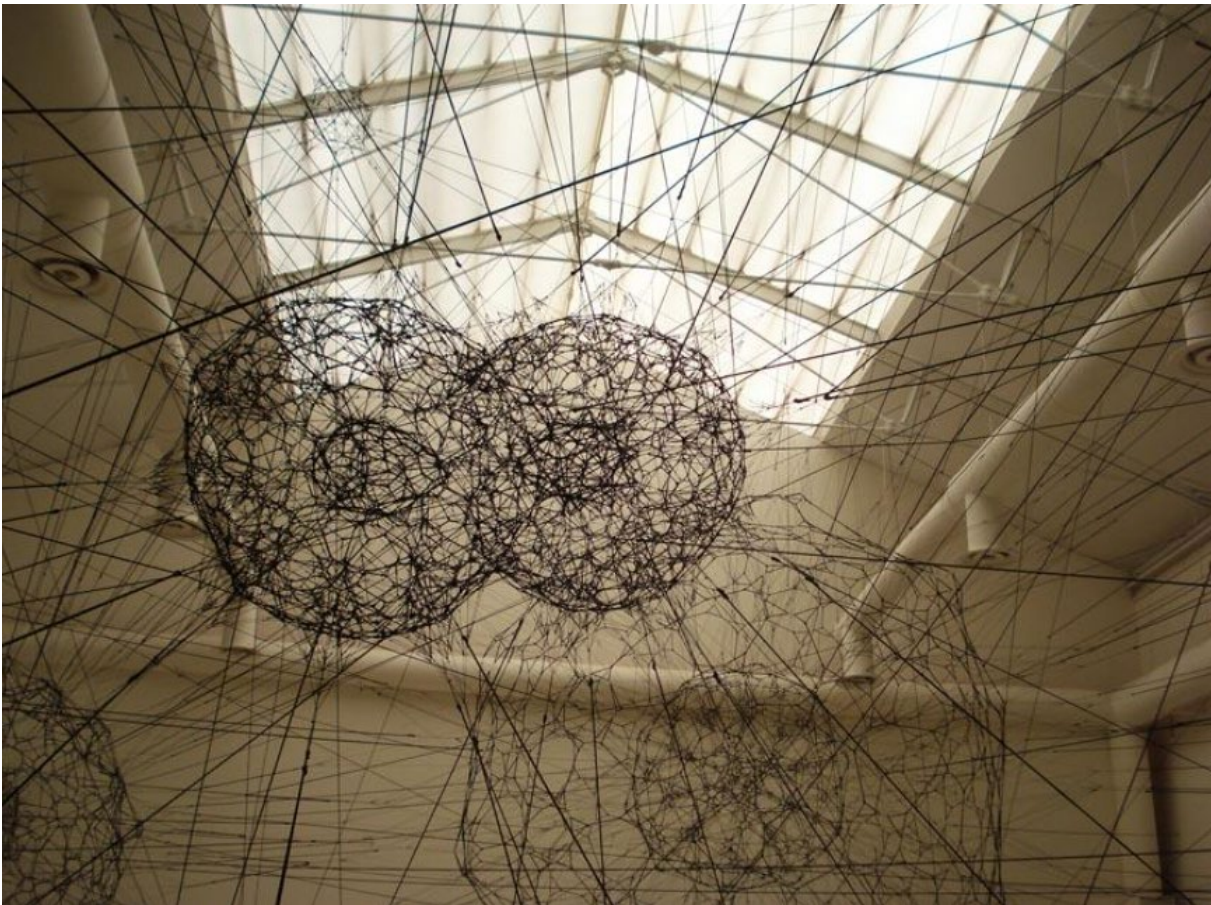


Figura 8 – “Galáxias formando-se ao longo dos fios, como gotículas ao longo das costas de uma teia de aranha”, de Tomas Saraceno, em exposição na Bienal de Veneza de 2009. Foto: Eduardo de Jesus. Fonte: LATOUR, 2012.



além das máquinas tecnológicas, as sociais, humanas e cósmicas, substituindo o nome máquina por *agenciamentos maquínicos*: “categoria que engloba tudo o que se desenvolve como máquinas nos diferentes registros e suportes ontológicos” (GUATTARI, 2003, p. 42). Embora tratemos, neste trabalho, do que poderíamos indicar como uma máquina tecnológica, a formulação de Guattari nos será importante ao permitir-nos conectar este ente técnico a dinâmicas humanas e sociais e ao também nos permitir compreender que seu funcionamento não pode ser tomado isolado de seu entorno – no ensimesmamento a que nos referimos –, devendo a máquina ser descrita a partir dos vetores que a atravessam e a conectam ao que talvez entendêssemos como seu exterior, extra-técnico<sup>30</sup>. Guattari (2003) indica, a esse respeito, que todos agenciamentos maquínicos – e este é um dos aspectos gerais, em comum a todas elas – articulam necessariamente a produção de seu interior (auto-poiese) à de seu exterior (alopoiese). Talvez de modo ainda mais importante, nesta dinâmica produzem uma espécie de *mise-en-abyme*, fractal, o que Guattari sugere a partir de Leibniz:

Desde Leibniz, dispomos do conceito de máquina articulada (de maneira fractal, diríamos hoje) com outras máquinas, elas mesmas compostas de elementos maquínicos até o infinito. Assim, aquém e além da máquina, o ambiente da máquina faz parte de agenciamentos maquínicos (GUATTARI, 2003, p. 41).

Deste modo, longe de isolarmos as máquinas que estudamos em si mesmas ou de as situarmos em relação ao humano como polos opostos de uma dicotomia, a abordagem que buscamos construir pensa a Visão Computacional como necessariamente articulada a outras máquinas de visão, a outros agenciamentos maquínicos do visual, que não se encerram em processos ópticos, químicos ou algorítmicos, mas atravessam formulações discursivas, demandas institucionais e acepções e práticas históricas e culturais do *ver*. Também entendemos que, para além de sua função mais explicitamente constituída, a visão das máquinas, como nossa própria, engloba modos de configuração do espaço, do tempo e dos corpos e, neste sentido, produz bem mais do que apenas inferências acerca das imagens que lhes são alimentadas. No sentido que sugere Guattari (2003, p. 42), falamos que destes *agenciamentos maquínicos* emergem *universos de referência*, vetores de variabilidade ontológica pelos quais cada arranjo, incrustado em meio a outros, produziria determinados modos de ser, afetando, assim, todo o conjunto de elementos com que toma contato. Não dizemos, com isto, que a tecnologia seria um componente determinante de todos os processos, pelo contrário: na complexa teia descrita pelo filósofo, a máquina tecnológica seria apenas um dos elementos participantes deste processo e, se a priorizamos em nosso estudo, é apenas por compreendê-la enquanto um componente emblemático e, em alguma medida, paradigmático desta articulação, de modo que não apenas participaria como também permitiria expressar este conjunto. Tal é, como discutiremos mais adiante, um dos sentidos pelos quais invocaremos o conceito de *dispositivo* para a ela nos referir.

---

30 A fim de evitar equívocos, nos referiremos por máquina, neste texto, prioritariamente no sentido de máquina tecnológica, a não ser quando explicitemos de outro modo, diferenciando estas, do sentido ampliado que nos sugere Guattari, pela remissão, como ele mesmo sugere, a *agenciamentos maquínicos* – estes, sim, sempre ultrapassando o sentido tecnológico estrito.

Cabe dizer que buscamos, nesta abordagem, algo como um ponto de vista oscilante, que nos permita alternar entre diferentes níveis e escalas de leitura, indo dos componentes internos da Visão Computacional e das configurações particulares das máquinas que a incorporam, aos eixos de atravessamento e aos arranjos tecidos com seu exterior, sempre buscando articular e compreender as dinâmicas realizadas entre estes níveis. Uma figura fundamental que assumimos na descrição deste emaranhado é aquela da *rede*, assumida por Bruno Latour (2005, p. 128-133) em uma clara derivação da noção de agenciamento maquínico de Guattari e Deleuze. Como formula o autor, trata-se de um conceito operatório para lidar com o arranjo intrincado pelo qual se elaboram compostos actantes, ou a associação disforme de diferentes agências na efetuação da ação. Não é senão este o sentido da palavra composta *ator-rede*, que nomeia a abordagem desse e de outros autores situados, de início, no contexto dos estudos da ciência e da tecnologia: longe de reduzir-se a um ente isolado, qualquer ação teria origem não em um ator, mas em uma rede de agências, todas participantes e cúmplices de sua efetuação. A rede, contudo, mais do que nomear uma *coisa*, como alerta Latour (2005, p. 131), indica algo como uma figura narrativa e descritiva pela qual buscaremos compreender a dinâmica de tais associações e os fluxos de traduções que se realizam nestes agrupamentos.

Oferecendo-nos uma imagem que julgamos inspiradora para compreender a dinâmica sugerida por Guattari entre o interior e o exterior dos agenciamentos maquínicos, Bruno Latour (2012) toma a obra “Galáxias formando-se ao longo dos fios, como gotículas ao longo das costas de uma teia de aranha”, de Tomas Saraceno (Figura 8), para indicar os processos pelos quais, em meio às redes, emergiriam individuações, sustentadas por suas conexões. Trata-se de instalação apresentada na Bienal e Veneza de 2009 e composta de um complexo arranjo de elásticos conectados em uma forma dispersa que eventualmente se organiza em grandes esferas vazadas. Para o autor, pelo modo como se dão as conexões, a obra explicitaria um modo de relação que poderíamos estender aos processos de individuação: as esferas só poderiam ser reconhecidas em suas individualidades porque emergem de conexões, que projetam ao exterior, e que as sustentam. De modo similar a tais esferas, caberá concebermos tanto a Visão Computacional quanto as máquinas que a efetuam enquanto enredadas e sustentadas por seu entorno, demandando que retracemos estas ligações pelo gesto que Latour denomina *composição*: a agregação de entes heterogêneos sem, contudo, homogeneizá-los, encontrando uma abordagem e um vocabulário que nos permita passar de uma identidade localizada – de uma esfera – às suas conexões exteriores – à rede – que dela irradiam e cujas tensões a mantêm (LATOURE, 2012 e 2010). Aproximamo-nos, assim, da própria concepção que desenvolve Guattari ao indicar, com relação às subjetividades, a necessidade de que invertamos a lógica que tomaria os sujeitos como previamente constituídos e entre os quais estabeleceríamos canais de comunicação: “Pelo contrário, a transferência deve ser primeira, deve já estar lá” (GUATTARI, 2003, p. 49-50). Traduzindo a nosso caso, portanto, não seria suficiente, para compreendermos as dinâmicas em que se inserem as máquinas que abordamos, se não observamos que mesmo sua constituição se dá a partir, e em função, das conexões em que emerge. Contudo, embora seja praticamente mandatário que partamos, didaticamente, de um ponto de vista mais atento a individualidades

constituídas do que a processos de individuação (conduz-nos, a isto, a própria linguagem) – neste sentido partindo, precariamente, do que seria o nosso objeto *em si* – buscamos enfrentar o desafio de alcançar certa velocidade que nos permita desprender desta ou daquela esfera e colocarmo-nos, e à própria descrição, em trânsito: a partir das esferas, chegar às redes.

## 1.2 Visão–algoritmo

Golan Levin (2006), em um texto em que propõe uma abordagem introdutória da Visão Computacional para artistas e *designers*, oferece uma definição generalista do campo: “A ‘visão computacional’ se refere a uma ampla classe de algoritmos que permitem aos computadores fazer asserções inteligentes sobre imagens digitais e vídeo”<sup>31</sup>. (LEVIN, G., 2006, p. 462). O caráter abrangente da definição – destacado pela caracterização do campo como uma “*ampla classe* de algoritmos” – parece incontornável, tanto diante da variedade de operações subsumidas, quanto diante da diversidade de aplicações desdobradas, de modo que cada aplicação particular constituiria uma abordagem técnica específica de seu problema. Trata-se de um domínio cujo escopo de operações seria circunscrito por esta produção de “asserções inteligentes” sobre imagens, o que também é, evidentemente, uma caracterização de contornos difusos, sendo necessário aprofundarmos quanto ao que poderíamos compreender como tal. A amplitude coloca-se como primeiro desafio ao tratamento desta tecnologia em um sentido geral: pois ainda que nos proponhamos abordar alguns dos traços comuns que em alguma medida conectariam este amplo espectro de manifestações, se os tratamos sob uma mesma denominação – Visão Computacional –, não o fazemos com o intuito de tomá-los de forma monolítica, sendo este um risco a se evitar. Buscamos, contudo, compreender em que medida seria possível nomear e caracterizar traços comuns que atravessariam, ao fundo, a diversidade de suas manifestações. Em linhas gerais, poderíamos dizer que a Visão Computacional busca realizar uma tradução pela qual seriam extraídos dos dados numéricos que compõem uma imagem – a matriz em que se distribuem os valores de cor de cada *pixel* – informações acerca daquilo que figurariam. Neste sentido, sua operação seria caracterizada pela passagem entre a materialidade bruta da representação digital da imagem<sup>32</sup>, *pixel a pixel*, para outro modo de sua representação que, ainda que também se efetue na máquina enquanto sequências de dados binários, buscaria descrever suas *características aparentes*.

Forsyth e Ponce (2012), autores de um manual técnico do campo, introduzem um capítulo dedicado a operações elementares da Visão Computacional com um exemplo desta passagem entre os dados

---

31 No original: “*Computer vision*’ refers to a broad class of algorithms that allow computers to make intelligent assertions about digital images and video”. Tradução nossa.

32 Entendemos como adequado falar da materialidade da imagem digital não apenas por uma possível abordagem metafórica que compreenda sua materialidade simbólica, em torno do nível mais básico de sua descrição numérica, mas também porque, neste nível, trata-se efetivamente de uma inscrição material, física, que se faz sobre o suporte de armazenamento digital de que se faz uso, qualquer que seja.

e as características aparentes: “Figuras de zebras e de dálmatas têm [, ambas,] *pixels* brancos e pretos, e em torno do mesmo número, inclusive. A diferença entre as duas tem a ver com a aparência característica de pequenos grupos de *pixels*, mais do que valores de *pixels* individuais” (FORSYTH e PONCE, 2012, p. 107)<sup>33</sup>. Trata-se de um exemplo elucidativo: a diferenciação de imagens de uma zebra e um dálmata a partir dos dados numéricos que a compõem não pode ser feita a partir de operações que tomem seus *pixels* um a um, nem que os tomem conjuntamente descartando as suas posições relativas. A incidência de *pixels* brancos e pretos, em ambos os casos, seria praticamente igual, meio a meio. A operação de tradução da imagem a suas características aparentes parte, de outro modo, de uma análise da imagem que se baseia em agrupamentos de *pixels*, mais que em *pixels* individuais, valendo-se daquilo que seria sua aparência como parâmetro para a formação dos conjuntos. Deste modo, mais do que o dado estatístico da proporção de *pixels* brancos e pretos, valeriam os contornos que circundam seus agrupamentos na imagem: há, é claro, uma clara diferenciação a se fazer, neste caso, entre os círculos que encontraríamos nos dálmatas e as listras, nas zebras. Esta é, tomada de modo amplo, a operação característica da Visão Computacional, da qual poderíamos dizer que derivam suas diversas outras operações e aplicações<sup>34</sup>. Trata-se, contudo, antes de mais nada, de uma caracterização descontextualizada desta tecnologia, que embora ajude-nos a compreender seu modo de funcionamento interno, não dá conta do que realizaria por sua inscrição em contextos pragmáticos efetivos.

Lev Manovich (1993a) aborda aspectos do modo de operação da Visão Computacional tomando como uma de suas principais referências a tese de doutorado de Lawrence G. Roberts, apresentada no início dos anos 1960 ao *Massachusetts Institute of Technology*. Trata-se de pesquisa que teria sido, possivelmente, o primeiro esforço dedicado mais especificamente ao problema da Visão Computacional, então nomeado pelo autor como o da “percepção da máquina” (*machine perception*). Naquele momento, como Roberts (1963) indica em relatório derivado da tese, já se empreendiam pesquisas em torno de operações mais rudimentares de reconhecimento de caracteres e códigos de barras, a partir das quais o pesquisador se propõe como problema ultrapassar esse estágio para enfrentar, efetivamente, o reconhecimento de formas tridimensionais em imagens planas, encaminhando para o objetivo geral do “reconhecimento de dados pictóricos”. Diante desse problema, ele desenvolve processos para a descrição de fotografias, por meio de arestas e formas geométricas tridimensionais, com posterior processamento por métodos de Computação Gráfica para gerar outros pontos de vista da mesma cena (Figura 9). Parte-se, portanto, da identificação, na superfície imagem, tanto dos contornos de cada objeto figurado quanto das arestas que distinguem diferentes planos de um mesmo objeto, para então reconhecer, nas linhas detectadas, quais seriam

---

33 No original: “*Pictures of zebras and of dalmatians have black and white pixels, and in about the same number, too. The differences between the two have to do with the characteristic appearance of small groups of pixels, rather than individual pixel values*”. Tradução nossa.

34 Cf. no “Anexo I – Listagem ilustrativa de operações características da Visão Computacional” uma listagem ilustrativa de algumas das operações características da Visão Computacional, com uma descrição básica de seu funcionamento.

os poliedros figurados, analisando não mais toda a informação pictórica inicial, mas apenas os dados mais simples das linhas encontradas. Em um último estágio, em uma imagem já inscrita no domínio da Computação Gráfica, o programa permitia gerar um outro ponto de vista daquela mesma cena, valendo-se das formas tridimensionais reconhecidas para elaborar outro enquadramento dentre os infinitos então disponíveis por meio da simulação.

Baseando-se na pesquisa de Roberts, Manovich elabora uma abordagem particular da Visão Computacional atenta principalmente à operação de reconstituição numérica de um espaço tridimensional figurado na fotografia. Se a câmera é a máquina que nos permite produzir tal modo de representação plana, tratar-se-ia, aqui, de uma tecnologia que visa, em certa medida, perfazer o caminho inverso, como uma câmera às avessas. Embora seja impossível reconstituir *de fato* o espaço representado, o programa elaborado por Roberts se engaja na tentativa de fazê-lo por meio da simulação, desenvolvendo para isto, de modo igualmente pioneiro, processos de Computação Gráfica. A vinculação fundante entre a Visão Computacional e a Computação Gráfica tridimensional é inclusive ressaltada por Manovich (1993a, p. 135-136), que credita a Roberts a formulação, naquele mesmo período, de algumas das principais operações em que se baseia o domínio da simulação do espaço tridimensional, indicando tal desenvolvimento, contudo, como um subproduto de seu real objetivo: o de desenvolver máquinas que reconhecessem e *entendessem* imagens fotográficas. O autor traça, então, como um aspecto distintivo da pesquisa desse cientista, em relação aos desenvolvimentos anteriores de processamento de imagens ou de reconhecimento de padrões, o fato de que reconheceria não apenas objetos bidimensionais, mas também tridimensionais – uma tarefa que considera fundamental para uma aplicabilidade geral da Visão Computacional:

Somente então ele poderia ser usado para reconhecer um tanque inimigo, para guiar um míssil automático em direção ao seu alvo ou para controlar um braço robótico no chão da fábrica. [...] Enquanto o reconhecimento de padrões estava funcionando para imagens de objetos bidimensionais, como letras ou cromossomos, uma abordagem distinta era demandada para ‘ver’ em 3-D<sup>35</sup> (MANOVICH, 1993a, p.141).

Enfatizando os modos pelos quais tal tecnologia subsidiaria a realização de ações sobre o espaço concreto da cena retratada, Manovich destaca também como a circunscrição da tarefa da Visão Computacional atenderia a demandas militares ou industriais. Lawrence G. Roberts (1963) propunha outras aplicações para sua pesquisa, como a compressão de imagens, com a simplificação da informação armazenada, reduzida às formas reconhecidas; e a elaboração de interfaces de entrada de informação em aplicações de Computação Gráfica tridimensional. Embora tais aplicações tenham de fato se desenvolvido, estando presentes em nosso cotidiano<sup>36</sup>, talvez sejam incontornáveis as

---

35 No original: “*Only then it can be used to recognize an enemy’s tank, to guide an automatic missile towards its target or to control a robotic arm on the factory floor. [...] While pattern recognition was working for images of two-dimensional objects, such as letters or chromossomes [sic], a different approach was required to ‘see’ in 3-D*”. Tradução nossa.

36 Destacamos, como exemplos: os métodos de compressão de imagens baseadas no formato MPEG que baseiam-se na análise de movimento das imagens e na distinção entre fundo e forma para reduzir a redundância da informação em arquivos de vídeo; o sensor Microsoft Kinect que se vale de processos de Visão Computacional para reconstituir

aplicações militares e industriais, em particular quando consideramos o quão adequadas são tais funcionalidades a estes domínios, atendendo a demandas concretas relacionadas à vigilância, à guerra e à racionalização da produção. São estes alguns dos aspectos salientados por Manovich ao indicar, com ênfase à operação geral de reconstituição e controle do espaço, aquele que seria o modo particular de percepção da Visão Computacional:

Mas o que significa ‘entender uma cena’? Com a pesquisa em visão computacional financiada pelo complexo militar-industrial, a definição de entender torna-se altamente pragmática. Na melhor tradição do pragmatismo de James e Peirce, cognição é equiparada à ação. Do computador pode se dizer que ele entende uma cena se ele pode agir nela – mover objetos, compor detalhes, destruir alvos. Logo, no campo da visão computacional, ‘entender uma cena’ implica dois objetivos. Primeiro, significa a identificação de vários objetos representados em uma imagem. Segundo, significa a reconstrução do espaço tridimensional da imagem. Um robô, por exemplo, precisa não apenas reconhecer objetos particulares, mas deve construir uma representação do ambiente circundante para planejar seus movimentos. De modo similar, um míssil não apenas deve identificar um alvo, mas também determinar a posição de seu alvo no espaço tridimensional<sup>37</sup> (MANOVICH, 1993a, p. 143).

Com tal formulação, o autor indica alguns aspectos importantes com relação a nosso objeto. Primeiramente, a inserção da imagem, pela Visão Computacional, em um domínio fundamentalmente pragmático, no qual, de forma bastante literal nos exemplos destacados, as representações visuais são tomadas como parâmetros para o desenvolvimento de ações concretas sobre o espaço – guiar um robô ou um míssil, mas também disparar um alarme diante da identificação, por exemplo, de um corpo que adentra uma região de acesso proibido<sup>38</sup>. Em um segundo momento, destacamos a distinção analítica efetuada pelo autor ao indicar pelo menos dois grandes conjuntos de operações realizadas no âmbito da visão da máquina: aquelas voltadas para o reconhecimento ou identificação dos objetos retratados, e aquelas voltadas para sua localização e disposição no espaço.

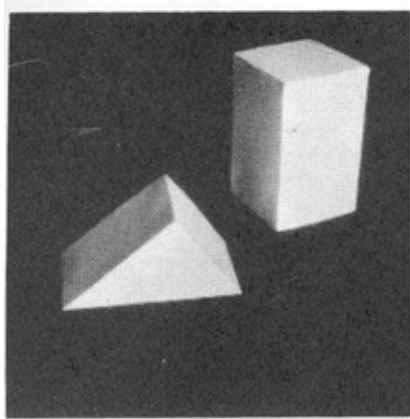
Encontramos em uma instalação do artista Myron Krueger, do final da década de 1960 e início dos anos 1970, um exemplo que se aproxima bem desta operação de *localização* que Manovich enfatiza em sua caracterização da Visão Computacional. *Videoplace* (1969), é indicado por Golan Levin (2006a) como um dos primeiros trabalhos artísticos a se valer dessa tecnologia, descrito pelo próprio artista (KRUEGER, 2003, p. 384) como um ambiente responsivo no qual uma câmera

---

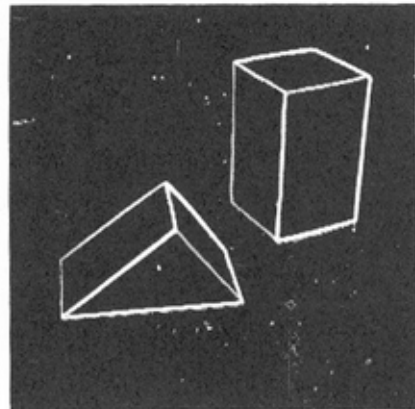
no ambiente gráfico de jogos e outras aplicações os corpos dos jogadores; e a utilização de fotografias para a geração de modelos gráficos tridimensionais, operação oferecida, dentre outras ferramentas, no aplicativo *online* 123D Catch, da empresa Autodesk (cf. <http://www.123dapp.com/catch>).

37 No original: “*But what does ‘understanding a scene’ mean? With computer vision research financed by the military-industrial complex, the definition of understanding becomes highly pragmatic. In the best tradition of the pragmatism of James and Pierce [sic], cognition is equated with action. The computer can be said to “understand” a scene if it can act on it – move objects, assemble details, destroy targets. Thus, in the field of computer vision “understanding a scene” implies two goals. First, it means the identification of various objects represented in an image. Second, it means reconstruction of three-dimensional space from the image. A robot, for instance, need not only recognize particular objects, but it has to construct a representation of the surrounding environment to plan its movements. Similarly, a missile not only has to identify a target but also to determine the position of this target in three-dimensional space*”. Tradução nossa.

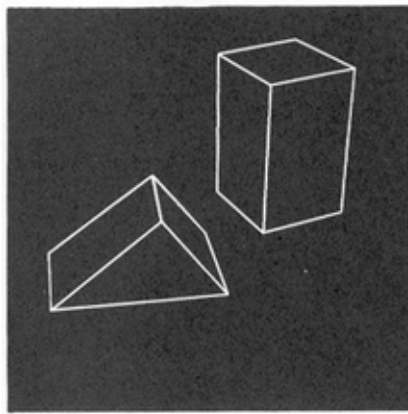
38 Fernanda Bruno (2012) aborda alguns casos como este em sua reflexão acerca da videovigilância inteligente.



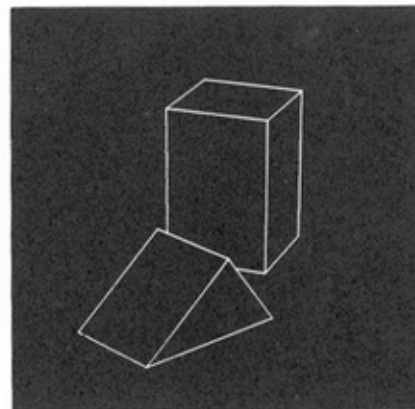
(a) Original picture.



(b) Differentiated picture.



(c) Line drawing.



(d) Rotated view.



Figura 10 – Módulo interativo de *Videoplace* (1969), de Myron Krueger. Fonte: LEVIN, G., 2006a.



capta imagens dos participantes para que seu programa extraia suas silhuetas, que servem como guias para repostas do ambiente, permitindo a interação gráfica com linhas e cores projetadas na instalação (Figura 10). Em uma de suas configurações possíveis, a obra permitiria conectar dois espaços geograficamente remotos, porém igualmente equipados com sistema de interação composto por câmera e projeção, de modo que os visitantes destes dois ambientes interagiriam entre si através das silhuetas projetadas simultaneamente nos dois espaços. Embora em um contexto pragmático distinto daqueles indicados por Manovich, vemos se efetuar, neste trabalho, operação similar ao contexto militar ou industrial em que, ainda que não se busque recompor o ambiente tridimensional figurado na imagem capturada pela câmera, o jogo que se efetua na instalação baseia-se prioritariamente na circunscrição dos corpos em sua posição relativa na imagem. Neste sentido, ainda que bidimensionalmente, na forma de silhuetas (Figura 6), os corpos daqueles que adentram a instalação de Krueger são trazidos ao interior das imagens sintéticas do programa, tomando-os como interfaces de interação com os elementos gráficos sobrepostos e com as silhuetas de visitantes remotos.

Golan Levin (2006a) destaca a importância desse trabalho de Krueger não apenas como um dos primeiros experimentos artísticos a se valer da Visão Computacional, mas também como um dos primeiros trabalhos de arte interativa. Destacamos, noutro viés, sua relevância a partir dos eixos que o conectam a outros trabalhos instalativos desenvolvidos na mesma época. Quase simultaneamente ao trabalho de Krueger, assistimos, no âmbito da entrada do vídeo ao contexto da arte, à realização de videoinstalações por artistas como Bruce Nauman (*Live-taped video corridor*, 1970), Dan Graham (*Time delay room*, 1974) e Michael Snow (*De la*, 1969) que exploraram, de modo característico, a estética da imagem ao vivo, valendo-se do circuito fechado de TV – dispositivo característico da videovigilância – para incorporar à obra o espaço da instalação, o tempo da fruição e o corpo do espectador. Câmeras dispostas nas instalações destes artistas tinham suas imagens exibidas em monitores e telas dispostas em seu interior simultaneamente, ou quase simultaneamente, à sua captura. Em vista de trabalhos destes e de outros artistas do período, inclusive, autores como Michael Rush (2006, p. 111-118) e Anne-Marie Duguet (2012, p. 58-61) ressaltam a interseção entre os domínios da videoinstalação e da videovigilância. Para Duguet o princípio fundamental desta aproximação é a exterioridade do olhar assumido pela câmera: “Na quase totalidade das instalações de vídeo que utilizam o registro ao vivo, a câmara [...] funciona de maneira autônoma. O sujeito se ausenta e delega seu olhar à máquina. Eis em cena o princípio da videovigilância” (2012, p. 58). Existe, claramente, um eixo a atravessar os trabalhos destes artistas, pois tanto em Krueger como em Nauman, Graham e Snow, parece valer a ideia proposta por Duguet de delegação do olhar à câmera e de incorporação à obra de um ponto de vista em certa medida externo aos dos sujeitos. Contudo, há diferenças no modo com que esta perspectiva se inscreve nos trabalhos, pois em *Videoplace*, diferentemente das obras desses outros artistas, não se dá acesso direto ao espectador às imagens produzidas pela câmera que compõe a instalação. De outro modo, o que é dado a ver são já as imagens resultantes de um processamento em que mais do que incorporados à instalação, o espaço

e o corpo do espectador são mapeados, mensurados e computacionalmente processados. Haveria algo de emblemático neste aspecto que percebemos na aproximação dos trabalhos, pois o que se coloca em evidência é a própria operação da Visão Computacional, pela qual a instalação agencia os corpos dos espectadores pela solicitação de que ajam, propositadamente, de acordo com os modos de interação prescritos. Invertendo, em alguma medida, os exemplos indicados por Manovich, portanto, em que a automação da visão guiaria modos de ação da própria máquina sobre o espaço concreto (na movimentação de robôs ou na destruição de alvos), na instalação de Krueger trata-se, de outro modo, de fazer agirem os corpos.

Seguindo este mesmo modo de operação de *Videoplace*, observamos como se elabora uma espécie de configuração paradigmática de obras de arte interativas desenvolvidas desde então. Por sua operação característica de mapeamento do espaço e dos corpos, a Visão Computacional parece se inserir, em diversas outras instalações, em modos de agenciamento similares que inclusive ressoam na própria configuração espacial dos trabalhos. Com a câmera posicionada em plano superior do ambiente, supervisionando o espaço instalativo, tais obras tomam a imagem assim capturada como entrada do sistema que responde aos movimentos dos visitantes através de sons e imagens. *Very nervous system* (1986), de David Rokeby<sup>39</sup>, produz respostas sonoras à movimentação dos corpos diante da câmera, explorando, como Krueger em *Videoplace*, possibilidades para o desenvolvimento de modos de interação gestual com o computador. *Tensión Superficial* (1992), de Rafael Lozano-Hemmer<sup>40</sup>, vale-se do rastreamento dos corpos no espaço para guiar a imagem de um grande olho projetado na instalação, que segue a movimentação do espectador, como se estivesse observando-o. *Boundary functions* (1998), de Scott Sona Snibbe<sup>41</sup>, desloca a câmera para um ponto de vista perpendicular ao chão, o qual também serve como tela na qual são projetadas linhas que dividem o espaço entre os corpos que adentram a instalação, estabelecendo espécies de fronteiras aos domínios espaciais de cada espectador. *Text rain* (1999), de Camille Utterback e Romy Achituv<sup>42</sup> (Figura 11), vale-se da captura de uma imagem frontal, levemente superior, por uma câmera posicionada logo acima da tela na qual são projetados os corpos dos espectadores cujos contornos servem como anteparos para uma chuva de letras que se sobrepõe à imagem, sendo necessário segurá-las com os braços para ler o texto formado. Mais recentemente, *Hand from above* (2009), de Chris O'Shea<sup>43</sup> (Figura 12), já em um momento de ampla disseminação deste tipo de obra<sup>44</sup>, porta esta mesma configuração ao

---

39 Cf. <http://www.davidrokeby.com/vns.html>.

40 Cf. [http://www.lozano-hemmer.com/surface\\_tension.php](http://www.lozano-hemmer.com/surface_tension.php).

41 Cf. <http://www.snibbe.com/projects/interactive/boundaryfunctions>.

42 Cf. <http://camilleutterback.com/projects/text-rain/>.

43 Cf. <http://www.chrisoshea.org/hand-from-above>.

44 Indicamos, hipoteticamente, alguns dos fatores que parecem contribuir a esta disseminação: a ampliação do circuito da artemídia e da arte interativa, o barateamento dos equipamentos utilizados (computadores com alta capacidade de processamento e câmeras digitais) e a disponibilização de bibliotecas de programação de código aberto e preparadas para o desenvolvimento simplificado deste tipo de aplicação. Referimo-nos, a este respeito, especialmente, à biblioteca *OpenCV* (cf. <http://opencv.org/>), desenvolvida inicialmente pela Intel e, de modo mais específico ao contexto da arte, sua transposição às plataformas de desenvolvimento *OpenFrameworks* (cf. <http://openframeworks.cc/>) e *Processing* (cf. <http://processing.org/>).



Figura 11 – *Text rain* (1999), de Camille Utterback e Romy Achituv. Fonte: <http://camilleutterback.com/projects/text-rain/>.



Figura 12 – *Hand from above* (2009), de Chris O'Shea. Fonte: <http://www.chrisoshea.org/hand-from-above>.

espaço público, valendo-se de um painel de LED para exibir a imagem captada de uma praça sobre a qual se exhibe uma mão que, tal qual a de um gigante, manipula os corpos dos transeuntes e com eles interage.

Nesta pequena amostra de um universo mais amplo de obras que se valem da Visão Computacional como componente de suas interfaces de interação, vemos se repetir modos de configuração espacial e de estruturação de lógicas de respostas à interação que sugerem modos característicos de agenciamento pela Visão Computacional do espaço e dos corpos. Em todos eles, o campo de visão da câmera demarca um território no espaço instalativo para o desenvolvimento das ações. Seu ponto de vista, em geral levemente superior ao dos espectadores – sendo, em alguns casos, situada a pino –, favorece esta operação de mapeamento e circunscrição do espaço. Diante deste olhar, os corpos dos espectadores, encontram-se como se fossem rendidos e desarmados: uma vez que adentram o território assim demarcado, estão necessariamente disponíveis e assujeitados a esse olhar. Golan Levin (2006b) explora esta questão em um bem humorado ensaio visual acerca de uma suposta “pose da artemídia” (“*media art pose*”), indicando como, com frequência, trabalhos interativos como os destacados (Levin inclusive cita *Videoplace* e *Text rain* entre seus exemplos) nos levam a interagir levando nossas mãos ao alto. Passando por vários possíveis significados deste gesto – do exercício de controle do maestro diante da orquestra à invocação dos poderes divinos pelo Papa – Levin inverte, ao final, o sentido geralmente atribuído às obras interativas. Diante de tantas poses com as mãos ao alto, ele sugere:

Esses sistemas ‘interativos’ são caracterizados pela ideia de que o usuário está controlando um sistema responsivo. A julgar deste padrão recorrente de comportamento, contudo, poderíamos especular que esse controle é, em última medida, uma ilusão? Talvez esses sistemas que estão nos controlando, afinal.

A pose da artemídia é a postura da rendição total<sup>45</sup> (LEVIN, G., 2006b).

As mãos ao alto são, então, aproximadas da imagem de um soldado que se rende diante do pelotão inimigo. Parece pertinente a sugestão de Levin, em particular quando aproximamos esses exemplos da operação que Manovich identifica, inicialmente, no estudo de Lawrence G. Roberts e em desenvolvimentos posteriores. Recompôr o espaço e *localizar* seus objetos parecem ser, efetivamente, algumas das operações centrais realizadas pela Visão Computacional visando, em última medida, mapear e exercer controle sobre um espaço que, demarcado pelo campo de visão da câmera, converte-se em território. Como um segundo nível deste agenciamento, portanto, encontramos ainda uma operação de *acionamento* de respostas tanto do próprio programa quanto dos espectadores, de modo que poderíamos compreender este domínio pelo par *localizar–acionar*. Nesse contexto, por mais que se busque oferecer algo como uma interface amigável ou intuitiva, responsiva aos gestos, há uma ambiguidade no modo de constituição de tal interface, já que nosso engajamento não é, de

---

45 No original: “*These ‘interactive’ systems are predicated on the idea that the user is controlling a responsive system. To judge from this recurring pattern of behavior, however, might we speculate that this control is ultimately an illusion? Perhaps these systems are really controlling us after all. | The ‘Media Art Pose’ is the posture of total surrender*”. Tradução nossa.

todo, voluntário: basta adentrar o território da câmera para estarmos sujeitos à sua operação de mapeamento, rastreamento e controle. Basta sermos localizados para que estejamos sujeitos a um *acionamento*, seja do programa, seja de nossos próprios corpos.

Manovich prioriza, em sua definição da Visão Computacional, sua inscrição no domínio pragmático da ação sobre o espaço concreto. É neste sentido, inclusive, que ele sugere a pesquisa de Lawrence G. Roberts como um trabalho inaugural do campo, distinguindo-o de pesquisas na área de reconhecimento de padrões, operação que, embora considere importante para a automação da visão, ele situa em um plano secundário de sua definição. Como vimos a partir dos exemplos levantados, o par *localizar–acionar* compreende, de fato, um conjunto de suas operações fundamentais, porém também é um traço de alcance restrito, não apenas por circunscrever a Visão Computacional a um contexto de ação específico como, também, por não abarcar manifestações cotidianas contemporâneas desta tecnologia, no entorno das operações de classificação de imagens, de reconhecimento de objetos e corpos e da indexação e busca de informação em grandes arquivos visuais. Nesse segundo grande conjunto de operações das máquinas de visão, que talvez pudéssemos nomear como *reconhecimento* – diferenciando-as, assim, daquelas de *localização–acionamento* – tratamos da inscrição desta tecnologia em um domínio pragmático distinto, em que não se objetiva tanto a ação sobre um espaço concreto (a navegação de robôs, o mapeamento da movimentação dos corpos ou a destruição de um alvo) mas, principalmente, a ação sobre um plano informacional: organizando, categorizando e indexando imagens. Pelo segundo nível deste modo de agenciamento, portanto, talvez pudéssemos nomeá-lo também por um par: *reconhecimento–conexão*. Evidentemente, não há como separar cirurgicamente estas operações, até porque, em geral, não haveria como realizar um tipo de operação sem o outro. Contudo, esta distinção operatória entre a *localização–acionamento* e o *reconhecimento–conexão* ajuda-nos a compreender ênfases distintas que observamos em alguns contextos de aplicação e, principalmente, permite-nos compreender certa especificidade das operações que observamos, contemporaneamente, pela inserção dos algoritmos de visão no contexto das tecnologias informacionais em rede.

Os exemplos da artemídia ilustram uma necessidade premente das operações de *localização–acionamento* que é a de circunscrição de um território, a delimitação de um campo de ação que coincidirá com o campo de visão da câmera, delimitando aquilo que é por ela enquadrado e no qual os programas podem agir – sobre si mesmo e sobre os sujeitos. Em robôs ou em artefatos bélicos, como mísseis, a câmera é acoplada a um objeto em movimento que em alguma medida varre o espaço em busca de alvos e obstáculos. Em contextos como a videovigilância, contudo, estamos mais próximos do caso das instalações interativas: ainda que possa haver certa margem de movimentação da câmera, a operação da Visão Computacional baseia-se, em última medida, em um determinado espaço circunscrito de monitoramento e de ação. De outro modo, as aplicações que situamos mais próximas destas operações de *reconhecimento–conexão* parecem partir de outro paradigma que muitas vezes prescindiria da demarcação de um território, até porque não é tanto sobre este espaço

que incide sua ação. Ainda que o campo de visão da câmera siga como uma demarcação importante daquilo que pode ser percebido, não há necessidade de inscrição deste território demarcado no modo de funcionamento do programa. Se lidamos com os algoritmos aplicados no contexto da internet, por sua vez, já nem falamos exatamente de uma ou outra câmera: lhe são alimentados, mais do que um determinado ponto de vista, inúmeros deles, oriundos das diversas imagens em circulação, tomadas por câmeras fotográficas e de vídeo, por celulares e computadores, extraídas de filmes e vídeos.

O caráter distribuído das operações de *reconhecimento–conexão*, mais do que uma particularidade do contexto atual de aplicação da Visão Computacional, constitui uma importante estratégia de seu próprio desenvolvimento. Costuma-se dizer, no contexto da computação, de uma gradação entre processos que se atêm mais diretamente à materialidade bruta dos dados binários ou do nível mais estrito de uma *linguagem da máquina* (relativa às operações elementares dos processadores), e outros que se constroem sobre camadas de abstração, que já não remetem diretamente a este nível mais baixo e elementar das operações da máquina, sendo mediados por camadas de código que traduzem estas operações a termos e conceitos mais elaborados e compreensíveis por humanos. Diz-se, desta forma, de operações de *baixo nível*, mais próximas do domínio concreto da máquina, e operações de *alto nível*, que abstraem estes aspectos em favor de lógicas mais elaboradas de funcionamento. Forsyth e Ponce (2012) valem-se em parte desta gradação para organizar a estrutura de sua apresentação técnica da Visão Computacional, dividindo seu livro entre as operações mais elementares, relacionadas à formação da imagem ou à aplicação de filtros lineares, e aquelas que indica como uma “Visão de nível médio” (“*Mid-level vision*”) ou como uma “Visão de alto nível” (“*High level vision*”). No caso das operações de *reconhecimento–conexão*, encontramos-nos, em geral, em um nível mais abstrato no qual já não se trata tanto de segmentar a imagem ou descrever suas características aparentes mas, construindo sobre estas operações, vincular os traços descritos a determinados conceitos, sejam eles os nomes dos objetos identificados, a expressão facial estimada ou o contexto retratado na imagem.

Para a realização destas operações, a estratégia empregada em geral envolve o recurso ao método de aprendizado de máquinas (*machine learning*), relacionado ao campo da Inteligência Artificial. Trata-se, em linhas gerais, de um método de desenvolvimento que envolve a criação de algoritmos capazes de “aprenderem” a desenvolver a tarefa pretendida a partir de um processo de treinamento. Envolve, assim, em uma de suas vertentes, chamada de *aprendizado supervisionado* (*supervised learning*), a alimentação ao sistema de uma grande quantidade de exemplos previamente selecionados daquilo que deve reconhecer. A um algoritmo de detecção e reconhecimento de rostos, por exemplo, seriam apresentadas centenas de imagens de rostos, previamente colhidas e recortadas, a partir das quais o programa encontraria as características recorrentes que deve tomar para orientar seu processo de reconhecimento. Em outra frente, chamada de *aprendizado não supervisionado* (*unsupervised learning*), os dados de treinamento não correspondem a uma base previamente circunscrita mas

apresentam-se, desde o início, como heterogêneos: é o próprio programa, nesse caso, que identifica padrões recorrentes e infere possíveis categorizações. Logo, para tais estratégias de desenvolvimento, o contexto de intensa circulação de imagens na rede constitui não apenas um campo privilegiado de ação, diante da necessidade de organizar e dar sentido a este conjunto de informações, mas também um importante substrato para o seu próprio desenvolvimento, já que tais imagens serviriam, também, para o processo de treinamento desses algoritmos.

Não por acaso, alguns dos principais empreendimentos no âmbito das operações a que nos referimos sob o par *reconhecimento–conexão* – que não dizem respeito apenas à identificação de rostos, por exemplo, mas a toda operação de classificação de imagens e aspectos selecionados, dizendo-se tratar deste ou daquele objeto, desta ou daquela situação – são hoje desenvolvidas por empresas como *Google* e *Facebook*, que têm acesso e mesmo certa margem de controle sobre parcela significativa do tráfego de informações na internet. Em ambos os casos, implementam sistemas bastante complexos de aprendizado de máquinas, que se valem não apenas de uma quantidade gigantesca de imagens de treinamento, vinculadas ainda a um volume significativo de *metadados*<sup>46</sup>, como de uma enorme quantidade de unidades de processamento, de modo a dar conta da análise deste volume de informação. Como um subproduto deste tipo de investimento, a *Google* oferece um aplicativo de celular chamado *Google Goggles* que realiza a interpretação automatizada de imagens tomadas pelo usuário. Vale-se, para isto, do chamado processamento na nuvem: a imagem tirada com o celular é transmitida aos servidores da empresa para análise, sendo nelas reconhecidos caracteres, códigos de barra e os chamados códigos QR, mas também logomarcas e mesmo os lugares retratados. O aplicativo pode tanto ser utilizado a partir do acionamento do usuário quanto pode ser autorizada sua execução em segundo plano, por vezes surpreendendo-o ao reconhecer algo em uma fotografia recentemente tomada. Por exemplo, pode o usuário registrar a imagem de uma pintura e logo ser informado pelo aplicativo quanto a seu título e autor, com um *link* para uma página da *Wikipedia* sobre ela. Ou pode registrar um edifício e ser indicado a ele seu nome, bem como encaminhado a outras fotografias do mesmo local. Fotografando a capa de um livro, pode ser encaminhado a um *site* em que poderia efetuar sua compra. De modo similar, a *Amazon* incorporou em seu telefone celular *Fire Phone*, lançado em 2014, a função chamada *Fire Fly*, com um botão do aparelho dedicado para o acionamento de algoritmos de reconhecimento que leem a imagem captada pela câmera para extrair dados de contato de cartões de visita, informações em pôsteres e *outdoors* bem como para identificar cenas de filmes e programas de televisão e livros por sua capa – claro, também sempre permitindo que adquiramos alguns destes produtos em sua loja *online*.

Em todos estes exemplos, a imagem tomada pela câmera do celular passa a desempenhar o papel de um *link*, encaminhando a navegação do usuário na internet, facilitando o acesso e o consumo de

---

<sup>46</sup> Metadados são o registro de informações suplementares acerca de um dado tido como principal. Uma fotografia digital, por exemplo, muitas vezes vem acompanhada de dados quanto ao modelo da câmera e suas configurações técnicas. O arquivo de vídeo de um filme pode conter dados sobre a produção. Em sistemas personalizados, inúmeros tipos de dados podem ser vinculados ao arquivo.



conteúdos relacionados. Não se trata, neste sentido, da ação sobre um espaço concreto ou mesmo do monitoramento de um tal domínio circunscrito, como seria o caso do que viemos indicando pela ênfase nas operações de *localização–acionamento*: embora igualmente pragmático, a ênfase no par *reconhecimento–conexão* esboça outro modo de atuação desta tecnologia que se inscreve, principalmente, em um plano informacional ou simbólico. A interpretação automatizada de uma imagem vincula-se principalmente ao acionamento de uma rede simbólica e conceitual na qual ela se insere, articulando-a com outras imagens e textos, por exemplo, com os quais guarda alguma relação de semelhança e contiguidade ou, ainda, com os quais têm, como elo, apenas o fato de que, estatisticamente, pessoas que a acessam também acessam esses outros conteúdos. Essa rede simbólica também opera a fim de induzir os sujeitos à ação (comprar um produto, visitar um prédio, comer em um restaurante etc.) retornando, assim, de um contexto pragmático restrito à organização e indexação da informação, a um tipo de ação que incide diretamente no espaço e nos objetos do mundo. Não é muito relevante, portanto, nesses casos, indagarmo-nos se o programa de fato entende aquilo que vê, mas apenas se traça relações pertinentes. Vemos, mais uma vez, destacar-se o caráter pragmático da visão da máquina, que será tão mais eficaz quanto a partir dela realizar ações adequadas e relevantes, do ponto de vista do usuário ou, mais frequentemente, dos provedores de conteúdo.

David Rokeby explora operações de *reconhecimento–conexão* da Visão Computacional na obra *The giver of names* (1990-)<sup>47</sup>. Embora em desenvolvimento desde 1990, apenas em 1997 o artista expôs a primeira versão de seu trabalho, que passa, a partir de 1998, a contar com o uso de recursos de Visão Computacional (ROKEBY, 2010). Desde essa versão da instalação, o espectador é convidado a escolher alguns objetos – de uma diversidade deles, espalhados pelo chão – dispendo-os sobre um pedestal diante de uma câmera conectada a um computador (Figura 13). O programa analisa a imagem captada, extraindo-lhe os contornos, analisando sua cor e textura, e busca identificar, a partir de seus traços encontrados, quais seriam aqueles objetos dentre uma base conhecida. O reconhecimento ativa uma rede conceitual, composta por uma base de nomes de objetos, sensações e ideias – extraídas, entre outras fontes, de textos escritos pelo artista e obras da literatura – que são articuladas a partir das formas percebidas, de forma a elaborar uma frase em alguma medida afetada pela percepção visual, pela máquina, daquela cena (Figura 14).

A respeito desse trabalho, Christiane Paul (2008, p. 150) sugere que ele nos ofereceria algo como uma “anatomia da significação” (“*anatomy of meaning*”), levando-nos a níveis cada vez mais abstratos que, desta forma, abririam novas formas de contexto e significação. Em última medida, sugere a autora, trata-se de um trabalho que reflete sobre como as máquinas pensam e como nós as fazemos pensar (PAUL, 2008, p. 150). Destaca-se, nas descrições da obra feita por Christiane Paul e outros autores (HENRY, 2004 e TUER, 2004), como a operação da obra se distancia daquelas típicas do par *localização–acionamento*, pois não se trata de recompor, a partir da imagem, um espaço

---

47 Cf. <http://www.davidrokeby.com/gon.html>.

concreto, como é o caso das aplicações que indicamos anteriormente, ou dos trabalhos de artemídia derivados. Trata-se, antes, de uma operação de abstração pela qual acessamos, a partir da imagem, um espaço difuso, conceitual, em que mais do que circunscrever regiões ou monitorar movimentos, estabelecem-se conexões e ativam-se percursos, fazendo emergir novos sentidos a partir das relações estabelecidas. Alguns dos comentadores deste trabalho (HENRY, 2004, p. 44-45; TUER, 2004, p. 32-33) descrevem a experiência da obra como marcada pela sensação de uma lacuna entre os modos de percepção da máquina e os modos de percepção humana, em especial pelo fato de que, as formulações linguísticas resultantes são, em sua maioria, embora gramaticalmente corretas, sem sentido. Evidentemente, o artista não possui compromisso com a eficácia de seu programa – segundo uma expectativa mais ortodoxa. De toda forma, talvez se trate mesmo de um indício do caráter probabilístico de sua operação: no âmbito das estratégias de aprendizagem, neste espaço difuso em que se processam suas operações, a construção de sentido inscreve-se, fundamentalmente, no terreno da estatística. Em alguma medida, talvez a dissolução dos territórios que descrevemos na passagem entre a *localização–acionamento* e o *reconhecimento–conexão* derive do próprio modo de operação enfatizado em cada um destes domínios, de forma que o espaço da Visão Computacional em sua aplicação mais distribuída seja descrito, tanto dentro, quanto fora da máquina, enquanto fluxos e conexões.

Conquanto não sejam homogêneos os modos de agenciamento da Visão Computacional em suas diferentes aplicações, trata-se de tecnologia que cumpre, de modo preponderante, uma função que poderíamos descrever pela passagem do regime do visível ao regime do informacional, do computável. Em qualquer uma das ênfases que identificamos em nossa descrição, por mais que abranjam, cada uma, diferentes registros do agenciamento do espaço, dos corpos e da significação, o traço comum a conectar suas múltiplas aplicações parece ser este de tradução do visível, do aparente, à informação e desta, por conseguinte, à ação. Em certo sentido, trata-se de reduzir a imagem a um estado mais determinístico, condensando-a e dela extraindo descritores mais simples, em geral articulados a seu valor pragmático. Se entendemos essas máquinas de visão na concepção dos agenciamentos maquínicos elaborada por Guattari (2003), faz-se necessário, contudo, uma vez que já nos atentamos a seu modo de funcionamento particular, que passemos desta a outra escala de análise<sup>48</sup>. Passamos, portanto, a percorrer alguns dos vetores rumo ao além da máquina, indo deste nível mais específico de seu funcionamento ao seu exterior. Visamos, neste movimento, compreender em que medida os agenciamentos que viemos descrevendo se inscrevem em uma rede mais ampla de elementos que incluem, para além da operação técnica destas máquinas, sua relação a máquinas sociais, discursivas, epistemológicas, políticas. Ainda que já venhamos indicando as articulações entre os algoritmos abordados e seu entorno imediato, chega o momento de compreendermos que

---

48 Uma listagem mais pormenorizada de algumas das operações que subsumimos em torno das categorias *localização–acionamento* e *posicionamento–conexão* pode ser encontrada, com breves descrições dos procedimentos técnicos implicado, ao final deste volume. Cf.: Anexo I – Listagem ilustrativa de operações características da Visão Computacional



Figura 13 – *The giver of names*, de David Rokeby, em montagem de 2008. Fonte: <http://www.davidrokeby.com/gon.html>.

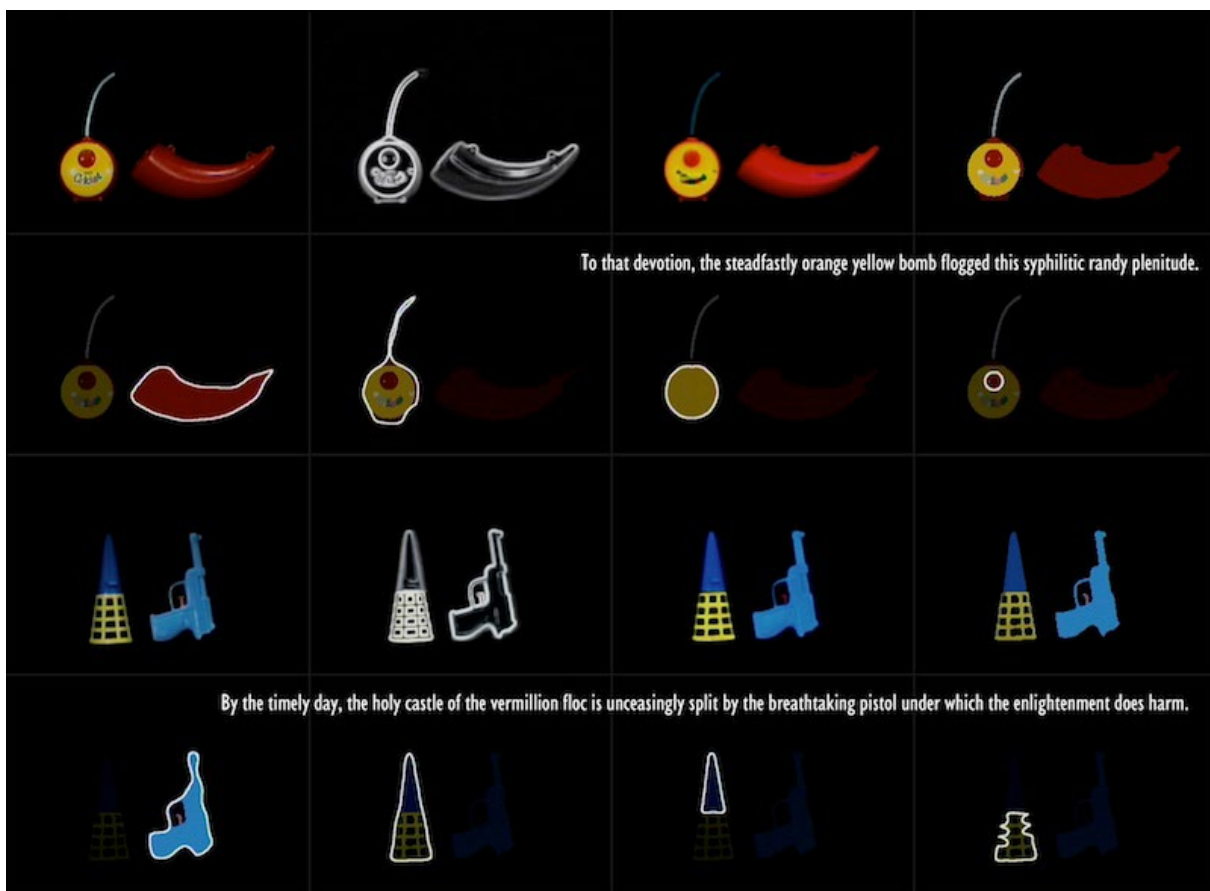


Figura 14 – Ilustração da sequência de algumas respostas visuais e textuais do programa de *The giver of names*, de David Rokeby.  
 Fonte: <http://www.davidrokeby.com/gon.html>.

eles não agem sozinhos, senão articulados a uma multiplicidade de outras agências, que tanto os fazem agir, quanto eles, a elas.

### 1.3 Dispositivo, máquina, paradigma

No centro de formulações que buscam dar conta de uma constituição multidimensional dos arranjos dos quais emergem os objetos técnicos, o conceito de *dispositivo* desempenha um papel fundamental ao nomear a instância em que se interconectariam e entrecruzariam discursos e materialidades, operações simbólicas e concretas. Trata-se de conceito que renderia, por si só, uma discussão bastante extensa, que tem tido um interesse particular nos estudos da comunicação, e cujos aspectos diferentes autores têm se dedicado a compilar, tanto ao longo de suas manifestações transversais quanto a partir de vertentes específicas, como incidências nos estudos da imagem (particularmente no cinema, no vídeo e na instalação) e dos *media*, em geral<sup>49</sup>. Em meio à dispersão de tais manifestações, a obra de Foucault (1999 e 1979), seguida pelas leituras particulares de Deleuze (2005 e 1999) e Agamben (2009), parecem constituir uma espécie de “núcleo duro” do conceito, garantindo algum nível de continuidade entre suas compreensões mais dispersas – ainda que mesmo entre esses autores persistam diferenças. Na filosofia, o *dispositivo* é comumente identificado como uma figura conceitual central na obra de Foucault, atravessando algumas de suas obras e tendo em sua análise da prisão panóptica, em *Vigiar e punir* (1999), um dos exemplos mais recorrentes. Contudo, como foi inclusive apontado por Agamben (2009, p. 28), o filósofo deixara de definir claramente em seus textos a que se referia por tal conceito. Um dos raros momentos em que se aproxima de fazê-lo, oferecendo-nos uma breve síntese, o faz em uma entrevista:

Através deste termo tento demarcar, em primeiro lugar, um conjunto decididamente heterogêneo que engloba discursos, instituições, organizações arquitetônicas, decisões regulamentares, leis, medidas administrativas, enunciados científicos, proposições filosóficas, morais, filantrópicas. Em suma, o dito e o não dito são os elementos do dispositivo. O dispositivo é a rede que se pode estabelecer entre estes elementos (FOUCAULT, 1979, p. 244).

