Media Visualization: Visual Techniques for Exploring Large Media Collections

Lev Manovich
Visual Arts Department, University of California, San Diego
manovich@ucsd.edu
www.softwarestudies.com

June 2011

Introduction: How to work with massive media data sets?

Early 21st century media researchers have access to unprecedented amounts of media – more than they can possibly study, let alone simply watch or even search. A number of interconnected developments which took place between 1990 and 2010 – digitization of analog media collections, decrease in prices and expanding capacities of portable computer-based media devices (laptops, tablets, phones, cameras, etc.), the rise of user-generated content and social media, and globalization which increased the number of agents and institutions producing media around the world – led to exponential increase in quantity of media while simultaneously making it much easier to find, share, teach with, and research. Millions of hours of television programs already digitized by various national libraries and media museums, million pages of digitized newspaper pages from 19th and 20th century (http://www.chroniclingamerica.loc.gov), 150 billion snapshots of web pages covering the period from 1996 until today (http://www.archive.org), and hundreds of billions of videos on YouTube and photographs on Facebook (according to the stats provided by Facebook in the summer 2011, every day users upload 100 million images) and numerous other media sources are waiting to be "digged" into. (For more examples of large media collections, see the list of repositories made available to the participants of Digging Into Data 2011 Competition.)

How do we take advantage of this new scale of media in practice? For instance, let's say that we are interested to study how presentation and interviews by political leaders are reused and contextualized by TV programs in different countries. (This example comes from our application for Digging Into Data 2011.) The relevant large media collections that were available at the time we were working on our application (June 2011) include 1,800 Barack Obama official White House videos, 500 George W. Bush Presidential Speeches, 21,532 programs from *Al Jazeera English* (2007-2011), and 5,167 *Democracy Now!* TV programs (2001-2011). Together, these collections contain tens of thousands of hours of video. We want to describe the rhetorical, editing and cinematographic strategies specific to each video set, understand how different stations may be using the video of political leaders in different ways, identify outliers, and find clusters of programs which share similar patterns, But how can we simply watch all these material to begin persuing these and other questions?

But even when we are dealing with large collections of still images – for instance, 167,00 images on "Art Now" Flickr gallery, 236,000 professional design portfolios on coroflot.com (both numbers as of 6/2011), or over 100,000 Farm Security Administration/Office of War Information photographs taken between 1935 and 1944 digitized by Library of Congress (http://www.loc.gov/pictures/) – such tasks are no easier to accomplish. The basic method which always worked when numbers of media objects were small – see all images or video, notice patterns, and interpret them – no longer works.

Given the size of many digital media collections, simply seeing what's inside them is impossible (even before we begin formulating questions and hypotheses and selecting samples for closer analysis). Although it may appear that the reasons for this are the limitations of human vision and human information processing, I think that it is actually the fault of current interface designs.

Popular interfaces for massive digital media collections such as list, gallery, grid, and slide (the example from Library of Congress Prints and Photographs site) do now allow us to see the contents of a whole collection. These interafces usually they only display a few items at a time. This access method does not allow us to understand the “shape” of overall collection and notice interesting patters.

Most media collections contain some kind of metadata such as author names, production dates, program titles, image formats, or, in the case of social media services such as Flickr, upload dates, user assigned tags, geodata, and other information (http://www.flickr.com/services/api/). If we are given access to such metadata for a whole collection in the easy to use form such as a set of spreadsheets or a database, this allows us to at least understand distributions of content, dates, access statistics, and other dimensions of the collection. Unfortunately, usually online collections and media sites do not make available complete collection metadata to the users. But even if they did, this still would not substitute for directly seeing/watching/reading the actual media. Even the richest metadata available today for media collections (for a good example, see http://www.gettyimages.com/EditorialImages) do not capture many patterns which we can easily notice when we directly watching video, look at photographs, or read texts – i.e. when we study the media itself as opposed to metadata about it.

The popular media access technologies of the 19th and 20th century such as slide lanterns, film projectors, microforms, Moviola and Steenbeck, record players, audio and video tape recorders, VCR were designed to access single media items at a time at a limited range of speeds. This went hand in hand with the organization of media distribution: record and video stores, libraries, television and radio would all only make available a few items at a time. For instance, you could not watch more than a few TV channels at the same time, or borrow more than a few videotapes from a library.

At the same time, hierarchical classification systems used in library catalogs made it difficult to browse a collection or navigate it in orders not supported by catalogs. When you walked from shelf to shelf, you were typically following a classiffication based on subjects, with books organized by author names inside each category.

Together, these distribution and classification systems encouraged 20th century media researchers to decide before hand what media items to see, hear, or read. A researcher usually started with some subject in mind – films by a particular author, works by a particular photographer, or categories such as “1950s experimental American films” and “early 20th century Paris postcards.” It was impossible to imagine navigating through all films ever made or all postcards ever printed. (One of the the first media projects which organizes its narrative around navigation of a media archive is Jean-Luck Godard’s *Histoire(s) du cinéma* which draws samples from hundreds of films.) The popular social science method for working with larger media sets in an objective manner – content analysis, i.e. tagging of semantics in a media collection by several people using a predefined vocabulary of terms (for more details, see Stemler, 2001) also requires that a researcher decide before hand what information would be relevant to tag. In other words, as opposed to exploring a media collection without any preconceived expectations or hypotheses

- just to “see what is there” - a researcher has to postulate “what was there,” i.e. what are the important types of information worth seeking out.

