Special – *peer-reviewed*
*Edited by Rossella Catanese, Adelheid Heftberger and Christian Gosvig Olesen*
https://doi.org/10.6092/issn.2280-9481/13210

Cinergie – Il cinema e le altre arti. N.20 (2021)
ISSN 2280-9481
https://cinergie.unibo.it/

# Digital Picklock for Video Art Exgesis: Reflections, Conditions and Possible Employment of Distant Viewing to Moving Image Datasets in Visual Arts Scholarship

Diego Mantoan*

Ca' Foscari University of Venice (Italy)

Submitted: June 30, 2021 – Revised version: August 7, 2021
Accepted: October 14, 2021 – Published: December 20, 2021

**Abstract**

With the advent of digital humanities new expectations and challenges are emerging for institutions harbouring video artworks, specifically in offering access and analytical tools to their archival collections. The paper argues for the possible employment of distant viewing to allow visual arts scholars an unprecedented take on video art, holding together both quantitative comparison and aesthetic considerations. In doing so, the paper addresses the peculiar conditions of video art that need proper consideration for a fruitful employment of distant viewing. Set on the background of the existing platforms for video-art consumption – such as UBU Network, JSC Media Centre, and Daata Streaming Platform that constitute true forerunners in this domain – the paper explores productive connections, synergies and frictions that might emerge with methods in digital humanities. In doing so, this research aims at setting early theoretical assumptions necessary to draw a methodological approach in the employment of distant viewing to video art. Accordingly, the paper reflects on the effectiveness of using thematic sub-sets based on categories already defined by visual arts, as well as on the possible implications of widespread practices such as manipulation and appropriation of video material.

**Keywords**: Video Art; Distant Viewing; Computer Vision; Pathosformeln; Image Detection.

* diego.mantoan@unive.it

## 1 Distant Viewing of a Virgin Territory and the Challenges to Visual Arts Scholarship

Changing the image ratio, tempering with the reproduction speed, deforming aspect patterns, altering the colour palette, and overlapping different frames are but a few aesthetic strategies employed during video art's relatively short history (Mantoan 2018: 106–10). The works of pioneers such as Nam June Paik and Wolf Vostell in the 1970s already presented a high degree of technical skills, which only grew stronger with subsequent generations. Such was the analytical awareness of the 1990s demonstrated by artists like Pipilotti Rist and Douglas Gordon, finally leading to younger artists such as Ed Atkins and Sondra Perry embracing computer generated images. Over the decades, video artists showed a heightened attention to technological aspects of their chosen medium, intended as a means for aesthetic expression (Comer 2009). Comparably, it is striking that the historiography of video art has been far less technologically oriented and hardly took the chance to employ analytical tools offered by twin disciplines such as moving image studies. Laying somewhere between a unique collectible that needs physical installation and a reproducible product that can be transferred to new carriers, caught up between an everyday appliance and a source of artistic possibility, the ambiguous nature of video art perhaps prevented the possibility to deploy computer-based techniques for analysing and visualising both quantitative and qualitative data (Heiser 2008: 195). While art museums are offering growing datasets of pictures and still images of their collections, often under Creative Commons, video artworks are hardly made accessible online, neither for scholarly purposes nor for the general audience (Navarrete and Villaespesa 2020: 232). Only recently various independent initiatives developed open platforms for viewing wide corpora of digitised video artworks, thus allowing a public insight on the products of celebrated artists, though still with little tools for analytical purpose. Set on the background of the rather few existing platforms for video-art consumption – such as forerunners like UbuWeb, JSC Video Lounge, and Daata Streaming Platform – productive connections, synergies, and frictions with methods in digital humanities might eventually emerge.

Among analytical tools that are gaining pace in computer vision for moving image studies, distant viewing appears to be a digital picklock in image analysis because of its ability to provide quantitative data relating to form aspects and their relationship within broad datasets that might offer scholars an insight into social, economic, and political values defining cultural products (Taylor and Tilton 2019: 6). Over the last decade, computer vision made its entrance in the domain of art history for good with several research teams constructing convolutional neural networks that allow fine-grained big data analysis of art historical corpora (Manovich 2013). Digital humanists focused their efforts on large scale retrieval – with distant viewing among its preferred tools – to be paired with individual analysis of selected samples, such as through close reading (Esser, Sutter and Ommer 2018). So far, the benefit of fine-tuned algorithms and trained machines for art historical research appear to be of a morphological kind, thus centred on finding objects, deciphering gestures, recognising positions, and comparing compositions (Bell and Ommer 2016). In the case of art history this leads to an extension of iconographic and, perhaps, even iconological considerations that allow a deeper analysis of differences and similarities in the dissemination or reception of images across time and space (Monroy, Bell and Ommer 2014).

