Alessandra Campos

Final Paper (Transforming Data)

Cultural Analytics, Cinema, and Surveillance:
Transforming Culture through the Vision of Data

In the fall of 2007, Lev Manovich drafted a proposal for what he termed "cultural analytics." Cultural analytics seeks to apply the model of visual analytics to research in the humanities. Visual analytics is a multidisciplinary field of study that combines information and scientific methods of analysis[1] in order to "synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data" (Thomas and Cooke, 10). The visual analytics method focuses on data representations that allow for massive amounts of information to be communicated and analyzed by the human visual field. The visual analysis of these massive data sets allow the interactive human eye to draw novel conclusions that may have otherwise been hindered by the sheer size of the data set. In other words, by enlarging the scope of data that may be accessed and analyzed through a single interface the analyst is able to detect patterns that emerge from the data. Manovich's idea of cultural analytics takes the visual analytics model and applies it to culture. Manovich expands the definition of culture, which usually consists entirely (in the humanities discipline) of historical objects such as painting, literature, and film. The culture that Manovich is primarily interested in is so-called "born digital" data[2], that is, data which is continuously generated in digital space through devices such as smartphones or computers; or user-driven sites such as Flickr and recommendation system modeled, web-based corporations such as Amazon.

Manovich interacts with a variety of media and theoretical frameworks in the formation of his idea for cultural analytics, including cyberinfrastructure, visual analytics, database and

interface, and cinema. The intersection of these theories creates a novel theoretical framework that may be necessary for an inclusive and precise analysis of digital culture; however, their emergent hybrid form (cultural analytics) transforms the shape of culture itself by conceiving of culture as data. In this essay, I will explore these media and media theories in order to trace the roots of the cultural analytics model. Additionally, I will explore the possibility of unintended consequences of a data-centric approach to culture. Issues such as surveillance and privacy, cross-cultural difference, and corporate ownership of knowledge must be considered when examining the application of cultural analytics to global sets of data. An explanation of cyberinfrastructure should begin this essay, as Manovich largely responds to the call to action put forth by the proponents of this model in his proposal.

According to the 2006 report from the American Council of Learned Societies Commission on the relevance of cyberinfrastructure to the humanities, at the time of writing the report "most expressions of human creative in the United States—writing, imaging, music—will be born digital" (Preface, ii). According to the report, it is necessary that the humanities and social sciences develop a model of theoretical engagement and analysis that is able to account for the speed and the shape of digital culture. The model that is proposed by the report is that of a "cyberinfrastructure." The report draws from computer engineer Dan Atkins, defining the term as a network of digital storage that is composed of "hardware, algorithms, software, communication, institutions and personnel" which in turn exist within layers of "base technologies…the integrated electro-optical components of computation, storage, and communications [as well as] software programs, services, instruments, data information, knowledge, and social practices" (6). According to Atkins, cyberinfrastructure consists of digitally stored information, along with the community that interacts with that information and

the tools that are used in the analysis of the stored data. Cyberinfrastructure fosters a collaborative space in which cross-disciplinary modes of understanding are able to interact with one another in shared inquiry. The report calls for humanities scholars to not only utilize this storage in their analysis of contemporary culture, but to adopt the form of that storage in order to foster spaces of collaboration in which information and research may be shared across communities, rather than remaining within the limited space of the academic institution. Manovich responds to this report by the ACLS, as well the related science and engineering based NSF report "Cyberinfrastructure Vision," in his proposal for a cultural analytics model of inquiry.

The intended purpose of a cultural analytics approach to data is to "analyze all existing human cultural records in all media as well as contemporary born-digital cultural data and cultural and social flows" (lab.softwarestudies.com). In order to realize this goal, Manovich expects to utilize both petascale computing, the high performance of which allows for the advanced computation of data that visual analytics requires, and wall-sized gigapixel displays. As a result, Manovich drafts his proposal to the Calit2 Institute, which is one of the four institutes commissioned by the University of California in order to uphold the state's position as a leader of global technological innovation. The institute, which has laboratories across UC campuses, aims to utilize cross-disciplinary research in order to meet its innovate goals. Manovich's proposal was accepted in 2007, and his cultural analytics research is currently being implemented under the name of the Software Studies Initiative at UCSD.