Unfortunately, the current standard in media access – computer search – does not take us out of this paradigm. Search interface is a blank frame waiting for you to type something. Before you click on search button, you have to decide what keywords and phrases to search for. So while the search brings a dramatic increase in speed of access, its deep assumption (which we may be able to trace goes back to its origins in 1950s when most scientists did not think how huge massive digital collecions can become) is that you know beforehand something about the collection worth exploring further.

The hypertext paradigm which defined web of the 1990s also only allows a user navigate though the web according to the links defined by others, as opposed to moving in any direction. This is consistent with the original vision of hypertext as articulated by Vannevar Bush in 1945: a way for a researcher to create “trails” though massive scientific information and for others be able to follow his traces later.

My informal review of the largest online institutional media collections available today (Library of Congress collections, archive.org, etc.) suggests that the typical interfaces they offer combine 19^{th} century technologies of hierarchical categories and mid 20^{th} century technology of information retrieval (i.e. search using metadata recorded for media items). Sometimes collections also have subject tags. In all cases, the categories, metadata, and tags were input by by the archivists who. This process imposes particular order son the data. As a result, when a user accesses institutional media collections via their web sites, she can only move along a fixed number of trajectories defined by the taxonomy of the collection and types of metadata.

In contrast, when you observe a physical scene directly with your eyes, you can look anywhere in any order. This allows you to quickly notice a variety of patterns, structures and relations. Imagine, for example, turning the corner on a city street and taking in the view of the open square, with passersby, cafes, cars, trees, advertizing, store windows, and all other elements. You can quickly detect and follow a multitude of dynamically changing patterns based on visual and semantic information: cars moving in parallel lines, house painted in similar colors, people who move along their own trajectories and people talking to each other, unusual faces, shop windows which stand out from the rest, etc.

We need similar techniques which would allow us to observe vast “media universes” and quickly detect all interesting patterns. These techniques have to operate with speeds many times faster than the normally intended playback speed (in the case of time-based media.) Or, to use an example of still images, I should be able to see important information in one million images in the same time it takes me to see in a single image. These techniques have to compress massive media universes into smaller observable media “landscapes” compatible with the human information processing rates. At the same time, they have to keep enough of the details from the original images, video, audio or interactive experiences to enable the study of the subtle patterns in the data.

Media Visualization

The limitations of the typical interfaces to online media collections which we already discussed also apply to interfaces of software for media viewing, cataloging, and editing. These applications allow users to browse through and search image and video collections, and display image sets in an automatic slide show or a PowerPoint-style presentation format. However, as research tools, their usefulness is quite limited. Desktop applications such as iPhoto, Picassa, and Adobe Bridge, and image sharing sites such as Flickr and Photobucket can only show images in a few fixed formats - typically a two-dimensional grid, a linear strip, or, a slide show, and, in some cases, a map view (photos superimposed on the world map). To display photos in a new order, a user has to invest time in adding new metadata to all of then. She can't automatically organize images by their visual properties or by semantic relationships, create animations, compare collections which may have hundreds of thousands of images to each other, or use various information visualization techniques to explore patterns across image sets.

Graphing and visualization tools that are available in Google Docs, Excel, Tableau (http://www.tableausoftware.com), manyeyes (http://www-958.ibm.com/software/data/cognos/manyeyes) and other graphing, spreadsheet, and statistical and software do offer a range of visualization techniques deseigned to reveal patterns in data. However, these tools have their own limitations. A key principle, which underlies the creation of graphs and information visualizations, is the representation of data using points, bars, lines, and similar graphical primitives. This principle has remained unchanged from the earliest statistical graphics of the early 19th century to contemporary interactive visualization software which can work with large data sets (Manovich, 2011a). Although such representations make clear the relationships in a data set, they also hide the objects behind the data from the user. While this is perfectly acceptable for many types of data, in the case of images and video this becomes a serious problem. For instance, a 2D scatter plot which shows a distribution of grades in a class with each student represented as a point serves its purpose, but the same type of plot representing the stylistic patterns over the course of an artist's career via points has more limited use if we can't see the images of the artworks.

Since 2008, our Software Studies Initiative (http://www.softwarestudies.com) has been developing visual techniques that combine the strengths of media viewing applications and graphing and visualization applications. Like the latter, they create graphs to show relationships and patterns in a data set. However, if plot making softwate can only display data as points, lines or other graphic primitives, our software can show all the images in a collection superimposed on a graph. We call this method *media visualization*.

Typical information visualization involves first translating the world into numbers and then visualizing relations between these numbers. In contrast, media visualization involves translating a set of images into a new image which can reveal patterns in the set. In short, pictures are translated into pictures.

