



Abstração, expropriação, antecipação: nota genealógica sobre as visões maquinicas da gestualidade¹

*Abstraction, expropriation,
anticipation: a genealogical
note on the machinic
visions of gesture*



Maurício Lissovsky²

Fernanda Bruno³

Icaro Ferraz Vidal Junior⁴

¹ Uma versão condensada desse texto foi publicada em 2018 na Revista Réseaux 2018/5 n° 211 (DOI: 10.3917/res.211.0105). A versão original, que ora publicamos no dossiê “Mauricio Lissovsky: as imagens nas experiências do contemporâneo”, conta com trechos inéditos, especialmente na introdução e nas considerações finais.

² Falecido em 2022, foi historiador, roteirista, pesquisador e professor do Programa de Pós-Graduação em Comunicação da Escola de Comunicação da Universidade Federal do Rio de Janeiro (ECO-UFRJ).

³ Professora Titular da Universidade Federal do Rio de Janeiro (UFRJ), Programa de Pós-Graduação em Comunicação e Cultura da UFRJ. E-mail: bruno.fernanda@gmail.com

⁴ Doutor em Comunicação pela Universidade Federal do Rio de Janeiro (UFRJ), Pesquisador colaborador do MediaLab, UFRJ. E-mail: vidal.icaro@gmail.com

Resumo: Este artigo propõe uma genealogia das visões maquínicas da gestualidade. O texto está estruturado em três partes. A primeira descreve como a reprodutibilidade técnica da imagem é posta a serviço da reduplicação do gesto, de sua abstração e de sua apropriação em diversos domínios. A segunda parte demonstra que a matriz do modelo técnico contemporâneo, voltada para a predição dos comportamentos, já estava parcialmente definida pela cibernética. Finalmente, a última parte analisa os dispositivos de visão computacional baseados na detecção automática de gestos e movimentos voltados para a predição e o controle da conduta. Esse percurso genealógico evidencia dois vetores de transformação: uma progressiva abstração e expropriação dos gestos pela imagem técnica e um relativo apagamento da identidade e da subjetividade à medida que o viés preditivo da imagem técnica é reforçado.

Palavras-chave: Gesto; visão computacional; imagem técnica; abstração; previsão.

Abstract: This study traces a genealogy of automatic vision of gestuality. Its text is organized into three parts. Its first part describes how the technical reproducibility of images came to be used to replicate gestures, to abstract them, and to expropriate them in several fields. Its second part shows that the matrix of the contemporary technical model, aimed at predicting behaviour, is already partially defined by cybernetics. Finally, the last part analyses computer-assisted vision devices using the automated detection of gestures and movements to predict and control driving. This genealogical outline highlights two vectors of transformation: a gradual abstraction and expropriation of gestures by technical imaging and a relative erasure of identity and subjectivity when the predictive bias of technical imaging is reinforced.

Keywords: Gesture; computer vision; technical image; abstraction; prediction.



Figura 1: J. Préview. *What shall we do next? (Sequence #2)*, 2014

Fonte: https://www.preview.net/fr/videos_WSWDNSeq2.html. Acesso em: 29 ago. 2024.

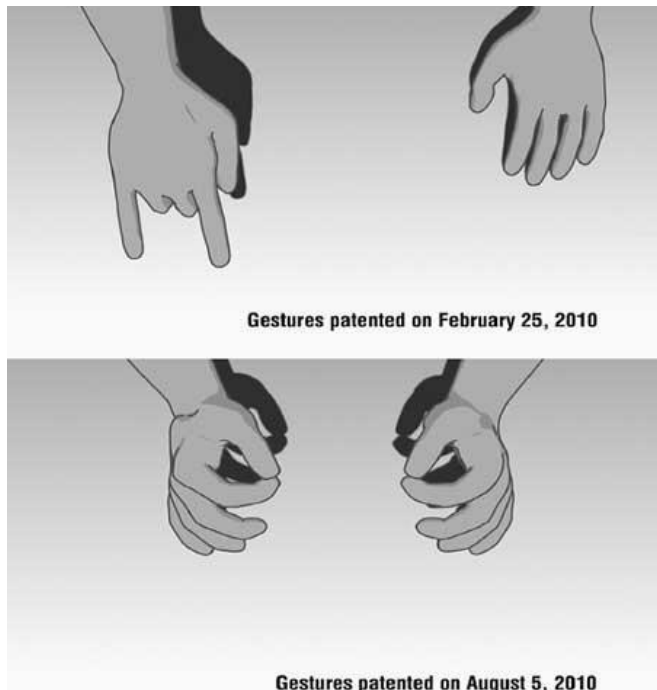


Figura 2: J. Préview. *What shall we do next? (Sequence #1)*, 2006-2011

Fonte: https://www.preview.net/fr/videos_WSWDNSeq2.html. Acesso em: 29 ago. 2024.

“O que faremos a seguir?” – a pergunta-título do trabalho de J. Préviex, do qual as Figuras 1 e 2 foram extraídas, aponta para uma questão central nos debates contemporâneos em torno da gestualidade. A Figura 1 apresenta cenas de um vídeo no qual vemos uma coreografia composta a partir de gestos e movimentos utilizados na interação com os mais variados *gadgets* e a Figura 2 é parte de uma animação que apresenta uma série de gestos patenteados no período de 2006 a 2011. Tais gestos compõem o crescente repertório de movimentos voltados para a interação com dispositivos eletrônicos como smartphones, computadores, tablets, consoles de games etc. Muitos desses aparelhos sequer existem no mercado, mas os movimentos que os acionam constituem, como caracteriza Préviex, em seu trabalho de 2006, um “arquivo dos gestos por vir”. Essa estranha forma de arquivo é singularmente sintomática de nosso tempo. Ela aponta para uma inscrição específica da gestualidade no presente e no futuro próximo, em que a relação homem-máquina ganha cada vez mais relevância. À linhagem do gesto mecanizado do trabalho industrial, da gestualidade acentuada na imagem em movimento do cinematógrafo, vieram se somar os gestos que nos conectam às máquinas com as quais cotidianamente percebemos, escrevemos, comunicamos, nos informamos, nos deslocamos etc. Reconfigura-se, aqui, não apenas o domínio do gesto, mas também da visão, especialmente da visão maquínica⁵. De fato, um dos campos mais promissores de pesquisa e desenvolvimento da visão computacional e das chamadas câmeras inteligentes⁶ é o da gestualidade. A proliferação de dispositivos que contam com a função *gesture recognition* atestam a centralidade do gesto na agenda dos desenvolvedores. Conforme estudo publicado pela agência Juniper (Moar, 2016), a estimativa era de que, no ano de 2020, disporíamos de meio milhão de dispositivos de detecção ou rastreamento de gestos.

A dimensão tecnopolítica desse crescente investimento das câmeras inteligentes sobre os gestos vincula-se amplamente à história das práticas policiais de arquivamento, categorização e reconhecimento de criminosos. Mas também diz respeito à própria absorção do gesto pela imagem técnica, inicialmente pela fotografia e posteriormente pelo cinema. Neste artigo, o papel do gesto nas cenas litigiosas, abordado modernamente a partir da promessa tecnológica de registro cinematográfico do movimento, constitui o contraponto histórico dos investimentos das máquinas de visão contemporâneas sobre o domínio da gestualidade.

⁵ Designamos por visão maquínica uma série de dispositivos com processos automatizados de visão (câmeras ou vídeo inteligentes, visão computacional etc.).

⁶ Câmeras que operam por algoritmos (acoplados, por vezes, a sensores de movimento e presença) voltados para a detecção automatizada de padrões na imagem.

As máquinas contemporâneas – de informação, de comunicação, de vigilância e de entretenimento – evoluem no campo da gestualidade. Por um lado, os aparatos eletrônicos e digitais são cada vez mais guiados por gestos que se incorporam progressivamente ao nosso repertório cotidiano: as diferentes maneiras de deslizar os dedos sobre as telas para acionar funções específicas; os movimentos orientados aos captosres que ativam portas, luzes, torneiras etc. Por outro lado, as máquinas são programadas para reconhecer, detectar e “ler” nossos gestos segundo esquemas que lhes permitem predizer nossas condutas e ações futuras. Há todo um campo da visão computacional que se volta para a detecção e predição de movimentos e gestos, particularmente no domínio da segurança.

Embora a intensificação das relações entre a máquina e o gesto se acelere hoje, essa relação tem uma história antiga. Propomos mostrar neste artigo como a decomposição, a interpretação e a antecipação dos movimentos interessam, desde a modernidade, a diferentes campos de conhecimento e de aplicação: a segurança, a indústria, a psiquiatria e a ciência da arte. Sublinhamos, nessa nota genealógica, dois vetores de transformação das visões maquínicas da gestualidade.

O primeiro é uma progressiva abstração e expropriação dos gestos. Observamos, por meio da fotografia e do cinema, como os gestos foram expropriados dos corpos e dos indivíduos com o objetivo de elaboração de um conhecimento específico a seu respeito e, sobretudo, com o objetivo de torná-los mais produtivos, expressivos ou eficazes. Há aqui um processo de apropriação do gesto pelo aparelho maquínico (a filmadora ou a máquina industrial) ao qual eles vão, enfim, pertencer. Isso quer dizer, segundo Adorno (2005, p. 40), seguir “a exigência das coisas”.

O segundo é um apagamento relativo da identidade e da subjetividade quando o viés preditivo da imagem técnica se fortalece. A decomposição do movimento e a leitura maquínica dos gestos mudam de alvo. Sobretudo no domínio da segurança e da vigilância, a dimensão preditiva se fortalece afastando-se do paradigma da identificação que era dominante na modernidade. A fotografia e o cinema prometiam o registro dos gestos visando o reconhecimento de criminosos. A expropriação progressiva dos gestos pelos aparatos modernos criou as condições para que as atuais câmeras inteligentes prometam evitar futuras infrações, sem passar, necessariamente, pela identificação. Os sistemas preditivos atuais, baseados na detecção algorítmica de gestos e movimentos, se voltam para o reconhecimento de modelos e de anomalias que permitem predizer e intervir sobre as ações e os comportamentos, sem conhecimento prévio a respeito da identidade de seus “autores”.

