sigradi2024
biodigital_intelligent_systems

iBAG
UIC
BCN

Beyond Parallax: Neural Radiance Fields and Invisual Representation

Thomas Parker[1]

[1]University College London (UCL), London, United Kingdon

Thomas-parker@ucl.ac.uk

Abstract. With the emergence of Neural Radiance Fields (NeRFs) in 2020 there have been a series of ongoing innovations in computer vision with a particular focus on new forms of representational practice. Through these cutting-edge techniques, which draw from traditionally photogrammetric processes underpinned by artificial intelligence, our conventional optical methods of reconstruction and representation in metrology have begun to undergo radical evolution. These innovative methodologies are challenging how we understand the operationalization of the image and are ushering in an 'invisual' era, redefining how we generate and deploy images with non-conventional paradigms of parallax, and resulting in a post-lenticular form of architectural representation. This paper aims to explore this paradigm shift, probing the operational essence of architectural imagery crafted through multiple generative A.I. systems and this post-digital practice, whilst highlighting the myriad of opportunities within an expanding invisual domain beyond our known parallax.

Keywords: Digital strategies, Neural radiance fields (NeRF), Parallax, Architectural representation, Phygital interaction

1 Introduction

In 1867 Albrecht Meydenbauer set out to develop a form of measuring the built environment based on photographic imaging and optical parallax, which would go on to be known as photogrammetry (Jones, 2023). Little would he know that over 150 years later his pursuit would lay the groundwork for forms of deep neural network modelling and representation in the form of Neural Radiance Fields (NeRFs) (Mittal, 2023) (Rabby et al., 2023). In setting out these principles, Meydenbauer was one of the first practitioners to also set out the positional operation of the photographic image in measuring the built environment. In this shift towards the operationalization of the image, the photograph elevated itself above a solely representational piece of work. Instead, it had been translated into a vessel by which discrete data could be defined and communicated. As Jessi Parikka describes, *Operational images organize our world, but they are also organizing our sense and skills in terms of how we are trained to approach such images, from the photogrammetric mapping of landscapes to pattern recognition* (Parikka, 2023). With recent advances in generative A.I. image models, such as OpenAI's Sora (Brooks & Peebles, 2024) or Runway's Gen-3 Alpha Model (Germanidis, 2024), the operational nature of our images, and how we might deploy them to measure the world, has fundamentally shifted away from our prior understandings. As a result, there is a need to consider how a future, post-lenticular and 'invisual', representational logic may underpin a range of alternate visual media practices and methodologies.

1.1 Operational Images & Invisual Parallax

As new forms of generative A.I. text and image-to-video models begin to output consistent and temporally stable imagery, a foundational shift occurs in how metrological practices can be harnessed to process this complex and parallactical material. The photogrammetric and lenticular-based processes, which previously required careful calibration and a consistent overlapping image-based parallax to reconstruct accurately measured physical space, are fundamentally challenged by these cutting-edge A.I. models that can now simulate and represent the underlying spatial characteristics of three-dimensional scenes. In the internal processes of these large data models, we find processed-based logics that Parikka describes as, *the boundaries of signal processing, [that] linger on the horizon of the visible/invisible, visual/invisual* (Parikka, 2023). These undulating 'invisual' boundary processes, in which images are produced without being rendered visible to the human eye, change not simply how we understand imagery and imagining, but also the range of optical and lenticular methodologies they currently underpin, such as image-based photogrammetry.

Simultaneously, the expanded media definition of the 'invisual' image opens potentially new dialogues, as well as some challenging conundrums, when operationalized in further complex workflows. The physical act of capturing lenticular data, for measured reconstruction, and the internal workings of the A.I. models, that are trained to generate digital imagery, are at odds when defined by our existing terminologies. As Parikka goes on to define, *the pairing of the visual-invisual [is] to draw this continuum as a territory of transformations that concerns images and their role as aesthetic-epistemic agents* (Parikka, 2023). It is in this state of territorial transformation, that we can position invisual forms of representation as epistemological agents, allowing us to expand the currently limited definitions of these new types of generative images.