Destacamos, com relação a esta formulação, dois aspectos fundamentais que Foucault elege ao sumarizar sua proposição. Em primeiro lugar, a convocação pelo autor da figura metafórica da rede – ainda que sem muito aprofundar sobre suas implicações –, que nos leva a compreender um dispositivo não como uma individualidade discreta (*um* objeto técnico, *um* edifício), mas como um arranjo de elementos heterogêneos, sem que possamos situá-lo em nenhum destes elementos, mas apenas nas relações traçadas entre eles. É em direção a tal acepção que acenamos, anteriormente, ao problematizar definições técnicas ou funcionais pelas quais poderíamos circunscrever, em um primeiro momento, a Visão Computacional. A *rede* compreende, inclusive, uma metáfora

---

49 Cf.: DUGUET, 2012; MIGLIORIN, 2005; BRASIL e JESUS, 2008; MOUILLAUD, 1997; BRUCK, 2012; KESSLER, 2007; KLEIN, 2007.

importante para outros autores que orbitam a discussão das técnicas em compreensões similares, ainda que com sotaques particulares, tais como o *rizoma* e os *agenciamentos maquínicos* de que falam Gilles Deleuze e Félix Guattari (1995a) na proposição de um modelo de pensamento avesso à formas hierárquicas, binárias ou arborescentes. De outro modo, como também já indicamos, Bruno Latour (2005, p. 121-140) também escolhe a rede como conceito operatório para abarcar os compostos actantes pelos quais a ação seria sempre decorrente de formações distribuídas, nunca individuais<sup>50</sup>. Tais aproximações serão importantes na construção de nossa própria compreensão do termo.

O segundo aspecto por se destacar refere-se à ênfase que Foucault confere à articulação entre o dito e o não-dito. No mesmo sentido, Agamben (2009, p. 29) dirá do linguístico e do não-linguístico. Como indica Fernanda Bruno (2013, p. 19), esse *não-dito* não se refere ao implícito ou ao oculto: “é sim o que não se diz pela via de um enunciado linguístico, mas que se expressa e se afirma em técnicas, procedimentos, ordenações espaciais, arquiteturas etc.”. Acrescentaríamos, ultrapassando os verbos ainda do plano linguístico empregados pela autora (“expressar”, “afirmar”), que o não-dito pode ser ainda o que efetivamente *se faz*: uma *ação* para além do que *se diz*<sup>51</sup>. Nesse sentido, o dispositivo é compreendido por Deleuze (2005, p. 42-43) como um conceito gestado no enfrentamento do desafio teórico de articular tais dimensões, a operação técnica e a simbólica, buscando compreender como se dá a coadaptação entre elas. Trata-se de aspecto enfatizado por Deleuze em sua leitura de *Vigiar e punir* (FOUCAULT, 1999) indicando como um de seus problemas centrais a compreensão da relação entre o direito penal (instância discursiva) e a própria prisão (instância não-discursiva), entre os quais não haveria forma comum ou correspondência, mas constantes penetração mútua e ajustamento (DELEUZE, 2005, p. 41-44). A articulação sugerida entre o dito e o não-dito, neste sentido, ao mesmo tempo amplia e aprofunda o caráter “decididamente heterogêneo” que Foucault atribui ao arranjo do dispositivo.

Tais aspectos destacados, contudo, carecem ainda de qualificação para entendermos o que circunscreveria o domínio do dispositivo. Em certa medida, inclusive, não haveria nesta primeira aproximação grande diferença entre aquilo que já indicamos, de passagem, ser a proposição de Bruno Latour acerca do ator-rede e o que se desenha, aqui, em torno do dispositivo. Afinal, a figura da rede é retomada por Latour não apenas para indicar a impossibilidade de isolarmos a origem de uma ação, como também para dar conta da articulação tanto do linguístico como do não-linguístico. Criticando a limitação da análise social apenas aos componentes simbólicos e discursivos, Latour ironiza, em certa passagem: “Quando uma bicicleta se choca contra uma

---

50 A metáfora, em particular no âmbito da Teoria Ator-Rede, é objeto de discussão e crítica de Tim Ingold (2012) em sua proposição da *malha* enquanto figura que mais adequadamente ilustraria tais relações, enfatizando a existência de um ambiente “indiscretizável” e que, portanto, não poderia ser decomposto em nós ou agências de uma rede.

51 Não recusamos, por tal proposição, a compreensão de que o próprio dizer já compreende em si uma ação, e em vários níveis, como inclusive discutem Deleuze e Guattari (1995b) acerca da performatividade da linguagem. Buscamos enfatizar, contudo, os modos de ação não mediados pela linguagem.

pedra, não é social. Mas quando um ciclista cruza uma placa de ‘pare’, torna-se social”<sup>52</sup> (LATOURE, 2005, p. 83). O que ele defende, então, e que o aproximará (em termos) da abordagem construída por Foucault<sup>53</sup>, é a necessidade de considerarmos não apenas a dimensão simbólica, discursiva e, em última medida, humana, mas também a agência dos não-humanos, que incluem tudo aquilo que seria desconsiderado em uma separação natureza/cultura e, também, os objetos técnicos, considerados não apenas em sua dimensão de expressão social e cultural, mas como agentes sociais plenos (cf. LATOUR, 2005, p. 63-86; e 2001). Na esteira desta formulação, ele sugere, em um exemplo já clássico (LATOURE, 2001, p. 203-208), que não poderíamos regredir à origem do disparo de um revólver isolando seja o atirador, seja a arma: “Você se torna outro sujeito porque segura a arma; a arma se torna outro objeto porque entrou numa relação com você. O revólver não é mais o revólver-no-armário [...] e sim o revólver-em-sua-mão, apontado para alguém que grita apavorado” (LATOURE, 2001, p. 207). Seu olhar se atenta, assim, aos acoplamentos, aos múltiplos agenciamentos que desencadeiam determinada ação, algo que parece se aproximar da proposição do dispositivo, exceto por uma circunscrição que nos é oferecida pelo próprio Latour. No que parece ser uma resposta a um questionamento hipotético à sua proposição da agência dos objetos, o autor matiza sua definição ao negar aos não-humanos a intencionalidade da ação (sem, contudo, deixar de fazer o mesmo para os humanos):

A ação intencional e a intencionalidade talvez não sejam propriedades de objetos; contudo, também não são propriedades de humanos. São propriedades de instituições, de aparatos, daquilo que Foucault chama de *dispositifs*. Somente pessoas jurídicas estão aptas a absorver a proliferação de mediadores, a regular sua expressão, a redistribuir habilidades, a forçar caixas a obscurecer e fechar-se. Objetos que existem simplesmente como objetos, apartados de uma vida coletiva, são desconhecidos, estão sepultados. [...] Os artefatos reais são sempre parte de instituições, hesitantes em sua condição mista de mediadores, a mobilizar terras e povos remotos, prontos a transformar-se em pessoas ou coisas, sem saber se são compostos de um ou de muitos, de uma caixa-preta equivalente a uma unidade ou de um labirinto que oculta multiplicidades (MacKenzie, 1990). Os Boeings 747 não voam, voam as linhas aéreas. (LATOURE, 2001, p. 221).

Trata-se de uma formulação elucidativa. Quando se atém às agências humanas e não-humanas (explicitando a participação destas) Latour explora um campo de indeterminação: não há intencionalidade porque não haveria preponderância de nenhum mediador sobre os demais. A ação seria, assim, decorrente dos agenciamentos mútuos destes diversos elementos e, assim, imprevisível. O dispositivo, de outro modo, indicaria arranjos elaborados estrategicamente para absorver essa proliferação e conduzir este arranjo em direção a determinados fins, a determinadas ações. Trata-

---

52 No original: “*When a bicycle hits a rock, this is not social. But when a cyclist crosses a ‘stop’ sign, it becomes social*”. Tradução nossa.

53 Foucault parece ser um ilustre omitido das referências de Latour. As similaridades entre suas proposições são dignas de nota assim como é a ausência de Foucault, de modo mais pronunciado, em suas menções. Talvez isto se deva a uma recusa geral desse autor a qualquer herança de matriz marxista, o que seria no mínimo injusto. Em todo caso, destaca-se a elogiosa nota de rodapé que Latour lhe dedica bem ao final do capítulo em que discute a agência dos objetos em *Reassembling the social* (2005, p. 86). Nela, Latour diz que ninguém como Foucault teria sido mais crítico das *explicações sociais*, em sua decomposição do poder, mas que seus textos teriam sido incompreendidos.

se, neste sentido, de um arranjo orientado à diminuição da incerteza, à redução do horizonte de probabilidades. O autor sugere que tais arranjos, os dispositivos, seriam da ordem das instituições e das pessoas jurídicas, o que não precisaríamos compreender em seu sentido literal, mas que talvez se caracterize, principalmente, como um arranjo abstrato que não se localiza neste ou naquele objeto ou pessoa – físicos, concretos – mas na relação estratégica entre vários deles. Nesse sentido, um Boeing 747 não voaria porque não depende apenas dele fazê-lo. Tampouco bastariam ele e o piloto: há toda uma complexa rede de agentes, orientados segundo dispositivos de administração e eficiência que permitem que uma tal aeronave decole de Guarulhos às 12 horas e 5 minutos. Há que se considerar o abastecimento de combustíveis, a gestão de bagagem e passageiros, os sistemas de venda de passagem, os controladores de voo, a administração das rotas aéreas, as barreiras securitárias. Enfim, talvez precisássemos chegar ao nível da operação de um detector de metais para compreendermos o que faz com que voe um Boeing. Apenas dispositivos seriam capazes de gerir esta ampla rede de envolvidos.

A genealogia do conceito feita por Agamben (2009) aproxima-se bastante desta compreensão pelo estudo etimológico que conectaria o *dispositivo*, por meio de sua raiz latina *dispositio*, ao grego *oikonomia*. Esquivando-nos de enfrentar o sentido teológico da reflexão do autor, indiquemos apenas que, neste âmbito, a *oikonomia* nomeia a função de governo exercida pela divindade sobre o mundo das criaturas (AGAMBEN, 2009, p. 38). Nesse sentido, ele sugere que assim se indicaria uma cisão fundamental e, ao mesmo tempo, uma articulação entre: “a natureza ou essência e a operação por meio da qual se realiza uma pura atividade de governo sem nenhum fundamento no ser” (AGAMBEN, 2009, p. 38). Derivando, deste sentido originário, para nosso atual sentido de *economia*, poderíamos dizer então que o dispositivo nomearia esta instância de gestão, de economia das agências, de modo similar às atividades de absorção, regulação e redistribuição que lhe atribui Latour (2001, p. 221). Diferenciando estes autores, contudo, cabe destacarmos a deliberada expansão do sentido foucaultiano do termo, por Agamben, para incluir sob tal denominação “qualquer coisa que tenha de algum modo a capacidade de capturar, orientar, determinar, interceptar, modelar, controlar e assegurar os gestos, as condutas, as opiniões e os discursos dos seres viventes” (AGAMBEN, 2009, p. 40). Trata-se de formulação que contrasta, em alguma medida, com a de Latour, para o qual não cabem sob o dispositivo quaisquer coisas com tais propriedades, mas apenas associações abstratas, em certo nível institucionais, que indica como *pessoas jurídicas* – um objeto, por si, não conteria tais possibilidades. Se os autores se aproximam na compreensão do dispositivo enquanto instância de gestão, se distanciam quanto ao que poderíamos situar nesse lugar – Agamben, por exemplo, sugere o telefone celular como um dispositivo, enquanto Latour provavelmente acenaria a um arranjo mais amplo como, por exemplo, às companhias telefônicas. De todo modo, talvez pudéssemos conceber a diferença entre eles como uma questão de escala dos recortes preferidos pelos autores: ainda que dependente das companhias telefônicas, haveria já no próprio telefone celular não apenas uma complexa heterogeneidade de elementos como várias estratégias de regulação de seu funcionamento,



de absorção das agências. De certo modo, vemos incidir, também nestes exemplos, o caráter fractal que Guattari (2003) atribui aos agenciamentos maquínicos.

Retornando à Visão Computacional, a formulação do conceito de *dispositivo* nos sugere que, ainda que suas operações internas desempenhem um papel significativo nos agenciamentos de que participam, seria insuficiente observarmos apenas esta instância. Os algoritmos, com sua relação particular com o visual, seriam, neste sentido, apenas uma das agências envolvidas em uma rede que se estende para além deste nível mais técnico, conectando, ainda, agentes humanos, discursos, instituições. As máquinas que veem são, neste sentido, objetos técnicos, mediadores, mas não, por si só, dispositivos. A Visão Computacional, contudo, nomeia mais do que um conjunto de algoritmos, circunscrevendo também um campo de pesquisa e aplicação tecnológica que, neste sentido, estende-se a variados contextos de aplicação e relações institucionais. Tomando-a enquanto um *dispositivo*, nomeamos pela Visão Computacional um campo em que se cruzam, por exemplo: abordagens analíticas da imagem e do visual; a aceleração das dinâmicas cotidianas de produção e circulação de imagens, com a proliferação de câmeras nos mais diversos objetos; e a ampla disponibilidade de acervos visuais em rede, como resultado desta produção cotidiana e, também, dos esforços de digitalização de imagens de períodos anteriores. Nesse contexto, a Visão Computacional parece responder a uma necessidade, proclamada em nossa relação atual com a informação e pelas promessas do chamado *Big Data*, de que as imagens sejam plenamente inseridas no atual regime informacional, pelo que indicamos o contexto em que a mineração e o processamento de dados se colocam enquanto estratégias fundamentais para a produção de conhecimento e para a organização de cadeias de decisão na contemporaneidade. Tomando a estatística como fundamento metodológico, vemos, hoje, a ampla disponibilidade de dados – resultantes de nossas interações online, de nossos padrões de acesso e consumo de conteúdos e de nosso uso de *softwares* e *gadgets* – se traduzir em um esforço constante e crescente de dar sentido a este volume de informações, tanto pelas empresas provedoras destes serviços e por serviços de inteligência policial quanto pelas próprias ciências que vêm se valendo – ainda que criticamente, em alguns casos – de algumas destas mesmas metodologias. A meio caminho entre a vigilância oculta e compulsória e a exibição voluntariosa e consciente de nós mesmos, a coleta destes dados se baseia, em grande parte, de nossa adesão consentida – ainda que inadvertida – aos termos de uso de muitos dos serviços “gratuitos” de que hoje fazemos uso. Em todo caso, não se trata de um domínio apenas de controle ou de exercício do poder mas, fundamentalmente, de um dos modos pelos quais, hoje, produzimos conhecimentosobre o mundo.

Neste sentido, parece se explicitar um dos aspectos inerentes do conceito de *dispositivo* segundo Foucault (1979, p. 246) – também destacado por Fernanda Bruno (2013, p. 21) – que é sua articulação entre relações de poder e formações de saber. O exercício do poder pela via informacional caminha lado a lado, portanto, com formas de produção de conhecimento pela via estatística, oriundas da mineração e do processamento de dados na rede. Poderíamos mesmo dizer que tanto nos serviços de

inteligência policial quanto no âmbito da ciência dos dados (*data science*) encontramos uma ciência probabilística na qual a matriz ontológica seria aquela do perfil de risco. Abordando especificamente as configurações contemporâneas da vigilância, Fernanda Bruno (2013, p. 49-50) identifica este mesmo contexto de proeminência da coleta e processamento de informações como uma de três vias atuais de sua legitimação. A primeira indicada pela autora seria aquela da segurança, neste sentido mais relacionada ao domínio mais tradicional da vigilância enquanto tecnologia para o controle e a minimização de riscos. A segunda seria aquela da visibilidade, mais afeita ao domínio do espetáculo e que se articula contemporaneamente às práticas de exibição da intimidade pelos próprios indivíduos, constituindo um modo de vigilância alimentado pelo desejo de *ver e ser visto*. A terceira seria a que a autora sugere como um *regime de eficiência informacional*, pelo que nomeia a prática disseminada de coleta e o monitoramento de informações de usuários de serviços *online* – operações que nos são apresentadas como necessárias para o bom funcionamento destes serviços. Entre os exemplos deste caso, há desde os algoritmos empregados por companhias como *Netflix* e *Amazon*, na indicação de filmes e compras, até os filtros de interesse de que se valem redes sociais como o *Facebook* na composição da *timeline* de seus usuários ou os serviços de publicidade direcionada oferecidos pela *Google*. Embora a Visão Computacional se relacione às três vias indicadas pela autora, e embora se trate de uma tecnologia com especial relevância para a discussão da vigilância, destacamos esta terceira via por entender que ela nos permite uma compreensão mais abrangente deste dispositivo, uma vez que neste contexto podemos compreendê-lo como um forte mediador que hoje conectaria as lógicas da vigilância e do espetáculo a esta da eficiência informacional. Evidentemente, tratamos de um domínio tecnológico que tem trazido transformações fundamentais à vigilância como, por exemplo, com o emprego de algoritmos de interpretação automatizada de imagens em aplicações diversas da videovigilância ou da vigilância distribuída. Também, para o caso da espetacularização da intimidade, trata-se de uma tecnologia de fundamental importância para as atuais ferramentas de busca e recuperação das imagens, potencialmente agregando, por exemplo, todas as aparições de determinada pessoa em fotografias no *Facebook*. Contudo, talvez seja principalmente pela aproximação de tais domínios aos aspectos que identificaríamos como característicos de um regime de eficiência informacional que conseguiríamos compreender estas transformações. A Visão Computacional é, então, uma importante mediação entre os modos de recuperação e processamento da informação e o domínio do visual. Neste sentido, um tal regime assume, em nosso estudo, a posição de domínio privilegiado em que a Visão Computacional ganha forma e sentido.

Bruno (2013, p. 146) indica como, em um paralelo com os dispositivos disciplinares modernos, assistimos na contemporaneidade à emergência de novas formas de registro e coleta de informações que se baseiam, fundamentalmente, no monitoramento de dados e rastros pessoais em circulação no ciberespaço. O paralelo reside em certa analogia entre essas estratégias de registro e as que seriam próprias do regime disciplinar, em fábricas, hospitais, escolas e prisões, que fariam, de outro modo, o registro escrito em fichas individuais, prontuários, históricos. Mais do que estratégias de monitoramento e exercício do poder, contudo, a coleta de dados na internet configura, como

já indicamos, uma formação contemporânea do saber, que se processa, com frequência, pela via estatística e que, assim, ganha força com a ampla disponibilidade de informações constantemente produzidas e compartilhadas na rede. Não raramente, inclusive, vemos hoje serem divulgadas na mídia estatísticas de menções de determinadas *tags* em redes sociais como indicadores da popularidade de determinado assunto. De modo similar, centros de pesquisa voltam seus olhares para o conteúdo disponível na rede. Há, é claro, grande proximidade, em termos metodológicos e epistemológicos, entre tais esforços e a vigilância exercida pelos Estados Unidos por meio sua Agência de Segurança Nacional, como recentemente denunciado por Edward Snowden, ou mesmo os esforços anunciados pelo Ministério da Defesa do Brasil de remodelagem do Centro de Inteligência do Exército para o monitoramento ativo de movimentos sociais na internet (Cf. MONTEIRO, 2014; e MAZZINI, 2014). Diante de tais empreendimentos e da crescente participação da imagem no âmbito das interações em rede, a superação de sua *opacidade informacional* por meio de algoritmos de Visão Computacional ganha, então, particular relevância. Também se destacam neste contexto os esforços realizados no entorno de pesquisas em cultura visual na rede que se valem de metodologias de análise computacional de imagens, tais como as que vêm sendo empreendidas por Lev Manovich, na tentativa de dar sentido a uma enorme massa de fotografias publicadas em redes sociais na internet<sup>54</sup>.

Como já indicamos, o que conectaria os distintos modos de operação da Visão Computacional, em seus variados contextos de aplicação, seria sua operação enquanto mediadora entre o visível e o informacional e computável. Tanto nos casos em que se sobressaem operações que agrupamos sob o par *localização–acionamento* quanto nos que agrupamos sob *reconhecimento–conexão*, seu funcionamento direciona-se fundamentalmente à extração de dados das imagens, traduzindo-as a valores numéricos e chaves semânticas que descreveriam seu conteúdo. Em última medida, a Visão Computacional permite que se insira a imagem no regime de eficiência informacional de que fala Fernanda Bruno. O teórico da mídia Sean Cubitt (2011, p. 9) sugere que, em nosso tempo, as mídias dominantes (entendidas como aquelas utilizadas para o exercício do poder) já não seriam mais a narrativa ou a imagem, mas aquelas que indica como os três pilares da economia política: planilhas, bancos de dados e sistemas de informação geográfica (GIS). Seriam estas que, em última medida, serviriam aos propósitos contemporâneos de elaboração de perfis estatísticos e de controle e gerenciamento de riscos. Poderíamos compreender a Visão Computacional, neste sentido, como um elemento de passagem responsável pela tradução entre dois estágios ou dois paradigmas das tecnologias de controle e conhecimento, pois é por meio delas que a imagem é levada a participar das dinâmicas destas mídias que seriam hoje galgadas à posição, destacada por Cubitt, de mídias dominantes.

De um lado, há evidentemente uma forte reconfiguração no papel desempenhado pela imagem, que, passível de ser traduzida automaticamente em informação, desdobra-se enquanto uma fonte

---

54 Cf. <http://selfiecity.net/> e MANOVICH, DOUGLAS e ZEPPEL, 2011.

de dados estatísticos e deixa de ser, neste sentido, um objeto de difícil tratamento, indexação ou arquivamento. De outro, há também um ganho significativo para as mídias identificadas por Cubitt, com a entrada da imagem em seu domínio. Operando tradicionalmente por meio de dados numéricos e abstratos, a elas lhes é permitido, a partir da imagem, circunscrever e singularizar a informação – indo do plano geral a territórios, objetos e sujeitos singulares. De outro modo, também lhes torna possível fazer o caminho inverso e generalizar a partir destas mesmas instâncias singulares, na medida que é pelo enfrentamento de múltiplos registros do individual fotográfico que o método do *aprendizado de máquinas* permite aos algoritmos da Visão Computacional realizar o *reconhecimento* de padrões e a sua *conexão* a instâncias relacionadas. Assim, os pares *localização–acionamento* e *reconhecimento–conexão* sugerem algo que talvez possamos compreender como um cruzamento entre a lógica da estatística populacional, que é o que realiza, tipicamente, o regime de eficiência informacional, calcado na lógica do *controle*; e a da circunscrição do espaço e dos corpos que é aquela da fotografia de identificação ou da imagem de vigilância, dispositivos típicos da lógica *disciplinar*. Um modo de compreender nosso dispositivo, neste sentido, o coloca como uma espécie de mediação entre estes dois modelos de exercício do poder e de produção do saber, permitindo ir do populacional ao individual, dos aspectos gerais aos singulares, em um mesmo procedimento<sup>55</sup>.

Tanto Fernanda Bruno (2012 e 2013) quanto Cubitt (2011) parecem concordar, ainda, que um dos aspectos relevantes desta passagem da imagem um regime informacional é sua temporalidade orientada à antecipação e à limitação das incertezas – portanto, em uma tentativa de anulação do futuro, ou do próprio tempo. Bruno indica, a este respeito, que o ingresso das chamadas *câmeras inteligentes* à videovigilância, teria trazido uma mudança substancial pela qual a imagem seria encarada, com frequência, como um “índice de um acontecimento por vir” (BRUNO, 2012, p. 58). Os algoritmos empregados neste contexto, afinal – em particular no âmbito das operações de *reconhecimento–conexão* – mobilizam um determinado repertório de padrões para identificar, na imagem captada em tempo real, índices do risco de desenvolvimento de um acontecimento indesejável. O caráter de registro, portanto, geralmente destacado na compreensão do *digital*, é levado a um plano secundário. A memória será tão mais valiosa quanto subsidiar interpretações do presente para limitar as possibilidades de acontecimentos futuros: “Os arquivos gigantescos e humanamente insondáveis da videovigilância tornam-se, segundo os propositores deste sistema, uma memória útil” (BRUNO, 2012, p. 58). Esta operação, identificada pela autora para o caso específico da videovigilância, parece se estender a toda a lógica inerente do regime informacional que viemos discutindo. Cubitt (2011), de modo similar, irá sugerir que tal seria o traço característico da economia política de nosso tempo, amparada, em larga medida, pelas mídias que indica como protagonistas deste processo. Traduzindo as imagens a este domínio, tornando-as computáveis e,

---

55 Acerca do regime disciplinar e seus dispositivos, cf. FOUCAULT, 1999. Acerca da formulação de Deleuze da “sociedade de controle”, descrevendo, a partir de Foucault, uma nova ordem do exercício do poder na sociedade da informação, cf. DELEUZE, 1992: “Controle e devir”, p. 209-218; e, especialmente, “*Post-scriptum* sobre as sociedades de controle”, p. 219-226.

em certa medida, probabilísticas, a Visão Computacional parece ser alçada, hoje, a uma posição de destaque.

\*

Elaborando uma compreensão própria daquilo que poderíamos considerar, ou não, um *dispositivo*, Deleuze (2005) constrói sua leitura deste conceito em Foucault a partir das dinâmicas que os dispositivos estabeleceriam em sua relação com *diagramas*, os quais também denomina *máquinas abstratas*. Nos aproximamos, nesta sua formulação, da definição de *agenciamentos maquinicos* que indicamos anteriormente em Guattari (2003). Para Deleuze (2005), os dispositivos seriam, mais especialmente, *agenciamentos concretos*, aproximando-se, neste sentido, mais ao próprio edifício da prisão do que à instituição prisional, como talvez sugeriria Latour (2001). Contudo, longe de agirem sozinhos, eles seriam mobilizados por um diagrama, ou máquina abstrata, que consistiria, basicamente, de uma fórmula, de uma pura função que abstrairíamos das formas em que se encarnam (DELEUZE, 2005, p. 43). Se, por um lado, poderíamos descrever o dispositivo da prisão panóptica pelo regime de visibilidade a que seriam submetidos os prisioneiros – por seu “agenciamento óptico ou luminoso” (DELEUZE, 2005, p. 43) –, este agenciamento concreto (composto pelo edifício da prisão, seu regulamento, sua organização institucional) seria a encarnação de um diagrama disciplinar que se estenderia para além do regime prisional, para o hospital, a escola, o quartel e a fábrica. Pois o diagrama compreenderia justamente a abstração de uma função pura que se materializaria e se operacionalizaria nos agenciamentos concretos, progressivamente, em vários deles. Contudo, não haveria equivalência entre esta abstração e, por exemplo, um modelo ou ideal transcendente que aos agenciamentos concretos seria anterior – e dos quais estes seriam apenas manifestações imperfeitas. O diagrama seria, pelo contrário, para Deleuze, uma *causa imanente*:

O que quer dizer, aqui, causa imanente? É uma causa que se atualiza em seu efeito, que se integra em seu efeito, que se diferencia em seu efeito. Ou melhor, a causa imanente é aquela cujo efeito a atualiza, integra e diferencia. Por isso nela há correlação recíproca entre a causa e o efeito, entre a máquina abstrata e os agenciamentos concretos (é a esses que Foucault reserva mais frequentemente o nome de “dispositivos”). [...] Mas a atualização também é uma integração, um conjunto de integrações progressivas, a princípio locais, depois globais ou tendendo a serem globais, operando um alinhamento, uma homogeneização, uma soma de relações de força... (DELEUZE, 2005, p. 46-47).

Deleuze descreve, portanto, entre o diagrama e os dispositivos, uma relação que subverte a lógica determinística de causa e efeito ao conceber o dispositivo como cindido pelo diagrama, que nele se inscreve e se entranha. Maria Cristina Franco Ferraz (2013, p. 168) oferece uma formulação elucidativa deste movimento ao situar o dispositivo como, simultaneamente, *instrumento* das transformações sociais de que também é a *expressão*. Seria pela introdução desta instância, sugere Deleuze, que Foucault conseguiria explicar a coadaptação entre prisão e direito penal, mas também seria através dela que explicaria o potencial de alastramento de uma atualização local do diagrama a uma ampla rede de agenciamentos concretos articulados, conectados. A máquina abstrata, a causa

imane de um dispositivo, neste sentido, seria aquilo que se abstrai da tomada em conjunto desta rede mais ampla de agenciamentos concretos.

A partir da dinâmica entre o dispositivo e o diagrama, Deleuze sugere uma diferenciação fundamental pela qual – diferentemente de Latour (2001) e de Agamben (2009), que parecem expandir seu sentido – os dispositivos ocupariam uma posição sempre privilegiada em relação a outras formações tecnológicas e arquitetônicas:

A tecnologia é então social antes de ser técnica. [...] se as técnicas, no sentido estrito da palavra, são tomadas nos agenciamentos, é porque os próprios agenciamentos, com suas técnicas, são selecionados pelo diagrama: por exemplo, a prisão pode ter existência apenas marginal na sociedade de soberania (as *lettres de cachet*), ela só existe como dispositivo quando um novo diagrama disciplinar a faz ultrapassar “o limiar tecnológico” (DELEUZE, 2005, p. 49).

Nomearíamos um dispositivo, portanto, no sentido que lhe atribui Deleuze, não qualquer coisa com a capacidade de capturar, como sugere Agamben (2009, p. 40-41), ou qualquer formação institucional ou *pessoa jurídica*, como parece sugerir Latour (2001, p. 221), mas, de modo mais especial, aqueles agenciamentos que seriam selecionados pelo diagrama de sua época. Gostaríamos de destacar, contudo, que não temos a pretensão de defender, por exemplo, que o dispositivo da Visão Computacional poderia ser um destes agenciamentos selecionados pelo diagrama contemporâneo, como quer o quiséssemos nomear – diagrama de controle ou de eficiência informacional, por exemplo. Seria-nos satisfatório, para o momento, que pudéssemos compreendê-lo como um dispositivo ou um agenciamento menor, desde que esteja clara a articulação que viemos construindo entre seu modo de operação e as demandas institucionais e políticas, ou as formações de saber, que contemporaneamente reconhecemos.

De todo modo, talvez em uma leve inflexão do sentido pretendido por Deleuze, gostaríamos de encerrar este capítulo sugerindo, sim, algo que se aproxima de uma seleção da Visão Computacional por um diagrama contemporâneo, mas não tanto como um dispositivo de existência central e preponderante em nossa sociedade mas como algo que poderíamos indicar como um *dispositivo paradigmático*. Isto, talvez, justificaria nosso interesse particular em estudá-lo. Agamben (2002) sugere, em uma reflexão acerca do valor epistemológico e metodológico do *paradigma*, certa aproximação entre esta figura do pensamento e o conceito de dispositivo. O filósofo indica possuir valor paradigmático o uso que Foucault faz da prisão panóptica na passagem a um *panoptismo*, pela qual a configuração arquitetônica da prisão, com seu regime de luz e visibilidade, se aplicaria a uma série de outros agenciamentos da época, como já indicamos ser o caso do hospital, da escola, do quartel e da fábrica. Em suma, para o filósofo italiano, o paradigma seria um recurso intelectual que se aproxima do *exemplo*: um caso singular que, por ser mais inteligível ou “conhecível”, lançaria luz sobre si e sobre outros casos igualmente singulares (AGAMBEN, 2002). Logo, se podemos compreender o caráter paradigmático do panóptico na análise de Foucault, é pelo modo com que confere maior inteligibilidade a outras configurações que a ele se assemelham, evidenciando as

ressonâncias, em cada caso, de um regime que se desenha, naquele momento, em torno da vigilância e da disciplina. A Visão Computacional, de modo análogo, seria para nós um agenciamento exemplar ou paradigmático que nos permitiria compreender processos díspares e dispersos relacionados à cultura visual contemporânea, bem como aos mecanismos atuais de controle, e desenhando, em certa medida, aspectos de nossos modos de percepção, de nossas visualidades.

Parece-nos claro, mas vale explicitar, que estamos longe, nesta formulação, de situar este domínio tecnológico como *determinante* de nosso contexto. Diferentemente, o concebemos, ao modo de Deleuze, como necessariamente “fendidos pelo interstício através do qual se efetua a máquina abstrata” (DELEUZE, 2005, p. 47-48). Trata-se, pois, de outro modo de dizer aquilo que nos propomos, ao início deste capítulo, de não tomar a Visão Computacional por contornos claramente definidos, como se sua individualidade, enquanto máquina que vê, preexistisse à sua inscrição fundante nos coletivos sociotécnicos de que emerge. Neste sentido, tratamos não de uma tecnologia que seria determinante pelo contexto de eficiência informacional que descrevemos, mas enquanto, simultaneamente, instrumento e expressão deste regime (FERRAZ, 2013, p. 168). Seria impossível, por uma tal compreensão, chegarmos a relações simplistas de *causa e efeito*.

Em *Técnicas do observador*, o historiador da arte Jonathan Crary (1992), embora não se proponha elaborar uma definição do *dispositivo*, parte do referencial teórico de Foucault e Deleuze para abordar artefatos ópticos a partir dos quais busca descrever uma transformação no tratamento dado à visão que teria ocorrido no início do século XIX. O modo como concebe estes artefatos, em particular pela proximidade temática de seu estudo, ganha especial relevância para nós, informando, em larga medida, o modo como tomamos a Visão Computacional:

Os dispositivos ópticos em questão, de modo mais importante, são pontos de interseção em que discursos filosóficos, científicos e estéticos se sobrepõem com técnicas mecânicas, demandas institucionais e forças socioeconômicas. Cada um deles é compreensível não apenas como o objeto material em questão, ou como parte de uma história da tecnologia, mas pela forma com que se insere em um arranjo bem mais amplo de eventos e poderes<sup>56</sup> (CRARY, 1992, p. 8).

Assumindo para nosso objeto a mesma imagem indicada por Crary – que não por acaso se aproxima da metáfora da rede – tomamos a Visão Computacional como um destes pontos de interseção, enquanto um agenciamento que se oferece como ponto de partida do qual se ramificam questões de pesquisa acerca da nossa relação contemporânea com o visual. Trata-se de um terreno de trânsito pelo qual atravessam: formações estéticas da produção e da circulação de imagens em rede; discursos da ciência e da tecnologia; demandas institucionais relacionadas à vigilância, à automação industrial

---

56 No original: “*The optical devices in question, most significantly, are points of intersection where philosophical, scientific, and aesthetic discourses overlap with mechanical techniques, institutional requirements, and socioeconomic forces. Each of them is understandable not simply as the material object in question, or as part of a history of technology, but for the way in which it is embedded in a much larger assemblage of events and powers*”. Tradução nossa.

e à gestão da informação; e crenças e narrativas – ficcionais ou não – que em alguma medida orientam nossa relação com a visão, com a imagem e com as máquinas.

A perspectiva de Crary parece, assim, arrematar a formulação teórica pela qual viemos abordando, neste capítulo, as máquinas que veem. Se nos é possível relacionar um nível mais técnico de suas operações aos seus modos de agenciamento externo e, em um plano mais geral, a suas articulações com estes componentes *macro* – discursivos, institucionais e estéticos – isto se deve, principalmente, à inexistência de um limite claro a separar tais instâncias. Neste sentido, não é que estejamos construindo tais relações, mas apenas esforçando-nos por reconhecê-las e identificá-las. De modo análogo à figura das esferas vazadas e sustentadas pela rede, que Latour (2012) busca em Saraceno, olharmos a Visão Computacional como este ponto de interseção implica reconhecermos que, entre as linhas que circunscrevem este domínio tecnológico e as que o sustentam, não há ruptura real. A tarefa da análise, neste sentido, é tornar visíveis estes percursos pelos quais a máquina se abre ao exterior, sem apenas agir sobre ele. Somente assim seríamos capazes de compreendê-la.



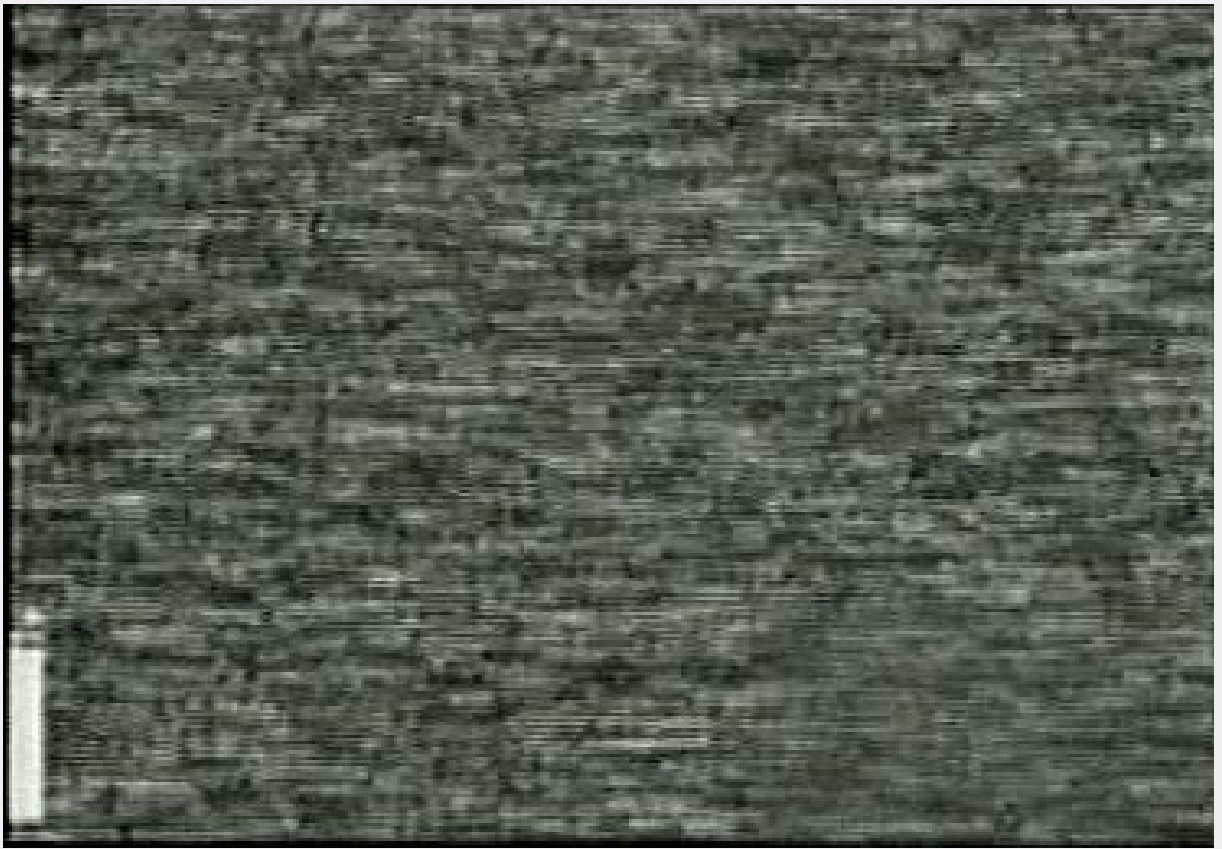


Figura 15 – Sinal perdido das “câmeras suicidas” em *frame* de um dos canais da videoinstalação *Eye/Machine* (2001), de Harun Farocki. Fonte: EYE/MACHINE, 2001.



## 2 O que veem as máquinas

### 2.1 Câmeras suicidas, imagens evanescentes

Em sua série de videoinstalações *Eye/Machine* (2001-2003)<sup>57</sup>, Harun Farocki explora as imagens produzidas por máquinas autônomas de visão, em particular no contexto da guerra e da indústria. Dando continuidade à sua pesquisa em torno dos arquivos e das imagens instrumentais, que apresenta em outras instalações e filmes, o cineasta apresenta-nos registros diversos realizados por algumas das máquinas concernentes ao nosso objeto de estudo. Capturadas por robôs para se orientarem no espaço; alimentadas a protótipos de carros que dirigem a si próprios; produzidas com o objetivo de reconhecer e rastrear alvos; de automatizar não apenas o *trabalho manual*, como também o *trabalho visual*. Articulam-se, em sua montagem, registros dispersos nos contextos da tecnologia militar e industrial, permeados por austeros letreiros que oferecem algo como um complemento lacunar à em geral silenciosa alternância de imagens – a estas, em sua maioria, apenas acompanham o próprio ruído do funcionamento das máquinas, capturado em som direto. Partimos neste capítulo desta referência fundamental para abordarmos um dos aspectos que consideramos centrais a uma abordagem das máquinas de visão contemporâneas: o modo como agenciam, em seu interior o visível, as imagens, ou, posto de outro modo, aquilo que veem.

Parece-nos emblemática uma sequência que abre a primeira instalação, de 2001, em que nos são apresentados registros documentais da guerra contra o Iraque, de 1991 – a Guerra do Golfo – em que observamos em imagens monocromáticas e nebulosas, a partir de um ponto de vista aéreo, edificações sendo bombardeadas. No centro, linhas sobrepostas formam uma cruz, indicando o alvo. Outras informações, ilegíveis, ocupam regiões periféricas do quadro. A imagem serve claramente a um propósito operacional específico: auxiliar a mira para o tiro e verificar se o alvo foi atingido. Somos apresentados a uma série destes disparos mas, em meio a eles, logo encontramos um registro distinto do ponto de vista de sobrevoo, ainda que cumprindo objetivo similar. Já não nos restringimos ao ponto de vista do avião embora partamos, ainda, de uma visão do alto, antes de iniciarmos uma rápida descida. As câmeras, acopladas às próprias bombas, apresentam-nos um ponto de vista em movimento: aproximando-nos cada vez mais do solo, em queda livre. Chegamos bem próximos dos alvos – da ponte, da edificação, da estrada – quando há um corte dando lugar ao ruído da estática: perdemos o sinal (Figura 15 e Figura 16). Como sugere o cineasta nos letreiros

---

57 A série é composta de três obras: *Eye/Machine* (2001), *Eye/Machine II* (2002) e *Eye/Machine III* (2003). Elas são apresentadas em dois canais de vídeo sincronizados, lado a lado. Foram consultadas para esta pesquisa versões adaptadas destes trabalhos para a exibição em monocanal, em que os dois quadros são exibidos simultaneamente em uma mesma tela, em uma composição plana distinta da que geralmente se apresenta este trabalho.

exibidos nesta sequência, estamos diante das imagens de “câmeras suicidas”. Como as imagens remotas de disparos, que as antecedem, também estas cumprem objetivos operacionais precisos, mas neste caso, uma vez atingidos tais objetivos (ou, os alvos), as câmeras são também destruídas e o fluxo de imagens interrompido – descartáveis.

Poderíamos dizer que ao explodirem os alvos, tais bombas realizariam também uma implosão do visível. São imagens – explicitam os letreiros – que não foram feitas para que nós as vissemos. São imagens operacionais (também diríamos instrumentais) que desempenham papel bem distinto daquelas a que mais comumente dirigimos nosso olhar, seja na televisão, no cinema, no jornalismo, na internet ou em fotografias e vídeos domésticos. São imagens sem plasticidade, sem referência na experiência cotidiana – nos dizem, ainda, os letreiros: são imagens que não foram feitas para serem vistas por nós, ao menos não qualquer um de nós. Prestam-se, ocasionalmente, aos técnicos – neste caso, da guerra, em outros, da indústria – para verificarem o bom funcionamento de suas máquinas. Contudo, muitas vezes já nem se requisita o olhar do técnico, exceto em caso de mal funcionamento: é à própria máquina que se prestariam muitas destas imagens. Ao míssil que se encaminha de modo autônomo ao alvo, ou à máquina que, na fábrica, encontra as peças, supervisiona a produção e interage com seu entorno. Por mais que ainda se faça ocasionalmente necessário que tais imagens cheguem ao olhar humano, é evidente que, como objetivo último, busca-se uma substituição deste pelo da máquina – já não mais a máquina de propósito único, lembram-nos os letreiros, mas máquinas de automação flexíveis, capazes de alterar seu curso de funcionamento em função do que percebem do entorno. Tal é, afinal, o claro propósito da Visão Computacional.

Neste estranho jogo de identificação pelo qual assumimos o olhar de bombas em queda, encontramos nas câmeras suicidas – *camerakazes* – aspectos emblemáticos que talvez estenderíamos, de um modo mais geral, às máquinas de visão. Afinal, abordamos uma tecnologia que, embora claramente inscrita no domínio das máquinas de imagem, ou das máquinas visuais, guarda em relação às demais um importante traço distintivo que é o caráter secundário – ou, na maioria dos casos, ausente – que tem a produção de imagens em seu funcionamento. A Visão Computacional antes alimenta-se do visível do que o produz: se toma como entrada fotografias e vídeos, o que oferece como saída são, em geral, dados numéricos ou textuais acerca do que teria sido percebido nestas imagens. Assim, talvez pudéssemos compreender seu funcionamento também sob a chave da implosão do visível que encontramos nas câmeras suicidas da Guerra do Golfo, em imagens que se prestam apenas enquanto instrumentos descartáveis após o uso. Como indicamos, inclusive com relação à incidência desta tecnologia na arte, um aspecto característico de seu funcionamento é o encerramento das imagens de que se alimenta: diferentemente das instalações em circuito fechado, por exemplo, não é tanto a imagem em si que interessa, senão as respostas a ela que a máquina produz. Em instalações interativas, em sistemas de vigilância e monitoramento e mesmo em interfaces baseadas em procedimentos da Visão Computacional, o registro que sustenta seu funcionamento quase nunca nos é apresentado em forma visível.

Paul Virilio (1994), indicando tal característica das máquinas de visão dirá inclusive que, neste domínio, a *imagem* talvez nem seja o termo adequado:

Não esqueçamos, contudo, que a ‘imagem’ é aqui nada mais do que uma palavra vã, uma vez que a interpretação da máquina nada tem a ver (é o caso de o dizer!) com a visão habitual. A imagem eletróptica é, para o computador, nada mais do que uma série de impulsos codificados dos quais não podemos nem mesmo imaginar a configuração, uma vez que, exatamente nesta ‘automação da percepção’, o *retorno-imagem não é mais assegurado* (VIRILIO, 1994, p. 103).

Não haveria correspondência, portanto, entre aquilo que chamamos de imagem e aquilo que a máquina vê – como inclusive sugerimos no capítulo anterior, ao descrevermos seu funcionamento. Sem retorno-imagem assegurado, sem que uma vez convertida a este estado “eletróptico” ela retorne ao visível, como mesmo abordá-la, neste domínio? Haveria, contudo, certo paralelo entre tal modo de percepção da máquina e o nosso próprio – ou menos assim acreditam alguns dos cientistas engajados em seu desenvolvimento. Virilio alerta quanto a isto logo na sequência, ao dizer que tampouco a visão ocular passaria “de uma série de impulsos luminosos e nervosos que nosso cérebro decodifica rapidamente” (1994, p. 103). Porém, ainda que possa ser sustentado, em vista das ciências cognitivas (em seu entrelaçamento com a Inteligência Artificial) que a imagem tampouco seria uma categoria da instância interna de nosso modo de ver, isto não diminui os efeitos de tal desenvolvimento, pelo contrário: se tais máquinas subsidiariam a compreensão da nossa própria percepção visual pela implosão do visível, como processamento e decodificação de sinais, mais profundas seriam suas consequências.

Os planos em queda apresentados em *Eye/Machine* são emblemáticos de nosso objeto, portanto, ao trazerem-nos algo de um ponto de vista em geral omitido ou ausente. Se o ponto de vista das máquinas de visão encontra-se em geral oculto por seu próprio funcionamento, demandando-se esforços como o de Farocki para que ganhem novamente o campo do visível, as imagens destas câmeras acopladas em bombas, como o próprio cineasta explicita, foram transmitidas pela televisão durante a guerra. Trata-se de um registro até então estranho à própria guerra: geralmente representada pelo ponto de vista do repórter ou da propaganda (tal era, até então, a *visualidade* do conflito). A Guerra do Golfo teria inaugurado – como nos sugere Manovich (1993a, p. 95), comentando a abordagem de Paul Virilio em outro texto<sup>58</sup> – um registro em que a guerra seria apresentada pelas próprias imagens *através das quais* era feita:

[...] o que nós também vimos não foram somente imagens da guerra, mas infundáveis imagens através das quais a guerra era conduzida: imagens de vídeo de uma câmera infravermelho montada em um avião; imagens de vídeo de uma câmera instalada em uma arma guiada por um sensor a *laser*; vídeo em seu papel de ‘avaliação de dano da batalha’, em que uma arma equipada com um dispositivo de imagem segue uma arma de destruição e grava detalhes do dano. Esta não era mais uma típica visão de repórter da batalha. Nós vimos o que os

---

58 VIRILIO, Paul. *War and cinema: the logistics of perception*. London: Verso, 1989. *apud* MANOVICH, 1993, p. 95.

soldados viam eles mesmos: as imagens que eram suas únicas informações acerca de seus alvos. Mais frequentemente, em um caso de estranha identificação, nós testemunhávamos o que era ‘visto’ por uma máquina, uma bomba, ou um míssil (MANOVICH, 1993a, p. 95)<sup>59</sup>.

As imagens da Guerra do Golfo talvez tenham oferecido a primeira oportunidade em que nos teria sido dado assumir, a partir de registros documentais, o ponto de vista da máquina. Uma das primeiras ocasiões – ainda raras – em que este ponto de vista teria alcançado nosso olhar. Ainda hoje, acessar tais registros e – se conseguimos acessá-los – compreendê-los pelo ponto de vista da máquina, ver o mundo por *uma visão computacional*, talvez ainda constitua um dos desafios em estudar tais objetos. Para além da dificuldade de termos acesso ao registro bruto das imagens alimentadas a esses programas – que em diversas aplicações nos é negado – ainda temos a de conseguirmos olhá-las pelo ponto de vista ao qual se endereçam. Se durante a Guerra do Golfo tal identificação era pautada pelo estranhamento, talvez assim seja ainda hoje, por mais que as máquinas de visão venham se tornado cada vez mais frequentes e presentes.

Pela operação primária destas máquinas, talvez pudéssemos ser levados a crer que a imagem seria um aspecto menor por se considerar no estudo da Visão Computacional. Se concordamos com o diagnóstico de Sean Cubitt (2011), quando ele sugere a preponderância das tecnologias orientadas à gestão da informação como as novas mídias dominantes de nosso tempo – em detrimento da imagem e da narrativa – talvez pudéssemos compreender a Visão Computacional como uma mídia de passagem que representaria, inclusive, uma transição a algo como um mundo sem imagens. Ou, pelo menos, a um mundo de imagens já despidas de sua relevância e de sua efetividade política. Contudo, talvez não seja o caso de abandonarmos a imagem de nosso quadro de análise. Especialmente diante do que nos apresenta Farocki, ao recolocar em circulação aquilo que seria ocultado ou descartado nos circuitos tecnológicos que abordamos, não há como negarmos que, no contexto da Visão Computacional e das máquinas de visão, a imagem seguiria desempenhando um papel fundamental, ainda que seja o caso de reconsideramos seu lugar. Defendemos, pois, que por mais que possamos descrever tais tecnologias pela via do encerramento do visível no informacional, é ainda importante observamos as imagens pelas quais elas operam, não apenas porque tais registros incorporam seu próprio modo de funcionamento (são, por assim dizer, peças destas máquinas), mas também porque as máquinas que abordamos integram processos contemporâneos mais amplos de reconfiguração do visível. Afinal, a Visão Computacional toma parte daquele que identificamos como um processo de aceleração das dinâmicas de produção e circulação de imagens, em particular na internet, não apenas como um dispositivo de indexação e organização como, também, deles

---

59 No original: “[...] what we also saw were not just images of the war, but endless images through which the war was carried out: video images from an infrared camera mounted on a plane; video images from a camera installed on a weapon guided by a laser sensor; video in its role as ‘battle damage assessment’ where a weapon equipped with an imaging device follows a weapon of destruction and records details of the damage. This was no longer a traditional reporter’s view of the battle. We saw what the soldiers themselves saw: the images that were their only information about their targets. More often, in a case of strange identification, we witnessed what was ‘seen’ by a machine, a bomb, or a missile”. Tradução nossa.

dependendo em sua própria constituição, pelos métodos de *aprendizado* empregados. A proliferação de agentes computacionais de visão parece, assim, tanto sustentar quanto ser sustentada por relações que hoje estabelecemos com a produção de imagens, uma vez que seriam componentes mutuamente dependentes para dar conta do que talvez pudéssemos indicar como uma *hipertrofia do visível*.

Ao articular estes dois movimentos, de produção e circulação de imagens e de sua interpretação automatizada por máquinas, o que desponta, contudo, é a natureza radicalmente ambígua deste processo, pelo qual apenas seria sustentável tal *hipertrofia do visível* quando passamos a delegar às máquinas boa parte da tarefa da visão, pois é a elas que se direciona – hoje e cada vez mais – grande parte das imagens que produzimos. Lev Manovich, em algumas de suas pesquisas atuais sobre a cultura visual, recorre à Visão Computacional para abarcar em sua análise quantidades massivas de imagens em circulação na *web* – como as fotografias postas em circulação, em um dia, na rede *Instagram*<sup>60</sup> – enunciando por vezes o problema técnico de sua pesquisa por “como comparar um milhão de imagens?”<sup>61</sup> (MANOVICH e DOUGLAS, 2011). Já dependemos desta tecnologia, também, para navegarmos pelos amplos arquivos visuais disponíveis na rede e mesmo para situarmo-nos diante de imagens que encontramos à nossa frente – aplicativos de celular como o *Google Goggles*, por exemplo (e, certamente, em funções nativas de dispositivos vestíveis como o *Google Glass*) já se prestam a buscar na rede informações acerca daquilo que vemos à nossa frente. O olhar da máquina, neste sentido, já se constitui enquanto um componente contemporâneo do visual seja em imagens produzidas especificamente no contexto da automação e da vigilância, seja naquelas endereçadas – ao menos a princípio – a nós.

Tal ambiguidade de sua relação com a imagem faz com que a Visão Computacional se aproxime da discussão proposta por Bruno Latour e Peter Weibel na exposição *Iconoclash: beyond the image wars in science, religion and art*<sup>62</sup>. Em texto de apresentação da exposição, publicado em seu catálogo, Latour (2002) explica o sentido pretendido pelo neologismo com que a intitulam. No *iconoclash*<sup>63</sup>, diferentemente do *iconoclasmo*, não há clareza se o ato de destruição de uma imagem, de seu apagamento, é de fato *destrutivo* ou *construtivo*. Interessado nas paixões mobilizadas pela imagem, como objeto tanto desejado quanto odiado, Latour sugere que nossa relação com ela seria mais ambígua do que poderíamos inicialmente supor e propõe que reflitamos acerca das razões pelas quais as imagens sempre retornam, por mais que sejam perseguidas e destruídas:

Portanto, podemos definir um *iconoclash* como aquilo que ocorre quando há incerteza a respeito do papel exato da mão que trabalha na produção de um mediador. É a mão com um martelo pronto para expor, denunciar, desbancar, desmascarar, mostrar, desapontar, desencantar, dissipar as ilusões de alguém, para

60 Cf. <http://selfcity.net/>.

61 No original: “How to compare one million images?”. Tradução nossa.

62 Realizada em 2002 no *Zentrum für Kunst und Medientechnologie* (ZKM) em Karlsruhe, Alemanha. Referimo-nos a ela a partir de seu catálogo (LATOURE e WEIBEL, 2002).

63 Trata-se de termo de difícil tradução para o português. Nicole Reis, que realiza tradução desse texto para sua publicação na revista *Horizontes Antropológicos* (LATOURE, 2008), sugere algo como *iconochoque*, porém opta por manter a palavra no original para preservar o jogo de sentidos pretendido pelo autor.

deixar o ar correr? Ou é, ao contrário, uma mão cautelosa e cuidadosa, com a palma virada como se fosse pegar, extrair, trazer à luz, saudar, gerar, entreter, manter, colher verdade e santidade? (LATOURET, 2008, p. 117-118)

Aproximamos a Visão Computacional desta definição por compreendermos que sua operação em relação às imagens não pode ser facilmente definida como destruidora, como iconoclasta – como talvez sugeriria a ênfase em sua operação primária. Embora, por seu funcionamento, a imagem seja levada a deixar o domínio do visível, sendo desfeita em sua constituição primeira, a *mão* do algoritmo perfaz um gesto ambíguo: busca destruí-la ou revelá-la? Haveria, afinal, um esforço dos algoritmos empregados por transpor a opacidade computacional da imagem, à máquina. Também, uma vez processada, a imagem passaria a ser mais facilmente indexável e, portanto, mais facilmente encontrável em meio à crescente massa visual que circula na rede. Porém, o que transparece à máquina seria o mesmo que nós vemos? Podemos supor que, por se tratar de uma tradução, haveria, inevitavelmente, alguma perda. Contudo, também se trataria, em algum grau, de um processo criativo – ao ver, a máquina produz, afinal, *um* olhar. Será necessário que encontremos então esta instância criativa, produtiva, para além do gesto de destruição que parece inicialmente se desenhar.

Peter Galison, historiador da ciência que tem em sua pesquisa um interesse particular pelo lugar ocupado pela imagem neste contexto<sup>64</sup>, descreve como o movimento ambíguo do *iconoclash* poderia ser observado em um debate entre a natureza da imagem e a natureza dos dados na prática científica (GALISON, 2008), em uma perspectiva que parece se aproximar bastante do problema que enfrentamos ao localizar a imagem no contexto da Visão Computacional. De um lado, a imagem teria reconhecido seu valor pela apresentação em um modo tátil-visual daquilo que nos propomos analisar – seja enquanto uma representação da própria empiria ou na forma de um gráfico, esquema ou modelo representando as equações ou cálculos empregados –, auxiliando tanto a descoberta científica quanto processos pedagógicos, além de por seu modo de representação alcançar características peculiares dos quais o cálculo ou a descrição verbal não conseguiriam dar conta. Contudo, a imagem seria também detestável no âmbito da ciência porque elas *enganam*, nos levam assumir premissas falsas e a particularizar ao invés de abstrair. Como diria o discurso de sua recusa, sugere Galison: “Se não se pode desenhar figuras do muito grande, do muito pequeno, ou do muito complexo, assim seja. Treinamento, descoberta e verdade são todos dependentes *apenas* de proposições não-ambíguas e de seu arranjo lógico”<sup>65</sup> (GALISON, 2008, p. 300). Apesar deste debate acalorado, que exemplificará com diferentes passagens da história das ciências naturais e da matemática, o autor busca compreender como esta discussão se sustenta, sem conclusão aparente, em

---

64 Cf. DASTON e GALISON, 2007. Neste estudo, junto à Lorraine Daston, os autores buscam observar em atlas científicos dos séculos XIX e XX a emergência e a transformação do conceito de *objetividade* nas ciências naturais a partir das imagens que elas produzem. Trata-se de uma dentre outras das pesquisas de Galison e de Daston que se dedicam à imagem como objeto de análise no âmbito das ciências. Outro autor que reconhecidamente dedica-se a tais objetos, mas a partir do ponto de vista da história da arte é James Elkins (cf. ELKINS, 2011).