Analisaremos esses vetores de transformação nas três seções deste artigo. A primeira é consagrada ao modo como a reprodutibilidade técnica da imagem – fotográfica e cinematográfica – é posta a serviço da replicação do gesto, de sua abstração e de sua expropriação em diversos domínios: segurança e autodefesa, ciência, arte e indústria. Em seguida, veremos como a matriz do modelo maquínico contemporâneo, voltada para a predição de comportamentos, é parcialmente definida pela cibernética. Por fim, analisaremos os atuais investimentos maquínicos sobre a gestualidade, mostrando mais uma vez sua transversalidade em diferentes domínios (design, comunicação, entretenimento, marketing, segurança e vigilância). Focalizaremos os dispositivos de visão computacional e as câmeras inteligentes baseadas na detecção e na leitura automatizada de gestos e movimentos, buscando comparar as máquinas modernas de antecipação e de identificação com as máquinas contemporâneas de predição e controle algorítmico das condutas.

A imagem viva do gesto e sua reprodutibilidade técnica

Em um artigo em *La Vie au Grand Air*, de 8 de dezembro de 1906, o campeão de esgrima Jean-Joseph Renaud alertava contra os truques da gangue dos “Apaches” que atemorizava os cavalheiros burgueses nas ruas de Paris: como saber se esse cidadão que nos atravessa o caminho e nos saúda tocando o chapéu não irá inclinar-se demasiadamente e arremeter-se contra nós em violenta cabeçada? Nos primeiros anos do século XX, o “apachismo” era considerado a maior ameaça à vida normal dos parisienses. Entre os “apaches”, tudo era gesto e trejeito, não apenas a abordagem, os golpes, mas também uma dança peculiar, provavelmente inspirada no tango argentino, em que o casal simulava uma briga de rua. A análise minuciosa de seus movimentos (seus “truques”), tanto para fins de reconhecimento como de autodefesa, era tema recorrente na imprensa. (Hoobler, 2013) Essa análise é, desde o início, indissociável de uma *maquinação*, pois o objetivo primordial do apache é “‘tombar’ suas vítimas”: ao aproximar-se para pedir uma informação ou fogo, ele o faz timidamente, com um semblante muito humilde, mas subitamente o ritmo muda, o chapéu é lançado contra o rosto, a cabeça desfere uma violenta cabeçada e as mãos procuram arrancar as pernas da vítima do chão. A segmentação desses movimentos (Figura 3), com o auxílio da fotografia, era considerada um método bastante eficaz para difundir técnicas de autodefesa, que, por sua vez, eram baseadas nos princípios mecânicos de ação e reação (Renaud, 1906).



Figura 3: J.-J. Renaud. “Trucs d’apaches et leur Parades”, 1906

Fonte: <http://gallica.bnf.fr/ark:/12148/bpt6k9607877r/f82.item>. Acesso em: 29 ago. 2024.

A decomposição, interpretação e antecipação dos movimentos não interessavam apenas aos defensores da ordem pública. Na primeira metade do século XX, as relações entre gesto e intenção eram objeto de interrogação nos mais diferentes campos do saber. Quando o historiador da arte Erwin Panofsky pretende explicar os fundamentos de seu método de investigação – que privilegia o sentido das obras de arte em detrimento da sua forma –, nas primeiras páginas de seu *Studies in Iconology*, de 1939, não recorre, como exemplo, a uma pintura do cânone ocidental, mas ao evento de um “conhecido” que “me saúda na rua removendo o seu chapéu” (Panofsky, 1972, p. 3). Esse simples gesto abriria três níveis de interpretação: o primeiro, que o autor chama “primário” diz respeito ao reconhecimento formal dos objetos – seu sentido “factual” ao qual vem acrescentar-se um sentido “expressional”, que nos permitiria supor se esse “conhecido” está de bom ou mau humor, ou se “sentimentos em relação a mim são indiferentes, amigáveis ou hostis” (Panofsky, 1972, p. 3). Mas, para entender efetivamente esse gesto, entraria em jogo um segundo nível de sentido, que Panofsky chama “secundário” ou “convencional”, pois é apenas conforme um código ocidental de comportamento que levantar o chapéu corresponde a uma forma polida de saudação. Apenas o terceiro nível de análise, no entanto, permitiria ao observador vislumbrar a “personalidade” do gesticulador – isto é, sua maneira de “ver as coisas e reagir ao mundo”. Panofsky (1972, p. 3-5) admite que um gesto isolado como uma saudação polida não pode manifestar tudo isso “compreensivamente”, mas apenas, “sintomaticamente”: “nós não podemos construir um retrato mental do homem com base em uma única ação, mas coordenando um grande número de ações similares e interpretando-as de acordo com a nossa informação geral acerca do período, nacionalidade, classe, tradições intelectuais do cavalheiro e assim por

diante.”. O terceiro nível seria o do “sentido intrínseco” ou “conteúdo”, objeto por excelência da iconologia como disciplina acadêmica.

Tanto para fins de autodefesa como de história da arte, reconhece-se a insuficiência preditiva do gesto tomado em si mesmo. Ele precisa ser associado a um amplo conjunto de informações para tornar-se plenamente legível. Uma legibilidade, no entanto, que só é possível quando se ultrapassa a dimensão meramente expressiva ou convencional do gesto e nos aproximamos de sua dimensão sintomática, inconsciente (Panofsky, 1972). O cinema foi, sem dúvida, o grande responsável por difundir a noção de que o gesto apresentava uma dimensão inconsciente. Sem a decupagem da tomada cinematográfica, é improvável que narrativas provindas de campos tão distintos, como as de Renaud e Panofsky, caminhassem pela mesma calçada e cumprimentassem uma à outra com um leve toque na aba do chapéu. De fato, desde os experimentos dos cronofotógrafos como Étienne-Jules Marey, Edward Muybridge e Albert Londe há um despertar para a legibilidade dos gestos em campos tão díspares como a indústria, a psiquiatria e a ciência da arte. Em 1931, Walter Benjamin (1985, p. 94) não fez mais do que dar um nome a esse fenômeno – “inconsciente ótico”:

Percebemos em geral o movimento do homem que caminha, mas nada percebemos de sua atitude na exata fração de segundo que ele dá um passo. A fotografia nos mostra essa atitude, através de seus recursos auxiliares: câmera lenta, ampliação. Só a fotografia revela esse inconsciente ótico, como só a psicanálise revela o inconsciente pulsional.

Não surpreende, portanto, que já no primeiro *Congrès de Police Judiciaire Internationale*, realizado em 1914, em Mônaco, no qual a “ficha parisiense” de Bertillon é adotada como modelo para a criação de uma “ficha internacional de dados de segurança”, uma das participantes, Agathe-Berthe Dyvrande, advogada no tribunal de Apelação de Paris, tenha sugerido que câmeras cinematográficas fossem utilizadas nos laboratórios de identificação criminal para filmar os indivíduos em “suas atitudes habituais”. Esse Congresso, cujas atas só foram publicadas em 1926, por causa da guerra, foi um impressionante esforço para criar uma plataforma europeia comum de combate ao crime, afinal, “o universo é um, e os criminosos não respeitam as fronteiras levantadas pela política ou pela história” (Premier Congrès..., 1926, p. X) O ânimo colaborativo dos delegados é tão elevado que eles chegam a aprovar a adoção do esperanto como “língua auxiliar” na comunicação entre as polícias do mundo. Mas o congresso também é o canto do cisne da Bertillonagem como sistema de identificação. A delegação francesa, numericamente muito

superior a dos demais países, pretende estabelecer um *Bureau Central de Indentité Judiciaire* em Paris e a criação de um sistema ideográfico universal que permitisse a padronização e a transmissão telegráfica dos dados antropométricos. O próprio Bertillon deveria ter estado presente, mas falecera poucos meses antes, sendo objeto de gongóricas homenagens por parte de vários oradores, as quais culminaram em um telegrama enviado à viúva do “inventor genial” (Premier Congrès..., 1926, p. 133). Sua “descoberta” imprescindível equiparava-se, “guardadas as devidas proporções, na ordem das técnicas policiais, à de Pasteur no domínio das ciências biológicas” (Premier Congrès..., 1926, p. 81-82). A despeito do sistema classificatório DKV (nariz/orelha), criado por Bertillon, ter se mostrado inviável à medida que o arquivo alcançava a dimensão de centenas de milhares de fichas, admitiu-se apenas sua inadequação quando se trata da identificação de mulheres e crianças. A delegação francesa aceitou, em virtude disso, que o grande arquivo judiciário universal fosse classificado pelo sistema datiloscópico argentino, criado por Vucetich, e já utilizado em vários países do mundo (Premier Congrès..., 1926, p. 47)

Nesse contexto, a proposta apresentada por Mlle. Dyvrande procurava restabelecer o pioneirismo perdido pela criminologia francesa. Não era, originalmente, uma ideia da polícia, mas de Albert Gaveau, diretor da *Pathé-Journal*, que a submeteu à consideração Célestin Hennion, prefeito de Polícia de Paris (entre 1913 e 1914) e que havia criado a primeira escola de formação de policiais, em 1910. O *Pathé-Journal*, pioneiro no ramo das atualidades cinematográficas, lançado em 1909, tinha por objetivo, igualmente, realizar filmes que servissem ao treinamento dos policiais nessa escola. Os congressistas aderiram a essa ideia com entusiasmo: filmes realizados na Rússia mostravam, por exemplo, como o comportamento de uma multidão dependia do posicionamento dos agentes; seria possível comparar boas e más atitudes dos policiais; demonstrar como se coletam provas em uma cena de crime; ou ilustrar o *modus operandi* de certos criminosos (Premier Congrès..., 1926, p. 93). Aparentemente, o *Pathé-Journal* chegou a produzir algumas imagens que, tendo desaparecido, não puderam ser exibidas em Montecarlo por Mlle. Dyvrande. Foram substituídas por uma retórica entusiasmada. A advogada admitia que a fotografia facilita a pesquisa e o reconhecimento, “mas ela não é mais do que a representação de um busto, uma face, um perfil – falta-lhe animação, a vida, que permite descobrir as atitudes, os gestos familiares, as marcas particulares, manias, tics do indivíduo” (Premier Congrès..., 1926, p. 209). Na rigidez da pose do estúdio antropométrico, essas “atitudes naturais” se perderiam. O cinema produziria uma “imagem concreta viva”, em vez de, apenas, “palavras abstratas inscritas sobre uma

ficha pouco evocativa”⁷. O caráter vivo da imagem, no entanto, era indissociável da natureza intrinsecamente mecânica do movimento, uma vez que com o concurso do cinematógrafo os gestos poderiam “ser lentamente decompostos e estudados” (Premier Congrès..., 1926, p. 210).

