Computer vision, human senses, and language of art

Lev Manovich

AI & SOCIETY
Journal of Knowledge, Culture and Communication

ISSN 0951-5666

AI & Soc
DOI 10.1007/s00146-020-01094-9

Springer

AI & SOCIETY
https://doi.org/10.1007/s00146-020-01094-9

ORIGINAL ARTICLE

Computer vision, human senses, and language of art

Lev Manovich[1]

Received: 1 June 2020 / Accepted: 14 October 2020

Abstract
What is the most important reason for using Computer Vision methods in humanities research? In this article, I argue that the use of numerical representation and data analysis methods offers a new language for describing cultural artifacts, experiences and dynamics. The human languages such as English or Russian that developed rather recently in human evolution are not good at capturing analog properties of human sensorial and cultural experiences. These limitations become particularly worrying if we want to compare thousands, millions or billions of artifacts—i.e. to study contemporary media and cultures at their new twenty-first century scale. When we instead use numerical measurements of image properties standard in Computer Vision, we can better capture details of a single artifact as well as visual differences between a number of artifacts–even if they are very small. The examples of visual dimensions that numbers can capture better then languages include color, shape, texture, contours, composition, and visual characteristics of represented faces, bodies and objects. The methods of finding structures and relationships in large numerical datasets developed in statistics and machine learning allow us to extend this analysis to very big datasets of cultural objects. Equally importantly, numerical image features used in Computer Vision also give us a new language to represent gradual and continuous temporal changes—something which natural languages are also bad at. This applies to both single artworks such as a film or a dance piece (describing movement and rhythm) and also to changes in visual characteristics in millions of artifacts over decades or centuries.

Keywords Computer vision · Digital humanities · Cultural analytics · Language of art

1 Computer vision and digital humanities

Researches in humanities research, write and argue about cultural images. They analyze and interpret content, visual style, author's intentions, audience reception, meanings, emotional effects, and other aspects of images' creation and circulation. Researchers in Computer Vision field also work with images, but their goals are very different—to teach computers to automatically understand images and enable automatic actions using visual information. The examples of these applications include their use in self-driving cars, industrial and home robots, medical diagnostics, content-based image retrieval.

What are the intellectual consequences of adopting Computer Vision methods in humanities research? What happens to humanists' understanding of images and assumptions about how to describe and study visual cultures in this meeting? How can we bring together assumptions and goals of AI research in general and the assumptions and goals of the humanities that think of the study of cultural artifacts as their exclusive domain? (In addition to humanities fields such as art history, musicology, performance studies, cinema studies, literary studies, digital culture studies and game studies, these questions are also relevant for social science fields that deal with visual culture such as cultural anthropology, sociology, culture studies, communication and media studies.)

In this article, I will discuss the most important consequence of using Computer Vision in humanities, as I see it. Certainly, the achievements of Computer Vision such as detection of objects and scene types, people, and faces, pose estimation or optical character recognition all have their uses in art history, cinema and media studies, game studies, archeology, and so on. Working with the researchers in these fields, computer scientists also develop new tools for specific problems (Visart 2018). These applications and tools allow answering existing and generating new questions, and this work is certainly important. But in my view, they don't affect

✉ Lev Manovich
manovich.lev@gmail.com

[1] Program in Computer Science, The Graduate Center, City University of New York, New York, USA

Published online: 22 November 2020

in a fundamental way how we see images in humanities. What does affects this is the way Computer Vision describes images, as I will explain below.

2 Computer vision and digital humanities

We can find examples of using computational techniques to analyze single artworks or small groups of artworks carried out already for a number of decades. In the case of visual arts, such work was aimed to help in restoration, conservation, material and structure characterization, authentication, and dating. It made a good use of Digital Image Processing techniques, but it did not challenge existing methods of classifying, describing and narrating and exhibiting art. We can find a similar logic in other fields that use computational methods such as archeology. For example, one recent paper presents a method for automatically fitting together available artifact pieces together. This is a useful application for archeology, but it does not lead to big new ideas for the field (Derech et al. 2018).

