How to Compare One Million Images?

Dr. Lev Manovich,
Professor, Visual Arts Department, University of California, San Diego (UCSD).
Director, Software Studies Initiative (softwarestudies.com) at California Institute for Telecommunication and Information (Calit2).

Dr. Jeremy Douglass,
Post-doctoral Researcher, Software Studies Initiative.

Tara Zepel,
Ph.D. candidate, Art History, Theory, & Criticism, UCSD.

Exploring one million manga pages on the 287 megapixel HIPerSpace (The Highly Interactive Parallelized Display Space) at Calit2, San Diego. HIPerSpace offers 35,840 x 8,000 pixels resolution (287 megapixels) on 31.8 feet wide and 7.5 feet tall display wall made from 70 30-inch monitors.

INTRODUCTION

The description of joint NEH/NSF Digging into Data competition (2009) organized by Office of Digital Humanities at the National Endowment of Humanities (the U.S. federal agency which funds humanities research) opened with these questions: "How does the notion of scale affect humanities and social science research? Now that scholars have access to huge repositories of digitized data—far more than they could read in a lifetime—what does that mean for research?" A year, later, an article in New York Time (November 16, 2010) stated: "The next big idea in language, history and the arts? Data."

While digitized archives of historical documents certainly represent a jump in scale in comparison to traditionally small corpora used by humanists, researchers and critics interested in contemporary culture have even a larger challenge. With the notable exception of Google Books, the size of digital historical archives pales in comparison to the quantity of digital media created by contemporary cultural producers and users – designs, motion graphics, web sites, blogs, YouTube videos, Flickr photos, Facebook postings, Twitter messages, and other kinds of professional and participatory media. This quantitative change is as at least as important as the other fundamental effects of the political, technological and social processes that start after the end of the Cold War (for instance, free long-distance multimedia communication). In an earlier article I described this in the following way:

The exponential growth of a number of both non-professional and professional media producers over the last decade has created a fundamentally new cultural situation and a challenge to our normal ways of tracking and studying culture. Hundreds of millions of people are routinely creating and sharing cultural content - blogs, photos, videos, online comments and discussions, etc. At the same time, the rapid growth of professional educational and cultural institutions in many newly globalized countries along with the instant availability of cultural news over the web and ubiquity of media and design software has also dramatically increased the number of culture professionals who participate in global cultural production and discussions. Before, cultural theorists and historians could generate theories and histories based on small data sets (for instance, "Italian Renaissance," "classical Hollywood cinema," "post-modernism," etc.) But how can we track "global digital cultures", with their billions of cultural objects, and hundreds of millions of contributors? Before you could write about culture by following what was going on in a small number of world capitals and schools. But how can we follow the developments in tens of thousands of cities and educational institutions? (Manovich, Cultural Analytics, 2009).

While the availability of large digitized collections of humanities data certainly creates the case for humanists to use computational tools, the rise of social media and globalization of professional culture leave us no other choice. But how can we explore patterns and relations between sets of photographs, designs, or video, which may number in hundreds of thousands, millions, or billions? (By summer 2010, Facebook contained 48 billion photos; deviantArt.com, the leading site for non-professional art, housed 100 million submissions; coroflot.com, a site used by professional designers had 200,00 portfolios.)

In 2007 we have set up a new lab called Software Studies Initiative (http://www.softwarestudies.com) at University of California, San Diego (UCSD) and California Institute for Telecommunication and Information (Calit2) to address these challenges. We developed a number of methods and techniques for the analysis and visualization of large sets of images, video, and interactive visual media. This article describes our key method that consists from two parts: 1) automatic digital image analysis that generates numerical descriptions of various visual characteristics of the images; 2) visualizations that show the complete image set organized by these dimensions.

We have already successfully applied this method for researching many types of visual media: comics, web comics, video games, virtual worlds, video, films, cartoons, motion graphics, print magazines, paintings, and photographs. Examples include 167,000 images from Art Now Flickr group, 100 hours of *Kingdom Hearts* video game play, and 20,000 pages of *Science* and *Popular Science* magazines (1872-1922). (For more details and other projects, see Digital Humanities section of our lab blog http://www.softwarestudies.com; you can also find our visualizations on YouTube and Flickr at http://www.flickr.com/photos/culturevis/collections/).

