ESTUDIS ESCÈNICS. núm. 50. 2025

Transformations in Image Extraction Regimes: from Marey's Photography to the Latent Vector

ESTAMPA

Roc Albalat. ORCID ID: 0009-0008-9209-795X. BAU, Barcelona, Spain.
Pau Artigas
Marc Padró. UB, Barcelona, Spain
Marcel Pié. UB, Barcelona, Spain
Daniel Pitarch. ORCID ID: 0000-0002-2406-2682. Escola Massana, Barcelona, Spain.
estampa@tallerestampa.com

BIOGRAPHICAL NOTE: Estampa is an artistic collective of programmers, filmmakers and researchers established in Barcelona. Their practice is based on a critical and archaeological approach to audiovisual and digital technologies, with a special interest in the archive and the experimental audiovisual. Since 2017, one of the main lines of work has been the research on the uses and ideologies of AI. Their projects can be consulted on the collective's website: tallerestampa.com.

English translation, Neil CHARLTON.

Abstract

This article explores the operational status of contemporary images linked to machine learning, revealing their historical continuity with data extraction methods developed in 19th-century scientific photography. The analysis begins with the experiments by Gilles de la Tourette and Étienne-Jules Marey, who transformed bodies into diagrams, eliminating the figurative to isolate patterns of movement and behaviour. These techniques, designed to transcend human perception and access "objective" knowledge, established a paradigm where the image functions as a surface for data inscription. The study demonstrates how this paradigm has reached its maximum expression in contemporary artificial vision systems, where images have ceased to be fundamentally objects of perception and become instruments of automated extraction. Unlike 19th-century photographic records — explicit in their technical mediation —, data registration now operates in opaque computational layers, shifting from the phenomenic to the statistical plane. Through the analysis of latent spaces in deep learning models, we show how images are reduced to mathematical vectors that encode both visual information and power structures. This transformation operates on three levels: technical (data compression into dimensions that are not humanly legible), epistemological (knowledge production through mass correlations), and political (naturalisation of social categorisations through seemingly neutral interfaces). The research concludes that this regime of post-retinal visuality

Reception: 24/4/2025 | Acceptance: 3/7/2025

inverts the traditional logic: the visible becomes an accessory residue, while real agency occurs in the calculation of the imperceptible.

Keywords: operational imaging, machine learning, artificial vision, latent space, data extraction, scientific photography

Reception: 24/4/2025 | Acceptance: 3/7/2025

ESTAMPA

Transformations in Image Extraction Regimes: from Marey's Photography to the Latent Vector

In 1886, the French neurologist Gilles de la Tourette published his analysis of human gait using graphic recordings. His method was as follows: he placed a long, white roll of paper on the floor and drew a longitudinal line indicating the direction his patients should follow. He had previously coated the soles of their feet with powdered iron sesquioxide, dyeing them red and allowing the footprints to be recorded as they walked along the canvas. The resulting image allowed Gilles de la Tourette to accurately measure human gait, recording step length, lateral deviation, or angle of inclination (fig. 1). The meticulous descriptions he wrote about each of these actions convey the scientific determination of the time to document, measure and calculate all types of observations, especially those that sought to capture human gesture,