Media visualization can be formally defined as *creating new visual representations from the visual objects in a collection* (Manovich, 2011a). In the case of a collection containing single images (for instance, the already mentined 1930s WPA photographs collection from Library of Congress), media visualization involves displaying all images, or their parts, organized in a variety of configurations according to their metadata (dates, places, authors), content properties (for example, presence of faces), and/or visual properties. If we want to visualize a video collection, it is usually more convenient to select key frames that capture the properties and the patterns of video. This selection can be done automatically using variety of criteria – for example, significant changes in color, movement, camera position, staging, and other aspects of cinematography, changes in content such as shot and scene boundaries, start of music or dialog, new topics in characters conversations, and so on.

Our media visualization techniques can be used independently, or in combination with *digital image processing* (Digital image processing, n.d.). Digital image processing is conceptually similar to automatic analysis of texts already widely used in digital humanities (Text Analysis, 2011). Text analysis involves automatic extracting various statistics about the content of each text in a collection such as word usage frequencies, their lengths, and their positions, sentence lengths, noun and verb usage frequencies, etc. These statistics (referred in computer science as "features") are then used to study the the patterns in an single text, relationships between texts, literary genres, etc.

Similarly, we can use digital image processing to calculate statistics about various visual properties of images: average brightness and and satuaration, the number and the properties of shapes, the number of edges and their orientations, key colors, and so on. These features can be then used for similar investigations - for example, the analysis of visual differences between news photographs in different magazines or between news photographs in different countries, the changes in visual style over the career of a photographer, or the evolution of news photography in general over 20th century. We can also use them in a more basic way – for the initial exploration of any large image collection. (This method is described in detail in Manovich, 2011b).

In this chapter we focus on media visualization techniques which are similarly suitable for initial exploration of any media collection but don't require digital image processing of all the images. We present our key techniques and illustrate them with examples drawn from different types of media. You can use variety of software tools and technologies to implement these techniques – scripting Photoshop, using open source media utilities such as ImageMagic (http://www.imagemagick.org), writing new code in Processing (http://Processing.org). In our lab we are rely on open source software imageJ (http://rsbweb.nih.gov/ij/). This softare is normally used in biological and medical research, astronomy, and other science felds. We wrote many custom macros which add new capacities to existing imageJ commands to meet the needs of media researchers. You can find these macros, these macros, the detailed tutorials on how to use them, and sample data sets on http://www.softwarestudies.com. These alow you to both create all visualization types described in this chapter, and also do basic visual feature extraction of any number of images and videos. Thus, if you are interested in using the techniques presented here with your own data sets, you can do this now.

In collaboration with Gravity lab at Calit2 (http://vis.ucsd.edu) where our lab is housed, we have also developed interactive media visualization software runining on the visual supercomputer HIPerSpace (http://vis.ucsd.edu/mediawiki/index.php/Research_Projects:_HIPerSpace). In its present configuration, HIPerSpace consists from seventy 30-inch monitors which offer the combined resolution of 287 megapixel. The size of this cobined display is 9.66 x 2.25 m. We can load up to 10,000 images at one time and then interactively create a variety of media visualizations in real time. (See Yamaoka, S, Cultural Analytics in Large-Scale Visualization Environments, 2011).

Figure:
http://www.flickr.com/photos/culturevis/5866777772/

caption:
Exploring a visualization of one million manga pages on HIPerSpace visual supercomputer.

The development of our tools was driven by research applications. During the last three years, we have used our software for the analysis and visualization of over 20 different image and video collections covering a number of fields. The examples include all paintings by van Gogh, every cover of *Time* magazine published between 1923 and 2009 (4553 images), one million manga (Japanese comics) pages, hundred hours of game play video, and 130 Obama weekly video addresses (2009-2011). The software has also been used by undergraduate and graduate students at the University of California, San Diego (UCSD) in art history and media classes. We also collaborate with cultural institutions that are interested in using our methods and tools with their media collections and data sets - the Getty Research Institute, Austrian Film Museum, Netherlands Institute for Sound and Image, and Magnum Photos. We have also worked with a number of individual scholars and research groups from UCSD, University of Chicago, UCLA, University of Porto (Portugal), and Singapore National University to apply our tools to visual data sets in their respective disciplines - art history, film and media studies, dance studies, and area studies.

Our software has been designed to work with image and video collections of *any* size. Our largest application to date is the analysis of one million pages of fan-translated manga (Douglass, Huber, Manovich, 2011); however, the tools do not have built-in limits and can be used with even larger image sets.

The dimensions of media visualization

In this section we will discusss media visualization method in relation to other methods for analyzing and visualizing data. We have already noted that media visualization can be understood as the opposite of statistical graphs and information visualization.

Media visualization can be also contrasted with content analysis (manual coding of media collection typically used to describe semantics) and automatic media analysis methods commonly used by commercial companies (video fingerprinting, content-based image search, cluster analysis, concept detection, image and video mining, etc.) In contrast to content analysis, media visualization techniques don't require time-consuming creation of new metadata about media collections. In contrast to automatic computational methods, media visualization techniques don't require specialized technical knowledge and can be used by anybody with only a basic familiarity with digital media tools (think QuickTime, iPhoto, and Excel.)

Media visualization method exploits the fact that image collections typically contain at least minimum metadata. This metadata defines the order of the images should be ordered, and/or groups them in various categories. In the case of digital video, the ordering of individual frames is built-in into the format itself. Depending on the genre, other higher-level sequences can be also present: shots and scenes in a narrative theme, the order of subjects presented in news program, the weekly episodes of a TV drama, and so on.

