

This is a forthcoming chapter accepted for the edited volume: *The Bridge: Connecting Perspectives: Visual Methods as Epistemological Practice*, edited by Nilay Ulusoy, Funda Kaya and Mark Dunford, expected to be published in 2027 by Peter Lang. This text is a pre-contractual draft and may undergo further revisions. Do not cite, quote, copy, or circulate this version in any form without explicit written permission from the author: regev@umich.edu

Principles for Multimodal Anthropological Videography (MAV)

Regev Nathansohn, PhD (regev@umich.edu)

Abstract

Multimodal Anthropological Videography (MAV) is an innovative platform for relational analysis and display of anthropological research. Instead of treating research materials (videos, images, sounds, texts) separately, presenting them sequentially, or emphasizing their collective value, MAVs creatively juxtapose and synchronize two or more materials on a single screen in ways that promote dialectic and even trialectic argumentations and modes of display. This chapter presents the epistemological and methodological foundations of MAVs, and through the case study of Uncoding the City – a MAV based on anthropological research on urban digitalization – demonstrates the practice of techno-analytical video editing as vidécriture.

Have you ever created and collected videos, photos, audio, texts, and other materials during your research project, but felt unsure how to weave them into a compelling argument? This is how I started my anthropological inquiry on urban digitalization. I wrote field notes, used my smartphone to record videos, images, and sounds, and saved multimedia materials from relevant websites. After several years of creating and collecting, the desire to do something meaningful with all this material led me to create "Uncoding the City", a trilogy of videographies, and to develop the Multimodal Anthropological Videography (MAV) as a platform for relational analysis and display of anthropological research.

MAV as Vidécriture

MAVs are multimodal in that they analyze and display multiple modalities, including text, images, sound, and their hybrid forms. MAVs are anthropological in the

sense that these modalities are anchored in participant observation and are prone to critical analysis, which may include various forms of collaborations in the inquiry of "the conditions and possibilities of human life in the world" (Ingold, 2017). MAVs are videographic in the sense that they require practices of analysis and presentation through writing with video. Borrowing from filmmaker Agnès Varda's concept of cinécriture (Smith, 1998), MAVs are a form of vidécriture, where the researchers use video-editing tools as their techno-analytical artisanal pen.

Technically simple and analytically challenging, MAV can serve researchers across disciplines and career stages for teaching, conferences, film festivals, public engagement, and research publications in various outlets. Additionally, it can serve various forms of creative collaborations with students, peers, and research partners (see Almenara-Niebla et al., 2026; Dattatreyan and Marrero-Guillamón, 2019). This chapter presents the epistemological and methodological foundations of MAVs and discusses the practice of techno-analytical video editing, namely the juxtaposing and synchronizing of anthropological materials of different modalities by means of videography, or vidécriture.

MAV as a Platform for Relational Analysis and Display

MAVs are based on two preliminary axioms: surprising things happen when two or more elements are brought together, and there is no one way to create a MAV. The idea of bringing two (or more) elements together to develop a new meaning is inspired by Sergei Eisenstein's dialectical montage (1949), which presents an argumentative strategy to filming that juxtaposes different scenes (thesis and antithesis) to allow a new meaning (the synthesis) to emerge (Polan, 1977). Building on this approach, rather than treating research materials separately, presenting them sequentially, or emphasizing their collective value, MAVs creatively juxtapose and synchronize two or more materials on a single screen in ways that promote dialectic and even trialectic argumentations and modes of display.

As a multimodal platform for analysis and display, MAV is inspired by videographic genres in film and media studies (Keathley & Mittell, 2016), which in the past two

decades analyze films and moving images by filmographic means (Avissar, 2023).¹ While videography is not common as a form of analysis and presentation of anthropological knowledge, multimodal and multisensorial approaches are gaining increasing legitimacy as experimental and inventive forms of creating and conveying anthropological knowledge (see Borpujari et al., 2024; Collins et al., 2017; Collins & Durlington, 2024; Dattatreya and Marrero-Guillamón, 2019; Favero, 2020; Murphy, 2025). In fact, from its beginning, material objects and audiovisual media (such as drawings, photos, films, and digital media) have been part of modern anthropological research and display, establishing the notion that "meaning is produced through the inter-relationships between and among different media and modes" (Dicks et al., 2006, p. 78). Moreover, the idea of video editing as a form of analysis is now established in scholarly and artistic fields (see Suhr & Willerslev, 2013). For the social sciences, MAV is most suitable as a platform for relational analysis, which stands at the core of anthropological inquiry (see Strathern, 2020). It can serve, for example, for spatiotemporal relational analyses, and for elaborating on the trialectic relation (attributed to anthropologist Margaret Mead) between what people say, what people do, and what people say they do.