This paper sets out a position around these invisual representational practices, the role of operational images in the future of three-dimensional metrology (photogrammetry), and how early forms of generative A.I. images can become operationalized via alternative methodologies. The paper will outline a novel methodology, that takes advantage of recent developments in the consistency of generative A.I. video models, whilst also harnessing the depth training and pose estimation of NeRF's, to consider how data might be handled in the pursuit of new representational practices. Finally, in the discussion, we will question the speed of the image, the infallible nature of image-based measurement, and what consequences, opportunities and future practices may emerge, when the previously lenticular practices, become a post-lenticular form of invisual operation. Although the paper will refer to various scientific methodologies throughout, the discussion will predominantly focus on expanding the work around the larger implications of these practice-based shifts in representation, with a particular focus on an invisual parallax.

2 Methodology

As part of this methodology, I will set out the practices of generating frame-consistent image-based content from language to video A.I. models, paying specific attention to the targeted photogrammetric language prompts and terminologies that enable accurate forms of reconstruction. I will then discuss the reconstruction process via sparse alignment in software such as RealityCapture. Finally, I will establish how this methodology enables training via Neural Radiance Fields, as a 'deeper' form of parallax, that compensates for the potential deinterlacing between footage and images.

2.1 Generating a Language of Invisual Parallax

The recent advances in consistent video generation have played a significant role in allowing for the parallactical reconstruction of post-generation A.I. imagery. Previously, these types of methodologies would fail due to inconsistency between frames, with deviations making reconstruction a major challenge. The consistency in frame-to-frame stability has enabled consistent camera tracking and alignment and opens up this dialogue with traditional parallactical reconstruction. For this case study, I will be using the recently released RunwayML Gen-3 Alpha, (RunwayML, 2024) which takes on textual prompts in the following format:

A`[Camera Movement]:` **B**`[Establishing Scene].` **C**`[Additional Details].`

Through the three-part textual prompting operation, there is an ability to generate consistent parallax between frames within a scene, which can then be used in traditional photogrammetric reconstruction.

Figure 1. Generation (Left to right, start to end) of consistent orbital video footage in Runway Gen-3 Alpha, using a series of parallax-based prompts. Authors Own, 2024.

From experimenting with prompt engineering within these generative video models (figure 1), it becomes clear that certain forms of camera description (A), and specific prompts (B, C) enable more consistent ‘parallax’ than others, and enable the ability to reconstruct three-dimensional scenes in more accurate, or stable, ways. The specific language prompt below has been used in the methodology section of this paper, showing a generated orbital camera path:

A`[A Camera orbits around the space with a Macro Lens]:`
B`[The object is a biological architectural model, atop a podium, the architecture is intertwined with different species of flora and complex biological material, made of various stone and natural materials].`
C`[As the camera orbits around the model it makes sure to keep every area in focus, there are also a series of survey calibrations located around the object and podium.]`

Each text prompt attempts to define further parallactical terms, and calibrations that support image-to-image consistency in future photogrammetric reconstruction. Further camera tests have developed a language that operates with similar prompts, and although not an extensive list, the following prompts will show the linguistic importance of generating this type of content:

A `[A FPV camera moves through the space slowly]:`
A `[A wide-angle camera with a 50mm lens]:`

Simultaneously, specific scene calibration prompts, help to define additional ways in which the generated content can be made more consistently stable.

C [Black and white calibrations markers are located evenly around the space].
C [Everything remains in clear and sharp focus as the camera moves around the scene].

2.2 Photogrammetric Reconstruction of Generated Content

In the second step of this methodology, a standard photogrammetric reconstruction is undertaken in RealityCapture to create a sparse camera alignment (figure 2). The spare alignment is solely used to generate the camera coordinates in 3Dimension space, which are then sequentially used to train the NeRF model.