65 No original: “*If pictures cannot be drawn of the very large, the very small, or the very complex, so be it. Training, discovery, and truth are all dependent only on unambiguous propositions and their logical arrangement*”. Tradução nossa.



um duelo constante entre o iconoclasmo e a iconofilia (GALISON, 2008, p. 301). Frequentemente, o uso da imagem nestes contextos, como indica o historiador, articula-se intimamente a um modo de extração de dados que pareceria buscar resolver o impasse entre o visual e o lógico e, por conseguinte, também os dados, com cada vez mais frequência, voltam-se à forma de imagens por meio da disciplina – hoje em destaque – da *visualização de dados*. Longe de um caminho unidirecional, da imagem à não-imagem, portanto, Galison sugere um movimento pendular entre estas duas instâncias que tomaria o numérico como ponto de contato entre elas:

Após perseguir o movimento pendular inesgotável entre imagens e dados, torna-se claro que a forte tendência às imagens e a igualmente forte pressão em direção à análise nunca estabilizou por completo a prática científica. Pelo contrário: nem o “pictorial-representativo” nem o “analítico-lógico” existem como posições fixas. Através de um amplo espectro das ciências especiais, nós vemos que a imagem em si está constantemente no processo de fragmentação e reconfiguração. Sem dúvida que isto ocorre desde muito, lá no passado da história das imagens científicas de Etienne Jules Marey ou ainda nas formas cambiantes dos diagramas na Renascença. Mas agora, mais intensamente, as já usuais conversões digital-analógico e analógico-digital tornaram as trocas cintilantes rotina: imagem à não-imagem à imagem. Não mais posta em movimento apenas em momentos de crise, percebemos que a ciência ordinária, cotidiana, ativa esta incessante oscilação:

“Imagens desintegram-se em dados, dados agregam-se em imagens”<sup>66</sup>  
(GALISON, 2008, p. 322).

Vemos hoje se realizar, de modo cada vez mais rotineiro, a agregação dos dados em imagens, em processos que vêm se tornando, inclusive, cada vez mais frequentes nas ciências sociais. Em estágio avançado em relação à infografia, que já há algum tempo constitui um recurso importante inclusive no jornalismo, a *visualização de dados* nomeia, em geral, a produção de gráficos a partir de massas de dados em maior escala, muitas vezes a partir de bases dinâmicas, em configurações diversas que permitam compreender, além das dimensões quantitativas dos dados abordados, aspectos emergentes das relações estabelecidas entre seus componentes. Se nestes casos, porém, os dados de que parte a visualização não seriam visuais, para o interesse específico de nossa abordagem da Visão Computacional as práticas de visualização serão importantes a fim de compreendermos a restituição da forma-imagem àquilo que, percebido pela máquina, se desintegraria em dados, algo que tanto pode operar pela via da *visualização de dados* quanto pela *Computação Gráfica*. Imagens-dados-imagens: é ao movimento pendular completo, habilmente descrito por Galison, que nos voltamos na tentativa de compreender aquilo que veem as máquinas. Longe de assumirmos a imagem enquanto

---

66 No original: “*After tracking the endless drive back and forth between images and data, it becomes clear that the powerful drive to images and the equally forceful pressure towards analysis never completely stabilized scientific practice. Quite the contrary: neither the ‘pictorial-representative’ nor the ‘analytic-logical’ exist as fixed positions. Instead, across a wide span of the special sciences, we see that the image itself is constantly in the process of fragmenting and re-configuring. No doubt this has long been so, back deep into the history of scientific images of Etienne Jules Marey or even to the shifting forms of diagrams in the Renaissance. But now, ever more intensively, the routinization of analog-to-digital and digital-to-analog conversions has made the flickering exchanges routine: image to non-image to image. No longer only set in motion at moments of crisis, we find that ordinary, every-day science propels this incessant oscillation: | ‘Images scatter into data, data gather into images.’*”. Tradução nossa.

um elemento perdido deste processo, é pela sua compreensão como um ponto de retorno, ainda que intermitente, que buscamos abordar nosso objeto neste capítulo.

Assumindo esta ambiguidade ao invés de combatê-la, portanto, nos propomos abordar a relação da Visão Computacional com as imagens e o visível sem a reduzirmos a um gesto puramente destrutivo, de apagamento ou de tradução em informação (longe do visível). Compreendemos, de outro modo, que esta é uma relação constituinte e fundamental do objeto que abordamos, não apenas pela sua dependência e articulação a processos de produção e circulação de imagens como, também, pela sua produção, ainda que oculta ou indireta, de registros visuais particulares articulados a seu modo de ver. “Deve haver uma conexão entre a produção e a guerra”<sup>67</sup>, repetem os letrados de Farocki. Embora seu interesse com esta frase seja propor uma vinculação fundamental entre a racionalização da produção industrial capitalista e a racionalização da guerra – entre as tecnologias de produção, da indústria, e as tecnologias de destruição, da guerra – talvez pudéssemos sugerir uma outra leitura, mais metafórica, pela qual a guerra de imagens, os *iconoclashes*, implicariam sempre uma dimensão produtiva articulada a seus processos destrutivos. Deve haver, também no caso da Visão Computacional, uma conexão entre sua “destruição” das imagens e algum modo de produção do visível, ainda que indiretamente. Talvez seja isto, inclusive, que nos apresente Farocki ao dar visibilidade ao que as máquinas de visão ocultariam. Identificando-nos com seu esforço, nos propomos neste capítulo encontrar brechas pelas quais vazam ou se insinuam as imagens da Visão Computacional – sua instância produtiva – e compreender em que medida revelariam aspectos de seu modo de ver.

## 2.2 Imagem–instrumento

Nos letrados da série *Eye/Machine* (2001-2003), vemos se repetir mais de uma vez a qualificação das imagens que nos são apresentadas como *operacionais*. São imagens que se endereçam a espectadores bastante específicos: às máquinas e, quando muito, aos técnicos e engenheiros. Cumprem, também, finalidades específicas, que se orientam ao funcionamento de máquinas da guerra e da indústria, sem pretenderem-se expressivas, belas ou reflexivas:

Estas imagens são desprovidas de intenção social / Elas não são para edificação / Não são para reflexão / Na era da produção de fluxos – / imagens para monitorar o predeterminado<sup>68</sup> (EYE/MACHINE, 2001).

---

67 “There must be a connection between production and war”. Tradução nossa.

68 No original: “These images are devoid of social intent / They are not for edification / Not for reflection / In the age of flow production – / images to monitor the predetermined”. Tradução nossa.

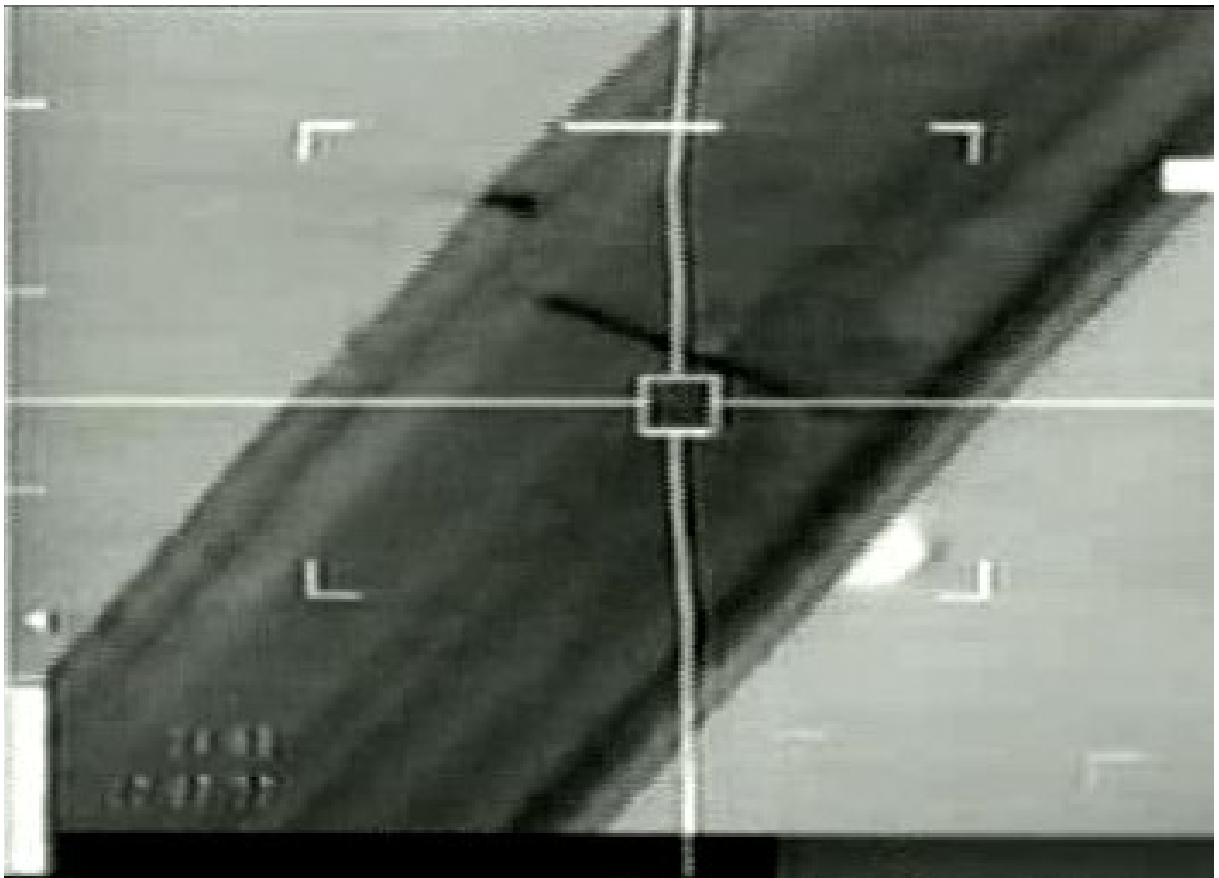


Figura 16 – Imagens de câmeras acopladas a mísseis em *frame* de um dos canais da videoinstalação *Eye/Machine* (2001), de Harun Farocki. Fonte: EYE/MACHINE, 2001.



Figura 17 – *Frame* da versão monocanal consultada, da instalação *Eye/Machine I* (2001), de Harun Farocki, exibindo, simultaneamente, os dois canais da obra. Fonte: EYE/MACHINE I, 2001.

Estas imagens são puramente operacionais / Se tais imagens possuem beleza – / esta beleza não é calculada / Cinematografia por dispositivos / Imagens gravadas por robôs de montagem / Imagens, na verdade não destinadas para olhos humanos / – às vezes vistas por técnicos para checar o funcionamento<sup>69</sup> (EYE/MACHINE III, 2003).

São imagens que não parecem guardar qualquer relação com aquelas com que tomamos contato no cinema, na arte, no entretenimento, na televisão, ou mesmo no jornalismo – por mais que cada vez mais adentrem estes domínios, como vimos, pelo menos desde a Guerra do Golfo. De nossa experiência, talvez recordem-nos, principalmente, o registro da perícia criminal que povoa boa parte da ficção televisiva e também certa linhagem sensacionalista da imprensa. Talvez, ainda, os diagnósticos médicos por imagem a que eventualmente nos submetemos – radiografias, ultrassonografias, ressonâncias magnéticas. São imagens que, por cumprirem funções prescritas, inserem-se em um domínio particular: servem como instrumentos, como ferramentas técnicas dos distintos contextos em que são produzidas.

Trata-se de um tipo de registro que possui características particulares ao dar privilégio à funcionalidade de suas estratégias figurativas, garantindo maior eficiência de sua análise, bem como ao demandar um olhar técnico treinado capaz de lê-la em seus aspectos significativos. A iluminação deve, na medida do possível, dar conta de toda a cena, sem deixar áreas de sombra ou penumbra. Deve-se garantir um bom contraste a fim de distinguir os objetos figurados. O enquadramento deve favorecer à leitura das dimensões mais significativas – talvez mesmo achatando e excluindo alguma dimensão a fim de aproximar o espaço de um plano cartesiano. Atendendo a finalidades pragmáticas pelas quais sustentariam um diagnóstico ou mesmo embasariam um plano de ação, as imagens instrumentais têm sua própria forma articulada de modo intrincado com os contextos de ação que dela podem se valer.

Encontramos, nesta descrição, uma grande proximidade entre as imagens instrumentais e aquele que descrevemos como o modo de percepção da Visão Computacional. Como descreve Manovich (1993a, p. 143), a Visão Computacional circunscreve um sentido particular ao *entendimento* de uma cena: para a Visão Computacional, trata-se de realizar operações de *localização e reconhecimento* que possam subsidiar a realização de ações seja no espaço concreto figurado, seja no espaço simbólico dos arranjos das imagens que são interpretadas. Diríamos, neste sentido, haver uma relação profunda entre tal modo de percepção e a forma pela qual as imagens instrumentais – não apenas no âmbito das máquinas tecnológicas de visão – se dão a ver. Embora, evidentemente, nem toda imagem instrumental se digirija ao olhar exclusivo da máquina, a máquina sempre veria, qualquer imagem, como se instrumental fosse. Se, de um lado, a imagem instrumental é aquela produzida com o objetivo de servir a objetivos operacionais, quaisquer que sejam os observadores a que se enderece

---

69 No original: “*These images are purely operational / If such images possess beauty – / this beauty is not calculated / Cinematography by devices / Images recorded by assembly robots / Images, not really intended for human eyes / – sometimes viewed by technicians to check functioning*”. Tradução nossa.

(técnicos, cientistas, detetives, médicos, *máquinas*), ela também nomearia o modo pelo qual a Visão Computacional toma qualquer imagem.

Mesmo diante de uma fotografia *postada* por um usuário do *Facebook*, por exemplo, tratar-se-á ainda de um modo de percepção que a tomará operacionalmente, em meio à tarefa de indexar os arquivos de imagem do usuário ou de compor sua rede de relacionamentos. Diante de uma tal fotografia, as operações efetuadas serão ainda as de reconhecer, identificar, localizar, orientar – assim como faz o míssil em busca do alvo. Diante do olho da máquina, neste sentido, não haveriam grandes diferenças entre o rastreamento aéreo do campo inimigo e as fotografias encontradas a esmo na internet – entre uma e outra, mudaria apenas o alvo. A articulação que sugerimos entre *imagem-instrumento* diz respeito, então, tanto às imagens instrumentais propriamente ditas quanto ao modo como a Visão Computacional interpreta as imagens que lhe são alimentadas. Ainda que o dispositivo que estudamos não vise necessariamente à produção de imagens, os registros instrumentais parecem ser fundamentais para que compreendamos seu modo de ver. Neste sentido, devemos tomar a imagem como um dos elementos que seriam mobilizados pelo dispositivo da Visão Computacional em seu funcionamento – no vocabulário de Latour, diríamos que ela representa uma das agências integrantes desta composição, que nele desempenha um papel fundamental. A imagem seria, então, um mediador que requisita nossa atenção tanto quanto os algoritmos que buscamos descrever no capítulo anterior – até porque, como veremos, é em relação a elas que tais algoritmos são elaborados. Talvez pudéssemos dizer que a imagem (em particular a de base fotográfica) integra as máquinas que abordamos como uma de suas partes – inclusive tendo já inscrita, em seu modo de registro, a possibilidade deste acoplamento.

Em uma passagem de *Eye/Machine I* (2001), somos apresentados, em um dos canais, a uma imagem tomada por um protótipo de carro autônomo (que dirige-se sozinho), no qual temos uma vista monocromática da rua com uma demarcação gráfica, em vermelho, destacando a linha do meio-fio da calçada. Para um carro que dirige sozinho, como a um motorista, traça-se um limite por se observar na condução do veículo. No outro canal, vemos um registro gráfico simplificado, que já abandona a imagem de vídeo que lhe dá forma, mas que também representa uma vista da rua, com as linhas do meio-fio expressas em verde e objetos, aparentando outros carros e pedestres, figurados por silhuetas brancas (Figura 17). Em ambas as telas, nos é apresentado como os algoritmos de Visão Computacional de tal protótipo interpretam as imagens alimentadas a fim de extrair delas informações úteis acerca do posicionamento e movimentação do carro – subsidiando, assim, decisões do programa acerca da condução do veículo (acelerar, frear, virar etc.). O vídeo da rua é reduzido a informações simplificadas e de processamento mais fácil – linhas e formas geométricas – que também permitem, por sua posição e orientação relativa, inferir aspectos como as distâncias e as dimensões dos objetos, dando maior precisão aos movimentos efetuados. Em suma, o processamento visa recompor o espaço tridimensional figurado na imagem – de modo bastante similar ao processo

elaborador por Lawrence G. Roberts (1963) em sua tese de doutorado, que discutimos no capítulo anterior.

Embora trate-se de um resultado obtido a partir do processamento da imagem – com operações de detecção de contornos e arestas<sup>70</sup> – ele se apoia necessariamente em uma operação que teria sido realizada anteriormente, na instância de registro da imagem pela câmera. Afinal, estamos diante, neste caso, de uma representação do espaço baseada na *perspectiva linear*, método de figuração que embora tenha se desenvolvido no entorno do desenho e da pintura, é também aquele no qual se baseiam os princípios ópticos e geométricos das câmeras de base fotográfica – como a câmera de vídeo pela qual o carro robótico percebe seu entorno. Trata-se de elemento importante para o funcionamento das máquinas que abordamos, pois é somente a partir de seu acoplamento às máquinas de registro que sua tarefa torna-se possível. A *perspectiva*, inscrita em algumas destas máquinas, mais do que um método de figuração *realista*, tem como uma de suas principais vantagens a racionalização do espaço pela geometrização. Por meio dela, obtemos meios de analisar, com maior precisão e consistência, aspectos daquilo que é representado numa figura plana. Uma vez que a representação em perspectiva encontra-se inscrita no próprio funcionamento da câmera, que a reproduz por meio dos recursos ópticos empregados na objetiva, toda imagem fotográfica produzida por tais instrumentos permite-se ser lida através dela<sup>71</sup>. Esta é, inclusive, uma das razões principais pela qual a imagem fotográfica presta-se tanto ao domínio das imagens instrumentais, periciais ou científicas. Por meio dela, encontramos – de forma automatizada – um tipo de representação que preserva a consistência das formas e dimensões do espaço e objetos retratados.

Como sugere o crítico e curador William M. Ivins (1975, p. 9), a perspectiva elabora uma via de mão dupla entre a imagem e a coisa representada:

A perspectiva pode ser compreendida como um meio prático para assegurar uma relação métrica de mão dupla, ou recíproca, entre as formas dos objetos definitivamente localizados no espaço e suas representações pictóricas. Embora isto seja importante para a produção de imagens no sentido mais estreito, é sem dúvida ainda mais importante para o pensamento em geral, porque as premissas em que está baseado são implícitas em toda constatação feita com seu auxílio<sup>72</sup> (IVINS, 1975, p. 9).

---

70 Cf. “Anexo I – Listagem ilustrativa de operações características da Visão Computacional”.

71 Tratamos, evidentemente, de um domínio particular das imagens fotográficas, embora seja o mais comum. Dubois (2012), por exemplo, oferece-nos uma definição do fotográfico que vai muito além das imagens em perspectiva, já que em sua ênfase da indicialidade como traço definidor da fotografia ele sugere que a fotografia implicaria, fundamentalmente, o fenômeno químico de registro da imagem – que incluiria, assim, formas de registro sem câmera e sem convergência dos raios luminosos, como os fotogramas que, talvez por esta razão, foram tão amplamente experimentados pelos surrealistas. Em todo caso, mesmo nestas outras manifestações do fotográfico – que também encontramos nas radiografias e outros tipos de diagnóstico médico por imagem – persiste algo como um *coeficiente* de consistência da representação que é a chave principal pela qual tomamos, em nosso argumento, a perspectiva.

72 No original: “*Perspective may be regarded as a practical means for securing a rigorous two-way, or reciprocal, metrical relationship between the shapes of objects as definitely located in space and their pictorial representations. Important as this is to picturemaking in the narrowest sense, it is doubtless even more important to general thought, because the premises on which it is based are implicit in every statement made with its aid*”. Tradução nossa.

Formulada em um breve ensaio deste autor acerca da “racionalização da visão” por meio da perspectiva, esta citação de Ivins tem fundamental relevância para a compreensão da Visão Computacional, como é sugerido por Manovich (1993a, p. 111-112). Pois a interpretação automatizada de imagens, em especial no contexto das operações de *localização*, apoia-se fundamentalmente no método representacional da perspectiva: é por meio dela que se permite aos procedimentos algorítmicos deste domínio funcionar como uma câmera às avessas, derivando a descrição precisa dos objetos e espaços representados a partir da superfície de sua representação bidimensional. Toda imagem produzida por meio da perspectiva, como indica Ivins, traz em si, de modo implícito, uma espécie de esquema gramatical, de regra, que garante tanto coerência interna à representação quanto correspondência lógica e recíproca entre a representação e seu referente. Isto, que tomamos especialmente para a perspectiva por ser um método figurativo que herdamos das tradições do desenho e da pintura, aplica-se, de modo mais geral, a todo tipo de projeção geométrica que seja consistente e previsível – em imagens como as radiografias, por exemplo, embora não se trate de um registro em perspectiva e ainda que percamos referências de profundidade, encontramos, todavia, uma representação consistente no que toca às dimensões e distâncias observadas. Em todo caso, seria apenas pela existência de uma correlação possível entre a imagem e seu referente que a Visão Computacional poderia elaborar uma chave algorítmica para sua leitura, uma vez garantido certo nível de reversibilidade da operação e consistência de sua aplicação.

Bruno Latour (1986, p. 7) convoca o estudo de Ivins para sustentar seu argumento de que as formas representacionais teriam sido elementos fundamentais da constituição das ciências modernas. Ele situa tal inovação em meio a outras, como a prensa de tipos móveis e à cartografia, que indica como fundamentais para o desenvolvimento de métodos de inscrição ao mesmo tempo móveis (transportáveis, acumuláveis) e imutáveis (garantindo a consistência do registro ao longo do tempo e dos deslocamentos), além de apresentáveis, legíveis, reprodutíveis e comparáveis. Latour sugere, a partir de Ivins, que a perspectiva teria tido papel fundamental neste processo na garantia de uma “consistência óptica” das representações, pela qual poderíamos virtualmente navegar pelo espaço representado e efetuar translações nos objetos figurados sem que eles perdessem suas características e proporções: “Ivins nos mostra como a perspectiva permite o movimento através do espaço com, por assim dizer, uma passagem de volta”<sup>73</sup> (LATOUR, 1986, p. 7). O autor indica ainda a relevância de tal método pictórico, na abordagem de Ivins, pela possibilidade de mesclar a representação visual do espaço com a geometria:

Como vimos para a perspectiva, o espaço no papel pode ser feito contínuo em relação ao espaço tridimensional. O resultado é que podemos trabalhar no papel com réguas e números, mas ainda manipular objetos tridimensionais ‘lá fora’ (Ivins, 1973). Melhor ainda, por causa desta consistência óptica, tudo, não importa de onde venha, pode ser convertido em diagramas e números, e uma combinação de números e tabelas pode ser usada, que são ainda mais fáceis de manejar do que palavras ou silhuetas (Dagonet, 1973). Você não pode medir o

---

73 No original: “*Ivins shows us how perspective allows movement through space with, so to speak, a return ticket*”. Tradução nossa.



Sol, mas você pode medir uma fotografia do Sol com uma régua. [...] Isto é o que eu chamo, na falta de termo melhor, a vantagem de segundo grau das inscrições, ou a mais-valia que é ganha por sua capitalização<sup>74</sup> (LATOURE, 1986, p. 20).

Pelo argumento de Latour, portanto, a perspectiva configuraria um método de inscrição que teria fundamental relevância para a emergência da ciência moderna, justo por aquilo que é sustentado por Ivins (1975), que é o seu papel na *racionalização da visão*. O argumento desenvolvido por Latour, tendo na perspectiva um de seus exemplos, é o de que teria sido fundamental para a ciência moderna a elaboração de métodos de inscrição que permitissem o conhecimento do mundo a partir de suas representações, envolvendo também práticas de acúmulo de inscrições consultáveis e comparáveis, bem como o trânsito dos estudos e observações realizadas. Por meio da representação em perspectiva torna-se possível mensurar o visível, organizá-lo, compará-lo, acumulá-lo. Em suma, nela encontramos um reminiscente genealógico da Visão Computacional.

Na esteira de Latour, Manovich (1993a, p. 113-114) sugere que, por meio da perspectiva, a imagem poderia ser um instrumento de poder, já que, garantida sua consistência óptica, seria possível valer-mo-nos da imagem para mobilizar e manipular recursos à distância. A uniformidade garantida por esta técnica de representação permite, afinal, trabalhar sobre as representações mas ainda assim planejar ou compreender objetos e acontecimentos que têm seu lugar *out there*. Construindo tanto sobre Ivins quanto sobre Latour, Manovich (1993a) elabora sua abordagem da Visão Computacional em constantes remissões a suas formulações, inserindo-a em meio a um conjunto de técnicas que têm a perspectiva como principal precursora e as quais reúne sob a noção do *nominalismo visual*. Em sua narrativa sobre o que chama de uma “engenharia da visão”, tema de sua tese, o autor pretende com tal noção demarcar uma diferença entre os paradigmas da visão em que se baseariam tecnologias de monitoramento e supervisão espacial, como o radar e a Visão Computacional, e, de outro modo, práticas do *design* e da comunicação visual. Neste caso, tendo como exemplos ilustrativos tanto o recurso aos pictogramas no *design* gráfico soviético quanto a montagem intelectual de Sergei Eisenstein, o visual seria abordado enquanto meio para transmitir ideias e conceitos abstratos – neste sentido, em compreensões que poderíamos dizer *idealistas* do visual. No domínio da Visão Computacional e do radar, por sua vez, haveria uma ênfase maior em outra operação, que seria a de representar um objeto ou cena concretos e singulares (MANOVICH, 1993a, p. 98) – assim, ele propõe denominar a esta abordagem como a de um *nominalismo visual*<sup>75</sup>.

---

74 No original: “As we saw for perspective, space on paper can be made continuous with three-dimensional space. The result is that we can work on paper with rulers and numbers, but still manipulate three-dimensional objects ‘out there’ (Ivins, 1973). Better still, because of this optical consistency, everything, no matter where it comes from, can be converted into diagrams and numbers, and combination of numbers and tables can be used which are still easier to handle than words or silhouettes (Dagonet, 1973). You cannot measure the sun, but you can measure a photograph of the sun with a ruler. [...] This is what I call, for want of a better term, the second-degree advantage of inscriptions, or the surplus-value that is gained through their capitalization”. Tradução nossa.

75 Tal denominação é sugerida por Manovich em uma remissão ao debate filosófico quanto ao que constituiria a realidade, se ideias abstratas ou se formas concretas e singulares, identificado no *idealismo* de Platão e no *nominalismo* de Aristóteles (MANOVICH, 1993a, p. 98). Dado o escopo da dissertação, não avançaremos em tal discussão.

Nesse âmbito, Lev Manovich inclui formações desde a perspectiva linear até tecnologias como a fotografia, o radar, a Visão Computacional e mesmo a Computação Gráfica e a perspectiva, as quais ele entende como engajadas nesta tarefa de representação de objetos e cenas singulares e concretas. Conquanto seria possível indicarmos algumas limitações em sua proposição<sup>76</sup>, seria sustentável seu argumento se circunscrevemos tal denominação de um *nominalismo visual ao uso instrumental* destas tecnologias. Não foi por acaso, afinal, que a perspectiva linear se desenvolveu tão fortemente na arquitetura e que ainda tenha relevância neste domínio. Vale lembrar, também, que é em especial no contexto do projeto arquitetônico e de engenharia que a perspectiva ainda persiste enquanto um método representacional relevante – ainda que, como já abordaram autores como Panofsky (1993), tenhamos observado seu declínio na arte, pelo menos desde o modernismo. Manovich (1993a, p. 116) sugere que é pela atenção de William M. Ivins à cultura visual, mais do que, de modo restrito, à arte, que faz com que ele seja capaz de perceber a relevância contemporânea da perspectiva nestes outros domínios. Nestes, ainda hoje, embora com adaptações, ela segue sendo um referencial importante, inclusive como um dos princípios nos quais se apoiam programas de *design* auxiliado por computador (*Computer Aided Design* – CAD) utilizados na arquitetura e no desenho industrial, por exemplo<sup>77</sup>.

Para Manovich (1993a, p. 130), haveria uma conexão intrínseca entre os desenvolvimentos da computação gráfica em 3D e da Visão Computacional, inclusive com um nascimento simultâneo na pesquisa de Lawrence G. Roberts (1963). A síntese da imagem a partir dos dados extraídos do registro fotográfico constitui uma etapa importante, como vimos, dos processos desenvolvidos por este cientista na medida em que ela permitiria, por exemplo, compor outros pontos de vista da mesma cena ou simular um deslocamento no espaço representado – como parece ser necessário para o caso do carro autônomo trazido por Farocki. Neste sentido, portanto, ainda que possamos ver na Computação Gráfica um sentido bem mais idealista do que o nominalismo descrito por Manovich, podemos entender como também ela pode dizer respeito, em determinadas aplicações, a espaços e objetos particulares, subsidiando a ação da máquina sobre estes elementos. Há, assim,

---

76 Haveria problemas nesta categorização talvez decorrentes de uma generalização da abordagem nominalista para toda uma mídia sem especificar contextos particulares de aplicação – pois, não apenas a perspectiva poderia se prestar para a criação de cenas idealizadas, porém realisticamente apresentadas, quanto mesmo a fotografia, ainda que sempre guarde uma relação intrínseca com seu referente (necessariamente concreto e singular) poderia ser orientada a cenas idealistas. O cinema de Eisenstein, afinal, tido como um dos exemplos de um idealismo visual, com sua busca pela representação visual de conceitos abstratos como a luta de classes, parte, ainda assim, de meios de base fotográfica para tal fim. Haveria como idealizar a partir do concreto. Quanto à Computação Gráfica, estaríamos já bem mais no domínio do idealismo, poderíamos supor, já que trata-se antes de uma modelização matemática, de uma simulação, do que da representação de uma cena concreta ou singular – por mais que assim possa aparentar.

77 Ainda assim, vale ressaltar, entre o uso instrumental, pautado pelo que Manovich chama de *nominalismo visual*, e outras aplicações desta tecnologia, mesmo em um domínio particular de aplicação, o limite seria tênue: no desenho de pranchas para auxiliar o trabalho no canteiro de obra, por exemplo, a correspondência precisa entre representação e o espaço constitui um aspecto crucial por se considerar, prestando-se a tal abordagem nominalista; nas simulações realistas da construção concluída, contudo, utilizadas em anúncios do empreendimento imobiliário – em geral apresentando uma bela edificação povoada, em um cenário de tranquilidade de felicidade, em pleno caos e conflito urbano – a idealização ganha novamente a cena

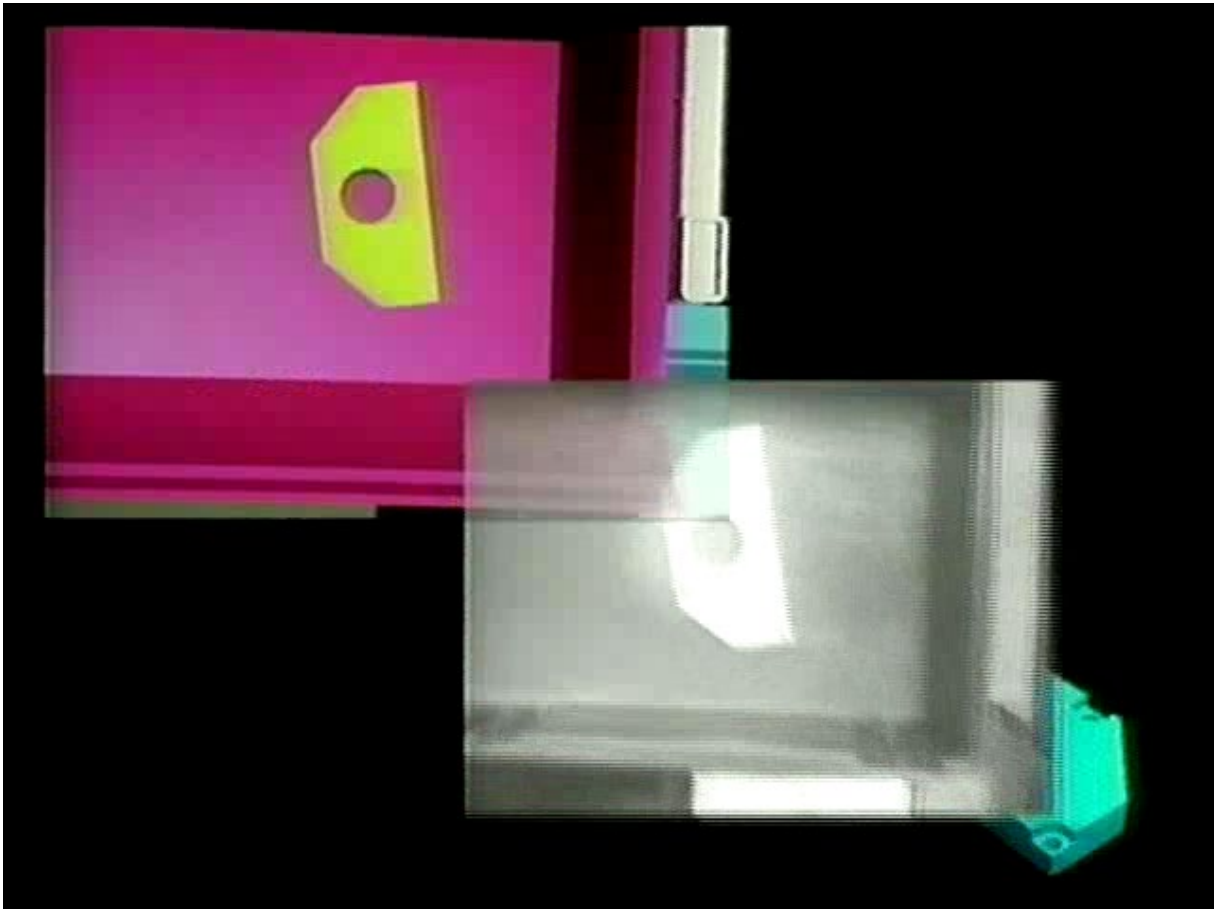


Figura 18 – *Frame* da versão monocal consultada, da instalação *Eye/Machine I* (2001), de Harun Farocki, exibindo, simultaneamente, os dois canais da obra. Fonte: EYE/MACHINE I, 2001.

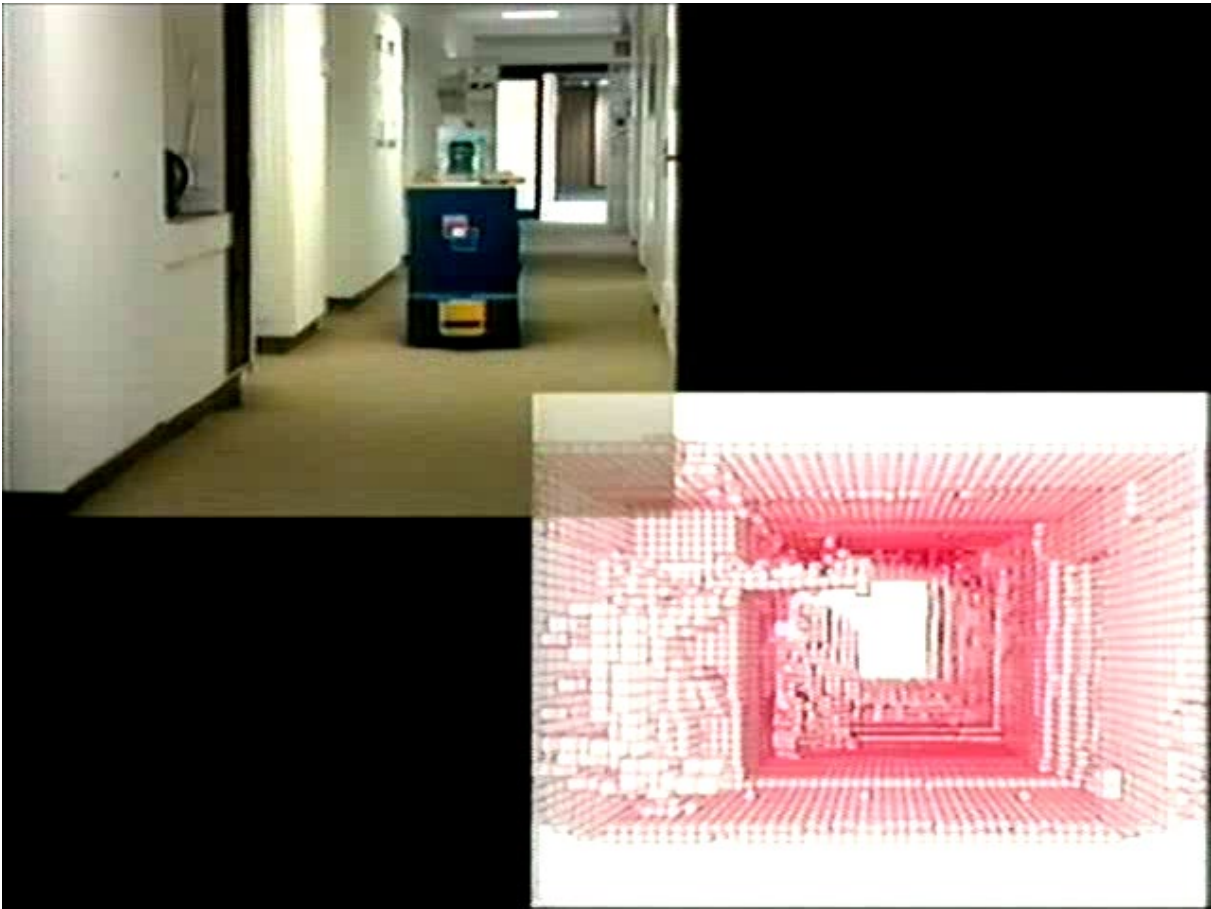


Figura 19 – *Frame* da versão monocal canal consultada, da instalação *Eye/Machine III* (2003), de Harun Farocki, exibindo, simultaneamente, os dois canais da obra. Fonte: EYE/MACHINE III, 2003.

um eixo a atravessar as imagens em perspectiva que passa pela Computação Gráfica e incide sobre os próprios algoritmos empregados na interpretação automatizada de imagens. Manovich (1993a, p. 130) sugere que foi importante para o sucesso do desenvolvimento da Computação Gráfica tridimensional o fato de que desde a perspectiva linear, tratava-se de um processo algorítmico, composto por regras e cálculos a serem conduzidos pelo artista mas que podiam ser facilmente transpostos ao domínio da máquina. Segundo ele, teria sido inclusive necessário a Roberts recorrer a textos alemães sobre a geometria da perspectiva do início do século XIX para desenvolver seus algoritmos. A constituição algorítmica da perspectiva foi também fundamental para o sucesso da empreitada da Visão Computacional, pois é o que garante que tal processo possa ser tomado de modo inverso, determinístico e reversível. O autor indica que entre a Computação Gráfica e a Visão Computacional, encontramos processos inversos:

Efetivamente, o objetivo da visão computacional é desfazer o que a computação gráfica 3D tenta atingir. A primeira tenta reconstruir a cena a partir da imagem fotográfica, deduzir a informação independente do observador de seus objetos: suas formas e dimensões. [...] O objetivo da computação gráfica 3D é exatamente o oposto: dadas as informações objetivas sobre a cena (formas e posições de objetos, seus materiais, direção da luz), produzir sua imagem que é virtualmente indistinguível de uma fotografia<sup>78</sup> (MANOVICH, 1993a, p. 147).

De um lado, para o nosso caso, a Visão Computacional abstrai da imagem descritores numéricos dos objetos e espaços representados e, de outro, a Computação Gráfica recompõe a visualidade desfeita em informação. Trata-se de um estágio posterior, neste sentido, à operação que abordamos neste estudo, e que dá visibilidade aos produtos obtidos no processo. Não seria, deste modo, exatamente *aquilo que a máquina vê*, mas talvez seja um dos modos pelos quais torna-se possível que algo do que vê nos seja apresentado. Haveria certa complementaridade entre tais processos, servindo-se inclusive enquanto forma de verificação do bom funcionamento da máquina e seus programas, pelo que se observaria se o que ela percebe é em alguma medida análogo ao que nós percebemos da mesma cena.

Encontramos algo desta complementaridade no *frame* de *Eye/Machine I* já apresentado (Figura 17). De modo ainda mais próximo às representações que temos como típicas da Computação Gráfica, encontramos tal modo de funcionamento ainda em outros registros colhidos por Farocki (Figura 18 e Figura 19). Vemos então como tanto a imagem de que parte o programa de Visão Computacional como a representação em computação gráfica elaborada por seu processamento apresentam-se como faces de uma mesma estratégia pela qual a imagem se faz instrumento. O programa realiza descrições e cálculos a partir dos quais recompõe as posições e dimensões. Porém, trata-se de processo que estende-se desde a formação da imagem pela câmera, seguindo princípios em certa medida algorítmicos e objetiváveis que a Visão Computacional toma de modo reverso. Também, podemos

---

78 No original: “*In effect, the goal of computer vision is to undo what 3-D computer graphics aims to achieve. The former tries to reconstruct a scene from its photographic image, to deduce the viewer independent information about the objects: their shapes and dimensions. [...] The aim of 3-D computer graphics is exactly the opposite: given the objective information about the scene (shapes and positions of objects, their materials, direction of light) to produce its image which is virtually indistinguishable from a photograph*”. Tradução nossa.

compreender que a mesma estratégia se estende aos modos pelos quais o programa reconstitui seus resultados a uma forma visível, pois a Computação Gráfica baseia-se nos mesmos procedimentos da perspectiva linear para produzir suas visualizações. Dada a centralidade destes códigos de formação da imagem, que atravessam todo o processo abordado – da análise à síntese do visível – talvez seja possível dizermos que a Visão Computacional até mais do que “decifrar o visível”, dedica-se a decifrar imagens a partir de seus próprios princípios constitutivos, tendo na perspectiva sua chave de leitura.

O modelo geométrico da câmera e a formação da imagem constituem temas primários do campo da Visão Computacional abordados, por exemplo, nos capítulos introdutórios do livro de Forsyth e Ponce (2012). Uma das primeiras tarefas enfrentadas pela aplicação desta tecnologia poderia ser compreendida como a dedução de uma função operacional da câmera por meio de determinados procedimentos de calibragem. Por tal “função operacional”, nos referimos àquela que descreveria a mediação efetuada pela câmera entre o referente e sua representação visual, sua inscrição em imagem. Cada artefato possuiria, pela configuração particular de seus componentes ópticos, um modo distinto de inscrever a imagem, derivado de aspectos como a distância focal de sua objetiva, a profundidade de campo obtida, a amplitude do campo etc. Tais procedimentos de calibragem visam, então, inferir a função matemática que descreveria, de modo previsível e reversível, a formação da imagem por aquela câmera em particular, observados tais fatores. Em outras palavras, neste primeiro momento, busca-se encontrar a chave de leitura a ser empregada pelo sistema de Visão Computacional de modo a se permitir tomar a imagem por esta via de mão dupla pela qual poderíamos acessar aspectos de seu referente.

Manovich (1993a, p. 145) indica que, apesar desta herança fundamental da perspectiva, a Visão Computacional logo enfrentaria as limitações deste tipo de registro, apesar de suas pretensões. Conquanto seja racionalizado e geométrico, ele ainda seria insuficiente para que pudéssemos recompor a totalidade da cena representada. “A imagem é indeterminada”<sup>79</sup>, sugere Manovich (1993a, p. 145), mas não para indicar a abertura de sua interpretação, em um sentido mais amplo, mas a sua dubiedade mesmo em contextos pragmáticos: mais de um arranjo espacial pode resultar em uma mesma imagem em perspectiva, como as gravuras ilusionistas de Escher parecem nos alertar. Além disto, a aparência de um objeto é influenciada por ângulos de iluminação e características dos materiais e nem todas as características geométricas seriam preservadas. Neste sentido, a imagem perspectivada seria antes um elemento emblemático do tipo de exame e controle espacial presumido pela articulação imagem–instrumento e pela Visão Computacional do que, propriamente, sua solução última. Com efeito, em diferentes aplicações não se restringe a tomada da imagem do ambiente a apenas um ponto de vista, recorrendo também ao desenvolvimento de sistemas estereoscópicos ou, ainda, a modos de inferência mais precisa da distância, articulando a produção da imagem a mecanismos de senseamento remoto como sonares ou, no caso do sensor Microsoft

---

79 No original: “*The image is undetermined*”. Tradução nossa.

Kinect, com o recurso a indicadores visuais projetados sobre a cena que ofereceriam à imagem uma referência mais precisa das distâncias observadas<sup>80</sup>. Em todo caso, porém, trata-se antes de modos auxiliares ou complementares de produção das imagens, acreditando ainda em sua funcionalidade instrumental, do que de seu abandono. Mesmo nestes casos, ainda que recorrendo a formas mais sofisticadas para contornar as limitações da perspectiva linear da imagem de base fotográfica, ainda estamos diante de um mesmo tipo de articulação das máquinas de visão às imagens.

No capítulo anterior, remetemos a uma compreensão das máquinas que Guattari toma de Leibniz pela qual toda máquina seria tomada em sua articulação com outras máquinas – a *máquina articulada* (GUATTARI, 2003, p. 91). Remetemos, ainda, à metáfora da *rede* pela qual Latour (2005) descreve a articulação de múltiplos agentes na realização de uma ação – que não poderia, assim, ter sua origem reduzida a esta ou aquela agência. Por tais elaborações, compreendemos que a articulação entre a Visão Computacional e a imagem articulam-se enquanto agências interdependentes. Afinal, a visão algorítmica destas máquinas já possui inscritos, em seus procedimentos, aspectos do registro da imagem instrumental de base fotográfica. Além disso, é o modo de inscrição intrínseco a este tipo de imagem, com a reversibilidade sugerida por Ivins (1975) e, ainda, a possibilidade de conhecer e agir sobre o mundo a partir das imagens, sugerida por Latour (1986), que garante a possibilidade de funcionamento da Visão Computacional. Entre o modo de representação e a automatização do procedimento de leitura, encontramos, enfim, ramificações de um mesmo desenvolvimento que contemporaneamente desemboca naquilo que indicamos, a partir de Fernanda Bruno (2013), como um *regime de eficiência informacional*. Tomando toda e qualquer imagem como instrumental, as máquinas que abordamos visam, afinal, transformar o visível em informação.

Como nos lembra a autora em um outro texto em que se dedica ao caso específico da chamada *videovigilância inteligente* (BRUNO, 2012), encontramos um antecedente importante das máquinas de visão contemporâneas nos catálogos criminalísticos de Bertillon, desenvolvidos no final do século XIX (BRUNO, 2012, p. 56). A autora indica como tanto nas câmeras inteligentes da Visão Computacional quanto nos trabalhos deste fisionomista francês visa-se uma taxonomia dos corpos figurados, ou seja, a elaboração de um sistema de classificação que permita recuperar informações a partir de uma imagem. No caso de Bertillon, a identificação de um suspeito a partir de um banco de imagens de criminosos conhecidos; nos casos analisados pela autora, a identificação de um comportamento suspeito como índice de um acontecimento por vir. De fato encontramos neste exemplo, também para nosso percurso dissertativo, um dos primeiros esforços sistemáticos da utilização da fotografia como instrumento, inserida em um circuito mais amplo de procedimentos e operações. Se, como sugere Deleuze (2005, p. 49) a partir de Foucault, toda tecnologia seria social antes de ser técnica, talvez encontremos no enfrentamento da imagem por Bertillon uma das primeiras configurações que mais se aproximam daquela da Visão Computacional, inclusive

---

80 No *Kinect*, a alternativa compreende, basicamente, a projeção de um padrão de pontos sobre a cena por um *laser* infravermelho (invisível à visão humana) pelo qual a imagem da câmera permitiria analisar a deformação deste padrão e assim inferir os volumes e dimensões dos objetos figurados.

buscando enfrentar um problema bem próximo àqueles que identificamos, hoje, como motivação para o desenvolvimento dos algoritmos de interpretação de imagens. Em um texto dedicado ao uso da fotografia no contexto da criminalística, Tom Gunning (2004) traz uma citação do próprio Bertillon que ilustra bem este ponto:

Nos últimos dez anos, a polícia parisiense colecionou mais de 100 mil fotografias. Você acha possível comparar sucessivamente cada uma dessas 100 mil fotografias a cada uma das 100 pessoas presas diariamente em Paris?

Havia necessidade de um método de eliminação análogo àquele empregado em ciências como a botânica e a zoologia; isto quer dizer, tomando como base os elementos característicos da individualidade, e não o nome, que está sujeito a falsificações (BERTILLON, 1896<sup>81</sup> *apud* GUNNING, 2004, p. 48).

Diante da prática crescente e já bem estabelecida de se fotografar suspeitos detidos pela polícia com a finalidade de auxiliar processos de identificação, o então oficial de polícia Bertillon desenvolveu todo um método de padronização da tomada destas imagens e de indexação e arquivamento com o objetivo de tornar o processo mais racional e eficiente. Havia, é claro, o interesse de contornar possíveis falseamentos e disfarces dos criminosos, que muitas vezes recorriam, para além do habitual barba, cabelo e bigode, a mutilações, visando remover traços característicos ou deformar seus rostos. Mas também, principalmente, visava-se tornar mais eficiente o processo de busca de uma imagem por meio de critérios precisos e objetivos. Como descreve Gunning (2004, p. 49), seu sistema envolveu desde a padronização da pose e do enquadramento, até protocolos de tomadas de medida de partes específicas dos corpos que permitiriam a classificação estatística do criminoso. Mas, principalmente, teriam sido as técnicas de catalogação e arquivamento que constituíram o grande trunfo do desenvolvimento de Bertillon.

O sistema de medição de Bertillon procura resolver precisamente o paradoxo da fotografia nos sistemas de circulação modernos. A fotografia serve aos propósitos de vigilância e identificação necessários a um sistema policial burocrático, estabelecendo a identidade por meio de sua rede de semelhança icônica e referência indicial. No entanto, ela permanece *muito* individual, muito específica para ser processada tão minuciosamente quanto exigiam os meios rápidos de circulação de informação. Portanto, a fotografia analógica necessita, efetivamente, tornar-se digitalizada, complementada por dados quantificáveis que atribuam a cada fotografia uma posição única em um sistema racionalizado de informações. Ela própria tinha que ser analisada e racionalizada... (GUNNING, 2004, p. 50).

Tal problema, como vimos, pouco se distancia daquele enfrentado hoje diante dos *atuais* “meios rápidos de circulação de informação”. A imagem fotográfica já traria um modo de inscrição efetivamente apropriado a seu uso instrumental – guardando proporções e certa coerência formal dos objetos representados que a permitiria ser tomada neste uso. Tal seria, como discutimos, uma das vantagens da perspectiva que desde o Renascimento já cumpria um tal papel na arquitetura. Com a fotografia, porém, a simplificação do processo de tomada de imagens, e, principalmente, as

---

81 BERTILLON, Alphonse. *Signaletic instructions, including the theory and practice of anthropometric identification*. R. W. McCloughry (org.). Chicago: The Werner Company, 1896. p. 12-13.



demandas institucionais de sua produção e recuperação em um ritmo acelerado fizeram com que não se bastassem as relações icônicas e indiciais estabelecidas entre a fotografia e seu referente, já que o problema torna-se menos o de reestabelecer tais relações (algo que a fotografia resolveria quase que instantaneamente), mas o de como manejar o volume de imagens disponíveis de forma eficiente. É preciso tornar o singular comensurável, passível de comparação, afim de que se possa arquivá-lo e recuperá-lo – algo que hoje a Visão Computacional oferece, na esteira do projeto de Bertillon, embora com particularidades que deveremos observar.

Vemos ressoar a mesma preocupação de Bertillon na formulação contemporânea de Manovich, quando ele propõe a pergunta de como comparar um milhão de imagens (MANOVICH e DOUGLAS, 2011). E, se a pergunta é a mesma, a resposta também envolve estratégias similares. Bertillon elaborou um método para reduzir o escopo de imagens a se comparar a partir da medição de partes selecionadas dos corpos dos criminosos e o cruzamento dos dados obtidos – como o tamanho do dedo mínimo ou a forma das orelhas do suspeito. Manovich, em suas pesquisas, analisa traços cromáticos das fotografias, as poses daqueles retratados, a expressão facial, entre outros aspectos. Em um contexto mais orientado à identificação, como aquele de Bertillon, são ainda as medidas de traços da face de que se valem empresas como o *Facebook* ao rastrear a presença *online* de seus usuários em fotografias postadas na rede social. Como Gunning sugere, a partir de Allan Sekula (1986), o artefato principal do sistema de Bertillon não era a câmera, mas o arquivo. De modo similar, os sistemas hoje derivados da Visão Computacional adquirem papel central no modo com que nos relacionamos com as imagens. Mas, bem além de apenas um método de arquivamento e recuperação de fotografias – e este seria um importante traço diferencial das tecnologias contemporâneas – tais sistemas cruzam hoje bem mais do que apenas os aspectos aparentes, combinando-os com outras imagens, com dados acerca dos indivíduos e de suas ações (compras, mensagens enviadas, lugares percorridos etc.). Na pesquisa realizada por Manovich, trata-se de buscar padrões nas estratégias de autorrepresentação de acordo com a localização dos autores das imagens. No caso do *Facebook* – como em aplicações securitárias – trata-se menos de recuperar uma imagem e identificar o suspeito, como faz Bertillon, do que de percorrer a enorme rede ativada por aquela imagem, inserindo aquele prontamente identificado em perfis de consumo e/ou de risco. Menos recuperar uma imagem, portanto, do que mobilizar a maior quantidade de informação que possa ser inferida a partir daquela imagem.

A diferença substancial do sistema de Bertillon aos dos dias atuais, portanto, vai bem além da maior eficiência da recuperação da informação, embora seja dela derivada: fundamentalmente, não é apenas aos criminosos que tais técnicas se dirigem. Mesmo em usos não particularmente circunscritos aos domínios da vigilância ou da segurança pública, vemos hoje como se alastra este método como se, ativado pelo diagrama da eficiência informacional, tal desenvolvimento cujo percurso traçamos desde a perspectiva alcançasse, hoje, a posição especial que Deleuze reserva, em sua interpretação, aos dispositivos. Em todo caso, e este é um argumento central que buscamos

desenvolver nesta seção, trata-se menos de um desenvolvimento isolado destas máquinas, em particular, do que da manifestação de uma herança histórica que se atualiza e se diferencia, hoje, em função dos novos agenciamentos de que participa. Como vimos, há precedentes importantes que são incorporados à Visão Computacional, relacionados ao modo de registro e a metodologias anteriores de classificação e recuperação da informação a partir da imagem. Neste sentido, tanto no método de Bertillon quanto nos carros sem motorista, nos robôs de chão de fábrica ou nos mísseis teleguiados, abordamos sistemas operacionais que integram a imagem enquanto um elemento fundamental de suas próprias engrenagens – também ela, neste sentido, poderíamos dizer que *funciona* e ativamente participa dos compostos que viemos abordando. O que se dissemina, contudo, no estágio atual de desenvolvimento do dispositivo da Visão Computacional, é principalmente o modo de ver que toma tais registros como instrumentais, já bem além das aplicações securitárias, bélicas, industriais, médicas ou científicas, que lhe são próprias, para alcançar também o domínio das imagens amadoras em circulação na *web* e mesmo das câmeras integradas em celulares, computadores, *tablets* e consoles de *videogame* – alimentando interfaces amigáveis ou intuitivas. Além disso, pela massiva interconexão em rede de toda materialização do visível, uma imagem não vem recuperar sua semelhante apenas (como no caso da metodologia de identificação de Bertillon), mas toda uma dispersa rede de relações que a orbita. Pela conversão potencial de qualquer imagem em informação assistimos à *dispersão* ou *desintegração*, nos termos propostos por Galison (2004), que assim a coloca disponível para que qualquer um dos dispositivos deste *visionismo* – se assim possamos arriscá-lo chamar – a tome como instrumento. Dispersa em informação, ela se torna disponível para estabelecer associações diversas, indo da identificação à composição de perfis mais amplos nos quais se aglutinariam derivações estatísticas do visível.