We can exploit these already existing sequence orders in two complementary ways. On the one hand, we can bring all images in a collection together in the order provided by metadata. For example, in the visualization of all 4535 *Time* magazine covers, the images are organized by publication dates. (The visualizations mentioned in this section are illustrated and discussed in more detail below.) On the other hand, to reveal patterns which such an order may hide, we can also place the images in new sequences and layouts. In doing this, we deliberately go against the conventional understanding of cultural image sets which metadata often reify. We call such conceptual operation "remapping." By changing the accepted ways of sequencing media artifacts and organizing them in categories, we create new "maps" of our familiar media universes and landscapes.

Media visualization techniques can be also situated along a second dimension depending on whether they use all media in a collection, or only a sample. We may sample in time (using only some of the available images) or in space (using only parts of the images). The example of a former technique is the visualization of Dziga Vertov's 1928 feature documentary *The 11th Year* that uses only the first frame of each shot. The example of the latter is "slice "of 4535 covers of *Time* magazine which uses only a single-pixel vertical line from each cover.

The third conceptual dimension which also helps us to sort out possible media visualization techniquea is the sources of information it uses along with the media files themselves. As I we already explained, media visualization exploits the presence of at least minimal metadata in media collection, so it does not require addition of new metadata about the individual media items. However, if we decide to add such metadata – the examples may be content tags created via manual content analysis, labels which divide the media collection into classes generated via automatic cluster analysis, automatically detected semantic concepts, the results of face detection, and visual features extracted with digital image processing – all this information can also be used in visualization.

In fact, media visualization gives a new way to work with this information. For example, consider information about covers content we have added manually added to our *Time* covers set, such as whether each cover features portraits of particular individuals or whether it is a conceptual composition which illustrates some concept or issue. Normally a researcher would use graphing techniques such as a line graph to display this information. However, we can also create a high resolution visualization which shows all the covers and uses red borders to highlight the covers which fall in one of these categories.

----

figure:
http://www.flickr.com/photos/culturevis/4347477551/in/set-72157623414034532/

Caption:
Image plot of 4535 covers of *Time* magazine. Jeremy Douglass and Lev Manovich, 2009. X-axis: publication dates. Y-axis: average saturation of each cover (mean value). Red borders are placed around the covers tagged as “compositions.”

The rest of this chapter discusses some of these distinctions in more details, illustrating them with visualization examples created in our lab. Section 1 starts with the simplest technique: using all images in a collection and organizing them in a sequence defined by existing metadata. Section 2 focuses on temporal and spatial sampling. Section 3 discusses and illustrates the operation of remapping.

Media visualization techniques

1. Zoom out (image montage)

Conceptually and technically, the simplest technique is to bring a number of related visual artifacts together and show them together. Since media collections always come with some kind of metadata such as creation or upload dates, we can organize the display of the artifacts by this metadata.

This technique can be seen as an extension of the most basic intellectual operations of media studies and humanities – comparing a set of related items. However, if 20th century technologies only allowed for a comparison between a small number of artifacts at the same time - for example, the standard lecturing method in art history was to use two slide projectors to show and discuss two images side by side – today's applications running on standard desktop computers, laptops, netbooks, tablets, and the web allow for simultaneous display of thousands images in a grid format. These image grids can be organized by any of the available metadata such as year of production, country, creator name, image size, or upload dates. As already discussed above, all commonly used photo organizers and editors including iPhoto, Picassa, Adobe Photoshop and Apple Aperture include this function. Similarly, image search services such as Google Image Search and content sharing services such as Flickr and Picasso also use image grid as a default interface.

However, it is more convenient to construct image grids using dedicated image processing software such as ImageMagic or imageJ since they provide many more options and control. ImageJ calls such the appropriate command image grids "make montage" ("Make Montage," imageJ User Guide, n.d.) and since this is the software we commonly use in our lab, we will refer to such image grids as *montages*.

This quantitative increase in number of images that can be visually compared leads to a qualitative change in type of observations that can be made. Being able to display thousands of images simultaneously allows us to see gradual historical changes over time, the images which stand out from the rest (so-called outliers), the differences in variability between different iamge groups, detect possible clusters of images which share some values in common, and compare the shapes of these clusters. In other words, montage is a perfect technique for "exploratory media analysis."

To be effective, even this simple technique may still require some transformation of the media being visualized. For example, in order to be able to observe similarities, differences, and trends in large image collections, it is crucial to make all images the same size. We also found that its best to position images side by size without any gaps (that's why typical software for managing media collections which leaves such gaps in the display is not suited for discovering trends in a image collections.)

The following example illustrates image montage technique.

----

figure:
http://www.flickr.com/photos/culturevis/4038907270/in/set-72157624959121129/

Caption:
Montage of all 4535 covers of *Time* magazine organized by publication date (1923 to 2009) from left to right and top to bottom. Jeremy Douglass and Lev Manovich, 2009. The large percentage of the covers included red borders. We cropped these borders and scaled all images to the same size to allow a user see more clearly the temporal patterns across all covers.