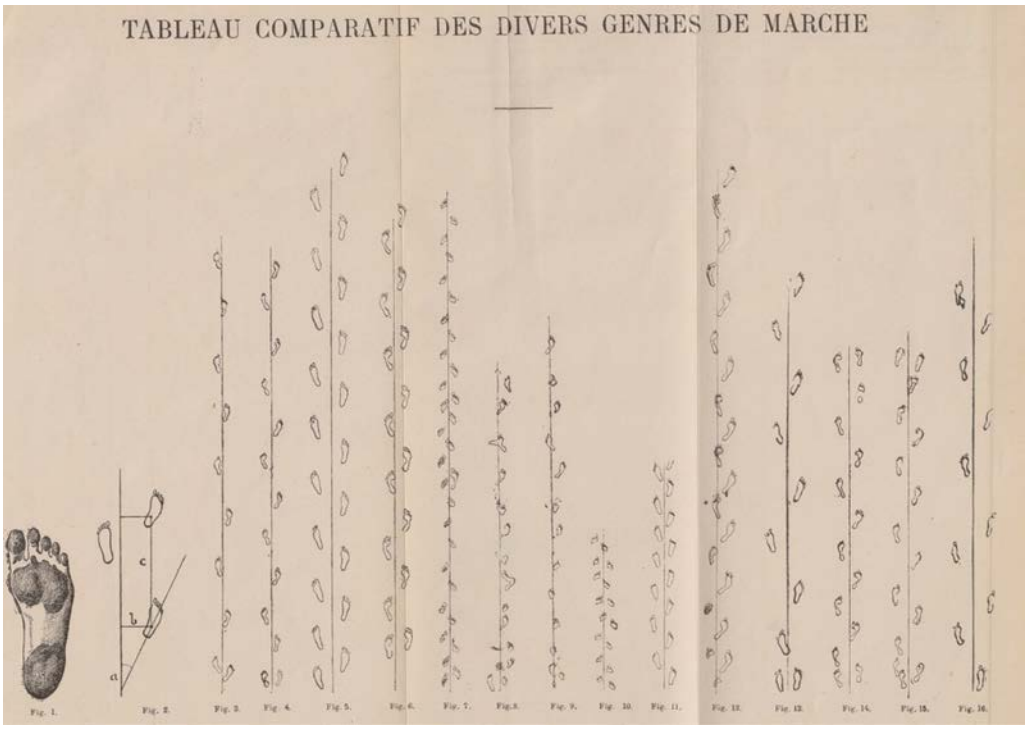

Figure 1. Études *cliniques et physiologiques sur la marche* (Tourette, 1886).

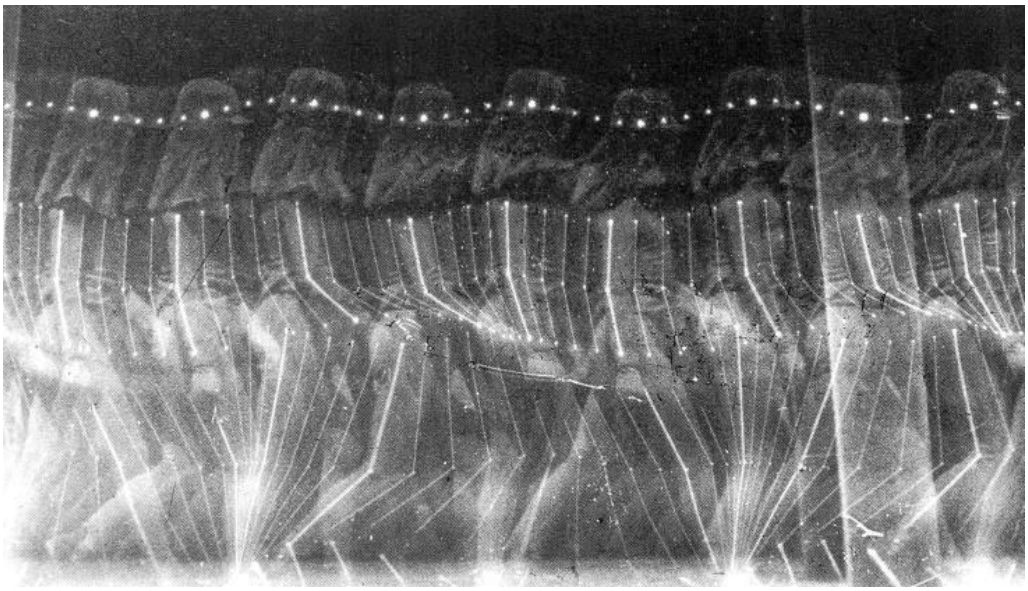

Figure 2. Chronophotograph *Walking Man*, by E. J. Marey, 1882.

movement, and behaviour. While this gait study was being conducted, the French physiologist Étienne-Jules Marey was using another method of graphic synthesis based on photography. He realised that, by superimposing prints, the phases of movement could be visualised and analysed, thus studying the relationships between them in a single frame. However, the human figures depicted on Jules Marey's photographic plate were too detailed; his object of study meant he had to distance himself from the content of the image to achieve a greater degree of infographic abstraction. He achieved this, in a demonstration of his inventiveness, by rendering the human figure invisible and capturing only a geometric synthesis of the movement of bodies (fig. 2). For this purpose, he designed a black suit that completely covered his model's body and printed the lines and dots in white so they would be recorded and contrasted in the photograph (fig. 3). In this way, the figure blended into the background, and the marks of his productive instance were recorded on the invisible body.