There are a number of theoretical frameworks that have contributed to Manovich's current research into cultural analytics. It is important that the development of this idea be understood, as his idea is already being applied to and transforming disciplines in the humanities.

For example, the paintings of Mark Rothko have been analyzed according to elements such as their "texture, brightness, number of shapes, [and] saturation" and from the computational analysis of these aspects, a visual representation of Rothko's opus is created (a video of cultural analytics being applied to Rothko's paintings can be seen here:

http://www.youtube.com/watch?v=-YlT1qFhJhk&feature=player_embedded#!). While the results of this image processing of Rothko's paintings are not interpretable unless they are then filtered through the imagination of the lived human analyst, the conversion of the image of the painting into data that is then processed by an algorithm implicitly relies on a mathematical model of culture. The resulting visualization does not describe the unmediated Rothko painting, and instead describes an algorithmic interpretation of the image of that painting. In focusing on the analysis of the paintings of a single artist, the visualization effectively removes the "history" portion of "art history." Rather than situating the paintings within a historical framework, which necessarily accounts for all previous and succeeding artistic movements and periods as well as a canon of artistic work, visual analytics places them within an enclosed system. The work of the specific artist is privileged over its place in a historicized timelines of artwork. This type of analysis utilizes the image processing abilities of the computer and takes a formalist approach to analyzing Rothko's work—the formal elements of his paintings, as well as the differences and similarities among them, in isolation describe his work to the analyst. While the ahistoricity of this approach may appear troubling when compared to a humanities approach to critical analysis of artwork, this may be the goal of cultural analytics. Manovich seeks to break from the sequential and hierarchal structure of historical narrative. While the introduction of engineering and mathematical-minded analysis into the humanities does serve to create a potentially beneficial relationship between the arts and sciences that has previously been lacking within the

academic sphere, it is necessary to critique these emerging tools of analysis so that the way in which those tools order culture is understood.

In his book *The Language of New Media*, Manovich configures the structure of the database by relating it to cinema. As he comes from a film theory background, heavily focused on Soviet Cinema and Sergei Eisenstein and Dziga Vertov's ideas concerning the film form, it is not surprising that Manovich draws on the cinematic framework in order to understand the medium of the database. He uses cinematic montage in order to outline what he views as the layered structure of the database, with new data constantly being generated over existing data, explaining that this layering effect allows the database to support both its own form and a narrative form, which occurs in opposition to the logic of the database. For him, the "default" language of image composition is montage—which utilizes techniques of superimposition and rapid intercutting of images in order to generate meaning. Manovich calls upon the semiotic theory of Ferdinand de Saussure in order to explicate his position. The term semiotics was coined by linguist Charles Sanders Peirce, who was writing in the same period as Saussure. Saussure's version of semiotic linguistic theory was termed "semiology," and while Saussure's model gained more prominence in the academic field of film studies over, Peirce's terminology for their branch of linguistic study was adopted to refer to the work of both scholars (Gill 45). Manovich's use of semiotics in his formation of an ontology of the database immediately situates that database within the history of cinema. By placing the database within the larger structure of cinema, and its attendant theoretical frameworks, the conclusions that he draws concerning it must align with a cinematic value system.

While pointing to the historical structure within which Manovich's ideas are situated is essential to understanding exactly what he means when he refers to a cultural analytics model of

knowledge, it does not negate the relevancy of his model. In Saussure's semiotic study of language, he theorized two structures—the syntagmatic and the paradigmatic. Saussure observed that "each word...enjoys a relationship of similarity with numerous other words which are associated with it in some way" (Silverman 102). Found within these relationships between words are the syntagm and the paradigm. The syntagm occurs at the level of discourse, or what is said. In relation to cinema, the syntagm is the diegetic story that is presented to the viewer. On the other hand, the paradigm occurs in the imaginative space of what is not said. Again, in relation to cinema, the syntagm are the images that are not chosen to be represented in the diegetic space. At times these images may be implied by the syntagm, but they do not appear in a visual space external to the viewer. In "The Tutor-Code of Classical Cinema," Daniel Dayan gives the example of the shot-reverse shot as exemplifying the cinematic paradigm. At the initial point of the shot-reverse shot, before the "looked at" character is revealed, there is a moment in which an "absent-one" or invisible other is perceived by the spectator (Dayan 29). This absent other occurs at the level of the paradigm, and remains invisible within the diegesis.