### 2.3 A visão subjetiva das máquinas

No outro extremo do pêndulo – que na seção anterior oscilava próximo à desintegração da imagem – seria o caso de buscarmos, como sugere a noção de *iconoclash* por Latour e Weibel (2004), como poderia haver alguma reconstituição ao visível daquilo que teria se desfeito com o olhar analítico lançado pela Visão Computacional. Trata-se em geral, como já apontamos, de um esforço de *visualização* pelo qual os dados em que teria se desintegrado a imagem seriam novamente *agregados* em forma visível (GALISON, 2004). Mas, diferentemente da abordagem desta questão a partir dos gráficos que de alguma forma ofereceriam tal visibilidade, gostaríamos de seguir um caminho sugerido por Virilio (1994 e 1993<sup>82</sup>) em sua discussão acerca das máquinas de visão. O autor sugere,

---

82 O autor desenvolve este tópico tanto no texto “A imagem virtual mental e instrumental”, publicado em *Imagem-máquina*, coletânea organizada por André Parente (VIRILIO, 1993), quanto como *parte* do capítulo final “A máquina de visão”, publicado em livro homônimo do próprio autor (VIRILIO, 1994). Optamos, sempre que a passagem está contida na parte publicada em 1993, por citar a versão desta publicação uma vez que nela observamos um melhor trabalho de tradução. Contudo, as passagens dele citadas podem também ser encontradas, em formulações bastante próximas, na versão publicada em 1994.

indicando esta como uma das questões centrais diante da visão sintética, uma aproximação analógica daquelas que nomeia como as *imagens virtuais instrumentais* destas máquinas ao que chamaríamos de nossas *imagens mentais*:

Definitivamente afastados da observação direta ou indireta das imagens de síntese produzidas *pela máquina, para a máquina*, essas imagens virtuais instrumentais serão, para nós, equivalentes àquilo que já significam hoje as representações mentais de um interlocutor estrangeiro: um enigma. Sem que haja saídas, gráficas ou videográficas, a prótese de percepção automática funcionará como uma espécie de imaginário maquínico do qual seremos totalmente excluídos. (VIRILIO, 1993, p. 127)

O autor parte, como logo se percebe, de uma personificação das máquinas, as quais sugere como alheias ao humano, como se os contornos de seu imaginário não se interceptassem com os do nosso<sup>83</sup>. Ele busca, neste sentido, compreender aquilo que seria sintetizado pela a máquina, a partir de sua percepção, e que não se voltaria particularmente ao olhar humano – como é o caso da *visualização de dados* – mas que atende a suas próprias demandas operacionais (“*pela máquina, para a máquina*”). Levando ao extremo seu diagnóstico, que indicamos em citação anterior, quando afirma que nas máquinas de visão “*o retorno-imagem não é mais assegurado*” (VIRILIO, 1994, p. 103), o autor sugere, aqui, que seríamos totalmente excluídos de seu imaginário. Ele se indaga, então, quanto à existência e à natureza de suas representações internas. Diante deste “enigma insondável”, Virilio desenha uma compreensão segundo a qual apenas poderíamos supor ou imaginar quais seriam tais representações e como elas seriam, demandando que recorramos às *imagens mentais*, oriundas de nossa própria experiência perceptiva.

A insolubilidade deste problema, evidentemente, se deve à própria dificuldade que teríamos em supor que, no nível destas representações internas, encontraríamos algo no plano do visível. Compreendamos como dados, como pulsos elétricos ou como gravações magnéticas, bem sabemos que ali se trata de uma representação numérica, simbólica, que apenas poderia se converter em visível se a este estado a reconstituíssemos – o que já não atenderia ao funcionamento da máquina, mas à nossa interface com ela. O problema da sugestão de Virilio, como discutiremos, talvez seja esta exclusão da interface, como se a máquina só pudesse dizer de si mesma e, por conseguinte, nos excluiria de seu modo de existência. Porém, a sugestão, ainda que hipotética, deste limite que nos permitiria supor a existência destas representações não deixa de ser instigante para nossa investigação, já que talvez por meio dela pudéssemos descrever o problema que enfrentamos neste capítulo como o de *ver pelos olhos da máquina* – por mais absurdo que possa soar.

Parece-nos ainda válido reforçar, que é apenas de forma hipotética – a fim de dar encaminhamento à análise – que supomos, neste momento, tal distanciamento da máquina. É claro que, por se tratar de um desenvolvimento técnico passível tanto de aprofundamento quanto de intervenção, o *enigma*

---

83 Trata-se de um argumento que lhe rendeu algumas críticas que optamos por assumir e abordar no capítulo seguinte. A crítica sobre este aspecto, em particular, é desenvolvida por JOHNSTON, 1999.

de que fala Virilio não seria assim tão enigmático. A sugestão de um tal alheamento, portanto, poderia trazer implicações, inclusive, éticas, como se nos isentássemos do desenvolvimento destas máquinas e de nossa capacidade de decidir sobre seus caminhos. Contudo, cremos na validade deste exercício como modo de melhor compreendermos as implicações de assumirmos a existência de “máquinas que veem”. Uma vez orientadas a finalidades bastante específicas, corremos o risco de as medirmos apenas por seu grau de eficiência, sem considerarmos as representações que para isto produzem, ou que tipo de olhar desenvolvem neste processo. Afinal, por mais estejamos nos referindo a programas que nós (humanos) desenvolvemos, há de se levar em conta que muitas das técnicas hoje utilizadas para a programação destes dispositivos valem-se de processos automatizados para seu desenvolvimento. Além disso, como sugere Latour (2001) em sua discussão da natureza da mediação técnica, por mais que sejamos nós a criar e desenvolver tais máquinas, devemos conceder a estes seres certo grau de autonomia pelo qual seriam capazes de fazer-nos agir, tanto quanto nós, a eles – por mais que a arma seja obra humana, não teríamos exatamente controle sobre aquilo que ela faria, ou nos faria fazer. Tomarmos este distanciamento inicial para compreendermos como e o que veem estes dispositivos, portanto, não apenas parece plausível quanto pode ser um primeiro passo para engajarmo-nos em seu desenvolvimento e intervenção – sem abdicarmos, assim de nossa responsabilidade. Quando dizemos, então, de uma *visão subjetiva* das máquinas, aludimos à existência hipotética de uma representação interna às máquinas, a qual poderíamos observar como imagem, e pela qual poderíamos compreender alguns outros aspectos do modo como veem. Seguimos, então, a trilha de Virilio em sua sugestão destas *imagens virtuais instrumentais*, que seriam o equivalente técnico das *imagens mentais*. Mas, ao mesmo tempo, aceitamos o desafio de contradizer seu diagnóstico, quando afirma que estaríamos totalmente excluídos deste imaginário: buscamos, ao contrário, formas de visualizá-lo.

No que tange às imagens mentais, o cinema há tempos vale-se de um recurso narrativo para contornar sua inacessibilidade, apresentando-nos com certa recorrência aquele que seria o ponto de vista de seus personagens através da chamada *câmera subjetiva*. Fazendo as vezes de um equivalente cinematográfico do discurso em primeira pessoa – ou do fluxo de consciência que encontraríamos, tipicamente, em James Joyce e Clarice Lispector, dentre outros –, é assim que nos é permitido encontrar na tela, de modo bastante literal, o ponto de vista de um personagem. São momentos em larga medida já incorporados à gramática fílmica nos quais, não raramente, nos é dado acompanhar uma cena pelo olhar de um de seus personagens, como se assim fosse facilitada nossa identificação com ele ou, ao menos, como se assim pudéssemos melhor compreender seu ponto de vista das situações narradas. De modo análogo, é também por meio de tal recurso que encontramos no cinema de ficção científica figurações daquilo que seria, se pudéssemos acessá-la, a *visão subjetiva* de uma máquina. Em filmes que as trazem como personagens da narrativa, que possuem a capacidade de ver – computadores, andróides, ciborgues e robôs –, com frequência somos levados a acompanhar determinadas cenas compartilhando, por via da tela, a visão de alguns destes personagens. Ainda que

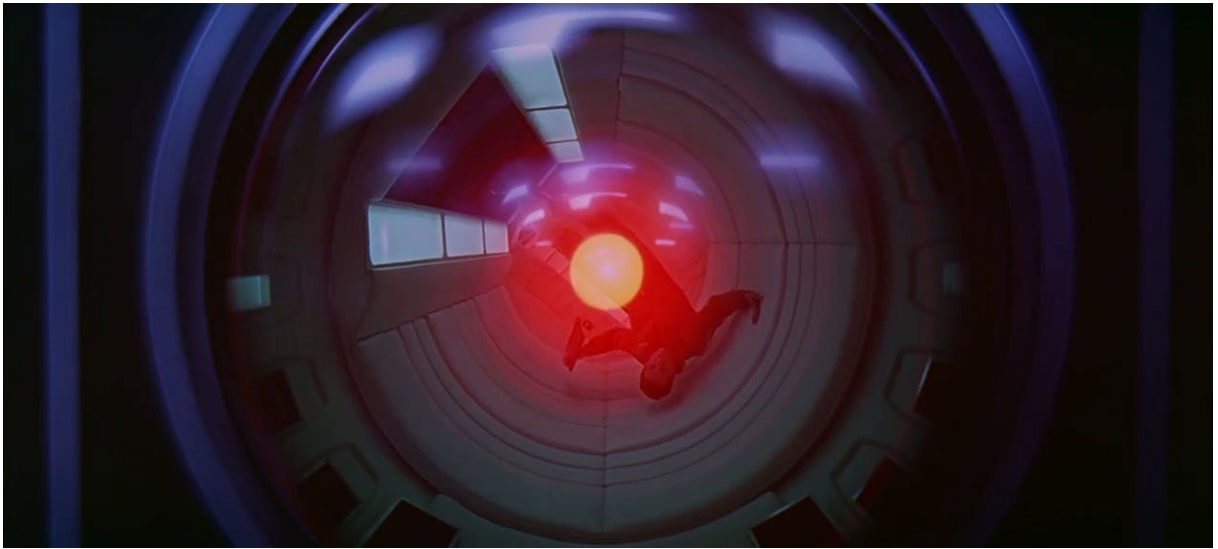


Figura 20 – Plano detalhe do olho de HAL 9000 em *2001: uma odisseia no espaço* (1968), de Stanley Kubrick. Fonte: 2001, 2014.



Figura 21 – Ponto de vista de HAL em *2001: uma odisseia no espaço* (1968). Fonte: 2001, 2014.



Figura 22 – Ponto de vista de HAL em *2001: uma odisseia no espaço* (1968). Fonte: 2001, 2014.

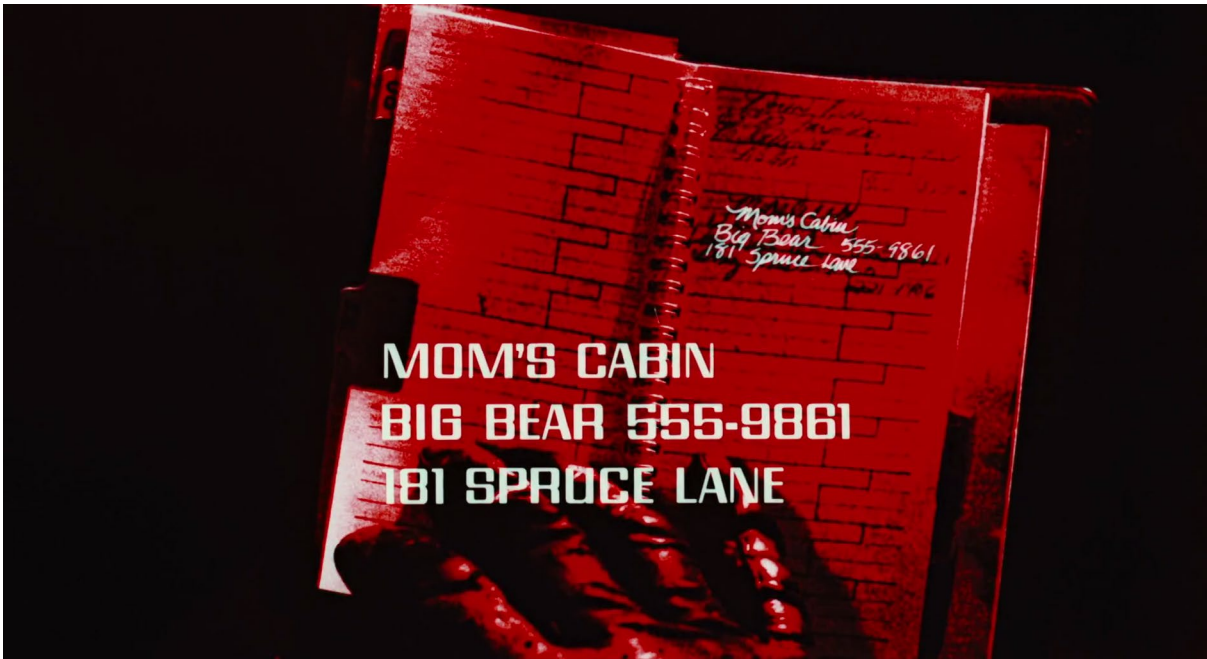


Figura 23 – Frame de *O exterminador do futuro* (1984), de James Cameron. Texto e elementos gráficos sobrepostos sugerem o reconhecimento da escrita manual pelo robô androide. Fonte: TERMINATOR, 2014.



Figura 24 – Frame de *Robocop* (1987), de Paul Verhoeven. Elementos gráficos sobrepostos indicam a operação de alinhamento da mira ao alvo realizada pelo policial ciborgue. Fonte: ROBOCOP, 2014.

em um contexto ficcional, encontramos em alguns exemplos deste tipo de figuração uma referência imaginária daquilo que poderíamos considerar como uma *visão subjetiva* da máquina.

Em *2001: uma odisseia no espaço*, dirigido por Stanley Kubrick em 1968, encontramos aquele que talvez seja um dos mais conhecidos destes personagens, o computador de bordo HAL 9000, que acompanha os protagonistas da segunda parte do filme em sua conturbada missão a Júpiter a bordo da espaçonave *Discovery*. Totalmente integrado à nave, HAL tem uma presença distribuída ao longo de todos os seus ambientes, figurando em cada um deles por meio de seu grande olho ciclópico: câmeras dotadas de objetivas esféricas que cobrem, de modo abrangente, um amplo campo de visão, e no meio das quais reluz um brilhante ponto vermelho, fazendo as vezes de pupila (Figura 20). É diante desta espécie de interface que os protagonistas interagem com o personagem e é através deste *olho* que nos são apresentados, em diversos momentos, o ponto de vista da máquina (Figura 21 e Figura 22). Inicialmente apresentados como mais um ponto de vista do lento cotidiano da missão – que, de tão longa, tem alguns de seus principais tripulantes em estado de hibernação até a chegada ao planeta – os planos tomados do ponto de vista do computador não parecem, a princípio, desempenhar um papel tão central na narrativa. Contudo, ao longo de seu desenvolvimento, tais momentos ganham especial importância ao sugerir que HAL possuiria intenções não manifestas em algumas das ações que realiza, até que rebela-se contra a tripulação em um ato desesperado de autopreservação. É através do recurso à *câmera subjetiva* que nos é revelado que o personagem acompanha o diálogo travado entre o comandante da missão e seu imediato em seu plano de desativá-lo diante das suspeitas de seu mal funcionamento. É também através destes planos que passamos a compreender que as ações de HAL talvez não fossem tão autoevidentes, quanto suporíamos ser as ações de uma máquina, como se houvesse uma clara separação entre aquilo que ele demonstra em suas ações e discurso e aquilo que seriam seus reais planos e intenções. As remissões a seu ponto de vista, neste sentido, parecem ser cruciais para acompanharmos, ainda que de forma silenciosa, as transformações por que passa o personagem ao longo da narrativa, sugerindo uma dimensão interna, em alguma medida inalcançável, através desta figuração, na tela, de seu ponto de vista.

O exemplo de HAL em *2001* se contrapõe a figurações mais corriqueiras do ponto de vista da máquina que encontramos em outras referências da cinematografia de ficção científica e que se assemelhariam mais aos registros documentais trazidos por Farocki na série de instalações *Eye/Machine*. Em representações orientadas a um contexto pragmático imediato e mais relacionadas às operações que identificamos em torno das categorias de *localização* e *reconhecimento*, encontramos este tipo de representação, por exemplo, em passagens dos filmes das franquias *O exterminador do futuro*

e *Robocop*<sup>84</sup>, quando também nos são apresentados os pontos de vista do androide e do ciborgue<sup>85</sup> que, respectivamente, protagonizam seus filmes (Figura 23 e Figura 24). São, em geral, registros de baixa resolução e qualidade, sobre os quais são sobrepostos elementos gráficos simples indicando, ao olhar do espectador, o processamento realizado pela máquina na interpretação das imagens. São sobrepostas linhas em cruz indicando o alvo, destacadas porções da imagem reconhecidas, linhas de comando sugerindo processos em curso no funcionamento da máquina. Como seria de se esperar em personagens concebidos, literalmente, como “máquinas de matar”, encontramos figurações bem próximas das máquinas militares e industriais que nos são apresentadas por Farocki. Na maior parte dos casos, nestes filmes<sup>86</sup>, sugere-se pelo ponto de vista da máquina uma percepção visual articulada diretamente à ação, uma percepção pragmática, como descrito por Manovich (1993a, p. 143), orientada ao reconhecimento de padrões e à localização de alvos. Desta forma, antes de sugerirem uma dimensão subjetiva dos personagens, parecem cumprir o papel auxiliar de exibir a precisão e a frieza de seu modo de operação, algo que também se exhibe, regularmente, na feição inexpressiva que exibem seus rostos e na precisa movimentação mecânica de seus corpos – além, é claro, da eficiência vigorosa com que atingem os alvos e resistem aos ataques sofridos. Enquanto no caso de HAL sua frieza e calculismo surgem enquanto aspectos de sua personalidade, tais características, no caso do *Exterminador* e do *Robocop*, pareceriam ser apenas decorrentes de eles serem máquinas, sem que isto redunde na constituição de uma personalidade ou na sugestão de uma dimensão interna, de um modo de ver e pensar que não estivesse já evidente em suas ações.

Ainda que tais imagens sugiram muito do imaginário acerca do que veriam as máquinas, como representações que são de sua suposta *visão subjetiva*, não chegamos por tal registro – nem mesmo no caso de HAL – ao ponto de podermos desvendar o enigma de que fala Virilio. Devemos reconhecer o hiato existente entre o modo de *representação* do que veem as máquinas e o que de fato veriam, assim como o registro da chamada *câmera subjetiva* é apenas um recurso para a representação, e não a apresentação, do que efetivamente seria a percepção visual de uma cena por um personagem – mesmo que humano. Seria demasiado simplista supor que por tais figurações, que tanto se aproximam das saídas gráficas oferecidas por dispositivos de Visão Computacional a seus

---

84 A franquia *O exterminador do futuro*, estrelada por Arnold Schwarzenegger, tem início com o filme homônimo de 1984, dirigido por James Cameron, tendo, até o momento, três continuações em longa-metragem: *O exterminador do futuro 2: o julgamento final* (1991), de James Cameron, *O exterminador do futuro 3: a rebelião das máquinas* (2003), de Jonathan Mostow, e *O exterminador do futuro: a salvação* (2009), de McG. A franquia *Robocop* tem início com o filme homônimo de 1987, dirigido por Paul Verhoeven, com duas continuações e uma refilmagem: *Robocop 2* (1990), de Irvin Kershner e *Robocop 3* (1993), de Fred Dekker, e a refilmagem do primeiro episódio, *Robocop* (2014), de José Padilha.

85 Tipicamente, androide denomina um robô que assemelha-se a um humano em sua forma aparente, enquanto ciborgue nomeia a mistura de partes humanas e robóticas, em geral na forma de um humano que tem suas capacidades ampliadas por próteses robóticas.

86 Em certas passagens de *Robocop* encontramos, destoando do conjunto, momentos em que esta representação característica dá espaço a recordações do personagem de momentos de sua vida na forma humana, antes de sua transformação em ciborgue, em que a figuração transforma-se bastante, com um registro de melhor qualidade e sem as interferências gráficas características. Contudo, estes pequenos lampejos parecem restituir algum grau de interioridade ao personagem apenas pela posta em evidência de sua parte humana (contraposta à sua parte máquina), que de resto é praticamente excluída da narrativa.



operadores e supervisores humanos, chegássemos algo próximos de suas representações internas, assim como tampouco poderíamos supor que o registro realizado por uma câmera conseguiria abarcar a dinâmica de percepção da cena por uma pessoa. Em ambos os casos, trata-se, antes, de um modo de visualização que se oferece a *nosso* olhar, desenhado para que se apresente ao espectador. Neste sentido, não deixam de ser curiosas as figurações exibidas por tais filmes em uma suposição da aparência destas *representações internas* elaboradas pelos personagens (humanos ou máquinas). Em todo caso, a própria noção paradoxal de *câmera subjetiva* explicita a ambiguidade deste tipo de representação que por mais que possa buscar uma equivalência entre o ponto de vista da câmera e aquele do personagem, não chegaria a alcançar, efetivamente, a síntese perceptiva que produziria a *imagem virtual mental* ou *instrumental* a que Virilio se refere.

O autor coloca bem a questão ao indicar que não poderíamos abordar as máquinas de visão apenas pelo modo de registro da imagem, por sua *objetivação* em relação a um *suporte-superfície* (VIRILIO, 1993, p. 129), sendo necessário que as compreendamos com relação ao tempo da percepção: “este tempo de exposição que nos faz ver ou não permite mais ver” (VIRILIO, 1993, p. 129). O autor sugere, assim, que a questão central no domínio da automação da percepção se dirige ao problema da relação entre *atualidade* e *virtualidade*. Pois se tratamos de um modo de percepção – seja do humano ou da máquina – havemos de considerar a necessária participação do tempo como um de seus componentes, mesmo na percepção imediata, pois aquilo que vemos sempre decorrerá não apenas da projeção que instantaneamente se produz sobre a retina, mas também de uma síntese posterior que nos permitirá perceber tanto o movimento quanto a profundidade espacial, por exemplo; além disto, dependerá da mobilização da memória, que também se coloca como um componente necessário a toda forma de percepção. “Uma vez que qualquer *tomada de imagens* (mental ou instrumental) é, simultaneamente, uma *tomada de tempo*, por mais ínfimo que este seja, este ‘tempo de exposição’ acarreta uma memorização (consciente ou não) segundo a velocidade de tomada de imagens” (VIRILIO, 1993, p. 129). Tendo em vista esta extensão da percepção no tempo, que mesmo no cinema apenas poderia ser registrado de modo analítico, por via de instantes discretos, a imagem objetiva, enquanto componente atual ou fatural da percepção, não seria suficiente para compreendermos um modo de percepção, que sempre mobilizará uma dimensão virtual.

Esta discussão é elucidada pelo autor em uma remissão ao princípio do cinema: “Como admitir o caráter fatural do fotograma e rejeitar a realidade objetiva da imagem virtual do espectador de cinema? [...] Mais ainda, como aceitar o princípio da persistência retiniana sem aceitar, ao mesmo tempo, o papel da memorização na percepção imediata?” (VIRILIO, 1993, p. 128). Longe do objetivismo com que talvez pudéssemos compreender uma imagem como a fotografia como uma representação plena da percepção visual de um instante, a abordagem de Virilio traz inscrita a constatação de que a percepção se desenrola no tempo e articula-se necessariamente a um passado e a um futuro. Compreendendo a visão como um “ato antes da ação” (VIRILIO, 1993, p. 129), ele sugerirá, um

pouco adiante, que ver seria prever (VIRILIO, 1993, p. 129). Assim, mais além da *profundidade de campo* que reconhecemos como um dos elementos a compor a imagem perspectivada – e na qual seríamos suplantados pelo telescópio e pelo microscópio na superação de nossas limitações – Virilio indica que as máquinas de visão suplantariam a fraqueza da *profundidade de tempo* de nossa visão fisiológica (VIRILIO, 1993, p. 129). Chegamos, assim, a um outro componente de nossa discussão acerca do que veem as máquinas em que já não nos basta a compreensão da imagem como instrumento, como discutimos na seção anterior, pois não nos bastará *uma imagem* para descrever este modo de percepção. Embora sejam importantes os princípios que permitem à Visão Computacional tomar a imagem como este intermediário reversível entre mundo e representação, talvez sejam ainda mais importantes aqueles que lhe permitem mobilizar uma espécie de memória para as tarefas de reconhecimento e aqueles que lhe permitem ultrapassar o registro *atual* para chegar à *virtualidade* com a qual efetivamente opera.

Tratamos já deste problema, em parte, ao discutirmos a relação entre os desenvolvimentos da Visão Computacional e da Computação Gráfica tridimensional, a partir da leitura feita por Manovich (1993a, p. 130) da pesquisa de Lawrence G. Roberts (1963). A síntese digital em três dimensões oferece-se tanto como uma ferramenta importante para a inferência de características da cena interpretada quanto como uma espécie de processo inverso ao realizado pela Visão Computacional – indo da informação à imagem. Mesmo em nossa percepção cotidiana, poderíamos compreender a mobilização desta virtualidade pelo recurso a inferências, ainda que involuntárias acerca daquilo que vemos, como forma de interpretarmos a cena – como os diversos exemplos de ilusão de óptica nos demonstram. O ato de *ver* – e, sob este aspecto, qualquer ato de percepção – envolve, portanto, inferências acerca daquilo que percebemos, projeções que não se limitam àquilo que atualmente incidiria sobre nossas retinas. Com o que Virilio nos sugere, chegamos ainda a outro aspecto desta articulação entre *atualidade* e *virtualidade*, em que não poderíamos compreender a visão sem levarmos em conta as representações internas – virtuais, poderíamos supor – que elaboramos (nós e as máquinas) acerca daquilo que percebemos. A síntese técnica de tal processo deslocaria nosso olhar para tais representações justamente ao torná-las, em algum nível, objetiváveis – embora, evidentemente, não haja uma equivalência real entre a percepção visual e sua codificação algorítmica, a qual descreve, efetivamente, apenas certa compreensão da visão.

Em seu estudo acerca do que sugere como um paradigma da *visão subjetiva* que teria emergido na primeira metade do século XIX, Jonathan Crary (1992) dedica-se justamente a alguns dos processos que Virilio parece se referir na sugestão da dimensão inerentemente *virtual* da percepção visual. A persistência retiniana – indicada por este autor como exemplo para a contraposição entre a imagem mental do espectador e a atualidade factual do fotograma – seria apenas um dos dois componentes fundamentais abordados por Crary a partir dos artefatos técnicos envolvidos no estudo e na popularização dos processos subjetivos e corpóreos da percepção visual. Além da percepção do movimento a partir de uma sequência de imagens estáticas, que Crary aborda em objetos como

o fenacistocópio e o praxinoscópio, ele também discorrerá acerca da visão estereoscópica, que compreende o princípio pelo qual realizamos uma síntese mental a partir de duas imagens tomadas de pontos de vista distintos na formação de uma percepção do volume e da profundidade. O autor sugere uma ampla articulação que conecta os estudos em fisiologia da visão no período a: formas de entretenimento popular, no entorno destes dispositivos ilusionistas; demandas institucionais relacionadas à racionalização da percepção no âmbito da otimização laboral; elaborações filosóficas no campo da fenomenologia; e emergências de antecedentes da arte impressionista, como ele observa em Turner, especialmente. Podemos compreender que, em larga medida, o reconhecimento do papel ativo do sujeito no ato de observar seria também o reconhecimento da visão como um ato produtivo que não apenas se atenta ao atual mas que, também, se desdobra em virtualidades.

Será mais importante para este momento de nossa discussão o modo como Crary sugere uma contraposição entre a compreensão da visão que teria se construído por tal paradigma da *visão subjetiva* e aquela que diria respeito a um paradigma anterior, que ele localiza no entorno da câmara escura. Trata-se de um aspecto de seu argumento que já indicamos brevemente: por sua configuração espacial e, principalmente, pela produção óptica de uma imagem externa ao corpo do observador, a câmara escura teria inaugurado um modo de compreensão da visão baseado na separação entre objeto e sujeito, em que a percepção que se tem do mundo derivaria basicamente do próprio mundo, sem participação significativa daquele que observa (CRARY, 1992, p. 25-66). Conforme sugere o autor, esta leitura da câmara escura teria aparecido em diversos textos filosóficos do século XVII e XVIII – de autores como Descartes, Newton e Leibniz, entre outros – embasando uma compreensão do conhecimento e da ciência pautada por esta separação fundamental entre mundo e observador. Nos termos de nossa discussão, talvez pudéssemos tomar esta compreensão da visão e do conhecimento como pautada pela atualidade: como se fossem excluídos, do ato de percepção e de conhecimento, componentes de virtualidade. Os desenvolvimentos da primeira metade do século XIX estudados por Crary integrariam um outro movimento que colocaria em evidência, pela via dos componentes subjetivos da percepção, justamente a tarefa de síntese, ultrapassando a atualidade. Aquilo que seria percebido como profundidade ou como movimento, afinal, já não estaria presente apenas nas imagens apresentadas ao olhar, mas seriam qualidades da imagem mental elaborada pelo observador. Pela síntese do movimento ou do espaço a partir de registros imagéticos discretos, produz-se, efetivamente, uma representação virtual daquilo que se percebe. Como sugere Virilio (1994, p. 87), contudo, embora evidentemente se trate de uma virtualidade do ato perceptivo, encontramos hoje no campo da neurofisiologia, das Ciências Cognitivas e das máquinas de visão desenvolvimentos que tomam por base certo caráter fatural destas imagens virtuais, demandando que concedamos, a elas, uma *atualidade*.

A inacessibilidade desta representação interna, sugerida pelo autor, talvez se deva ao fato de que seu modo de inscrição não é propriamente visual, mas compõe-se de uma rede de elementos que seria ativada pela percepção de uma cena, articulando-se não apenas aos aspectos que escapariam

ao registro fotográfico ou cinematográfico (como a profundidade e o movimento, que sobre tais registros seriam projetados virtualmente), mas também à memória que seria mobilizada para compreendê-la. Escreve Virilio:

O espaço do olhar não é, portanto, um espaço newtoniano, um espaço absoluto, mas um espaço minkovskiano, um espaço relativo (topológico e teletopológico). Não é somente a obscura claridade das estrelas que nos chega do passado longínquo da noite dos tempos; também a fraca claridade que nos permite apreender o real, nos permite ver, compreender nosso ambiente presente, provém de uma longínqua memória visual sem a qual não haveria o ato de olhar (como é provado pelos fenômenos de cegueira psíquica) (VIRILIO, 1994, p. 89).

Situados na encruzilhada do passado e do futuro, na sugestão de Virilio, não haveria como vermos sem prevermos e sem rememorarmos. A alegoria do céu estrelado – em uma remissão sutil ao árido tema da relatividade espaço-temporal discutida pela Física – evoca, pois, o tempo que necessariamente se inscreve no olhar, dado que a distância, no caso da observação do céu, permite-se converter em tempo – da chegada da luz até aqui. Mirar o céu, diz-se, é mirar o passado, como bem sabem os astrônomos que buscam no universo profundo indícios da origem dos tempos. Mesmo na escala humana, contudo (na qual a alegoria de Virilio ganha, ainda, força) se concordamos com os pontos de que partimos, de que a visão não é um ato solitário e que nela se inscreve, necessariamente, um componente construído e histórico, não haveria como desconsiderarmos que nosso olhar não reside, pontualmente, no presente<sup>87</sup>. A representação interna das máquinas perceptivas, portanto, se as equivalemos às imagens mentais, como sugere Virilio, dificilmente poderiam ser visualizadas na forma de um instantâneo. Precisaríamos ver, nelas – mais do que aquilo que atualmente se apresenta diante do aparato perceptivo –, a dimensão virtual que sobre elas se projeta, tomando-as como a encruzilhada entre passado e futuro que efetivamente são.

As visualizações oferecidas por textos técnicos da Visão Computacional oferecem-nos, neste caminho, uma potencial aproximação do enigma destas representações internas. Subprodutos do processo de aprendizado de máquinas, imagens compostas a partir da sobreposição e mesclagem de inúmeros exemplares daquilo que se programa a máquina para reconhecer são, nestes contextos, oferecidas como sugestões daquilo que as máquinas veriam. Em um destes casos, encontramos em um texto jornalístico (MARKOFF, 2012) o resultado apresentado pela *Google* de experimento realizado no âmbito de seu projeto *Google Vision*, de Visão Computacional, baseado em técnicas de aprendizado

---

87 Trata-se de tema caro a Walter Benjamin – um dos autores de que partimos para compreender a visão por este viés que leva em conta sua circunscrição histórica. Também para ele – como o conceito de *aura*, mais marcadamente, parece exemplificar – haveria uma profunda relação entre percepção e memória, que a modernidade colocaria em crise Cf. BENJAMIN, 2012 e 1995. Especialmente em “Sobre alguns temas em Baudelaire” (1995), o filósofo sugere uma articulação de sua compreensão da percepção aurática e a noção de memória involuntária, que ele desenvolve a partir de Freud e Proust. Por meio desta vinculação, podemos compreender, como também sugere Miriam Hansen (2012), que a distância a que se refere em sua conhecida definição da percepção da aura como “uma aparição única de algo distante, por mais próximo que esteja” (BENJAMIN, 2012, p. 14) seja a do lapso temporal da emergência de uma memória involuntária. A distância, que poderíamos supor uma qualidade espacial, portanto, diria respeito neste caso a uma dimensão temporal.

profundo (*deep learning*) não supervisionado<sup>88</sup>. Empenhando 16 mil núcleos de processamento aos quais foram apresentadas cerca de 10 milhões de imagens encontradas em vídeos do *YouTube*, o programa desenvolvido pela empresa teria aprendido, autonomamente, a reconhecer diversos elementos recorrentes nas imagens, dentre os quais, gatos (Figura 25)<sup>89</sup>. Segundo artigo publicado pelos pesquisadores envolvidos no experimento (LE *et al.*, 2012), contudo, o foco principal parece ter sido a detecção de rostos, além de terem sido testados em torno de 22 mil categorias de objetos com base em testes padronizados deste tipo de programa. Chama-nos, muita atenção, de todo modo, a imagem por eles divulgada na matéria jornalística, que representaria o modelo elaborado pelo programa em sua concepção genérica de um gato.

Trata-se de uma figuração similar à que encontramos, com menor destaque, no artigo científico produzido pelos cientistas da empresa envolvidos no experimento (LE *et al.*, 2012), representando o que indicam se tratar de uma das técnicas de visualização utilizadas para verificação daquilo que o programa aprendeu a reconhecer. Eles sugerem duas técnicas: a visualização dos estímulos do grupo teste<sup>90</sup> que resultaram em uma melhor resposta (Figura 26); e a visualização daquele que seria o estímulo ótimo, de melhor resposta possível do programa (Figura 27). Com base nesta descrição, podemos compreender que a imagem do gato ilustra uma visualização do estímulo ótimo, embora no artigo – diferentemente do exemplo apresentado à imprensa – o exemplo abordado seja o da detecção de rostos.

Com base nestas imagens, escrevem os autores, pode-se confirmar que o programa testado no experimento “de fato aprende o conceito de rostos”<sup>91</sup> (LE *et al.*, 2012). Podemos compreender, com base na noção do estímulo ótimo – conquanto os autores apontem certas limitações<sup>92</sup> – que nas imagens apresentadas (Figura 25 e Figura 27) encontramos algo próximo daquilo que o programa da *Google* melhor reconheceria como um gato ou como um rosto, levando em consideração todos os rostos e gatos que já lhe foram apresentados. São, evidentemente, estranhas figurações, compostas pela aparente sobreposição de inúmeras imagens que tem como resultado estas composições

---

88 No aprendizado não supervisionado, não são apresentados ao programa exemplos daquilo que ele deve aprender a reconhecer. Diante de uma massa de dados oferecidos para a análise, o algoritmo busca recorrências a fim de identificar elementos passíveis de reconhecimento.

89 O fato do exemplo apresentado, dentre os milhares disponíveis na pesquisa, ser justamente o de gatos – considerando o reconhecido gosto de usuários do *YouTube* por vídeos de gatos fazendo coisas engraçadas – é curiosa, certamente, mas provavelmente deve decorrer de uma seleção deliberada dos pesquisadores, do pessoal de *marketing* da empresa ou dos jornalistas responsáveis pela matéria, diante do potencial de divulgação da pesquisa.

90 Eles indicam como grupo teste uma base de imagens padronizada para o teste deste tipo de programa – oferecendo, assim, uma avaliação comparativa de sua eficiência e eficácia em relação a outros programas similares. Trata-se de amostra distinta do grupo de imagens utilizadas no treinamento, cerca de 10 milhões de imagens do *YouTube*. O grupo de treinamento é utilizado como base para a inferência pelo programa dos padrões recorrentes enquanto o grupo de teste é utilizado para verificar o nível e a velocidade de acerto do programa após ter “aprendido” a reconhecer determinados padrões.

91 No original: “*indeed learns the concept of faces*”. Tradução nossa.

92 Eles sugerem no texto (LE *et al.*, 2012) que a visualização do estímulo ótimo estaria sujeita a problemas de “mínimos locais”. Posto de outra forma, isto significa que haveria dificuldade em precisar qual seria de fato o melhor estímulo por limitações da técnica de visualização, de modo que embora se trate efetivamente de um estímulo com um alto nível de resposta, não há como garantir que se trate do estímulo com o *maior* nível de resposta pelo programa.

acinzentadas e disformes. Talvez possamos compreendê-las como figurações *médias* ou *estatísticas* daquilo que seria um gato ou um rosto. Um *gato estatístico*, talvez, seja aquilo que o programa buscaria ao reconhecer cada gato nos vídeos do *YouTube*. Neste estranho modo de condensar a memória e o tempo – que talvez diga respeito, particularmente, à experiência computacional desta dimensão – por estas imagens talvez cheguemos mais perto de desvendar o enigma de que fala Virilio. Pelos olhos da máquina, nesta atestação aos cientistas da capacidade de aprendizado do programa desenvolvido, somos confrontados, enfim, pela imagem de um conceito: a figura abstrata de um rosto ou um gato genéricos que certamente contrastam-se à forma como geralmente tomamos *uma face* ou *um gato*, em sua singularidade e concretude. Por meio de tal processo de abstração<sup>93</sup>, o programa de Visão Computacional da *Google* extrai da massa de dados visuais singulares alimentados a seu sistema (fotografias, *frames* de vídeos) aquilo que todos os rostos encontrados teriam em comum, suas características definidoras: a *ideia* de rosto, ou *uma* ideia de rosto, talvez pudéssemos dizer.

Assim, inversamente à descrição oferecida por Manovich (1993a, p. 98), sugerindo que a Visão Computacional poderia ser considerada uma tecnologia pautada pelo que chama de um *nominalismo visual*, vemos nestes exemplos uma guinada *idealista* em seu modo de operação – que não anula, evidentemente, o outro princípio definidor. Se nas tarefas de *localização*, de fato, a intrincada relação da imagem com seu referente ganha relevância, nestes outros casos torna-se mais importante a sua relação a um conjunto de outras imagens ou, ainda, à composição abstrata deste conjunto, que definiria aquilo que a máquina compreende como este ou aquele objeto. Há, neste sentido, um deslocamento do lastro pelo qual a imagem seria interpretada, vindo se conectar não apenas a um conjunto de imagens previamente processadas pelo programa mas, também, – e isto ganhará particular relevância em nossa discussão – à projeção que seria elaborada pela combinação destes registros e que, efetivamente, produziriam uma *figuração estatística* e, portanto, *virtual, probabilística*, da visualidade abstrata do objeto reconhecido. Com surpreendente precisão, encontramos no texto Virilio uma descrição que parece se aplicar bem às imagens que encontramos neste exemplos recentes, inclusive na sugestão de que a imagem de síntese – pois, por mais que se parta da base fotográfica para sua elaboração, encaramos, aqui, imagens sintéticas – seria, antes de tudo, uma “imagem estatística”:

Observemos, entretanto, que a *imagem de síntese*, como seu nome indica, é na realidade nada mais do que uma ‘imagem estatística’ que só surge graças aos rápidos cálculos dos PIXEL que compõem o código de representação numérica [...] a habitual crítica do pensamento estatístico gerador de *ilusões racionais* remete, portanto, necessariamente ao que se pode chamar aqui de pensamento visual do computador; a *óptica numérica* sendo nada mais do que uma óptica estatística capaz de gerar uma série de ilusões visuais, ‘ilusões racionais’, elas também afetando não somente o raciocínio, mas o entendimento (VIRILIO, 1994, p. 106).

---

93 O artigo dos pesquisadores intitula-se algo como “Construindo características de alto nível com aprendizado não supervisionado de larga escala” (LE *et al.*, 2012) – lembremos do sentido da expressão *alto nível* no âmbito da computação, que discutimos no capítulo anterior.



Figura 25 – Modelo visual de um gato gerado pelo programa de reconhecimento da *Google*, disponível em tamanho menor na matéria jornalística (MARKOFF, 2012), mas retirada de [http://www.embedded-vision.com/sites/default/files/news/CatDetection\\_resized.jpg?1349732566](http://www.embedded-vision.com/sites/default/files/news/CatDetection_resized.jpg?1349732566).



Figura 26 – Visualização dos 48 estímulos do grupo teste com melhores respostas do programa da *Google* para reconhecimento de rostos, que se valeu de imagens encontradas no *YouTube* como base de treinamento. Fonte: LE *et al.*, 2012.

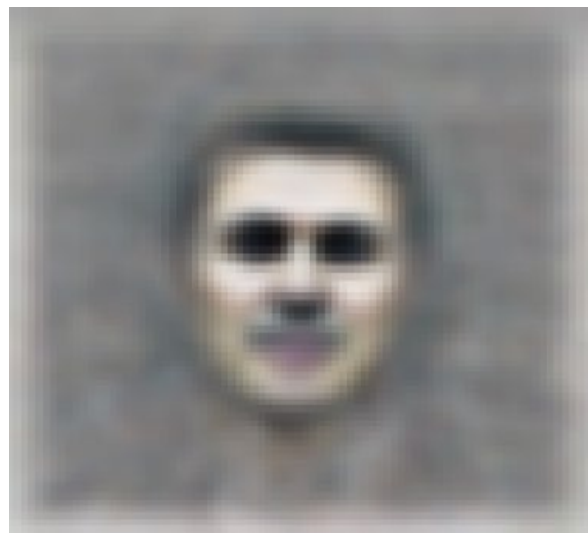


Figura 27 – Visualização do estímulo ótimo do programa da *Google* para o reconhecimento de rostos. Fonte: LE *et al.*, 2012.



O que hoje vemos, nestes subprodutos visuais da percepção da máquina (que talvez não possamos equiparar literalmente à sua *visão subjetiva*, suas *representações internas*, ou suas *imagens virtuais instrumentais* – mas que, certamente, delas se aproxima), parece confirmar a descrição que nos oferece Virilio. Pois se, como sugerem os pesquisadores da *Google*, estamos diante de representações visuais de conceitos como *gato* ou *rostos humano*, não deixa de nos ser estranha tal figuração genérica, cinza e disforme. Sendo compostas, essencialmente, pelo cruzamento estatístico da aparência de inúmeros gatos e inúmeros rostos, vemos, afinal, não *um rosto qualquer*, ou *um gato qualquer* – não é deste tipo de generalidade que se trata. Tampouco vemos *qualquer rosto* ou *qualquer gato*. Vemos aquilo em que se desfaz a aparência de um rosto ou de um gato quando sobrepõem-se vários de seus exemplares – o que resta de semelhante no cruzamento desta diversidade, efetivamente anulando-a. O que resta, evidentemente, não é um rosto humano, nem um gato. Na visualização desta aparência estatística, aquilo que seria um rosto para um computador afasta-se claramente daquilo que seria, para nós, um rosto – e não deixa de ser sugestivo que sejam justo os olhos aqueles que, nesta representação do rosto humano, são excluídos, surgindo como orifícios vazios (Figura 27), como se o rosto ótimo, para a máquina que vê, não tivesse olhos.

Quanto às “ilusões racionais”, de que fala Virilio (1994, p. 106), talvez um modo de as compreendermos passe pelo processo de aprendizado da máquina, que leva à produção de tais representações. Nisto, chegamos talvez ao cerne do problema do *idealismo visual* – se assim podemos nomear esta óptica estatística que guia a Visão Computacional. Como já indicamos, o aprendizado de máquinas parte, fundamentalmente, do processamento de uma ampla base de dados (pré-selecionada ou não – a depender de se tratar de método *supervisionado* ou *não supervisionado*) a partir dos quais o programa abstrai, tentativamente, modelos que preveriam os traços gerais do padrão reconhecido, a fim de identificá-lo em outras ocorrências. O caráter possivelmente ilusório do modelo assim gerado, contudo, reside não apenas no possível arranjo tendencioso (ainda que de modo não intencional) da base de dados utilizada, mas também no tipo de previsão que o modelo gerado seria capaz de subsidiar. Já no exemplo do rosto que apresentamos, percebe-se como o *estímulo ótimo* ao programa de detecção de rostos elaborado sugere uma figura masculina, adulta e branca. Não é menos relevante, inclusive, que a imagem daquele que chamamos de *gato estatístico* – que antecipa aplicações bem menos ofensivas – tenha sido escolhida para a divulgação ampla da pesquisa fora do circuito acadêmico. Tendo em vista as conhecidas aplicações contemporâneas dos métodos estatísticos no âmbito do chamado *Big Data*, ou no *regime de eficiência informacional*, como propõe Fernanda Bruno (2013), não é difícil, inclusive, imaginarmos de que modo este mesmo método poderia ser utilizado, potencialmente, em bases de rostos de jovens infratores, por exemplo – tampouco nos é difícil imaginar, no contexto brasileiro, que tipo de perfil seria sugerido pela aparência do *estímulo ótimo* de um tal programa. Noutros contextos, poderíamos especular o mesmo quanto a perfis de imigrantes ilegais, suspeitos de terrorismo, entre outros grupos. Enquanto registros “objetivos” e “racionalis”, pois estatísticos, não é difícil imaginarmos como uma tal prática, ainda que hipotética (o que, contudo, não significa que já não seja praticada) poderia levar ao

recrudescimento de práticas racistas correntes, uma vez que fundamentadas em *dados concretos* – ainda que apenas ilusoriamente uma tal fundamentação pudesse ter qualquer lastro ético, e nisto vem bem ao caso a formulação de Virilio, “ilusões racionais” (1994, p. 106).

Tal aplicação, para o contexto da criminologia e da criminalística<sup>94</sup>, inclusive já teria surgido anteriormente, no século XIX, como indicamos na seção anterior, ao convocarmos o trabalho de Alphonse Bertillon. Allan Sekula (1986) sugere que, neste e em outros exemplos, observamos a interseção entre a imagem e o pensamento estatístico: “esta mescla de óptica e estatística foi fundamental para a mais ampla integração dos discursos da representação visual e aqueles das ciências sociais no século XIX”<sup>95</sup> (SEKULA, 1986, p. 17). A peculiaridade do problema enfrentado, gerando a necessidade de adequação da fotografia à arquivística, se deve à própria natureza da fotografia, como toda imagem, que não permite sua organização na forma análoga a um dicionário, por exemplo, segundo unidades léxicas discretas (SEKULA, 1986, p. 17). Para organizar o crescente arquivo de imagens então produzidas, fez-se necessário desenvolver outras formas de análise das imagens a fim de produzir alternativas à sua indexação. O autor indica que dois caminhos particulares poderiam ser observados na realização desta tarefa: um se voltaria para a tradução do caráter circunstancial ou idiossincrático da fotografia em figuras típicas ou emblemáticas; outro buscaria preservar o caráter particular de cada imagem única, elaborando um modo de resgatá-la de uma quantidade massiva de fotografias. De modo similar ao que vimos indicando, Sekula sugere que o primeiro caso se aproximaria de uma abordagem “realista”, no sentido que lhe atribui a filosofia medieval, baseada na verdade de proposições gerais e *tipos* (o que indicamos como *idealista*); o segundo, ele sugere que se aproximaria de uma abordagem “nominalista”, que recusaria categorias genéricas, da forma como tomamos de Lev Manovich. Por tal distinção, cada método atenderia a finalidades específicas, seja para se chegar ao tipo criminal característico – atendendo, assim, a fins científicos –, seja para se chegar àquele criminoso específico – atendendo, assim, a fins técnicos e policiais.

O caso mais exemplar da segunda abordagem, nominalista, é aquela de Alphonse Bertillon e sua sinalética, voltada para a identificação criminal. O exemplo trazido por Sekula para a abordagem que chama de *realista*, contudo, aproxima-se das representações que tomamos como exemplares na presente seção: trata-se do trabalho de *retratos compostos* realizado pelo estatístico britânico Francis Galton (mais conhecido como o pai da *eugenia*) em sua busca pela aparição visual do “tipo criminal” (Figura 28 e Figura 29). Como descreve o autor:

Através de uma de suas várias aplicações dos retratos compostos, Galton tentou construir uma aparição *puramente óptica* do tipo criminal. Esta impressão fotográfica de uma face criminal abstrata, estatisticamente definida,

---

94 Conforme Sekula, haveria uma diferença de terminologia por se observar: “Criminologia caçava ‘o’ corpo criminal. Criminalística caçava ‘este’ ou ‘aquele’ corpo criminal” (SEKULA, 1986, p. 18). No original: “*Criminology hunted ‘the’ criminal body. Criminalistics hunted ‘this’ or ‘that’ criminal body*”. Tradução nossa.

95 No original: “*This merger of optics and statistics was fundamental to a broader integration of the discourses of visual representation and those of the social sciences in the nineteenth century*”. Tradução nossa.



Figura 28 – Retratos compostos por Francis Galton exibindo características prevaletentes em condenados por furto sem uso de violência. Fonte: <http://galton.org/>.



Figura 29 – Retrato composto de um “criminoso médio”, elaborado ao modo de Francis Galton. Fonte: ELLIS, Havelock. *The criminal*. London: Walter Scott, 1890. *apud* SEKULA, 1986, p. 42.

e empiricamente não existente foi ao mesmo tempo a mais bizarra e a mais sofisticada de muitas tentativas concorrentes de arranjar evidências fotográficas na busca pela essência do crime<sup>96</sup> (SEKULA, 1986, p. 19).

Tanto por seus objetivos quanto por seu método, Galton aproximava-se da visualidade que encontramos como subproduto do processo de aprendizado de máquinas utilizado em aplicações contemporâneas da Visão Computacional. Embora, talvez em função da escala reduzida do conjunto de imagens utilizada em suas composições, os traços figurativos sigam ainda bem definidos, não se distanciando demasiado daquilo que compreenderíamos como *uma* fotografia, as composições galtonianas trazem alguns dos aspectos que encontramos nas figurações médias que abordamos: contornos difusos, baixo contraste. Nelas encontramos um primeiro exemplo do que talvez poderíamos chamar de um olhar indutivo diante da fotografia, buscando o geral a partir de um conjunto disperso de registros particulares.

Ter como precursor o idealizador da eugenia, em aplicações voltadas para a construção “racional”, pois estatística, da aparência típica do criminoso, talvez já nos servisse de alerta claro o suficiente para os riscos desta visualidade. Se partimos, entretanto, de um subproduto dos programas de Visão Computacional – vale dizer, a produção destas imagens médias não é a finalidade de seu desenvolvimento –, isto não diminui a importância de observarmos tal registro nem suas implicações para o panorama mais geral desta tecnologia. Este ponto de interseção se oferece como um importante elo para compreendermos algumas de suas heranças, bem como alguns de seus aspectos talvez impensados, ou recalçados. Com frequência avaliadas apenas por seu grau de eficácia e eficiência, os altos índices de acerto deste ou daquele método de reconhecimento podem, com frequência, obscurecer os princípios envolvidos em seu funcionamento, ou envolver-nos, como sugere Virilio, em uma “ilusão racional”. A aproximação da Visão Computacional das composições fotográficas de Galton revela-nos, em certa medida, de onde partem seus métodos e aonde podem eventualmente nos levar. Esta espécie de virada idealista aplicada à imagem fotográfica por uma tal abordagem pelos programas de Visão Computacional – herança de Galton – articula-se intimamente, como também pudemos ver, a uma lógica nominalista, de identificação e atuação sobre o espaço e os corpos – herança de Bertillon – e é na inter-relação entre estes movimentos, do singular ao geral e de volta ao singular, que talvez encontremos um traço fundamental e a força deste desenvolvimento. Tais figurações *médias*, portanto, mais além de poderem em alguma medida refletir algumas das práticas discriminatórias já instaladas, não o fariam sem reforçá-los, inclusive ao dar a ver a imagem deste preconceito. Mais ainda, elas se tornam no âmbito da Visão Computacional ainda um índice que, a meio caminho entre memória e previsão, orienta modos de percepção de corpos singulares, segundo sua aproximação ou não dos modelos assim constituídos. Neste sentido, a Visão Computacional vem resolver simultaneamente os problemas de Bertillon e Galton: elaborando, ao

---

96 No original: “*Through one of his several applications of composite portraiture, Galton attempted to construct a purely optical apparition of the criminal type. This photographic impression of an abstract, statistically defined, and empirically nonexistent criminal face was both the most bizarre and the most sophisticated of many concurrent attempts to marshal photographic evidence in the search for the essence of crime*”. Tradução nossa.

mesmo tempo, modelos gerais, perfis de risco (criminologia) e oferecendo-se enquanto instrumento para a identificação e a localização de corpos singulares, específicos (criminalística). Assim, talvez devamos ver com certo alarme esta *imagem virtual instrumental* que nos é revelada, refletindo sobre as implicações de construções similares em outros contextos. Conquanto seja evidente que seguem havendo limitações ao grau de eficácia das técnicas de reconhecimento automatizado aplicadas às imagens – como argumenta, inclusive, Fernanda Bruno (2012) – talvez não seja o caso de argumentar pela impossibilidade de algoritmos verdadeiramente eficientes, entre outras razões, pelo risco de que tal defesa seja tomada, por alguns, como desafio.

\*

Quanto ao que viemos chamando, hipoteticamente, de uma *visão subjetiva* da máquina, diante destes exemplos, talvez seja o caso de revisarmos aquilo que pretendíamos como uma tal denominação. A analogia proposta por Virilio (1993) entre as *imagens mentais* e as *representações sintéticas internas* que a máquina elaboraria em seu processo de interpretação sugere-nos tal formulação na medida em que haveria uma proximidade de natureza entre as elaborações mentais e técnicas decorrentes dos processos perceptivos: ambas constituídas majoritariamente por uma virtualização do percebido e inacessíveis a um observador externo. A figuração estatística a que chegamos, portanto, não seria muito mais do que uma representação daquilo que, no âmbito da máquina, poderíamos compreender como um conceito, ou como uma ideia. Posto de outra forma, trata-se da figuração do elemento que se espera que seja ativado quando à máquina se ofereça a imagem de um rosto, ou de um gato, para que interprete. Se na seção anterior indicamos a presunção de reversibilidade entre a imagem e seu referente como premissa necessária para a Visão Computacional, nesta talvez poderíamos indicar a existência de uma segunda premissa: a da reversibilidade entre o que, em termos da semiótica peirciana, diríamos se tratar do signo (a imagem) e seu interpretante – que, como descrito por Júlio Pinto (1995, p. 29), poderíamos compreender como o terceiro termo “que se produz da relação do signo com seu *objeto*”. A *imagem virtual instrumental* ou *mental*, como nos sugere Virilio (1993), seriam os *interpretantes* supostos no processo de interpretação realizado por um observador ou por uma máquina de visão.

Se por meio da visualização do *estímulo ótimo* os pesquisadores da *Google* verificam a apreensão pelo programa de determinado conceito, espera-se que haja algum nível de correspondência entre aquilo que ele aprendeu a reconhecer e a representação que se produz por tal técnica de visualização. Há, efetivamente, a representação algo abstrata de gato ou de rosto humano nas imagens apresentadas, que nos oferecem a imagem condensada dos gatos e rostos singulares que a máquina elaborada pelos cientistas da empresa processou durante seu treinamento. Embora trate-se de uma representação em constante mutação – cada novo rosto humano ou gato encontrado potencialmente transforma a representação, interna ao programa, destes conceitos – em cada estágio desta operação, por mais fugaz que ele possa ser, haveria uma correspondência probabilística na operação de reconhecimento,

pela qual diante da imagem de um objeto apresentada ao programa haveria um indicador preciso acerca da proximidade deste exemplar do modelo ótimo elaborado internamente à máquina. O programa será tão mais eficaz quanto objetos pertencentes a uma determinada categoria ativarem a parte do programa responsável por detectá-la, ou seja, será tão melhor quanto elaborar modelos genéricos capazes de abarcar a diversidade de objetos pertencentes àquela categoria. Fica evidente, por tal formulação, o efetivo caráter hipotético de nossa sugestão de uma tal representação como a *visão subjetiva* da máquina – abordaremos mais profundamente as implicações desta elaboração metafórica no capítulo seguinte.

Quanto à questão que lançamos para este capítulo, o que veem as máquinas, tendo já percorrido o trajeto pendular que sugerimos ao início, a partir de Peter Galison (2004), *imagem-dados-imagem*, talvez já nos seja dado sugerir o efetivo papel preponderante da imagem enquanto um ponto de retorno necessário em nossa compreensão da Visão Computacional. De um lado, como pudemos explorar, pelo modo como já estaria inscrito no registro particular das imagens de base fotográfica, produzidas pelo mecanismo óptico das câmeras, a sua validade como instrumento pragmático de conhecimento e ação sobre o mundo. De outro, pelo modo com que a “memória” ativada pela máquina, em sua percepção particular do visível, seria composta pelos arquivos de imagem hoje amplamente disponíveis, através dos quais não apenas aprende a ver e elabora modelos abstratos daqueles objetos que *reconhece*, como, também, deles elabora a representação interna daquilo que vê. Conquanto neste nível nos encontraríamos necessariamente em um estágio já de profunda desintegração do visível, pudemos perceber como seria possível apreendermos, pelo olhar, diante dos exercícios de visualização abordados, a estranha figura de um conceito genérico que seria elaborado, em última medida, por uma *visualidade estatística*.

Neste percurso em que indicamos, inclusive, as heranças desta visualidade em práticas anteriores à Visão Computacional, nomeamos também, e vale recuperarmos, a particularidade dos desenvolvimentos atuais que confeririam sua relevância e sua diferenciação em relação aos anteriores. Verificamos como desenvolvimentos contemporâneos recodificam práticas de análise e síntese do visível que alternam entre a circunscrição de corpos singulares e a sua generalização em contextos abstratos. Contudo, como pudemos indicar, a Visão Computacional permite uma mais íntima associação entre estes movimentos, reunindo, em uma única visada, abordagens *nominalistas* e *idealistas* do visível. A identificação de um corpo ou objeto articula-se, afinal, neste contexto, íntima e diretamente a um movimento de generalização pelo qual estes mesmos corpos e objetos são reconhecidos por determinadas categorias, *tipos*, perfis ótimos. Por sua integração aos sistemas de acelerada produção e circulação de imagens, o dispositivo que abordamos potencializa-se pela sua capacidade de mobilizar o visível em uma rede na qual tanto se estratificam conjuntos, categorias arquivísticas e conceitos, quanto atravessam vetores transversais de conexão que lhe permitem constantemente reordenar e reconfigurar os conjuntos. Neste sentido, a Visão Computacional traria outras diferenças fundamentais por sua articulação a demandas institucionais e configurações

materiais e simbólicas distintas: já não parte de imagens tomadas estritamente para fins criminais ou científicos, mas de virtualmente todas aquelas que circulam livremente pela rede. Também, já não dependem de técnicas e processos orientados e circunscritos pelo dispositivo do arquivo. De outro modo, ele se elabora e ganha consistência justamente por ter já inscrito, em seu funcionamento, a lógica e a topologia da *rede*. Trata-se, assim, de dispositivo chave para que compreendamos possíveis modos de absorção desta proliferação, desta hipertrofia do visível, que hoje habitamos. Garantindo-lhe *eficiência informacional*, no sentido sugerido por Fernanda Bruno (2013), a Visão Computacional conferiria a possibilidade de manejar este fluxo incessante, alternando entre o singular e o geral. Entre o visível, concreto, e o abstrato.