Methodologically and epistemologically, creating MAVs requires considering which materials and modalities stand against what, and how: visually, analytically, aesthetically, poetically, and politically. The affordances and constraints of audiovisual thinking through videographic practices allow reframing analytical procedures (Avissar & Laird, 2026). They enable researchers to creatively transcend limitations of written formats while also amplifying an emotional affect as part of the analytical process, also reflexively (Avissar, 2023). Moreover, much like analytical writing, creating MAVs is an iterative process that develops with, not before, the juxtaposition of videos, images, and sounds (words) into shots (sentences) and scenes (paragraphs).

An inspiring example of thinking about MAV is Lucrecia Martel's short documentary "La ciudad que huye" (2006),² edited by Ariel Ledesma Becerra. In this

¹ See, for examples, the videographies and audiovisual essays published in *tecmerin* (<https://tecmerin.uc3m.es/revista>), *Fragments* (<https://fragments.video>), *[in]Transition* (<https://intransition.openlibhums.org>), and *NECSUS* (<https://necsus-ejms.org>).

² <https://youtu.be/1grtEpw0q9c>.

videography, Martel focuses on gated communities of upper- and middle-class residents in the Buenos Aires area. The videography is composed of segments in which Martel juxtaposes footage of walls and fences surrounding these areas with footage from the opposite neighborhoods. Another juxtaposition is presented at the bottom of the screen, with three narrow strips of footage showing a drive along the wall, and a meter counter indicating the length of the wall. At the same time, the center of the screen shows satellite images and dynamic maps with infographics that describe the expansion of the city, with a voiceover that contextualizes it socio-historically. Between these juxtapositions are footage and sounds describing Martel's attempts to enter the gated communities and the guards' reactions to them. Towards the end of the videography, the center of the screen is divided into six parts, with each showing footage of a drive along a wall surrounding one of the gated communities. On the top and bottom of this composition are narrow stretches of wall footage. Such videography requires a different kind of attention while watching. Not having a single frame to look at, watching becomes an active practice of searching for the relations between the different audiovisual materials presented simultaneously.

In the next part of this chapter, I will present several techno-analytical editing tools relevant for MAVing, and will exemplify how I used some of them in Part One of *Uncoding the City*.³ This trilogy of 15-minute MAVs is based on my anthropological research on urban digitalization. In this multi-sited urban field, "smart cities" are believed to be more efficient, sustainable, and predictable thanks to their translation of urban practices and interactions into digital data that feeds digital algorithms and artificial intelligence (AI). Between 2017 and 2025, I created and collected videos, photos, audio recordings, and fieldnotes from smart city exhibitions, conferences, meetups, and hackathons, as well as from smart city webinars and websites. At the same time, I created multimodal vignettes depicting everyday urban life in several cities around the world. Each videography proposes a way to un-code the city by focusing on one of these three research questions: (1) How do technology companies transform the urban sphere

³ See trailer of part one here: <https://vimeo.com/1164594566> [the full version is currently under review in an academic publication; its link will be added to the final version of this manuscript].

into a digitally codable one? (2) What social interactions are digitally coded by smart city products? (3) Why do technology companies view urban social interactions as digitally codable?

Theoretically, the project draws on a neo-Bourdieuian approach (Wacquant, 2023) to analyze urban digitalization as a social field that unveils the trialectics of symbolic, physical, and social spaces. Each of these spaces corresponds with one of the project's videographies. Part One of this trilogy answers the first question by focusing on the symbolic space: the mental categories through which we perceive and organize the world. This videography analyzes how technology companies re-symbolize the city as a codable space in which humans, objects, practices, and interactions can be digitally coded. It critically examines the marketing strategies of nine technology companies and shows how they seek to re-symbolize urban space as future-oriented and digitally codable to promote their products and deepen their intervention in cities. Juxtaposing their promotional videos with four anthropological vignettes of everyday life in Barcelona reveals how the quantitative bias embedded in smart city products disregards qualitative characteristics of human practices, decontextualizes them, and oversimplifies them.

This videography presents the uncoding of cities by editing marketing videos against the grain and juxtaposing them with urban anthropology. Showing how both digital and cultural algorithms (Nathansohn, 2022) play out in human interactions offers a gateway from the companies' posthumanist approach. By redefining the urban symbolic space as future-oriented and digitally codable, technology companies frame their intervention in urban life as both unquestionable and inevitable, resulting in transformations of the city's physical and social spaces.