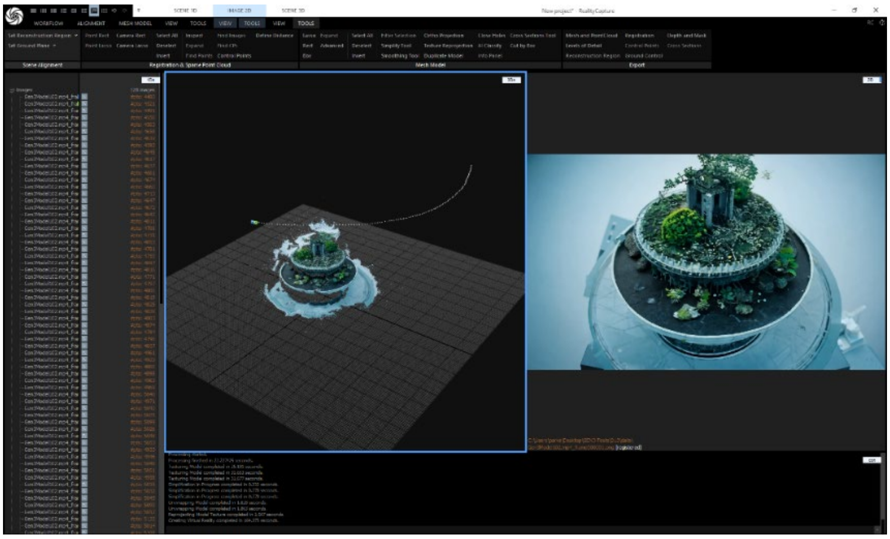

Figure 2. Three-dimensional Reconstruction of the generated scene, showing the consistency of camera work based on a sparse alignment method. Authors Own, 2024.

2.3 Neural Radiance Fields and Training an Invisual Parallax

The final step in this methodology is to train the NeRF model on the generated and aligned images. The NeRF model updates the camera positions slightly during the optimization of training the scene, which allows for a more complete visual scenography. This stage currently works with a series of open source applications, and in this case, the described methodology deploys Nerf-Studios platform due to the speed and computation efficiency (Tancik et al., 2023). The specific model used is the Nerfacto implementation, which is focused on a complex balance between pose refinement in the generation of accurately aligned content, and the overall generation of the density field.

The Nerfacto implementation developed by Nerf-Studio has been shown to have the clearest refinement in the consistent generation of spatial scenes in this methodology, whilst other methods, such as Instant-NGP and Mip-Nerf have struggled with this type of image data. Figure 3 shows the previously defined prompt and its final three-dimensional outcome in a stereoscopic pairing, having been trained on the default Nerfacto method to 30,000 steps.

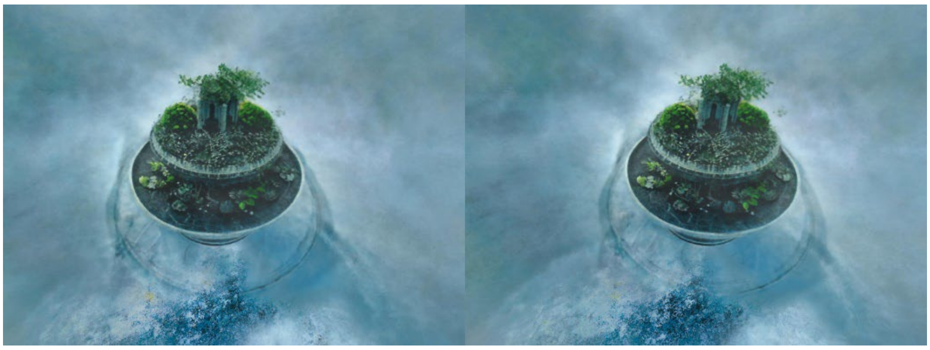

Figure 3. Initial test models within the proposed workflow, are shown as a stereoscopic pair. To view these images without a dedicated stereoscopic viewer, the viewer must first go cross-eyed until identical objects from each image register with each other and then hold the image until it finds its full depth. The viewer will then see three images - concentrate on the middle one. Authors Own, 2024.

2.4 Alternative Methodologies (Gaussian Splatting)

For the sake of conceptual clarity, this paper is focused on Neural Radiance Field methodologies, even though there have been numerous advances in the past years concerning Gaussian Splatting (GS) techniques. The primary focus of this research is not simply to represent the underlying data in the most computationally efficient way, which is how a large amount of GS methodologies work. Instead, the focus is on understanding the complex pose optimization and training that is occurring across the NeRF neural network, and the consequences this has for spatial reconstruction and further visualization. The focus on understanding this process, as opposed to simply representing data most efficiently, will be expanded upon in the discussion section. This distinction is important to highlight to help contextualize the position of the 'invisual' as a realm of opportunity for future epistemological investigation, rather than a singular pursuit of computational optimization.

3 Results

The results of this methodology have opened up a question about the fundamental nature of photogrammetric reconstruction, and how this field might evolve based on A.I. generative image models and advances in computer vision techniques via Neural Radiance Fields. This methodology shows how language-based prompt engineering can create new ways of altering our existing parallactical practices, in both image making and metrology, with a focus on invisual methodologies that can produce new forms of representation.