Placing 4535 *Time* magazine covers published over 86 years within a single high-resolution image reveals a number of historical patterns. Here are some of them:

Medium: In the 1920s and 1930s *Time* covers use mostly photography. After 1941, the magazine switches to paintings. In the later decades the photography gradually comes to dominate again. In the 1990s we see emergence of the contemporary software-based visual language which combines manipulated photography, graphic and typographic elements.

Color vs. black and white: The shift from early black and white to full color covers happens gradually, with both types coexisting for many years.

Hue: Distinct "color periods" appear in bands: green, yellow/brown, red/blue, yellow/brown again, yellow, and a lighter yellow/blue in the 2000s.

Brightness: The changes in brightness (the mean of all pixels' grayscale values for each cover) follow a similar cyclical pattern.

Contrast and Saturation: Both gradually increase throughout the 20th century. However, since the end of the 1990s, this trend is reversed: recent covers have lower contrast and lower saturation.

Content: Initially most covers are portraits of individuals set against neutral backgrounds. Over time, portrait backgrounds change to feature compositions representing concepts. Later, these two different strategies come to co-exist: portraits return to neutral backgrounds, while concepts are now represented by compositions which may include both objects and people – but not particular individuals.

The visualization also reveals an important "meta-pattern": almost all changes are gradual. Each of the new communication strategies emerges slowly over a number of months, years or even decades.

Montage technique has been pioneered in digital art projects such as *Cinema Redu*x (Dawes, 2004) and *The Whale Hunt* (Harris, 2007). To create their montages, artists wrote their custom code; in each case, the code was used with particular data sets and not released for others to use with other data. ImageJ software however allows easy creation of image montages without the need to program. We found that this technique is very useful for exploring any media data set that has a time dimension such as shots in a movie, covers and pages of magazine issues, or paintings created by an artist in the course of her career.

Although montage method is the simplest technically, it is quite challenging to characterize it theoretically. Given that information visualization normally takes data which is not visual and represents in visual domain, is it actually appropriate to think of montage as a visualization method? To create a montage, we start in the visual domain and we end up in the same domain – that is, we start with individual images and we put them together and zoom out to see them all at once.

We think that calling this method "visualization" is justified if instead of focusing on the transformation of data characteristic of information visualization (from numbers and categories to images), we focus on its other key operation: arranging the elements of visualization in such a way as to allow the user easily notice the patterns which are hard to observe over wise. From this perspective, montage is a legitimate visualization technique. For instance, the current interface of Google Books can only shows a few covers of any magazine (including *Time*) at a time, so it is hard to observe the historical patterns. However, if we gather all the covers of a magazine and arrange them in a particular way (making their size the same and displaying them in a rectangular grid in publication date order - i.e. making everything the same so eye can focus on comparing what is different - their content), these patterns are now easy to see. (For further discussion of the relations between media visualization and information visualization, see Manovich 2011a).

2. Temporal and Spatial Sampling

Our next techniques ads another conceptual step. Instead of displaying all the images, we can *sample* them using some procedure to select their subsets and/or their parts.

Temporal sampling involves selecting a subset of images from a larger image sequence. We assume that this sequence is defined by existing metadata such as frame numbers in a video, upload dates of user-generated images on a social media site, or page numbers in a manga chapter. The selected images are then arranged in the order provided by the metadata back to back.

Temporal sampling is particularly useful in representing cultural artifacts, processes, and experiences which unfold over significant periods of time – such as playing a video game. Completing a single player game may take dozens and even hundreds of hours. In the case of Massively multiplayer online role-playing games (MMORPG), users may be playing for years. (2005 study by Nick Yee found the MMORPG players spend on average 21.0 hours per week in the game world.)

In the following example, two visualizations together capture 100 hours of game play.

Left figure:
http://www.flickr.com/photos/culturevis/4039126932/in/set-72157624959121129/

Right figure:
http://www.flickr.com/photos/culturevis/4038975476/in/set-72157624959121129/

Caption:
Kingdom Heart*s* video game play visualizations. William Huber and Lev Manovich, 2010. Left image: *Kingdom Hearts* (2002, Square Co., Ltd.). Right image: *Kingdom Hearts II* (2005, Square-Enix, Inc.).

Each game was played from the beginning to the end over a number of sessions. Kingdom Hearts game play: 62.5 hours of game play, in 29 sessions over 20 days. Kingdom Hearts II game play: 37 hours of game play, in 16 sessions over 18 days. The video captured from all game sessions of each game were assembled into a singe sequence. The sequences were sampled at 6 frames per second. This resulted in 225,000 frames for *Kingdom Hearts* and 133,000 frames for *Kingdom Hearts II*. The visualizations use every 10th frame from the complete frame sets. Frames are organized in a grid in order of game play (left to right, top to bottom).

Kingdom Hearts is a franchise of video games and other media properties created in 2002 via a collaboration between Tokyo-based videogame publisher Square (now Square-Enix) and The Walt Disney Company, in which original characters created by Square travel through worlds representing Disney-owned media properties (e.g., *Tarzan*, *Alice in Wonderland*, *The Nightmare before Christmas*, etc.). Each world has its distinct characters derived from the respective Disney-produced films. It also features a distinct color palettes and rendering styles, which are related to visual styles of the corresponding Disney film. Visualizations reveal the structure of the game play which cycles between progressing through the story, and visiting various Disney worlds.