        According to Manovich, narrative-based cinema (he claims that nearly all cinematic works follow a narrative logic) privileges the syntagmatic layer of language. Cinema focuses on the explicit and material existence of shots in order to communicate. However, new media such as the database reverses this, instead privileging the virtual or imagined paradigmatic structure. The database does away with narrative in its creation of an interactive space in which potentially infinite instances of information are presented to the user at once. He says, "On the level of an individual screen, these choices form a paradigm of their own which is explicitly presented to the user…she is selecting one trajectory from the paradigm of all trajectories which are defined" (Manovic 204). The sheer storage capacity of the database, along with the formal logic of the

interface, allows for all possible linguistic combinations to occur simultaneously. So, what the user is presented with is not the closed space of syntagmatic narrative, but the protean space of the imaginary.

While this paradigmatic privileging does create novel possibilities for the way in which data may occur and be ordered, it remains limited to the informational capacity of the database and to the manner in which this data is retrieved—for example, though pre-determined metadata tags. This is the point at which cultural analytics is conceived. Manovich seeks to create an interactive interface in which "the paradigm [is made] even more explicit by presenting the user with an explicit menu of all available choices…all of the categories are always available, just a mouse click away" (204). This may be read as his thesis statement for cultural analytics; he attempts to shatter the linear and sequential structure of narrative by making all existing data available all at once on one interface. He breaks from the cinematic idea of the simultaneous, in which a sequence of images is merely sped up to create the illusion of simultaneous action, and instead conceives of structure which is truly able to present the entire paradigm of human-generated data (or culture) at once. The paradigmatic database would develop an unending catalog of visualizations, and would be presented within an interface that allows for the exploration of all possibilities of the visual moment—in opposition to cinema where the visual moment (the shot) acts as a fleeting hallucination of a frozen history.

The primary hindrance to contemporary users being able to access large sets of data is not due to the limitations of human information processing; according to Manovich, the lack of interface designs that are able to take into account the massive influx of data that is generated daily are the reason that visual analytics has not yet been applied to culture in the social science and humanities fields. Manovich is currently working on a solution to this at his Calit2 lab on the

UCSD campus. His interactive data visualization software runs on the high resolution display

HIPerSpace, which consists of "seventy 30-inch monitors which offer the combined resolution of

287 megapixel (Manovich 7/21). While petascale computing has not yet been introduced to his

lab, the access to the latest hi-res graphic display that he has been given acts as an important step

towards his realization of a global cultural analytics of all human data. Currently ten thousand

images may be displayed at once, allowing the user to detect novel or interesting patterns across

enormous expanses of data territory. Manovich explains that this sort of data-centric logic has

become self-evident in modern culture. The ubiquity of software such as the plug-in weaves the

logic of the database within the fabric of culture—"the ability of a computer to produce endless

variations of elements and to act as a filter, transforming its input to yield a new output —

becomes the logic of culture at large" (Manovich 207).

Those in the field of data visualization might agree with Manovich's sentiment. In

Parsons graduate Manuel Lima's book *Visual Complexity: Mapping Patterns of Information*, his

chapter entitled "Seeing the World in Data" explains

> Current technological developments offer the opportunity to collect data in the same vein
> as Mass-Observation [an early 1900s data collection project], at an even more detailed
> and a much larger scale…An individual can collect thousands of data points during a
> during a single day without even batting an eye or picking up a pencil and a
> notepad…one of the keys to data collection is to make the process easy and to intertwine
> it with daily activities. (246)