As fichas criminais francesas, depois de uma longa resistência de Bertillon, passaram a incluir, em 1902, as impressões datiloscópicas. Em virtude disso, clamava-se, a ficha sinalética francesa havia se tornado a “mais completa do mundo”. Agora, graças à proposta de Dyvrande, essa completude avançava ainda mais, pois passaria a incorporar as “fotografias cinematográficas”. O valor da proposição era tão evidente (“é um progresso”), que o presidente do Congresso não via o “porquê não seria aceita” (Premier Congrès..., 1926, p. 93). Um mês depois, o jornal *Le Film* informava que a polícia de Paris iria usar o cinema em um novo curso para formar seus “agentes da paz”. Graças a Pathé, o agente aprenderia como proteger “a criança perdida”, “o velho aleijado”, ou como conduzir em segurança os “alienados perigosos”. A revista saúda a iniciativa: “o futuro confirmará que o cinema é o único capaz de instruir sem entediar” (Bernard, 1914, p. 7).

Nessa que talvez tenha sido a primeira formulação de uso sistemático da imagem em movimento para fins policiais, observamos claramente dois impulsos antagônicos: um de decomposição do movimento que conduz à abstração do gesto, e outro de recomposição que pretende reencontrar a imagem viva. Porém, ambos estão claramente subordinados ao paradigma da identificação. Não há representantes ingleses e norte-americanos no congresso, bem como estão ausentes as potências imperiais (Alemanha, Áustria-Hungria, Turquia), com exceção da Rússia. Não é por acaso, portanto, que a única objeção levantada no congresso de 1914 dissesse respeito ao custo elevado da proposta. Em um contexto em que a ficha sinalética antropométrica perdia aceleradamente terreno para o modelo anglo-saxônico, que privilegiava e reduzia a identificação judiciária às impressões digitais, os defensores de Bertillon, recorrendo ao cinema, apegavam-se à dimensão inefável do gesto, isto é, à própria “vida”.

Três décadas depois do Congresso, Léon Lerich, prestigiado Juiz de Instrução Adjunto no Tribunal do Sena, lamenta que o projeto proposto por Mlle. Dyvrande

⁷ Como já foi observado, o tema da produção de uma “imagem viva” a partir de elementos mecanicamente decompostos ocorre pela primeira vez no romance de Auguste Villiers de L’Isle-Adam, *A Eva Futura*, de 1880. Em uma bela passagem, o inventor promete ao amigo desesperadamente apaixonado apoderar-se “matematicamente” e “com os recursos da ciência” da graça dos gestos, do perfume da carne, do timbre da voz, da luz dos olhos, dos movimentos e do modo de andar e tudo mais o faz de uma mulher a mulher ideal para então reproduzi-la com exatidão, “com a ajuda sublime da luz” (Oubiña, 2009, p. 28).

não tenha tido qualquer tipo de continuidade (Lerich, 1949). De fato, o processo de abstração do gesto, sem o qual as máquinas preditivas contemporâneas não teriam como operar, avançou muito mais lentamente na polícia do que em outros campos. Desde meados de 1870, a fotografia, a estereoscopia e a cronofotografia estiveram a serviço da elaboração do “grande ataque histérico”, como uma sucessão de figuras e poses exibidas pelos pacientes do Dr. Charcot (Didi-Huberman, 2003). De um paciente, Charcot dirá, por exemplo, que “parece um autômato que está sendo puxado por um fio” (Agamben, 2000, p. 51). Na década seguinte, em 1889, Mauricio Vidal Portman inicia seu exaustivo trabalho de documentação dos Andamaneses, no qual dezenas de fotografias são empregadas para “mostrar cada passo da feitura de uma arma, etc., tão claramente, [...] que seria possível para um trabalhador europeu imitar o seu trabalho” (Pinney, 2011, p. 38). A reprodutibilidade técnica da imagem é colocada a serviço da replicação do gesto. A decomposição do movimento – condição tanto da reprodução do artefato como da generalização da sintomatologia histérica – assume claramente seu caráter de expropriação aqui. Era de se esperar, portanto, que a realização mais bem-acabada da decomposição do gesto – como expropriação e abstração – ocorra no campo do “gerenciamento científico” do trabalho industrial, no âmbito do taylorismo. Nas *ciclografias* de Frank B. Gilbreth, a trajetória dos gestos é marcada por luzes associadas às extremidades dos membros dos trabalhadores (Figura 4). Há uma dupla maquinação aqui: a primeira diz respeito a esse corpo, cujos traços individuais, em particular as feições do rosto, se apagam, para que a cinemática do gesto possa ser registrada; a segunda é a contraposição das linhas luminosas ao diagrama antropométrico – de uso corrente tanto nos estudos antropológicos como na identificação criminal. A projeção sobre o diagrama não visa aqui, porém, a singularidade do indivíduo ou a definição de um tipo, mas o mapeamento dos “movimentos torpes ineficientes ou mal dirigidos” que, a despeito do prejuízo que causam à indústria “não deixam nenhum rastro tangível ou visível atrás deles” (Oubiña, 2009, p. 36). O desenho do gesto eficiente resulta não apenas de uma despersonalização do trabalhador, mas da depuração de todos os seus trejeitos e tiques idiossincráticos que, no contexto da linha de produção, contribuem para o desperdício de tempo e de material. A diferença com o paradigma da identificação mostra-se claramente aqui, pois são esses mesmos trejeitos e tiques que Agathe Dyvrande pretendia registrar para melhor reconhecer os criminosos.

Nos usos posteriores da técnica criada por Gilbreth, a ambivalência do registro visual do gesto é percebida com muita clareza. Associando a longa exposição à iluminação estroboscópica (basicamente uma combinação dos procedimentos de

Gilbreth e Jules-Marey), Gjon Mili retratou não apenas operários como maestros, bailarinos, jazzistas e, em sua sequência mais famosa, o pintor Pablo Picasso (Figura 5). O gesto desperdiçado na linha de montagem agora é o excesso – de talento e de movimento – que singulariza um artista e sua obra. Um ano antes dessa fotografia, Siegfried Gideon havia observado, em seu ensaio sobre a mecanização da vida cotidiana, a semelhança entre as ciclografias de Gilbreth e as pinturas de Paul Klee, Wassily Kandinsky e Joan Miró, assinalando que para o cientista as “trajetórias” se tornaram “entidades com leis independentes” enquanto para os pintores a “pura forma do movimento converteu-se em objeto artístico de pleno direito” (Gideon, 1948, p. 104-113).

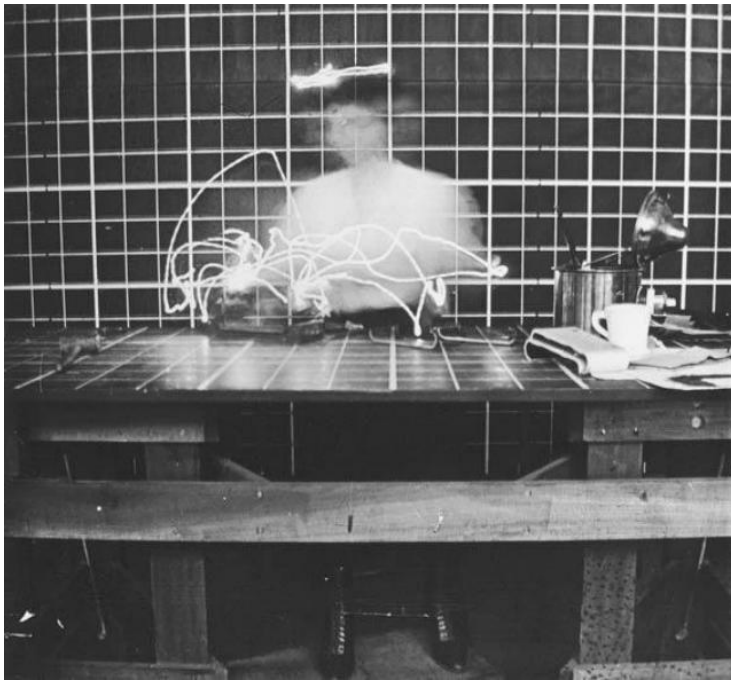


Figura 4: Frank Gilbreth. Ciclografia. Estudo sobre a eficiência da movimento
Fonte: National Museum of American History.

Entre os muitos personagens retratados por Mili, um nos interessa em particular: o agente do Departamento Federal de Investigação; em inglês, *Federal Bureau of Investigation* (FBI), Delf “Jelly” Bryce (Figura 6), considerado o “gatilho mais rápido” da América – era capaz de sacar um revólver e atirar, acertando o alvo, em 0.4 segundos. Reza a lenda que, nos anos 1940, bastava anunciar que “Jelly” estava

chegando ao local em que os bandidos se encontravam que, para evitar enfrentá-lo, entregavam-se imediatamente. Ao contrário de Picasso, cujo gesto sugere imaginação e espontaneidade, a ação de Bryce é minuciosa e milimétrica. Costumava treinar seu movimento por horas seguidas diante do espelho e, por mais absurdo que isso possa nos parecer, procurava ser mais rápido que si próprio.