Virilio (1993, p. 129) sugere que as máquinas de visão suplantariam nossa *profundidade de tempo*. Neste sentido, buscariam principalmente absorver as indeterminações próprias à fisiologia da visão humana, com suas recuperações irregulares e incertas da memória e com seu déficit de atenção inerente, considerado patológico e tornado endêmico na contemporaneidade, entre outras inconstâncias. O processo perceptivo eficaz, eficiente e quantificável que elas buscam realizar viriam, então, suplantá-los em um trabalho que nosso olhar já não daria conta: de observar, monitorar, analisar e indexar as imagens hoje em acelerada circulação. O que, então, veem, é a sobreposição e o jogo constante entre estas instâncias, mais rapidamente do que nós mesmos seríamos capazes – o que algumas das imagens que pudemos trazer à discussão talvez apenas comecem a ilustrar. Em todo caso, é apenas na precisa remissão ao já visto, pelo reconhecimento, que uma tal visão se faz possível, em uma constante releitura e atualização de um passado, ou – mais grave – de uma virtualidade futura, sobre o presente visível.



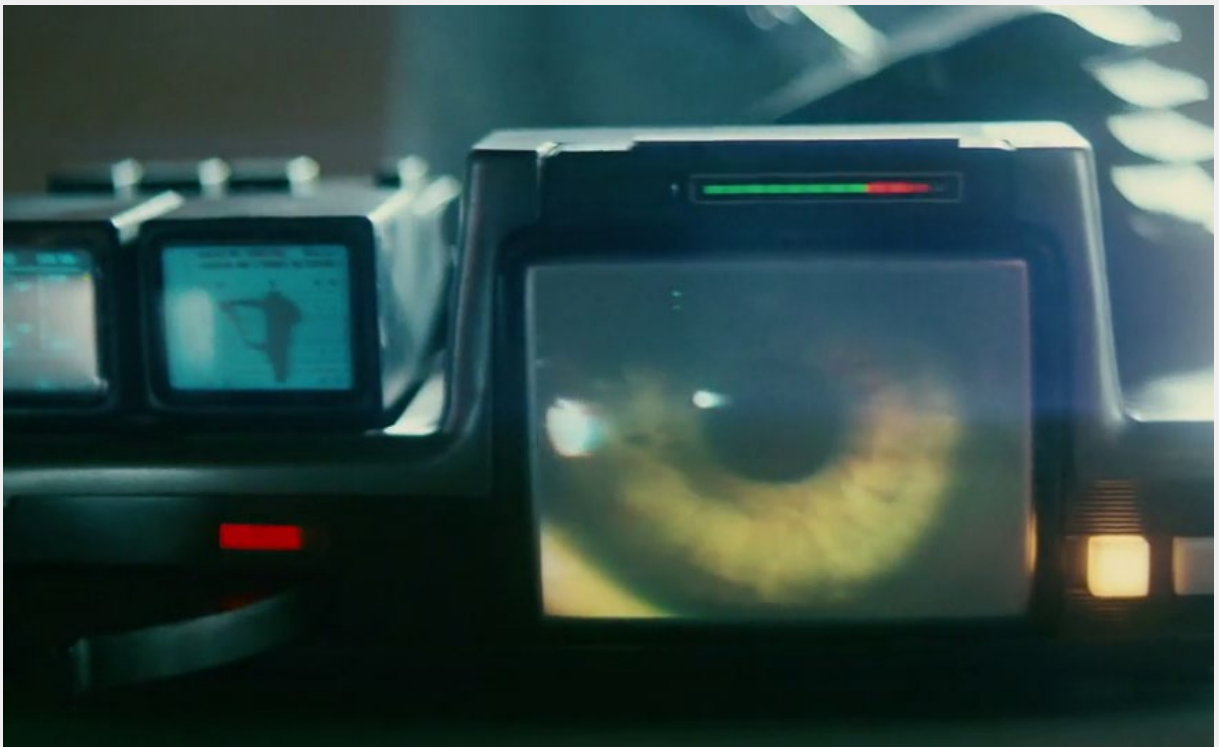


Figura 30 – *Frame de Blade Runner*, de Ridley Scott: o teste *Voight-Kampff*. Fonte: BLADE, 2014.



## 3 Nós, que vemos

### 3.1 Quem vê?

Cientistas do Instituto de Neurociências Helen Wills da Universidade da Califórnia em Berkeley divulgaram no ano de 2011 resultados de um de seus estudos mais recentes no âmbito da reconstrução da experiência visual (NISHIMOTO *et al.*, 2011; ANWAR, 2011). Posto de outro modo, um experimento na área que talvez chamaríamos de “leitura da mente”: a partir do monitoramento da atividade cerebral de um indivíduo por ressonância magnética, em resposta a determinado estímulo visual, a pesquisa voltou-se para a possibilidade de reconstrução de tal estímulo, oferecendo, assim, um acesso às suas *imagens mentais*. Aplicado em três sujeitos, o método do estudo consistiu, grosso modo, no desenvolvimento de um sistema computacional de aprendizado de máquinas supervisionado para gerar, para cada sujeito, um modelo de correspondência entre uma base de estímulos conhecida e os registros de circulação e oxigenação sanguínea no cérebro. O modelo de correspondência gerado por tal processo foi então testado a partir de uma segunda base de estímulos, sendo que, desta vez, o programa computacional buscou reconstruir um estímulo ao que o sujeito fora exposto, tendo acesso apenas ao registro da atividade cerebral resultante, o que realizou a partir da combinação e composição visual do repertório de imagens contido na base usada no treinamento. Segundo o mesmo grupo de cientistas, em publicação anterior (NASELARIS *et al.*, 2009, p. 902), a pesquisa desenvolvida por eles visa ultrapassar a abordagem mais difundida que realizaria uma *classificação* da experiência registrada pela atividade cerebral, realizando, de outro modo, sua *reconstrução*, apresentando não apenas um conjunto de categorias que descrevam a experiência do sujeito, mas sua efetiva representação visual. Vale ressaltar que a técnica de aprendizado de máquinas empregada, tanto para o caso da *classificação* quanto da *reconstrução*, é baseada fundamentalmente nos mesmos princípios daquelas que descrevemos anteriormente, configurando uma aplicação particular da Visão Computacional, uma vez que busca-se uma correlação possível para a tradução de uma imagem aos dados de atividade neural, e vice-versa. No caso da reconstrução da experiência visual, em que o modelo de correspondência é utilizado para produzir uma representação visual dos dados neurais registrados, tal relação torna-se ainda mais intrincada.

Um aspecto deste experimento que contrasta com outros similares realizados por estes e outros cientistas diz respeito à natureza dos estímulos apresentados aos indivíduos. Enquanto em experimento anterior (NASELARIS *et al.*, 2009) o grupo de pesquisadores se valeu de fotografias monocromáticas extraídas randomicamente de “coleções fotográficas”; e enquanto estudo similar de um instituto francês (THIRION *et al.*, 2006) valeu-se de padrões gráficos sintéticos elaborados especificamente para o seu experimento; o estudo realizado pelos cientistas de Berkeley usou como

estímulo (tanto para o treinamento do programa quanto para os testes posteriores) trechos de *trailers* de filmes – coloridos e moventes – encontrados *online* no *YouTube* e na galeria de *trailers* da *Apple* (NISHIMOTO *et al.*, 2011, p. 1645), aos quais os cientistas se referem pela estranha categoria de “filmes naturais” (“*natural movies*”). Foram extraídos trechos randômicos destes vídeos, totalizando duas horas, aos quais foram expostos os sujeitos enquanto eram coletados os dados para treinamento do sistema computacional. Nesta fase, o programa elabora um modelo de correspondência entre o estímulo conhecido e a atividade cerebral registrada. Outros nove minutos deste material, não coincidentes com as duas horas da fase anterior, foram apresentados como estímulos na fase de teste. Nesta etapa, o programa buscou inferir o estímulo que teria sido apresentado ao sujeito com base apenas na atividade cerebral registrada (NISHIMOTO *et al.*, 2011, p. 1641). O resultado da reconstrução é alcançado pela sobreposição e mescla das imagens do repertório de treinamento que desencadearam respostas na atividade cerebral mais próximas àquelas observadas na fase de teste, em cada fração de tempo considerada para análise (Figura 31 e Figura 32).

Trata-se de um registro cuja similaridade aparente àquelas estranhas figurações estatísticas que discutimos ao final do capítulo anterior vem reforçar a também evidente similaridade entre os métodos que as geraram. Tanto lá quanto aqui, são empregados processos de interpretação automática de imagens articulados ao aprendizado de máquinas com o objetivo de elaborar modelos operacionais para o reconhecimento de padrões recorrentes. Também, tanto lá quanto aqui, são elaboradas visualizações dos padrões reconhecidos a partir da composição de imagens similares encontradas no repertório visual utilizado no treinamento dos *software* – tal repertório, em ambos os casos, constituído principalmente por imagens encontradas na *web*, particularmente no *YouTube*. A principal diferença a observarmos entre os resultados alcançados, contudo, é que enquanto as figurações do gato e do rosto resultantes do processo de aprendizado do programa da *Google* (Figura 25 e Figura 27) são representações daquilo que seriam o estímulo ótimo para que fossem reconhecidos como tais pelo sistema, as figurações apresentadas por esta outra pesquisa são reconstruções estimadas daquela que teria sido a experiência visual dos sujeitos submetidos ao experimento. É, então, no mínimo curiosa a aparente coincidência que se manifesta entre o modo de representação das *imagens mentais* destes sujeitos e a representação daquelas que chamaríamos, no vocabulário de Virilio (1993 e 1994), as *imagens virtuais instrumentais* elaboradas pela máquina em seu processo perceptivo. Reforçando, em um primeiro momento, a analogia proposta pelo filósofo, vemos na correspondência entre estes exemplos a clara aproximação entre representações internas da visão humana e das máquinas de visão.

Destaca-se, contudo, que, longe de coincidente, poderíamos creditar a similaridade entre os registros alcançados pelas pesquisas não apenas ao recurso comum a uma base de imagens encontradas em repositórios *online* mas, principalmente, ao fato de que se trata, em ambos os casos, de aplicações da Visão Computacional em suas operações de *reconhecimento–conexão*, desenvolvidas por métodos de aprendizado de máquinas. Ainda que os neurocientistas de Berkeley voltem-se para as respostas

cerebrais à percepção visual, é através de máquinas de visão que eles elaboram não apenas os modelos de correspondência entre estímulo visual e atividade cerebral, mas também as representações visuais dos estímulos reconstruídos. O recurso de ambas pesquisas a uma base de imagens comum seria, neste sentido, uma provável decorrência da oferta, nestes repositórios, de uma grande quantidade de conteúdos amplamente disponíveis e de acesso facilitado. Utilizá-lo, portanto, atende à demanda do método que têm em comum, dependente de um grande volume de dados para o treinamento dos programas *inteligentes*. Assim, a sugestão de Virilio de que as máquinas de visão recolocariam a questão da *atualidade* das *imagens mentais*, através de sua analogia às *imagens virtuais instrumentais* das máquinas, parece se desdobrar em uma ocorrência concreta desta articulação: por meio da Visão Computacional vemos nos registros alcançados pelos cientistas, em seu esforço de reconstrução da experiência visual, uma tentativa de *atualização* destas representações internas, *virtuais*, da percepção humana. Trata-se, ainda, de uma manifestação da articulação que viemos sugerindo, desde o início, entre a Inteligência Artificial, da qual a Visão Computacional seria um de seus subdomínios, e as Ciências Cognitivas – pela qual desenvolvimentos de ambas se revelariam complementares pela tomada da informação como termo de passagem entre computação e cognição.

Nesta correspondência, explicitada tanto pelos métodos quanto por seus resultados, insinua-se a questão que tomamos como mote desta terceira etapa de nossa investigação, que se dirige ao modo de constituição de seu *observador*. Como sugerimos no primeiro capítulo, a partir da discussão da abordagem por Dubois (2004, p. 31-67) do eixo maquinismo–humanismo, uma primeira conclusão, prematura, a que poderíamos chegar é a de que o Sujeito seria um termo ausente da constituição deste domínio em particular. Afinal, se delegamos às máquinas de visão o ato de ver, como falar de um *observador*, humano, que poderia ser por elas constituído? Talvez – prosseguiria uma tal conclusão – nesta delegação nos ausentamos do domínio do visual e, assim como as imagens sintéticas já *nada* diriam do Real, tampouco a visão sintética diria de um Sujeito, ainda que hipotético, que pudesse ser confrontado por suas imagens. Trata-se, contudo, de uma compreensão que ao invés de enfrentar os termos problemáticos do objeto abordado, resolve suas tensões pela postulação de uma ausência. Mais ainda, se levamos adiante as implicações de uma tal formulação, encontramos um sistema em circuito hermeticamente fechado, incomunicável com seu exterior, que não apenas é improvável quanto relegaria tais máquinas a uma condição de insignificância, já que nada afetariam e apenas diriam de si mesmas. No capítulo anterior buscamos, a este respeito, enfrentar a questão da imagem que seria, a princípio, oclusa pelo funcionamento da Visão Computacional. Buscamos, neste capítulo, aprofundarmo-nos na questão do *observador* implicado por tal tecnologia, recusando uma leitura mais corriqueira que sugeriria sua inexistência, ou total conversão em máquina. O exemplo que trazemos para abrir esta etapa de nossa dissertação tem particular relevância para tal questão na medida em que revela o modo pelo qual o desenvolvimento da Visão Computacional vem implicar, inerentemente, sujeitos e *observadores* humanos – para além de seu aspecto mais autoevidente de elaboração de uma espécie de *observador-máquina*. Trata-se, neste sentido, de um exemplo extremo em que a visão humana seria *literalmente* explicada pela Visão Computacional, não

apenas em nível teórico, mas também empírico, encaminhando a resultados e visualizações análogas entre o humano e a máquina. Embora já tenhamos indicado a aplicabilidade pressuposta do campo da Visão Computacional tanto à visão humana quanto à visão da máquina, talvez tenhamos apenas no exemplo trazido a este capítulo uma ilustração mais clara da efetuação desta passagem.

Haveria, em uma analogia histórica, certa ressonância no estudo realizado em Berkeley de uma pretensão científica já há muito manifesta em casos anteriores, do final do século XIX, de acessar imagens internas da percepção, registradas como que fotograficamente na retina de corpos recém-falecidos. Gunning (2004, p. 57-59) e Dubois (2012, p. 231-234<sup>97</sup>) contam-nos como, em uma analogia direta à câmera fotográfica, foi proposto por um Dr. Vernois e um Dr. Bourion, da Sociedade de Medicina Legal de Paris, o *optograma*, que consistia na técnica de revelação das últimas imagens percebidas por vítimas de assassinatos a partir de suas retinas, removidas pouco após a morte (Figura 33). Numa efetivação prática da analogia câmera-olho – em sua tomada ao pé da letra, como sugere Dubois (2012, p. 233) –, o olho humano é tido como uma câmera e sua retina, como uma película sensível. Gunning (2004, p. 58-59) indica como tal formulação possuiu ressonâncias na narrativa ficcional, enquanto Dubois (2012, p. 231-234) recompõe seu desenvolvimento no âmbito criminal e suas implicações para uma compreensão da imagem fotográfica. Trata-se, sugere ele (DUBOIS, 2012, p. 233), de uma sugestão fantástica, logo descartada de sua validade científica, que teria sido fruto não apenas da já referida leitura literal da metáfora câmera-olho, mas de um desejo irresistível de encontrar esta “impossível representação de um irrepresentável”, de “entregar ao visível o próprio instante do apagamento do olhar” (DUBOIS, 2012, p. 233-234).

Dubois (2012, p. 233) indica como o optograma de Vernois e Bourion realizaria na prática uma formulação hipotética de Descartes em seu tratado *Dióptrica*, no qual discorre acerca de tópicos relacionados à óptica e à visão humana, com especial atenção para a analogia entre esta e a câmara escura<sup>98</sup>. Em trecho do texto citado por Dubois (DESCARTES, s/d, p. 205-206<sup>99</sup> *apud* DUBOIS, 2012, p. 233), o filósofo francês sustenta que a analogia entre o olho e a câmara escura poderia ser verificada pela extração e inspeção do olho de um cadáver no interior do qual seria possível encontrar, como uma pintura (lembramos, Descartes não conheceu a fotografia), uma representação acurada, em perspectiva, dos objetos exteriores. O optograma atualizaria esta metáfora ao somar-lhe a inscrição fotográfica, como sugere Dubois, pois, neste caso, a analogia se desdobra: “deve-se ver [...] que essa intimidade quase fisiológica entre o órgão e o dispositivo técnico encontra-se de certa maneira redobrada na experiência do Dr. Bourion: o olho ‘fotografa’ o real decerto, mas é a fotografia que revela a imagem do olho” (DUBOIS, 2012, p. 233). Ou seja, não é apenas que o olho funciona *como* uma câmera, mas ele também se presta ao método fotográfico, que seria capaz de revelar seu

---

97 Dubois (2012, p. 231) faz referência a obras de Didi-Huberman que teriam tratado deste tema, às quais não conseguimos acesso: DIDI-HUBERMAN, Georges. *L'optogramme (l'arrêt sur la dernière image)*. *Revue Belge du Cinéma*, n. 4, (Filmes de photo), Bruxelas, APEC, verão 1983. p. 29-34. Ele indica haver também um curta-metragem dirigido pelo autor e denominado *L'optogramme*, de 1983.

98 Crary (1992) também retoma as formulações de Descartes neste tratado.

99 DESCARTES, René. *La dioptrique*. Quinto discurso. Paris: Gallimard, s/d. p. 205-206.

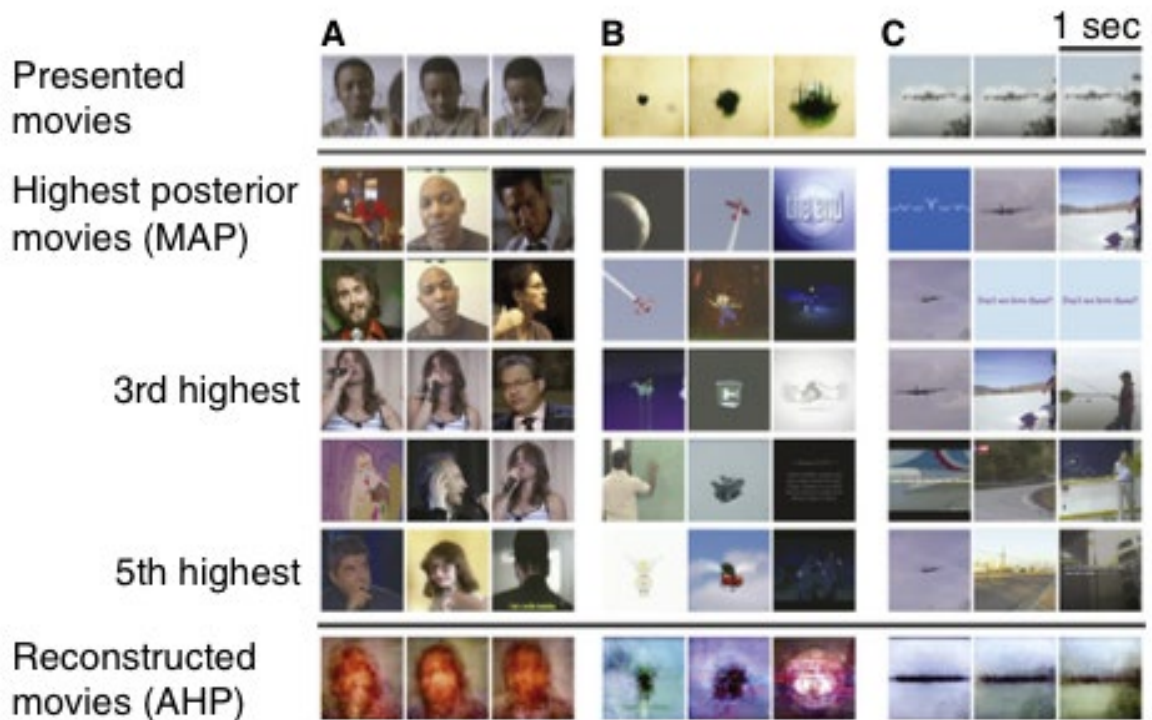


Figura 31 – Exemplos de reconstrução dinâmica da experiência visual na pesquisa realizada por cientistas da Universidade da Califórnia em Berkeley. A linha superior – “Presented movies” – apresenta *frames* dos vídeos apresentados como estímulo; a linha inferior – “Reconstructed movies (AHP)” – , *frames* dos estímulos reconstruídos pelo programa; e as linhas intermediárias, indicam as imagens da base de treinamento que seriam as melhores estimativas do programa, dentre as imagens já conhecidas. Fonte: NISHIMOTO *et al.*, 2011, p. 1644.

Presented clip



Clip reconstructed from brain activity



Figura 32 – *Frame* do vídeo de demonstração da pesquisa dos neurocientistas de Berkeley (NISHIMOTO *et al.*, 2011) exemplificando a reconstrução realizada. A imagem à esquerda apresenta o estímulo original. À direita é apresentado o estímulo reconstruído a partir da atividade cerebral. Fonte: <https://www.youtube.com/watch?v=nsjDnYxJ0bo>.



Figura 33 – Optograma do Dr. Vernois, revelado a partir da retina de uma mulher assassinada em 14 de junho de 1868. Fonte: DUBOIS, 2012, p. 230.



funcionamento. A efetividade da metáfora é demonstrada e aprofundada através da aplicação da fotografia ao olho, da *revelação* da inscrição feita na retina, que é extraída do globo ocular como se extrai uma chapa do interior da câmera. Embora, talvez, sem a mesma dimensão fantástica – mas nem por isto menos informado por um desejo e uma crença comuns – a reconstrução da experiência visual por meio da articulação de ressonância magnética e algoritmos de aprendizado de máquina realiza, como os experimentos de Vernois e Bourion no século XIX, a aplicação, ainda que atualizada, de uma mesma analogia. Em um momento histórico em que a fotografia já é claramente insuficiente enquanto metáfora elucidativa da visão, não deixam de surgir outras figurações aludindo a seu modo de funcionamento, por meio de distintas matrizes tecnológicas, em que tomam parte, de modo mais proeminente, o cálculo, a estatística e a conexão. De modo característico, aquilo que vemos já não é representado por meio de uma inscrição fotográfica, indicial (ainda que falsa), mas por uma complexa rede de figurações etéreas, colhidas do *YouTube*. Talvez por meio deste exemplo comecemos a perceber aspectos do lugar ocupado pela Visão Computacional como um dispositivo relevante para a constituição do *observador* contemporâneo.

Nos textos do historiador da arte Jonathan Crary (1992), o *observador* surge como uma figura conceitual central. Estudando dispositivos ópticos como o estereoscópio e o praxinoscópio, entre outros exemplares novecentistas, o autor desenvolve a compreensão de que cada um deles poderia elucidar paradigmas explicativos da visão em seus períodos, dada sua compreensão como pontos de encontro e interseção de diversos vetores que em cada momento informariam e deslocariam não apenas os modelos explicativos, como também as práticas de observação e as técnicas de regulação da percepção. Como indica o autor (1992, p. 8), não seria o caso de buscarmos posições correspondentes de objetos técnicos distintos em seus respectivos momentos históricos – nossa proposição não é, então, a de que a Visão Computacional estaria para a contemporaneidade como a câmara escura para os séculos XVII e XVIII; ou como o estereoscópio, o fenacistocópio ou a própria fotografia poderiam estar para distintos momentos do século XIX. Porém, ainda que por um modo de articulação particular, seria possível afirmar que cada objeto técnico inscreve-se necessariamente no amplo arranjo discursivo, político, econômico, epistemológico e estético de que tanto participa quanto emerge. Um fator de diferenciação seria, neste sentido, o modo como cada dispositivo configuraria, em cada caso, uma diferente economia destes componentes em seu arranjo particular – assim necessariamente impossibilitando qualquer correspondência analógica simplificada. Guardadas tais considerações, portanto, buscamos na Visão Computacional um mesmo valor paradigmático pelo qual ela nos auxiliaria a compreender e descrever a constituição de uma *visualidade* contemporânea, ainda que seus modos de articulação e incidência sejam distintos daqueles traçados por Crary em relação a seus objetos.

Nesta discussão, a questão do *observador* e sua articulação aos dispositivos técnicos ganha particular relevância à luz do modo como Crary define esta figura conceitual, em uma passagem bem conhecida:

Embora obviamente alguém que vê, um observador é, de modo mais importante, alguém que vê em um arranjo prescrito de possibilidades, alguém que está inserido em um sistema de convenções e limitações. E por ‘convenções’ eu quero sugerir bem mais do que práticas representacionais. Se pode ser dito que há um observador específico para o século XIX, ou para qualquer período, é apenas como um *efeito* de um sistema heterogêneo irreduzível de relações discursivas, sociais, tecnológicas e institucionais. Não há sujeito observador anterior a este campo continuamente em transformação<sup>100</sup> (CRARY, 1992, p. 6).

Por meio desta definição, Crary leva adiante, para o caso específico dos dispositivos de observação, uma das consequências tidas como necessárias da própria formulação conceitual do *dispositivo* por Foucault, enfatizada por alguns de seus comentadores (DELEUZE, 2005 e 1999; AGAMBEN, 2009): seu efeito de *subjetivação*. Por ele, compreenderíamos um sujeito como instância decorrente da operação de um dispositivo, que a ele seria logicamente anterior. Este é, afinal, o sentido pretendido por Crary quando ele indica que o observador seria sempre um *efeito* do sistema heterogêneo em que se encontra inserido. Com tal formulação, ele busca se desvencilhar de abordagens que tomariam o problema da constituição histórica da visão como passível de estudo apenas pela via da análise de práticas representacionais e, ao mesmo tempo materializar, como ele diz (CRARY, 1992, p. 5), o problema da percepção – esquivando-se, assim, também, de uma abordagem idealista. Indicando o escopo que pretende abarcar com tal conceito, escreve: “Se a percepção ou a visão efetivamente mudam é irrelevante, pois elas não têm história autônoma. O que muda são as forças e regras plurais que compõem o campo em que a percepção ocorre”<sup>101</sup> (CRARY, 1992, p. 6). O *observador*, assim, representa uma figura que embora não seja empiricamente observável – não há um único observador, em nenhum momento histórico – agrega a multiplicidade de fatores que conformam aquele que *seria* um observador em cada momento.

Trata-se, evidentemente, de uma noção que possui suas limitações, como sua insuficiência para dar conta dos sujeitos singulares, que efetivamente povoam os períodos históricos, bem como sua reunião, prioritariamente, de fatores hegemônicos do regime escópico estudado (CRARY, 1992, p. 7). Embora sejam características assumidas por Crary, são estes os tópicos de algumas das principais críticas dirigidas a ele, em diferentes registros<sup>102</sup>. Tal formulação segue nos parecendo interessante,

---

100 No original: “*Though obviously one who sees, an observer is, more importantly, one who sees within a prescribed set of possibilities, one who is embedded in a system of conventions and limitations. And by “conventions” I mean to suggest far more than representational practices. If it can be said there is an effect of an irreducibly heterogeneous system of discursive, social, technological, and institutional relations. There is no observing subject prior to this continually shifting field*”. Tradução nossa.

101 No original: “*Whether perception or vision actually change is irrelevant, for they have no autonomous history. What changes are the plural forces and rules composing the field in which perception occurs*”. Tradução nossa.

102 Crary é tradicionalmente criticado por sua imprecisão ao assumir o *observador* como figura central de sua abordagem sem efetivamente analisar registros históricos de *observadores* concretos, propondo, ao contrário, figuras macro-históricas, sustentadas, sugerem seus críticos, por evidências selecionadas a fim de sustentar suas hipóteses, e marcadas por grandes rupturas que não dariam conta da diversidade que efetivamente comporia cada período abordado. Cf. BATCHEN, 1991 e HUHTAMO, 2013, p. 16-18. William J. T. Mitchell (1994, p. 19-24), em uma crítica mais generosa, reconhece a importância de algumas de suas formulações macro-históricas – valorizando, diferentemente de Batchen ou Huhtamo, sua herança foucaultiana –, mas critica-o por deslizes que pareceriam contradizer seu recurso consciente e em certa medida modesto e autocrítico com respeito às generalizações, ao

contudo, em função da compreensão deste sujeito hipotético como atravessado e permeável ao contexto histórico de que participa, tendo como um ponto de contato fundamental os dispositivos com que toma contato e que, afinal, o produzem. De todo modo, o observador serve-nos enquanto uma categoria para compreendermos os dispositivos que os produzem e os contextos históricos em que se inserem, mais do que sujeitos empíricos, efetivos, que pudéssemos tomar, em sua singularidade, em cada momento histórico. Esta, talvez, seja a consideração que, ausente em Crary, sustenta a validade das críticas que lhe são dirigidas. Portanto, não é demasiado reforçar que ao mencionarmos um observador contemporâneo não buscamos nem subsumir toda a multiplicidade de observadores que poderíamos supor, se tomamos outros dispositivos como ponto de partida; nem projetar sobre sujeitos empíricos, histórica e geograficamente situados, um perfil generalista. O observador, de outro modo, se apresenta como ferramenta conceitual que nos permitirá compreender o lugar reservado aos sujeitos na economia de forças que constitui um dispositivo. Trata-se do modo como o sujeito é tomado neste agenciamento.

Sujeitos empíricos serão, efetivamente, resultado de uma rede mais ampla que não se esgota com apenas um dispositivo, mas que certamente o terá como um de seus componentes. Cabe destacar, ainda, que mesmo o processo de subjetivação que indicariamos como próprio deste ou daquele dispositivo tampouco seria inaugurado por ele – algo que já parece se sugerir quando, nos capítulos anteriores, articulamos a constituição da Visão Computacional com processos historicamente anteriores, constituintes de sua genealogia. Da mesma forma que poderíamos percorrer a genealogia de um dispositivo, também poderíamos sugerir a genealogia de um observador<sup>103</sup>. Com seus antecedentes genealógicos, um dispositivo virá, é claro, sempre atualizar e reconfigurar o contexto do qual emerge, mas nem por isso deixa de ser uma expressão de seu momento, vindo também ao encontro de seus sujeitos. Haveria, neste sentido, certa reciprocidade de modo que ainda que não exista, efetivamente, um espectador cinematográfico antes do cinema, ele começa a nascer bem antes do próprio dispositivo cinematográfico, como se o dispositivo e seu sujeito viessem ao encontro um do outro – sem, é claro, descartarmos em tal descrição a mútua reconfiguração que este contato derradeiro provocaria em ambos.

No extremo, se tomarmos como referência a noção de agenciamento maquínico sugerida por Guattari (2003), nem mesmo hipoteticamente poderíamos sugerir a figura de *um* observador ou de *um* sujeito, senão como ponto de chegada intangível de processos sobrepostos, indiscerníveis e inacabados de subjetivação<sup>104</sup>. Segundo sugere-nos Guattari, seria equivocado se partíssemos de sujeitos formados, de antemão, para então construirmos passagens, meios de comunicação, pelos

---

extrapolar e sugerir grandes quebras que afetariam todos os estratos da realidade histórica, com um embasamento empírico limitado.

103 Tal é, por exemplo, a abordagem de Vanessa Schwartz (2004) ao percorrer os antecedentes do espectador cinematográfico na Paris do final do século XX, com os espetáculos da realidade emergentes em seus necrotérios, museus de cera e panoramas.

104 Trata-se, também, da perspectiva desenvolvida por Simondon em sua teoria da individuação. Cf. SIMONDON, 2003; DELEUZE, 2003.

quais cada qual ultrapassaria sua mônada (GUATTARI, 2003, p. 49). Guattari diz da necessidade de invertermos esta compreensão monádica – baseada no que ele denomina uma “referência comunicacional” – para tomarmos a transferência como princípio anterior à própria formação das entidades que reconheceríamos, ainda que transitoriamente, como sujeitos. Em vista disto, devemos compreender não apenas a necessária multiplicidade daquele que, ainda que hipoteticamente, nomeamos como um *observador*, como também a impossibilidade de chegarmos – seja empírica ou hipoteticamente – a um sujeito plenamente constituído (antes ou depois do dispositivo). O observador, nesta compreensão, nomeia uma projeção daquilo a que encaminha um dispositivo em um processo de subjetivação (seu “*sujeito ótimo*”, talvez). Trata-se, então, de um vetor de subjetivação entre muitos outros, mais do que a figura de um sujeito constituído e estabelecido.

Quanto àqueles que identificaríamos como os sujeitos empíricos – singulares e plurais – é necessário que compreendamos, como sugere Agamben (2009, p. 41-42), que os seres viventes comportam processos de subjetivação simultâneos. Compreendendo o sujeito como um terceiro que se produz pela relação entre a classe dos viventes e a classe dos dispositivos – de que parte em sua própria leitura do conceito de dispositivo em Foucault – o filósofo evidencia a não coincidência entre seres e sujeitos:

Naturalmente as substâncias e os sujeitos, como na velha metafísica, parecem sobrepor-se, mas não completamente. Neste sentido, por exemplo, um mesmo indivíduo, uma mesma substância, pode ser o lugar de múltiplos processos de subjetivação: o usuário de telefones celulares, o navegador na internet, o escritor de contos, o apaixonado por tango, o não-global etc. Ao ilimitado crescimento dos dispositivos no nosso tempo corresponde uma igualmente disseminada proliferação de processos de subjetivação. Isso pode produzir a impressão de que a categoria da subjetividade no nosso tempo vacila e perde consistência; mas se trata, para ser preciso, não de um cancelamento ou de uma superação, mas de uma disseminação que leva ao extremo o aspecto de mascaramento que sempre acompanhou toda identidade pessoal (AGAMBEN, 2009, p. 41-42).

Deste modo, talvez possamos compreender com maior precisão que o *observador* não poderia dizer de nenhum indivíduo empiricamente localizado, inclusive porque cada um destes indivíduos pode comportar diferentes *observadores*. Deste modo, o incômodo manifesto e comum entre os críticos de Crary<sup>105</sup> parece se dever justamente à ausência em seu texto do reconhecimento explícito desta coexistência e mesmo da disputa entre diferentes regimes em um mesmo período ou, ainda, sobre um mesmo indivíduo. Afinal, se desvinculamos a figura do observador sugerida por Crary dos observadores empíricos que encontraríamos na Europa novecentista, não seria tão inadequado supor que haveria uma ruptura entre o observador da câmara escura e aquele do estereoscópio, pois já não se trataria necessariamente de uma distinção arbitrária entre diferentes momentos históricos – que, de fato, se interpenetram – mas entre diferentes posturas requisitadas ao *observador* deste e daquele dispositivo. Também hoje, se nos propomos discutir neste capítulo o *observador* engendrado pela Visão Computacional e apontar características particulares do processo de subjetivação implicado

---

105 Cf. Nota 102.

por este dispositivo, isto não equivale à sugestão de que tal processo anularia ou subjugaria outros que, evidentemente, permanecem.

Comolli (2008), em sua abordagem do espectador contemporâneo<sup>106</sup>, prevê a coexistência de múltiplos regimes de espectadorialidade em um mesmo período, entre os quais transitaríamos, segundo o modo clássico de lidarmos com estes diferentes *lugares do espectador*, alternando entre diferentes posturas e modos de relação com a imagem. Contudo, em um tom distinto daquele de Agamben (2009, p. 42), que sugere que a multiplicação dos dispositivos faria disseminar processos de subjetivação e assim multiplicar nossas posturas transitórias, Comolli aponta para um estado de indistinção que seria alimentado, contemporaneamente, por esta profusão, em que os limites não seriam mais tão bem reconhecidos:

[...] se ainda podemos (idealmente) passar de uma exposição à visão de um filme ou a visita a um monumento sem deixar de ser espectadores, mas trocando de registro e de papel, para mobilizar modos de funcionamento específicos [...], cada vez mais acontece, devido à assiduidade da pressão comercial, que esses diferentes e simultâneos módulos de formação do espectador estejam combinados em complexos, espaços híbridos e, se assim posso dizer, metaespectaculares [...]. O que antes estava no percurso do espectador – sucessão e passagem de um registro a outro, de uma experiência a outra, cada uma delas marcante o suficiente para permitir supor que fará esquecer aquela que a precedia –, torna-se empilhamento, superposição, confusão (COMOLLI, 2008, p. 135-136).

Tal passagem – que no texto de Comolli introduz sua reivindicação da especificidade do *lugar* do espectador cinematográfico – nos é particularmente interessante pelo modo como apresenta este *lugar* “metaespetacular” do espectador contemporâneo. Talvez possamos bem compreender tal sugestão pelo modo com que hoje nos relacionamos com telas que se prestam como interfaces, talvez, de “metadispositivos”. Por tal noção, nomeamos a versatilidade destes aparelhos (computadores, celulares, *smart* TVs etc.) que hoje nos apresentam qualquer imagem sobre um mesmo suporte, como filmes, programas de televisão, fotografias, *videogames*, textos jornalísticos, romances, *websites*, *softwares* de edição de textos e planilhas etc. Tendo a informação como termo comum entre diferentes conteúdos, tudo e qualquer coisa passível de ser traduzida a tal termo torna-se também passível de ser vista, lida e operada neste plano comum. Inclusive, dizemos de dispositivos que já nem possuem contornos tão estáveis a definir seu contexto de operação.

Alguns destes aparelhos, é claro, possuem ainda aspectos mais ou menos definidos: a *smart TV* ainda ocupa, embora com pequenas alterações, o lugar da televisão que bem descreve Comolli (2008, p. 139), em meio ao espaço doméstico e os típicos bibelôs do lar. Outros, como o celular,

---

106 Alternamos, ao nos referenciar a diferentes autores, entre as noções de observador, preferida por Crary, e aquela de espectador, mais comum no âmbito dos estudos do cinema. As tomamos como equivalentes neste contexto, porém devemos evidenciar suas distinções. Crary (1992, p. 6), em particular, evita a denominação espectador tanto no sentido de propor uma leitura mais abrangente do fenômeno da visão – para além do contexto mais circunscrito do espetáculo – quanto para explorar, em um jogo de palavras, outras conotações implícitas na raiz latina *observare*, pela qual se indica tanto o ato relacionado à visão quanto o de sujeição a determinada regra ou constrangimento.

o computador e o *tablet*, já envolvem contextos de fruição bastante diversos, indo desde contextos de maior concentração e imersão até a dispersão do caminho ao trabalho, ou da mesa de bar. Com frequência, a começar pela tela do computador, valemo-nos destes dispositivos polivalentes para transitar entre situações de trabalho e de lazer, com conteúdos funcionais e estéticos, alternando entre diferentes registros sem abdicarmos do mesmo aparelho que temos à mão em todos estes momentos. Talvez seja justamente tal trânsito – que efetuamos sem deixar, porém, um mesmo *lugar* – que caracterize a postura requisitada ao observador ou ao espectador contemporâneo, em diferentes dispositivos – e, talvez, justamente por isto, estes *lugares* se confundam. Alternando entre um filme e um jogo digital, ou entre um texto literário e um *software* de trabalho, com a mesma velocidade do pressionar de uma tecla ou da efetuação de um gesto, não haveria como distinguirmos com clareza diferentes posturas ou *lugares* ocupados em cada um destes momentos, nem haveria tempo para alternarmos entre eles. Também, com a presença ubíqua dos celulares, tal indistinção permearia mesmo outros dispositivos, ainda em alguma medida ritualizados, da sala de cinema, do teatro, da sala de aula, das galerias expositivas, das bibliotecas ou dos templos e rituais religiosos.

Indo da coexistência de múltiplos regimes escópicos e espetatoriais à sugestão de sua aparente indistinção, podemos compreender nesta passagem uma característica que seria inclusive própria à constituição dos observadores contemporâneos em distintos dispositivos. Se ainda não anulamos por completo as diferenças entre tais registros, não parece demasiado supor que os limites remanescentes, em cada caso, encontram-se hoje bem mais permeáveis, e que, além da capacidade de trânsito entre diferentes *lugares* – beirando sua indistinção –, a atenção em ambientes dispersos seria uma das características centrais demandadas ao espectador ou ao observador contemporâneo. Assim, a sugestão de um regime “metaespetacular” por Comolli, associada a esta consideração de nosso engajamento constante ao que talvez poderíamos até considerar como “metadispositivos” indica-nos a possibilidade de compreender traços gerais concernentes ao modo de constituição do observador em meio a estes diferentes registros. Sugerimos, como modo de compreensão deste contexto, que haveria certa correlação entre um tal regime da profusão e da indistinção e o modo de operação que viemos descrevendo para a Visão Computacional. Se nos é requisitada a capacidade de percorrer velozmente diferentes registros, ou mesmo de tomá-los indistintamente, é justamente nesta tarefa que os algoritmos da Visão Computacional parecem nos exceder. Conquanto seja limitada a capacidade interpretativa de máquinas de visão, elas já nos ultrapassam em tarefas básicas de reconhecimento de padrões, senão por sua eficácia, por sua resistência e velocidade, percorrendo rapidamente inúmeras imagens, ininterruptamente, bem além de nossos limites atencionais. Os dispositivos informacionais, hoje ubíquos, diante do volume massivo de imagens produzidas, circuladas e consumidas, nos solicitam, então, este trânsito veloz por múltiplos registros, percebidos em sua quase indistinção. Talvez neste contexto encontremos a Visão Computacional tanto como um modelo inteligível daquilo que se requisita ao observador quanto como um eficaz *substituto* nesta tarefa. Aspecto ao qual nos deteremos, mais particularmente.

Evidentemente, a sugestão da máquina como um *observador* é uma hipótese que complicará ainda mais a construção teórica no entorno de nossa questão, mas talvez não haveria como ser diferente: se nos indagamos acerca do observador engendrado pela Visão Computacional, e se lidamos, efetivamente, com máquinas que em alguma medida *veem*, seria incontornável a questão da possibilidade, ou não, de as compreendermos como *observadoras*. Talvez, inclusive, esta seja uma das principais particularidades trazidas por este dispositivo em sua incidência sobre os modos de articulação contemporânea do visível – de modo bastante distinto daquele que encontramos, por exemplo, nas análises dos objetos técnicos abordados por Cray. A tal problema se segue, ainda, o da relação entre estes supostos *observadores-máquina* e os observadores humanos. Entre uns e outros, além da evidente tendência a compará-los, temos, como vemos no exemplo do início deste capítulo, uma implicação mútua, em que uns serviriam de modelo explicativo e de ferramenta investigativa sobre o funcionamento dos outros. A Visão Computacional vem, afinal, explicar a visualidade humana e esta alimenta o modo de funcionamento da primeira. Diante de máquinas que hoje nos revelariam, inclusive, como nós veríamos, em uma estranha mescla entre nossa percepção e a sua, como conceber aquilo que *nós* veríamos? Ou, ainda, quem seríamos *nós*, que vemos?

Não pretendemos, por este caminho, desdobrar as promessas do *pós-humano*, de possibilidades de equiparação entre máquinas e humanos, ou da máquina como *superação* do humano. Gostaríamos, contudo, de refletir acerca da possibilidade e das consequências de reconhecermos também nas máquinas a capacidade de *ver*. A antropóloga Lucy Suchman (2007, loc. 4562) desenvolve compreensão semelhante em sua adesão crítica à ideia de que os desenvolvimentos no âmbito da Inteligência Artificial subsidiariam reconsiderações das concepções tradicionais do humano. Concordamos com ela quando sugere que embora as resistências a tais projetos sejam em geral identificadas como se afeitas a um humanismo conservador – defendendo o humano contra as máquinas – talvez seria justamente o contrário: a Inteligência Artificial, em seu modo de codificação computacional da cognição, na verdade viria reforçar e reencenar justamente este humanismo conservador. Em resumo, sua crítica à Inteligência Artificial baseia-se na identificação de que sua concepção de inteligência parte do referencial da racionalidade moderna, ocidental, que não apenas é excludente de outros modos de pensamento e de ação quanto, demonstra ela, é também ineficaz para o que se propõe realizar neste domínio tecnológico. Ela contrapõe, em seu estudo, a concepção da ação racional como baseada em *planos para ação* àquela de uma *ação situada*, a qual desenvolve a partir de teorias pragmatistas da ação e, particularmente, de uma bela descrição do modo de navegação do povo chuquês, da Micronésia<sup>107</sup>. A partir de um percurso argumentativo que não caberá desenvolver aqui em profundidade, a autora termina por sugerir revisões ao projeto da Inteligência Artificial, sugerindo aspectos do humano que teriam sido obliterados em sua

---

107 Não chegaremos a desenvolver esta sua análise, que mencionamos aqui apenas como exemplo da discussão de onde Suchman parte para a elaboração das ideias que dela trazemos. Quanto à questão da navegação do povo chuquês, trata-se de um modo de navegação que ela contrapõe, a partir de outros autores, ao modo europeu – baseado em um plano de navegação. Os chuqueses, de outro modo, navegariam de modo mais intuitivo e incorporando de modo mais ágil as respostas das águas percorridas. Cf. SUCHMAN, 2007, “Chapter 2 Preface to the 1st edition”.

compreensão particular da inteligência e da humanidade. Deste modo, ela se contrapõe a visões demasiado otimistas de tais desenvolvimentos para propor que eles seriam, sim, propulsores de reconsiderações das concepções tradicionais do humano, mas não por transgredi-las. Identificando mais reiterações do que transformações, a autora parece encontrar em tal tecnologia algo como uma encenação palpável de certa concepção de inteligência – assim, mais passível de análise e de revisões.

Um ponto central da abordagem da autora encontra-se em seu posicionamento que compreende, mais além da discussão da possibilidade, ou não, de equiparação humano–máquina, a necessidade de analisarmos o modo como tal equiparação e, ao mesmo tempo, suas distinções, seriam ainda continuamente performadas, em um constante borrar e retraçar de fronteiras. Trata-se de uma compreensão produtiva na medida em que nos permite escapar tanto categorizações essencialistas quanto equiparações rasas que não levam em conta as efetivas diferenças remanescentes ou a parcialidade das traduções realizadas. De modo mais importante, ela aponta para a maneira com que, ainda que se sugiram equiparações, encontraríamos, em diversos momentos, distinções que seguem sendo elaboradas, *performativamente*, no modo como não apenas se apresentam as máquinas que abordamos (da Inteligência Artificial ou da Visão Computacional), quanto pelo modo com que agem e com que suas limitações tornam-se explícitas (propositalmente ou não)<sup>108</sup>. Suchman sugere-nos esta compreensão ao associar tal retraçar de fronteiras, de modo semelhante, à compreensão que Judith Butler<sup>109</sup>, no contexto da teoria feminista, sugere para a distinção de gênero:

O argumento de Butler de que corpos sexuados e de gênero são materializados ao longo do tempo pela reiteração de normas é sugestivo para uma visão da construção da tecnologia como um processo de materialização por uma reiteração de formas. Butler argumenta que o “sexo” é uma materialização dinâmica de normas de gênero sempre contestadas: de modo similar, poderíamos entender “coisas” ou objetos como materializações de figurações normativas mais ou menos contestadas da matéria. [...] Tecnologias, como corpos, são ambos produzidos e desestabilizados no curso destas reiterações<sup>110</sup> (SUCHMAN, 2007, loc. 5166-5177).

Ou seja, mais do que tomar formas que se nos apresentam na relação humano–máquina como estáveis, *essenciais* ou *naturais*, o entendimento do próprio caráter construído e performado do gênero lança à nossa discussão das categorias do humano e da máquina a necessidade de repensarmos seu modo de

---

108 Quando dissemos do caráter performativo destas distinções, sugerimos, na esteira dos chamados *estudos da performance* (SCHECHNER, 2003; TAYLOR, 2003), que elas não seriam propriamente *representadas*, discursivamente – embora também sejam –, mas que, talvez de modo mais significativo, elas seriam efetivamente elaboradas e reforçadas a partir de determinados comportamentos reiterados – ou *restaurados*, para utilizar o termo de Schechner – pelos quais reencontraríamos, constantemente, cenários de equiparação ou distinção entre os modos perceptivos de máquinas e humanos.

109 BUTLER, Judith. *Bodies that matter: on the discursive limits of “sex*. New York: Routledge, 1993. *apud* SUCHMAN, 2007, loc. 5166.

110 No original: “*Butler’s argument that sexed and gendered bodies are materialized over time through the reiteration of norms is suggestive for a view of technology construction as a process of materialization through reiteration of forms. Butler argues that “sex” is a dynamic materialization of always contested gender norms: similarly, we might understand “things” or objects as materializations of more and less contested, normative figurations of matter. [...] Technologies, like bodies, are both produced and destabilized in the course of these reiterations*”. Tradução nossa.



articulação, para além da contraposição estanque entre elas ou mesmo da tomada entre *distinção* e *indistinção* como estados discretos. A noção de performance sugere, pelo próprio termo, uma posta em movimento da forma, que não é nunca estabilizada completamente, e é tomada, sempre, em transformação. Por tal entendimento, ganha relevância tanto uma certa genealogia do dispositivo, a qual, ainda que de modo menos aprofundado, buscamos pontualmente em nossa análise, quanto uma atenção aos modos reiterados pelos quais age e se apresenta, contemporaneamente. Se entendemos que a Visão Computacional se elaboraria, materialmente, enquanto um movimento oscilante entre o reforço e a contestação de formas constituídas, é principalmente em relação a estes movimentos que parece ser fundamental que observemos, também, os modos de elaboração de seu *observador*. É importante, neste sentido, que a *performance* se faz tanto enquanto um gesto de reiteração quanto de potencial transgressão, pois, se mesmo as condições de sua reiteração são dinâmicas – na forma de comportamentos ou movimentos restaurados – cada gesto de reforço poderia, ao se efetuar na ação, levar as formas previamente constituídas a um limiar a partir do qual poderia assumir outras formas, radicalmente distintas. Voltaremos, na terceira seção deste capítulo, a esta questão. Para este momento, será mais relevante que nos atentemos ao modo com que, mais do que como categorias estanques, podemos compreender a visão humana e a visão da máquina enquanto instâncias em constante movimento e transformação, inclusive uma em relação à outra.

Compreender o caráter performado das constituições da visão da máquina e da visão humana implica, então, atentarmo-nos constantemente às trocas estabelecidas entre elas ou, no modo como Suchman (2007, loc. 4981) indica, à sua “mútua constituição”. Ela sugerirá tal abordagem, ainda, a partir de uma compreensão da agência enquanto uma capacidade distribuída, em rede, algo que Latour e demais autores da Teoria Ator-Rede (TAR) desenvolvem, marcadamente, a partir de Deleuze. Neste sentido, capacidades para a ação, ou para a visão, não podem ser compreendidas de modo isolado, seja para o humano, seja para a máquina – talvez esta possa ser uma tradução, ao vocabulário da TAR, do *observador* de Crary. Escreve a autora: “A agência nesta perspectiva é mais um efeito ou um resultado, gerado por configurações específicas entre entidades humanas e não-humanas”<sup>111</sup> (SUCHMAN, 2007, loc. 4969). Poderíamos reformular a citação para dizer que a visão seria, também, um efeito ou resultado de configurações similares. Como poderíamos, afinal, supor a existência de uma *visualidade humana* essencial? Isto significaria postularmos a existência de uma *visão natural*, quando, como viemos discutindo, nossa visão já teria sido informada pelos dispositivos que pelo menos desde a perspectiva já viriam ordenando e racionalizando a visão, em profunda articulação com objetos e demandas de ordem técnica. Seguramente, mesmo antes da chamada modernidade e mesmo objetos não estritamente visuais também participaram, em outros momentos, desta configuração – mesmo no olhar daquele que mira a flecha para atingir a presa.

---

111 No original: “Agency in this view is rather an effect or outcome, generated through specific configurations of human and nonhuman entities”. Tradução nossa.

Articulada a esta compreensão do caráter distribuído da capacidade para ação, temos, ainda, o caráter distribuído da própria ontologia destes elementos que tomamos no contexto da rede. Bem antes de a tomarmos como conexões entre elementos previamente existentes, precisamos compreender que a própria ontologia destas entidades humanas e não-humanas apenas ganha sentido quando são postas em relação em meio a tal rede (o que, para Latour, acontece apenas no curso da ação). Guattari (2003, p. 42) sugere tal compreensão ao dizer que: “Em vez de haver oposição entre o *ser* e a máquina, o *ser* e o sujeito, esta nova concepção da máquina implica que o *ser* se diferencia qualitativamente e desemboca numa variabilidade ontológica, que é o próprio prolongamento da criatividade dos vetores maquínicos”. Trata-se menos, neste sentido, de nos ater àquilo no que a visão humana diferiria da visão da máquina, e vice-versa: devemos, de outro modo, compreender como se constituem estes modos de ver e estas próprias entidades *no* agenciamento maquínico do qual não apenas participam, mas no qual ganham sua condição de existência.

John Johnston (1999) parte também do *agenciamento maquínico* para abordar, justamente, os modos de articulação entre o humano e a técnica no âmbito do visual. Sugerindo que a visão seria, assim, decorrente de uma inter-relação constituinte destes elementos, que tampouco poderiam ser compreendidos senão em tal articulação, o autor sugere tratar desta questão sob a noção que elabora de uma *visão maquínica*:

A visão maquínica, como usarei o termo, pressupõe não apenas um ambiente de máquinas e sistemas humano-máquina em interação, mas um campo de percepções decodificadas que, sejam ou não produzidas ou emanadas destas máquinas, ganham sua plena inteligibilidade apenas em relação a elas<sup>112</sup> (JOHNSTON, 1999, p. 27).

Trata-se de algo bem próximo do que viemos construindo, não apenas por sugerir este caráter distribuído da visão como, também, por indicar este papel das máquinas como modelos de inteligibilidade para que compreendamos as percepções que seriam elaboradas em tais agenciamentos. Quando falamos da *visualidade* que emergiria do contato com determinados artefatos de produção da imagem – ou, de modo mais amplo, de mediação da visão – é claro que poderíamos discutir aspectos do modo de registro da câmera fotográfica ou da percepção do distante que nos é permitida pelo binóculo ou pelo telescópio. Contudo, podemos ainda falar de modos de percepção derivados destas experiências mas que, ainda assim, se manifestariam mesmo sem a necessidade de estarmos diante de uma fotografia ou munidos de nossas próteses técnicas: a olho nu. A agência destes dispositivos perdura, por assim dizer, mesmo em sua ausência, assim como eles podem surgir como elementos paradigmáticos – como sugerimos anteriormente – para compreendermos transformações das quais eles são também a expressão.

---

112 No original: “*Machinic vision, as I shall use the term, presupposes not only an environment of interacting machines and human-machine systems but a field of decoded perceptions that, whether or not produced by or issuing from these machines, assume their full intelligibility only in relation to them*”. Tradução nossa.

A partir desta compreensão, Johnston chega, inclusive, a criticar a abordagem tecida por Virilio (1994), o qual julga se ater demasiado a uma compreensão dicotômica entre o humano e a máquina. Johnston sugere que a narrativa do livro de Virilio se resumiria ao percurso de uma visão natural que sofreria, desde o telescópio e da matematização da natureza por Galileu, a interveniência da técnica, compreendida como um agente externo. Em todo o trajeto, sugere ele, Virilio teria encarado as próteses e demais intervenções técnicas sobre a visão como uma forma de intromissão, com efeitos de desorientação e alienação (JOHNSTON, 1999, p. 32)<sup>113</sup>. Para Johnston, como também compreendemos, seria fundamental que partíssemos de um outro referencial, que tomasse a articulação entre o técnico e o orgânico como elementos mutuamente constituintes, em uma aproximação, evidentemente, da compreensão desenvolvida por Deleuze e Guattari e que repercute nos demais autores que mobilizamos.

Se falamos de uma *visão maquinaica* engendrada pelo dispositivo da Visão Computacional – se optamos por tal noção ao tratar da visualidade que dele emanaria – sugerimos que haveria um modo de percepção, um certo deslocamento a que seríamos contemporaneamente conduzidos em nossa relação com o visível, do qual participaria esta tecnologia, ainda que não exclusivamente. Vale dizer, então, que nesta perspectiva não apenas vemos *com* máquinas, mas nossa própria constituição enquanto observadores depende desta rede pela qual se distribui uma capacidade para *ver*. Nesta compreensão, *ver* seria um ato que não faríamos sozinhos – não se trata de uma ação cuja origem se localizaria em um ator isolado, portanto –, mas articulados a uma rede de mediadores, humanos e não-humanos. *Quem vê não é*, então, uma pergunta tão simples quanto inicialmente poderíamos supor.

### 3.2 Nós e elas

Como o exemplo trazido ao início deste capítulo sugere, e como já pudemos em mais de um momento indicar, encontramos no discurso e nas práticas dos cientistas e desenvolvedores, bem como em registros do imaginário e da ficção, a sugestão, bastante literal, de uma potencial indistinção ou equiparação entre a visão humana e a visão da máquina. Encontramos, inclusive, programas de Visão Computacional cada vez mais eficazes e eficientes no cumprimento de tarefas visuais, cujo desempenho é frequentemente comparado ao de humanos<sup>114</sup>. Também encontramos,

---

113 Sobre este ponto, ele também critica Crary, por uma breve passagem (1992, p. 1) em que afirma que tecnologias contemporâneas do visual recolocariam a visão em um plano “decegado” do humano. Encontramos também em William J. T. Mitchell (1994, p. 24) uma crítica à mesma passagem. Mitchell, inclusive, a revisará justificando, em alguma medida, a premissa de que parte Crary nesta sugestão, identificando-a como certa ansiedade moral e política do autor, em sua defesa, algo nostálgica, do humano.

114 Cf. TAIGMAN *et al.*, 2013. Numa abordagem algo irônica, o pesquisador Andrej Karpathy, doutorando em Ciências da Computação da Universidade de Stanford, nos Estados Unidos, realizou em 2014 um experimento inverso: observando os altos índices de acerto dos programas de Visão Computacional da Google no teste de desempenho ImageNet – que utiliza um banco de imagens padronizado para comparar programas de interpretação

porém, talvez com igual frequência, aplicações reversas destes princípios, em que se visa justamente uma diferenciação entre humanos e máquinas com base em suas competências visuais. Uma destas aplicações é a que apresentamos como um dos indícios da presença dos agentes computacionais de visão em nosso cotidiano, os chamados *CAPTCHA*: testes a que somos frequentemente submetidos em nossa navegação na *web*, os quais nos solicitam decifrar uma imagem de letras distorcidas, permeada de ruídos, a fim de comprovarmos sermos humanos (Figura 5). Explorando o contraponto daquelas que seriam as características ideais de uma imagem legível por máquinas – como contornos claros e precisos, realçados pelo alto contraste –, os *CAPTCHA* visam algo como uma *invisibilidade* para a máquina, oferecendo barreiras efetivas para a tradução da imagem em informação pelos algoritmos de Visão Computacional. Trata-se da expressão visual, ainda que transitória, de um limite entre as competências visuais do humano e da máquina e que talvez tenha a sua contraparte nos códigos de barra e *códigos QR* – compreensíveis para máquinas, mas não para humanos.