To test how this method would work with larger datasets, in the Fall 2009 we downloaded the complete runs of 883 different manga series from the most popular web site for "scanlations" OneManga.com. ("Scanlation" refers to to manga publications which are digitized and translated by fans.) We obtained all pages available for these series along with the user-assigned tags indicating the genres and intended audiences for them (See Douglass, Huber, Manovich, Understanding Scanlation).

The resulting data set contains *1,074,790 manga pages*. Each page is in the form of a JPEG image; average image resolution is 850 x 1150 pixels. We used our custom digital image processing software to measure a number of visual features of every page (shading, presence of fine details, texture properties) using supercomputers at the National Energy Research Scientific Computing Center (NERSC). (A "feature" is the common term in digital image processing; it refers to a numerical description of some image property such as average grayscale value, or the number of lines. For the readers familiar with computational text analysis: image features are structurally similar to text features; they offer a compact description of the data.)

In this article we use the challenge of working with this set of one million manga pages to motivate the need for a computational approach for the exploration of collections of such size, and to explain our particular method that combines digital image analysis and a novel visualization technique.

Any reflection on culture begins with an informal comparison between multiple objects in order to understand their similarities and differences. For instance, if we want our analysis to properly reflect the range of graphical techniques used today by manga artists across thousands of manga books, millions of pages in these books, and tens of millions of individual panels, we need to be able to examine details of individual images and to find patterns of difference and similarity across large numbers of images. To do this, we need a mechanism that would allow us to precisely compare sets of images of any size – from a few dozens to millions. We discuss how our method, which combines automatic digital image analysis and media visualization, addresses these requirements.

HOW TO COMPARE ONE MILLION IMAGES?

Today, a typical publication in humanities is based on the detailed examination of a small number of artifacts (which, depending on the field can be literary texts, TV programs, films, video games, etc. Of course, this does not mean that the author only considered these artifacts in isolation. Usually the detailed analysis of these artifacts is performed against the larger horizon - the knowledge of the wider cultural field(s) which is acquired both directly (for instance, watching films) or indirectly (for instance, reading publication in film studies). But how reliable or complete is this background knowledge? For instance, to continue with the example of cinema, IMDb (www.imdb.com) contains information for over a half a million films produced since the beginnings of cinema; how many of them were seen by academic film scholars and film critics? (The same database lists 1,137,074 TV episodes as of summer, 2001; see IMDb Database Statistics).

The fact that using tiny samples has been a default method of humanities until now does not mean that we should keep using it by default. If Google can analyze billions of web pages and over a trillion links several times each day, we should be able to do better than simply consider a handful of artifacts and generalize from them – even if we don't have the same resources. The key reason for the huge infrastructure maintained by Google is the need to index the web in real time; in the case of cultural analysis, we don't have the same requirements, so we should be able to analyze large cultural data sets with much smaller computer resources. Indeed, today computer scientists who study online social media typically capture and analyze dozens or hundreds of millions of separate objects - photos, twitter updates, blog posts, etc. - with very small resources. (See Cha, Kwak, Rodriguez, Ahn, Moon, I Tube, You Tube, Everybody Tubes; Crandall, Backstrom, Huttenlocher, Kleinberg, Mapping the world's photos; Kwak, Lee, Park, Moon. What is Twitter?).

Having at our disposal very large cultural data sets which can be analyzed automatically with computers and explored using interactive interfaces and visualization opens up existing new possibilities. These possibilities can potentially transform our understanding of cultures in many important ways. Instead of being fuzzy and blurred, our horizon (knowledge of a cultural field as whole) can become razor sharp and at the same time acquire a new depth (being able to sort and cluster millions of artifacts along multiple dimensions). This would enrich our understanding of any single artifact because we would see it in relation to precisely delineated larger patterns. It would also allow us to make more confident statements about the field at large. Perhaps most importantly, it will erase the distinction between the precision of "close reading" and imprecision of a "zoomed out" view " – between a detailed understanding of a few works and very approximate ideas about the field as a whole which we normally form by mentally interpolating between a small number of facts and artifacts we studied. It will also erase the separation between "close reading" (detailed analysis of small parts of texts) and Franco Moretti's "distant reading" (as it is commonly understood - which is not the same as how Moretti defines it – see his Conjectures on World Literature, 2000): analysis of large scale patterns in the development of entire genres, literary production of whole countries, and the like using a whole novel as an atom of analysis (for instance, counting a number of novels in different genres published over a historical period.) Rather than choosing one scale of analysis, we would be able to easily traverse between all them at will, observing patterns at any scale.