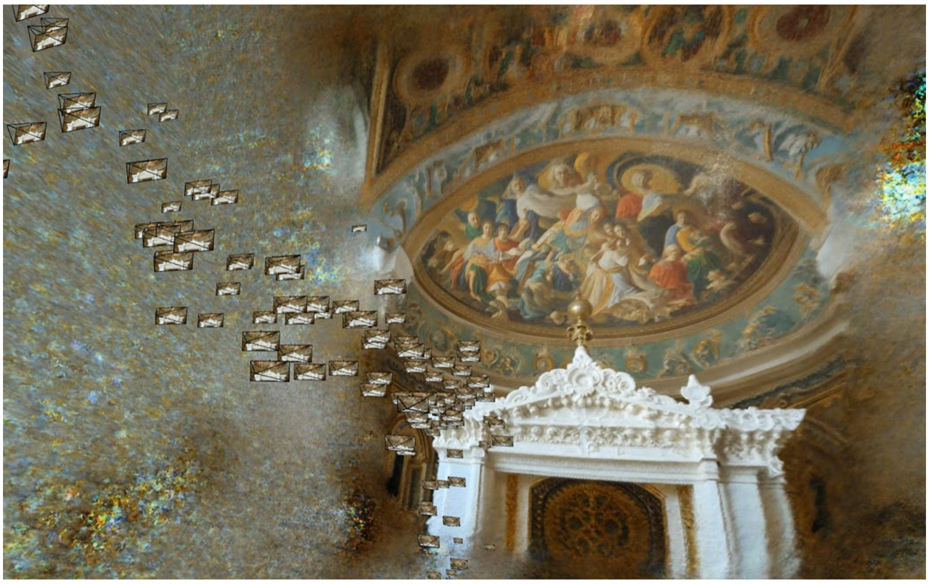

Figure 4. *A Post-Lenticular Baroque*, *Seed #1238779122*, Early study showing the placement of trained cameras within the NeRF scene. Authors Own, 2024.

3.1 A Post-Lenticular Baroque

The project we will describe is an ongoing series of representational experiments, (figures 4-6), which are currently titled 'A Post-Lenticular Baroque', due to the nature of their generation via non-lenticular means. As John May would posit, *images are inherently dynamic, and our tendency to think of them as static of fixed is a result of the psychohistorical residue of drawings and (chemomechanical) photographs* (May, 2019). These resultant representations are framed in such a manner to challenge this psychohistorical overlap between drawing, photograph and image. Instead, they embrace the representational framing of matter that sits between generated baroque scenography and the simultaneously abstract statistical data and stochastic noise.

In developing these hybrid images, it is important to also note that they have a paradoxical positionality and agency, operating somewhere between historical measurement and contemporary computer vision, or between representational image and generative language model. In visualizing this data, there is an intention to find resonance for these multiple agencies in vastly differing aesthetic and epistemological positions. As Francesca Hughes would define in the history of architectural representation, there is a tendency to conflate *message and matter* (Hughes, 2014). In the production of these images, what might at first be seen as erroneous matter, might be more accurately defined as the alternative message. The play between perspectival rules and representational content is subject to broader questions about where accurate spatial construction ends, and where the value of complex multilayered trained spatial data, and its complex resultant computational noise, begin.

Figure 5. *A Post-Lenticular Baroque, Seed #3834899512*, Stereoscopic Image Pair, showing not only the reconstructed scene but the NeRF noise field in the background. Authors Own, 2024.

In Figure 5 there is a focus placed on using a parallax underpinned representational practice in the form of the stereoscopic image pair. This image is also an attempt to momentarily freeze the dynamic nature of both the data and digitally 'reconstructed' scene, in an uncomfortable optical dialogue, creating a solitary position for the viewer between representational clarity and the need to discern spatial hierarchy from abstract forms in both foreground and background. As Massimo Scolari notes on the perspectival subversions of geometry, *Optical Illusions of this kind have always been used in scenography* to effectively counteract the *very shallow space of a stage* (Scolari, 2015). All of the outputs of this project operate in this currently still shallow field, challenging a post-lenticular engagement, whilst undulating between known pictorial conceits and emerging forms of computer vision-related logics.