Figure 3. Illustration of the volume *Le mouvement*, by E. J. Marey, 1894.

Host images

While it is difficult today not to perceive a certain beauty in the visualisation systems created by Jules Marey, he had no intention of endowing them with aesthetic qualities. Images had an instrumental function, based on their potential for analysis and data extraction. Today, his obfuscation of photographic details, as well as the primacy of annotation, take on new meanings, as these image detection processes have become automated and widespread. As digital images tend to become part of new networks of online social communication, automatic annotation is becoming more prominent. In each new online publication, operations, invisible to the human eye, are activated that identify and label what they recognise in the images. These processes, embedded in the digital image, are today the reverse side of visual culture, a two-way agency in which whoever looks at the images is also looked at. In an analogy with other types of ecosystems, it could be said that, in a digital environment, images are hosts, because they house all kinds of parasites (operations) that depend on them to function. [1] In the extraction process applied to the images we share on social media, echoes of 19th-century photographic experiments resonate, but with one key difference: while in Jules Marey's images the photographic representation was rendered invisible, making the

Figure 4. *Fall Detection Based on Key Points of Human-Skeleton Using OpenPose* (Chen [*et al.*], 2020).

extractive objective evident, in today's online images what remains invisible is the data annotation. Digital images have added a layer of detections and operations that are invisible to the human eye. Perhaps some models have a graphic interface that allows us to better understand the recording processes that take place in that layer (fig. 4). This interface, however, is nothing more than an accessory visualisation, deferential to human vision. In the resulting images, the detection of bodies goes back to Jules Marey's diagrams, to such

1. This metaphor comes from the piece *Imágenes huéspedes* (Estampa, 2022, available at: ‹https://tallerestampa.com/es/estampa/imagenes-huespedes/›). On the processes of automatic annotation of images, see also the video essay *¿Qué es lo que ves, YOLO9000?* (Estampa, 2019, available at: ‹https://tallerestampa.com/es/estampa/que-es-el-que-veus-yolo9000/›).

an extent that, if we superimpose the representation of both operations, we can study them iconologically, understanding them as cultural mutations of the same extractive drive.

Designed to imitate human cognition, the automatic recognition of objects and bodies in images has begun to be understood and equated with a particular form of vision. American artist Trevor Paglen has studied the ubiquity of machine vision and its profound implications for visual culture. In his article "Invisible Images (Your Pictures Are Looking at You)", he poses the question this way: "Human visual culture has become a special case of vision, an exception to the rule. The overwhelming majority of images are now made by machines for other machines, with humans rarely in the loop" (Paglen, 2016). The image we see, the one signified before our eyes, is a secondary and temporary application of the digital image. It appears on the screen for a few seconds, as one of the many operations it performs outside our field of vision. Furthermore, the digital image does not require the human gaze to perform its function as an interface for detection, recording and transmission; rather, it operates in a loop that, most often, only involves communication between machines. From the field of photographic theory, it had already been noted some time ago that the image was mutating in a direction beyond human perception. The historian John Tagg (2009: 24) writes: "The visual is also emptied of any content of palpable sensation, of stimulus and response, of the body as a surface in touch with the world. There is nothing being imaged that can touch the inner surface of the eye. There is nothing to be seen. The process of seeing is no more than a hook-up: a means to plug an organ into a circuit." Tagg suggests the example of the licence plate reading systems that control access to central London.

The system works by using images captured by the nearly seven hundred cameras located at the entry points to the restricted area (fig. 5). They record and recognise vehicle licence plates and send them to verify the fee the drivers should have paid (fig. 6). What strikes Tagg about the London traffic control system is that the entire cycle of capturing, reading, transmitting, and correlating data, and imposing sanctions, is automated, involving only machines. And, in this image-dependent circuit, the shift toward visibility almost never happens: "[It] does not involve anything that might be

Figure 5. *CCTV Camera in London*.
Wikimedia Commons, 2007. Public domain.