Figura 5: Gjon Mili.
Pablo Picasso, pintor, 1949

Fonte: © Gettyimages.



Figura 6: Gjon Mili. Delf Bryce,
agente do FBI, 1945

Fonte: © Gettyimages.

Tanto o pintor como o detetive têm seus gestos aprisionados pela fotografia em seu “estranho espaço confinado” (Price, 1994), mas a imagem do agente Bryce diante do espelho nos é, hoje, demasiadamente familiar: “*Are you talking to me?*” – talvez dissesse para o outro de si, desafiadoramente, como Robert de Niro, em *Taxi Driver*, de Martin Scorsese. Pois seu espaço de confinamento não é outro senão o próprio espaço fílmico, tal como ele é originalmente constituído pela alternância entre plano e contraplano. De fato, Walter Benjamin já havia assinalado que a invenção da fotografia fora tão significativa para a criminologia moderna quanto a imprensa foi para a literatura – e, *evidentemente*, para a combinação das duas no desenvolvimento da novela policial. Desde então, observa, “não tiveram fim os esforços de capturar um homem na sua fala e nas suas ações” (Benjamin, 2003, p. 27). Capturado pelo espelho cinematográfico, o agente Bryce está condenado a repetir infinitamente o mesmo gesto nesse duelo imaginário. Como agudamente

sugere Giorgio Agamben, “para seres humanos que perderam qualquer senso de naturalidade, cada gesto singular se torna um destino” (Agamben, 2000, p. 52). A ambivalência do registro visual do gesto, apontada acima, também é ressaltada pelo filósofo, para quem “toda imagem é animada por uma polaridade antinômica: por um lado, imagens são a reificação e a obliteração de um gesto [...]; por outro, preservam a *dynamis* intacta” (Agamben, 2000, p. 54). Por isso, o cinema pode ser o lugar em que uma sociedade que “perdeu seus gestos tenta ao mesmo tempo reivindicar o que perdeu e registrar sua perda” (Agamben, 2000, p. 52). Mas de que perda se trata afinal? Perder-se de que e para quem?

Quando Elia Kazan cria, em 1947, o *Actors Studio*, em Nova York, coloca o método de ações físicas de Stanislavsky a serviço de uma técnica de atuação que permite conectar diretamente um gesto a um “personagem” ou “papel”. Sua adequação à produção cinematográfica não decorre, apenas, do caráter naturalista das lições do mestre russo. Ao admitir, no interior de cada cena, uma variedade de “unidades”, correlacionando “objetivos” e “ações” a cada uma delas, o método, de fato, promete reunir ações decupadas e produzidas sem linearidade cronológica ou causal em alguma instância superior (o filme, a narrativa ou a humanidade). Cada gesto do ator ou atriz, decomposto pelo cinematógrafo (quadro a quadro e, principalmente, plano a plano) é resgatado, acolhido e tem sua vida e seu sentido restabelecido pelo espírito que habita o rolo do filme como habitava antes o interior do corpo do ator (Gordon, 2010; Whyman, 2008).

Fora do cinema, no entanto, a decomposição mecânica do gesto e a perda a ela associada não encontram conforto em qualquer forma de transcendência. Theodor Adorno talvez tenha sido o primeiro filósofo a nos alertar a respeito do automatismo das portas. Em “Entre sem bater”, apontamento de 1944, incluído em *Minima Moralia*, afirma que “a tecnologia está tornando os gestos precisos e brutais, e os homens com eles” (Adorno, 2005, p. 40). Despidos de toda hesitação, os gestos teriam se convertido em “exigências das coisas”. Portas de geladeiras e de automóveis tinham que ser “batidas” e “ninguém mais sabe fechar uma porta gentilmente, mas de modo firme” (Adorno, 2005, p. 40). As portas com molas, que se tornaram populares no início dos anos 1940, impunham maus modos às pessoas, como não olhar para ver se vem alguém atrás. Sob a lei da funcionalidade, o contato com objeto estaria limitado à mera operação e tudo que não pode ser “consumido” nesse momento – tudo aquilo de onde poderia provir uma “experiência” – é banido da ação (Adorno, 2005). Poucos parágrafos adiante, observa, a respeito da perda de gentileza:

Toda membrana interposta entre os homens e suas transações é sentida como uma perturbação ao funcionamento do dispositivo, ao qual estão não apenas objetivamente incorporados, mas com os quais orgulhosamente se identificam. Assim, ao invés de levantar seus chapéus, cumprimentam-se uns aos outros com olhares de familiar indiferença (Adorno, 2005, p. 41).

Controle e predição: a cibernética

A mecanização do gesto – indissociável de sua decomposição e condição de sua replicabilidade – visava, em última instância, sua conformação a uma finalidade. Tratava-se fundamentalmente da economia de um ajuste biodinâmico à máquina de aparafusar, de escrever ou de duelar, sem o qual os objetos técnicos não poderiam propriamente operar como “dispositivos”, no sentido estrito que esse conceito vai adquirir na obra de Foucault – isto é, como dimensão positiva na qual as relações de poder se materializam em instituições, discursos, técnicas, regras e processos de subjetivação (Agamben, 2009). O apogeu paradoxal dessa mecanização do gesto foi a cibernética. Na síntese brilhante produzida por Norbert Wiener (1989, p. 16-17), “a teoria do controle na engenharia, seja humana ou animal ou mecânica, é um capítulo na teoria das mensagens”. Na utopia cibernética, a porta automática da *Pennsylvania Station*, em Nova York (o exemplo é do próprio Wiener), é o paradigma das máquinas sensíveis: da mínima informação proveniente do “mundo exterior” decorre a ação mais exata e precisa do mecanismo. As máquinas automáticas evoluíam segundo o mesmo princípio dos animais: “controlar a entropia por meio da retroalimentação” (Wiener, 1989, p. 26). Com a cibernética, de fato, o último vestígio do paradigma cartesiano dos organismos vivos como relógios desaparece, pois a própria máquina deixa de ser uma simples engrenagem. O autômato cibernético coabita nosso universo termodinâmico e assimila as propriedades do vivo.

Retornemos ao agente do FBI que permaneceu todo esse tempo diante do espelho, procurando sacar sua arma cada vez mais rápido. De que vale todo esse aprimoramento se o que ele tem agora diante dos olhos é um caça-bombardeiro e, nas suas mãos, um canhão antiaéreo? O que pode fazer a não ser seguir o alvo procurando em vão atingi-lo em uma ação que parece ocorrer em câmera lenta? Mas, Wiener sugere, em 1950, que um novo canhão antiaéreo pode ser construído – um que “observe por si mesmo as estatísticas concernentes ao avião alvo” (Wiener, 1989, p. 62), pois suas manobras evasivas, por mais variados que sejam os pilotos e as aeronaves, estão sujeitas a regularidades. Desse modo, o canhão estaria

engajado não apenas em uma ação de combate, mas em uma ação de aprendizado que lhe permitiria antecipar as eventuais manobras do piloto. Para que as máquinas fossem capazes de antecipar os movimentos, para que se tornassem “*prediction machines*”, deveriam ser capazes de aprender (Wiener, 1989, p. 86-88). Esse estranho duelo, que coloca, de um lado, um piloto em uma cabine ergonomicamente otimizada para reagir ao seu menor comando – isto é, para que cada movimento seu seja tão preciso, exato e desprovido de desperdício quanto possível – e, de outro, uma máquina capaz de antecipar sua errática trajetória a partir de padrões estatísticos inferidos de experiências anteriores, é a cena de origem da autonomização do gesto na contemporaneidade.

Um dos antepassados do canhão sonhado por Wiener foi o *Kerrison Predictor*, computador analógico criado no final da década 1930, na Inglaterra, para orientar os tiros de um canhão antiaéreo. A *British Pathé* produziu, em 1939, uma peça de propaganda sobre ele, exibida nos cinemas ingleses às vésperas da Segunda Guerra Mundial. Chamava-se *Predictions while you wait*, e seu objetivo era claramente duplo: apresentar uma máquina capaz de *prever* trajetórias de aviões inimigos e, simultaneamente, demonstrar que o governo se prevenia para a guerra. Diante de um diagrama da máquina, o locutor informa que se trata de um *Predictor*, “um dos mais engenhosos equipamentos em nosso esquema de defesa”. E, provavelmente, antecipando a perplexidade da audiência diante de seu formato inusitado, acrescenta: “parece com uma câmera de filmar” (*Predictions...*, 1939). A semelhança entre os dois dispositivos dissolvia o estranhamento, diluía a intranquilidade resultante do encontro entre máquinas bizarras e os temores da guerra, remetendo o *Predictor* ao universo familiar do entretenimento de massas.

À decomposição do gesto na fotografia e no cinematógrafo, que simultaneamente mecanizava o vivo – para fins de disciplina, controle, exame, diagnóstico e desempenho – e nos encantava com a efêmera restituição de seu movimento nas salas de cinema, a cibernética veio agregar a animação termodinâmica das máquinas automáticas e, posteriormente, preditivas. Que esse confronto dramático tenha ocorrido em um campo de batalha e não em uma sala de identificação judiciária revela os limites que a criminologia historicamente enfrentou para operar fora do paradigma da identificação (uma vez que o direito lhe impunha necessariamente a imputação de um determinado crime a um culpado em particular).