While these applications were and continue to be dominant, some researchers were also using methods from Image Processing, Computer Vision and Computer Graphics to do something new for art history—come up with mathematical descriptions of various characteristics of art images such as brushstrokes, lighting, and composition. As the key researcher in the area David G. Stork pointed out in his 2009 overview of this research, "In some circumstances, computers can analyze certain aspects of perspective, lighting, color, the subtleties of the shapes of brush strokes better than even a trained art scholar, artist, or connoisseur. Rather than replacing connoisseurship, these methods—like other scientific methods such as imaging and material studies—hold promise to enhance and extend it, just as microscopes extend the powers of biologists" (Stork 2009). Today the use of computers to mathematically describe cultural artifacts and analyze quantitatively and interpret cultural patterns based on such descriptions has become popular in some areas of humanities such as literary studies and history. However, this did not happen yet on any significant scale in art history, film and media studies, game studies or other fields that analyze visual culture. However, there have been a few inspiriting researches projects done by computer scientists working together with humanists. Among them, I want in particular mention work by Impett and Moretti (2017). They carefully translate ideas of early twentieth century art historian Aby Warburg into an interactive tool while probing theoretically and critically Warburg ideas. Such work stands in contrast to more common references to books in art history, media production, graphic design and other fields in computer science research that borrow ideas from these books to build automatic systems.

Although the humanists that study the visual have been slow to make use of computers, many researchers in Computer Vision and also other areas of Computer Science, Artificial Intelligence and Machine Learning have started to do exactly this after 2006. The rise of social media platforms including Flickr and Instagram (launched in 2004 and 2010, respectively) and availability of their big data via APIs led to the burgeoning research presented in hundreds of thousands of conference papers and journal articles. While the earlier paradigm I described above subordinates the possibilities of Computer Vision to the goals and ways of working in art history discipline and museums, most of the relevant research in Computer Science does the opposite. The researchers take the goals of their field such as classification and prediction and apply them to a new type of data—large samples of visual user-generated content. So here cultural artifacts often are not approached in any different way than any other kind of data.

Both paradigms have their limitations. In this article, I take the outsider position. And this is why I start with the following question: How can we bring together assumptions and goals of two very areas of human knowledge which are fundamentally different, as opposed to subordinating one to another?

This article develops the following arguments: (1) Data representations of analog cultural artifacts used in Computer Vision, Music Information Retrieval, and Geospatial Computing give us a new and a better language for describing these artifacts in comparison to human natural languages; (2) These data representations are also closer to how human senses and central system encode analog signals. This provides another justification for the use of computer methods to analyze culture in general and using Computer Vision to see" visual culture in particular.

If Stork suggested that computers can analyze some aspects of art images better then human experts in some circumstances, I claim that computers are always more precise in their descriptions of characteristics of analog cultural artifacts. However, in the case of art historical images, the use of computation for analysis is one option because historical collections are small enough for us to study them directly. In the case of contemporary digital visual culture, using computer methods is the only way to see even small samples because of its scale. (Billions of images are shared every day on Facebook alone.)

My arguments presented in this article reflect my own practical experience of using Computer Vision with dozens of cultural datasets after I co-founded Cultural Analytics Lab at the University of California, San Diego (UCSD) in 2007. At that time, I defined "cultural analytics" as "the analysis of massive cultural data sets and flows using computational and visualization techniques" (Manovich 2007b). For about 10 years, our lab was the only one focusing on using

Computer Vision to study visual culture at large scale from the perspectives of humanities. To the best of my knowledge, the first such project done outside of our lab in the U.S. that received attention was only published in 2017 (Yale Digital Humanities Lab).

Cultural Analytics is only one among a number of research paradigms that emerged in the second part of the 2000s to take advantage of the availability of large cultural and social data. They include Digital Humanities, Computational Social Science, Social Computing, Digital Anthropology, Digital History, The Science of Cities, Urban Informatics, and Culturomics. In the same time, big cultural datasets started to be analyzed by computer scientists working in Machine Learning, Computer Vision, Natural Language Processing, Music Information Processing, Computer Multimedia, and also in Communication Studies. In the early 2010s, the "quantitative turn" begun in art history, and *International Journal for Digital Art History* was established in 2015. In 2020, the first large volume on digital art history was published:

The Routledge Companion to Digital Humanities and Art History (Brown 2020). In film studies, the first monograph that uses quantitative methods and data visualization to analyze works of a single film director appeared in 2019 (Heftberger 2019).