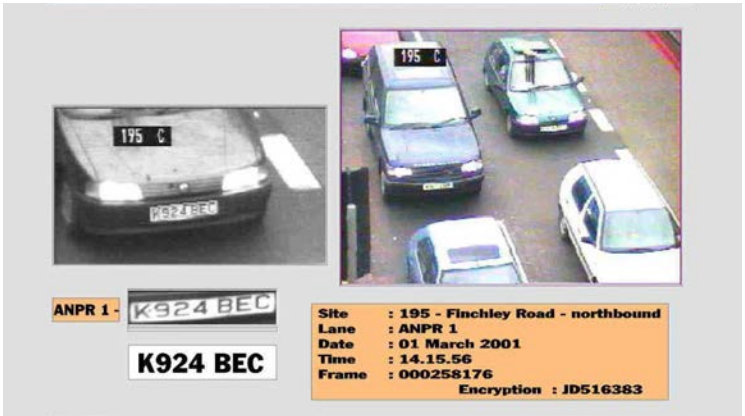

Figure 6. *ARS (Automotive Recall Statistics) – Executive Summary*, by the European Commission – Directorate-General for Mobility and Transport, 2010.

seen as entailing communication, psychic investment, a subject, or even a bodily organ" (Tagg, 2009: 21). Issues concerning the digital image shift to extra-retinal matters, and the toolbox for understanding it — representation, mimesis, and iconographic motifs — are relegated to measurement, statistical modelling, and pattern analysis.

Filmmaker and writer Harun Farocki began using the term *operational images* even before the development of the networked image. Farocki's operational images were not necessarily representative, but rather functioned as part of an infrastructure. The audiovisual montage of the video installation trilogy *Eye / Machine* (2001) (fig. 7) emphasises the role of these images in military-technological development and various industrial applications, connecting regimes of visibility to the materialisation of institutional structures of command and control. Media theorist Jussi Parikka (2023: 11) highlights the way in which Farocki's montages have investigated "how contemporary images are intimately tied with the modern forms of industrial production, detaching from a history of images only as visual culture, to histories of chemistry, violence, labor, exploitation, and data." Within the context of the Gulf War (1991), Farocki uses the innovative shots of the camera located in the so-called "smart bombs", which replace the human eye as historical testimony of a war conflict, to warn of a new politics of images: "The Gulf Wars have changed the status of images, because images are no longer just a means for entertainment or education. They are *tools*" (Griffin, 2004). The necropolitical status of the image inaugurated by the Gulf War is related to the current institutional management of borders in the West, which depends on an ecology of observation, analysis and surveillance. Remote-controlled military imagery, integrated into the Israeli drones used by the EU (fig. 8 and 9), today fulfils the strategy of preventing migrants and asylum seekers from reaching Europe by boat from the coast of Libya by all possible means. Remote viewing provided by unmanned aerial vehicles allows law enforcement agencies to circumvent international law, which dictates that a ship's captain has a duty to assist people in distress, and enables a return circuit (Heller and Pezzani, 2021). The image here works as a prosthesis for a police view that acts by proxy, while simultaneously disabling the figure of a subject who can assume responsibility.

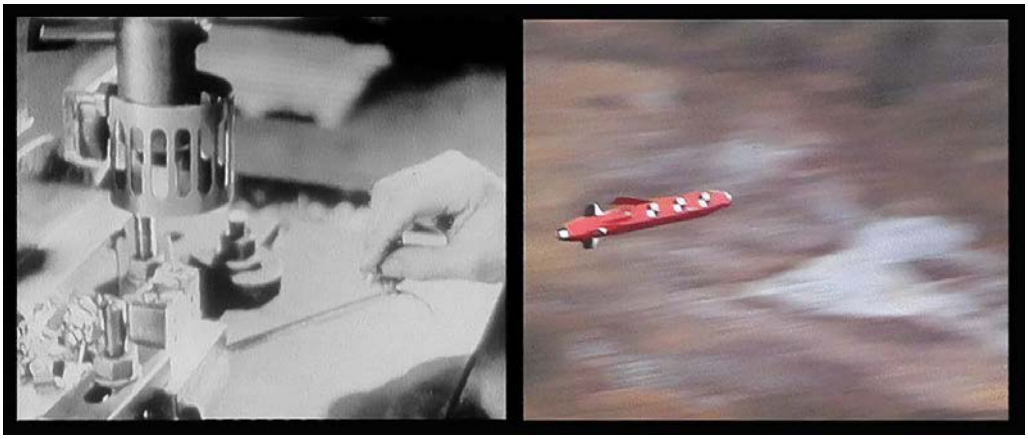

Figure 7. *Eye / Machine I*, by Harun Farocki, 2001.