4 Discussion

This paper has set out to frame a conversation around advances in generative A.I. imaging and the potential of future methodological shifts around parallax-based forms of representational practices. What this has uncovered is an unconventional form of production, which priorities a process-based understanding of the work of both architectural representation and the underlying digital models that are no longer simply designed geometry in cartesian coordinate systems. Instead, it probes the depths of multiple neural networks, which are both linguistically, optically, and visually complex, creating a multiplicity of dialogues between different media. This process allows a new form of practice to emerge for discussion, one which is both invisual and yet visible, operational yet aesthetically driven, and a positionality by which we might question our future practices with A.I. methodologies.

Figure 6. *A Post-Lenticular Baroque, Seed #2854811993*, Film still of animated output, showing the centrally emerging baroque space amongst the Radiance Fields surrounding noise. Authors Own, 2024

4.1 An Invisual-Visual Practice

To state the obvious, the shift that is occurring in image-based production within all forms of contemporary media is incredibly fast (Steyerl, 2009). Architecture's responses to A.I.-generated images are not uncommon amongst the wider creative industries, as emerging new visual fields begin to challenge the principles of our existing practices. The problematic position that emerges in these practices, around the roles of images and data, is not easily challenged by a singular approach either, and this paper does not intend to provide a simple answer for a singular future representational practice. Instead, it sets out to define how an acknowledgement of our ‘invisual’ practices might enable new forms of dialogue between the tools and methodologies by which we communicate, design and even generate, architectural ideas.

The emergence of stable image generation models, alongside advances in neural radiance fields, is changing how we generate and digest images and construct digital space. How we then measure such content is already proving to be an alarming shift in the operational logic of metrological practices. These practices have historically required an uneasy level of labour, calibration, time, and precise measurement to be able to discern useful clarity about the underlying geometry of a three-dimensional scene (Ray, 1999), and the role of the perspectival lenticular image in this type of process is now open for discussion. The underlying optical parallax, that all able-bodied viewers can perceive, is simply no longer a given aspect of how A.I. models are trained to generate this type of data. These models have no sense of a physical lenticular source and instead work through complex invisual data-based processes that outstrip that of the human oculus.

Although we refer to the process as a ‘trained parallax’ it is also important to note that we do not mean to imply that this is similar to a human-centric learning process, instead, it is as Michael Young describes, a *statistical optimisation*, not cognitive learning or competency (Young, 2021). This is an important definition, as this type of practice opens up larger fundamental epistemological concerns around the nature of how architects are now required to learn in dialogue with emerging generative tools. As Mario Carpo summarized at the end of his writings on the second digital turn, these new practices are *also, surreptitiously, pushing us towards a new way of thinking: [they are] training us to think in a different way, following a new, post-scientific logic* (Carpo, 2017). These post-scientific logics do not simply erase the historical context of the discipline that came before it, but instead challenge us to question the formerly rigid hierarchies in parallel with emerging and changing tools and media.

We might then find solace for our current dilemma in similarly technologically underpinned parallels that historically emerged in the invention of media such as photography. The invention of the media not only created new forms of representation, but also entire subcultures of experimentation around practices such as spirit-photography and thoughtography, and the imponderable media that they then enabled (Christian, 2018). Even though not formally scientifically underpinned pursuits, these endeavors gave rise to new forms of representational epistemologies in dialogue with the resultant viewer. We can also find answers in other forms of epistemological and representational shifts that have already occurred in the architectural discipline before. Questions of the geometric value of the perspectival drawing were challenged at the Ecole des Beaux-Arts, but instead of simply erasing these practices, it resulted in a *dichotomous merger of geometry with pictorialism* (Thomine-Berrada 2013).

These moments of epistemological challenge create a need for architects to construct simultaneous representational and geometrical practices, that move beyond the singular confines of an existing technique.

Invisual modes of representation may move our practices into complex dialogues, but primarily away from the long shadow of the optical, and instead towards their operation within the *spectrum [of the] neural network calculated models of the thousands of tiny futures just a fraction ahead of direct perception* (Parrakis, 2023). This type of practice would need to understand its paradoxical position, as an operation before visual perception and representation. Instead, it would need to prioritize the knowledge of measured registration of spatial information, the applications of generative data, and even the formal definition of the linguistics of spatial complexity. This palimpsest of information, which may not yet be achievable on a day-by-day scale, would open up the political, social, cultural, and aesthetic aspects of data and the digital in representational practices (Vesna, 2017). Allowing us to question more thoroughly how emerging forms of A.I. representation might engage in expansive epistemological challenges to our existing practices and workings, rather than simply seeking to automate or replace their underlying principles and labour.