The Techno-Analytical Video Editing

As a vidécriture, MAV requires a holistic approach to videography, where content is expressed also by form, and vice versa. While in standard films the screen replaces the written page and the timeline replaces the pace of reading, in videography this spacetime becomes richer, opening endless opportunities to use the screen's space and

the progression within the timeline. The sound layer adds another dimension that can be aligned or disaligned with the visual elements. Video-editing a MAV is therefore a techno-analytical practice through which every editing choice has analytical meanings, and every analytical move has its editing manifestations. Thus, when creating a MAV, equal attention is given to the how and the why of the editing technique in order to promote a specific analytical move.

Although creating a MAV is technically simple and accessible for researchers and their collaborators, regardless of their experience with filmmaking or editing,⁴ analytically, it requires intellectual effort to consider how best to juxtapose the multimodal materials and reflect on what is afforded by such juxtaposition. MAV's techno-analytical video editing fosters active human creativity through an iterative practice, from creating and collecting source materials to critically reflecting on the theory-based analysis that various forms of juxtaposition and synchronicity afford. While various AI tools can assist with some video-editing techniques, human creativity in the process, much like in anthropological fieldwork itself, cannot be replaced by digital algorithms. Such creativity requires considering the following key questions regarding content and form: What is the preliminary research argument? What is the added value of juxtaposing and syncing two or more modalities in one frame? How can the research argument be developed into a visual argument through multimodal videography, keeping in mind that each editing technique has its own meaning, explicit or implicit?

To give a sense of how technical simplicity can serve for analytical complexity, let us consider some of the basic audiovisual and temporal elements.

The Screen as a Canvas: composition, relation and movement

Think of the full screen as a canvas, where endless forms of juxtaposing and synchronizing multimodal materials can create a meaningful assemblage. Consider, for example, dividing the screen into two sides, each of which shows a different perspective on the same topic, simultaneously. This simple editing technique invites a comparative

⁴ Most of the free video-editing tools will suffice for creating a MAV.

observation. If the materials focus on differences, the analytical meaning of this editing technique is to conceptualize the presented topic as composed of binary oppositions (as in "La ciudad que huye"). One way to complicate such analysis is to focus on audiovisual similarities. In *Uncoding the City*, what may seem like a binary opposition between online and offline practices and interactions is broken in several instances, for example, in a scene where pigeons are presented on one side and drones on the other, implying that they share similar features. Other examples show similarities between interactions in a flea market or in a card-trading street market, and interactions between presenters in a smart city exhibition.

Clearly, there are an infinite number of other possible compositions. The screen can be divided into more than two sections; visual materials can be placed on the screen next to each other or on top of each other, with different degrees of transparency; a single source can be used several times; a still image or a full-size clip can serve as a background for other materials presented in a smaller size. And clearly, materials can be put in motion, or zoom in and out to attract viewers' attention.

In *Uncoding the city*, I decided to divide the screen into four parts: the top part for a video strip, the bottom part for subtitles, and the center of the screen for the main juxtaposition, with a division to the right and left. At the beginning, the left side shows footage of urban vignettes, and the right side shows scenes from promotional videos for smart city products. At a certain point, as the analysis focuses on the 2025 power outage that hit southern Europe, the smart city materials disappear, and footage of urban vignettes fills both parts. When smart city footage reappears, it is presented on the other side, and the footage of urban vignette shows how technology is being used in various forms in everyday life (see Image 1). Such a shift in composition highlights the relational aspect of technology: its embeddedness in everyday life, simultaneously with offline interaction in high-tech fields. In the last part of this videography, when the focus is on what is required for uncoding the city, the two sides change again, emphasizing the conjunctive, nonbinary relations between digital and cultural algorithms in urban life.



Image1 : Screenshot from Uncoding the City, Part One

Time, Speed and Rhythm

While in the physical world one second lasts a second, in anthropological inquiry the difference between a split-second twitch and a wink can be the subject of lengthy written analyses. In videography, a whole set of editing techniques invites different analytical moves that can allow slowing down a documented event, accelerating it, freezing a moment in time, reversing it, making it repeat, or play in a certain rhythm. In relational analysis, each of these, as well as the synchronizing (or desynchronizing) of media, carries its own meaning.

In *Uncoding the City*, it was important for me to have the upper part of the screen devoted to a narrow strip of long footage showing a single section of a road, where it meets the sidewalk. While events in the other parts of the screen change rapidly, this upper strip remains constant in the scale of events presented but surprising in the scope and rhythm of interactions involving humans, animals, and vehicles, thereby emphasizing the futility of the desire to digitally code everything that happens in the city.