Figure 8. *IAI Heron TP at ILA 2018*, by Bernd Gröning. Wikimedia Commons, 2018.

Figure 9. «*Unwitnessing the border*», by Border Criminologies Blog, by Giovanna Reder, 2024.

Resonance models

Today, images are becoming an automation tool through machine learning techniques. This automation is achieved not by programming the specific steps to be performed, but rather by using examples. If we have many annotated images, these tools are configured using statistical processes to replicate the annotation for new images. These examples for machine learning are what is called a dataset. The images in the dataset enable training that transforms the signifiers into data and, in turn, the data into statistical models. This training is based on tracking large-scale patterns and correlations, enabling tasks that previously could not be automated. However, the abstraction processes that occur in this circuit do not separate the data from the communities from which they were extracted. The translation of images into data codifies complex worldviews, reducing their meanings and subjecting them to Western taxonomies, categories, and archetypes — including racial and gender stereotypes —, which, from now on, in a world made of automation, will work in the opposite direction. As biometric-based AI applications have been implemented, concerns about their biases have grown: cases of racism and classism embedded in facial recognition systems, border management, job interviews, and prisoner reintegration (O'Neil, 2017). Faced with these cases, technology providers are taking advantage of the controversies and waves of protest to work on new updates, refining their algorithms with the promise of working on a bias-free model. What is key here, however, is questioning whether such a thing can exist (Crawford, 2021). On the contrary, it should be borne in mind that, in order for computers to see, these systems are built from a particular perspective. What the dataset reflects — its image archive, its categories — will always be an incomplete document, if not a document of barbarism. There will always be some value or underrepresented political identity. The ideology of big data assumes that everything can be converted and processed into data, but this is not ultimately true. Visual artist Mimi Ọnụọha's installation *The Library of Missing Datasets* (2016) focuses on the gaps that exist in our data-saturated world (fig. 10). Where are we not looking? the artist seems to ask. What is ignored by the data reveals our hidden social biases and indifferences: white

Figure 10. *The Library of Missing Datasets*, by Mimi Ọnụọha, 2016.

children adopted by people of colour, trans people murdered in hate crimes, people excluded from public housing due to criminal records. Faced with these non-existent datasets, we question the tensions present in the original archive and, therefore, the limits of AI. Any output from an AI tool should interrogate the dataset that created it, because it is still a reflection of its dataset, a resonance model. Viewed against the current, AI tools reflect the extractive utopias of the present, and it is difficult not to question the very desire for social relations to be more abstract, traceable, and machine-readable.

In his analysis of the Panopticon, philosopher Michel Foucault (1975: 200; trans. 1977) asserts that "visibility is a trap" that materialises in the form of architectures. Through the visual field, operational images capture communication flows, encoding their contents in order to train new models. In this process, input data is arranged in a mathematical space, a compressed, numerical representation of the image: the latent space. The latent space is a vector mapping of datasets (images, semantics, values) that extracts the most relevant features from the dataset in order to replicate them as faithfully as possible once the model has been trained. In data science, and in certain models, this intermediate stage is called the *bottleneck* (fig. 11).

What happens with images in this *bottleneck* is key to understanding the ulterior applications and automations. If in the field of the current operative images, the dataset determines the limits of what is recognisable (the visible), the latent space constitutes its operative syntaxis (the assignment of signifiers and values). Although it is a virtual space, the characteristics of the image are indeed latently inscribed. In this respect, after agreeing that today we must learn to read the images beyond their visible surface, it is also necessary to question the type of representations that take place beyond this surface, in that statistical space; understanding, in the first place, that it is a complex representation. Since it is a space that human eyes cannot perceive, its modes of calculating the images need another type of scrutiny, one that