Any automatic computational analysis of large samples of human cultures will have many limitations of its own, and therefore it will not replace human intuition and experience. However, while we should keep in mind these various limitations, the opportunities that it offers are still immense. For example, having access to a million manga pages should allow us, in principle, to pretty reliably map the full spectrum of graphical possibilities used by contemporary commercial Japanese manga artists. Such a mapping would also allow us to understand which manga series are most typical and which are most unique stylistically; to find all series where graphical language significantly changes over time (today all top series have been running for a number of years); to investigate if shorter series and longer series have different patterns; to separate the artists who significantly vary their graphical languages from series to series from the artists who do not; etc. We also would be able to take any hypothesis or observation we may make while informally looking through a small samples of images – for instance, we may observe that the manga aimed at different genders and age groups has distinct graphical languages – and see if it holds across our whole set.

But how can we accomplish this in practice? How can we possibly compare one million manga pages?

WHAT EYES CAN'T SEE

1. Let us start by selecting only two pages from our manga image set set and examining them directly without any software tools. We will take these pages from *One Piece* and *Vampire Knight* series. The first is one of the best selling and top rated *shounen* (teenage boys) manga; the second is among the top *shoujo* (teenage girls) manga.

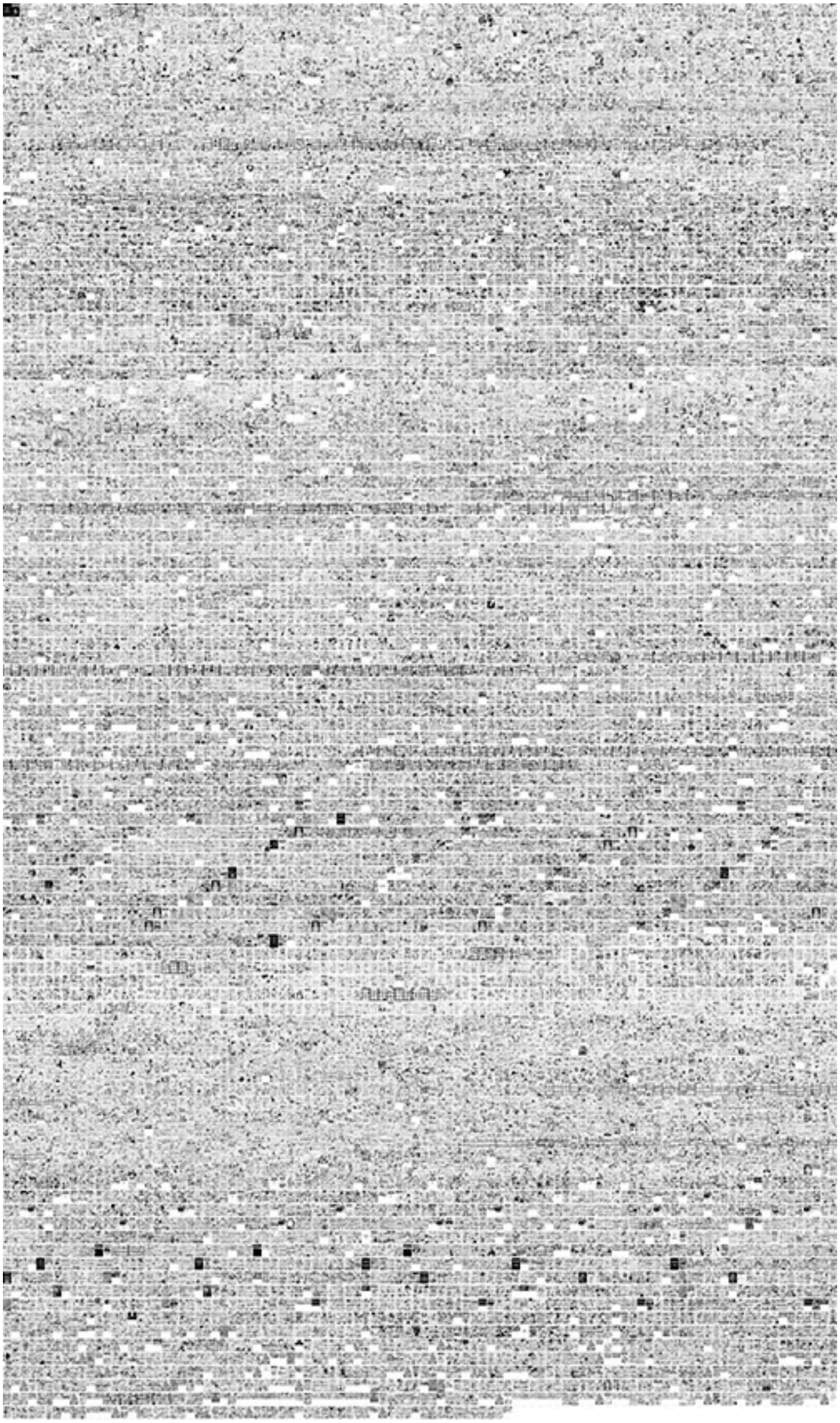
Left image: sample page from *Vampire Knight*. Right image: sample page from *One Piece*. According to the OneManga.com June 2010 list of top 50 series titles accessed by site users, *One Piece* was no. 2, while *Vampire Knight* was no. 13. According to www.icv2.com, during Q3 2010, One Piece was no. 2 in Japan, and Vampire Knight was no. 4. The difference in *Vampire Knight* ranking is likely to reflect different proportions of male/female manga readers inside and outside of Japan.