In parallel, the research in humanities using computational tools also started to grow. In 2003 it received the name Digital Humanities in 2003. In 2010s Digital Humanities kept growing, attracting more and more attention. However, the larger portion of the computational work in humanities so far focused on literary texts, historical text records and spatial data. In contrast, other types of media such as still and moving images and interactive media received relatively little attention. This situation is gradually improving but as I am writing this, analysis of visual media is still a small part of Digital Humanities (Digital Humanities Conference 2019). You can see this yourself by browsing programs of annual conferences organized by The Alliance of Digital Humanities Organizations or looking at the field journals that include *Digital Humanities Quarterly, International Journal of Digital Humanities*, and *Digital Scholarship in the Humanities.* The field limitations are well summarized by the title of the article published in 2017 in *Digital Scholarship in the Humanities*: "Digital humanities is text heavy, visualization light, and simulation poor" (Champion 2017).

This is surprising because computer scientists started to develop methods for the analysis of images already at the end of 1950s. Today they are implemented in numerous digital services and devices, including web image search engines, stand-alone photo cameras and cameras in mobile phones, widely used image editing software such as Photoshop, Pixelmator, Affinity Photo, and Luminar, image sharing services such Google Photos, and also available as programming libraries (OpenCV, MATLAB). In Computer Vision and Multimedia Computing, researchers have been publishing for many years new algorithms for automatic detection of image content, artistic styles, photographic techniques, user-generated and professional video and TV programs, and photos that are more interesting, memorable, or original than others, and applying these algorithms to progressively larger datasets (Redi et al. 2017). In our lab we have been using some of these methods to analyze many types of both historical and contemporary visual media—20,000 photographs from the collection in Museum of Modern Art (MoMA) in New York, films by the pioneer of documentary filmmaking Dziga Vertov from Austrian Film Museum, sixteen million images shared on Instagram in seventeen global cities, one million Manga pages, one million artworks from popular art network DeviantArt, and other datasets.

3 Describing images with words and numbers

Most representations of physical, biological and cultural phenomena constructed by artists, scholars and engineers so far only capture some characteristics of these phenomena. Linear perspective represents the world as seen from a human-like viewpoint, but it distorts the real proportions and positions of objects in space. Contemporary 100-megapixel photograph made with a professional camera captures details of human skin and separate hairs—but not what is inside the body under the skin.

If the artifacts are synthetic, sometimes it is easy to represent them more precisely. Engineering drawings, algorithms, manufacturing details used to construct such artifact are already their representation in the finished state—however, we can't predict human sensations and experiences of these artifacts only from these representations. But nature's engineering can be so complex that even all representational technologies at our disposal can barely capture a miniscule proportion of information. For example, currently best fMRI machines can capture the brain at a resolution of 1 mm. This may look like a small enough area—yet it contains millions of neurons and tens of billions of synapses. The most detailed map of the universe produced in 2018 by Gaia (the European Space Agency craft) shows 1.7 billion stars—but according to estimates, our own galaxy alone contains hundreds of billions of stars.

And even when we consider a single cultural artifact created by humans and existing on a human scale—a photograph you took, a mobile phone you used to take it with, or your outfit consisting from items you purchased at Zara or COS—data representations of these artifacts often can only capture some of their characteristics. In the case of a digital

photograph, we have access to all the pixels it contains. This artifact consists from 100% machine data. These pixels to us will look a bit different from one display to the next, depending on its brightness, contrast, and color temperature settings, and its technology. And if we want to edit this image, what is possible is defined by particular software. (In my article *There is Only Software* (Manovich 2009) I argued that "depending on the software I am using, the "properties" of a media object can change dramatically. Exactly the same file with the same contents can take on a variety of identities depending on the software being used.")

Digital pixel image is a synthetic artifact fully defined by only one type of data in a format ready for machine processing (e.g., an array of numbers defining pixel values). But what about physical artifacts, such as fashion designs that may use fabrics with all kinds of non-standard finishes, combine multiple materials, textures, and fabrics, and create unusual volumes? (This applies to many collections produced since the early 1990s the 1980s by Rei Kawakubo, Dres Van Noten, Maison Margiela, Raf Simons, Issey Miyake, among others, and also to many fashion designers working in countries such as South Korea today.) How do we translate cloves into data? The geometries of pattern pieces will not tell us about visual impressions of their cloves, or experience wearing them. Such garments may have unique two-dimensional and three-dimensional textures, use ornament, play with degrees of transparency, etc. And many fashion designs are only fully "realized" than you wear them, with the garment taking on particular shape and volume as you walk.