A cibernética, no entanto, estabelece a base de um modelo preditivo que prevalecerá sobre o paradigma da identificação, mesmo no domínio da segurança. A máquina preditiva idealizada por Wiener – o “preditor antiaéreo (AA)” –

era programada pelos dados estatísticos das performances de voos passados para prever matematicamente a posição futura do avião inimigo. Peter Galison (1994), em sua excelente análise da visão cibernética, mostra como o sistema preditivo concebido por Wiener (válido para humanos e para máquinas) estava estreitamente ligado a uma ontologia do inimigo. No campo de batalha mecanizado da Segunda Guerra Mundial, um “abismo de distância, de velocidade e metal” nos afasta do piloto inimigo que lança os mísseis do alto (Galison, 1994, p. 233). Trata-se de um inimigo escondido, de tal modo misturado a sua armadura maquinaica, que as fronteiras entre humano e não-humano tornam-se fluidas. A construção de uma máquina capaz de prever a posição do avião-piloto a tempo de abatê-lo dependia do conhecimento de seu comportamento, apesar de sua opacidade. Wiener e Bigelow notam que o piloto inimigo sob estresse apresenta regularidades e esquemas em sua história comportamental, semelhante a um servomecanismo. O servomecanismo, já presente nos estudos de Wiener sobre os dispositivos capazes de autorregular suas ações em função de um objetivo e de suas trocas informacionais com o ambiente, torna-se o modelo, ao mesmo tempo, do piloto inimigo e da máquina que o vencerá⁸.

É importante sublinhar que o conhecimento do comportamento da máquina-piloto inimiga e os cálculos matemáticos concernentes a sua posição futura são estabelecidos sem nenhuma referência à sua constituição e/ou a seus estados internos. A ênfase sobre o comportamento manifesto fundamenta-se em uma abordagem behaviorista que privilegia as manifestações diretamente observáveis, em detrimento de toda referência a estados mentais ou internos animados por intenções, crenças ou desejos. Galison sublinha que essa opacidade do Outro está na base da perspectiva cibernética do inimigo e também do organismo humano e não-humano, da máquina e do mundo. “Nós somos verdadeiramente, nessa visão de mundo, como as caixas pretas com entradas e saídas e sem acesso à vida interior de outra pessoa” (Galison, 1994, p. 256).

Esse modelo de máquina preditiva, visando antecipar as ações futuras de atores inicialmente concebidos como caixas pretas, persiste, com algumas variações, nos atuais dispositivos de predição de gestos e comportamentos em diversos domínios, notadamente no setor de segurança.

⁸ O preditor AA enfrentou um problema de retroação diferente dos servomecanismos estudados precedentemente por Wiener, tais como os termostatos e os torpedos autoguiados. O tempo entre a ação e o efeito era maior e tinha que ser calculado e controlado, já que o projétil levava alguns segundos para atingir o seu alvo.

A cibernética estabelece assim as bases de um modelo preditivo que prevalece sobre a identificação, mas não a anula. Mostraremos que a ênfase colocada na predição da ação (mais do que na identificação do agente) anda de mãos dadas com um distanciamento relativo dos conhecimentos disponíveis sobre a identidade ou a subjetividade dos indivíduos observados. Notaremos igualmente que os modelos maquínicos atuais de detecção e predição de comportamentos radicalizam a expropriação e a abstração do gesto tanto no domínio da segurança quanto no do consumo e das interações humano-máquina.

Detecção e antecipação automatizada dos gestos

Retomemos o método empregado para decompor e antecipar os gestos da gangue dos Apaches em contraste agora com os mecanismos automatizados de “detecção de brigas”, hoje incorporados a sistemas de videovigilância em presídios, escolas, hospitais psiquiátricos ou mesmo no espaço urbano. Tais mecanismos consistem em programas computacionais capazes de reconhecer gestos e condutas que indiquem situações de briga. A maior parte dos programas dessa natureza volta-se para o monitoramento e detecção de comportamentos no seio de pequenos grupos ou de multidões⁹. O processo de inscrição dessa função de reconhecimento na máquina envolve dotá-la da habilidade para diferenciar, na imagem, o regular e o irregular no plano da gestualidade e do movimento. A irregularidade, no caso em questão, indicaria uma briga por vir ou em curso. Tal habilidade implica, como no método aplicado à Gangue dos Apaches, a decomposição do movimento, mas aqui o princípio é sobretudo algorítmico e não mais visual. Não se trata de segmentar para ver melhor e mais lentamente os truques escondidos nos contratempos do gesto coreografado, mas de rastrear a imagem para detectar mais rápido a sua mudança de padrão. A abstração do gesto se radicaliza e se traduz em características espaço-temporais da imagem de vídeo, legíveis aos olhos da máquina como vetores de movimento, direção, magnitude, velocidade ou textura (Figura 7).

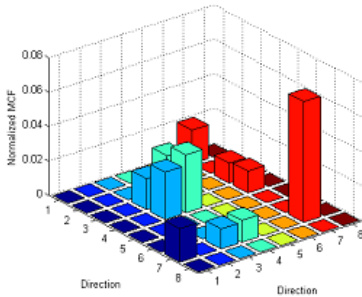
⁹ Outro contexto no qual encontramos os detectores de briga é o da classificação etária automatizada de vídeos em plataformas on-line. O elevado volume de produtos audiovisuais produzidos e colocados em circulação atualmente coloca o problema da força de trabalho empregada na classificação dessas imagens em função de seus conteúdos. As cenas de violência e sexo, consideradas impróprias para certos públicos, podem ser hoje classificadas a partir dos mesmos algoritmos de reconhecimento de gesto e movimento em operação nos sistemas de vídeo-vigilância. Cf. E. Bermejo, O. Deniz, G. Bueno, R. Sukthakar, “Violence detection in video using computer vision techniques”, *14th Int. Congress on Computer Analysis of Images and Patterns*, p. 332-339, 2011.



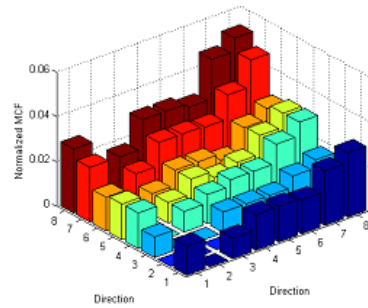
(a)



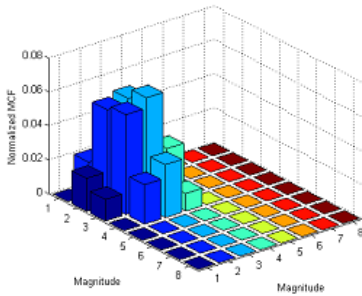
(a')



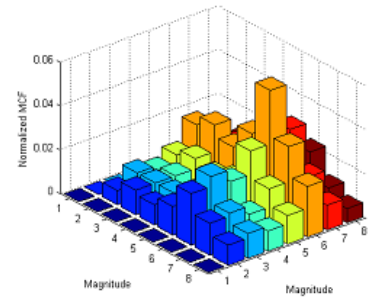
(b)



(b')



(c)



(c')

Figura 7: Detecção automatizada de brigas por videovigilância

Fonte: Esen; Arabaci; Soysal (2013).

A detecção da irregularidade na imagem deve ser quase simultânea à intervenção, que pretende, neste caso, desviar o curso da ação. Conectado a instâncias de segurança, o sistema faz soar o sinal de alarme no campo atencional do vigilante, permitindo uma intervenção no momento ou antes do incidente previsto. A imagem, já provida de um tempo real de observação, agora almeja um tempo real de reação. Mais ainda: o ideal dessa detecção automatizada é estar um passo à frente do desastre,

de modo a antecipar e evitar o incidente por vir. Ainda que o horizonte de todo sistema de detecção de brigas seja, em certa medida, preditivo, há projetos que se voltam especificamente para a detecção de condutas e gestos que seriam próprios ao momento que antecede uma briga ou situação de violência, como na pesquisa *Pre-fight detection: Classification of Fighting Situations Using Hierarchical AdaBoost* (Blunsden; Fisher, 2009). A forma potencial do gesto violento deve ser detectada a tempo de conjurá-lo.

Os mecanismos de detecção de brigas nos oferecem um contraponto à leitura moderna dos gestos potencialmente violentos, permitindo-nos compreender alguns dos princípios que dirigem o atual processo de automatização e predição de condutas. Tais mecanismos integram toda uma geração de sistemas de videomonitoramento e de visão computacional providos de uma camada dita “inteligente”, capaz não apenas de observar, transmitir e registrar imagens em tempo real, mas de “ler” essas imagens segundo parâmetros previamente definidos. A proliferação, no início do século XXI, de câmeras de videovigilância ao redor do mundo pode ser vista como uma vitória parcial e tardia da proposta de Mlle. Dyvrande e M. Gaveau, uma vez que a imagem em movimento finalmente incorporou-se à rotina da segurança pública. Entretanto e apesar de sua quase onipresença, a videovigilância convencional mostrou-se pouco eficaz no combate ao crime, entre outras razões, pelos limites da percepção e da atenção humanas em lidar com o volume e a monotonia de tais imagens. O conhecimento dessa limitação é muito anterior à popularização dos sistemas de videovigilância e remonta aos problemas relacionados ao monitoramento por radar na Segunda Guerra Mundial. A fadiga e as falhas dos operadores de radar na detecção de sinais mobilizaram estudos sobre os limites da atenção que, nesse mesmo período, se vinculam ao termo vigilância no campo da psicologia. Ao tipo de atenção empenhada por tais operadores deu-se o nome de *vigilância* para designar uma *atenção sustentada*; tais termos ainda hoje são usados como sinônimos em algumas teorias da atenção. Até então, nesse campo de estudos, vigilância designava um estado ou sentido fisiológico de ativação e receptividade relativo ao funcionamento geral do sistema nervoso (Head, 1923 *apud* Gómez-Iñiguez; Carbonelli-Vayá; Rio, 1999). Pela via dos operadores de radar (Mackworth, 1948 *apud* Gómez-Iñiguez; Carbonelli-Vayá; Rio, 1999), o termo vigilância passou a significar uma forma específica de atenção, que envolve focalização, seleção, continuidade, finalidade e antecipação. Os estudos de Mackworth mostram como, no homem, a atenção sustentada entra em declínio após meia hora de atividade focalizada. Os radares modernos passam, assim, a incorporar a detecção automatizada de qualquer sinal inesperado ou não familiar.