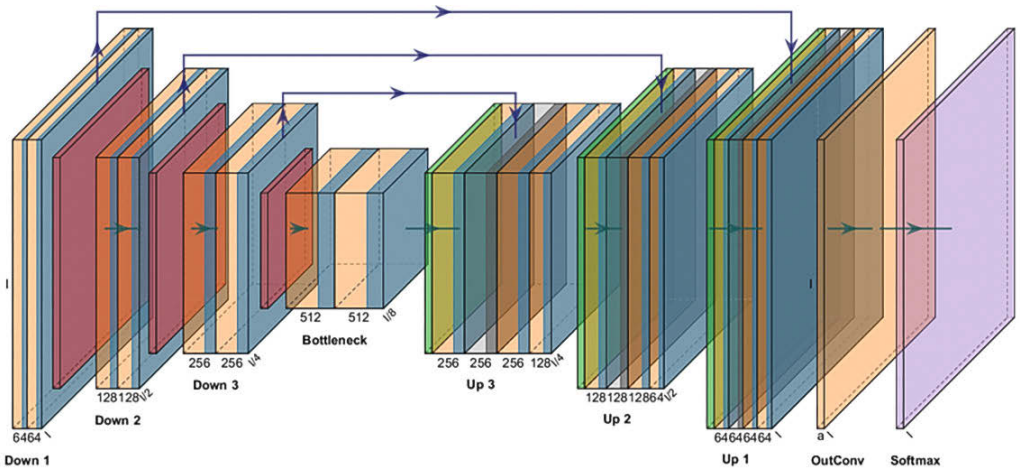

Figure 11. «Multi-label semantic segmentation of magnetic resonance images of the prostate gland» (Locherer [*et al.*], 2024).

concerns its phenomenology. After the *bottleneck*, in the return trip to the abstractions of what is material, new automated technical systems are set in motion: recursive loops, interferences and even decision-making processes, in such different applications as the systems of medical diagnosis, the identification of images via satellite, the ship's navigation systems, facial recognition, driverless cars or migration control in the EU. The question about the operational image should also include the models that depend on it: which infrastructure it unleashes, which laws it inscribes in the objects, which conducts it manages, and which logistics are being naturalised through instrumental visibility.

At this point, we can go back again to the images of Étienne-Jules Marey. To the synthesis of movement, to the recordings and annotations that illustrate the first extractive drive through photography (fig. 12). What we see in these images helps us understand the cultural archetypes that took us to the *bottleneck*, the gaze that constructs a technical interval that translates the *modes of seeing* into *modes of calculating*. The body appears as a

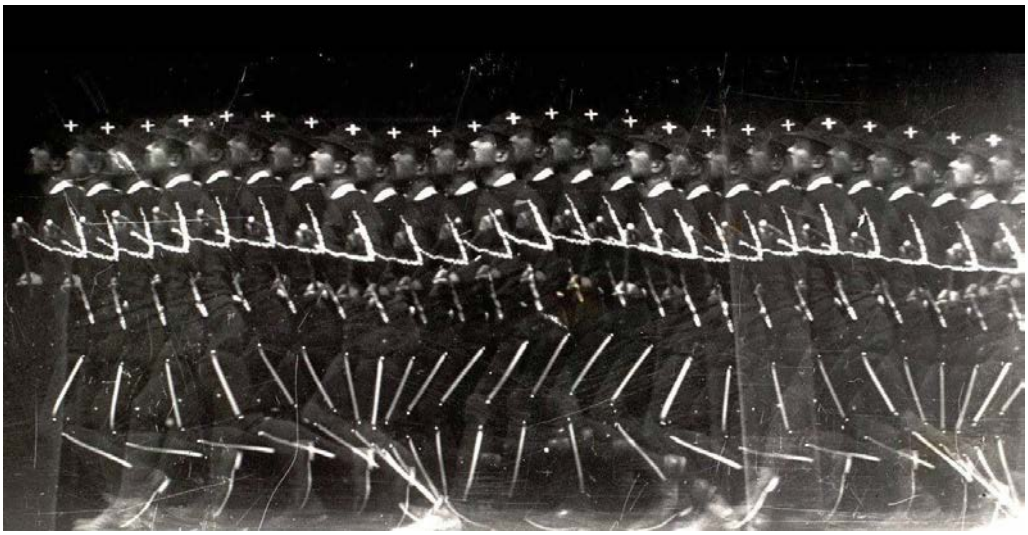

Figure 12. *The Running Lion Tamer*, by Étienne-Jules Marey, 1886.

representation of itself: Jules Marey synthesised it through schematic lines and, to do so, he had to print these marks straight on the clothing, on that body. The trace of the codes, their index, did not only remain on the surface of the image, but was also written on the bodies of the models (similarly as the dye impregnated the feet of Gilles de la Tourette's patients). Faced with that gesture, the question is whether he would not already be anticipating how the technological interpretation codes are inscribed on the bodies detected. In that case, the image takes us to the other side of the *bottleneck*, in which the said codes ultimately determine the spaces of action of the communities, individuals, organisms and environments from which they were extracted.