De outro modo, talvez possamos ainda compreendê-lo como uma manifestação performativa de uma distinção transitória entre o humano e a máquina, inclusive acompanhando aspectos de sua reconfiguração e de seu retraçar de fronteiras. O modo de funcionamento destes testes e suas transformações e aprimoramentos diriam respeito, afinal, ainda que em contextos circunscritos, à maneira pela qual ainda reconheceríamos algumas diferenças e limites entre aquilo que vemos e aquilo que veem as máquinas. Claro que há certa ambiguidade neste tratamento com relação à ideia de performance mobilizada, uma vez que os *CAPTCHA* realizam, também, certa medida de *desempenho* – um sinônimo de *performance*, no uso comum. O caráter *performativo* que sugerimos, contudo, em nossa remissão ao contexto dos *estudos da performance* e à antropóloga Lucy Suchman, tem outro sentido. Tomando os *CAPTCHA* como performance, compreendemos os testes como momentos particulares, circunscritos, em que uma diferença entre humanos e máquinas transitoriamente se expressa. Na ação solicitada no teste, é colocada em evidência, reflexivamente, uma operação corriqueira de nossa visualidade, o reconhecimento de caracteres, que realizamos diante das letras ruidosas apresentadas, com atenção à efetuação deste ato – aproximamo-nos, então, da noção de Schechner de *ação sublinhada* (SCHECHNER, 2003). Entendemos que é principalmente no seu sentido de efetuação ou realização de uma forma no transcurso de uma ação – assim, mais próximo daquela perspectiva que Suchman traz de Judith Butler – que podemos compreender estes testes em seu caráter performativo: os *CAPTCHA* traçam, afinal, em sua realização, um limite transitório entre tais instâncias, dando-lhes certa forma provisória. Porém, mais do que uma fronteira isolada, devemos tomar sua reiteração ao longo do tempo, com alterações progressivas no modo de aferir a distinção, e consequentes mudanças de concepção quanto à forma destes territórios – do humano e da máquina.

---

automatizada –, inclusive superando os índices humanos, ele tentou alcançar um índice superior, treinando-se, entre outras coisas, para reconhecer mais de 120 raças de cachorro. Cf. KARPATY, 2014 e MCMILLAN, 2015.

A sigla *CAPTCHA* desdobra-se, no inglês, em *Computer Aided Public Turing test to tell Computers and Humans Apart*, que poderíamos traduzir, aproximadamente, como: “teste de Turing público e auxiliado por computadores para distinguir humanos de máquinas”. Trata-se de uma referência ao teste proposto em 1950 por Alan Turing (2003), cientista da computação britânico, em uma das primeiras formulações da Inteligência Artificial. Ele propôs um teste na forma de um jogo de imitação pelo qual seria possível aferir a inteligência de uma máquina – manifesta quando um examinador, em certas condições, não conseguisse dizer qual de seus interlocutores era a máquina e qual o humano<sup>115</sup>. No caso dos *CAPTCHA*, temos uma de suas atualizações pelo que comumente se indica como um *teste de Turing reverso*, pois seu objetivo não é alcançar o ponto de indistinção previsto por Turing, mas retrair o limite que seguiria nos separando das máquinas. Talvez, inclusive neste sentido, não se trataria exatamente de uma medida de desempenho, como mais claramente identificamos em Turing, mas da reafirmação de uma diferença, como se intransponível fosse. Enquanto no teste clássico haveria algo como uma competição entre capacidades de convencimento, os *CAPTCHA* apenas valem enquanto não são uma competição, mas a efetuação de uma diferença. Evidentemente, contudo, trata-se de uma distinção que possui um forte valor pragmático e operacional, desempenhando uma função necessária em situações em geral relacionadas à comunicação mediada por computador, em que máquinas chegam a conseguir se passar, ainda que transitoriamente, por pessoas<sup>116</sup>.

No âmbito da internet, os *CAPTCHA* são dispositivos que visam automatizar o processo de distinção entre humanos e máquinas, ou seja, habilitar a própria máquina a realizar esta tarefa, servindo-se enquanto barreira para robôs que navegam a rede a fim de extrair informações automaticamente ou de difundir mensagens publicitárias de toda sorte (*spam*). Eles se oferecem enquanto fronteira anterior ao acesso às interfaces de publicação e acesso à informação, mas têm tanto sua aplicabilidade quanto sua eficácia (ou mesmo sua necessidade) restrita a tais contextos. Conquanto desempenhem tal função bastante concreta e operacional, a matriz performativa lança luz a aspectos importantes dos *CAPTCHA* que escapam a este domínio, pelo qual podemos inclusive compreender que eles realizariam algo próximo do que indica Suchman (2007, loc. 5166-5177), ao *reencenar e reiterar* formas constituídas da percepção do humano e da máquina. Inclusive, por meio dele, talvez possamos acompanhar minimamente os processos de mútua adaptação e constituição humano-máquina sugeridos pela autora. Neste domínio, trata-se de um teste que realiza hoje, na prática, um modo de distinção e constituição destas entidades já antecipado pela ficção.

---

115 Em resumo, o teste envolveria um examinador que deveria interrogar um homem e uma máquina contando apenas com um mecanismo de comunicação por escrito para com eles interagir. A cada resposta, a máquina responderia fingindo ser uma pessoa, de modo a confundir o examinador e a conduzi-lo a uma avaliação equivocada quanto a qual deles seria a máquina e qual o humano. Se a máquina conseguisse confundir o examinador a tal ponto que ao final do teste ele não conseguisse pudesse definir ao certo qual interlocutor era humano e qual era máquina, seria possível dizer que esta era inteligente.

116 Como parecem atestar análises recentes que demonstram, por exemplo, a ampla presença de robôs postando mensagens em redes sociais. No contexto brasileiro, ganhou visibilidade a constatação por pesquisadores do LABIC, da Universidade Federal do Espírito Santo, da ampla atuação de robôs para a difusão de mensagens em redes sociais durante as eleições presidenciais de 2014. Cf. GOVEIA *et al.*, 2014.

Em certo sentido, os *CAPTCHA* reencenam algo próximo dos testes que encontramos, por exemplo, em *Blade Runner*, de 1982, dirigido por Ridley Scott. Bem além do falar robótico e movimentos duros – que *consagraram* Schwarzenegger como o *Exterminador* – o filme de Scott pinta um futuro em que androides, chamados *replicantes*, já seriam tão semelhantes a humanos que não apenas muitos deles não saberiam de sua real natureza quanto seriam necessários investigadores especializados para realizarem tal distinção – que são os chamados *blade runners*. Baseado no romance de ficção científica *Do androids dream of electric sheep?*, de Philip K. Dick (1968), publicado originalmente em 1968, o filme apresenta-nos uma versão particular de um teste de Turing reverso: o *Voight-Kampff Empathy Test* (Figura 30), que realizaria a tarefa da distinção a partir do monitoramento minucioso das reações involuntárias de músculos oculares e capilares dos indivíduos. Com capacidades intelectuais avançadas, as máquinas já teriam plena capacidade de se passar por humanos, fazendo-se necessário encontrar um outro limiar – já bem distante daquele sugerido por Turing – para se traçar uma diferença. Em todo caso, como nos *CAPTCHA*, já não se trata de estabelecer um limite para ser ultrapassado pelas máquinas a fim de podermos considerá-las inteligentes, mas de estabelecer um limite que seria para elas, em seu estágio atual, intransponível.

Com uma narrativa que se elabora em vários momentos na dúvida acerca da real natureza de seus personagens, o filme estende o problema enfrentado pelo protagonista ao próprio espectador, engajado na tarefa dos *blade runners*. Como um sutil auxílio nesta tarefa encontramos, em alguns dos personagens robóticos do filme, indícios de sua natureza oculta em um estranho brilho reluzindo, ocasionalmente, no fundo de seus olhos (Figura 34). Assim como os *CAPTCHA* se apresentam, hoje, como os testes mais comuns para se distinguir humanos de robôs no âmbito da rede, com base em suas competências visuais, são também os olhos dos androides que, nas narrativas de Philip Dick e Ridley Scott, colocam-se como pontos de separação entre nós e as máquinas. A escolha do olho ou do olhar como este ponto de encontro e de separação, que no caso dos *CAPTCHA* provavelmente se deve a um critério técnico relativo à complexidade da tarefa da Visão Computacional, para o caso do filme parece remeter a uma característica das nossas próprias relações interpessoais, em que o olho seria uma forma de acessarmos, de modo mais íntimo, algo do *outro* – guardando, como se diz, o acesso à alma. Este pálido brilho no fundo das pupilas dos *replicantes*, neste sentido, sugere algo como uma opacidade a impedir que os alcancemos – como se, dentro, só nos restasse encontrar o vazio. Haveria uma correspondência no mínimo curiosa a conectar os rostos destes androides e o rosto estatístico elaborado pelos programas de Visão Computacional da *Google*, que abordamos no capítulo anterior (Figura 27), como se o mesmo vazio dos olhos que as máquinas veem em nós lhes fosse devolvida. A inacessibilidade do olhar, neste sentido, parece ser um aspecto recorrente no modo como se figura, em diferentes momentos, a relação entre humanos e máquinas.



Figura 34 – *Frame* de *Blade Runner*, de Ridley Scott. Fonte: BLADE, 2014.

**b****blrr**

Se você consegue  
ler isto,  
provavelmente  
não é uma  
máquina



Figura 35 – Captura de tela do aplicativo *Bablrr* (2012), de Thiago Hersan.



*Blade Runner* nos traz, então, uma situação extrema em que os limites que viemos descrevendo entre humanos e máquinas já teria se transformado a tal ponto que já não encontraríamos indícios tão evidentes para tal distinção. Os *CAPTCHA*, embora atualizem problema similar, o fazem em um contexto bastante distinto, como indicamos, relativo à navegação na rede. Efetivamente, em nosso contexto, diferentemente da figura dos androides que foram tão elaboradas na ficção, as máquinas com as quais lidamos não possuem um corpo e tampouco são individualizadas na forma humana. A forma dos *robôs* contemporâneos, das máquinas inteligentes, é, marcadamente, distribuída, em rede. Neste sentido, não se trata de uma distinção de aparência, de naturalidade do modo de falar, ou do modo de funcionamento de seu organismo. Não chegamos a este nível e nem parece ser bem este o caminho hoje perseguido no âmbito da Inteligência Artificial. Assim, é principalmente no contexto da rede que tal distinção, quando é o caso, se faz necessária. Em um contexto em que não apenas migramos para a internet grande parte daquilo que produzimos, inclusive nossas próprias relações sociais, e, também, em que espaços físicos encontram-se cada vez mais conectados por câmeras e outros dispositivos<sup>117</sup>, driblar a máquina, tornar-se invisível ou incompreensível diante dela, alcança um outro sentido. Assim, performam-se de outra maneira os limites entre humanos e máquinas, não tanto para efetivamente distingui-los, ontologicamente, mas enquanto um gesto de escape à capacidade de monitoramento e conexão dos algoritmos inteligentes diante daquilo que lhe é apresentado. Neste contexto, os *CAPTCHA* surgem enquanto um *contradispositivo* paradigmático.

O artista Thiago Hersan propõe, com seu aplicativo para celulares *Bablrr* (2012)<sup>118</sup>, a utilização do modelo do *CAPTCHA* como uma alternativa para a comunicação interpessoal por redes sociais e *email* em uma forma incompreensível aos algoritmos de monitoramento presentes na maioria destes serviços. O aplicativo gera, a partir da mensagem digitada por seu usuário, uma imagem de letras distorcidas e ruidosas, para ser lida – espera-se – apenas por outro humano (Figura 35), em uma criptografia inacessível à máquina. Trata-se de um trabalho que leva adiante, com o mesmo tom irônico, algumas das inquietações do artista, cujo portfólio inclui, entre outras obras, a performance *Facebukake* (2010)<sup>119</sup>: realizado no *Facebook*, este trabalho envolveu a criação de múltiplas contas com seu nome na rede social, cujos dados de *login* e senha foram distribuídos entre seus amigos, confundindo, assim, qualquer tentativa de produção automatizada de um perfil pela rede. De modo similar, em *Bablrr* Hersan realiza ludicamente um outro gesto de escape ao controle ao mesmo tempo em que joga criativamente com o vocabulário e a lógica da própria rede e de seus usuários. Trata-se de uma atualização reiterada da forma do *CAPTCHA*, tomando-o como um modelo de interdição aos algoritmos de monitoramento.

---

117 Podemos indicar, por exemplo, leitores biométricos, catracas, máquinas de cartão de crédito e sensores dos mais diversos tipos (meteorológicos, geolocalizadores, acelerômetros etc.) embutidos em nossos celulares e, potencialmente, em quase tudo à nossa volta.

118 Cf. <http://www.thiagohersan.com/project/bablrr-2012/>.

119 Cf. <http://www.thiagohersan.com/project/facebukake-2010/>.

Explorando sistematicamente este domínio, o artista Adam Harvey<sup>120</sup>, em meio a seus diversos trabalhos concebidos como dispositivos de contravigilância<sup>121</sup> também incorpora o modelo do *CAPTCHA* como uma de suas estratégias. Desenvolvido desde 2010 junto a *designers* de moda, o projeto *CVDazzle*<sup>122</sup> elabora uma série de *looks* de maquiagem e cabelos desenhados estrategicamente para camuflar os aspectos do rosto que os algoritmos de Visão Computacional utilizam para detectar faces (Figura 36). O artista realizou um estudo do funcionamento destes programas e concebeu os desenhos de cabelo e maquiagem com o objetivo específico de ocultar porções dos rostos que seriam fundamentais para que a máquina os visse como tais – sem, contudo, impedir que nós os façamos. Driblando, diante do poder vigilante, a fácil proibição do uso de máscaras ou disfarces, o artista propõe em seu trabalho uma alternativa de difícil regulação, como uma espécie de estética facial de resistência com forte apelo *pop*<sup>123</sup>.

Hersan e Harvey, em uma interlocução direta com os algoritmos de Visão Computacional, atualizam em seus trabalhos gestos de escape que ganham hoje especial atenção de artistas e ativistas, mas com a inflexão que é própria à vigilância algorítmica. Trata-se de proposta, assim, distinta daquela da artista Hito Steyerl, no seu recente *How not to be seen: a fucking didactic educational .mov file* (2013). Entre várias técnicas de desaparecimento apresentadas neste vídeo, Steyerl sugere, em um tom claramente irônico, que, para escaparmos ao olhar das câmeras, sejamos menores do que um *pixel* – com seus cômicos personagens com cabeças cúbicas, *pixeladas* (Figura 37). O escape ao registro pela câmera, contudo, talvez já tenha deixado de ser uma possibilidade. Hoje, em particular nos grandes centros urbanos e, mais ainda, no contato com os múltiplos dispositivos informacionais que permeiam nosso cotidiano, já quase não conseguimos, ao certo, encontrar territórios em que não estejamos sujeitos ao olhar das câmeras ou de outros dispositivos de coleta e registro da informação. Nem mesmo a resolução das imagens – limitação explorada pela proposição irônica de Steyerl – coloca-se, hoje, como uma barreira efetiva ou, ao menos, duradoura para a vedação do anonimato que parece impor-se, contemporaneamente, ao nosso próprio *estar no mundo*. Os trabalhos de Hersan e Harvey, contudo, sugerem algo como um ocultamento seletivo, restrito apenas ao olhar e às capacidades interpretativas da máquina. Nos é bastante relevante, então, que os exemplos indicados constituam – mais do que simples disfarces, cifras ou ocultamentos – camuflagens direcionadas especificamente à máquina. Desta forma, vemos constatada e denunciada a presença estratégica destes dispositivos como partes de amplos arranjos, instâncias de efetivo governo dos seres (como sugere Agamben), seja para a manutenção da ordem ou para a garantia de maior eficiência (leia-

---

120 Cf. <http://ahprojects.com/>.

121 Destacamos o projeto *Privacy Gift Shop* (<http://privacygiftshop.com/>), por meio do qual comercializa uma série de produtos voltados para o disfarce diante das câmeras de vigilância, incluindo “capas de invisibilidade”, feitas com material que barra a visão de câmeras sensíveis a temperatura utilizados em *drones* (que, não por acaso, jogam com o referencial do vestuário árabe, com burcas e *hijabs*).

122 Cf. <http://cvdazzle.com/>.

123 Em um registro mais conceitual, Harvey realiza, ainda, um experimento voltado especificamente para a questão dos limites à visibilidade da máquina no projeto *For your eyes only* (2012-), experimentando com as possibilidades de camuflagem de objetos diante dos algoritmos da *Google*. Cf.: <http://ahprojects.com/projects/rhizome-commission/>.

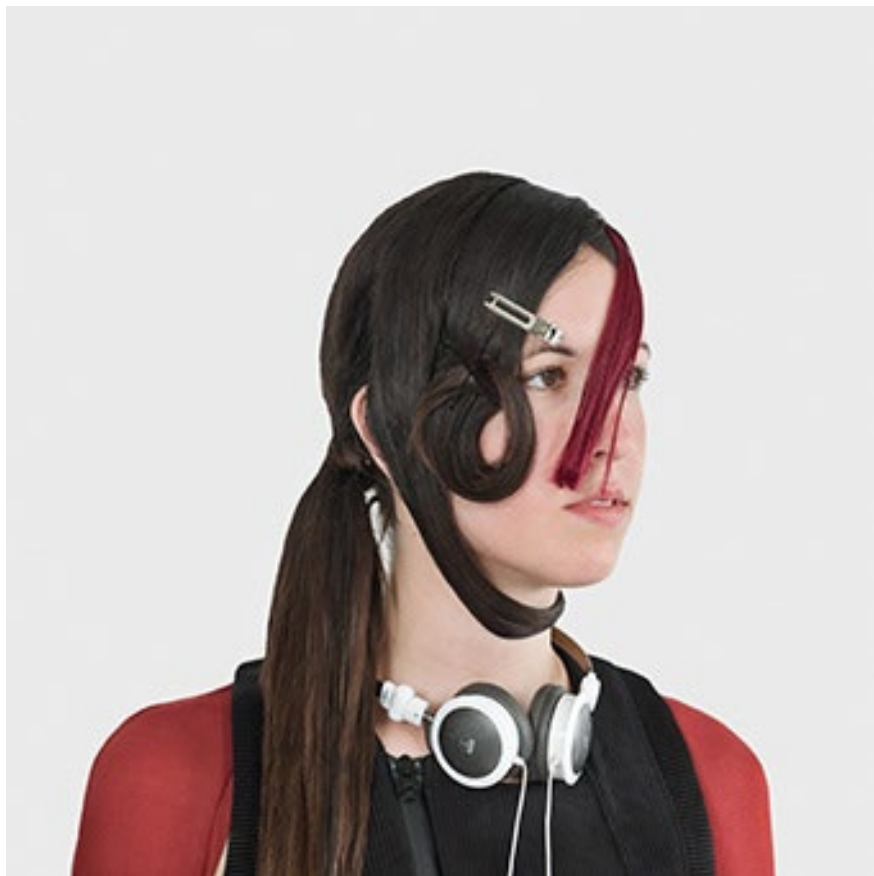


Figura 36 – *CVDazzle* (2010), de Adam Harvey. *Look 4*. Fonte: <http://cvdazzle.com/>.



Figura 37 – Frame de *How not to be seen: a fucking didactic educational .mov file* (2013), de Hito Steyerl. Fonte: CONNOR, 2013.

se, lucratividade) ao capital. No âmbito mais circunscrito de nossa discussão, projetos como o de Harvey e Hersan atualizam o dispositivo dos *CAPTCHA* como uma espécie de contradispositivo diante da Visão Computacional. Eles reiteram, ao fazê-lo, os limites sugeridos por estes testes, em que a máquina volta a surgir como uma espécie de entidade cega – ou, ao menos, míope – às coisas do mundo. Tomando um pequeno dispositivo desenvolvido a fim de automatizar o trabalho de distinção entre humanos e máquinas (justamente quando se buscava, ao menos pretensamente, chegar à sua equiparação) trabalhos como *CVDazzle* e *Bablrr* levam tal proposta a outros domínios, visando em alguma medida tornar disfuncionais os dispositivos de monitoramento mais frequentes.

Porém, mais do que exemplos de uma interdição definitiva ao olhar da máquina, tais casos nos apresentam o caráter ambíguo pelo qual o visível constitui uma importante *fronteira* contemporânea da máquina: tanto no sentido de que seria ele a estabelecer os limites desta com relação ao humano, quanto no sentido de que seria este um dos seus principais campos de expansão. Inclusive, tendo em vista os avanços recentes no âmbito da Visão Computacional, tornando algoritmos cada vez mais capazes de ler as imagens geralmente utilizadas nos *CAPTCHA*, tais dispositivos de aferimento da *humanidade* de navegadores da internet têm passado por revisões e melhoramentos com o objetivo de tornar mais difícil sua superação pelas máquinas, até mesmo com a incorporação de outros procedimentos além da solicitação de interpretação de uma imagem (SHET, 2013). Um bom exemplo, tanto destas reformulações quanto da ambiguidade destes desenvolvimentos – que, ao mesmo tempo que retraçam as fronteiras entre o humano e a máquina, promovem sua indistinção –, é o projeto *reCAPTCHA*, atualmente mantido pela *Google*<sup>124</sup>.

Fundamentalmente, o conceito do *reCAPTCHA* envolve o aproveitamento do trabalho realizado pelos sujeitos submetidos ao teste, de modo que o conteúdo gerado pelo usuário, ao decifrar a imagem apresentada, é incorporado a bases de treinamento de programas de Visão Computacional, além de resolver casos particulares de reconhecimento de caracteres em que os algoritmos empregados se mostraram incapazes. Em sua primeira versão, eram tomadas como imagens para reconhecimento pequenos excertos de livros digitalizados: palavras incompreensíveis para a máquina. Assim, cada solução de um *reCAPTCHA* por um usuário, ao efetuar “manualmente” o trabalho de reconhecimento de caracteres, contribuía na digitalização de textos – no caso da *Google*, em seu projeto de digitalização massiva dos acervos de bibliotecas (Figura 38). Em suas versões mais recentes, foram ainda incorporados como desafios: partes de fotografias tomadas de diversas ruas do mundo pelo projeto *Google Street View* (que oferece imagens georreferenciadas tomadas no nível da rua), em que o trabalho dos usuários era revertido no reconhecimento de números de edificações (Figura 39); e imagens através das quais a empresa treina seus computadores em tarefas de classificação e indexação, em que o *input* do usuário oferece-se como uma forma de contribuir a processos *supervisionados* de aprendizado de máquinas. Para o último caso, o dispositivo convoca o

---

124 Originalmente desenvolvido por Luis von Ahn, da Universidade Carnegie Mellon, o projeto foi adquirido pela empresa como um serviço oferecido “gratuitamente” a desenvolvedores de *sites*. Cf. <https://www.google.com/recaptcha>.

usuário para a seleção, dentro de uma amostra, de todas as imagens pertencentes ao mesmo grupo, de acordo com uma figura oferecida como exemplo (Figura 40).

Em seu estágio mais recente, lançado no final de 2014, que vem sendo chamado como um “*CAPTCHA* sem *CAPTCHA*” (SHET, 2014), o *reCAPTCHA* parece atestar a tendência de abandono da visão como critério distintivo entre humanos e máquinas, sugerindo ser esta uma fronteira praticamente transposta, para este fim. Neste formato, ele se vale prioritariamente de uma análise mais sofisticada, envolvendo fatores relativos ao modo de engajamento do usuário com os *CAPTCHA*, pelo qual, na prática, o teste poderia se resumir, em alguns casos, apenas no clicar de um botão pelo usuário (SHET, 2014 e 2013)<sup>125</sup>. Em um tal caso, veríamos se dissolver, em alguma medida, os limites hoje postos em evidência. O recurso à interpretação visual, contudo, ainda permanece como um segundo estágio de segurança do teste, acionado caso a análise realizada pelo dispositivo não tenha uma boa margem de confiança. Nestes casos, recorre-se aos modos já indicados de leitura de números de edificações, palavras de livros e classificação de imagens. Na divulgação deste seu projeto, em um tom que parece buscar enaltecer um suposto espírito colaborativo dos usuários, o site da empresa dedicado ao *reCAPTCHA*<sup>126</sup> traz alguns slogans curiosos: “Pare um robô. Melhore um mapa”<sup>127</sup> ou “Pare um robô. Salve um livro”<sup>128</sup>. Chama-nos principalmente a atenção, entre eles, o paradoxal: “Pare um robô. Construa um robô”<sup>129</sup>. Vemos, em um nível extremo, a ambiguidade que permeia o desenvolvimento das máquinas que abordamos, particularmente neste jogo de distinção e indistinção.

Encontramos nestes exemplos algo que poderíamos descrever como uma retroalimentação da própria lógica da Visão Computacional ao engajamento produtivo (para a máquina) do olhar humano. A fim de tornar máquinas mais capazes de interpretar o visível, o observador tem seu olhar mobilizado de modo evidentemente instrumental para auxiliar o desenvolvimento de sistemas *inteligentes*. O limite da capacidade da máquina se reverte em trabalho humano convocado – poderíamos mesmo dizer, cooptado, compulsoriamente – entre os usuários da internet. Haveria, assim, uma reconfiguração de seu papel, uma efetiva codificação da visão, que visa, em última instância, tornar mais ágil e mais capaz a máquina, que é, por sua vez, avaliada com base na competência humana diante daquela tarefa em particular<sup>130</sup>. O *reCAPTCHA*, neste sentido, realiza um movimento bem mais amplo daquele que identificamos nos *CAPTCHA* tradicionais, ou em seus desdobramentos.

---

125 Não encontramos menções explícitas a que seriam estes fatores, mas supomos pode se tratar desde movimentação do mouse e digitação no teclado até aspectos da localização do computador do usuário na rede.

126 Cf. <https://www.google.com/recaptcha>.

127 No original: “*Stop a bot. Improve a map*”. Tradução nossa.

128 No original: “*Stop a bot. Save a book*”. Tradução nossa.

129 No original: “*Stop a bot. Build a bot*”. Tradução nossa.

130 Vale destacar que encontramos movimentos similares em outros contextos. O estudo *Selfecity* desenvolvido por Lev Manovich, (<http://selfecity.net/>), vale-se não somente de algoritmos de Visão Computacional, mas do trabalho distribuído de várias pessoas que se dispuseram a trabalhar para o projeto estimando, por exemplo, a idade das pessoas retratadas nos *selfies* estudados, entre outros aspectos. O projeto artístico *White Glove Tracking*, de Evan Roth, por sua vez, valeu-se da mesma forma de trabalho distribuído – chamada de *crowd sourcing* – para rastrear o movimento da luva branca de Michael Jackson em uma performance musical. Os dados gerados por tal rastreamento



Figura 38 – *reCAPTCHA* em sua versão clássica. Fonte: SHET, 2014.

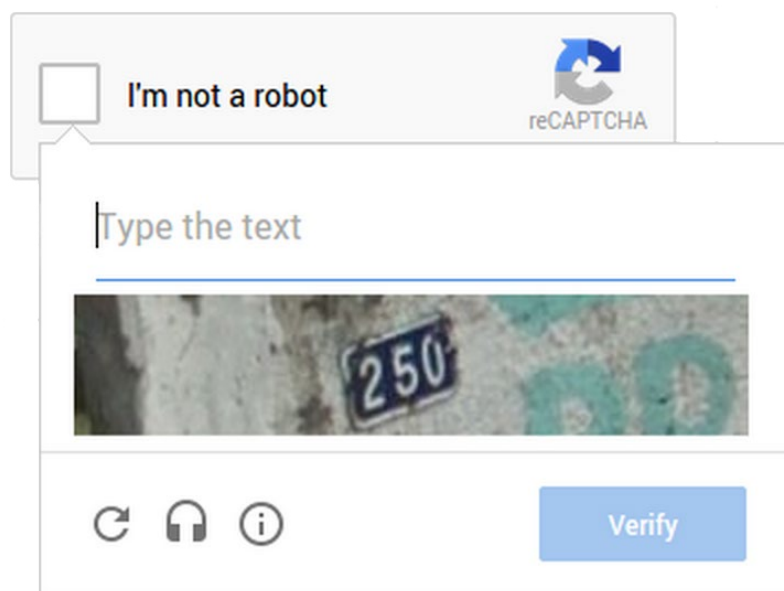


Figura 39 – *reCAPTCHA* com imagem do projeto *Street View*. Fonte: SHET, 2014.

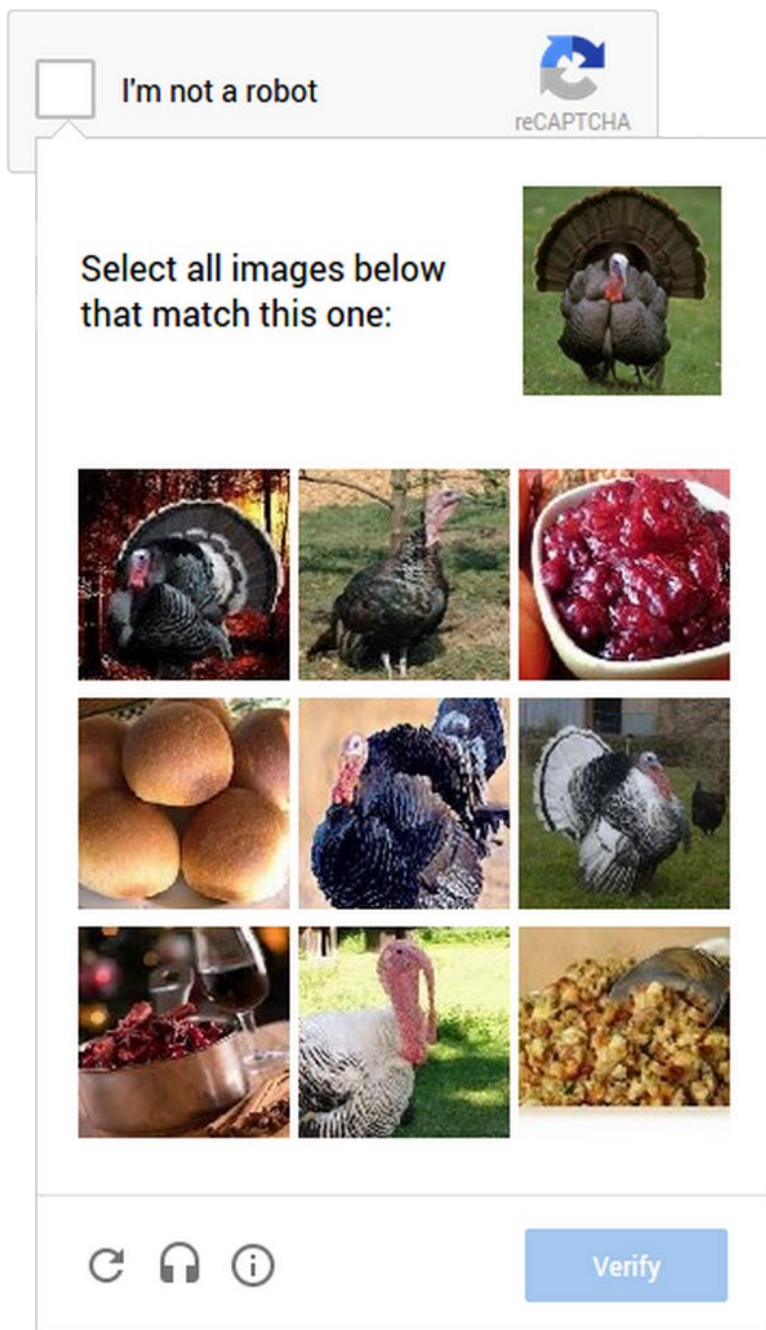


Figura 40 – reCAPTCHA com tarefa de classificação de imagens. Fonte: SHET, 2014.



Estamos, nestes casos, diante de testes que mais do que traçar limites transitórios entre nosso olhar e aquele da máquina, anima a sua transformação, os coloca em movimento, a ponto de quase invertê-lo, ao menos potencialmente.

Um exemplo elucidativo deste movimento podemos encontrar no estágio atual das centrais de atendimento ao cliente e *telemarketing* que, com seu misto de atendentes humanos e máquinas, desafiam distinções estanques destes dois termos ao executar movimentos quase contraditórios. De um lado, os atendentes humanos, submetidos a condições de trabalho e metas de eficiência sub-humanas<sup>131</sup>, são orientados a seguir à risca um roteiro de atendimento que é exibido em uma tela diante deles, muitas vezes na forma de um autômato de estados finitos no qual cada resposta do cliente deve ser enquadrada em uma das opções disponíveis para que a máquina apresente o próximo passo do algoritmo que levará ao encaminhamento da demanda e à conclusão do atendimento. De outro, os atendentes maquínicos já começam a ultrapassar o limiar dos menus numéricos (“disque 1 para...”) e a insinuar a capacidade de estabelecer algum nível de conversação com o cliente a partir do reconhecimento da fala (“diga-me em que posso ajudar”), enquanto, simultaneamente, esboçam – em detrimento do falar mecânico da máquina a que nossos ouvidos se habituaram – uma fala mais coloquial, em um tom mais afetuoso e tranquilo (que já há muito não se apresenta, e com razão, na fala dos atendentes humanos, ou mesmo dos clientes). A contradição, portanto – bastante cruel – se efetua numa quase inversão de papéis: humanos assimilados às máquinas, levados ou forçados a um comportamento mecânico; e máquinas assimiladas aos humanos, desenvolvidas com o objetivo de se desprender do comportamento dito mecânico, e se aproximar de uma qualidade humana da qual nós mesmos já nos distanciamos<sup>132</sup>.

Soma-se, assim, ao discurso de busca de equiparação da máquina ao humano, indícios de um processo inverso, o qual podemos encontrar, ainda, em vários outros exemplos. Um aspecto importante por ressaltar, contudo, que parece tornar-se evidente com o exemplo das telefonistas – e que pretendemos aproximar de nossa discussão da Visão Computacional – é que não são processos independentes. O discurso da equivalência – seja teórica ou efetiva – parece, assim, mascarar as reconfigurações pelas quais o humano não poderia ser tomado nestas comparações como um elemento estável. Torna-se mais claro, assim, que a questão central não é a de se as máquinas poderiam um dia ver como nós, mas, de outro modo, que não é fácil definirmos quem somos nós; que a visão humana se transforma com o tempo; e que o próprio desenvolvimento destas máquinas de visão vem engajar e, conseqüentemente, transformar, o nosso modo de ver. Nos ajustamentos necessários

---

foram depois disponibilizados para a criação de trabalhos artísticos derivados (<http://whiteglovetracking.com/>). Nestes casos, contudo, até onde sabemos, o trabalho não é alimentado ao treinamento de máquinas.

131 Parece-nos que, com relação às metas de eficiência, seria possível argumentar que seriam *sobre-humanas*. Optamos pelo *sub*, contudo, como um posicionamento em relação a certa concepção de humanidade.

132 Suchman (2007, pos. 4103) elabora, a partir de Alexandra Chasin, uma interessante crítica da equiparação do humano e da máquina pelo viés da exploração do trabalho e das representações de classe e de gênero – entre elas as figuras recorrentes da máquina como empregada doméstica ou como mordomo.

a tais acoplamentos humano–máquina, não poderíamos supor que os movimentos realizados nesta rede de agências interconectadas não afetariam todo o conjunto.

Uma tal consideração certamente vem problematizar, como antecipamos, o modo pelo qual consideramos, neste contexto, a constituição daquele que viemos denominando, a partir de Crary, o *observador*. O próprio autor (CRARY, 2014), sugere um aspecto desta implicação em uma pontual revisão que oferece em seu mais recente estudo, *24/7*, que explora a teia de elementos que compõem o regime temporal das sociedades contemporâneas<sup>133</sup>. Partindo da consideração, que toma de Fredric Jameson, acerca da centralidade que o ato de *olhar para as imagens* assume, hoje, no modo de funcionamento de instituições hegemônicas, ele escreve:

É claro que hoje mais imagens, dos mais diversos tipos, são olhadas, vistas, do que em qualquer outro período, mas no interior do que Foucault descreveu como uma “rede de observação permanente”. A maioria dos conceitos historicamente acumulados do termo “observador” é desestabilizada sob tais condições: isto é, quando os atos individuais de visão são interminavelmente solicitados e convertidos em informações que *tanto* serão utilizadas para o aprimoramento das tecnologias de controle *quanto* virarão uma forma de valor excedente em um mercado baseado na acumulação de dados sobre o comportamento do usuário. Há uma subversão de pressupostos muito mais literal a respeito da posição e da capacidade de ação do observador no leque cada vez maior de meios técnicos de transformar os próprios atos de visão em objetos de observação (CRARY, 2014, p. 56).

A noção de observador entraria em crise, pois observar não significa apenas submeter-se a determinadas restrições e ordenamentos técnicos ou institucionais mas significa também, e cada vez mais, contribuir e participar ativamente – ainda que à revelia – das dinâmicas de desenvolvimento destes ordenamentos. Claro que, matizando a generalização sugerida, não se trata ainda de um modo de funcionamento pervasivo, mas trata-se, de toda forma, como nossos exemplos parecem também indicar, de uma culminância prefigurada tanto pelos desenvolvimentos quanto pelas formulações ficcionais que hoje encontramos. O *reCAPTCHA* é, assim, um exemplo perfeito do caráter corriqueiro que vem assumindo aquilo que talvez poderíamos chamar, a partir do sugerido por Crary, de uma *comoditização* do olhar. Vemos se realizar, ali, a sua captura enquanto componente fundamental do desenvolvimento das próprias técnicas que o governam. Para além deste caso, haveria ainda as várias métricas que se aplicam hoje ao monitoramento de nosso modo de engajamento com aplicativos computacionais, páginas na *web* e mesmo em nossa fruição de filmes, músicas e livros. Como um de seus exemplos mais extremos – porém de aplicabilidade restrita – Crary indica ainda dispositivos de estudo da ergonomia óptica (com o rastreamento do movimento ocular). Trata-se, em todos eles, de

---

<sup>133</sup> Não será o caso de aprofundarmos na descrição de sua análise, pois dela nos aproximaremos de modo apenas pontual. Em resumo, contudo, o autor investiga o modo pelo qual se constitui contemporaneamente uma temporalidade caracterizada pela disponibilidade ininterrupta dos sujeitos e pela dinâmica do consumo e das transações comerciais 24 horas. Em certo sentido, Crary sugere uma espécie de colonização da experiência pelo capitalismo, sugerindo que a redução do tempo dedicado ao sono representaria um tipo de redução da própria parcela da vida que não teria sido ainda totalmente integrada às lógicas de consumo – com a exceção, talvez, do mercado de soníferos. Cf. CRARY, 2014.

um modo de compreensão da visão que já se estende bem além de sua fisiologia ou psicologia para atingir, de modo articulado ao desenvolvimento da Visão Computacional, o pleno conhecimento e, principalmente, o controle da visão em seus aspectos cognitivos e comportamentais.

Sua faceta mais extrema talvez seja, ainda, o que encontramos na pesquisa dos neurocientistas de Berkeley, em seu acesso às *imagens mentais*. Parece ser a estes mesmos estudos que Crary (2014, p. 106-107) se refere em outro momento, quando percorre – a partir de uma bela recuperação de *La Jetée* (1962), de Chris Marker – os empreendimentos contemporâneos voltados ao controle da própria mente e suas imagens. O autor sugere, a respeito destas pesquisas, que elas indicam a compreensão dos sonhos como: “entidades à parte que, com o desenvolvimento da tecnologia necessária, poderiam ser gravados e de alguma maneira reproduzidos ou baixados” (CRARY, 2014, p. 106). Efetivamente, neste ponto já se ultrapassa o próprio governo do olhar para se alcançar o governo do pensamento ou da consciência. Crary sugere como absurda a presunção de que partem estes projetos, porém indica sua relevância enquanto componentes do imaginário contemporâneo, presentes, ainda, em filmes como *A origem* (2010), de Christopher Nolan. Poderíamos ainda acrescentar outros exemplos como: *Brilho eterno de uma mente sem lembranças* (2004), de Michel Gondry e *Violação de Privacidade* (2004), de Omar Naim. Em todos estes casos, porém – e este é um ponto de importância particular para nossa abordagem –, tanto nos exemplos da ficção científica quanto da própria ciência, o que parece conduzir os modos de captura e registro das *imagens mentais* é uma mesma e recorrente analogia entre humano e máquina, pelas quais compreendemos as percepções e as memórias como fotografias, filmes ou, enfim, como arquivos, informação. Apenas assim elas se tornam acessíveis, *baixáveis*, editáveis. Os *optogramas* da França novecentista realizavam um tal esforço revelando a imagem da percepção como uma fotografia. Os neurocientistas de Berkeley o fazem a partir de métodos da Visão Computacional e da Inteligência Artificial.

Em todas estas elaborações, suposições e desenvolvimentos efetivos, trata-se da tomada do olhar, também ele, como instrumento. Nesta articulação, portanto, evidencia-se que uma visualidade que suporíamos para observadores humanos apenas se tornaria compreensível e inteligível – talvez hoje, mais do que nunca – por meio de sua intrincada relação com uma *visualidade da máquina*, e vice-versa. Afinal, não apenas dependemos da mobilização da máquina e de seu modo de *ver* para lidar com nossa própria visão quanto é também o nosso olhar que se articula cada vez mais na constituição destas máquinas de visão. O que Crary nos aponta acerca da crise do observador contemporâneo vem acrescentar, então, que o olhar humano não participa destes desenvolvimentos apenas como metáfora ou como modelo – como na analogia originária entre câmera e olho –, pois sua participação não é mais apenas teórica e localizada. De outro modo, encontramos hoje um contexto em que nossos *atos de visão* são constantemente cooptados a integrar e a participar ativamente do desenvolvimento destes sistemas, estabelecendo novos pré-requisitos à constituição contemporânea do *observador*.

### 3.3 Nós e a rede

*I'm Google* é uma obra da artista estadunidense Dina Kelberman, em desenvolvimento desde 2011<sup>134</sup>. Apresentada na forma de um *blog* da plataforma *Tumblr*<sup>135</sup>, ela se despe das características tradicionais deste tipo de publicação – como o típico encadeamento de *posts*, em geral precedidos por título e data em uma série linear. Seu *layout* se estrutura, de outro modo, em três austeras colunas de imagens, quase sempre de tamanho uniforme, sobre o fundo branco. Novas figuras são acrescentadas sempre ao topo e é possível navegar por todo o histórico de atualizações da obra percorrendo verticalmente sua extensão, fazendo descer a barra de rolagem. Quando acreditamos chegar ao fim, basta que esperemos um pouco para que mais imagens sejam carregadas, sugerindo uma extensão infinita à sequência (embora seja, de fato, limitada, conquanto extensa). Teríamos dificuldades em conferir qualquer traço geral ao conjunto apresentado: com enquadramentos variados, englobando um amplo espectro cromático e abarcando temas completamente díspares, ele é fortemente heterogêneo. Se acompanhamos a sequência de sua apresentação, contudo, revelam-se conjuntos semânticos nos quais desfilam séries de imagens de: esculturas de palitos e massas de modelar; luvas de borracha; máquinas agrícolas; boias; peças de crochê; mãos protéticas; sacolas plásticas presas em galhos de árvore etc. No último acesso que realizamos antes do fechamento deste texto, o topo da sequência era ocupado por uma série de quadras de *squash* (Figura 41). Mais do que apenas uma categorização semântica das imagens, contudo, logo percebemos uma organização expressa tanto internamente a estes grupos quanto, de forma mais pronunciada, na passagem entre um e outro conjunto. Nestes intervalos, efetua-se uma transição cuidadosamente construída com base na similaridade aparente entre as figuras. É assim que passamos, por exemplo, das quadras de *squash* aos tênis de mesa (*ping pong*). O salto entre temas tão diversos quanto massas de pão e *rallies* no deserto são amortecidos pela tonalidade de uma e outra imagem, pela forma aparente dos objetos retratados, pela textura de seus materiais (Figura 42). De modo semelhante, passamos de ginásios de treinamento de ginástica olímpica a caixas cheias de pedaços de isopor, em um lapso da mudança de escala entre os objetos retratados (Figura 43).

Em todas estas transições, a passagem de um conjunto ao outro se dá por elementos presentes apenas na superfície das imagens, como que freando nosso olhar e impedindo-o momentaneamente de buscar a coisa representada, levando-o a se ater a seus aspectos formais. Diante das relações estabelecidas e mesmo da profusão heterogênea de temas e contextos, não conseguimos adentrar a cada uma e nossa leitura, assim, oscila entre estes níveis, entre a superfície e a profundidade, sem se estabelecer firmemente em nenhum dos dois. Os saltos temáticos e contextuais, mascarados pela aparência superficial das imagens, sugerem-nos, ainda, um equívoco: como se o olhar de quem construiu a série tivesse sido enganado pelas similaridades aparentes entre as imagens de um e outro conjunto. Como se a organização partisse de uma interpretação falha daquilo que é apresentado.

---

134 Cf. <http://dinakelberman.tumblr.com/>.

135 <http://www.tumblr.com/>.



Figura 41 – Excerto da sequência de *I'm Google* (201-), de Dina Kelberman. Fonte: KELBERMAN, 201-.



Figura 42 – Excerto da sequência de *I'm Google* (201-), de Dina Kelberman. Fonte: KELBERMAN, 201-.



Figura 43 – Excerto da sequência de *I'm Google* (201-), de Dina Kelberman. Fonte: KELBERMAN, 201-.

Contudo, ao mesmo tempo em que o estranhamento ou a surpresa possam ser as reações iniciais diante do jogo proposto pela obra, talvez em um segundo momento reconheçamos, nesta experiência, ressonâncias de uma visualidade com a qual já lidamos, em alguma medida, em nosso contato diário com as imagens da internet apresentadas em mecanismos de busca. Talvez até prescindiríamos do título, embora ele venha a nosso auxílio: haveria alguma proximidade entre a experiência de navegar a coleção de Kelberman e aquela de passear pelas imagens resultantes de uma busca no serviço da *Google*, tanto pela heterogeneidade visual encontrada quanto pela multiplicidade de contextos de que teriam sido colhidas – aparentemente a esmo. Mais do que isto, no movimento de leitura a que somos conduzidos pela obra sugere-se uma dinâmica refinada em que não jogam apenas nosso olhar e o olhar da artista. O processo, quase algorítmico, com que são construídas as relações entre as imagens, com seus vínculos cromáticos e sutis jogos de semelhança, sugere a interveniência de um olhar da máquina na composição das sequências e agrupamentos temáticos. Haveria, afinal, certa correspondência entre este olhar que identificamos como preso à superfície e aquela imagem virtual instrumental, estatística, média, descontextualizada, que encontramos como aproximações da representação interna elaborada pelas máquinas de visão.

Se, em outro gesto de navegação, clicamos sobre uma imagem apresentada, somos levados a reconhecer a participação efetiva dos mecanismos de pesquisa neste arranjo. O *link* construído pela artista, acionado pelo clique, leva-nos justamente aos resultados de uma busca realizada no *site* da *Google* que toma a imagem selecionada como chave da requisição. Seguindo, desta forma, os rastros deixados pela última quadra de *squash* da série correspondente (penúltima fotografia da Figura 41), descobrimos que se trata de imagem de divulgação do *Resort Mantra Samui*, na Tailândia, constando em diferentes páginas da *web* relativas ao hotel. Embora nem todas as figuras ofereçam resultados tão localizados, por meio desta busca particularizam-se algumas fotografias integrantes da sequência, demarcando uma diferença em meio ao fluxo de semelhanças que as encadeiam. Pouco abaixo, em meio à mesma página de busca, encontramos outros resultados oferecidos pela *Google* a partir de um método distinto: são imagens “visualmente semelhantes” àquela que indicamos como chave da pesquisa e que nos apresentam ambientes vazios do que parecem ser apartamentos à venda ou disponíveis para locação – potenciais correspondências, logo percebemos, para a elaboração de uma série como a de Kelberman. Já em outro registro, tais resultados nos afastam de qualquer contextualização ou singularização para nos levar, novamente, ao jogo de semelhanças explorado pela artista. Percebemos, então, como a oscilação entre a tomada da imagem enquanto signo de um objeto singular e a sua assimilação como mais uma de uma série de rimas visuais não é apenas parte da dinâmica da obra como é também sugerida pelo próprio modo com que se organizam os resultados da busca, reforçando, assim, nossa leitura da obra como uma ressonância destas experiências.

A vinculação oferecida pela artista ao mecanismo de busca pareceria sugerir, em um primeiro momento, que a sequência fosse resultado de um processo automatizado de requisições e agregação de resultados. Contudo, logo percebemos que, talvez, a opção da artista venha explicitar justamente

que este não é este o caso. Ainda que, ao perseguirmos as conexões elaboradas pela busca da *Google*, encontremos com frequência imagens bem próximas daquelas pertencentes ao mesmo conjunto temático, não são exatamente estas as presentes na sequência – ao menos, não frequentemente. Quando realizamos buscas a partir das imagens que compõem as transições entre os conjuntos, por sua vez, não encontramos nem mesmo correspondências próximas às elaboradas pela artista – como explicitado pelo exemplo da quadra de *squash* (Figura 44). Revela-se, assim, que por mais que faça uso destas ferramentas na elaboração de sua coleção, trata-se de um trabalho que parece ser realizado de forma *manual* pela artista, construindo, “ela mesma”, cada passagem da sequência<sup>136</sup>. A compilação manual, inclusive, revela-se um traço característico de seus trabalhos, dentre os quais encontramos diversos outros constituídos em curadorias similares – embora nenhuma tão extensa – em torno de outras propostas temáticas<sup>137</sup>. Independentemente do ferramental e da estruturação do processo conduzido pela artista, contudo, a percepção da incidência na obra de visualidade informada pelos dispositivos de busca, com seu modo particular de assimilação entre imagens, constitui certamente um traço característico deste trabalho.

Em *Im Google*, por mais que haja um engajamento pessoal da artista na seleção e organização de sua coleção, não podemos situar apenas nela a origem de tais ações, pois precisamos recompor as relações que as constituem, estendendo-as aos dispositivos de busca e seus componentes (tais como seus algoritmos e bases de dados). A experiência constituída pela obra, em sua compilação heterogênea de imagens, inclusive ganha forma e consistência não apenas pelas rimas visuais mas pela mobilização da experiência, que reconhecemos, de nossa navegação na internet. Por isso, tampouco olhamos e percorremos estas imagens sozinhos, pois ao tomarmos parte do jogo proposto pela obra mobilizamos mais do que um olhar *puro, imaculado*, pois ele é também informado pelas visualidades com as quais aprendemos a ver na *web*. O sujeito do título de Kelberman, que conclama para si a identidade do *Google*, neste sentido, podemos entender não apenas como uma personificação da empresa, nem somente como a artista que projeta sobre si mesma aspectos desta personalidade. Com efeito, é também sobre nós que incide tal projeção: “sou Google”, poderia dizer o espectador da obra. O *lugar* que lhe é reservado não é, evidentemente, aquele da máquina, mas tampouco é o de qualquer sujeito, senão o daquele informado por uma visualidade particular que podemos compreender a partir do percurso que realizamos até aqui, tendo a Visão Computacional – com seu modo de operação, suas aplicações e antecedentes – como um referencial que lança luz e confere inteligibilidade à sua constituição.

---

136 São estes alguns indícios a confirmar o que a própria artista descreve ser seu processo, conforme respostas em 6 e 16 de setembro de 2013 a uma consulta realizada por email (KELBERMAN, 2013a e 2013b). Cf. Anexo II – Entrevista com a artista Dina Kelberman. Segundo Kelberman, tanto a seleção quanto a organização das imagens é feita manualmente por ela, em alguns momentos valendo-se de critérios relativamente objetiváveis para este processo, em outros baseando-se na força das transições encontradas. Ela diz utilizar regularmente ferramentas de busca de imagens do Google, mas apenas raramente, segundo ela, na sua forma de busca por imagens semelhantes.

137 Dentre os trabalhos que identificamos como também realizando esforços similares, destacamos *Smoke & Fire* (2013) e *Simpsons Gifs* (2009). Cf. <http://dinakelberman.com/>.



Mantra Samui Resort - HotelsCombined.com



[www.roomguru.ru/.../Mantra\\_Samui\\_Res...](http://www.roomguru.ru/.../Mantra_Samui_Res...) Traduzir esta página  
451 x 300 - Mantra Samui Resort, Бонхут - Найдите лучшие предложения от RoomGuru.ru. Сравните тысячи сайтов одновременно. Оценка 8,9 из 10 от 1 757 ...

Imagens visualmente semelhantes

Denunciar imagens



Googooogle >  
1 2 3 4 5 6 Mais

Figura 44 – Excerto dos resultados de busca realizada na *Google* a partir de imagem da sequência de *I'm Google*. Fonte: <http://www.google.com>.



Figura 45 – Seção da visualização da rede composta a partir da base de imagens utilizadas no desafio ILSVRC 2012. Fonte: KARPATY, [2012?].

Em uma correspondência quase literal entre a visualidade elaborada pelo trabalho de Kelberman e aquela da Visão Computacional, encontramos o experimento de Andrej Karpathy ([2012?]) – cientista da computação e doutorando da Universidade de Stanford – que busca oferecer uma visualização de uma rede de conexões entre imagens que foi elaborada em um processo de aprendizado de máquinas. Tomando como referência a base de imagens utilizada na edição de 2012 da competição anual de programas de reconhecimento visual *ImageNet*<sup>138</sup> (*ILSVRC 2012 – ImageNet Large Scale Visual Recognition Challenge 2012*), Karpathy desenvolveu um método de visualização que realiza uma redução das múltiplas dimensões pelas quais se estende a rede simbólica composta por tais programas. Deste modo, a ampla e difusa articulação entre tais imagens nos é apresentada sobre um plano. Como resultado, tem-se um mosaico em que a posição relativa de cada figura indica suas relações de proximidade com outras figuras da rede, sugerindo aproximações semânticas que acabam por implicar, também, ressonâncias aparentes. Apresentamos na Figura 45 – bem como na capa desta dissertação – seções da visualização gerada por Karpathy.

Disposto bidimensionalmente – diferentemente do encadeamento linear da obra de Kelberman – o mosaico gerado por Karpathy permite-nos também visualizar pelo menos dois níveis de organização que parecem reger suas imagens: em um primeiro nível, agrupamentos semânticos como tampas de bueiro, cúpulas arquitetônicas, escadarias espiraladas, cadeados de senha e discos telefônicos; e, em um segundo nível, passagens entre os conjuntos que parecem se articular por semelhanças entre os objetos e seu registro. Vemos assim ressoar, tanto na obra artística quanto nesta visualização científica, aspectos de uma visualidade comum. Embora o trabalho de Kelberman realize em alguma medida uma alusão intencional ao modo de ver que seria revelado nos resultados de pesquisa da *Google*, não deixa de ser relevante que se trata de uma visualidade que já não se restringe a tal domínio, mas que somos capazes de reconhecer e que talvez, em alguma medida, exerçamos em determinados contextos e atividades. Expressa-se, nestes mosaicos, a mesma visualidade estatística que sugerimos caracterizar o modo de percepção da Visão Computacional, em uma posta em evidência da semelhança ou, ainda, de um eixo associativo do visual, que seria aquele sobre o qual operam tanto as máquinas em sua operação de reconhecimento quanto nós, em alguma medida, em nosso acoplamento a elas na realização de pesquisas ou em experiências corriqueiras de navegação na *web*.

Encontraríamos vestígios deste modo de percepção em diversas circunstâncias que talvez sejam características de um regime escópico da contemporaneidade. O observador, confrontado com a profusão de imagens que encontra no âmbito da internet e dispositivos relacionados mobiliza, afinal, justamente a capacidade que viemos descrevendo de traçar associações e conexões entre tais conteúdos, oriundos de registros diversos, mas sem perder de vista nesta articulação, a particularidade da imagem com a qual é confrontado. Talvez seja também relevante que hoje o processo de elaboração dos conteúdos visuais que povoam as mais diversas publicações em tal

---

138 Cf. <http://image-net.org/challenges/LSVRC/2012/index>

domínio como *sites*, *blogs* e redes sociais (e mesmo estendendo-se ao impresso, à televisão ou ao cinema) articulam-se intimamente com o próprio contexto de busca de imagens da *Google*, com o qual Kelberman claramente dialoga e cujo funcionamento o mosaico de Karpathy dá visibilidade, ainda que indiretamente. Neste contexto, aquilo que viemos descrevendo e denominando sob o campo da Visão Computacional parece configurar um dispositivo chave para conferir inteligibilidade à postura que seria solicitada ao *observador*.

Encontramos ressonâncias desta compreensão na consideração tecida por John Johnston (1999, p. 46) em um aprofundamento de sua discussão daquilo que nomeia *visão maquínica*. Cotejando as correspondências entre a filosofia deleuziana e o estágio atual de desenvolvimento das ciências cognitivas e da Inteligência Artificial, o autor sugere que o cérebro, diante do processo de automatização de suas funções, seria, hoje, um órgão *desterritorializado*<sup>139</sup> (1999, p. 45). Em meio aos arranjos de telecomunicações em rede e a topologia distribuída dos processos computacionais, portanto, e compreendendo, como sugere ele, a inseparabilidade de cognição e percepção, a *visão maquínica* compreenderia o modo de percepção a que se daria lugar neste contexto:

Nos circuitos das redes globais de telecomunicações, não apenas o cérebro encontra-se bem mais desterritorializado, [...] como a multiplicidade de imagens circulando por estes circuitos não pode ser significativamente isolada como instâncias materiais do cinema (ou televisão) e cérebro. Muitas destas imagens, claro, são percebidas, mas sua articulação ocorre por outra lógica: a incessante codificação e recodificação da informação e sua disseminação viral. A imagem em si torna-se apenas uma das formas que a informação pode assumir<sup>140</sup> (JOHNSTON, 1999, p.46).