We can certainly note many kinds of stylistic distinctions by comparing these two pages. For instance, we may observe that a page from *Vampire Kinight* contains dramatic diagonal angles created by both panel divisions and lines inside the panels, the full palette of grey tones from white to grey to black; the large black letters representing sounds and contributing additional visual energy. In contrast, the page from *One Piece* has very little shading; the lines have curves; the panels are neat rectangles. However, how do we know if these stylistic distinctions hold for all of the 10562 *One Piece* pages and 1423 *Vampire Knight* pages which we have available? Similarly, if we want to talk about a graphical style of an artist who may have produced tens of thousands of pages across dozens of titles, would selecting and examining a few dozen pages be sufficient? And what if we want to compare all shounen and shoujo manga in our data set? How many pages should we pull out from each of these categories to reach authoritative conclusions about the possible difference in their graphical languages?

Here is another example. Let’s say we want to compare *One Piece* with another highly popular *shounen* (teenage boys) manga: *Naruto*. Here are the two sample pages from the two series.

Left image: sample page from *Naruto*. Right image: sample page from *One Piece*.

We can notice certain differences in graphical style between these pages – but how typical are these differences for all of 10562 *One Piece* pages and 8037 *Naruto* pages we have available? In contrast to the first example where the styles varied dramatically, here the differences are more subtle – which makes it even more problematic to generalize what we see in these two pages to the complete series.