The challenge of representing the experience of material artifacts as data is not unlike calculating an average for a set of numbers. While we can always mechanically calculate an average, this average does not capture the shape of their distribution, and sometimes it is simply meaningless (Desrosières 1998). In a Gaussian distribution, most data lie close to the average, but in a binomial distribution, most data are away from it, so the average does not tell us much.

Similarly, when we try to capture our sensorial, cognitive and emotional experience of looking at or wearing a fashion garment, all methods we have available—recording heartbeat, eyes movements, brain activity, and other physiological, cognitive and affective processes, or asking a person to describe her subjective experience and fill out a questioner—can only represent some aspects of this experience.

But this does not mean that any data encoding automatically loses information, or that our intellectual machines (i.e., digital computers) are by default inferior to human machines, i.e. our senses and cognition. For example, let's say I am writing about artworks exhibited in a large art fair that features hundreds of works shown by hundreds of galleries across a large space.

What I can say depends on what I was able to see during my visit and what I remembered—and therefore constrained by the limitations of my senses, cognition, memory, and body, as well as by the language (Russian, Spanish, Indonesian, etc.) in which I write.

In the twentieth century, modern humanities, the common method of describing artifacts and experiences was to observe one own reaction as filtered by one's academic training and use natural language for describing and theorizing these experiences. In social sciences and practical fields concerned with measuring people attitudes, taste and opinions, researchers used questioners, group observations and ethnography, and these methods remain very valuable today. Meanwhile, since the 1940s engineers and scientists working with digital computers have been gradually developing a very different paradigm—describing media artifacts such as text, shapes, audio, and images via numerical features. Humanities studies of visual art, architecture, design, video games, films, user-generated video and all other visual forms can adopt the same paradigm. Why it is such a good idea? My explanation is summarized in the next paragraph.

Numerical measurements of cultural artifacts, interactions and behaviors give us a new language to talk about cultural artifacts and experiences. This language is closer to how the senses represent analog information (sounds, music, colors, spatial forms, movement, etc.) The senses translate their inputs into quantitative scales, and this is what allows us to differentiate between many more sounds, colors, movements, shapes, textures than natural languages. So, when we represent analog characteristics of artifacts, interactions and behaviors as data using numbers, we get the same advantages. This is why a language of numbers is a better fit than human languages for describing analog aspects of culture.

Using natural languages was the only mechanism humanities have been using for describing all aspects of culture until the recent emergence of Digital Humanities. *Natural* or *ordinary language* refers to a language that evolved in human evolution without planning. While the origins of natural languages are debated by sciences, many suggest that it developed somewhere between 200,000 and 50,000 years ago. Natural languages cannot represent small differences on analog dimensions which define aesthetic artifacts and experiences such as color, texture, transparency, types of surfaces and finishes, visual and temporal rhythms, movement, speed, touch, sound, taste, etc. In contrast, our senses capture such differences quite well.

Aesthetic artifacts and experiences human species were creating during many thousands of years of their cultural history exploit these abilities. In the modern period, the arts started to systematically develop new aesthetics that strives to fill every possible "cell" of a large multi-dimensional space of all sense dimensions, taking advantage of the very high fidelity and resolution of our senses. Dance innovators from Loie Fuller and Martha Graham to Pina Bausch, William Forsythe, and Cloud Gate group defined new body

movements, body positions, compositions and dynamics created by groups of dancers or by parts of a body such as fingers or speeds and types of transitions. Such dance systems are only possible because our eye and brain abilities to register tiny differences on these dimensions of dance.

In visual arts, many modern painters developed lots of variations of a *white on white* monochrome painting—images that feature only one field of a single color, or a few shapes in the same color that differ only slightly in brightness, saturation, or texture. They include Kazimir Malevich (*Suprematist Composition: White on White*, 1918), Ad Reinhardt ("black paintings"), Agnes Martin, Brice Marden, Lucio Fontana, Ives Klein, and many others.