Temporal sampling is used in interfaces of some media collections – for instance, Internet Archive (http://www.archive.org) includes regularly sampled frames for some of the videos in its collection (sampled rates vary depending on the type of video). In computer science, many researchers are working on video summarization: algorithms to automatically create compact representations of video content which typically produce a small set of frames (the search for "video summarization" on Google Scholar returned 17,600 article). Since applying these algorithms does require substantial technical knowledge, we promote much simpler techniques which can carried without such knowledge using built-in commands in imageJ software.

Spatial sampling involves selecting parts of images according to some procedure. In fact, our visualization of *Time* covers already used spatial sampling since we cropped images to get rid of the red borders. However, as the example below demonstrates, often more dramatic sampling which leaves only a small part of an image can be quite effective in revealing additional patterns which a full image montage may not show as well. Following the standard name for such visualizations as used in medical and biological research and in imageJ software, we are going to refer to them as *slices* ("Orthogonal Views", imageJ User Guide, n.d.)

----

figure:
http://www.flickr.com/photos/culturevis/4040690842/in/set-72157622525012841

Caption:
Slice of 4535 covers of *Time* magazine organized by publication date (from 1923 to 2009, left to right). Jeremy Douglass and Lev Manovich, 2009. Every one-pixel wide vertical column is a sampled from a corresponding cover.

Naturally, the question arises about the relations between the techniques for visual meda sampling as illustrated here, and the standard theory and practice of sampling in statistics and social sciences. (For a very useful discussion of general sampling concepts as they apply to digital humanities, see Kenny, 1982). While this question will require its own detailed analysis, we can make one general comment. Statistics is concerned with selecting a sample in such a way as to yield some reliable knowledge about the larger population. Because it was often not practical to collect data about the whole population, the idea of sampling was the foundation of 20th century applications of statistics.

In some application of media visualization, we face the same limitations. For instance, in our visualizations of *Kingdom Hearts* we sampled the complete videos of game play using a systematic sampling method**.** Ideally, if imageJ software was capable of creating a montage using all of video frames, we would not have to do this.

However, as the *Time* covers slice visualization demonstrates, sometimes dramatically limiting the part of the data is actually better for revealing certain patterns than using all the data. While we can also notice hisorically changing patterns in covers layout (placement and size of the word "Time," the size of portraits in the center in relation to the whole covers) in full image montage, using a single pixel column from each cover makes them stand more clearly.

The following two visualizations further illustrate this idea. The first uses regular sampling (1fps) to represent a 60-minute film – The *11 Year* directed by Dziga Vertov (1928). The resulting representation of the film is not very useful. The second visualization follows semantically and visually important segmentation of the film – the shots sequence. Each shot is represented by its first frame. (The frames are organized left to right and top to bottom following shot order in the film.) Although this visualization uses much smaller number of frames to represent the film than the first, it is much more revealing. We can think of it as reverse engineered imaginary storyboard of the film – a reconstructed plan of its cinematography, editing, and content.

----

figure:
http://www.flickr.com/photos/culturevis/4052703447/in/set-72157622608431194

Caption:
Visualization of *The Eleventh Year* (Dziga Vertov, 1928). The film is sampled at 1 frame per second. The resulting 3507 frames are organized left to right and top to bottom following their order in the film.

figure:
http://www.flickr.com/photos/culturevis/3988919869/in/set-72157622608431194

Caption:
Visualization of *The Eleventh Year* (Dziga Vertov, 1928). Every shot in the film is represented by its first frame. Frames are organized left to right and top to bottom following the order of shots in the film.

----

3. Remapping

Any representation can be understood as a result of a mapping operation between two enities: the sign and the thing being signified. Mathematically, mapping is a function which creates a correspondence between the elements in two domains. A familiar example of such mapping is geometric projection techniques used to create two-dimensional images of three-dimensional scenes such as isometric projection and perspective projection. Another example is 2D maps which represent physical spaces. We can think of well-know taxonomy of signs defined by philosopher Charles Pierce - icon, index, symbol, and diagram - as different types of mapping between an object and its representation. (Twentieth century cultural theory often stressed that cultural repesentations are always partial maps since they can only show some aspects of the objects. However, given the dozens of recently developed technologies for capturing data about physical objects and the ability to process massive amounts of data to extract features and other information - something which, for instance Google does a few times a day than it analyzes over a trillion web links - this assumption needs to be re-thought. For further discussion of this point, see Manovich, 2011c.)

While any representation is a result of a mapping operation, modern media technologies - photography, film, audio records, fax and copy machines, audio and video magnetic recoding, DVD, the web - led to a new practice of cultural mapping: using already an existing media artifact and creating a new meaning or aesthetic effect by sampling and re-arranging parts of this artifact. While this strategy was already used since the second part of the 19th century (photomontage), it became progressively more central to modern art since the 1960s. Its different manifestations include pop art, compilation film, appropriation art, remix, and a large part of media art - from Bruce Conner's very first compilation film *A Movie* (1958) to Dara Birnbaum's *Technology/Transformation: Wonder Woman* (1979), Douglas Gordon's *24 Hour Psycho* (1993), Joachim Sauter and Dirk Joachim' *The Invisible Shapes of Things Past* (1995), Mark Napier's Shredder (1998), Jennifer and Kevin McCoy' *Every Shot / Every Episode* (2001), *Cinema Redux* by Brendan Dawes (2004), and many others. While earlier remappings were done manually and had small number of elements, use of computers allowed artists to automatically create new media representations containing thousands of elements.