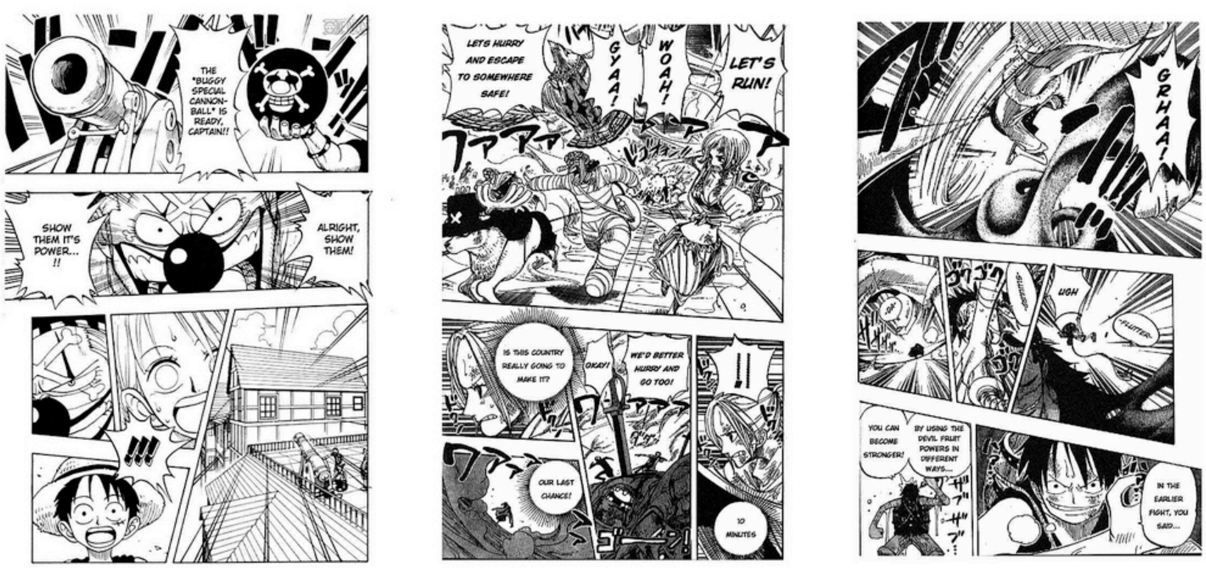


10461 scanlation pages from *One Piece* which were available on onemanga.com in the Fall 2009, organized by sequence of publication (left to right, top to bottom). This composite image (we call such images “montages”) includes special pages inserted by scanlation groups (some of them appear as dark black squares when you look at a small version of the visualization). Note: To fit all pages in a rectangular grid, some of the pages were cropped.

If we only examine a small number of images from a much larger image set at random, we can't be certain that our observations will hold across the complete set. For example, in the case of our one million pages collection (which itself is only a sample of all manga being published commercially), 100 pages is %0.0001 of all available pages. We don't have any reason to believe that whatever we may observe in these 100 pages is typical of our manga set a whole.

Thus, *the first problem with using our native perception to compare images in order to notice differences and similarities is that this approach does not scale.* If we only examine a small selection of manga pages, this does not qualify us to make general statements about the graphical style of "best-selling manga," "shounen manga," or even of a single long running series such as *Hajime no Ippo* (15978 pages in our data set.)

Examining only a small sample of a larger image set also precludes us from understanding detailed patterns of change and evolution. To illustrate this, we pulled out three pages from a best-selling *One Piece* manga series drawn. The pages come from chapters 5, 200, and 400. The series started in 1997; approximately 600 chapters were published by the end of 2010, with new chapters appearing weekly. Thus, the time passed between chapter 5 and chapter 400 is approximately eight years; during these periods, the artists created over 7000 pages. (Our data source did not have exact information on publication dates of every chapter, and that's why we have to estimate them in this way). As we can immediately see by comparing these three pages, the graphical language of *One Piece* apparently changed significantly during these eight years - but *how* did it change? Was it a gradual transition, a series of abrupt changes, or some other temporal pattern? Unless we have a mechanism to compare a much larger number of pages, we can't answer this question.



Sample pages from *One Piece* manga series drawn from the 5th, 200th, and 400th chapters.

In this example, it will be beneficial to pull out a larger set of sample pages using some systematic procedure. For example, we can select every 100th page to get a better idea of how series' visual language is changing over time. This approach can also be applied for comparing different manga series. For instance, we could take every 200th page of *Naruto* and every 200th page of *One Piece*. (*Naruto* is the most popular manga series around the world today.) Since our data set contains approximately 10,000 pages for each of these series, we would end up with 50 pages for each. We could then examine these 100 pages in order to describe the differences between the styles of the two series. Or, if our series are not very long, we can use a different procedure. We can select one page from each chapter of a series and then use such sets of pages to compare series to each other.

However, such an approach will fail if the style within each chapter *varies*. The pages we may select using our sampling procedure may not properly reflect all these variations. This is a fundamental problem with the idea of selecting a number of "representative pages" and only working with this smaller set.

For instance, consider all pages from one of *Abara's* chapters (see illustration below). Which page in each chapter best represents its style? Regardless of which page we select to stand in for the whole chapter, it will not adequately convey the stylistic variations across all the pages. Some pages consist mostly from black and white areas, with little texture; other pages have lots of shading and use mostly grey; still others combine panels in both styles.

Of course, if we are using a computer to browse our image set, we are no longer limited to pulling out a small sample of images for close examination. Modern operating systems (OS) such as Windows, Mac OS, iOS, and Android, image organizer software such as Google's Picasa and Apple' iPhoto, and web photo sharing services such as Photobucket and Flickr all provide the options to browse through lots of images. So if we, for example, add all 10,461 *One Piece* pages to iPhoto, we can pan, zooming in and out at will, quickly moving between individual images and the whole set.