The question arises whether it is fruitful or even possible to employ a specific method of computer vision, such as distant viewing, to video art allowing visual arts scholars an unprecedented take on this pre-defined category, holding together both quantitative comparison and aesthetic considerations. First, structural and institutional conditions of video art need proper consideration, in order to set the initial clauses necessary to draw a methodological line for the employment of distant viewing to this domain (Moretti 2007). Relying on an art historical framework centred on technological and aesthetic aspects of video art, as well as on practical experience in the development of digital archives of time-based art, the paper examines a variety of research questions that shall define the boundary conditions for distant viewing in video art analysis. Accordingly, it is necessary to reflect on the effectiveness of research categories, labels, annotations, and general metadata already defined by visual arts scholarship. One may wonder what to make of labels referred to the type of content such as narrative, non-narrative, performance based, optical, naturalistic, black and white, colour, palimpsest or animated videos. A similar problem may arise with genre, thematic subsets or stylistic categories that sometimes lie at the crossroads of different conceptual as well as formal positions, as is the case with art feminism, appropriation art, and installation art.

Furthermore, it is relevant to ponder the possible implications of widespread technological practices peculiar to the field of video art that might hinder the employment of distant viewing or adulterate its results (Kimmelmann 1998). Technical manipulation of moving images and cross-media adoption of source material lie at the heart of video art's history since its inception, though they could severely distort the temporal scope of computer vision if not properly trained (Tilton and Taylor 2019: 7). Distant viewing could thus constitute a new way to perform video art exegesis and, perhaps, offer a different perspective on the medium, depending on the extent at which museums and collections will open their datasets. The following paragraphs shall explore the institutional and digital conditions for the employment of distant viewing, as well as ponder on the specificity of video art as a technological and aesthetic medium to be seized by means of computer vision methodology.

## 2 Structural Provisions and Archival Conditions for an Analytical Hypothesis

Given the varying technological stances of video art and the uncertain nature of the term itself in visual arts scholarship, a round of definitory reflections is due. It is necessary to identify the exact object of analysis and it better be something that can effectively undergo a computational evaluation. Finding a unitary definition of video art – although a quite reductionist one – already hints at the questions that can be asked by means of computer vision and at the analytical procedures allowed by distant viewing. Considering the many contrasting views in recent scholarly writing, there is a variety of definitions to resort to when trying to set boundaries to the field of video art. Furthermore, the use of video technology found acceptance in the visual arts much quicker than earlier technologies such as photography (Newman 2009). Indeed, video came at a convenient time finding wider acceptance due to the dawn of Postmodernism that disrupted longstanding conceptions of High Art, introducing new materials, processes, and devices (Foster 1996, 1–71). Nevertheless, since its inception there has hardly been a unitary approach, since visual artists resorted to video technology according to their peculiar understanding of the artistic process. Already pioneers such as Nam June Paik, Wolf Vostell, Joan Jonas, Dan Graham, Susan Hiller, and Bruce Nauman embraced it for completely different aims, ranging from the mere recording of performances to the creation of complex environmental installations, or going as far as the disruption of our temporal appreciation. Even works that appear structurally similar – such as multiple-channel video-projections in a closed environment – can have completely different conceptual objectives and practical effects, like a subtle perceptual estrangement in Graham's *Time Delay Room* (1974), a harsh psychological involvement in Hiller's *An Entertainment* (1990), or an annoying repetition in Nauman's *Anthro/Socio* (1992) (Schwarze 2012). Subsequent generations of technology natives employed the moving image with open-ended processes superseding the limits of human condition embracing computer generated contents as with Atkins' avatar in *Warm, Warm Warm Spring Mouths* (2013) or allowing direct interaction as in Perry's *Graft and Ash for a Three Monitor Workstation* (2016) (Bier 2013, Schriber 2017).

Considering the variety of techniques and processes developed over just a few decades, it comes as no surprise that scholars are still struggling to find a unitary definition to this artistic phenomenon. Perhaps the most restrictive perspective is that of Michael Fried, who applies the term photography to all mechanical extensions of technically reproduced pictures, be they still frames or moving images (Fried 2008). Although historically penned to the second half of the 20th century and rather bound to analogue technologies, the term video art still finds wide acceptance and is used by authors such as Michael Rush and Tanya Leighton to highlight the predominance of the audio-visual component in the referenced artworks, even if they comprise articulated spatial installations (Rush 2007, Leighton 2008). The latter's counterpart in the new millennium is the label Digital Art, a term which Oliver Grau praises for the ability to evoke "interaction and variability, simultaneity, sequentiality and narration, connected polydimensional visual spaces of happening, experience and immersion" (Grau 2017: 109). A broader term is media art, which avoids emphasising technological distinctions and rather points at the intermingle of novel and obsolete devices or interfaces used in its production, an intermingle that Rosalind Krauss defines *post-medium condition* (Krauss 2000). Following this line of thought, Cosetta Saba stresses that, in order to analyse media artworks, they should be broken down into descriptive elements, which constitute documentary traces thereof (Saba 2013: 102). Faced with practical problems of archiving information, compiling metadata, and storing files, archives of media art indeed set the conceptual framework and processual standards for the reduction of such artworks to a set of instructions, images, or textual elements (Cocciolo 2014: 247). To apply computer vision to media art, it is relevant to acknowledge that

artworks in this domain are archived by means of the construction of information objects, which are digital documents that preserve the technological and material dimensions of these complex works (Saba 2013: 104). It is further important to stress that such practices are not epistemologically neutral but also retain a glimpse of the cultural contexts in which these works emerged and were experienced (Foster 2002: 81–22). Over the decades, digital archiving in the case of media art allowed to spread a set of documentary practices, which through shared protocols could potentially store and process any artwork despite its complexity, as well as guarantee the preservation and migration of its data and metadata (Segre 1979).