If the original media artifact such as a news photograph, a feature film, or a web site – can be understood as "map" of some "reality," an art project which re-arranges the elements of this artifact can be understood as "remapping." Such projects derive their meanings and aesthetic effects from systematically re-arranging the samples of original media in new configurations.

In retrospect, many of these art projects can be also understood as media visualizations. They examine ideological patterns in mass media, experiment with new ways to navigate and interact with media, and defamiliarize our media perceptions.

Our techniques described in this chapter already use some of the techniques explored earlier by artists. However, the practices of photomontage, film editing, sampling, remix, photomontage, and digital art are likely to contain many other strategies which can be appropriated for media visualization. Just as with the question of sampling, a more detailed examination of the relations between these artistic practices of media remapping and media visualization as a research method for media studies requires its own detailed analysis, but one difference is easy to describe.

Artistic projects typically sample media artifacts, selecting parts and then deliberately assembling these parts in new orders. For example, in a classical work of video art *Technology/Transformation: Wonder Woman* (1979), Dara Birnbaum sampled TV series *Wonder Woman* to isolate a number of moments such as the transformation of the woman into a super-hero; these short clips were then repeated many times to create a new narrative. The order of the clips did not follow their order in original TV episodes.

In our case, we are interested in revealing the patterns across the complete media artifact or a series of artifacts. Therefore, regardless of whether we are using all media objects (such as in *Time* covers montage, *Kingdom Hearts* montages) or their samples (*Time* covers slice), we typically start by arranging them in the same order as in the original work. Additionally, rather than selecting samples to support our new message (communicating a new meaning, expressing our opinion about the work), we start by following a systematic sampling method – for instance, selecting every 2nd frame in a short video, or a single frame in every shot in a feature film.

As we already argued earlier, we also on purpose experiment with new sampling orders, and different spatial arrangements of the media objects in visualization. Such deliberate remappings are more close to the artistic practice of aggressive rearrangements of media materials – although our purpose is revealing the patterns, which are already present as opposed to making new statements about the world using media samples. Whether this purpose is realizable is an open question. It is certainly possible to argue that our remappings are reinterpretations of the media works which force viewers to notice some patterns at the expense of others.

In the following example, we selected the first and last frame of every shot in *The Eleventh Year*.

----

figure:
http://www.flickr.com/photos/culturevis/4117658480/in/set-72157622608431194/

Caption:
Visualization of *The Eleventh Year* (Vertov, 1928). Each column represents one shot in the film using its first frame (top row) and last frame (bottom row). The shots are organized left to right following their order in the film.

”Vertov” is a neologism invented by director who adapted it as his last name early in his career. It comes from the Russian verb *vertet* which means “to rotate something.” “Vertov” may refer to the basic motion involved in filming in the 1920s - rotating the handle of a camera – and also the dynamism of film language developed by Vertov who, along with a number of other Russian and European artists, designers and photographs working in that decade wanted to defamiliarize familiar reality by using dynamic diagonal compositions and shoting from unsual points of view. However, our visualization suggests a very different picture of Vertov. Almost every shot of *The Eleventh Year* starts and ends with practically the same composition and subject. In other words, the shots are largely static. Going back to the actual film and studying these shots further, we find that some of them are indeed completely static – such as the close-ups of a people faces looking in various directions without moving. Other shots employ a static camera which frames some movement – such as working machines, or workers at work – but the movement is localized completely inside the frame (in other words, the objects and human figures do not cross the view framed by the camera.) Of course, we may recall that a number of shots in Vertov's most famous film *Man with A Movie Camera* (1929) were specifically designed as opposites: shooting from a moving car meant that the subjects were constantly crossing the camera view. But even in this most experimental of Vertov's film, such shots constitutes a very small part of a film.

Conclusion

In this chapter we described visual techniques for exploring large media collections. These techniques follow the same general idea – use the content of the collection, i.e. all images, their subsets (temporal sampling) or their parts (spatial sampling) – and present it in various spatial configurations in order to make visible the patterns and the overall "shape" of a collection. To contrast this approach to more familiar practice of information visualization, we call it "media visualization."

We included examples which show how media visualization can be used with already existing metadata. However, if researchers add new information to the collection, this information can also be used to drive media visualization. For example, images can be plotted using manually coded or automatically detected content properties, or visual features extracted via digital image processing. In the case of video, we can also use automatically extracted or coded information about editing techniques, framing, presence of human figures and other objects, amount of movement in each shot, etc.

Conceptually, media visualization is based on three operations: zooming out to see the whole collection (image montage), temporal and spatial sampling, and remapping (re-arranging the samples of media in new configurations). Achieving meaningful results with remapping techniques often involves experimentation and time. The first two methods usually produce informative results more quickly. Therefore, every time we assemble or download a new media daat set, we next explore them using image montages and slices of all items.