Reencontramos problema e solução similares no campo da videovigilância nas primeiras décadas do século XXI. A queda no desempenho atencional do operador de câmera, após certo período, pode comprometer a sua percepção quando algo efetivamente relevante acontece. A adição de uma camada “inteligente” destina-se justamente a contornar esse limite e selecionar nas imagens momentos e regiões para onde deve se dirigir a atenção dos operadores humanos. A utilidade pedagógica da imagem em movimento para a defesa social, entrevista por M. Gaveau, destina-se agora não aos policiais, mas às próprias máquinas. São elas que devem aprender a reconhecer na cena situações suspeitas, de risco ou de perigo. Não se trata mais de tornar visíveis ações que transcorrem aquém dos limiares perceptivos humanos (“inconsciente ótico”) ou aprimorar, por meio da observação dos movimentos, as habilidades cognitivas dos agentes. O ideal de uma máquina ergonômica tão rápida quanto o pensamento humano estava fadada ao fracasso como o sonho do agente Jelly, que pretendia ser mais rápido do que ele mesmo. Esse ideal sucumbiu diante da fatal distração que, mais cedo ou mais tarde, acometeria os vigilantes.

Só a videovigilância “inteligente” poderia sustentar o ideal comercial-securitário de monitoramento 24 horas desperto e maximamente espreado, sem ser assombrada pelo seu próprio excesso. A visão algorítmica das câmeras promete uma vigilância contínua e eficaz, permitindo uma maior mobilidade tanto aos agentes de segurança quanto aos corpos vigiados. A publicidade do grupo *Thalès*, que desenvolveu o sistema inteligente de detecção de comportamentos suspeitos empregado pela *SNCF*, a companhia ferroviária nacional da França, anunciou a seguinte promessa para a vigilância dos aeroportos:

Graças a essa solução inovadora, o operador de segurança – por trás de suas telas ou circulando no aeroporto, pois o sistema permite igualmente mostrar e controlar os fluxos de vídeos sobre terminais móveis para uma vigilância “nômade” – dispõe de agora em diante de uma poderosa ferramenta de detecção e monitoramento de comportamentos anormais para assegurar uma segurança ideal a milhares de pessoas circulando cotidianamente nos grandes aeroportos internacionais (Thales Group, 2015)¹⁰.

A inspeção dos corpos segue sem que se interrompa, salvo em caso de suspeita ou irregularidade, o curso de suas ações e de sua mobilidade cotidiana. Além do monitoramento em tempo real, a observação dos corpos moventes em

¹⁰ Disponível em: <https://www.thalesgroup.com/fr/worldwide/securite/news/smart-videoprotection>. Acesso em: 29 ago. 2024.

seus deslocamentos habituais tem uma função cognitiva importante, pois permite conhecer os padrões de conduta de indivíduos e massas em seu ambiente. Conhecimento decisivo para garantir o aprendizado das máquinas na tarefa de reconhecer desde brigas e situações de violência até diversos tipos de comportamento considerados suspeitos, de risco ou simplesmente indesejados: corpos andando no contrafluxo, ingressando nas chamadas “zonas de interesse”, permanecendo por tempo demasiado em locais de intensa mobilidade (aeroportos, estações de trem ou metrô, centros comerciais), efetuando trajetos atípicos em contextos específicos (como estacionamentos ou em vias urbanas engarrafadas). Ou ainda, no caso das máquinas voltadas para o monitoramento de multidões, alterações na forma, no ritmo ou na densidade das aglomerações (Shao; Change Loy; Wang, 2014) (Figura 8).



Figura 8: Detecção automatizada de comportamento de multidões

Fonte: http://www.ee.cuhk.edu.hk/~jshao/papers_jshao/jshao_group_in_crowd_cvpr14.pdf. Acesso em: 1 out. 2024.

No domínio da segurança pública, observamos a importância mediadora dos atores humanos. Jean-Yves Dufour (2012, p. 47-48), ao escrever sobre a implementação dos sistemas inteligentes de videovigilância nas redes de transporte público na França, aposta na irredutibilidade da faculdade humana de julgar uma situação detectada como suspeita pela máquina.

Essa comunicação com os atores humanos é representativa de uma abordagem na qual o captor é utilizado para detectar uma situação de risco potencial, a decisão e os meios de intervenção cabem ao homem, único julgado competente para avaliar finamente a situação real sobre a qual o sistema automático oferece, muito eficazmente, indícios. As situações críticas e as funções identificadas quando da análise da necessidade do utilizador são a detecção de queda sobre o solo (mal-estar

ou agressão), a detecção de briga, o monitoramento de um passageiro na rede (útil, por exemplo, para interpellar um indivíduo que tenha cometido um delito).

Se a avaliação *final* das imagens de videovigilância fica sempre a cargo de atores humanos, é preciso sublinhar uma camada mais sutil de antecipação que permanece, mesmo no domínio securitário, delegada às máquinas inteligentes. O artigo intitulado “*Pre-Emptive camera activation for Video-Surveillance*” apresenta “uma nova técnica de visualização da informação que visa reduzir o esforço mental dos operadores de segurança” (Martinel; Micheloni; Piciarelli, 2011, p. 189). A partir do diagnóstico da desumanidade do esforço requerido para vigiar um objeto ou um indivíduo em movimento através dos espaços decompostos em diversos campos de visão, os pesquisadores desenvolveram um sistema automático de videovigilância.

Em vez de apresentar as imagens de todas as câmeras disponíveis, são as anomalias dos fluxos que serão exibidas aos agentes de segurança. A fim de determinar tais fluxos, o sistema deve prever as trajetórias regulares dos indivíduos com as câmeras que os gravam (Martinel; Micheloni; Piciarelli, 2011). A capacidade de antecipação nesse sistema não é orientada exclusivamente no sentido de uma análise automatizada das ações efetivamente realizadas diante das câmeras. Trata-se mais de uma estratégia ligada a um alto grau de normalização das condutas, que indica aos operadores do sistema os deslocamentos suspeitos sobre o espaço vigiado. O sistema se dirige àquilo que não está previsto pela arquitetura de um tal espaço.

Dois imagens (Figuras 9 e 10) apresentadas no artigo citado ilustram bem o tipo de lógica antecipadora em operação.

Na Figura 9, vemos as duas camadas que integram esse sistema visual de videovigilância. A primeira figura é a imagem de um estacionamento tal como ele é capturado pelo vídeo em tempo real, ao passo que a segunda e a terceira imagens (na horizontal, da esquerda para a direita) correspondem aos diagramas que confrontam algoritmicamente a imagem registrada pelas câmeras de vídeo, sinalizando eventuais desvios dignos de uma atenção suplementar. No estacionamento sob vigilância, o pedestre que se encontra no ponto C1 da imagem tende a se deslocar em direção a C2, C3 e C4, e cada uma dessas possibilidades é estimada por um algoritmo. Na segunda linha da Figura 9, vemos o procedimento análogo de estimativa da trajetória e cálculo de probabilidade para um veículo que entra no mesmo estacionamento. Como os deslocamentos de pedestres e veículos seguem diferentes protocolos (o espaço pelo qual uma pessoa pode passar não permite necessariamente a passagem de um veículo), somos confrontados a dois diagramas de monitoramento e predição de movimento. Eles orientam a atenção dos agentes de segurança em função de qualquer desvio não previsto por tais parâmetros.

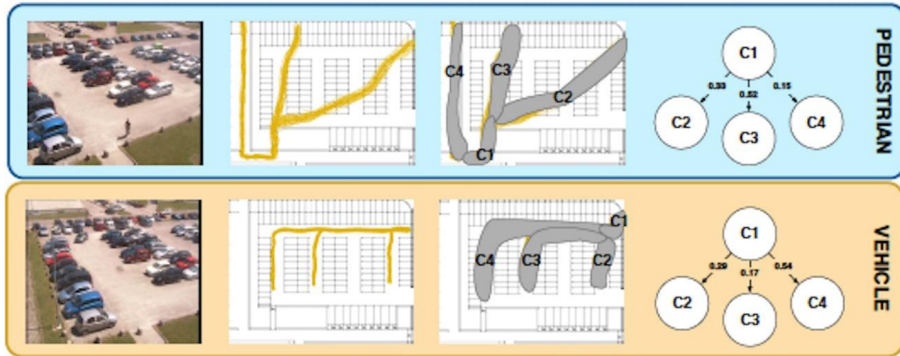


Figura 9: Exemplo de associação da trajetória e a probabilidade do caminho futuro

Fonte: Martinel, Micheloni e Piciarelli (2011)

A Figura 10 apresenta o modo como a antecipação das trajetórias dos indivíduos vigiados por esse circuito pode servir ao desenvolvimento de uma interface mais fácil de utilizar pelo agente de segurança. Os pesquisadores tiveram que resolver o problema da automatização da seleção das imagens a serem exibidas em um circuito de videovigilância de grande extensão, contando com diversas câmeras. Nessa segunda imagem, vemos o diagrama do deslocamento de um objeto ou de um indivíduo móvel atravessando um campo de visão abarcado por diversas câmeras. Em um dado momento da trajetória existe a possibilidade de uma bifurcação na qual um dos caminhos possíveis é previsto como o mais provável pelo sistema, o que faz com que a câmera 3 tenha prioridade em relação à câmera 14. No campo inferior direito do diagrama, vemos a interface tal como visualizada pelo operador, em que as imagens apresentam onde se encontra o objeto em tempo real e onde ele *provavelmente* estará em breve. Esse dispositivo nutre um estado de atenção e de espera do operador do sistema que também focaliza o caminho que pode, alternativamente, ser percorrido pelo indivíduo vigiado.

Retornando à questão da gestualidade, observamos que a recodificação da gestualidade e dos movimentos dos corpos é central para a automatização da previsão. A animação e a vitalidade do gesto, ressaltadas por Mlle. Dyvrande, não se voltam mais para o desvelamento da particularidade e dos tiques do indivíduo, mas para o reconhecimento automatizado, em larga escala, de padrões de conduta que não mais informam, com o auxílio da máquina, o olhar de um homem sobre outro. Agora, tais padrões devem ser detectados diretamente pela visão maquínica e na ausência de qualquer conhecimento sobre o indivíduo em questão.