This should help us to notice additonal visual differences and affinities across large sets of images beoynd whose picked up by a close analysis of a small sample. Unfortunately, the display modes offered by existing consumer software and web services are rather inflexible. Typically, the only modes available are a slide show and an image grid. Moreover, usually the images can be sorted by only a few parameters such as uploaded dates, or names, and the user can't easily change the order in which images are displayed. In order to display images in a different order, you have to first to assign new keyword(s) to every image. This prevents spontaneous discovery of interesting patterns in an image set. Instead of reorganizing images along different dimensions and quickly trying different possibilities, a user would have to know first the exact order in which to display images.

Of course, if our image set has some very obvious patterns – lets say, it consists from images in three distinct styles only – these limited browsing modes and fixed order would be still sufficient, and we will easily notice these patterns. But such cases are exceptions rather than the norm. (While Picasso worked in a number of dramatically different styles, he is not typical.)

An alternative to examining a set of images informally - regardless of whether we are looking formally at a few, or use software to browse through many – is to systematically describe each using a number of terms, and then analyze the distributions of these terms. In humanities, this process is called "annotation." A researcher defines a dictionary of descriptive terms and then tags all images (or film shots, transitions between comic frames, or any other visual objects). A parallel to popular practice of users tagging media objects in social media sites (think of tags in Flickr), or adding keywords to one's blog post is obvious – however while users are free to add any keywords they want, in academic studies researchers typically employ "closed vocabularies" where a set of terms is defined beforehand. Once all images are annotated, we can look at all the images that have particular tags; we can plot and compare the tag frequencies and other statistical descriptions. For instance, if we annotate manga pages using a set of tags describing visual style, we can then compare how often each stylistic feature was used for shounen vs. shoujo pages.

Barry Salt pioneered the use of this method to study visual media. He annotated all shots in first 30 minutes of a few hundreds of 20th century feature films using a number of characteristics: shot scale, camera movement, and angle of shot. Balt used a small number of categories for each characteristic. For example, possible camera movement types were pan, tilt, pan with tilt, track, etc. (Barry Salt's Database). He also recorded shot duration (Salt, The Statistical Style Analysis; Film Style and Technology.) Salt then used descriptive statistics measures and graphs to analyze this data. In his very influential book *Understanding Comics* (1993) Scott McLoud employed a similar method compare the visual language of Japanese manga and comics from the West. He annotated types of transitions between frames in a number of manga and comic books, and then used histograms to visually explore the data.

Communication and media studies field have a similar method called "content analysis." If humanists usually are concerned with works of a particular author(s), communication researchers typically employed content analysis to analyze representations in mass media, and, more recently, user-generated content. Therefore, they more carefully determine their samples; they also employ multiple people to "code" (the term used in content analysis to refer to annotation) media material, and then calculate the degree of agreement between the results of different coders. Here is couple of recent applications of this method. Herring, Scheidt, Kouper, and Wright analyzed the content of 457 randomly selected blogs collected at roughly six-month intervals during 2004-2004 "to assess the extent to which the characteristics of blogs themselves remained stable or changed during this active period." Williams, Martins, Consalvo, and Ivory analyzed characters in 150 video games; the total of 8572 characters were coded to "answer questions about their representations of gender, race and age in comparison to the US population."

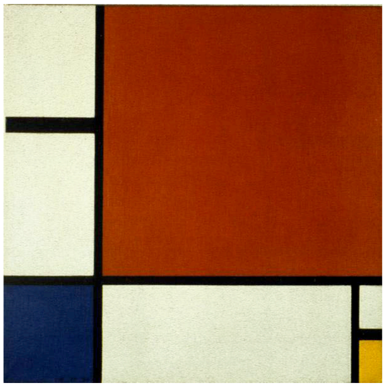
This method is more powerful than informal examination of media, but it suffers from the same problem – it does not scale to massive data sets. McLoud, for instance, only annotated a small number of comic books and manga titles. Would he obtain the same results with a much larger set – such as our collection of one million manga pages? And even if his results would be confirmed, what about all possible exceptions? To find them, we need to tag every page.