Archival practices condition the possibility to employ computer vision to media art, in so far as they clarify that major definitory differences and simple typological distinctions may cause an under-appreciation or even a full loss in the interdisciplinary nature of such artworks (Saba 2013: 111). To allow big data analysis, it is thus unavoidable to reduce media art's characteristics to a relatively stable dimension that can be seized with computational tools, although risking losing a single artwork's complexity and integrity (Grau 2017: 101-102). Particularly in the case of distant viewing, what counts is the presence of moving images, a countless amount of them, regardless of the peculiar technology adopted for filming or displaying, as well as heedless of whether the audio-visual content is the sole component of a work or rather a part of a larger installation or environment. For this very reason, the term applied to the object of analysis for the scope of this paper shall be video art, thus stressing the temporal and audio-visual dimension of these artworks that can be conveniently addressed with computer vision (Wang and Gupta 2015: 2796). This kind of reductionist approach comprehends single-channel video, in particular, and may cater to an analytical purpose, but it is consistent with the practices of digital platforms providing access and dissemination to such art forms (Blume 2017). Web archives and other sorts of digital repositories necessarily reduce media artworks to their audio-visual components, be they commercial and authorised such as Daata Streaming Platform or be they public and unauthorised like UbuWeb.[1] This is a central argument for employing distant viewing techniques, despite their focus on certain aspects of this art form that necessarily leaves the complexity of larger environmental installations aside. Hence, this specific interpretation of the term video art may seem reductionist, but it stems from an empirical observation grounded in the theoretical categorisation and digital organisation that has emerged autonomously in the art field itself (Becker 2008: 39).

If anything, the problem regards availability of source material, in a twofold way: on the one hand there are insufficient open platforms providing authenticated video art, like the JSC Video Lounge,[2] while on the other hand major projects for infrastructures, networks, and virtual museums of media art have been terminated due to a lack of financing (Grau 2017: 113). This contrasts with the widespread dissemination of two-dimensional artworks, which over the last decade saw the participation of major museums providing high quality files under Creative Commons (Navarrete and Villaespesa 2020). Of course it is possible to retrieve much video art on YouTube, Vimeo or other streaming platforms, but the provenance and integrity of the material is hardly verifiable, as can be tested in the case of Eduardo Paolozzi's experimental films *History of Nothing* (1962) and *Kakafon Kakkoon* (1965).[3] A preliminary condition to start using computer vision in the case of video art definitely comprises the creation of open access repositories that certify the wholeness of content and structural integrity of the preserved audio-visual material.

## 3 Potential and Limitations of Computer Vision in the Domain of Visual Arts Studies

Having identified video art as the object of computer vision – a specific kind thereof being single-channel videos – and having assessed the premises for big data analysis in the case of video art, it is now useful to understand what art history can expect from the employment of distant viewing to this visual art form. In fact, despite the reticence of traditional art historians to adopt digital tools, the last decade provided utter enthu-

1. Ref.: https://daata.art/ and https://www.ubu.com/

2. Ref.: https://www.jsc.art/jsc-video-lounge/

3. In this instance, the YouTube entry does not show any provenance information: https://www.youtube.com/watch?v=mkuJD-7vbt4&list=PL7LjOUbyOfgyKeeHuFO0RmQg0FyWThtC4 (Last accessed: 22-06-2021).

siasm across visual arts scholarship for the analytical potential of computer vision deployed at ever growing rates in the service of connoisseurship.[4] The quantitative advantage to swiftly process or retrieve thousands of images and bring them into visual correspondence proved very promising to assist the recognition of similarities and support the authentication of originals (Bell and Ommer 2016: 188). Analysing huge datasets promised to relativise established canons and categories, although case studies were mostly focused on sampling pre-classified material such that the thematic collections were intrinsically biased, at least in part (Bender 2015: 101–106). Especially in art historical analysis early computational experiments demonstrated hierarchy-aware classifications allow obtaining more informative results (Deng, Berg, Li and Fei-Fei 2010: 72). Adopting commercial software solutions in the digital humanities proved hardly valuable, given that art historic image processing needs specific competences to steer computational tools towards the resolution of art historical problems (Bell Ommer 2016: 189). Hence, most art historians working on computer vision called for establishing code systems for specific computational interpretations by means of metadata extraction algorithms prior to exploratory data analysis of visual corpora, and not the other way round (Taylor and Tilton 2019: 11).