In the twenty-first century, works by contemporary product designers often continue the explorations that preoccupied so many twentieth century artists. For example, in the second part of 2010s top companies making phones—Huawei, Xiaomi, Samsung, Apple—became obsessed with the sensory effects of their designs. The designers of phones started to develop unique surface materials, unique colors, levels of glossiness of a finish, surface roughness and waviness. As the phone moves closer and closer to becoming a pure screen a or transparent surface, this obsession with sensualizing still remaining material part may be the last stage of phone design before the phone becomes complete screen—although we may also get different form factors in the future, where small material parts become even more aestheticized (Manovich 2007c).

For instance, for its P20 phone (2017) Huawei created unique finishes each combining a range of colors. Huawei named them Morpho Aurora, Pearl White, Twilight and Pink Gold. When looking at the back of a phone at different angles, different colors would appear. (Peckham 2018). The company proudly described the technologies used to create these finishes on its website: "The Twilight and Midnight Blue HUAWEI P20 has a high-gloss finish made via a 'high-hardness' vacuum protective coating and nano-vacuum optical gradient coating." (Huawei 2019) (The P30 Mate Pro I have been using during 2019 had one of these screens.)

What about minimalism that has become the most frequently used aesthetics in the design of spaces in the early twenty-first century exemplified by all-white or raw concrete spaces, with black elements or other contrasting details? From the moment such spaces started to appear in the West in the second part of the 1990s, I have been seeking them so I can work there—hotel areas, cafes, lounges. Today you can find it everywhere from but in the late 1990s they were just a few such spaces. In my book *The Language of New Media* (Manovich 2001) completed in Fall 1999, I have thanked two such hotels because large parts of that book were written in their spaces—The Standard and Mondrian in Los Angeles. While not strictly minimalist in a classical way (they were not all white), the careful choice of textures, materials, and elimination of unnecessary details was certainly minimalist in its thinking. (Later in 2006–2007 I have been spending summers in Shanghai working on a new book and moving between a few large minimalist cafes—at that point, Shanghai had more of them then Los Angeles. Today, a city like Seoul probably has over 100,000 such cafes, each unique in its design.)

On first thought, such spatial minimalism seems to be about overwhelming our perception—asking us to stretch our limits, so to speak, to take in simultaneously black and white, big and tiny, irregular and smooth. I am thinking of famous Japanese rock gardens in Kyoto (created between 1450 and 1500), an example of *kare-sansui* ("dry landscape"): large black rocks placed in the space of tiny grey pebbles. In 1996 a store for Calvin Klein designed by London architect John Pawson opened in New York on Madison Avenue around the 60th Street, and it became very influential in the minimalist movement. Pawson was influenced by Japanese Zen Buddhism, and an article in the New York Times called his store "Less is Less." (Goldberger 1996). The photographs of the store show a large open white space with contrasting with dark wood benches (Pawson 2020). So what is going on with these examples?

I think that minimalist design uses both sensory extremes for aesthetic and spatial effect, and small subtle differences that are our senses are so good at registering. The strong contrast between black and white or smooth and textured, or wood and concrete, and so on helps us to better notice the variations in the latter—i.e., the differences in shapes of tiny pebbles in Kyoto Garden, or all white parts of the 1996 Calvin Klein store space which all have different orientations to the light coming from very large windows.

The famous early twenty-first century examples of minimalist design are all white and or silver-grey Apple products designed by Jonathan Ive in the 2000s. The first in this series was iPod in 2001, followed by PowerBook G4 in 2003, iMac G5 in 2004, and iPhone in 2007. In his article "How Steve Jobs' Love of Simplicity Fueled A Design Revolution," Walter Isaacson quotes Jobs talking about his Zen influence: "I have always found Buddhism—Japanese Zen Buddhism in particular—to be aesthetically sublime," he told me. "The most sublime thing I've ever seen are the gardens around Kyoto" (Isaacson 2012). In the most famous Kyoto garden, which I was lucky to visit, the monochrome surface made from small pebbles contrasts with a few large black rocks. In Apple products of the 2000s, the contrast between all-white object and the dark almost black screen when the device is turned off made from different material works similarly. It makes us more attentive to the roundness of the corners, the shadows from the keys, and other graduations and variations in tone and shape of the device.