Here, Lima echoes Manovich's claim that data mining has become a part of our daily cultural

activity—the Western man brushes his teeth with one hand and with the other uploads an image

of the play he attended the night before into the ether. For both Lima and Manovich, these

mundane activities of data collection should not only be encouraged, but should be made

available to select professionals and experts who are able to use those collections for the

purposes of data visualization and data analysis. Manovich takes this a step further, proposing that a theoretical model that capitalizes on the popularity of real-time data generation and collection be formulated. While this is certainly on track, as new ways of knowledge production emerge so must new ways of analyzing that knowledge, there are some issues with Lima and Manovich's enthusiastic and perhaps premature proposals. To start, Manovich's plan to apply his method of analysis to all cultural data on a global scale does not seem to take into account its inherently Eurocentric model. Data that is not generated from technology such as smartphones and computers, or not present in spaces such as the internet, will occupy a marginal space in relation to these hi-tech methods of data creation and collection. Furthermore, when applied to non-Western data sets (Manovich operation is based in California) the same method of analysis may serve to homogenize that data, rather than discovering novel patterns within it. The material location of the data being culled within the Software Studies Initiative should be taken into consideration before applying the cultural analytics method.

Aside from issues of Western imperialism, there also may be negative consequences to cultural analytics within its California hometown. In his proposal, Manovich indicates that companies such as Sony Online Entertainment have expressed interest in his cultural analytics model. This is not surprising, as his proposal often speaks to corporate agendas involving recommendation systems and contextual advertising. He gives the example of cultural "earthquakes" (adopting the geographical language of California) that occur when a popular product or major motion picture is released. These are very specific aspects of "culture," and in this case, it is unlikely that novel visualizations are analytical conclusions will be drawn from the data received. It is most likely that a user wishing to explore this data further will almost always be directed back to a host site (for example, Sony.com). The presumed interactivity of cultural

analytics combined with its links to corporate mass media brings to mind media scholar Mark Andrejevic's idea of the "digital enclosure" comes to mind.

According to Andrejevic, the proliferation of interactive media in the modern space, which he describes as a digital enclosure, has served to normalize surveillance tactics at both the state/commercial level and amongst peer groups. Here, Lima's linking of data mining to the "Mass Observation" project of early 20th century Britain takes on a new meaning. The digital enclosure is defined as "the creation of an interactive realm wherein every action and transaction generates information about itself" (Andrejevic 2). This digital record of activity acts as a means by which the user's activity may be tracked, and the act of surveillance becomes a built-in feature of interactive media. This inherent surveillance and complicity in the act of surveilling and being surveilled by average users creates an environment in which the monitoring of what might otherwise be private activity goes unnoticed. Recent bills such as SOPA (Stop Online Privacy Act) and trade agreements such as ACTA (Anti-Counterfeiting Trade Agreement) speak to ubiquity of, and expected compliance with, surveillance tactics at all levels. While Manovich certainly does not intend for this hierarchical structure of top-down surveillance to be supported by cultural analytics (his intention is to break down the hierarchical structure of historical and narrative-based models of culture), the potential for violation of user's privacy is embedded within his model.

In order to understand cultural analytics, it is necessary to understand the cinematic background of Lev Manovich. In formulating his idea of an interface large enough to deal with the massive amounts of data being generated on the internet, Manovich goes back to semiotic models of filmic language. He reverses cinema's preferences for the syntagm over the paradigm, conceiving of a visual space in which all possibilities for meaning are presented at once. He also

draws these ideas from the ACLS report that encourages scholars in the humanities field to apply visual analytics techniques to their analysis of culture. Manovich draws support from the corporate sphere, which brings up issues of surveillance and privacy. Ultimately, he seeks to transform culture into data—literally, conceiving of culture as data that is able to be mined and visualized. While he seeks to comprehend the shape of culture through visualization, he also shapes that culture by comprehending it. As data mining and visualization become more necessary as digital information proliferates, Manovich's introductory model of a theoretical framework must be refined and reconsidered as cultural analytics makes its way into academic thought.