Digital art historians agree that computational issues need to be approached at the outset of algorithm design before classifying large scale datasets, because their evaluation would otherwise be infeasible (Deng, Berg, Li and Fei-Fei 2010: 76). Quantitative research in this domain clearly provides a type of data, which is independent of interpretations, although it must rely on selected populations of artworks or thematic research collections (Moretti 2007: 119). As a matter of fact, distant viewing itself sets the framework for interpretative tasks that need some sort of supervised, assisted, or automated coding function (Taylor and Tilton 2019: 12). A paramount example of what can be achieved in digital art history are the widely known experiments of Lev Manovich, who used preconfigured form and colour categories to train image matching machines in clustering the known paintings of Vincent van Gogh, in order to recognise his stylistic evolvement from the Paris period to his stay in Arles (Manovich 2013). Computer vision recognises similarities even of other kinds – such as posture, viewpoint, texture, and iconological components – though being reliable mostly on shape-based and appearance-based relationships (Esser, Sutter and Ommer 2018: 8864). Nevertheless, shape-understanding and synthesis enables robust retrieval practices on a range of varied digital corpora, even in the case of low resolution images (Esser, Rombach and Ommer 2021: 12879). Image recognition steered by convolutional neural networks encompasses specific tasks, which include classification, detection, viewpoint understanding, segmentation, verification, and many more that are useful to assign unique labels even in the case of multiple instances (Deng, Berg, Li and Fei-Fei 2010: 72). So far, the most relevant results in art historical research were obtained mainly by pre-emptive work of labelling and annotating data, which in turn helped to fine-tune image retrieval on subsets and to train neural networks towards self-supervision (Monroy, Bell and Ommer 2014: 420–422).

Numerous experiments by digital art historians cooperating with IT specialists demonstrated that computer vision can retrieve details or motives across different techniques, as well as to find variations in a particular iconography and identify the characteristic style of an artist or workshop (Bell and Ommer 2016: 193). Training machines to become modern connoisseurs went so far as to create commercial softwares such as ArtPI that consists of an Application Programming Interface optimised for art recognition through deep learning models aware of the concepts of period, genre, subject matter, composition, light, space, colour, and other elements distinctive to the visual arts.[5] Leaving aside the question of the right degree of supervised or unsupervised learning to train artificial intelligence in the case of art, such ambitious projects clearly revolve around visual similarity, thus relying completely on iconographic classifiers like colour, shape, and texture to produce a topography of similarities, which could help identifying even anonymous artworks (Bell and Ommer 2016: 191). Insofar as computer vision can handle unreliable similarities and consider the multivariate nature of similarity, for instance by incorporating partonomy and context, images can be rendered meaningful through

4. In this instance, see the symposium 'SEARCHING THROUGH SEEING: OPTIMIZING COMPUTER VISION TECHNOLOGY FOR THE ARTS' presented by The Frick Collection and the Frick Art Reference Library on Thursday and Friday, April 12-13, 2018. https://www.youtube.com/watch?v=sZJcCtAgMpo&t=11s (Last accessed: 15-05-2021).

5. Ref.: Speech by Ahmed Elgammal: "ArtPI–The Art API: Artificial Intelligence for Art Recognition" (April 13, 2018) as part of a series of lectures from the symposium 'SEARCHING THROUGH SEEING: OPTIMIZING COMPUTER VISION TECHNOLOGY FOR THE ARTS' presented by The Frick Collection and the Frick Art Reference Library. https://www.youtube.com/watch?v=jq7cDXG-zqU&t=12s (Last accessed: 10-06-2021).

big data analysis at a pre-linguistic level (Scott 1999: 20). This sets the use of computer vision predominantly at an iconographic level, thus offering a kind of interpretation of visual materials that may fall short of decoding deeper layers of meaning (Taylor and Tilton 2019: 2). Transferring these computational models to video art analysis, they might be relevant to investigate image composition and tonal patterns as well, while the interest in defining attribution and detecting provenance might perhaps be of little value given the relatively short history of the field. However, computer vision could be trained to recognise technical elements that give a better insight into aesthetic matters pertaining to the domain of moving image alone, such as the use of blurred transitions, zooming in and out, deliberately employing low-definition images or framing out of focus. As a matter of fact, such fine-grained details are not just a matter of iconography, but rather extracted semantics that hint at meaningful aesthetic elements akin to Aby Warburg's *Pathosformeln* (Collins, Bala, Price and Süsstrunk 2020: 5777). This could lead to recognise signature style, canonised signifiers and period related schemes, although these very terms need to be contextualised as they are still a matter of debate among art historians (Blunk and Michalsky 2018: 8–9). What could emerge is a topography of similarities that might contribute to this scholarly discussion, perhaps shifting opinions on what constitutes style or which period schemes can be drawn.