In general, minimalism is everything but minimal. It would be more precise to call it "maximalism." It takes

small areas on sensory scales and expands it. It makes you see that between two grey values there are, in fact, many more variations than you knew (I call this aesthetics common today in Korea "50 shades of grey"); that the light can fall on a raw concrete surface in endless ways; that the edge in the textured paper cut into two parts by hand contains fascinating lines, volumes, and densities. Our senses delight in these discoveries. And this is likely to be one of the key functions of aesthetics in human cultures from prehistory to today—giving our sense endless exercises to register some small differences, as well as bold contrasts. And to clean visual, spatial and sound environment from everything else, so we can attend to these differences. To enjoy "less is less."

4 Human senses, numbers and the arts

For thousands of years, art and design have thrived on human abilities to discriminate between very small differences on analog dimensions of artifacts and performances, and to derive both pleasure and meaning from this. But natural languages do not contain mechanisms to represent such nuances and differences. Why? Here is my hypothesis for why this is the case.

Natural languages emerged much later in evolution than the senses—to compensate for what the latter cannot do—represent the experience of the world as categories. In other words, *human senses and natural languages are complementary systems*. Senses allow us to register tiny differences in the environment, as well as nuances of human expressions (face expressions, body movements, etc.), while languages allow us to place what we perceive into categories, to reason about these categories and communicate using them.

Evolution had no reason to duplicate the already available functions, and that is why each system is great at one thing and very poor at another. The senses developed and continued to evolve for billions of years—for instance, the first eyes developed around 500 million years ago during the Cambrian Explosion. In comparison, the rise of human languages with their categorization capacities is a very recent development (sometime between 200,000 and 50,000 years ago).

When we use a natural language as a *metalanguage* to describe and reason about an analog cultural experience, we are doing something strange: *forcing it into small number of categories which were not designed to describe it*. In fact, if we can accurately and exhaustively "put into words" an aesthetic experience, it is likely that this experience is an inferior one. In contrast, using numerical features instead of linguistic categories allows us to much better aspects of an analog experience.

Our sensors and digital computers can measure analog values with even greater precision than our senses. You may not be able to perceive a 1% difference in brightness between two image areas or 1% difference in the degree of smile between two photos of people, but computers are able to measure these differences. For example, for *Selfiecity* we used online computer vision service that measured the degree of smile in each photo on 0–100 scale. I doubt that you will be able to differentiate between smiles on such a fine scale.

Consider another example—representation of colors. In the 1990s and 2000s, digital images often used 24 bits for each pixel. In such format, each pixel can encode grayscale using 0–255 scale. This representation supports 16 million different colors—while human eyes can only discriminate between approximately 10 million colors. As I am writing this, many imaging systems and image editing software use 30, 36 or 48 bits per pixel. With 30 bits per pixel, more than 1 billion different colors can be encoded. Such precision means that if we want to compare color palettes of different painters, cinematographers, or fashion designers using digital images of their works, we can calculate it with more than sufficient accuracy. Certainly, this precision goes well beyond what we can do with small number of terms for colors available in natural languages (Gibson and Conway 2017). Certainly, some natural languages have more terms for different colors then other languages, but no language can represent as many colors as digital image formats.

In summary, a data representation of a cultural artifact or experience that uses numerical values or features computed from these values can capture analog dimensions of artifacts and experiences with more precision than a linguistic description. However, remember that a natural language also has many additional representation devices besides single words and their combinations. They include the use of metaphors, rhythm, intonation, stream of consciousness and other strategies that allow us to describe experiences, perceptions and psychological states in ways that single words and phrase can't. So, while natural languages are categorical systems, they also offer rich tools to go beyond the categories. Throughout human history poets, writers, and performers using speech (and best hip-hop and spoken word artists) today create exceptional works by employing these tools.

Not everybody can invent great metaphors. Numerical features allow us to measure analog properties of the scale of arbitrary precision and do this automatically at scale using computers. But this does not mean that data representations of aesthetic artifacts, processes, and performances that use numbers can easily capture everything that matters.