Conclusions

This article has drawn a critical genealogy of the image extraction regions, from the meticulous annotations of Étienne-Jules Marey – where the body was decomposed into legible traces, recorded on the photographic surface as scientific evidence – to the abstraction of the latent vectors in contemporary artificial vision systems. The move from Marey's traces to the Marey deep learning vectors marks a radical shift in the status of the image as a means of knowledge. In Marey's experiment, the abstraction operated through a controlled synthesis: the bodies were reduced to geometric lines, but these marks preserved an indexical relational with what was observed. The photographic trace, albeit stylised, related to a concrete reference in the physical world, enabling a two-fold reading: scientific (as metric data) and phenomenological (as an imprint of the body that existed before the camera).

In contrast, computer vision systems perform a radical abstraction: they compress the visual into latent spaces of high dimensionality, where image is no longer structured as representation but as statistical correlation. Consequently, in the contemporary visual regime, the images we perceive have lost the representational status to become territories of calculation: a surface of statistical operation and computational processes that operate in the background. In the framework of this new post-retinal regime, the images that appear on screens and interfaces are a phenomemic residue of technical operations that favour the supposed extraction of value (economic, cognitive, political). The visible, in this context, is the last link of a chain in which the decisive will always be elsewhere. This residual condition of the image poses a paradox: we have never produced more visual representations and, yet, the fights for their representation have never been so opaque. The cultural critique thus faces the challenge of tracking the political there where it is no longer shown: in the datasets that precede the image, in the vectoral spaces that cross it, and in the models that instrumentalise it.

Bibliographical references

Chen, Weiming; Jiang, Zijie; Guo, Hailin; Xiaoyang, Liu. "Fall Detection Based on Key Points of Human-Skeleton Using OpenPose". *Symmetry*, vol. 12, no. 5, 2020, art. 744. <https://www.mdpi.com/2073-8994/12/5/744> [Last accessed: 1 January 2025].

Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

Foucault, Michel. *Discipline and Punishment: The Birth of the Prison*. Translated from French by Alan Sheridan. New York: Vintage Books, 1977. [Orig. ed.: *Surveiller et punir*, 1975].

Griffin, Tim. "Viewfinder: A Conversation with Harun Farocki". *Artforum*, no. 43:3, 2004, pp. 162-163. <https://www.artforum.com/features/viewfinder-a-conversation-with-harun-farocki-170105/> [Last accessed: 1 January 2025].

Heller, Charles; Pezzani, Lorenzo. "Airborne Complicity: Frontex Aerial Surveillance Enables Abuse". *Border Forensics*, 2021. <https://www.borderforensics.org/investigations/airborne-complicity/> [Last accessed: 1 January 2025].

Marey, Étienne-Jules. *Le mouvement*. Paris: Gauthier-Villars, 1894.

Marey, Étienne-Jules. *Walking Man (Chronophotography)*, 1882. Photographic archive, public domain.

O'Neil, Cathy. *Weapons of Math Destruction*. London: Penguin Books, 2017.

Paglen, Trevor. «Invisible Images (Your Pictures Are Looking at You)». *The New Inquiry*, 2016. <https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/> [Last accessed: 1 January 2025].

Parikka, Jussi. *Operational Images: From the Visual to the Invisual*. Minneapolis: University of Minnesota Press, 2023.

Reder, Giovanna. "Unwitnessing the Border". *Border Criminologies Blog*, University of Oxford, 2024. <https://www.law.ox.ac.uk/research-subject-groups/centre-criminology/centreborder-criminologies/blog/2024/05/unwitnessing> [Last accessed: 1 January 2025].

Tagg, John. "Mindless Photography". In: Long, J. J.; Noble, Andrea; Welch, Edward (eds.). *Photography: Theoretical Snapshots*. London: Routledge, 2009, pp. 16-30.

Tourette, Georges Gilles de la. *Études cliniques et physiologiques sur la marche*. Paris: Rueff, 1886.