If transferred to the analysis of video art, this would mean that organising thematic collections complying with the conditions of homogeneity and size could be of some advantage for the fruitful employment of computer vision (Bender 2015: 108). Grouping or comparing the works of artists that scholars already put together content-wise or in accordance with an aesthetic relationship, for instance, might show unexpected as well as deeper connections in form, structure, and narrative. In this regard, parallels could be drawn between artists making frequent use of blurred or disrupted materials, such as with Susan Hiller's and Douglas Gordon's enlarged frames, or others centred on shootings of task accomplishments, like in many videos of Bruce Nauman and Cheryl Donegan. Moreover, a formal analysis may compare works that are thematically akin, like with Pipilotti Rist and Mona Hatoum insisting on the female body, or with Matthew Barney and Stan Douglas exploiting the aesthetics of music videoclips. Eventually, colour gradients and pose detection would make good categories to discern the monumental videos of Bill Viola and Charles Atlas. A better understanding of the underlying regularity or differentiation in such audio-visual corpora could give rise to unprecedented questions and offer new ways to decipher the cross-contamination of artistic productivity in video art, laying at the crossroads of visual arts and mass media. Hence, to conduct a computational analysis on large collections of audio-visual materials, a code system must be developed to fill the semantic gap between elements contained in the raw image and the extracted syntax used in its digital representation, even though it might start from a biased collection labelled with pre-emptive metadata (Taylor and Tilton 2019: 3).

## 4 Video Parsing, Visual Tracking, Action Detecting and the Benefits of Distant Viewing

Analysing two-dimensional artworks and still frames proved challenging enough for digital art historians, but parsing video art only makes things more complicated given the amount of (moving) images to be processed through computer vision. Over the last decade researchers approached video analysis to teach machines a kind of feature representation that superseded static images, since visual tracking is indeed one of the first capabilities that human infants develop, often before semantic representations are learned (Wang and Gupta 2015: 2795). Another quest was exploring at a computational level what different dimensions were at stake in the evaluation of salient-object detection within a set of images, which comprises several tasks such as object localisation, generic target detection, visual description, compression, and segmentation (Yildirim, Sen, Kankanhalli and Süsstrunk 2020: 2259). Consequently, scholars first focussed on procedural questions to understand how to employ computational methods in video parsing and, subsequently, how to train computer vision to recognise relevant features. Several approaches tried to exploit sports datasets because of the temporal information and spatial context contained in the actions captured in these videos, which allowed performing posture analysis and retrieval by learning a general and transferable feature representation for human poses (Sanakoyeu, Bautista and Ommer 2018: 332–335). However, this research was limited to the recognition and classification of single actions in pre-segmented clips, while real-world video streams, full-length films, and video artworks comprise multiple actions with variable durations and start/end times, such that recent

experiments tried to split instances in a set of ordered sub-actions to let internal temporal structure emerge (Pirsiavash and Ramanan 2014: 612). To avoid the manual labelling of millions of images, a useful technique of video analysis proved to be visual tracking of moving patches through subsequent frames, which allows convolutional neural networks to perform unsupervised object and action detection (Wang and Gupta 2015: 2796).

Interestingly, the concerted effort of various research groups was set on transferring knowledge to machines in such a way that video analysis could abandon the Mechanical Turk and become unsupervised, hence adapting deep learning methods derived from semantic domains. Performance improvements were initially obtained through supervised labelling of image similarities on millions of image samples, often using large labelled datasets, which however were not designed for artistic purposes, such as PASCAL VOC, HOG-LDA, Imagenet, and Alexnet (Sanakoyeu, Bautista and Ommer 2018: 341). Until recently semantic supervision on videos was considered a strong requirement for training neural networks, but visual tracking and action parsing proved instrumental to proceed towards an automated iconographic and structural recognition that opens to the possibility of interpreting deeper layers of meaning (Wang and Gupta 2015: 2801). Focusing on efficient segmental regular grammars and capturing temporal action constraints also resulted in quite practical tools for the unsupervised analysis of long-scale videos (Pirsiavash and Ramanan 2014: 618). Starting with annotated big data, neural networks were trained in video recognition up to a point to overcome noisy scenarios and perform event and face detection, text overlay and scene text differentiation, captioning of single objects as well as of groups of people (Bartz, Herold, Yang and Meinel 2017: 887). The future looks promising for the employment of distant viewing to video art inquiry, provided that digital art historians consider it a methodological and theoretical framework for moving image interpretation – that is, for studying and coding large collections of visual material through the extraction of semantic metadata (Taylor and Tilton 2019: 1–2). In doing so, distant viewing could help aggregating and interpreting elements that are meaningful for video art at an aesthetic level or content-wise, such as detection of camera movement and moving image shot breaks, recognition of lighting position and dominant colour pallets, object localisation and facial recognition (Redmon, Divvala, Girshick and Farhadi 2016). However, the detectable features refer to a very specific kind of video art, which are single channel videos containing figurative elements set in a real-world scenario.[6] In fact, computer vision is not yet sufficiently trained to make meaning of collage-like compositions, as can be found in Eduardo Paolozzi's *History of Nothing* (1962) (Stallschus 2019: 110–114), or image palimpsests like those in Andy Warhol's experiments with Norleco recorders for *Outer and Inner Space* (1965) (Mantoan 2018: 106-107).