Sound

Like prose in writing, sound is a foundational, non-secretive element that can shape the video's deeper meanings. The sound layer of a videography includes voiceover, music, and, clearly – the sound relevant to the analyzed topic. Sound can be synchronous

or asynchronous with the visuals. In filmographic language, the sound can be diegetic or nondiegetic: it can stem from the visually recorded and presented footage, or it can exist outside of the presented footage, or be related to another spatiotemporal event. In *Uncoding the City*, for example, in the background for my voiceover, I sometimes added another sound level of urban noises that I audio-recorded, particularly of people chatting in the public sphere, along with various sounds of transportation, in order to deliver the sense of offline interactions in the city, even when they are not presented in the visual footage. Like visuals, sound can also be manipulated to achieve a certain goal, for example, by playing it in a loop or changing its speed and volume.

The audiovisual Text

Relative size, font, color, transparency, location on the screen, and speed of presentation of audiovisual text may matter. In most cases, subtitles will be required, even if only for accessibility reasons, but they, too, can take many forms. Subtitles can be presented in the standard way, or in any other way that foregrounds a specific meaning. Meaningful text, that is not part of the subtitles, can be presented on the screen as part of the multimodal assemblage. Text can also be presented on top of video scenes, to break the division between the recorded scene and the related text. In *Uncoding the City*, I divided the subtitles area to conform with the division into the side of technology companies and the side of everyday life, with each side having its own subtitles. I also used a specific font for the subtitles of companies' voiceovers, a different font (handwriting style) for my fieldnotes, and yet another font for my analysis.

Transitions

Audiovisual transitions, such as cuts, dissolves, fades, and various movements, may signify fragmentation, transition, and connection of the materials presented in a MAV. While transition galleries of editing programs can be enormous, it is the analytical and aesthetic value that should determine which one to choose. In *Uncoding the City*, I used a cut to black when discussing the major power outage. Towards the end of the videography, I used a smear transition to switch the sides of technology companies and everyday urban life back to where they were at the beginning, to stress their interconnectedness before discussing what uncoding the city requires (see Image 2).



Image 2: Screenshot from Uncoding the City, Part One

AI and Additional Features

There are many other editing features that can carry analytical meanings, such as playing with colors, textures, and even disruptive editing that can draw the viewers' attention to the techno-analytical editing. Disruptive editing can, for example, be used for decontextualizing and recontextualizing research materials. Clearly, the more editing skills one has, the more precise and sophisticated they can be in crafting an argument videographically, using fewer words to explicitly explain analytical moves. However, overuse of editing techniques can also distract audiences.

Being creative also requires minimizing the use of AI tools, which become more and more integrated into video-editing programs. While some of these tools can save a lot of time doing strictly technical work, others may have analytical implications that may be better not delegated. It is also important to note that AI tools may require rights to use the materials for other purposes as part of their *modus operandi*.

Let's Start MAVing

With endless ways of techno-analytical video editing, there is clearly no one way or right way for MAVing. Moreover, as mentioned above, its production is not linear but iterative. You may have an idea of how to begin, but you cannot anticipate what it will look like in the end. During MAVing, materials will be transformed, theory will be developed, and research arguments will be crystallized.

For that to happen effectively, you need to begin by reviewing your raw materials, get familiarized with the relevant theory, and draft the research arguments you wish to convey. Reviewing the materials has three interrelated goals: analytical, organizational, and operational. Inspired by a relevant theory and driven by your research questions (How urban space becomes digitally-codable?), getting a deeper sense of your materials and their modalities is a necessary step in refining your research arguments and in thinking of how to deliver them videographically. I found it helpful to transcribe my materials, describe in words what I find significant in them (content-wise, form-wise, analytically, and aesthetically), and note how I wish to treat them. While reviewing your materials, you will also develop a better sense of the editing strategy and technique that will best deliver your argument. Think about how to best use juxtaposing and synchronizing of your materials to convey one (or more) of your arguments. To add another analytical level of sociopolitical criticism, consider being reflexive about the techno-analytical editing choices, as well as about absences and silences in the materials and their videographic treatment.

Just like in the film industry, and much like in writing, the process of MAVing will lead you from a rough cut to a fine cut to the final cut. In the rough cut, you will organize all the materials you wish to edit (footage, images, sounds, texts) sequentially and proportionally on the timeline, in separate layers according to your editing strategy. The fine cut will present a more advanced edit, including relevant graphics, transitions, music, voice-over, and subtitles, so it can be shared with others for review. The final cut will be the one ready for export and presentation, after implementing all relevant comments, color, and sound corrections.