In the beginning of the twentieth century, modern art rejected figuration and narration, and decided instead to focus on the sensorial communication—what Marcel Duchamp referred to as "retinal art." But over the course of the twentieth century, as more possibilities were fully explored and became new conventions, *artists started to*

create works that are harder and harder to describe using any external code, be it language or data. For example, today we can easily represent flat geometric abstractions of Sonia Delaunay, František Kupka, and Kasimir Malevich as data about shapes and colors and sizes of paintings and drawings; and we can even encode details of every visible brushstroke in these paintings. (Computer scientists have published many papers that describe algorithmic methods to authenticate the authorship of paintings by analyzing their brushstrokes.) But this becomes more difficult with new types of art made in the 1960s–1970s: light installations by James Turrell, acrylic 3D shapes by Robert Irvin, "earth-body" performances by Ana Mendieta, happenings by Alan Kaprow (to mention only most canonical examples), as well as works of thousands of other artists in other countries, such as Движение art movement in USSR. Their works included *Cybertheatre* staged in 1967 and described in their article published in *Leonardo* journal (Nusberg 1969). The only actors in this theatre performance were 15–18 working models of cybernetic devices (referred as "cybers") capable of making complex movements, changing their interior lighting, making sounds, and omitting color smoke. For something less technological, consider *Imponderabilia* by Marina Abramović and Ulay (1977): for 1 h, the members of public were invited to pass through the narrow "door" made by naked bodies of the two performers.

The experience of watching documentation left after an art performance is different from being present at this performance; and what can we measure if an artwork is designed to deteriorate over time or quickly self-destructs like Jean Tinguely's "Homage to New York" (1960)? Similarly, while the first abstract films by Viking Eggeling, Hans Richter, and May Ray made in the early 1920s can be captured as numerical data as easily as geometric abstract paintings by adding time information, how do we represent Andy Warhol's *Empire* (1964) that contains a single view of the Empire State Building projected for 8 h? We certainly can encode information about every frame of a film, but what is crucial is the physical duration of the film, its difference from the actual time during shooting, and very gradual changes in the building appearance during this time. The film was recorded at 24 frames per second, and projected at 16 frames per second, thus turning physical 6.5 h into 8 h and 5 min of screen time. (Very few viewers were able to watch it from beginning to end, and Andy Warhol refused to show it in any other way.)

5 Conclusion

In this article, I have argued that the use of numerical representation and data analysis and visualization methods offers a new language for describing cultural artifacts, experiences and dynamics. The human languages such as English or Russian that developed rather recently in human evolution are not good at capturing analog properties of human sensorial and cultural experiences. These limitations become particularly worrying if we want to compare thousands, millions or billions of artifacts—i.e. to study contemporary media and cultures at their new twenty-first century scale. When we instead use numbers, numerical summaries such as Computer Vision features and also data visualization, we can better capture small differences between a few or very many artifacts. The methods of finding structures and relationships in large numerical datasets developed in statistics and machine learning such as cluster analysis, dimension reduction, and other fields such as network science allow us to extend the analysis to very big datasets of cultural artifacts. Equally importantly, numbers, features and data visualization also give us a language to represent gradual and continuous temporal changes—something which natural languages are also bad at.

Having a better language to describe the analog dimensions of visual culture including single images, video, or a dance performance is invaluable. Digital computers that work on numerical representations are better at capture many dimensions which natural languages can't describe in enough detail, such as motion or rhythm. We can now describe the characteristics of cultural processes which are hard to capture linguistically—for example, gradual historical changes in any visual culture over long periods, changes in visual form over the career of an artist, changes in cinematography over the course of a feature film or a music video.

And this is what the phrase "language of art" in the title of this article refers to. In the twentieth century, many artists, filmmakers, architects and theorists—especially within semiotics paradigm—were proposing that different arts and culture areas have their own languages comparable to human natural languages (Barthes 1997). In my view, these explorations did not reach satisfactory results partly because these theories were using natural languages to try to describe analog dimensions of art and culture. And as I argued here, such an attempt is inherently problematic.

I don't want to argue for or against the idea that painting, fashion, food or space design communicate like languages. In fact, works by Goodman (1968), Sonesson (1989) and by other theorists developed more precise and productive concepts and theories that describe about the differences between languages and various art and cultural forms. What I did claim here is that now we can use digital computers to capture analog dimensions of artifacts and our aesthetic experiences as numbers. This numbers can use continuous scales that allows us to capture tiny differences between artifacts and details of artifacts with as much precision as we want. And we do can this for arbitrary large numbers of artistic and cultural artifacts.

In other words, we now possess a new language for describing and talking about art and culture. In my view, this is very important because being able to describe any phenomenon more precisely than we could earlier is the first step for expanding our knowledge in any domain.

Funding No funding supported writing this article.

Compliance with ethical standards

Conflicts of interest All authors declare that they have no conflict of interest.