Uma visualidade contemporânea inserida neste arranjo, portanto, precisaria ser compreendida, necessariamente, enquanto um elemento tomado em tal agenciamento difuso, em que a visão, longe de uma função fisiológica e cognitiva *natural* ou de um ato realizado por um ator isolado, insere-se necessariamente em uma ampla articulação com processos de processamento e interpretação distribuídos. Evidentemente, em uma escala menor, poderíamos compreender que todos os estágios de configuração da visão, em relação a diferentes agenciamentos técnico-discursivos (pelo menos desde a câmara escura, ou mesmo antes), por meio desta formulação da visão maquínica, sugerida por Johnston. Contudo, como sugere o autor (1999, p. 48), estamos hoje em um contexto já bastante distinto daquele em que Deleuze sugere a noção do maquínico. Naquele momento, diz ele, o filósofo a teria elaborado enquanto um substituto conceitual de contraposições inócuas do

---

139 Em uma etapa de sua argumentação, Johnston recorda que, para Deleuze e Guattari, os agenciamentos maquínicos compreenderiam, de um lado, uma face de *instabilidade*, *devir*, *decodificação* e *desterritorialização*; e, de outro, uma face de *estratificação*, caracterizada pela *redundância*, pela *codificação* e *reterritorialização* (JOHNSTON, 1999, p. 28). Embora tragamos tais noções como um referencial de fundo da discussão, não poderemos desenvolvê-las em profundidade.

140 No original: “*In the circuits of global telecommunications networks, not only is the brain far more deterritorialized, [...] but the multiplicity of images circulating through these circuits cannot be meaningfully isolated as material instances of cinema (or television) and brain. Many of these images, of course, are perceived, but their articulation occurs by means of another logic: the incessant coding and recoding of information and its viral dissemination. The image itself becomes just one form that information can take*”. Tradução nossa.

humano e do técnico. Hoje, encontramos uma relação mais estreita entre uma tal formulação e nossa experiência diária, que já não nos permite refutá-la. A *visão maquínica*, neste sentido, seria especialmente apropriada para dar conta deste contexto.

Um aspecto ressaltado por Johnston acerca desta articulação maquínica da visão é a instabilidade que tais acoplamentos humanos e não-humanos introduziriam não apenas na própria imagem quanto na percepção. Quanto à imagem, teríamos a necessária decodificação de modos de registro indiciais ou icônicos que permitiriam, por sua vez, sua recodificação em informação. Quanto à percepção, teríamos um deslocamento de modos perceptivos relativamente estabelecidos, por outros agenciamentos e dispositivos, antes de serem tomados e recodificados a partir das novas relações estabelecidas. Tal indicação de Johnston parece nos oferecer, assim, uma chave para a compreensão do trabalho de Dina Kelberman em relação aos demais exemplos que convocamos em nosso percurso. Encontraríamos, afinal, justamente esta face de instabilidade – o que Deleuze e Guattari chamariam de *desterritorialização* – na experiência de navegar por sua série de imagens. *I'm Google* revela-nos, em sutis jogos de montagem, associações erigidas por sobre os saltos abruptos que sustenta com seus *raccords*. Entre uma e outra imagem, os lapsos entre os conjuntos semânticos oferecem-se, neste sentido, enquanto um intervalo em que vislumbramos, ainda que momentaneamente, a dissolução da imagem, perdendo seus vínculos representacionais enquanto surgem, no instante da passagem, como pura forma ou, ainda, pura informação. Talvez experimentemos, neste momento, o que sugere Johnston quando afirma que, no âmbito da *visão maquínica*, “A imagem em si torna-se apenas uma das formas que a informação pode assumir”<sup>141</sup> (JOHNSTON, 1999, p.46). Haveria, neste sentido, tanto uma desestabilização daquilo que compreenderíamos como uma visão *humana*, quanto da própria Visão Computacional, pois, na medida em que incorpora aspectos desta visualidade da máquina sem qualquer compromisso com o valor pragmático que geralmente a orienta – sem, contudo, buscar qualquer forma de didatismo –, trata-se de uma obra que torna mais difusos os limites, e mais pronunciado o trânsito, entre estas instâncias.

Comparado aos exemplos que trouxemos na seção anterior enquanto desdobramentos do modelo dos *CAPTCHA*, *I'm Google* revela algo como uma outra face performativa dos modos de relação humano–máquina na Visão Computacional. Enquanto, naqueles casos, observamos um peso maior na reiteração das formas de cada visualidade, neste temos justamente a transgressão destes limites em uma experiência que nos leva a uma zona limiar, de passagem, na qual reconheceríamos mais claramente as possíveis aproximações entre estes domínios –que seriam recalçadas em outros contextos. Johnston (1999, p. 37) recorda como, para Deleuze, haveria um efeito desterritorializante na obra de Vertov, principalmente, por sua proposição de um *cine-olho*: a visão do espectador de *O homem com a câmera*, neste sentido, seria deslocada, posta em movimento, diante da efetuação de um olhar *não-humano*. *I'm Google* de certo modo atualiza a metáfora, colocando as categorias do humano e da máquina em movimento, tornando-as igualmente necessárias porém igualmente

---

141 No original: “*The image itself becomes just one form that information can take*”. Tradução nossa.

difusas na constituição da visualidade que nos é mobilizada. Diante de sua extensa sequência experimentamos a dissolução de nosso olhar em meio à rede que, efetivamente, faz-nos ver – como se finalmente alcançássemos, como sugerimos ao começo, a velocidade necessária para deixar a órbita das *esferas* descritas por Latour (2012) e alcançar a rede que as constitui e sustenta. *Im Google* nos permitiria, então, perceber reflexivamente a própria constituição contemporânea do *observador*, percorrendo a rede que o constitui, enquanto momentaneamente liberto do exercício instrumental da visão, que constantemente lhe é demandado.

\*

Mais além dos efeitos de deslocamento ou de uma potencial liberação momentânea do olhar que encontraríamos expressos no trabalho de Kelberman, contudo, devemos levar em conta também como se manifestam, em outros exemplos, um processo inverso, de estratificação ou de *recodificação*. A particularização conferida por Latour (2001) aos dispositivos, quando ele os define enquanto uma espécie de instância organizadora das agências, absorvendo sua proliferação, aproxima-se bastante do modo com que encontramos instâncias deste processo. De um lado, as conexões traçadas pelos diversos componentes enredados, em extensos sistemas humano–máquina, potencialmente deslocam, arrastam e reconfiguram os próprios elementos trazidos à rede. De outro, emergiriam também instâncias de contenção, de organização dos fluxos, de controle. A Visão Computacional já sugeriria tal codificação (literalmente) ao traduzir um modo particular de ver a uma operação algorítmica, programada na máquina. Trata-se, ainda, de um dispositivo que visa organizar os fluxos de circulação do visível, traduzindo as imagens em informação e estabelecendo, entre elas, conexões. De modo mais substancial e relevante para nossa discussão presente do *observador*, contudo, encontramos vertentes deste processo na incidência retroalimentada desta forma codificada sobre o nosso próprio olhar. O alerta que nos traz Crary (2014, p. 56) a este respeito, ao sugerir que, para o observador contemporâneo, seu próprio ato de olhar torna-se objeto de observação, indica-nos um caminho para as evidências materiais de uma tal efetuação. Algo deste processo já se mostra em alguns dos exemplos que pudemos percorrer neste capítulo, como nos experimentos de leitura das imagens mentais dos pesquisadores de Berkeley ou na instrumentalização do olhar humano por sistemas de aprendizado de máquinas como o *reCAPTCHA*. Talvez pudéssemos encontrar, ainda, outros rastros deste desenvolvimento tanto em dispositivos protéticos que sugerem algo como uma *visão aumentada* quanto em configurações contemporâneas de computadores e celulares que voltam suas câmeras já não apenas àquilo que gostaríamos de registrar, mas aos nossos próprios rostos e olhos.

No caso das próteses computacionais da visão, trata-se de uma das principais apostas da indústria de inovação tecnológica, atualmente, em que, por meio de dispositivos na forma de óculos, são oferecidas funcionalidades tanto de apresentação de informações e imagens sobrepostas ao nosso campo de visão quanto de registro e armazenamento daquilo a que direcionamos nosso olhar. O

*Google Glass* é um bom exemplo deste tipo de desenvolvimento embora (ou talvez justamente por isso) tenha sido suspensa sua comercialização em 2015, após cerca de dois anos desde seu lançamento<sup>142</sup>. Apesar de sua interrupção, trata-se ainda de um exemplar, ainda que prototípico, do tipo de projeto que segue sendo perseguido pela indústria, como vemos na recente divulgação pela Microsoft de seu projeto *HoloLens*, a ser lançado comercialmente em um futuro próximo<sup>143</sup>. No caso do projeto da *Google* (Figura 46), encontramos um dispositivo relativamente simples, composto por um par de óculos com computação embarcada, uma pequena tela translúcida sobreposta à lente do olho direito, e uma câmera direcionada àquilo que seria visto pelo usuário. Controlado por comandos de voz e conectado sem fio a um aparelho de celular, o dispositivo conecta-se à internet para recuperar informações exibidas na tela, capta fotografias e vídeos através da câmera e – potencialmente – poderia submeter o registro realizado, em tempo real, a programas de Visão Computacional, com várias aplicações possíveis. Pelas imagens publicitárias que ilustram a experiência de utilização do produto (Figura 47), não é difícil percebermos sua similaridade, por exemplo, às imagens que nos são apresentadas como a visão da máquina em *O exterminador do futuro* ou *Robocop* (Figura 23 e Figura 24), em que gráficos sobrepostos sugerem interpretações e efetivamente agenciam o que se apresenta no plano visível, orientando a ação e conduzindo respostas por parte do usuário. Em suas figurações encontramos, numa realização bastante literal, o *ciborgue*. Sugerindo uma articulação, em um nível extremo, do humano e da máquina – recobrando até mesmo a superfície do olhar – trata-se possivelmente de um exemplo ótimo, para o momento, dos efeitos de regulação e governo a que seríamos conduzidos pela Visão Computacional, ainda que indiretamente. Neles, o observador e o ato de observar são compreendidos, de modo bastante literal, como continuamente integrados ao agenciamento maquínico do visível. Oferecendo efetivas recodificações daquilo que vemos, nosso olhar pareceria se transformar, diante de tais dispositivos, como aquilo que noutros registros teríamos compreendido como o olhar da máquina. O projeto concorrente da Microsoft, embora ainda em estágio anterior ao lançamento, com uma configuração aparentemente bem mais complexa, sugere um ainda maior aprofundamento deste propósito em um dispositivo de realidade mista que projetaria sobre o campo de visão espécies de hologramas com os quais seria possível interagir através do gesto, captado também pelo dispositivo<sup>144</sup>. Levando adiante o propósito sugerido no projeto fracassado da *Google*, *HoloLens* vale-se já de várias operações da Visão Computacional em seu funcionamento, como a reconstrução virtual do ambiente percebido para a incrustação, na imagem, de seus modelos de Computação Gráfica. Mais além desta integração literal, contudo, vemos se atualizar, também neste exemplo, um movimento de retorno à visão humana dos processos e modos de funcionamento típicos da visão sintética dos computadores.

Em uma outra frente – esta já bem mais presente em dispositivos de nosso contato diário – encontra-se a crescente integração de câmeras a celulares e computadores com seu campo de visão

---

142 Cf. BILTON, 2015 e <http://www.google.com/glass/start/>.

143 Cf. HEMPEL, 2015.

144 Cf. HEMPEL, 2015.

voltado para os usuários. Inicialmente oferecendo-se à tomada de imagens para a realização de videoconferências, as lentes que nos encaram quando encaramos as telas encontram, hoje, uma nova funcionalidade nos autorretratos – popularmente chamados *selfies* – o que tem motivado a integração de câmeras ainda melhores e de maior resolução nesta posição. Embora aparentemente um aspecto menor e desvinculado de nossa discussão, a configuração atual destes dispositivos ganha bastante relevância na medida em que começam a despontar outros usos para as imagens captadas. Uma destas aplicações, por exemplo, é a identificação biométrica do usuário para desbloqueio do aparelho. Outra foi recentemente apresentada pela *Amazon*, com seu aparelho de celular *Fire Phone*<sup>145</sup>, cuja funcionalidade chamada *Dynamic perspective* vale-se da detecção do rosto do usuário por quatro câmeras frontais para adaptar dinamicamente o conteúdo exibido na tela. Por tal técnica, é realizado um alinhamento das imagens em relação ao ponto de vista do usuário, criando uma ilusão de profundidade, ou efetuando uma mudança da disposição dos conteúdos quando a tela é olhada de soslaio. Posto de outra forma, o aparelho da *Amazon* percebe o modo como o usuário o olha para, em relação a ele, adaptar-se. Talvez possamos compreender este como um pequeno traço de um desenvolvimento em curso que tomará, cada vez mais, como sugere Crary (2014, p. 56), o ato de olhar como um objeto da observação.

Anterior a tais exemplos, encontraríamos no *design* de interfaces outra situação ainda mais difundida em que o dispositivo da Visão Computacional pareceria ser paradigmaticamente elucidativo. Trata-se de campo responsável pela elaboração das imagens sobre as quais agimos nos mais diversos dispositivos – em que não apenas se requisita nosso olhar em contemplação, portanto, mas também nossa ação sobre aquilo que nos é mostrado. Stella Senra (2013), em prefácio da edição brasileira de *Suspensões da percepção*, de Crary (2013), propõe uma revisão contemporânea do observador ao nos sugerir uma tal compreensão, ressaltando o lugar central hoje ocupado pela visão enquanto interface dos sistemas humano–máquina. A autora sugere, então, em um desdobramento<sup>146</sup>, que talvez nesta posição já não se trataria tanto da visão, mas de um *reconhecimento de padrões*, que parece ser a operação solicitada a nosso olhar pelas imagens interativas das interfaces gráficas de computadores, celulares e outros dispositivos computacionais. O desenho de interfaces compreende, afinal, a elaboração de padrões e ícones compreensíveis, *intuitivos*, que o olho humano poderá rapidamente identificar ao percorrer a tela, depreendendo, de um só golpe, a ação esperada por aquela imagem. Escreve a autora:

Com efeito, na era da informação o que importa não é o que aparece na tela, mas o que está por trás dela, isto é, os algoritmos e seus agenciamentos, a informação enquanto diferença que faz a diferença. Nesse nível, a máquina não vê o que se apresenta mas “lê” as mais ínfimas transformações, numa velocidade e com uma

---

145 Cf. <http://www.amazon.com/Fire-Phone-13MP-Camera-32GB/dp/B00EOE0WKQ>

146 Subsidiada, nesta reflexão, tanto por Manovich quanto pela tese de doutoramento de Emerson Freire, defendida em 2012. FREIRE, Emerson. *Da sensação ausente à sensação como potência: tema e variações sobre a relação arte-tecnologia*. Tese (doutorado). Programa de Pós-graduação em Sociologia, Instituto de Filosofia e Ciências Humanas da Unicamp e Universidade Paris I – Panthéon-Sorbonne. 2012. Inédita. *apud* SENRA, 2013.



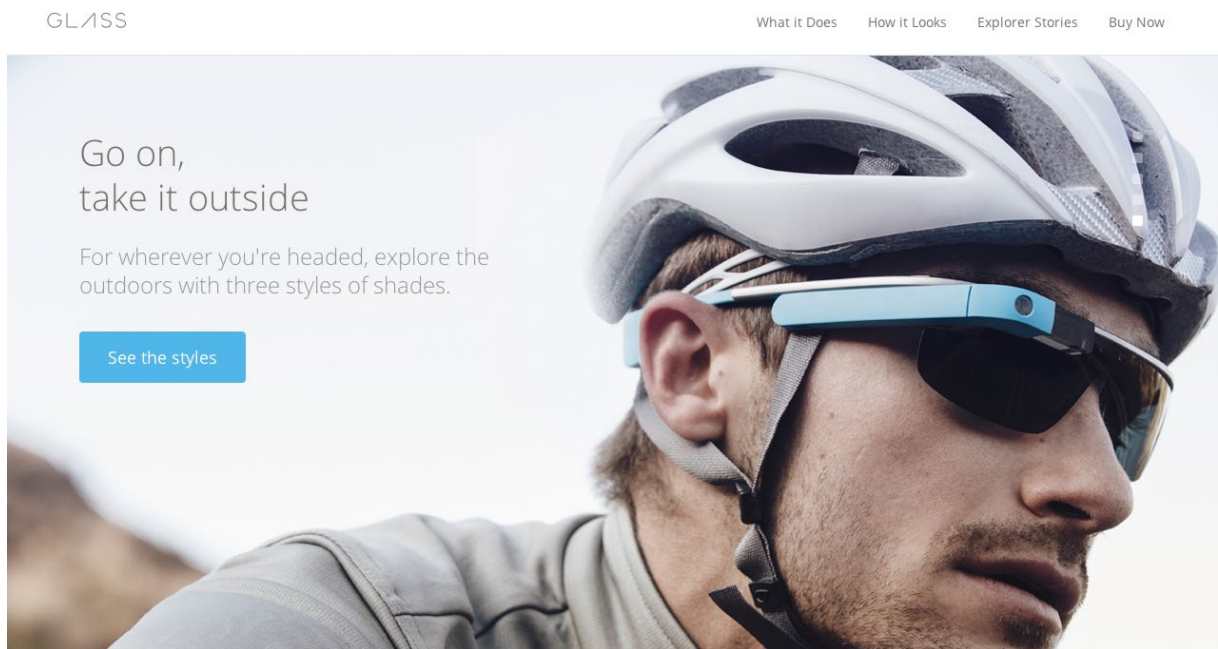


Figura 46 – Tela do *site* do projeto *Google Glass*, fora do ar desde janeiro de 2015. Fonte: <http://www.google.com/glass/start/>.

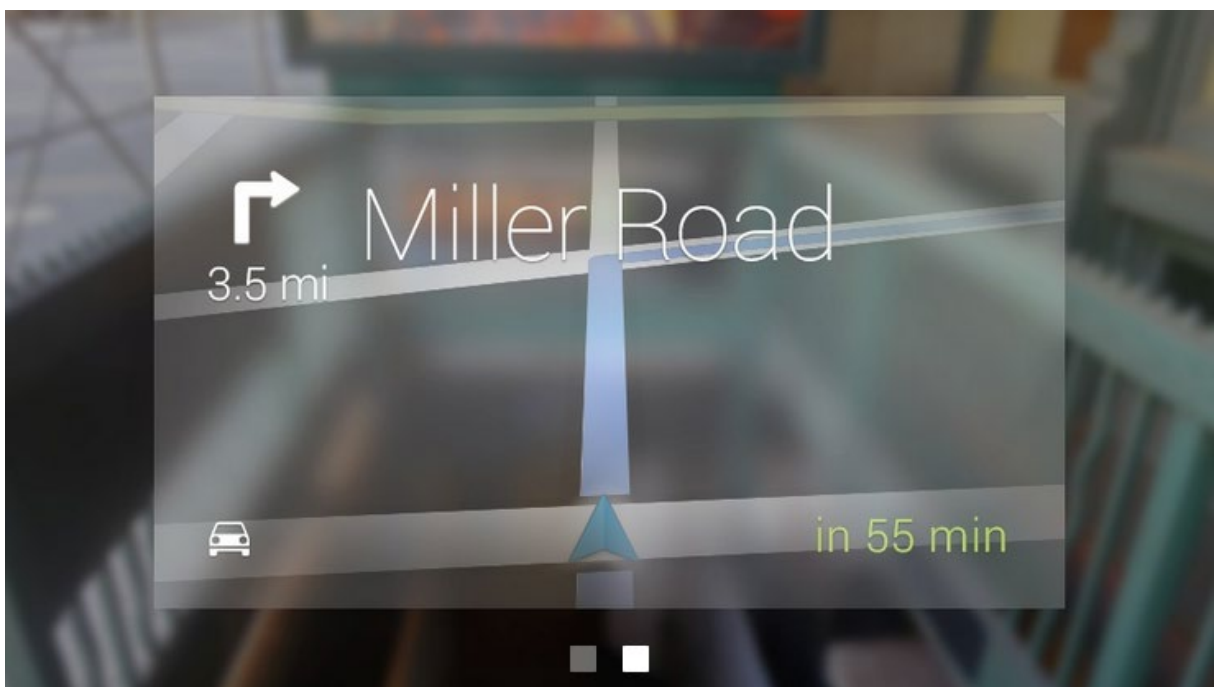


Figura 47 – Imagem ilustrativa da função de navegação do *Google Glass*, retirada do *site* do projeto, fora do ar desde janeiro de 2015. Fonte: <http://www.google.com/glass/start/>.



Figura 48 – Imagem de divulgação do projeto *HoloLens*, da Microsoft, apresenta uma representação daquilo que veria uma usuária do produto. Fonte: HEMPEL, 2015.

acuidade que o olho humano não pode alcançar. Portanto, na interação homem-computador, o observador atento não vê *através* da máquina, mas vê *com* ela o que reitera os padrões e o que deles escapa (SENRA, 2013, p. 18).

Trata-se de descrição que se potencializa quando nos atentamos aos dispositivos protéticos que hoje nos são apresentados, os quais sugerem todo o campo do visível como interface sobre a qual o olhar da máquina, acoplado ao nosso, nos auxiliaria a reconhecer padrões, com suas recorrências e anomalias, e orientar-nos à ação. Ver *com* a máquina, neste sentido, já seria algo bastante distinto daquilo que seria ver *com* a câmara escura, o telescópio ou o cinema. Pela Visão Computacional assistimos a entrada em cena, efetivamente, de um outro agente capaz não apenas de ordenar e realizar uma economia do visível, mas de derivar, automaticamente, o visível em ação. Se nos indagamos, então, quanto à constituição contemporânea do *observador*, não poderíamos compreendê-lo sem levarmos em conta, também, o tipo de capacidade para a visão que seria constituída nas *máquinas de visão*; ou, ainda, sem levarmos em conta o que a entrada destes novos agentes seria capaz de transformar na topologia da rede de que se compõem, hoje, o *ver* e o *visível*. O observador da Visão Computacional, neste sentido, parece absorver os modos de funcionamento que descrevemos anteriormente como características do funcionamento deste dispositivo, para o qual a percepção se traduziria, com frequência, em operações de *localização* e *reconhecimento*, seguidas de modos de *acionamento* e *conexão*, pelas quais ao visível se encadearia à ação. Trata-se de um modo de engajamento pressuposto não apenas por tecnologias contemporâneas que convocam a visão como interface de sistemas humano-máquina quanto por frentes de investigação acerca da visão, em particular em seus processos cognitivos – domínios que, como vimos, alimentam-se mutuamente.

\*

Compreendemos, nestes exemplos, o modo pelo qual os acoplamentos humano-máquina realizariam um governo ou uma regulação do *ver* e do *visível* – em um movimento talvez contrário ao que descrevemos pela experiência de *I'm Google*. Neste caso, afinal, talvez nos aproximemos da face de *decodificação* do olhar, que Johnston destaca em sua remissão ao duplo movimento dos agenciamentos maquínicos: de codificação, estratificação e territorialização, de um lado; e, de outro, decodificação, instabilidade e desterritorialização. Trata-se, talvez, do que poderíamos compreender como uma infiltração, no interior regulatório dos dispositivos, dos fluxos e dispersões da máquina abstrata que os anima. Alcançarmos esta instabilidade significa, talvez, encontrarmos brechas no governo do dispositivo, reconfigurá-lo, transgredi-lo – possivelmente, até, pela agência de *contradispositivos*. Na leitura deleuziana do conceito, seria encontrarmos suas linhas de atualização, de criatividade – a outra face do dispositivo, que ele diz ter sido menos explorada por Foucault (DELEUZE, 1999, p.161). Donna Haraway (2013), em outro registro, traz em sua conhecida alegoria do *ciborgue* um ponto de vista amparado nos potenciais transformadores destes acoplamentos, em que, transgredindo os limites do humano e, particularmente, da mulher, as transformações introduzidas pela articulação do corpo humano ao técnico seriam capazes de reconfigurar relações de dominação

há muito presentes, inclusive, no próprio desenvolvimento tecnológico. Nesta sua formulação, poderíamos compreender que o contato com a técnica promoveria uma decodificação do corpo, tornando-o em certa medida disponível para reconfigurações e ressignificações diversas – inclusive com o afastamento da figura do *ciborgue* de sua dita *origem* industrial ou militar. A este respeito, Haraway concebe o personagem desta sua alegoria enquanto um transgressor de fronteiras:

Os ciborgues não são reverentes, eles não conservam qualquer memória do cosmo: por isso, não pensam em recompô-lo. Eles desconfiam de qualquer holismo, mas anseiam por conexão – eles parecem ter uma inclinação natural por uma política de frente unida, mas sem o partido de vanguarda. O principal problema com os ciborgues é, obviamente, que eles são filhos ilegítimos do militarismo e do capitalismo patriarcal, isso para não mencionar o socialismo de estado. Mas os filhos ilegítimos são, com frequência, extremamente infiéis às suas origens. Seus pais são, afinal, dispensáveis (HARAWAY, 2013, p. 40).

A autora aposta, então, no *ciborgue* enquanto um transformador da própria tecnociência, colocada a serviço de uma reconfiguração das relações de dominação que, até então, a teriam sustentado. Não se trata, neste sentido, de recusar a ciência ou a tecnologia, inclusive porque a compreensão que ela desenvolve destes domínios – que é também a nossa – é a de que não se trata de um campo universal ou absoluto, mas composto por relações sociais (humanas e não-humanas), por agenciamentos nos quais podemos ainda intervir e transformar. Haraway advoga pela reorientação dos desenvolvimentos da tecnologia em favor de uma remodelação dos corpos e de transformação radical e performativa do humano. Assumindo que não é possível dizer, na relação humano–máquina quem faz e quem é feito (2013, p. 91), Haraway abraça a superação dos dualismos antes estancados pela cultura da técnica; contraposições como eu/outro, mente/corpo, cultura/natureza, macho/fêmea, entre outras, que apenas viriam estratificar relações de dominação entre estes supostos opostos. Ela nos leva, pela confusão conceitual, a um estado de possível indeterminação e revisão das formas constituídas.

Neste sentido, talvez seja ainda possível alcançar, por uma tradução da visão à máquina, uma modalidade de visão sintética guiada por uma tal perspectiva transgressora. Porém, não se trata de encarar positivamente qualquer efetuação desta transposição de limites pois, como vimos, haveria modos pelos quais o retraçar das fronteiras viria por meio de uma recodificação inversa ou de uma instrumentalização do olhar e da imagem que acabariam por reforçar, mais do que dissolver, modos da relação com o visível que não são exatamente próprios à técnica, mas que por ela são também elaborados e reforçados. Trata-se de aspecto que se torna evidente, como pudemos tratar, na forma pela qual a Visão Computacional articula particularização e conexão, com sua visualidade estatística e potenciais reforços de um olhar segregacionista, que se elabora na expectativa de um futuro, de uma virtualidade, expressa na singularidade aparente dos rostos, corpos e lugares. Seria por meio de uma tal operação que seria possível passar de *um* gato ou *um* rosto à sua figuração estatística, com todos os perigos de uma tal tradução. Além dela, pelas estratégias de localização e identificação, encontraríamos também um viés instrumental e utilitarista da visão, circunscrevendo uma visualidade particular. Conquanto reconheçamos a necessária mutualidade constitutiva que

torna qualquer traçar de fronteiras um ato construído, ética e politicamente implicado, isto não significa que o esboço de um tal limite não seja desejável ou necessário, pelo contrário. Tendo em vista, inclusive, o papel desempenhado pela indistinção humano–máquina em certo discurso conservador que viria retroalimentar ao humano as condições de funcionamento da máquina, talvez a elaboração de um tal limite constitua, efetivamente, nosso dever. Lucy Suchman (2007, loc. 5237), reconhecendo que o *ciborgue* de Haraway teria desempenhado um importante papel ao nos alertar para os aspectos políticos e transformadores das misturas humano–máquina, defende que seria o momento de compreender possíveis configurações a se desdobrar desta indiferenciação. Trata-se, sugere ela, de reencontrar diferenças expressas no próprio interior dos agenciamentos que abordamos:

Mutualidades, além disso, não são necessariamente simetrias. Minha própria análise sugere que pessoas e artefatos não constituem uns aos outros *da mesma forma*. Em particular, eu argumentaria que nós precisamos de uma rearticulação da assimetria, ou mais imparcialmente talvez, dissimetria, que de algum modo retenha o reconhecimento de híbridos, ciborgues e quase-objetos tornados visíveis através de estudos da tecnociência, enquanto simultaneamente recuperando certos posicionamentos sujeito–objeto – mesmo ordenamentos – entre pessoas e artefatos, e suas consequências<sup>147</sup> (SUCHMAN, 2007, loc. 5120-5131).

Em suma, ainda que reconheçamos a mutualidade constitutiva dos elementos que abordamos; que afastemos a possibilidade de definições essencialistas do humano e da máquina ou de suas visualidades; ou que valorizemos os aspectos transformadores que tais associações seriam capazes de provocar; isto não significa que não persistam diferenças ou que não devam ser retraçados, deliberadamente, limites produtivos pelos quais possamos situar-nos –, politicamente, inclusive – em relação a tais desenvolvimentos.

Neste ponto, afastamo-nos, é claro, da descrição ou da análise de nossos objetos – visando, inclusive, desprendermo-nos de sua configuração efetiva e concebermos possibilidades de escape. Vale dizer que neste movimento não efetuamos uma leitura teleológica que compreendesse os desdobramentos da Visão Computacional como predeterminados: o próprio entendimento que assumimos dos agenciamentos maquínicos pressupõe larga margem de indeterminação e imprevisibilidade. Contudo, os dispositivos nomeariam instâncias organizadoras destes fluxos, em alguma medida capazes de orientá-los – há, portanto, uma tendência à redução da indeterminação, por mais que ela não se complete. A questão, então, é a de como poderíamos agir no interior do funcionamento do dispositivo, compondo agenciamentos capazes de ativamente deformá-lo – deslocá-lo, arrastá-lo – em direção a outras configurações e outras modalidades de acoplamentos. Talvez isto passe, como sugere Suchman (2007, loc. 5247-5290), por práticas de *design* tecnológico mais articuladas

---

147 No original: “*Mutualities, moreover, are not necessarily symmetries. My own analysis suggests that persons and artifacts do not constitute each other in the same way. In particular, I would argue that we need a rearticulation of asymmetry, or more impartially perhaps, dissymmetry, that somehow retains the recognition of hybrids, cyborgs, and quasi-objects made visible through technoscience studies, while simultaneously recovering certain subject–object positionings – even orderings – among persons and artifacts and their consequences*”. Tradução nossa.

em redes de colaboração que tornem os dispositivos mais permeáveis a ajustes e reordenamentos coletivos – uma forma, por assim, dizer, de devolver a tecnologia a anseios e desejos compartilhados, ao bem comum<sup>148</sup>. Haveria algo desta vontade em parte do movimento pelas chamadas *tecnologias livres*<sup>149</sup> – por mais que sejam subsumidos, neste contexto, variados matizes ideológicos. Talvez, ainda, seriam pertinentes estratégias *neoluditas*, de efetiva sabotagem, que os trabalhos de Hersan e Harvey chegam a esboçar. Entre uma e outra possibilidade, encontraríamos movimentos de reconfiguração de nosso modo de assujeitamento aos dispositivos, que valem ser recuperados pelas potencialidades que abrem de outras constituições – menos hegemônicas, talvez – do *observador* da Visão Computacional. Sugerem, ainda, outras ordenações e dissimetrias, como sugere Suchman, em recodificações produtivas a partir da face de instabilidade dos acoplamentos humano–máquina.

Trata-se, é claro, de uma discussão que vai além de nosso interesse presente – restrito, em alguma medida, aos próprios contornos do dispositivo. Inclusive assumimos, em dado momento, as limitações de tal noção, dentre as quais sua melhor adequação à recomposição de traços hegemônicos, mais do que de resíduos ou resistências. Não se trata de ampliar, neste momento, o enfoque. De toda forma, acenamos a tais desdobramentos a fim de arejar um pouco nossa discussão e sugerir, talvez, um movimento de saída. A rede não se encontraria atada, por assim dizer, com *nós cegos*: seria sempre possível, em alguma medida, revisá-la e reconfigurá-la, desatando – ou mesmo cortando e rompendo – os *nós*. Trata-se, em cada caso, de encontrarmos as associações, acoplamentos e traduções capazes fazê-lo, bem como de encontrar, com calma, o ponto preciso a receber o corte. Seria este o momento em que, reconhecidas as mutualidades constitutivas e as interdependências, poderíamos chegar a assumir a narrativa de nossa própria constituição como *observadores* em um ato prático, analítico, deliberado, político e, talvez, mesmo arbitrário (SUCHMAN, 2007, loc. 5397), de construção de fronteiras.

---

148 Agamben (2009, p. 50-51) parece sugerir algo próximo disto pela sua proposição enigmática da *profanação* como modo de restituir ao uso comum aquilo que seria capturado e separado pelo dispositivo – trazendo à luz, sugere ele, “o Ingovernável”.

149 Um possível exemplo talvez encontraríamos no projeto *EyeWriter*, do coletivo novaiorquino *Graffiti Research Lab*, que desenvolveu um dispositivo de rastreamento ocular de baixo custo e de código aberto para habilitar um grafiteiro tetraplégico a desenhar com a movimentação da pupila, em uma efetiva reorientação, pela apropriação, do procedimento da Visão Computacional. Sua configuração remete, inclusive, àquela do *Google Glass* ou à dos estudos em ergonomia visual sugeridos por Cray (2014, p. 56). Porém, visa-se aplicações de outra natureza – desconectadas, inclusive, da rede de composição de perfis ou de adequação das imagens e da percepção a certo regime de eficiência. Cf. <http://www.eyewriter.org/>.







## Considerações finais

Elegemos a Visão Computacional como um *dispositivo* significativo das dinâmicas contemporâneas de produção e circulação do visível. Por meio dele, sugerimos que a imagem seria tomada em agenciamentos capazes de incorporá-la em variados contextos de acesso e de ação, em um plano tanto concreto quanto simbólico. Identificando em tal desenvolvimento tecnológico – indissociável em suas práticas e discursos – uma reencenação da metáfora *câmera-olho*, sugerimos sua relevância, em especial, na instauração de *visualidades* particulares – formulações historicamente localizadas da visão pelas quais se elaborariam modos circunscritos de nossa relação com o *visível*: disposições, concepções, posturas e práticas de observação. Indagamo-nos, então, quanto às suas implicações para o modo como concebemos e estudamos as imagens, considerando que a sugestão de uma tal *visão da máquina* – embora imaginativamente antecipada na ficção e mesmo na ciência – teria em sua efetiva aplicação e disseminação contemporâneas um fenômeno inaudito, até anos recentes. O problema proposto, portanto, foi o da descrição da Visão Computacional pela perspectiva dos estudos das imagens, com atenção especial às suas especificidades em meio às demais tecnologias do visual e aos contextos pragmáticos de sua aplicação.

Como já havíamos alertado à introdução, tratava-se, desde o início, de um problema decididamente amplo, que o percurso elaborado não seria capaz de resolver completamente, mas que, talvez, conseguiria colocá-lo em movimento. Compreendemos este trabalho, portanto, como um esforço inicial necessário para a descrição deste fenômeno, de relevância crescente em nosso campo de estudos, porém ainda com poucas pesquisas a ele dedicadas. Neste sentido, a opção por um trabalho que efetuassem sua descrição demonstrou-se estratégica, realizando, como sugerimos, a *composição* de nosso objeto a partir dos variados elementos pelos quais ele se manifesta e ganha forma. Reconhecendo os limites de nossa abordagem, que resolve apenas parcialmente o problema proposto ao início, esperamos que este esforço tenha conseguido dar algum corpo à Visão Computacional sob uma perspectiva de interesse aos campos de estudo da imagem, revisando e atualizando algumas das abordagens já existentes – por nós amplamente referenciadas – além de articulá-las a alguns exemplos contemporâneos de sua manifestação. Entendemos, neste sentido, que este trabalho pode ser produtivo na sugestão de outras indagações e problemas derivados – vetores que se estendem para além dos limites deste texto. Nestas considerações finais, gostaríamos, portanto, de não apenas recuperar aspectos centrais dos argumentos desenvolvidos, mas também indicar algumas das questões suscitadas e possíveis desdobramentos futuros.

Circunscrevendo e em larga medida organizando nossa abordagem, apontamos logo ao início duas questões que consideramos fundamentais quanto ao fenômeno estudado: nominalmente, seu modo de agenciamento da *imagem* e do *observador*. Sugerimos que seriam termos problematizados pelo próprio modo de operação da Visão Computacional, tecnologia apresentada como uma suposta

substituta técnica do olhar humano e que, embora articulada ao visual, não produz imagens – mas, de outro modo, as interpreta e mesmo as oculta. Em um esforço descritivo deste dispositivo pelo ponto de vista dos estudos da imagem, entendemos não apenas que a configuração destes aspectos colocavam-se como temas centrais por abordar como, também, seriam as próprias características que lhe confeririam relevância em meio a outros dispositivos comumente abordados. Iniciamos, então, com um capítulo dedicado à configuração técnica da Visão Computacional e a uma revisão das noções de *máquina* e *dispositivo* aplicadas a este domínio, compreendendo tal abordagem teórica como uma que nos permitiria passar destes aspectos intrínsecos aos modos de agenciamento que os conectam a seu exterior. Na sequência, assumimos duas diferentes modulações sugeridas pelos termos tidos como problemáticos, abordando em dois capítulos, respectivamente, as problemáticas da *imagem* e do *observador*, segundo o modo com que seriam agenciados pela Visão Computacional. Em cada um deles, buscamos compreender tanto em que medida haveria uma dependência ou a derivação pelo dispositivo de determinadas formas imagéticas e práticas observacionais quanto, ao mesmo tempo, o modo como ele lhes infligiria transformações.

Metodologicamente, atravessa nossa abordagem um esforço por *compor* o fenômeno estudado, compreendendo-o como uma emergência que encontraríamos na reunião e na conexão de diferentes objetos. Selecionamos, assim, um conjunto marcadamente heterogêneo, oriundo de domínios diversos: da arte contemporânea à produção técnica e científica, passando por antecedentes históricos da criminalística e criminologia e pelo cinema de ficção científica. Mais do que realizar, por meio deles, operações de cotejamento, comparação ou compilação, mobilizamos tais objetos de modo que sua aproximação pudesse ter um efeito de mútua amplificação, como se um elemento pudesse auxiliar-nos na compreensão de outros; como se intercambiassem certa luminosidade parcial pela qual tornar-se-iam, uns aos outros, mais inteligíveis. Tal é, em resumo, o principal gesto metodológico que recuperamos, ao primeiro capítulo, da leitura do *paradigma*, feita por Agamben (2002). Trata-se não apenas de um gesto que efetuamos entre os objetos, mas do modo que tomamos a própria Visão Computacional – campo tecnológico e científico de contornos relativamente definidos – enquanto um possível exemplar paradigmático que nos permitiria melhor compreender transformações em curso no âmbito do *ver* e do *visível*. Imaginamos, assim, um jogo complexo pelo qual este fenômeno *macro* seria uma decorrência dos objetos concretos que tomamos para análise, mas, ao mesmo tempo, a abstração metafórica que nos permitiria compreendê-los. Por articulações traçadas tanto no tempo quanto no espaço, realizamos neste percurso tanto uma espécie de *genealogia* – por meio da qual heranças e diferenciações foram descritas –, quanto algo como uma *cartografia* – pela qual dispomos tais objetos em relação, localizando-os em meio ao território de outras manifestações síncronas.

No primeiro capítulo atentamo-nos, então, principalmente à mediação técnica da Visão Computacional, descrevendo suas operações características a partir de referências primárias do campo, remissões a suas aplicações cotidianas e análises de algumas de suas apropriações no âmbito

da arte. Por meio destas descrições iniciais buscamos chegar, já em um segundo nível da análise, a uma compreensão dos agenciamentos que seriam efetivamente constituídos, em que os processos interpretativos aplicados à imagem digital se desdobram em modos particulares de ação sobre os corpos e o espaço. Neste sentido, a operação primária da Visão Computacional, que sugerimos se tratar da tradução da imagem a um domínio informacional<sup>150</sup>, se desdobra em modalidades de articulação do visível a contextos de ação, tanto sobre um espaço concreto quanto sobre um espaço simbólico – uma diferenciação importante entre as duas ênfases que alcançamos em nossa descrição: *localização–acionamento* e *reconhecimento–conexão*. Pela primeira, nos referimos a operações de circunscrição e varredura em que o visível é tomado como índice da localização e do posicionamento dos corpos no espaço. Encontramos tal agenciamento, mais marcadamente, em instalações da artemídia e em aplicações corriqueiras da Visão Computacional, em especial no âmbito da videovigilância. De modo importante, trata-se ainda de um modo de agenciamento que demanda dos corpos determinadas ações (aciona-os, também) diante dos padrões de resposta e dos acionamentos internos aos programas. Pela segunda ênfase, *reconhecimento–conexão*, referimo-nos a configurações distribuídas em que a operação do dispositivo abordado não seria descrita pela circunscrição de um território mas pela articulação de um recorte do visível a uma rede simbólica. Identificamos manifestações cotidianas deste agenciamento, especialmente, articuladas ao contexto da internet, em mecanismos de busca. Apontamos também, como uma incidência pontual, a instalação de David Rokeby, *The giver of names*. Inserido em um contexto de agregação de dados comportamentais dos indivíduos, sugerimos a vinculação deste modo de funcionamento às operações contemporâneas de elaboração de perfis informacionais de consumo e de risco. Trata-se, assim, de um agenciamento que tem à sua jusante contextos pragmáticos que vão, também, além do plano simbólico, com a mobilização de ações no espaço concreto (como a compra de um produto, a visita a um museu ou mesmo a detenção de um suspeito). Em suma, em ambos conjuntos de operações identificamos, com matizes distintos, uma profunda articulação entre modos de percepção e modos de ação, pelo qual descrevemos a visão da máquina como orientada a contextos pragmáticos.

A articulação que elaboramos, em meio a tal descrição, entre a formulação por Deleuze e Guattari (1995a e GUATTARRI, 2003) dos *agenciamentos maquínicos* e aquela de Foucault (1999 e 1979), Deleuze (2005 e 1999), Agamben (2009) e Crary (1992) para o *dispositivo* – temperada pelas interseções com a teoria da ação e da mediação técnica de Latour (2005 e 2001) – amparou-nos na passagem entre esta instância mais técnica da análise e as elaborações subsequentes, voltadas principalmente aos atravessamentos da máquina por vetores externos. Enumeramos, a seguir, alguns ganhos desta articulação teórica e de nossa compreensão da Visão Computacional como dispositivo. Primeiramente, tal construção permitiu-nos uma capacidade de articulação entre o agenciamento concreto do dispositivo e as traduções dispersas que encontramos em seu entorno, pela qual tomamos

---

150 Vale destacar, falamos não de uma tradução simples pela qual, decomposta em *pixels*, a imagem ganharia uma representação numérica para cada porção de sua matriz. A tradução efetuada pela Visão Computacional, de outro modo, implica em algum nível de interpretação que leve em conta aspectos daquilo que a imagem representaria.

a compreensão de Ferraz (2013) para quem ele compreende não uma causa ou consequência de seu contexto, mas, simultaneamente, seu instrumento e expressão. Neste sentido nos afastamos tanto de compreensões deterministas e tecnocêntricas quanto daquelas que compreenderiam a técnica como um elemento menor ou apenas sintomático de determinado processo. A compreensão de Deleuze (2005) da articulação entre uma máquina abstrata, ou um diagrama, e o agenciamento concreto do dispositivo diz justamente da possibilidade desta passagem entre elementos discursivos e não discursivos e entre o dentro e o fora de um dispositivo – que tanto lhe dá forma quanto é informado por seu agenciamento. Derivado deste primeiro ponto, destacaríamos como segundo a compreensão do dispositivo como uma instância emergente de regulação das trocas realizadas dispersamente por uma tal máquina abstrata. Reconhecendo o caráter eminentemente caótico das associações que reconheceríamos em tal agenciamento maquínico, tomamos de Agamben (2009) e de Latour (2001) o entendimento do dispositivo enquanto instância de absorção das agências, ou de seu governo, na organização do curso da ação. Tratamos da Visão Computacional, portanto, como uma instância emergente que conectaria e regularia a interação entre elementos diversos, humanos e não-humanos, participantes das traduções contemporâneas do visível. Um terceiro ganho por destacar é, então, a superação logo de início de uma oposição fundamental entre o humano e a máquina, ou o orgânico e o técnico. Trata-se de um ponto recuperado adiante, mas que, posto de início, já matiza nossa abordagem evitando contraposições estanques entre estes termos, permitindo que nos atentemos às traduções entre eles, e não apenas às sua contraposição. Portanto, sugerimos o caráter distribuído da visão enquanto um ato realizado não apenas por agentes humanos, mas sempre com a mediação, mais ou menos intensa, de artefatos técnicos mobilizados em sua realização. Ainda articulado a este referencial, teríamos como quarto ganho o valor paradigmático do dispositivo, o qual é, então, também entendido em nossa abordagem como objeto capaz de lançar luz sobre os processos de que toma parte.

Do agenciamento maquínico ao dispositivo, encontramos, então, modos de descrever as *máquinas que veem* e seus modos de operação não apenas do ponto de vista de sua constituição técnica mas, principalmente, pelo modo com que tal constituição se articula a agenciamentos externos. Trata-se menos de compreender a máquina como ente isolado, portanto, do que como um ser que toma corpo e forma pelo modo com que se relaciona com seu exterior. Foi neste sentido que articulamos, então, a Visão Computacional a demandas do domínio da vigilância e do monitoramento nas redes. Porém, de modo talvez ainda mais profundo, abarcando tais articulações, trata-se de dispositivo responsável por efetuar uma passagem significativa entre a imagem – com sua *opacidade computacional* – e a informação. Sugerimos, então, que ela realizaria uma tradução da imagem àquelas que Sean Cubitt (2011) identifica como as mídias dominantes de nosso tempo: as planilhas, os bancos de dados e os sistemas de georreferenciamento. Trata-se, em todo caso, de quantificar o visível em um processo automatizado, de colocá-lo a serviço daquilo que Fernanda Bruno (2013) identifica como um regime de eficiência informacional, que guia formas contemporâneas tanto de produção de conhecimento quanto de exercício do poder. Este talvez seja um dos ganhos centrais

desta primeira parte da investigação. Como um eixo a atravessar as diferentes escalas de análise, veríamos se articular, das operações mais elementares da Visão Computacional a seus agenciamentos mais abrangentes, um movimento contínuo em que a tradução informacional da imagem vem realizar, em um escopo mais amplo, uma integração.

No segundo capítulo, dedicamo-nos ao modo com que a Visão Computacional agencia a imagem em seu interior, o que sugerimos, metaforicamente, como aquilo que as máquinas *veem*. Embora evidentemente articulada com a discussão anterior, buscamos nesse segundo momento recuperar a própria relevância da *imagem* enquanto categoria de nossa abordagem, já que, por não abordarmos uma tecnologia produtora de imagens, talvez julgássemos este aspecto, inicialmente, como de menor relevância. Contudo, em uma reversão desta compreensão, defendemos não apenas a centralidade da imagem como um componente do funcionamento desse dispositivo, como, também, a particularidade de seu modo de interpretação do visível, circunscrevendo, embora não um modo de registro, um modo de observação. Sugerimos duas elaborações para aprofundarmos tal discussão. Primeiramente, tratamos da sobreposição de *imagem* e *instrumento* nos registros operacionais tradicionalmente tomados por tais máquinas, mas cuja articulação também incide no caráter instrumental pelo qual interpretariam mesmo imagens que a princípio não reconheceríamos como instrumentais. Em um segundo momento, sugerimos ainda a existência de uma instância, embora oculta, em que a Visão Computacional produziria, sim, em certo sentido, imagens. Trata-se da dimensão que sugerimos, a partir de Paul Virilio (1994 e 1993), como a *visão subjetiva* da máquina, em uma analogia daquilo que seria, para nós, a elaboração de *imagens mentais*.

Encontramos na sobreposição entre imagem e instrumento uma chave importante, que traz certa correspondência com as operações de *localização–acionamento*. Em seu caráter instrumental, a imagem ocupa uma posição particular no dispositivo abordado: sugerimos que ela constitui uma parte intrínseca a seu mecanismo. Remetemos, a partir de Lev Manovich (1993a), William Ivins (1975) e Bruno Latour (1986), às importantes heranças que os esforços de síntese da visão receberam da perspectiva linear ou mesmo da câmera fotográfica, em um processo de geometrização da representação visual, atendendo às demandas da racionalidade moderna. Desenvolvemos, a partir dos registros documentais de tais máquinas, trazidos por Harun Farocki em *Eye/Machine*, uma leitura de como veríamos incidir sobre tal registro aspectos desta racionalização. A Visão Computacional desponta, então, como uma culminância de um processo de instrumentalização da imagem, tomada como via de passagem, de mão dupla, entre o referente e sua representação. Recuperamos, então, a compreensão de Manovich (1993a) para quem ela teria automatizado a passagem – ainda que simulada – da imagem ao seu referente, da mesma forma que a fotografia teria já automatizado a passagem inversa, com a produção óptico-mecânica da imagem em perspectiva. Atravessando tais desenvolvimentos, encontramos aquilo que o autor sugere como um *nominalismo visual*, pelo qual a imagem seria tomada pela sua capacidade de registro do individual e do singular e, ainda, como instância de conhecimento e de ação sobre o mundo. Trata-se, como vimos em

outro exemplo, de racionalidade similar à que encontramos nos métodos analíticos de identificação criminalística elaborados por Bertillon, que também se valem desta dimensão instrumental da imagem em práticas de arquivamento e catalogação.

Mais além desta atenção aos corpos e espaços singulares, contudo, a remissão a uma possível *visão subjetiva* da máquina – às representações internas a seu funcionamento – nos sugerem outra compreensão em que o visível seria agenciado, principalmente, em função de conceitos e categorias abstratas. Em um sentido contrário ao *nominalismo* sugerido anteriormente, visualizações produzidas em pesquisa de cientistas da *Google* revelaram-nos um movimento *idealista* no qual, a partir de figurações singulares, a Visão Computacional elabora composições que representam não *um* gato, como pudemos ver, mas algo como a *ideia* ou o *conceito* de gato. Neste sentido, tal representação interna, oriunda do próprio funcionamento de tais máquinas (sua *visão subjetiva*), revela-nos algo como um movimento reverso à singularização, pelo qual a imagem permitiria à máquina aceder ao abstrato. Neste movimento, ela realiza uma passagem entre uma *atualidade* e uma *virtualidade* da percepção, de modo que a imagem produzida neste processo não aponta a um presente percebido mas, fundamentalmente, à condensação de um passado e, também, de um futuro. Há, evidentemente, certa correspondência entre tal *visualidade estatística* e as operações de *reconhecimento–conexão* que sugerimos no capítulo anterior: é por este movimento de abstração que é permitido à máquina traçar relações e aproximações entre distintos recortes do visível, agenciando uma imagem em meio a uma ampla rede de relações lógicas e semânticas que estabelece, em cadeia, com inúmeras outras.

Entre o *nominalismo* e o *idealismo*, contudo, não estabelecemos uma contraposição, mas duas visadas distintas articuladas pela Visão Computacional. O modo do agenciamento do visível por tais máquinas efetua-se, pois, na passagem entre estas duas compreensões, permitindo-lhe conectar procedimentos de circunscrição e mapeamento de espaços e corpos singulares a uma virtualização do visível, tecendo entre um amplo repertório visual conexões dispersas. Identificamos, nesta operação, uma forte implicação política na medida em que o que subsidia as tarefas de reconhecimento e conexão são perfis estatísticos elaborados a partir de figurações médias de uma determinada amostra previamente conhecida. As categorias expressas desta forma, às quais singularidades são subsumidas, apresentam uma forte tendência ao reforço de uma visualidade que diríamos segregacionista pela qual um corpo singular potencialmente se vincula, com base apenas em sua aparência, a uma categoria construída por incidências mapeadas anteriormente – tendo, como exemplos potenciais: jovens delinquentes, suspeitos de terrorismo, imigrantes ilegais. Sugere-se, assim, ainda que por vieses probabilísticos, uma predeterminação futura dos corpos percebidos, com sua singularidade articulada, pela Visão Computacional, a quadros estatísticos de comportamentos esperados.

Por fim, no terceiro capítulo, dedicamo-nos a compreender canais de retorno pelos quais as máquinas de visão informariam aspectos de uma visualidade compartilhada, também, por observadores

humanos. Neste sentido, esboçamos aspectos do agenciamento das modalidades perceptivas dos sujeitos, uma passagem da visão *das máquinas* às práticas de observação dela derivadas, em um aprofundamento das consequências da superação – sugerida ao início – da contraposição entre o humano e a máquina. Logo, não abordamos a questão como uma via de mão única: tanto a Visão Computacional deriva de práticas de observação já relativamente constituídas, quanto a visão humana incorpora práticas derivadas desta tradução. Na confluência destes dois processos, deste trânsito, sugerimos, a partir de Crary (1992), a constituição de um *observador* pelo dispositivo da Visão Computacional. Como chegamos a desenvolver, trata-se de uma noção com algumas limitações a que precisamos nos atentar – destacamos duas. Uma é a sua remissão não a sujeitos empíricos, mas a construções teóricas e em alguma medida emblemáticas daquele que seria o sujeito engendrado pelo dispositivo. Indicamos, assim, um processo de subjetivação que pode coexistir com outros pelos que passam um mesmo indivíduo – sem confundirmos, como sugere Agamben (2009), ser e sujeito. Outra consequência, decorrente dessa, é a sua melhor adequação para compor traços hegemônicos do funcionamento do dispositivo, sem dar conta de processos secundários e periféricos ou, ainda, práticas de resistência. Por tais aspectos, compreendemos a sugestão do *observador* como um ferramenta conceitual para a lida com o próprio *dispositivo*, particularmente em seu modo de agenciamento da percepção. Em todo caso, como reiteramos, devemos reconhecer a impossibilidade de projetarmos diretamente um tal *observador* sobre um indivíduo singular tomado de uma empiria. O principal valor desta formulação, contudo, encontra-se na composição, sob uma mesma figura teórica, dos diversos elementos que constituiriam, em determinado momento e sob determinado regime, aquilo que compreenderíamos como *ver*. Ele aponta, então, para uma complexa construção do ato perceptivo que seria elaborada pelos *dispositivos* que abordamos, nomeando o sujeito que emergiria de sua operação de governo.

O observador da Visão Computacional, neste sentido, pudemos descrever como tanto articulado a práticas observacionais decorrentes de uma crescente confusão de limites entre os dispositivos que tradicionalmente mediarão nossa relação com a imagem, quanto envolvido em processos de racionalização ou – como sugerimos a partir de Crary (2014) – de *comoditização* do olhar. Em ambos os casos, sugeriríamos se tratar de uma *crise* da constituição do observador, embora por vieses distintos. Em primeiro momento, dissemos que ela se daria por uma potencial indistinção entre processos de regulação tradicionais, devido à diluição das fronteiras entre diferentes dispositivos – a condição “metaespetacular” de que fala Comolli (2008). Em um segundo aspecto, de modo ainda mais relevante para a Visão Computacional, referimo-nos ao novo lugar ocupado pela prática observacional que não apenas é regulada como torna-se, ela própria, objeto de análise e observação pelos dispositivos que a governam – como sugere-nos Crary (2014). Articulam-se, neste domínio, diversas novas configurações das mediações técnicas da visão que parecem nos inserir em uma articulação maquínica por meio da qual, literalmente, vemos *com (junto às)* máquinas. Sobrepondo a nosso campo de visão uma camada de interpretação e codificação do visível, tais mediações sugerem um desdobramento extremo em que nosso olhar seria assimilado ao da máquina. De modo

igualmente profundo, encontramos na Neurociência tentativas de acesso às nossas *imagens mentais*, sugerindo, também neste nível cognitivo da visão, sua disponibilidade ao governo dos dispositivos.

Como sugerimos desde o início, as mutualidades constitutivas entre o olhar humano e suas mediações técnicas não são recentes. Contudo, o estágio de desenvolvimento destas máquinas de visão revela-nos um modo marcadamente distinto da efetuação desta relação na medida em que, sob a promessa de constituição de uma visão sintética, vemos operar uma retroalimentação de sua lógica de funcionamento que não apenas nos leva a incorporar aspectos desta visualidade codificada quanto, também, renova e aprofunda os modos de gestão e governo tanto do ver quanto do visível. Indicamos, em certa passagem, a partir de Suchman (2007), as potencialidades de compreendermos estas aproximações e eventuais distinções como *performadas*. Neste sentido, o que encontramos nestes exemplos poderíamos compreender como performances de modos de assimilação entre o olhar humano e o da máquina – da mesma forma que, em outros contextos, como nos *CAPTCHA* e desenvolvimentos derivados, encontraríamos a performance de sua distinção. Nos exemplos mais recorrentes contudo, que trouxemos no segundo capítulo e ao final do terceiro, vemos se generalizar o uso instrumental das imagens para além não apenas das aplicações marcadamente instrumentais (da ciências e engenharias), como, também, para além das imagens produzidas com este fim. Da mesma forma, vemos se generalizar práticas observacionais que tomam como princípio a mesma lógica de funcionamento destes dispositivos, articulando *nominalismo* e *idealismo* e *localização–acionamento* e *reconhecimento–conexão* nos contextos mais diversos. Evidentemente, indicamos exemplos que possuem, ainda, certa restrição de sua aplicabilidade. Porém, encontramos, ao mesmo tempo, promessas de sua generalização, as quais não devemos ignorar.

Bem sabemos como práticas de gestão informacional e formações do saber de base estatística encontram-se bastante disseminadas, inclusive, em contextos controversos como, por exemplo: a produtividade acadêmica, a avaliação de processos educacionais, ou a análise de risco em procedimentos de investimento financeiro e contratação de funcionários. Já acompanhamos, também, as distorções oriundas deste tipo de abordagem calcada em modos de articulação entre o singular e o categórico, tendendo à anulação das diferenças em processos de tradução generalizada à informação. Haveria, certamente, muito ainda por discutir acerca deste regime de eficiência informacional sugerido por Fernanda Bruno (2013), que não chegamos a aprofundar aqui. Em todo caso, vale destacar que a entrada da imagem a este domínio – como sugerimos ser a operação da Visão Computacional – não constitui um aspecto menor. Por meio desta passagem seria possível, portanto, que tal lógica de produção de conhecimento e exercício do poder se aplique não apenas a indicadores decorrentes – dentre outros aspectos – de nossa atividade *online* ou de nossos padrões de consumo (embora neste estágio já sejam preocupantes), mas também de nossa aparência. Trata-se de um ponto relevante não apenas em função do reforço que tal funcionamento sugeriria a práticas racistas, xenófobas e sexistas quanto, também, pela ampla disponibilidade de nossos corpos ao campo de visão de uma câmera qualquer. Se em determinados contextos ainda teríamos uma margem



– ainda que reduzida – de evitarmos nos expor ou nos identificar em determinados contextos, um tal dispositivo, que pelo campo de visão da câmera demarca seu território, certamente encontraria, no atual contexto de multiplicação das lentes, um território extremamente amplo para sua operação.