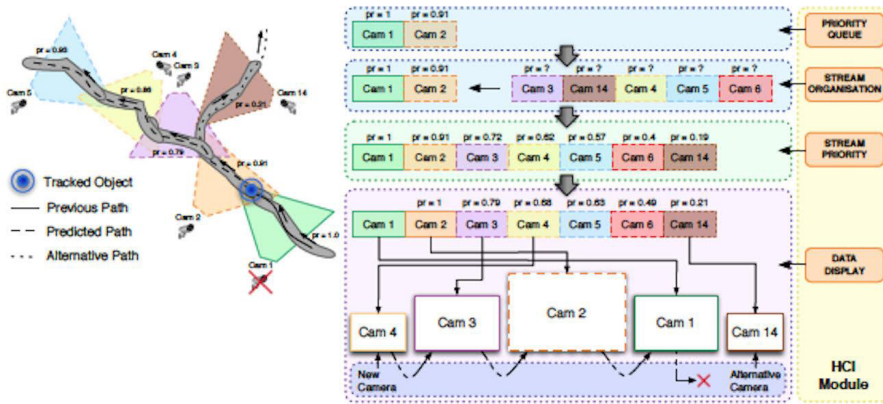


Figura 10: Módulo de interação humano-computador, exemplo de organização de fluxo, ativação e visualização de dados

Fonte: Martinel, Micheloni e Piciarelli (2011)

A abstração do gesto é um pré-requisito para a antecipação almejada pelas máquinas inteligentes. Os modelos preditivos são herdeiros diretos das técnicas de abstração do gesto que priorizam a análise de condutas, comportamentos e trajetórias. Vimos como essas técnicas foram inicialmente desenvolvidas na psiquiatria e, acima de tudo, na indústria. Na cibernética, a análise e a previsão de condutas se dissociam de qualquer referência aos estados internos do agente (humano ou não humano) que realiza a ação. Nas máquinas preditivas já descritas neste artigo, o reconhecimento como uma ação de segurança preventiva (geralmente destinada a surpreender e banir “infratores reincidentes” das ruas) não é mais direcionado a um determinado indivíduo previamente conhecido. A detecção (e o controle) da ação torna-se mais relevante que a identificação, dispensando-se o conhecimento prévio das motivações ou traços psicológicos que subjazem ao comportamento ou ao gesto.

A ênfase na atividade e na conduta não pode ser separada da racionalidade maquínica que a acompanha e tampouco da escala que ela pretende abarcar. Na linguagem técnica dos desenvolvedores de algoritmos preditivos no campo da segurança, a visão computacional é especialmente programada para a dupla função de reconhecimento de padrões e detecção de anomalias. Funções que máquinas podem operar em escalas massivas, de modo ininterrupto, veloz e relativamente eficaz. Além disso, tais funções alinham-se com os ideais preditivos e seus princípios atuariais, baseados em modelos matemáticos voltados para a antecipação sobretudo de ocorrências e incidentes que não estão vinculados a indivíduos previamente identificados. Na visão algorítmica e preditiva, o realce da gestualidade, da conduta e

do movimento dos corpos está relacionado com um relativo apagamento da identidade e da subjetividade. Não se trata mais de reconhecer ou diagnosticar a personalidade criminosa sob a gestualidade e os trejeitos corporais capturados na imagem, mas de antever, na superfície da imagem e no movimento dos corpos, um possível crime ou incidente por vir, detectando comportamentos ou situações anômalas. O indivíduo visado pelos sistemas de videovigilância inteligentes é o resultado de cálculos nos quais ele não consta como um elemento *a priori*. Ele emerge de uma zona pixelada anômala ou atípica na imanência da imagem.

Essas máquinas preditivas continuam e sofisticam os princípios dos dispositivos antimísseis criados por Wiener e Bigelow, ampliando sua margem de automação, a ponto de o piloto sumir. No cenário militar contemporâneo, os drones – veículos aéreos não tripulados – são a aplicação limite dos princípios preditivos automatizados, que já estão presentes na videovigilância algorítmica. Chamayou (2013, p. 64) mostra como o paradigma da inteligência baseada na atividade reside no coração da recente doutrina “contra insurrecional”:

Contrariamente ao que poderíamos pensar, o objetivo principal desses dispositivos de vigilância persistente é menos monitorar os indivíduos já conhecidos do que ver emergirem os elementos suspeitos sinalizando-se através de comportamentos anômicos. Porque esse modelo de informação é “fundado sobre a atividade”, ou seja, sobre uma análise das condutas mais do que sobre o reconhecimento de identidades nominais, ele pretende paradoxalmente “identificar” os indivíduos que permanecem anônimos, ou seja, qualificá-los pela tipicidade de seus comportamentos como emergindo de um perfil determinado: identificação não mais singular, mas genérica.

A ênfase na atividade também é, aqui, acompanhada do princípio de detecção de anomalias e de antecipação preventiva (Chamayou, 2013). Na própria linguagem militar, trata-se de monitorar e compreender “padrões de vida” e seus desvios, de modo a antecipar ameaças potenciais. Os padrões de vida aprendidos e reconhecidos pelos drones derivam do cruzamento de imagens e de uma série de dados e metadados extraídos de trajetórias, comunicações e ações de massas de indivíduos que permanecem anônimos até se tornarem um alvo privilegiado no campo de visão das máquinas preditivas. Vale lembrar que essa modalidade de olhar preditivo e de detecção de padrões e anomalias não é uma exclusividade do campo policial ou da guerra. Ela está presente de modo relativamente similar no rastreamento cotidiano de nossas ações online e em todo o comércio de dados e serviços a ele atrelado, nos modelos de gestão da chamada cidade inteligente,

nas oscilações do mercado financeiro, na circulação de notícias, nas campanhas políticas etc. O rastreamento, o arquivo e a antecipação dos gestos e atividades cotidianas assumem escalas vertiginosas e níveis cada vez mais altos de abstração: grafos, curvas, ondas, histogramas, vetores, nós e arestas pouco legíveis aos olhos humanos tornam-se a imagem contemporânea dos rastros e dados extraídos dos gestos e movimentos de populações conectadas a dispositivos informacionais. Nela não vemos mais os contornos do indivíduo e sua identidade, mas projeções de ações e comportamentos que são, ao mesmo tempo, alvos potenciais de intervenção. O alvo – termo caro tanto ao setor militar quanto ao publicitário – é sobretudo o efeito e não o referente dessa visão preditiva.

Essa expropriação e abstração da gestualidade aponta para uma possibilidade sem precedentes de convergência do mundo visível com bancos de dados de outra natureza, que integram os regimes contemporâneos de vigilância. Esses regimes não estão mais restritos exclusivamente ao domínio óptico, como Michel Foucault propôs em sua leitura do *panopticon* de Bentham.

Karl Palmãs (2011) evoca a noção de *Panspectric Surveillance* para analisar a transição do modo de produção industrial para o cenário pós-industrial, que seria marcado pelo surgimento de novas formas de organização corporativa, cujas encarnações seriam Google, Amazon, Wal-Mart e assim por diante. No seio dessa transição, propõe Palmãs (2011, p. 347), está a entrada de tecnologias de mineração de dados nos campos comercial e de marketing.

Em suma, a mineração de dados tornou-se imprescindível para muitas empresas, e um best-seller recente na área de negócios intitula-se “*Competing on Analytics: the new science of winning*”. Os autores argumentam que, usando o tipo de algoritmos de autoaprendizagem que o Google desenvolveu, as empresas de vários setores podem fazer previsões sobre o comportamento futuro dos usuários. [...] os dados coletados estudam os sujeitos não como indivíduos, mas como padrões e propensões de comportamento extraídos de vastos bancos de dados.

A crescente complexidade dos sistemas de vigilância preditivos pan-espectrais, estruturados em torno de parâmetros e dados heterogêneos, talvez seja um indicativo da intencionalidade não subjetiva subjacente aos processos de abstração de gestos que acabamos de descrever. Seja para intervir antes que um acidente ou assalto ocorram em uma estação de metrô, ou para prever a propensão dos consumidores a comprar quando eles já estão na fila para pagar suas compras, a observação de gestos e movimentos corporais integra os cálculos algorítmicos do futuro próximo. A visão maquínica dos gestos não é um capítulo à parte nos sistemas de previsão baseados

em dados. Os gestos e os movimentos corporais são abstraídos na forma de dados e expropriados dos corpos que os executaram para serem reimplantados nos cálculos multiparamétricos em vias de modelizar o futuro.

O estado da arte das máquinas preditivas contemporâneas nos permite compreender o escopo dos investimentos tecnopolíticos na gestualidade. Esses investimentos concernem não apenas ao campo da segurança pública, mas também às operações militares, às estratégias de marketing, ao design e às interfaces dos dispositivos de comunicação. Vimos, neste artigo, que a imagem técnica, de diferentes maneiras e em diferentes momentos, pretende ser capaz de capturar ou antecipar gestos, revelando suas dinâmicas, ritmos, trajetórias, intenções ou desenvolvimentos futuros. Em cada um desses casos, foi possível observar as mudanças na relação entre visão, máquina, corpo e tempo.

Na breve genealogia que propusemos, o processo de abstração e expropriação do gesto que começou na modernidade chega ao nosso tempo como antecipação de ações futuras por meio da detecção automatizada de comportamentos suspeitos. A contrapartida política desse processo é o surgimento de um novo estado policial, no qual a antecipação de gestos precede a identidade – e pode prescindir dela.

O gesto se torna cada vez mais abstrato à medida que a imagem técnica se afasta da dimensão humana e de uma referência prévia a um indivíduo identificado. Vimos que, num primeiro momento, a decomposição do gesto pela imagem técnica está ligada a mecanismos de controle nos quais a identificação ocupa um lugar central. Tanto na fotografia quanto no cinema, a visão maquínica dos gestos decompostos, abstraídos e expropriados dos corpos que os executavam, facilitou o reconhecimento dos indivíduos. Entre os exemplos mencionados neste artigo, as únicas exceções modernas são as ciclografias de Frank e Lillian Gilbreth, destinadas a otimizar o trabalho industrial no contexto do taylorismo. Nesse caso, a abstração do gesto registrado por uma cinemática de linhas luminosas acabava por apagar os traços singulares do indivíduo.