[1]In the article "Data Analysis, Proceedings of Information Visualization," written by Daniel A. Keim, Florian Mansmannm Jörn Schneidwind, and Hartmut Ziegler, the authors define visual analytics as an integrated approach to data visualization which combines "methodology from information analytics, geospatial analytics, and scientific analytics."

[2]In the "Cultural Analytics" tab of the Software Studies Initiative's website, Manovich says that (unlike culturomics which focuses on historical data) cultural analytics has the ultimate goal of "analyz[ing] all existing human cultural records in all media as well as contemporary born-digit cultural data and cultural and social flows." Link here:
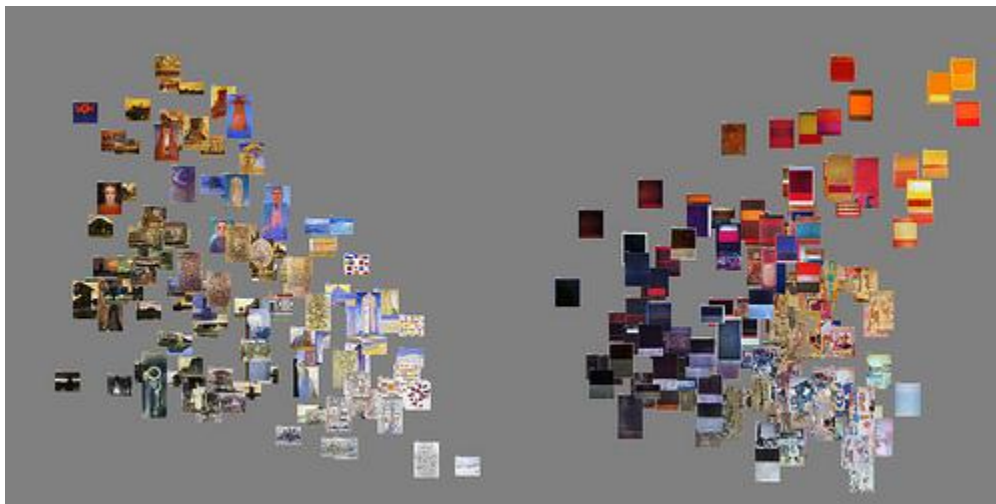http://lab.softwarestudies.com/2008/09/cultural-analytics.html

Cultural analytics HIPerSpace wall display at Software Studies Initiative; researchers analyze 1,000,000 images from manga comics. Image grabbed from cultural analytics Flickr page:
http://www.flickr.com/photos/culturevis/5866777772/

UCSD students view HIPerSpace wall display. Image grabbed from cultural analytics Flickr page: http://www.flickr.com/photos/culturevis/5879009048/in/photostream/

Mondrian vs. Rothko; paintings analyzed using Manovich's visual analytics method. . Image grabbed from cultural analytics Flickr page:
http://www.flickr.com/photos/culturevis/5883371358/in/photostream/

Works Cited

American Council of Learned Societies. "Our Cultural Commonwealth." 2006.

Andrejevic, Mark. Surveillance and Power in the Interactive Era. Kansas: University Press of
        Kansas, 2007.

Dayan, Daniel. "The Tutor-Code of Classical Cinema." Movies and Methods 1 (1976): 438-451.

Gill, Rosalind. Gender and the Media. Massachusetts: Polity Press, 2007.

Keim, D.A.; Mansmann, F. and Schneidewind, J. and Ziegler, H., Challenges in Visual Data
        Analysis, Proceedings of Information Visualization (IV 2006), IEEE, p. 9-16, 2006.

Lima, Manuel. Visual Complexity: Mapping Patterns of Information. New York: Princeton
        Architectural Press, 2011.

Manovich, Lev. The Language of New Media. Massachusetts: MIT Press, 2001.

Manovich, Lev. "Media Visualization: Visual Techniques for Exploring Large Media
        Collections." June 2011.

Manovich, Lev. Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets.
        September 2007.

Silverman, Kaja. The Subject of Semiotics. New York: Oxford University Press, 1983.

Software Studies Initiative. (2008/2009). Cultural Analytics Overview. Retrieved from:
        http://lab.softwarestudies.com/2008/09/cultural-analytics.html