Despite the mentioned limitations, at a morphological level distant viewing could offer insight in the representation of content and the stylisation process of an artist's work, given that videos of the same author can be analysed in a bulk and compared to that of others, be they video artists or even filmmakers (Kotovenko, Sanakoyeu, Lang and Ommer 2019: 4429). In this regard, pivotal experiments conducted by Arnold Taylor and Lauren Tilton with shot detection algorithms on homogenous corpora of television shows from the Network Era in the USA (1954–1975) allowed the identification of relevant differences in the aesthetic aspects and narrative arcs of situational comedies that cultural studies formerly believed to serve similar cultural purposes (Taylor and Tilton 2019: 6–7). To test this approach, they used two comedy shows of the 1960s with leading female characters – *Bewitched*, running from 1964 to 1972, and *I Dream of Jeannie*, aired from 1965 to 1970 – such that they were supposedly comparable contentwise. In trying to track the quantitative appearance of the leading characters throughout each episode, the results showed completely different narrative structures that had different interpretations at a cultural level regarding the traditional role of women in American society. However, it is important to stress that distant viewing always requires a code system that is culturally and socially constructed to analyse and explore artistic features, which in turn can be interpreted as proper *Pathosformeln*. As a tool for video art exegesis, the approach is encouraging, provided researchers ask meaningful questions and use the quantitative results of this digital picklock as basis for interpretative tasks to be carried out mutually by art historians and IT experts.

6. Ref. Speech by Christoph Meinel: "Machine Learning: The Reality Behind Artificial Intelligence" (April 12, 2018), as part of a series of lectures from the symposium 'SEARCHING THROUGH SEEING: OPTIMIZING COMPUTER VISION TECHNOLOGY FOR THE ARTS' presented by The Frick Collection and the Frick Art Reference Library (Last accessed: 10-06-2021).

## 5 Some Conclusions and the Interpretative Questions to be Finally Asked

Without a doubt, computer vision opened exciting new perspectives for art historians willing to embrace quantitative analysis, especially for the sake of iconographic understanding, recognition, and distribution among big data collections. Going a step further into image interpretation, latest research showed the potential of distant viewing in helping to bridge the semantic gap, thus pushing towards the need for a coding system of cultural and social relevance. While scholars are still debating whether it is likely – or even desirable – to train and fine-tune machines for complete unsupervised tasks in the domain of image recognition and interpretation, in visual arts what appears to really count are the questions that need be asked (Sanakoyeu, Bautista and Ommer 2018: 341). So far computer vision cannot but tentatively substitute human expertise in the exegesis of artistic products, as it rather assists in the queries that hold meaning for a specific coded system. Considering the case of video art, distant viewing may help building more robust categories, finding similarities between artists, or even detecting recurring aesthetic patterns across artist generations. Furthermore, it can help to highlight the aesthetic and technical elements in the visual and narrative construction that constitute the signature style of specific video artists (Porton and Hogg 2019). Transferring pose, genre, and style recognition from pre-Modern art historical datasets of two-dimensional works could also be fruitful, because automatic retrieval tools could then compare it to video artworks and detect formal consistencies that may belong to long-lasting iconographic formulae in the artistic field. Of even bigger advantage could be a comparison with moving image studies, such as to find formal parallels and subject similarities with television shows or feature films, which constitute a moving image repertoire that many video artists eagerly exploit. Drawing parallels to art history and moving image scholarship against the background of quantitative visual research should allow for a better understanding of the peculiar niche that video art carved out for itself, halfway between visual arts and mass media. What could emerge from distant viewing are especially recognisable *Pathosformeln* – be they centred on human poses and gestures, or rather relying on visual technicalities such as scene transition, blurring, camera movement – that make explicit reference to art historical tradition or the wider media culture. Such *Pathosformeln* need not be devised in advance, at least not fully, but could surface from loosely labelled image samples applied through distant viewing on large corpora of video artworks (Bell and Ommer 2015: 418). Of some benefit could eventually be the inclusion in the formal repertoire of video art analysis also of alternative practices, such as the kind of net art produced in image-laden blogs like *Screenfull* (2004) by Abe Linkoln and Jimpunk, because this would allow to explore the permeability of iconographic references across different media employed in the visual arts.