References

Barthes R (1997) Elements of semiology. Hill and Wang, New York. Originally published in France in 1962

Brown K (2020) The Routledge companion to digital humanities and art history. Routledge, London

Champion E (2017) Digital humanities is text heavy, visualization light, and simulation poor. Digital Scholarship Humanities 32, issue supplement 1: 25–32. https://academic.oup.com/dsh/article/32/suppl_1/i25/2957402. Accessed 1 July 2020

Derech N, Tal A, Shimshoni I (2018) Solving archeological puzzles. https://arxiv.org/pdf/1812.10553.pdf. Accessed 1 July 2020

Desrosières A (1998) The politics of large numbers: a history of statistical reading. Harvard University Press, Cambridge

Digital Humanities Conference (2019) https://dh2019.adho.org. Accessed 1 July 2020

Gibson T, Conway BR (2017) The world has millions of colors. Why do we only name a few? Smithsonian Magazine. https://www.smithsonianmag.com/science-nature/why-different-languages-name-different-colors-180964945/. Accessed 1 July 2020

Goldberger P (1996) On Madison avenue, sometimes less is less. The New York Times October 27, 1996

Goodman, N (1968) Languages of art: an approach to a theory of symbols. Bobbs-Merrill, Indianapolis

Heftberger A (2019) Digital humanities and film studies. Springer, Berlin

Huawei (2019) Huawei P20. consumer.huawei.com. http://consumer.huawei.com/en/phones/p20/. Accessed 1 July 2020

Impett L, Moretti F (2017) Totentanz. Operationalizing Aby Warburg's Pathosformeln. Stanford Literary Lab. https://litlab.stanford.edu/LiteraryLabPamphlet16.pdf. Accessed 1 July 2020

Isaacson W (2012) How Steve jobs' love of simplicity fueled a design revolution. Smithsonian Magazine, September 24, 2012. http://www.smithsonianmag.com/arts-culture/how-steve-jobs-love-of-simplicity-fueled-a-design-revolution-23868877/. Accessed 1 July 2020

Manovich L (2001) The language of new media. The MIT Press, Cambridge

Manovich L (2007b) Cultural analytics: about. Software Studies Lab. http://lab.softwarestudies.com/p/overview-slides-and-video-articles-why.html. Accessed 1 July 2020

Manovich L (2007c) Information as an aesthetic event. Receiver, n.p. http://manovich.net/index.php/projects/information-as-an-aesthetic-event. Accessed 1 July 2020

Manovich L (2009) There is only software. In: Lee Y, Henk Slager H (eds) Nam June Paik reader—contributions to an artistic anthropology. NJP Art Center, Yongin, pp 26–29

Redi M, Liu FZ, O'Hare NK (2017) Bridging the aesthetic gap: the wild beauty of web imagery. In: ICMR'17: proceedings of the 2017 ACM international conference on multimedia retrieval. ACM, New York, pp 242–250

Nusberg L (1969) Cybertheater. Leonardo 2: 61–62. http://monoskop.org/images/a/af/Nusberg_Lev_1969_Cybertheater.pdf. Accessed 1 July 2020

Pawson J (2020) Calvin Klein Collections Store. Johnpawson.com. http://www.johnpawson.com/works/calvin-klein-collections-store. Accessed 1 July 2020

Peckham J (2018) Huawei P20 and P20 pro colors: what shade should you buy? Techradar. http://www.techradar.com/news/huawei-p20-and-p20-pro-colors-what-shade-should-you-buy. Accessed 1 July 2020

Sonesson G (1989) Pictorial concepts: inquiries into the semiotic heritage and its relevance to the interpretation of the visual world. Lund University Press, Lund

Stork D (2009) Computer vision and computer graphics analysis of paintings and drawings: an introduction to the literature. In: Xiaoyi J, Nicolai P (eds) CAIP'09: proceedings of the 13th international conference on computer analysis of images and patterns. Springer, Berlin, pp 9–24

VISART IV (2018) 4th workshop on computer vision for art analysis, 9th September 2018, Munich, Germany. https://visarts.eu/past-workshops/2018. Accessed 1 July 2020

Yale Digital Humanities Lab (2017) Yale DHLab—robots reading vogue. http://dhlab.yale.edu/projects/vogue/. Accessed 1 July 2020