In summary, there is a dialogue yet to be had within this next phase of the digital turn, one that is framed against the background of our traditional practices becoming questioned by the emerging generative tools that subvert the foundational nature of our existing media landscape. This dialogue can be a productive moment of inflexion with the right epistemological lens to challenge and question the foundational logic underpinning our architectural practices of modelling and representation. This might require us to no longer simply look towards our existing aesthetic languages and judgements, but instead, allow these assumptions to retreat, for a position more able to engage with the invisual processes that now dominate our visual culture.

Acknowledgements. Financial research support is generously provided by the London Arts and Humanities Partnership (LAHP) Open Student Scholarship.

References

Jones, E.L. (2023) 'The Art of Measuring Images: Albrecht Meydenbauer and The Invention of The Photographic Survey', DMJournal–Architecture and Representation: No. 2: Drawing Instruments: Instrumental Drawings [Preprint].

Mittal, A. (2023). *Neural Radiance Fields: Past, Present, and Future*. https://doi.org/10.48550/ARXIV.2304.10050

Rabby, A. S. A., & Zhang, C. (2023). *BeyondPixels: A Comprehensive Review of the Evolution of Neural Radiance Fields*. https://doi.org/10.48550/ARXIV.2306.03000

Parikka, J. (2023). *Operational Images: From the Visual to the Invisual*. University of Minnesota Press. https://doi.org/10.5749/9781452970929 p18.

Brooks, T. and Peebles , B. (2024) Creating video from text, OpenAI. Available at: https://openai.com/index/sora (Accessed: 03 July 2024).

Germanidis, A. (2024) Introducing gen-3 alpha: A new frontier for video generation, Runway. Available at: https://runwayml.com/blog/introducing-gen-3-alpha/ (Accessed: 03 July 2024).

Parikka, J. (2023). *Operational Images: From the Visual to the Invisual*. University of Minnesota Press. https://doi.org/10.5749/9781452970929 p202.

Parikka, J. (2023). *Operational Images: From the Visual to the Invisual*. University of Minnesota Press. https://doi.org/10.5749/9781452970929 p22.

RunwayML. (2024) Documentation: Gen-3 Alpha Prompting Guide. Available at: https://help.runwayml.com/hc/en-us/articles/30586818553107-Gen-3-Alpha-Prompting-Guide (Accessed: 03 July 2024).

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., & Kanazawa, A. (2023). *Nerfstudio: A Modular Framework for Neural Radiance Field Development.* https://doi.org/10.48550/ARXIV.2302.04264

May, J. (2019). *Signal. Image. Architecture: Everything is already an image*. Columbia Books on Architecture and the City. p47.

Hughes, F. (2014) *The architecture of error: Matter, measure, and the misadventures of Precision.* Cambridge, MA: The MIT Press. p164.

Scolari, M. (2015) *Oblique drawing: A history of anti-perspective*. Cambridge: MIT Press. pp27-8.

Steyerl, H. (2009). In Defense of the Poor Image. *E-Flux Journal*, *Issue #10*.

Ray, S. F. (1999). *Scientific photography and applied imaging.* Boston : Focal Press.

Young, M. (2021). *Reality Modeled after Images: Architecture and Aesthetics after the Digital Image* (1st ed.). Routledge. https://doi.org/10.4324/9781003149682 p157.

Carpo, M. (2017) *The Second digital turn: Design beyond intelligence.* Cambridge, MA: The MIT Press. p164.

Christian, M. I. (2018). Cameraless Photography and Its Imponderable Media. *History of Photography*, *42*(4), 319–337. https://doi.org/10.1080/03087298.2018.1540182

Vesna, V. (Ed.). (2007). Database Aesthetics: Art in the Age of Information Overflow. University of Minnesota Press. https://doi.org/10.5749/j.cttts7q7

Parikka, J. (2023). *Operational Images: From the Visual to the Invisual*. University of Minnesota Press. https://doi.org/10.5749/9781452970929 p.198.