The final assumption is that distant viewing will allow to detect meaningful patterns in video art, but the biggest limitation at present it still the lack of available material. If museums do not provide open access to high quality videos, scholars can rely on very narrow datasets, whose provenance is sometimes even uncertain. There are of course a few reliable video art repositories, but they usually display a handful of video artworks of specific artists, which makes it impossible to analyse an artist's entire oeuvre. Another limitation surely is the fact that distant viewing can be employed merely to audio-visual material, which makes it impossible to grasp the complexity of environmental installations and the multi-sensory experience triggered in the audience. Despite these difficulties, the opportunities provided by distant viewing to video art exegesis remain considerable, provided art historians employ it for problems, which this digital tool can conceptually and technically handle.

## References

Bartz, Christian, Tom Herold, Haojin Yang and Cristoph Meinel (2017). "Language Identification Using Deep Convolutional Recurrent Neural Networks." In *Neural Information Processing. ICONIP 2017*, edited by Derong Liu, Shengli Xie, Yuanqing Li, Dongbin Zhao and El-Sayed M. El-Alfy (Lecture Notes in Computer Science, 10639), 880–889. Cham: Springer. https://doi.org/10.1007/978-3-319-70136-3_93.

Becker, Howard (2008). *Art Worlds*. Berkeley, University of California Press.

Bell, Peter and Björn Ommer (2015). "Training Argus. Ansätze zum automatischen Sehen in der Kunstgeschichte." *Kunstchronik* 68(8), 414–420.

Bell, Peter and Björn Ommer (2016). "Digital Connoisseur? How Computer Vision Supports Art History." In *Il metedo del conoscitore: Approcci, limiti, prospettive*, edited by Stefan Albl and Alina Aggujaro, 187–200. Rome: Editoriale Artemide.

Bender, K. (2015). "Distant Viewing in Art History: A Case Study of Artistic Productivity." *Digital Art History Journal* 1(June). https://doi.org/10.11588/dah.2015.1.21639.

Bier, Arielle (2013). "Reviews: Ed Atkins." *Frieze* (10). https://www.frieze.com/article/ed-atkins-1. (Last accessed 20-06-2021).

Blume, Jonas (2017). "Exploring the Potentials and Challenges of Virtual Distribution of Contemporary Art." In *Digital Environments*, edited by Urte Undine Frömming, Steffen Köhn, Samantha Fox and Mike Terry, 97–116. Bielefeld: Transcript Verlag.

Blunk, Julian and Tanja Michalski (editors) (2018). "Stil als (geistiges) Eigentum". *STUDI DELLA BIBLIOTHECA HERTZIANA* 43: 7-16. Munich: Hirmer Verlag

Cocciolo, Anthony (2014). "Challenges to born-digital institutional archiving: the case of a New York art museum". *Records Management Journal*, 24(3), 238–250. https://doi.org/10.1108/RMJ-04-2014-0023.

Collins, Edo, Raja Bala, Bob Price and Sabine Süsstrunk (2020). "Editing in Style: Uncovering the Local Semantics of GANs." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5770-5779. https://doi.org/10.1109/CVPR42600.2020.00581.

Comer, Stuart (Ed.) (2009). *Film And Video Art*. London: Tate Publishing.

Deng, Jia, Alexander Berg, Kai Li and Li Fei-Fei (2010). "What Does Classifying More Than 10,000 Image Categories Tell Us?" *ECCV*, 71-84. https://doi.org/10.1007/978-3-642-15555-0_6.

Esser, Patrick, Ekaterina Sutter and Björn Ommer (2018). "A Variational U-Net for Conditional Appearance and Shape Generation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 8857-8866.

Esser, Patrick, Robin Rombach and Björn Ommer (2021). "Taming Transformers for High-Resolution Image Synthesis." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 12873-12883.

Foster, Hal (1996). *The Return of the Real: The Avant-Garde at the End of the Century*. Cambridge: MIT Press.

Foster, Hal (2002). "Archives of Modern Art." *October* 99: 81-95.

Fried, Michael (2008). *Why Photography Matters As Art As Never Before*. New Haven: Yale University Press.

Frohen, Ursula (2008). Dissolution of the Frame: Immersion and Participation in Video Installations. In *Art and the Moving Image: A Critical Reader*, edited by Tanya Leighton, 355–370. London: Tate Publishing.

Grau, Oliver (2017). "Digital Art's Complex Expression and Its Impact on Archives and Humanities." In *Museum and Archive on the Move*, edited by Oliver Grau, 99–117. Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110529630-007.

Heiser, Jörg (2008). All of a sudden: Things that Matter in Contemporary Art. New York: Sternberg Press.

Kimmelmann, Michael (1998). "Installation Art Moves On, Moves In." *New York Times* (August, 9), 1.

Kotovenko, Dmytro, Artsiom Sanakoyeu, Sabine Lang and Björn Ommer (2019). "Content and Style Disentanglement for Artistic Style Transfer." *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 4422-4431.

Krauss, Rosalind (2000). *A Voyage on the North Sea. Art in the Age of the Post-Medium Condition*. London: Thames & Hudson.

Leighton, Tanya (2008). *Art and the Moving Image: A Critical Reader*. London: Tate Publishing.