Esse apagamento da identidade aumenta progressivamente à medida que a visão preditiva da máquina se torna mais forte. Com o desenvolvimento da cibernética, a figura humana, ainda central na escala de apreensão do gesto, se tornará mais fraca. As técnicas de combate aéreo durante a Segunda Guerra Mundial já haviam tornado impossível avaliar as intenções dos atores humanos. A cibernética, portanto, propõe uma ontologia do inimigo (Galison, 1994) e uma forma de prever o comportamento que não está mais ligada à leitura do corpo humano e de sua subjetividade. O processamento algorítmico da gestualidade talvez seja o mais recente estágio desse processo de abstração e afastamento da escala humana. Na última parte

deste artigo, apresentamos várias aplicações de detecção automatizada de gestos e movimentos, destacando aquelas voltadas para a previsão de comportamento. Em cada uma delas, vemos que a visão algorítmica se volta para o reconhecimento de anomalias e padrões gestuais que não estão atrelados a procedimentos de identificação, com base em parâmetros de direção, densidade, textura e ritmo das imagens técnicas analisadas pelos chamados dispositivos “inteligentes”. A previsão automatizada de comportamento agora assume formas que só podem se referir a um indivíduo identificado *a posteriori*. Essa ausência relativa do indivíduo e da identidade não diminui de forma alguma o controle sobre as ações e a conduta daqueles que se tornam os principais alvos dessas máquinas de visão. Resta compreender quais são as possibilidades de subversão desse tipo de controle que, paradoxalmente, se alimenta de nossas condutas e intervém sobre elas, alegando não saber quem somos.

Considerações finais

Nos últimos anos, a detecção de violência (ou, de maneira mais precisa, brigas) atraiu cada vez mais interesse e investimento no campo da visão computacional. Já existem alguns aplicativos disponíveis no mercado e dezenas de outros são desenvolvidos e testados anualmente por empresas e universidades. A detecção ideal deve operar rápido e com imagens de baixa resolução. Curiosamente, um bom campo de provas para esses sistemas, longe de serem as guerras civis do Oriente Médio ou os estádios de futebol, tem sido o cinema (Gracia *et al.*, 2015). Já ocorrem, inclusive, competições entre jovens desenvolvedores, como o *MediaEval Affect Task*, iniciado em 2011. Na edição de 2012, por exemplo, os sistemas de detecção de ações violentas deveriam examinar automaticamente 18 filmes (15 para treinamento, cujo espectro ia de *Armagedon* a *O Mágico de Oz*; e 3 para a competição: *Sociedade dos poetas mortos*, *Clube da Luta* e *Independence Day*). O julgamento dos competidores levava fundamentalmente em consideração falsos positivos (identificação equivocada de uma briga) ou falsos negativos (uma cena violenta que passava incólume pela detecção). Há basicamente dois desafios a superar: discernir ações violentas de não violentas; e, alertar rapidamente quando uma ação violenta está em curso. De modo geral, quanto mais rápido for um sistema de monitoramento e alerta, menos acuidade tem. Quando se usam cores e sons, por exemplo (ocorrência de sangue – vermelho – ou de gritos e estampidos), o tempo de resposta tende a se tornar mais lento. Por outro lado, uma solução pode mostrar-se muito rápida no reconhecimento de ações violentas (apenas 50 frames, cerca de 2”), mas confundir-se com a vegetação movendo-se bruscamente com o vento.

Todos sabemos que há enormes diferenças entre cenas violentas no cinema e brigas no cotidiano das ruas ou no interior de instituições. O sangue, por exemplo, nunca jorra tão rápido ou se faz imediatamente perceptível. Explosões e tiros, por sua vez, são menos frequentes do que os espectadores supõem, principalmente no início dos conflitos. No *MediaEval Affect Task*, os algoritmos das máquinas preditivas pagam seu tributo ao cinema, cujos procedimentos técnicos e estéticos foram decisivos para a elaboração dos processos de abstração dos gestos que estão na sua origem. Porém, essas competições deixam algumas perguntas sem resposta: que capacidade teria o mais acurado e veloz sistema de detecção de brigas para distinguir visualmente a briga real de uma simulada? O sangue de uma vítima de um banho de ketchup? O som de um disparo verdadeiro de um tiro de festim? O uso do cinema como campo de provas para máquinas preditivas nos fornece um pungente testemunho de quanto nossas vidas foram colonizadas pela imagem – e quanto nossos corpos, gestos e movimentos estão sujeitos ao confinamento que ela induz. Mas, por outro lado, deixa entrever o calcanhar de Aquiles desses dispositivos: seu mais verdadeiro ponto cego é a fabulação.

Referências

- ADORNO, T. *Minima Moralia*. London: Verso, 2005.
- AGAMBEN, G. *Means without end*. Minneapolis: University of Minnesota Press, 2000.
- AGAMBEN, G. . *What is an Apparatus? and other essays*. Stanford: Stanford University Press, 2009.
- BENJAMIN, W. *Obras Escolhidas I*. São Paulo: Brasiliense, 1985.
- BENJAMIN, W. *Selected Writings*. Cambridge: Belknap Press, 2003.
- BERMEJO, B. *et al*. Violence detection in video using computer vision techniques, INTERNATIONAL CONGRESS ON COMPUTER ANALYSIS OF IMAGES AND PATTERNS, 14., Berlin, 2011. *Anais [...]*. Berlin: Springer, 2011. p. 332-339.
- BERNARD, H. L'Education des agents par le cinéma. *Le Film* (Paris), n. 14, p. 7-8, 1914.
- BLUNSDEN, S. J.; FISHER, R. B. Pre-fight detection: Classification of Fighting Situations Using Hierarchical AdBoost. INTERNATIONAL CONFERENCE ON COMPUTER VISION THEORY AND APPLICATIONS, 4., Lisboa, 2009. *Anais [...]*. Lisboa: VISAPP, 2009.
- CHAMAYOU, G. *Théorie du drone*. Paris: La Fabrique, 2013.

- DIDI-HUBERMAN, G. *Invention of Hysteria*. Cambridge: MIT Press, 2003.
- DUFOUR, JY. (org.) *Outils d'analyse vidéo: pour une pleine exploitation des données de vidéo-protection*. Cachan: Lavoisier, 2012.
- ESEN, E.; ARABACI, M. A.; SOYSAL, M. Fight detection in surveillance videos. INTERNATIONAL WORKSHOP ON CONTENT-BASED MULTIMEDIA INDEXING, 11., Veszprem, 2013. *Anais [...]*. Piscataway: IEEE, 2013. p. 131-135.
- GALISON, P. "The ontology of the enemy: Norbert Wiener and the cybernetic vision", *Critical Inquiry*, Chicago, v. 21, n. 1, p. 228-266, 1994.
- GIDEON, S. *Mechanization takes command*. New York: Oxford University Press, 1948.
- GORDON, M. *Stanislavsky in America*. London: Routledge, 2010.
- GRACIA, I. S. *et al.* "Fast Fight Detection". *PLoS ONE*, San Francisco, v. 10, n. 4, p. e0120448, 2015. <https://doi.org/10.1371/journal.pone.0120448>
- GÓMEZ-IÑIGUEZ, C.; CARBONELLI-VAYÁ, E. J.; RIO, B. M. "Análisis conceptual del término vigilancia desde principios del s.XX hasta la actualidad: Una perspectiva histórica", *Revista de historia de la psicología*, [s. l.], v. 20, n. 3-4, p. 415-428, 1999.
- HOUBLER, D.; HOUBLER, T. *Os Crimes de Paris*. O roubo da Mona Lisa e o nascimento da criminologia moderna. São Paulo: Três estrelas, 2013.
- LERICH, L. *La Police Scientifique*. Paris: Presses Universitaires de France, 1949.
- MARTINEL, N.; MICHELONI, C.; PICIARELLI C. "Pre-emptive Camera Activation for Video-Surveillance HCI". In: MAINO, G; FORESTI, G. L. (eds.). *Image Analysis and Processing-ICIAP 2011*. ICIAP 2011. *Lecture Notes in Computer Science*. Berlin-Heidelberg: Springer, 2011.
- MOAR, J. *Gesture, Motion & Haptics: Future Use Cases, Shipments & Revenue Forecasts 2016-2021*. *Juniper*, 2016.
- OUBIÑA, D. *Una Juguetería filosófica*. Buenos Aires: Manantial, 2009.
- PALMÁS, K. "Predicting What You'll Do Tomorrow: Panspectric Surveillance and the contemporary Corporation", *Surveillance & Society*, Chapel Hill, v. 8, n. 3, p. 338-354, 2011.
- PANOFSKY, E. *Studies in iconology: humanistic themes in the art of the renaissance*. New York: Harper and Row, 1972.

PINNEY, C. *Photography and Anthropology*. London: Reaktion Books, 2011.

PREDICTIONS WHILE YOU WAIT. London: British Pathé, 1939. Disponível em: <https://www.britishpathe.com/video/predictions-while-you-wait>. Acesso em: 29 ago. 2024.

PREMIER CONGRÈS DE POLICE JUDICIAIRE INTERNATIONALE (Mônaco). *Actes de Congrès*. Paris: Marchall et Billard, 1926.

PRICE, M. *The Photograph: a strange, confined space*. Stanford: Stanford University Press, 1994.

SHAO, J.; CHANGE LOY, C.; WANG, X. "Scene-independent group profiling in crowd". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, [S. l.], p. 2219-2226, 2014.

WHYMAN, R. *The Stanislavsky system of acting: legacy and influence in modern performance*. Cambridge: Cambridge University Press, 2008.

WIENER, N. *The Human use of human beings: cybernetics and society*. Avon: Bookcraft, 1989.

submetido em: 31 ago. 2024 | aprovado em: 09 set. 2024