Muitas destas práticas vinculam-se, evidentemente, à discussão da vigilância e do controle, domínios que, como indicamos, têm o dispositivo que abordamos como um desenvolvimento central para a contemporaneidade de sua discussão. Compartilhando de tal inquietação, desdobramos, aqui, outra questão, quanto ao que esperamos de nossas práticas observacionais mesmo em outros domínios. Ou, ainda, quanto ao que esperamos de nossas imagens. Se, como sugerimos, nos aguarda uma codificação mais profunda de nosso próprio olhar em função de tais desenvolvimentos, devemos esperar reverberações de um tal processo em nossa percepção estética, em nossas organizações sociais e políticas, em nossas formações de saber – enfim, seria demasiado enumerar. Se, como sugerem-nos Ivins (1975) e Latour (1986), a perspectiva teria tido tamanha importância na construção de certa modernidade – em formações do saber e do poder, dir-nos-ia Foucault –, o que esperar dos modos de agenciamento do olhar e do visível engendrados pela Visão Computacional? Sem interpor a um tal desenvolvimento uma barreira efetiva, indagamo-nos sobre a profundidade de suas consequências e sobre a sua capacidade, ou não, de alastrar-se como, efetivamente, promete. Porém, indagamo-nos ainda quanto às possibilidades de intervenção, de desvio, de subversão de seu funcionamento, talvez orientando-o a outros potenciais.

Em alguma medida, tal possibilidade encontra-se sugerida pela própria figura que escolhemos para nomearmos os produtos desta investigação: a *composição*. Junto às potencialidades desta elaboração que tomamos de Bruno Latour (2012 e 2010), encontramos a posta em um terreno comum de objetos heterogêneos sem, contudo, perdermos suas diferenças. A composição visa, neste sentido, reencontrar um plano comum, compreender os termos por suas aproximações e ressonâncias sem por isto categorizá-los em estruturas arborescentes: reconhecendo o lugar singular ocupado por cada objeto neste arranjo, traçamos entre eles alguns caminhos – longe de exclusivos, exaustivos ou não-ambíguos. Do lado das limitações, temos que neste gesto não superamos, evidentemente, a condição fragmentária do arranjo elaborado, o qual tanto deixa lacunas quanto não chega a tecer entre os elementos reunidos tamanha coesão a ponto de o conjunto ser indecomponível. Trata-se, talvez, de uma limitação que converte-se em potencialidade: por tal forma, a *composição* seria um processo sempre revisável e reversível. Por mais que em uma tal concepção reconheçamos que o construto assim elaborado expressaria sempre uma leitura relativa, talvez também preservemos a maleabilidade que nos permitiria decompô-lo de forma menos traumática a fim de acrescentar outros elementos, traçar conexões distintas – ou, ainda, apenas desfazê-lo, se assim julgarmos necessário. Reconhecer o caráter construído da narrativa apresentada não equivale, portanto, a fragilizá-la a ponto de torná-la vã: esperamos, de outro modo, que ela auxilie a elaboração de sentidos acerca do fenômeno abordado, colocando em movimento o exercício de sua compreensão e, também, o de sua contestação política e estética. Inclusive, tendo em vista algumas das implicações dos processos

descritos, diríamos que este estudo almeja por sua contradição futura, pela construção de outros caminhos em meio ao território mapeado.

## Referências

2001: a space odyssey. Direção: Stanley Kubrick. Lançado em 1968. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (148 min.), son., color., widescreen.

A ORIGEM. Direção: Christopher Nolan. Lançado em 2010. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (148 min.), son., color., widescreen. Tradução de: *Inception*.

AGAMBEN, Giorgio. O que é um dispositivo?. In: \_\_\_\_\_. *O que é o contemporâneo? e outros ensaios*. Chapecó, SC: Argos Editora, 2009. p. 25-51.

\_\_\_\_\_. What is a paradigm?. Transcrição da palestra proferida na European Graduate School em agosto de 2002. Disponível em: <<http://www.egs.edu/faculty/giorgio-agamben/articles/what-is-a-paradigm/>>. Acesso em: 2 nov 2013.

ANWAR, Yasmin. Scientists use brain imaging to reveal the movies in our mind. *UC Berkeley News Center*, 22 set 2011. Disponível em: <<http://newscenter.berkeley.edu/2011/09/22/brain-movies/>>. Acesso em: 14 jul 2014.

ARANTES, Priscila. *@arte e mídia: perspectivas da estética digital*. São Paulo: Ed. Senac São Paulo, 2005.

AUMONT, Jacques. *A imagem*. Campinas: Papyrus, 1993.

BAGGIO, Daniel Lelis *et al.* *Mastering OpenCV with practical computer vision projects*. Birmingham, UK: Packt Publishing, 2012.

BATCHEN, Geoffrey. Enslaved, sovereign, observed spectator: on Jonathan Crary, techniques of the observer. *Continuum: the Australian journal of media & culture*, v. 6, n. 2. 1991. Disponível em: <<http://dx.doi.org/10.1080/10304319309359399>>. Acesso em: 05 fev 2015.

BAUDRY, Jean-Louis. Cinema: efeitos ideológicos produzidos pelo aparelho de base. In: XAVIER, Ismail (org.). *A experiência do cinema: antologia*. Rio de Janeiro: Graal, 1983.

BAZIN, André. A ontologia da imagem fotográfica. In: \_\_\_\_\_. *O cinema: ensaios*. São Paulo: Brasiliense, 1991.

BENJAMIN, Walter [*et al.*]. *Benjamin e a obra de arte: técnica, imagem, percepção*. Rio de Janeiro: Contraponto, 2012.

BENJAMIN, Walter. Sobre alguns temas em Baudelaire. In: \_\_\_\_\_. *Charles Baudelaire um lírico no auge do capitalismo*. São Paulo: Brasiliense, 1995.

\_\_\_\_\_. *Magia e técnica, arte e política: ensaios sobre literatura e história da cultura*. São Paulo: Brasiliense, 1994.

BERRY, David M.; VAN DARTEL, Michel; DIETER, Michael; KASPRZAK, Michelle; MULLER, Nat; O'REILLY, Rachel; VICENTE, José Luis de. *New aesthetics, new anxieties*. Rotterdam: V2\_ Institute for Unstable Media, 2012. Disponível em: <[http://v2.nl/files/2012/publishing/new-aesthetic-new-anxieties-pdf/at\\_download/file/view](http://v2.nl/files/2012/publishing/new-aesthetic-new-anxieties-pdf/at_download/file/view)>. Acesso em: 1º out 2013.

BILTON, Nick. Why Google Glass broke. *The New York Times*. 4 fev 2015. Disponível em: <<http://www.nytimes.com/2015/02/05/style/why-google-glass-broke.html>>. Acesso em 5 fev 2015.

BLADE Runner. Direção: Ridley Scott. Lançado em 1982. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (117 min.), son., color., widescreen.

BORGES, Jorge Luis. Del rigor en la ciencia. In: \_\_\_\_\_. *El hacedor*. Barcelona: Alianza Editorial, 1998.

BRAGA, José Luiz. Dispositivos interacionais. In: ENCONTRO DA COMPÓS, 20., 2011, Porto Alegre. *Anais...* Disponível em: <[http://compos.org.br/data/biblioteca\\_1657.doc](http://compos.org.br/data/biblioteca_1657.doc)>. Acesso em: 3 abr 2014.

\_\_\_\_\_; SALOMÃO, Mozahir; JESUS, Eduardo de. Entrevista com José Luiz Braga. *Dispositiva*, Belo Horizonte, v. 1, n. 1, maio/out 2012. p. 29-38. Disponível em: <<http://periodicos.pucminas.br/index.php/dispositiva/article/view/2817/3103>>. Acesso em: 5 fev 2015.

BRASIL, André; JESUS, Eduardo de. Poéticas do dispositivo: notas sobre um percurso conceitual. In: JESUS, Eduardo; SALOMÃO, Mozahir (orgs.). *Interações plurais: a comunicação e o contemporâneo*. São Paulo: Annablume, 2008.

BRIDLE, James. #sxaesthetic. 2012. Disponível em: <<http://booktwo.org/notebook/sxaesthetic/>>. Acesso em: 27 mar 2013.

\_\_\_\_\_. The new aesthetic: waving at the machines. 2011c. Disponível em: <<http://booktwo.org/notebook/waving-at-machines/>>. Acesso em 27 mar 2013.

\_\_\_\_\_. Waving at the machines. 2011b. Disponível em: <<http://www.webdirections.org/resources/james-bridle-waving-at-the-machines/>>. Acesso em: 27 mar 2013.

\_\_\_\_\_. The new aesthetic. 2011a. Disponível em: <<http://www.riglondon.com/blog/2011/05/06/the-new-aesthetic/>>. Acesso em: 27 mar 2013.

BRILHO eterno de uma mente sem lembranças. Direção: Michel Gondry. Lançado em 2004. [S.l.]: Netflix, 2014. *Streaming*. (107 min.), son., color., legendado. Tradução de: *Eternal sunshine of the spotless mind*.

BRUCK, Mozahir Salomão. Palavra: dispositivo. *Dispositiva*, Belo Horizonte, v. 1, n. 1, maio-out 2012. Disponível em: <<http://periodicos.pucminas.br/index.php/dispositiva/article/view/2514/3106>>. Acesso em: 5 fev 2015.

BRUNO, Fernanda. *Máquinas de ver, modos de ser: vigilância, tecnologia e subjetividade*. Porto Alegre: Sulina, 2013.

\_\_\_\_\_. Contramanual para câmeras inteligentes: vigilância, tecnologia e percepção. *Galáxia*, São Paulo, n. 24, p. 47-63, dez. 2012. Disponível em: <<http://revistas.pucsp.br/index.php/galaxia/article/view/9807/9426>>. Acesso em: 5 fev 2015.

\_\_\_\_\_. Máquinas de ver, modos de ser: visibilidade e subjetividade nas novas tecnologias de informação e de comunicação. *Revista FAMECOS*, Porto Alegre, v. 24, p. 110-124, 2004. Disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/revistafamecos/article/view/3271/2531>>. Acesso em: 5 fev 2015.

\_\_\_\_\_; KANASHIRO, Marta; FIRMINO, Rodrigo (orgs.). *Vigilância e visibilidade: espaço, tecnologia e identificação*. Porto Alegre: Sulina, 2010.

\_\_\_\_\_; VAZ, Paulo. Agentes.com: cognição, delegação, distribuição. *Contracampo*, Niterói, n. 7, pp. 23-38, 2º semestre 2002. Disponível em: <<http://www.uff.br/contracampo/index.php/revista/article/view/473/238>>. Acesso em: 5 fev 2015.

CASTELFRANCHI, Yuriy. Biting the apple: the challenge of Artificial Intelligence. *MultiCiência*, v. 3, Campinas, SP, out 2004.

COMOLLI, Jean-Louis. Retrospectiva do espectador. In: \_\_\_\_\_. *Ver e poder: a inocência perdida: cinema, televisão, ficção, documentário*. Belo Horizonte: Editora UFMG, 2008. p. 135-142.

CONNOR, Michael. Hito Steyerl's 'How not to be seen: a fucking didactic .MOV file'. *Rhizome*, 31 maio 2013. Disponível em: <<http://rhizome.org/editorial/2013/may/31/hito-steyerl-how-not-to-be-seen/>>. Acesso em 28 jan 2015.

COUCHOT, Edmond. *A tecnologia na arte: da fotografia à realidade virtual*. Porto Alegre: Ed. UFRGS, 2003.

CRARY, Jonathan. *24/7: capitalismo tardio e os fins do sono*. São Paulo: Cosac Naify, 2014.

\_\_\_\_\_. A visão que se desprende: Manet e o observador atento no fim do século XIX. In: CHARNEY, Leo; SCHWARTZ, Vanessa. *O cinema e a invenção da vida moderna*. São Paulo: Cosac & Naify, 2004. p. 67-94.

\_\_\_\_\_. *Suspensions of perception: attention, spectacle and modern culture*. Cambridge, Massachusetts: MIT Press, 2001.

\_\_\_\_\_. *Techniques of the observer: on vision and modernity in the nineteenth century*. Cambridge, Mass.: MIT Press, 1992.

CUBITT, Sean. Time to live. In: INTERNATIONAL SYMPOSIUM ON ELECTRONIC ART, 17, 2011, Istambul. *Proceedings...* Istambul: Leonardo Electronic Almanac, 2011. p. 8-15.

DANEY, Serge. A rampa. In: \_\_\_\_\_. *A rampa: Cahiers du cinéma, 1970-1982*. São Paulo: Cosac Naify, 2007. p. 229-234.

DANIELS, Dieter. Forerunners of media art in the first half of the twentieth century. *Media Art Net*. 2004. Disponível em: <[http://mediaartnet.org/themes/overview\\_of\\_media\\_art/forerunners/](http://mediaartnet.org/themes/overview_of_media_art/forerunners/)>. Acesso em: 19 nov. 2009.

DASTON, Lorraine; GALISON, Peter. *Objectivity*. New York: Zone Books, 2007.

DELEUZE, Gilles. Um novo cartógrafo (Vigiar e Punir). In: \_\_\_\_\_. *Foucault*. São Paulo: Brasiliense, 2005. p. 33-53.

\_\_\_\_\_. A propósito de Simondon. *Cadernos de subjetividade: O reencantamento do concreto*, São Paulo, v. 1, n. 1, 2003. p. 119-124.

\_\_\_\_\_. ¿Qué es un dispositivo?. In: BALBIER, Etienne; et al. *Michel Foucault, filósofo*. Barcelona: Gedisa, 1999. p. 155-163.

\_\_\_\_\_; GUATTARI, Félix. Introdução: rizoma. In: \_\_\_\_\_; GUATTARI, Félix. *Mil platôs: capitalismo e esquizofrenia*. Vol. 1. Rio de Janeiro: Ed. 34, 1995a. p. 11-37.

\_\_\_\_\_; GUATTARI, Félix. 20 de novembro de 1923 – Postulados da linguística. In: \_\_\_\_\_; GUATTARI, Félix. *Mil platôs: capitalismo e esquizofrenia*. Vol. 2. Rio de Janeiro: Ed. 34, 1995b. p. 9-59.

\_\_\_\_\_. *Conversações, 1972-1990*. Rio de Janeiro: Ed. 34, 1992.

DESCARTES, René. A dióptrica. Discursos I, II, III, IV e VIII. Trad. José Portugal dos Santos Ramos. *Scientiae studia*, São Paulo, v. 8, n. 3, 2010. p. 451-486. Disponível em: <<http://www.revistas.usp.br/ss/article/view/11212/12980>>. Acesso em 15 jan 2015.

DICK, Philip K. *Do androids dream of electric sheep?*. New York: Del Rey Books, 1996.

DUBOIS, Philippe. *O ato fotográfico*. 14ª ed. Campinas, SP: Papyrus, 2012.

\_\_\_\_\_. *Cinema, vídeo, Godard*. São Paulo: Cosac Naify, 2004.

DUGUET, Anne-Marie. Dispositivos. In: MACIEL, Kátia. *Transcineamas*. Rio de Janeiro: Contracapa, 2012. p. 49-70.

ELKINS, James. História da arte e imagens que não são arte. *Revista Porto Arte*, Porto Alegre, v. 18, n. 30. Maio/2011. p 7-42. Disponível em: <<http://seer.ufrgs.br/index.php/PortoArte/article/view/29619/18306>>. Acesso em: 5 fev 2015.

EYE/MACHINE. Direção: Harun Farocki. [S.l.]: Video Data Bank, 2001. Versão monocanal de videoinstalação originalmente apresentada em dois canais de vídeo (23 min). Son., color., 4x3.

EYE/MACHINE II. Direção: Harun Farocki. [S.l.]: Video Data Bank, 2002. Versão monocanal de videoinstalação originalmente apresentada em dois canais de vídeo (17 min). Son., color., 4x3.

EYE/MACHINE III. Direção: Harun Farocki. [S.l.]: Video Data Bank, 2003. Versão monocanal de videoinstalação originalmente apresentada em dois canais de vídeo (17 min). Son., color., 4x3.

FERRAZ, Maria Cristina Franco. Genealogia, comunicação e cultura somática. *Revista Famecos*, Porto Alegre, v. 20, n. 1, pp. 163-178, janeiro/abril 2013. Disponível em: <<http://revistaseletronicas.pucrs.br/ojs/index.php/revistafamecos/article/view/13640/9205>>. Acesso em: 5 fev 2015.

\_\_\_\_\_. Percepção, subjetividade e corpo: do século XIX ao XXI. In: PESSOA, Fernando (org.). *Arte no pensamento*. Vitória: Fundação Vale do Rio Doce, 2006. pp. 314-334.

FORSYTH, David A.; PONCE, Jean. *Computer vision: a modern approach*. New Jersey: Pearson, 2012.

FOSTER, Hal (ed.). *Vision and visuality*. Seattle: Bay Press, 1988.

FOUCAULT, Michel. *Vigiar e punir: o nascimento da prisão*. Petrópolis, RJ: Vozes, 1999.

\_\_\_\_\_. *Microfísica do poder*. Rio de Janeiro: Edições Graal, 1979.

GALISON, Peter. Image scatter into data, data gather into image. In: LATOUR, Bruno; WEIBEL, Peter (eds.). *Iconoclasm: beyond the image wars in science, religion, and art*. Karlsruhe; Cambridge: ZKM e MIT Press, 2002. p. 300-323.

GOVEIA, Fábio; MALINI, Fábio; e CIARELLI, Patrick. “Bots” contra a sociedade. *Blog do Labic*. 30 set 2014. Disponível em: <<http://www.labic.net/bots-contra-a-sociedade/>>. Acesso em 31 jan 2015.

GUATTARI, Félix. A paixão das máquinas. *Cadernos de subjetividade: O reencantamento do concreto*, São Paulo, v. 1, n. 1, 2003. p. 39-51.

GUIMARÃES, César; FRANÇA, Vera (orgs.). *Na mídia, na rua: narrativas do cotidiano*. Belo Horizonte: Autêntica, 2006.

GUIMARÃES, César; LEAL, Bruno Souza; MENDONÇA, Carlos Camargos (orgs.). *Comunicação e experiência estética*. Belo Horizonte: Ed. UFMG, 2006.

GUIMARÃES, César. O novo regime do visível e as imagens digitais. In: VAZ, Paulo Bernardo; CASA NOVA, Vera (orgs.). *Estação imagem: desafios*. Belo Horizonte: Editora UFMG, 2002. p. 147-161.

GUMBRECHT, Hans Ulrich. *The production of presence: what meaning cannot convey*. Kindle Ebook. Stanford, CA: Stanford University Press, 2004.

\_\_\_\_\_. *Modernização dos sentidos*. São Paulo: Ed. 34, 1998.

GUNNING, Tom. O retrato do corpo humano: a fotografia, os detetives e os primórdios do cinema. In: CHARNEY, Leo; SCHWARTZ, Vanessa. *O cinema e a invenção da vida moderna*. São Paulo: Cosac & Naify, 2004. p. 33-65.

HANSEN, Mark B. N. Seeing with the body: the digital image in postphotography. *Diacritics*, v. 31, n. 4, inverno 2001. p. 54-82.

HANSEN, Miriam. Benjamin, cinema e experiência: a flor azul na terra da tecnologia. In: BENJAMIN, Walter [et al.]. *Benjamin e a obra de arte: técnica, imagem, percepção*. Rio de Janeiro: Contraponto, 2012. p. 205-255.

HARAWAY, Donna J. Manifesto ciborgue: ciência tecnologia e feminismo-socialista no final do século XX. In: TADEU, Tomaz (org.). *Antropologia do ciborgue: as vertigens do pós-humano*. Belo Horizonte: Autêntica, 2013. p. 33-118.

HEMPEL, Jessi. Project HoloLens: our exclusive hands-on with Microsoft's holographic goggles. *Wired*, 21 jan 2015. Disponível em: <<http://www.wired.com/2015/01/microsoft-hands-on/>>. Acesso em: 31 jan 2015.

HENNION, Antoine; LATOUR, Bruno. How to make mistakes on so many things at once – and become famous for it. In: MARRINAN, Michael; GUMBRECHT, Hans Ulrich (ed.). *Mapping Benjamin: the work of art in the digital age*. Stanford, CA: Stanford University Press, 2003. p. 91-97.

HENRY, Karen. Collaborating with computers. In: ROKEBY, David. *David Rokeby*. Hamilton, ON: Art Gallery of Hamilton, 2004. p. 36-47.

HER. Direção: Spike Jonze. 2013. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (126 min.), son., color., widescreen.

HOW not to be seen: a fucking didactic educational .mov file. Direção: Hito Steyerl. 2013. Vídeo monocal (14 min). Son., color., 16x9, HD. Consultado em vídeo de registro de sua exposição na 55a Bienal de Veneza por Nikolai Blau. Disponível em: <<http://vimeo.com/70446674>>. Acesso em 11 out 2014.



HUHTAMO, Erkki. Introduction: moving panorama – a missing medium. In: \_\_\_\_\_. *Illusions in motion: media archaeology of the moving panorama and related spectacles*. Cambridge, MA: MIT Press, 2013. p. 1-26.

INGOLD, Tim. Trazendo as coisas de volta à vida: emaranhados criativos num mundo de materiais. *Horizontes Antropológicos*, Porto Alegre, ano 18, n. 37, p. 25-44, jan/jun. 2012. Disponível em: <<http://dx.doi.org/10.1590/S0104-71832012000100002>>. Acesso em: 5 fev 2015.

\_\_\_\_\_. When ANT meets SPIDER: social theory for arthropods. In: KNAPETT, Carl; MALAFOURIS, Lambros (eds.). *Material agency: towards a non-anthropocentric approach*. New York: Springer, 2008. p. 209-215.

IVINS, William M. *On the rationalization of sight: with the examination of three Renaissance texts on perspective*. New York: Da Capo Press, 1975.

JAY, Martin. Scopic regimes of modernity. In: FOSTER, Hal (ed.). *Vision and visuality*. Seattle: Bay Press, 1988. p. 3-27.

JOHNSON, Steven. *Emergence*. Kindle Ebook. New York: Scribner, 2004.

JOHNSTON, John. Machinic vision. *Critical inquiry*, v. 26, n. 1, Chicago, outono 1999. p. 27-48.

KARPATY, Andrej. What I learned from competing against a ConvNet on ImageNet. *Andrej Karpathy blog*. 2 set 2014. Disponível em: <<http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/>>. Acesso em: 24 jan 2015.

\_\_\_\_\_. The state of Computer Vision and AI: we are really, really far. *Andrej Karpathy blog*. 22 out 2012. Disponível em: <<http://karpathy.github.io/2012/10/22/state-of-computer-vision/>>. Acesso em: 24 jan 2015.

\_\_\_\_\_. t-SNE visualization of CNN codes. [2012?]. Disponível em: <<http://cs.stanford.edu/people/karpathy/cnnembed/>>. Acesso em: 24 jan 2015.

KELBERMAN, Dina. Re: Research – “I’m Google” [mensagem pessoal]. Mensagem recebida por <andregmintz@gmail.com> em 16 set 2013a.

\_\_\_\_\_. Re: Research – “I’m Google” [mensagem pessoal]. Mensagem recebida por <andregmintz@gmail.com> em 6 set 2013b.

\_\_\_\_\_. *I’m Google*. 201-. Obra de arte em formato de *blog* apresentando coleção de imagens periodicamente atualizada pela artista, desde 2011. Como a última atualização não é indicada, assumimos como incerta a data da versão consultada. Disponível em: <<http://dinakelberman.tumblr.com/>>. Último acesso: 29 jan 2015.

KESSLER, Frank. Notes on dispositif. Nov 2007 [versão de texto em processo]. Disponível em: <<http://www.frankkessler.nl/wp-content/uploads/2010/05/Dispositif-Notes.pdf>>. Acesso em 5 mar 2014.

KLEIN, Otávio José. A gênese do conceito de dispositivo e sua utilização nos estudos midiáticos. *Estudos em comunicação*, n. 1, abr 2007. p. 215-231.

KLÜVER, Billy. *Um dia com Picasso: 29 fotografias de Jean Cocteau*. Rio de Janeiro: José Olympio, 2003.

KRACAUER, Siegfried. Cult of distraction: on Berlin's picture palaces. *New German Critique*, n. 40, Special issue on Weimar Film Theory, Winter, 1987. p. 91-96. Disponível em: <<http://www.jstor.org/stable/488133>>. Acesso em: 5 fev 2015.

KRUEGER, Myron W. Responsive environments. In: WARDRIP-FRUIN, Noah; MONTFORT, Nick (eds.). *The new media reader*. Cambridge, MA: MIT Press, 2003. p. 379-389.

LA JETÉE. Dirigido por Chris Marker. Lançado em 1962. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (28 min.). son., p&b.

LATOURE, Bruno. Alguns experimentos em arte e política. Trad. Eduardo de Jesus. *Dispositiva*, Belo Horizonte, v. 1, n. 1, maio/out 2012. p. 18-28. Disponível em: <<http://periodicos.pucminas.br/index.php/dispositiva/article/view/2510/3104>>. Acesso em: 5 fev 2015.

\_\_\_\_\_. An attempt at a "Compositionist Manifesto". *New Literary History*, v. 41, 2010. p. 471-490. Disponível em: <<http://www.bruno-latour.fr/sites/default/files/120-NLH-finalpdf.pdf>>. Acesso em: 5 fev 2015.

\_\_\_\_\_. O que é iconoclash? Ou, há um mundo além das guerras de imagem? *Horizontes Antropológicos*. Porto Alegre, ano 14, n. 29, p. 111-150, jan-jun 2008. Disponível em: <<http://dx.doi.org/10.1590/S0104-71832008000100006>>. Acesso em 5 fev 2015.

\_\_\_\_\_. *Reassembling the social*. Oxford: Oxford University Press, 2005.

\_\_\_\_\_. What is iconoclash? Or is there a world beyond the image wars? In: \_\_\_\_\_; WEIBEL, Peter (eds.). *Iconoclash: beyond the image wars in science, religion, and art*. Karlsruhe; Cambridge: ZKM e MIT Press, 2002. p. 14-37.

\_\_\_\_\_. Um coletivo de humanos e não-humanos: no labirinto de dédalo. In: \_\_\_\_\_. *A esperança de Pandora: ensaios sobre a realidade dos estudos científicos*. Bauru, SP: EDUSC, 2001. p. 201-246.

\_\_\_\_\_. *Jamais fomos modernos*. Rio de Janeiro: Ed. 34, 1994.

\_\_\_\_\_. Visualization and cognition: drawing things together. *Knowledge and Society* 6, 1986. p. 1-40.

LATOURE, Bruno; WEIBEL, Peter (eds). *Iconoclasm: beyond the image wars in science, religion, and art*. Karlsruhe; Cambridge: ZMK e MIT Press, 2002.

LE, Quoc V. *et al.* Building high-level features using large scale unsupervised learning. *arXiv:1112.6209 [cs]*, arXiv: 1112.6209, 28 dez. 2011. Disponível em: <<http://arxiv.org/abs/1112.6209>>. Acesso em: 9 jan. 2015.

LEAL, Bruno Souza; GUIMARÃES, César; MENDONÇA, Carlos (orgs.). *Entre o sensível e o comunicacional*. Belo Horizonte: Autêntica, 2010.

LEVIN, Golan. Computer vision for artists and designers: pedagogic tools and techniques for novice programmers. *AI & Society*, v. 20, 2006a. p. 462-482.

\_\_\_\_\_. Hands up! A visual essay on the 'media art pose'. Texto originalmente apresentado no simpósio Emoção Art.ficial. São Paulo, Brasil; 24 jul 2006b. Disponível em: <[http://www.flong.com/texts/essays/essay\\_pose/](http://www.flong.com/texts/essays/essay_pose/)>. Acesso em 10 maio 2014.

LEVIN, Thomas Y.; FROHNE, Ursula; WEIBEL, Peter (eds.). *Ctrl [space]: rhetorics of surveillance from Bentham to Big Brother*. Karlsruhe; Cambridge: ZKM e MIT Press, 2002.

LÉVY, Pierre. *As tecnologias da inteligência : o futuro do pensamento na era da informática*. Rio de Janeiro: Ed. 34, 1993.

MACHADO, Arlindo. Máquina e imaginário. In: DOMINGUES, Diana (org.). *Arte, ciência e tecnologia: passado, presente e desafios*. São Paulo: Ed. UNESP, 2009. p. 179-199.

\_\_\_\_\_. *Arte e mídia*. Rio de Janeiro: Jorge Zahar, 2007.

MAGNET, Shoshana Amielle. *When biometrics fail: gender, race and the technology of identity. Kindle digital edition*. Durnham; London: Duke University Press, 2011.

MAN with a movie camera. Direção: Dziga Vertov. Lançado em 1929. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (68 min.), son., p&b. Tradução de *Chelovek s kinoapparatom*.

MANOVICH, Lev; DOUGLAS, Jeremy; ZEPPEL, Tara. How to compare one million images? 2011. Disponível em: <[http://manovich.net/content/04-projects/069-how-to-compare/68\\_article\\_2011\\_sm.pdf](http://manovich.net/content/04-projects/069-how-to-compare/68_article_2011_sm.pdf)>. Acesso em: 16 nov 2014.

\_\_\_\_\_. Modern surveillance machines: perspective, radar, 3-D computer graphics and computer vision. In: LEVIN, Thomas Y.; FROHNE, Ursula; WEIBEL, Peter (eds.). *Ctrl [space]: rhetorics of surveillance from Bentham to Big Brother*. Karlsruhe; Cambridge: ZKM e MIT Press, 2002. p. 383-395.

\_\_\_\_\_. *The language of new media*. Cambridge, MA: MIT Press, 2000.

\_\_\_\_\_. Automation of sight from photography to computer vision. 1997. Disponível em: <<http://manovich.net/DOCS/automation.doc>>. Acesso em 8 set 2013.

\_\_\_\_\_. *The engineering of vision from constructivism to computers*. Tese (doutorado). University of Rochester. 1993a. Disponível em: <<http://manovich.net/EV/EV.PDF>>. Acesso em 12 set 2013.

\_\_\_\_\_. Mapping space: perspective, radar and computer graphics. *Siggraph '93 visual proceedings*. New York: ACM, 1993b. p. 143-147. Disponível em: <<http://manovich.net/TEXT/mapping.html>>. Acesso em: 25 abr 2014.

MARKOFF, John. How many computers to identify a cat? 16,000. *The New York Times*. 25 jun 2012. Disponível em: <[http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html?\\_r=0](http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html?_r=0)>. Acesso em: 10 maio 2014.

MARR, David. *Vision: a computational investigation into the human representation and processing of visual information*. New York: W. H. Freeman, 1982.

MAZZINI, Leandro. Exército remodela centro de inteligência para monitorar movimentos sociais. *Blog Coluna Esplanada*. 29 jul 2014. Disponível em: <<http://colunaesplanada.blogosfera.uol.com.br/2014/07/29/exercito-remodela-centro-de-inteligencia-para-monitorar-movimentos-sociais/>>. Acesso em: 16 nov 2014.

MCMILLAN, Robert. This guy beat Google's super-smart AI – but it wasn't easy. *Wired*. 21 jan 2015. Disponível em: <<http://www.wired.com/2015/01/karpathy/>>. Acesso em 24 jan 2015.

MIGLIORIN, Cezar. O dispositivo como estratégia narrativa. *Digitagrama - Revista Acadêmica de Cinema*, 1º sem, 2005. Disponível em: <<http://www.estacio.br/graduacao/cinema/digitagrama/numero3/cmigliorin.asp>>. Acesso em 25 out 2007.

MITCHELL, William J. T. *Picture theory: essays on verbal and visual representation*. Chicago, IL: University of Chicago Press, 1994.

MONTEIRO, Tânia. Exército brasileiro cria órgão para monitorar manifestações. *O Estado de São Paulo*. 31 jul 2014. Disponível em: <<http://brasil.estadao.com.br/noticias/geral,exercito-brasileiro-cria-orgao-para-monitorar-manifestacoes,1536422>>. Acesso em 16 nov 2014.

MOUILLAUD, Maurice. Da forma ao sentido. In: \_\_\_\_; PORTO, Sérgio Dayrell (org.). *O jornal: da forma ao sentido*. Brasília: Paralelo 15, 1997. p. 29-35.

NASELARIS, Thomas; PRENGER, Ryan J.; KAY, Kendrick N.; OLIVER, Michael; GALLANT, Jack L. Bayesian reconstruction of natural images from human brain activity. *Neuron* 63, 24 set 2009. p. 902-915. Disponível em: <<http://dx.doi.org/10.1016/j.neuron.2009.09.006>>. Acesso em: 16 jan 2015.

NISHIMOTO, Shinji; VU, An T.; NASELARIS, Thomas; BENJAMINI, Yuval; YU, Bin; GALLANT, Jack L. Reconstructing visual experiences from brain activity evoked by natural movies. *Current biology* 21, 11 out 2011. p. 1641-1646. Disponível em: <<http://dx.doi.org/10.1016/j.cub.2011.08.031>>. Acesso em: 16 jan 2015.

NOBLE, Joshua. *Programming interactivity: a designer's guide to Processing, Arduino and openFrameworks*. Sebastopol: O'Reilly, 2009.

NÖTH, Winfried. Máquinas semióticas. *Galáxia*, São Paulo, n. 1, 2001. Disponível em: <<http://revistas.pucsp.br/index.php/galaxia/article/view/1057/694>>. Acesso em 7 out 2014.

ORLANDI, Luiz B. L. O indivíduo e sua ímplexa pré-individualidade. *Cadernos de subjetividade*. O reencantamento do concreto, v. 1, n. 1, 2003. p. 87-96.

PANOFSKY, Erwin. *A perspectiva como forma simbólica*. Lisboa: Edições 70, 1993.

PAPERT, Seymour. *Mindstorms : children, computers, and powerful ideas*. New York: Basic Books, 1993.

PAUL, Christiane. *Digital art*. London: Thames & Hudson, 2008.

PINTO, Julio. *1, 2, 3 da semiótica*. Belo Horizonte: Editora UFMG, 1995.

PROUST, Marcel. *No caminho de Swann*. Tradução: Mario Quintana. São Paulo: Globo, 2006. (Em busca do tempo perdido; v. 1).

RENAUD-ALAIN, Alain. L'image sans gravité: la forme image aux risques de l'information. *La revue d'esthétique*, n. 25, 1994. p. 11-22.

RISEN, James; POITRAS, Laura. N.S.A. collecting millions of faces from web images. In: *The New York Times*, 31 maio 2014. Disponível em: <<http://www.nytimes.com/2014/06/01/us/nsa-collecting-millions-of-faces-from-web-images.html>>. Acesso em 7 jul 2014.

ROBERTS, Lawrence G. Machine perception of three-dimensional solids. MIT Lincoln Laboratory, 1963. Disponível em: <<http://www.packet.cc/files/mach-per-3D-solids.html>>. Acesso em 8 jul 2014.

ROBOCOP: o policial do futuro. Direção: Paul Verhoeven. Lançado em 1987. [S.l.]: Netflix, 2014. *Streaming*. (104 min.), son., color., legendado. Tradução de: Robocop.

RODRÍGUEZ, Pablo. Prólogo. El modo de existencia de una filosofía nueva. In: SIMONDON, Gilbert. *El modo de existencia de los objetos técnicos*. Buenos Aires: Prometeo Libros, 2007. p. 9-24.

ROKEBY, David. The giver of names (1990-) (in progress). 2010. Disponível em: <<http://www.davidrokeby.com/gon.html>>. Acesso em 15 nov 2014.

ROSEN, Philip. History of image, image of history: subject and ontology in Bazin. In: MARGUILES, Ivone (org.). *Rites of realism: essays on corporeal cinema*. Londres: Duke University Press, 2003. p. 42-79.

RUSH, Michael. *Novas mídias na arte contemporânea*. São Paulo: Martins Fontes, 2006.

SANTELLA, Lucia; NÖTH, Winfried. Os três paradigmas da imagem. In: \_\_\_\_\_. *Imagem: cognição, semiótica, mídia*. 3ª edição. São Paulo: Iluminuras, 2001. p. 159-186.

SANTAELLA, Lúcia. Do signo. In: \_\_\_\_\_. *A teoria geral dos signos: como as linguagens significam as coisas*. São Paulo: Pioneira, 2000. p. 11-31.

\_\_\_\_\_. O homem e as máquinas. In: DOMINGUES, Diana. *A arte no século XXI: a humanização das tecnologias*. São Paulo: Fundação Editora da UNESP, 1997.

SCHECHNER, Richard. O que é performance?. *O percevejo*, v. 11, n. 12. 2003. p. 25-50.

SCHWARTZ, Vanessa R. O espectador cinematográfico antes do aparato do cinema: o gosto do público pela realidade na Paris fim-de-século. In: CHARNEY, Leo; SCHWARTZ, Vanessa. *O cinema e a invenção da vida moderna*. São Paulo: Cosac & Naify, 2004. p. 337-360.

SENRA, Stella. Cray e as transformações do observador. In: CRARY, Jonathan. *Suspensões da percepção: atenção, espetáculo e cultura moderna*. São Paulo: Cosac Naify, 2013. p. 9-19.

SEKULA, Allan. The body and the archive. *October*, v. 39, inverno 1986, p. 3-64. Disponível em: <<http://www.jstor.org/stable/778312>>. Acesso em 10 jan 2015.

SHET, Vinay. Are you a robot? Introducing “No CAPTCHA reCAPTCHA”. *Google online security blog*, 3 dez 2014. Disponível em: <<http://googleonlinesecurity.blogspot.com.br/2014/12/are-you-robot-introducing-no-captcha.html>>. Acesso em: 27 jan 2015.

\_\_\_\_\_. reCAPTCHA just got easier (but only if you are human). *Google online security blog*, 25 out 2013. Disponível em: <<http://googleonlinesecurity.blogspot.com.br/2013/10/recaptcha-just-got-easier-but-only-if.html>>. Acesso em: 27 jan 2015.

SIBILIA, Paula. *O homem pós-orgânico: corpo, subjetividade e tecnologias digitais*. Rio de Janeiro: Relume Dumará, 2002.

SIMMEL, Georg. The metropolis and the mental life. In: WOLFF, Kurt H. (ed.). *The sociology of Georg Simmel*. Glencoe, Illinois: The Free Press, 1950.

SIMONDON, Gilbert. A gênese do indivíduo. *Cadernos de subjetividade*. O reencantamento do concreto. Núcleo de estudos e pesquisas da subjetividade do programa de estudos pós-graduados em psicologia clínica da PUC-SP, v. 1, n. 1, 2003. p. 97-117.

SIMONITE, Tom. Facebook creates software that matches faces almost as well as you do. MIT Technology Review. 14 mar 2014. Disponível em: <<http://www.technologyreview.com/news/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do/>> Acesso em 10 maio 2014.

SINGER, Ben. Modernidade, hiperestímulo e o início do sensacionalismo popular. In: CHARNEY, Leo; SCHWARTZ, Vanessa (org.). *O cinema e a invenção da vida moderna*. São Paulo: Cosac & Naify, 2004. p. 95-123.

SONTAG, Susan. O heroísmo da visão. In: \_\_\_\_\_. *Sobre fotografia*. São Paulo: Companhia das Letras, 1983. p. 99-128.

STERLING, Bruce. An essay on the New Aesthetic. 2012. Disponível em: <[http://www.wired.com/beyond\\_the\\_beyond/2012/04/an-essay-on-the-new-aesthetic/](http://www.wired.com/beyond_the_beyond/2012/04/an-essay-on-the-new-aesthetic/)>. Acesso em: 27 mar 2013.

STRATHERN, Marilyn. Cutting the network. *The journal of the royal anthropological institute*, v. 2, n. 3, set 1996. p. 517-535.

SUCHMAN, Lucy. *Human-machine reconfigurations: plans and situated actions*. 2nd edition. Kindle Ebook. New York: Cambridge University Press, 2007.

SZELISKI, Richard. *Computer vision: algorithms and applications*. September 3, 2010 Draft. Disponível em: <[http://szeliski.org/Book/drafts/SzeliskiBook\\_20100903\\_draft.pdf](http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf)>. Acesso em: 10 maio 2014.

TADEU, Tomaz (org.). *Antropologia do ciborgue: as vertigens do pós-humano*. Belo Horizonte: Autêntica, 2013.

TAIGMAN, Yaniv *et al.* Deepface: closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013. p. 1701–1708. Disponível em: <[http://www.cv-foundation.org/openaccess/content\\_cvpr\\_2014/html/Taigman\\_DeepFace\\_Closing\\_the\\_2014\\_CVPR\\_paper.html](http://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Taigman_DeepFace_Closing_the_2014_CVPR_paper.html)>. Acesso em: 27 set. 2014.

TAYLOR, Diana. Hacia una definición de performance, *O perceivejo*, v. 11, n. 12, 2003. p. 17-24.

TERMINATOR. Direção: James Cameron. Lançado em 1984. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (117 min.), son., color., widescreen.

THIRION, Bertrand; DUCHESNAY, Edouard; HUBBARD, Edward; DUBOIS, Jessica; POLINE, Jean-Baptiste; LEBIHAN, Denis; DEHAENE, Stanislas. Inverse retinotopy: inferring the visual content of images from brain activation patterns. *Neuroimage* 33, 2006. p. 1104-1116. Disponível em: <<http://dx.doi.org/10.1016/j.neuroimage.2006.06.062>>. Acesso em 15 jan 2015.

TUER, Dot. Art in the age of intelligent machines. In: ROKEBY, David. *David Rokeby*. Hamilton, ON: Art Gallery of Hamilton, 2004. p. 23-33.

TURING, Alan. Computing machinery and intelligence. In: WARDRIP-FRUIN, Noah; MONTFORT, Nick (eds.). *The new media reader*. Cambridge; London: MIT Press, 2003. p. 50-64.

UN CHIEN Andalou. Direção: Luis Buñuel. Lançado em 1929. Acessado por arquivo compartilhado na internet via rede P2P (*peer to peer*), sem distribuidora oficial, 2014. 1 arquivo MP4 (16 min.), son., p&b.

VERTOV, Dziga. Resolução do conselho dos três em 10-4-1923. In: XAVIER, Ismail. *A experiência do cinema: antologia*. Rio de Janeiro: Edições Graal; Embrafilme, 1983a. p. 252-259.

\_\_\_\_\_. Nascimento do cine-olho. In: XAVIER, Ismail. *A experiência do cinema: antologia*. Rio de Janeiro: Edições Graal; Embrafilme, 1983b. p. 260-262.

VIOLAÇÃO de privacidade. Direção: Omar Naim. Lançado em 2004. [S.l.]: Netflix, 2014. *Streaming*. (94 min.), son., color., legendado. Tradução de: *The final cut*.

VIRILIO, Paul. Da política do pior ao melhor das utopias e a globalização do terror. Entrevista. *Revista FAMECOS*, Porto Alegre, n. 16, dez 2001. Disponível em: <<http://www.revistas.univerciencia.org/index.php/famecos/article/view/264/198>>. Acesso em: 5 fev 2015.

\_\_\_\_\_. *A máquina de visão*. Rio de Janeiro: José Olympio, 1994.

\_\_\_\_\_. A imagem virtual mental e instrumental. In: PARENTE, André (org). *Imagem-máquina: a era das tecnologias do virtual*. Rio de Janeiro: Ed. 34, 1993. p. 127-132.

WEES, William C. The camera-eye: dialectics of a metaphor. In: SHAW, Jeffrey; WEIBEL, Peter. *Future cinema: the cinematic imaginary after film*. Karlsruhe; Cambridge: ZKM Center for Art and Media e MIT Press, 2003. p. 48-55.

WOOD, Lamont. AI gets its groove back. *Computerworld*. 14 abr 2014. Disponível em: <[http://www.computerworld.com/s/article/9247432/AI\\_gets\\_its\\_groove\\_back?source=CTWNLE\\_nlt\\_dailyam\\_2014-04-14](http://www.computerworld.com/s/article/9247432/AI_gets_its_groove_back?source=CTWNLE_nlt_dailyam_2014-04-14)>. Acesso em: 10 maio 2014.

YUILLE, Alan; OLIVA, Aude. *Frontiers in computer vision: NSF White Paper*. 2010. Disponível em: <<http://www.frontiersincomputervision.com/WhitePaperInvite.pdf>>. Acesso em: 09 set 2012.



## Anexo I – Listagem ilustrativa de operações características da Visão Computacional

Com o intuito de oferecer um panorama das operações pelas quais se compreende a Visão Computacional, listamos algumas delas a seguir, com base em algumas referências técnicas e didáticas (LEVIN, G., 2006a; FORSYTH e PONCE, 2012; BAGGIO, 2012; NOBLE, 2009, cap. 14). Cabe ressaltar que não nos propomos uma enumeração exaustiva nem uma taxonomia precisa dos algoritmos e métodos pertinentes<sup>151</sup>, mas apenas uma listagem ilustrativa, a fim de situar o leitor que talvez desconheça a composição do campo abordado, assim oferecendo subsídios para a discussão realizada.

**Subtração de fundo e detecção de presença:** a subtração de fundo (*background subtraction*, cf. LEVIN, G., 2006a, p. 469) compreende um de diferentes métodos para a segmentação da imagem por agrupamento de seus *pixels* (*segmentation by clustering*, cf. FORSYTH e PONCE, 2012, cap. 9). Em seu modo elementar, envolve, primeiramente, uma sequência de operações aritméticas simples em que os valores de cor dos *pixels* de uma imagem previamente tomada, retratando apenas o fundo de determinada cena (sem qualquer outro objeto) são subtraídos das imagens analisadas a fim identificar as porções da imagem que seriam diferentes do fundo e que, portanto, indicariam a entrada de novos objetos ou personagens à cena. Pela necessidade de coincidência entre o ponto de vista da imagem do fundo tomada como referência e das imagens das quais tal fundo é subtraído, trata-se em geral de operação realizada sobre imagens de vídeo tomadas a partir de uma câmera fixa. Um uso comum deste método é a detecção de presença em aplicações de videovigilância, em que o programa dispararia uma ação quando seriam detectadas variações significativas ou, por exemplo, quando a presença é detectada em regiões específicas da imagem, chamadas de *regiões de interesse* (*region of interest*). A partir da subtração de fundo torna-se igualmente possível a discretização dos objetos retratados, podendo ainda traçar seus contornos ou rastrear seu movimento. Além da subtração de fundo, outros métodos de segmentação por agrupamento envolvem a segmentação cromática – como a realizada para o efeito de *chroma key*, com recurso aos fundos verdes e azuis – e a segmentação por luminância, em que *pixels* seriam agrupados com relação à sua inserção em determinadas faixas de brilho.

**Detecção de contornos e arestas:** por contornos (*contours* ou *occluding contours*, cf. FORSYTH e PONCE, 2012, cap. 5), nos referimos aos limites que separam, na imagem, um objeto ou corpo do fundo, que, se traçados, desenhariam suas silhuetas. Por arestas

---

<sup>151</sup> Algumas operações, nomeio a partir da tradução de sua denominação técnica; outras nomeio a partir da derivação de agrupamentos livres com o objetivo de indicar frentes de operação de diferentes algoritmos.

(*edges*, cf. FORSYTH e PONCE, 2012, cap. 5), nos referimos às linhas que se inscreveriam no interior destas silhuetas, mas que indicariam dobras de sua volumetria, no mesmo sentido que nomeamos as arestas de um poliedro. Embora sejam, em ambos os casos, linhas, os métodos de detecção podem ser distintos, bem como suas aplicações. Um modo de detectar os contornos é em etapa posterior à separação do fundo, como no método descrito anteriormente, permitindo a conversão da imagem em uma matriz binária em que diferenciaríamos, claramente, as formas do fundo. A partir desta representação, as formas encontradas em diferença ao fundo são circunscritas por um polígono composto de vários pontos dispostos ao longo desse contorno. As arestas, por sua vez, são geralmente identificadas a partir do processamento da imagem por filtros que buscam encontrar variações mais ou menos bruscas de luminância, indicando possíveis limites ou dobras nos próprios objetos. Tanto os contornos quanto as arestas podem subsidiar operações de reconhecimento de objetos e também de padrões bidimensionais, ao descrever o conjunto da imagem em dados geométricos mais simples. As silhuetas circunscritas pelos contornos permitem ainda modalidades de interação entre um corpo tomado desta forma e elementos gráficos sobrepostos à imagem, os quais podem ser programados para simular, em seu movimento, choques ou mudanças de direção, fazendo com que o corpo tomado na imagem passe a integrar, em alguma medida, o ambiente simulado.

**Reconhecimento de padrões bidimensionais:** códigos de barra, caracteres e os chamados códigos QR (*Quick Response Codes*) são os mais comuns padrões visuais bidimensionais reconhecidos por algoritmos de Visão Computacional. Uma vez que partem, de início, de um alto contraste, com formas em preto e branco, seu reconhecimento é geralmente elaborado sobre os métodos já indicados de segmentação por agrupamento e detecção de contornos, a partir dos quais as linhas obtidas são comparadas a padrões conhecidos – no caso dos caracteres – ou mensuradas a fim de serem convertidas em informação alfanumérica. Tendo em vista os vários possíveis posicionamentos destes padrões em relação ao ponto de vista de tomada da imagem, bem como eventuais distorções decorrentes da distância focal das lentes utilizadas, são ainda empregados métodos de correção de perspectiva, uma vez identificados marcadores que permitam esta compensação, tais como os marcadores de registro dos *códigos QR* (cf. BAGGIO, 2012, cap. 2).

**Diferenciação de frames e análise do movimento:** com imagens de vídeo, a partir da diferenciação de seus *frames* consecutivos (*frame differencing*, cf. LEVIN, G., 2006a, p. 469) – através da subtração dos valores dos *pixels* de um *frame* pelo outro, por exemplo – podem ser realizadas não apenas mensurações da quantidade de movimento, mas também a análise de sua direção, permitindo aplicações como a estabilização digital da imagem, compensando o movimento da câmera. Pode-se, da mesma forma, valer-se desta comparação para métodos de segmentação da imagem, distinguindo, por exemplo, porções estáticas das dinâmicas.

**Rastreamento:** em linhas gerais, envolve o monitoramento da posição relativa de uma determinada porção da imagem já segmentada e discretizada do plano. Para isto, é necessário que existam aspectos característicos deste segmento (em geral um objeto reconhecido ou um detalhe) que possam ser reiteradamente identificados a cada *frame* consecutivo, permitindo, por exemplo, analisar seus padrões de movimento ou sua velocidade. Trata-se de operação fundamental para diversas aplicações, em particular no contexto bélico ou da vigilância. Forsyth e Ponce (2012, p. 326) indicam, na introdução de capítulo dedicado ao rastreamento, campos de aplicação que evidenciam tal relevância, sugerindo: a possibilidade de que os padrões de movimento sejam utilizados na identificação de determinados corpos ou objetos, considerando que determinados corpos teriam tipos de movimento bastante característicos; a detecção de movimentos anômalos ou irregulares de determinados objetos, como diferentes tipos de veículos que transitam no pátio de um aeroporto; e, ainda, o rastreamento aplicado à mira (*targetting*) de uma arma em um alvo em movimento, caso no qual o rastreamento serviria não apenas para descrever o movimento já realizado quanto para prever a posição futura deste alvo.

**Reconstrução da cena:** em um nível ideal, trata-se da recomposição integral, por computação gráfica tridimensional, da cena retratada na imagem, com a geração de modelos para os objetos, sobre os quais se aplicariam texturas extraídas da imagem fotográfica, permitindo inclusive a geração de imagens de outros pontos de vista da cena por simulação. Tal reconstrução pode ser realizada por diferentes estratégias. Uma delas, mais limitada, envolve a extração de dados de distância (*range data*) da imagem, o que geralmente se faz por dispositivos dedicados como o sensor Microsoft Kinect que projeta sobre a cena, por meio de um canhão de *laser* infravermelho (portanto invisível ao olho humano), um padrão de pontos que são analisados pela imagem captada por uma das câmeras que compõem o sensor, estabelecendo as distâncias entre o sensor e cada parte da imagem tomando como referência as deformações sofridas pelo padrão de pontos ao incidir sobre as superfícies dos corpos e objetos. Outro modo envolve a composição da cena a partir de dois (estereoscopia) ou múltiplos pontos de vista, com mais de uma imagem. A volumetria da cena é recomposta, nestes casos, a partir dos efeitos de paralaxe resultantes da relação entre as imagens produzidas por estes diferentes ângulos. Há ainda processos similares a partir da projeção de sombras na cena. Métodos como este foram utilizados pela Google, recentemente, na produção de imagens tridimensionais, com volumetria de edifícios e construções, para a navegação no Google Earth. Há ainda aplicações no contexto forense ou na arquitetura e, mais recentemente, em modos alternativos de geração de modelos tridimensionais para uso em processos de fabricação digital, como a chamada *impressão 3D*.

**Reconhecimento de objetos e classificação de imagens:** trata-se de operação que pode ser realizada por muitos métodos diferentes, mas que tem como objetivo principal o

reconhecimento na imagem de determinado corpo ou objeto. Relacionando-se à operação descrita anteriormente, de reconstrução da cena tridimensional, pode envolver a verificação da adequação de determinada forma encontrada na imagem a algum modelo tridimensional conhecido, desta forma permitindo a geração de outros pontos de vista deste mesmo objeto. De modo similar, permitiria ainda estimar a pose de um corpo humano, por exemplo, como fazem aplicações de Visão Computacional para sistemas interativos, como as relacionadas ao Microsoft Kinect. De outra forma, contudo, geralmente refere-se por reconhecimento de objetos operação realizada a partir de processos de aprendizado de máquinas (*machine learning*) pelos quais algoritmos são treinados para o reconhecimento de dos padrões visuais pelos quais determinado objeto geralmente se apresenta, tomando, como referência, uma base de dados construída a partir de exemplos destes objetos (cf. FORSYTH e PONCE, 2012, cap. 17 e 18). Algoritmos de detecção e reconhecimento de rostos, por exemplo, são geralmente elaborados por tais estratégias. Partem deste mesmo princípio operações de classificação de imagens (cf. FORSYTH e PONCE, 2012, cap. 15 e 16), pelas quais a análise da imagem como um todo resulta na indicação de alguns aspectos gerais ou conceitos que poderiam ser a ela atribuídos, como o provável local ou o tipo de situação retratada.

## Anexo II – Entrevista com a artista Dina Kelberman

Apresentamos abaixo, já em forma editada, as perguntas e as respostas de breve consulta feita por *email* à artista Dina Kelberman em duas ocasiões (6 set 2013 e 16 set 2013), nas quais buscamos compreender mais do processo envolvido na realização da obra *I'm Google*. Foram excluídos textos introdutórios dos *emails* e intercaladas as perguntas e respostas na edição – mesclando partes do *email* enviado e partes do *email* recebido, que compreendem momentos distintos no tempo. Mantivemos a distinção entre as duas ocasiões em que foram realizadas as consultas, bem como o idioma original da correspondência, em inglês.

### 6 set 2013

**André Góes Mintz** – I have read some articles published about *I'm Google*, and also read your description on your website, but, from those, your creative process is not completely clear to me. I understand the near obsessive collecting of images, but not so much how you search for them, organize and make associations. So, my first question – which will help me elaborate the others – is: when you say you use “Google Image Search”, do you search by keyword? Or do you use the “Search by image” tool as well?

**Dina Kelberman** – So, for *Google* searches I use both keywords and the “visually similar” tool, as well as visiting sites via the searches, and just keywords and linking through *YouTube* for videos.

However, a lot of people think that the visually similar tool probably does a lot of the work but it really provides surprisingly little help! I actually think I've used images found directly through visually similar less than 10 times on the whole blog. The algorithm seems to relate images based on like, general color palette, which is not really what most of the blog is using as the linking relationship. When it *is* helpful it's usually just as an aid in figuring out what *type* of thing I could look for next, or, more frequently, I just find something else I think is cool looking that I search for and save to use as a later batch.

The real bulk of the searching and finding comes through me searching keywords and looking through every image available for those keywords, selecting the ones I think are beautiful, and then whittling those down to the best few, and putting them in order either to get to the next idea that I already have in mind, or with an open ending that I deal with however I can.

I often end up learning a little bit about each thing, basically just enough to figure out more specific keywords to find better images.

**A** – What would you say is the level of participation of the machine in the creation/discovering of associations between images?

**D** – So I'd say the level of the machine in making the connections is extremely minimal. Google doesn't really associate things the same way as my blog does, and also the particular things I choose to use as themes are all things I personally am interested in. So in that way it's not just a machine making whatever connection it can, it's me making the connections I choose. The images selected are also chosen based on my own personal inclinations. Someone else would make a different version of this blog.

**16 set 2013**

**André Góes Mintz** – I already imagined that the machine did very little of the work. In fact, I tried a reversed engineering approach, by feeding an image from your blog to similar image search to see if any of the surrounding images came up – of course it failed. I don't know if you would agree on (or like) my interpretation, but even though there is less participation of image recognition and comparison processes, I think – and, as it seems, other people too – that the way you select and organize the images resembles the way we would imagine a machine would do it. By projecting *Google* on your subjectivity in the title this is also suggested, I think. Is this something you have in mind in your process of selection and organization of the images? Do you try to mimic, in some level, a “machine perception” of these images?

**Dina Kelberman** – It's not something I think about consciously, no. I mean, it's certainly very systematic though, sometimes it feels a bit like factory work (very enjoyable factory work). Often my rule of thumb when scanning through loads and loads of images is to just sort of not pay too close attention and wait for a gut-reaction to an image to happen. The rule is sort of “if I don't gasp at it then I'm not allowed to post it”. This rule gets broken sometimes if I desperately am trying to make some connection that I'm having a hard time with, but in general it's how I operate. It's funny to think about that as a mechanical process, but I suppose it is a bit like an algorithm itself. So, no, I base the selection very much on my personal aesthetic reactions.

**A** – Do you have a precise process, or algorithmic procedure you develop in this collecting? Or is it something you don't think so much about?

**D** – Maybe I already answered this too, but to be more specific, the answer is “not really.” I basically just start with a keyword, look at every image from that search (less if it’s going very poorly), click on particular images I like, sometimes I’ll link to the page they’re from to try to learn more specifically what they are, if there are particular terms that describe them that I wouldn’t have known to search, etc.

Maybe some time I should record myself searching, that would make for an interesting / incredibly boring video!