Juxtaposing multiple materials on a single screen demonstrates that the result is not simply the sum of its parts. This approach enables consideration of possible thematic and visual relations as well as fissures between elements presented simultaneously. Rather than presenting a closed, didactic analytical argument, this method opens possibilities for diverse modes of viewing, interpretation, and the generation of new questions.

MAV is a straightforward and accessible method, that provides social scientists with new tools to expand the boundaries of their thinking, representation, and argumentation. These tools are designed to engage a growing audience, both within and beyond academia, in a contemporary, relevant, and compelling way. Moreover, it can and should also be a legitimate form of an academic theses and articles.

MAVing requires analytical thinking, which can go hand in hand with low-key, low-budget production. Today, everyone can adopt the method, make multimodal videographies on their own or as a collaborative project, and share their knowledge on whichever platform they desire. MAVing invites creativity. It calls for non-didactic reasoning and linear argumentation. It calls for using all the modalities at hand and, when possible, letting their assemblage, through juxtaposition and synchronization, do its magic.

Work Cited

- Almenara-Niebla, S., Smets, K., & Alencar, A. (2026). Reorienting Collaborative–Creative Methodologies: Reflexive Devices in the Interplay of Ethnographic and Mobility Research. *Qualitative Inquiry*, online, 1-10. <https://doi.org/10.1177/10778004251412872>
- Avisar, A. (2023). A Thousand Words, 24 Times a Second: The Videographic Essay. *Media Frames*, 23, 125-131. <https://doi.org/10.57583/MF.2023.23.10033>
- Avisar, A., & Laird, C. (2026). Reframing the Argument Special Issue Introduction. *[in]Transition*, 13(2), 1-7. <https://doi.org/10.16995/intransition.29322>
- Borpujari, P., Cook, I. M., İlengiz, Ç., Murphy, F., Öffen, J., Sander Puustusmaa, J., ... Wardell, S. (2024). Empathy and Dialogue: Embracing the Art of Creative Review. *Anthropology and Humanism* 49(2), 83-87. <https://doi.org/10.1111/anhu.12536>
- Collins, S. G., Durlington, M., & Gill, H. (2017). Multimodality: An Invitation. *American Anthropologist* 119(1), 142-46. <https://doi.org/10.1111/aman.12826>
- Collins, S. G., & Durlington, M. S. (2024). *Multimodal Methods in Anthropology*. New York: Routledge.

- Dattatreya, E. G., & Marrero-Guillamón, I. (2019). Introduction: Multimodal Anthropology and the Politics of Invention. *American Anthropologist*, 121(1), 220-228. <https://doi.org/10.1111/aman.13183>
- Dicks, B., Soyinka, B., & Coffey, A. (2006). Multimodal Ethnography. *Qualitative Research* 1(1), 77-96. <https://doi.org/10.1177/1468794106058876>
- Eisenstein, S. (1949). *Film Form: Essays in Film Theory* (edited and translated by Jay Leyda). New York: Harcourt Brace Jovanovich.
- Favero, P. SH. (2020). Audio-visual-sensory Essays in Post-Digital Times: Communicating Anthropological Knowledge Across the Multimodal, Multisensory, Participatory and More-than-human. *Anthrovision* 8.2, 1-18. <https://doi.org/10.4000/anthrovision.7934>
- Ingold, T. (2017). Anthropology contra ethnography. *HAU: Journal of Ethnographic Theory* 7(1), 21-26. <https://doi.org/10.14318/hau7.1.005>
- Keathley, C., & Mittell, J. (2016). *The Videographic Essay: Criticism in Sound and Image*. Montreal: Caboose books.
- Murphy, F. (2025). An Ode to Being Wrong About Creativity. *Allegra Lab*. <https://doi.org/10.65268/VTWN5404>
- Nathansohn, R. (2022). Invisible Smart Cities. In B. Linder (Ed.), *Invisible Cities and the Urban Imagination* (pp. 293-301). Cham: Palgrave Macmillan.
- Polan, D. B. (1977). Eisenstein as Theorist. *Cinema Journal* 17(1):14-29. <https://doi.org/10.2307/1225467>
- Smith, A. (1998). *Agnès Varda*. Manchester: Manchester University Press.
- Strathern, M. (2020). *Relations: An Anthropological Account*. Durham, NC: Duke University Press.
- Suhr, C., & Willerslev, R. (Eds.). (2013). *Transcultural Montage*. New York: Berghahn.
- Wacquant, L. (2023). *Bourdieu in the City: Challenging Urban Theory*. Cambridge: Polity Press.