Consider this defintion of "browse": "To scan, to casually look through in order to find items of interest, especially without knowledge of what to look for beforehand" ("Browse", n.d.) Also relevant is one of the meanings of the word "exploration": "to travel somewhere in search of discovery" ("Exploration", n.d.) How can we discover interesting things in massive media collections? I.e., how can we browse through them efficiently and effectively, without a knowledge of what we want to find? Media visualization techniques give us some basic ways of doing this - although certainly this is only a start.

Acknowledgments

The research presented in this chapter was funded by Interdisciplinary Collaboratory Grant "Visualizing Cultural Patterns" (UCSD Chancellor office, 2008-2010), Humanities High Performance Computing Award "Visualizing Patterns in Databases of Cultural Images and Video" (NEH/DOE, 2009), Digital Startup level II grant (NEH, 2010), and CSRO grant (Calit2, 2010). We are also very grateful to California Institute for Information and Telecommunication (Calit2) and UCSD Center for Research in Computing and the Arts (CRCA) for their ongoing support.

All the visualizations and software development is a result of systematic collaboration between the key lab members: Lev Manovich, Jeremy Douglass, Tara Zappel, and William Huber. HyperSpace software for interactive media visualization has been developed by So Yamaoka.

References

"Art Now" Flickr group. (n.d.). [Flickr group]. Retreived June 26, 2011, from http://www.flickr.com/groups/37996597808@N01/

Browse. (n.d.) In *Wiktionary*. Retreived June 26, 2011, from http://en.wiktionary.org/wiki/browse

Bush, V. (1945). As We May Think. *The Atalantic Monthly* (July).

Dawes, B. (2004). *Cinema Redux* [digital art project]. Retrieved from http://www.brendandawes.com/project/cinema-redux/

Digging Into Data Challenge 2011 (2011, June 6). List of Data Repositories. Retrieved from http://www.diggingintodata.org/Repositories/

Digital image processing. (n.d.). In *Wikipedia*. Retrieved June 26, 2011, from http://en.wikipedia.org/wiki/Digital_image_processing

Douglass, J., Huber, W., Manovich, L. (2011). Understanding scanlation: how to read one million fan-translated manga pages. *Image and Narrative* 12 (1), 206-228. (2011). Retrieved June 26, 2011, from http://www.imageandnarrative.be/index.php/imagenarrative/article/viewFile/133/104

Explore. (n.d.) In *Wiktionary*. Retreived June 26, 2011, from http://en.wiktionary.org/wiki/explore

Flickr. (n.d.) *The Flickr API*. Retreived June 26, 2011, from http://www.flickr.com/services/api/

Gettyimages.com. (n.d.) *Editorial images*. Retreived June 26, 2011, from http://www.gettyimages.com/EditorialImages

Harris, J. (2007). *The Whale Hunt*. [digital art project]. Retrieved from http://thewhalehunt.org/

Library of Congress. (n.d.). *Farm Security Administration/Office of War Information Black-and-White Negatives* [online photo collection]. Retreived June 26, 2011, from http://www.loc.gov/pictures/collection/fsa/about.html

imageJ introduction. (n.d.) Retrived June 26, 2011, from http://rsbweb.nih.gov/ij/docs/intro.html

Kenny, A. (1982). *The Computation of Style: An Introduction to Statistics for Students of Literature and Humanities*. Oxford: Pergamon Press.

Library of Congress. (n.d.). *Prints & Photographs Online Catalog*. Retreived June 26, 2011, from http://www.loc.gov/pictures/

Make Montage. (n.d.). In *imageJ User Guide*. Retreived June 26, 2011, from http://rsbweb.nih.gov/ij/docs/guide/userguide-25.html#toc-Subsection-25.6

Manovich, L. (2011a). What is Visualization? *Visual Studies*, 26 (1), 36-49. Available at http://lab.softwarestudies.com/2008/09/cultural-analytics.html

Manovich, L. (2011b). How to compare one million images?" In D. Berry (Ed.), *Computational Turn*. Macmillan, forthcoming 2012. Available at http://lab.softwarestudies.com/2008/09/cultural-analytics.html

Manovich, L. (2011c). Trending: The Promises and the Challenges of Big Social Data. In M. Gold (Ed.), *Debates in Digital Humanities*. Minnesotta Univeristy Press, forthcoming 2012. Available at http://lab.softwarestudies.com/2011/04/new-article-by-lev-manovich-trending.html

Orthogonal Views. (n.d.). In *imageJ User Guide*. Retreived June 26, 2011, from http://rsbweb.nih.gov/ij/docs/guide/userguide-25.html#toc-Subsection-25.6

Stemler, Steve (2001). An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17). Retrieved June 25, 2011, from http://PAREonline.net/getvn.asp?v=7&n=17

Text Analysis. (2011). In *Tooling Up for Digital Humanities*. Retrived from http://toolingup.stanford.edu/?page_id=981

Yee, N. (2005). *The Deadalus Project*. May 23, 2011 Retreived June 26, 2011, from http://www.nickyee.com/daedalus/

Yamaoka, S., L. Manovich, J. Douglass, F. Kuester. (2011). Cultural Analytics in Large-Scale Visualization Environments. Submitted to *Computer Graphics Forum*. (http://www.wiley.com/bw/journal.asp?ref=0167-7055).