Manovich, Lev (2013). "Museum without Walls, Art History without Names: Visualization Methods for Humanities and Media Studies." In *Oxford Handbook of Sound and Image in Digital Media*, edited by Carol Vernallis, Amy Herzog and John Richardson. Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199757640.013.005.

Mantoan, Diego (2018). "Video art between aesthetic maturity and medium immersion: Patterns of change and generational shift in moving image technology". In *Immersion – Design – Art, Revisited: Transmedia Form Principles in Contemporary Art and Technology*, edited by Lars C. Grabbe, Patrick Rupert-Kruse and Norbert M. Schmitz, 98–117. Marburg, Büchner Verlag.

Monroy, Antonio, Peter Bell and Björn Ommer (2014). "Morphological analysis for investigating artistic images." *Image and Vision Computing* 32: 414–423.

Moretti, Franco (2007). *Graphs, Maps, Trees – Abstract Models for Literary History*. London, New York: Verso.

Navarrete, Trilce, and Elena Villaespesa (2020). "Digital Heritage Consumption: The Case of the Metropolitan Museum of Art". *magazén* 1(2): 223-248. https://doi.org/10.30687/mag/2724-3923/2020/02/004.

Newman, Michael (2009). "Moving the Image in the Gallery since the 1990s". In *Film And Video Art*, edited by Stuart Comer, 86–121. London: Tate Publishing.

Porton, Richard, and Joanna Hogg (2019). "Coming of Age in Knightsbridge: An Interview with Joanna Hogg." *Cinéaste* 44(4): 4-7.

Redmon, Joseph, Santosh Divvala, Ross Girshick and Ali Farhadi (2016). *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788.

Rush, Michael (2007). *Video Art: Revised Edition*. London: Thames & Hudson.

Saba, Cosetta G. (2013). "Media Art and the Digital Archive". In *Preserving and Exhibiting Media Art Book: Challenges and Perspectives*, edited by Julia Noordegraaf, Cosetta G. Saba, Barbara Le Maître and Vinzenz Hediger, 102–121. Amsterdam: Amsterdam University Press.

Sanakoyeu, Artsiom, Miguel A. Bautista and Björn Ommer (2018). "Deep unsupervised learning of visual similarities." *Pattern Recognition* 78(C–June 2018), 331–343. https://doi.org/10.1016/j.patcog.2018.01.036.

Schriber, Abbe (2017). "Sondra Perry. The Kitchen." *Artforum* (March). https://www.artforum.com/print/reviews/201703/sondra-perry-66704 (Last accessed 20-06-2021).

Schwarze, Dirk (2012). *Meilensteine: Die Documenta 1–13*. Berlin: Verlag B&S.

Scott, Clive (1999). *The Spoken Image: Photography and Language*. London: Reaktion Books.

Segre, Cesare (1979). *Semiotica filologica. Testo e modelli culturali*. Turin: Einaudi.

Stallschus, Stefanie (2019). "Poetic Metaphor: Paolozzi's Animated Films and Their Relation to Wittgenstein." In *Paolozzi and Wittgenstein. The Artist and the Philosopher*, edited by Diego Mantoan and Luigi Perissinotto, 109–124. London: Palgrave MacMillan.

Taylor, Arnold and Lauren Tilton (2019). "Distant viewing: analyzing large visual corpora." *Digital Scholarship in the Humanities* 0(0). https://doi.org/10.1093/digitalsh/fqz013.

Wang, Xiaolong and Abhinav Gupta (2015). "Unsupervised learning of visual representations using videos." *Proceedings of the 2015 IEEE International Conference on Computer Vision*, 2794–2802.

Yildirim, Gökhan, Debashis Sen,Mohan Kankanhalli and Sabine Süsstrunk (2020). "Evaluating salient object detection in natural images with multiple objects having multi-level saliency". *IET Image Process* 14(10), 2249-2262. https://doi.org/10.1049/iet-ipr.2019.0787.

**Diego Mantoan** – Ca' Foscari University of Venice (Italy)
✉ diego.mantoan@unive.it
Diego Mantoan is Assistant Professor in digital and public art at Ca' Foscari University of Venice with a PhD at FUBerlin. He is a founding member of the Venice Centre for Digital & Public Humanities, Advisory Editor of Vernon Press Academic (USA), Associate Editor of *magazén* (Italy). Among his books, *The Road to Parnassus* (Vernon Press 2015) was long-listed for the Berger Prize 2016 and recently edited *Paolozzi & Wittgenstein* (Palgrave MacMillan 2019). He delivered speeches at UBern, Bibliotheca Hertziana, UCL, VUAmsterdam, Sotheby's Institute of Art, NYU, Galerie Belvedere, maat Lisboa. He was director assistant at Venice Biennale, later working as art archive curator for Douglas Gordon Studio (Berlin), Sigmar Polke Estate (Cologne), Julia Stoschek Collection (Düsseldorf) and Museo Mario Rimoldi (Cortina).