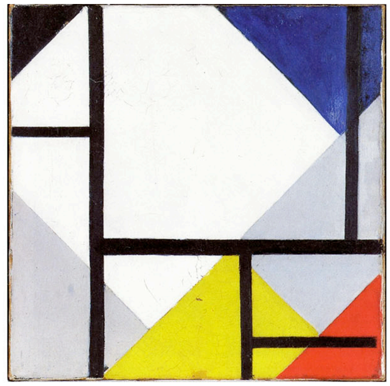
Let us say we have a trained viewer who can examine a manga page and select the relevant tags. If this viewer spends 1 minute for each page and works 8 hours per day, it would take almost 6 years to annotate one million pages.

(Recently, it became possible to use crowd sourcing in order to speed up this process. Since we cannot expect every person to have the same judgments about visual form or to use tags in the same way, researchers use statistical techniques to calculate consistency between the judgments of all participating workers, and to disregard low quality results. However, this approach has a fundamental limitation – like any other attempt to describe images using natural languages, it is much better at capturing images content than form. We discuss this in more detail below.)

Summary: When we examine only a small subset of large image set, our sample may not be representative of the whole set; it may not properly reflect all variations in a set; and we may not be able to study gradual changes over time.

2. So what if we assume that our data set contains not a million images, but only a hundred? Not that the problem of scale goes away, is it sufficient to use our eyes only? No. *The second problem with using our eyes is that we are not very good at registering subtle differences between images*. If you are confronted with a small number of images that have substantial differences, your brain can easily organize these images according to their visual differences. (Here I am not concerned with differences in content, which are easy to see, but with the differences in visual language). This means being able to separate them into groups, rank them according to one or more kinds of visual characteristics, notice the outliers (images which stand out from the rest), and complete other tasks. For instance, we have no problem distinguishing between paintings by Piet Mondrian and Theo van Doesburg created after 1925. Mondrian used exclusively horizontal and vertical line orientations, while van Doesburg also only used diagonals. These properties clearly mark Mondrian's and van Doesburg's paintings.





Left: Piet Mondrian. Composition II in red, blue and yellow. 1930.
Right: Theo van Doesburg. Simultaneous Counter Composition. 1929.

But with large number of images, which have smaller differences, we can no longer perform these tasks. The following example illustrates this problem. The first composite image contains all pages from a single chapter of *Abara* by Tsutomu Nihei. The second composite image contains all pages from a single chapter of *BioMega* by the same artist. Do *Abara* and *BioMega* share the same style, or do they have subtle but important differences (beside the size of the panels)? For example, does one title has more stylistic variety than the other? Which page in each title is the most unusual stylistically? Even with this small number of manga pages, these questions are already impossible to answer.

A chapter from *Abara* manga by Tsutomu Nihei.

A chaper from *BioMega* manga by Tsutomu Nihei.

Summary: even with a small image sample, we may not be able to notice small visual differences between images.

3. *Abara* and *BioMega* have only a few hundred pages each. Out of 883 manga series in our collection, 297 series contain over 1000 pages, while a number of series have more than 10,000 pages each. If we have difficulty comparing only a few dozen pages from *Abara* and *BioMega*, how can we possibly do this with manga series which are much longer?

Annotation/content analysis methods will not help here. To use it, we need to have enough tags to comprehensively describe visual characteristics of manga. However, creating a vocabulary which we can use to label all types of visual differences in manga – or in any other form of visual culture - is problematic by itself. We do not have enough words in our natural languages to adequately describe visual characteristics of *even a single manga image* – let alone all other kinds of human-created images. Consider a sample page from *Vampire Knight* (the image on the left in the first illustration above). Can we describe all variations in background textures in its four panels? Or the differences between the rendering of hair in each of these panels?

This is *the third problem* with studying visual art, visual culture, and visual media using traditional humanities approaches. Regardless of the methodologies and theories being employed in a given case, all of them use one representational system (a natural language) to describe another (images). But as the last example shows, we will not be able to give names to all of the variations of textures, compositions, lines, and shapes used even in a single chapter of *Abara*, let alone one million manga pages. We can proceed with traditional approaches so long as we limit ourselves to discussing manga iconography and other distinct visual elements which have standardized shapes and meanings: water drops signifying stress, throbbing veins signifying anger, and so forth. But if we actually want to start discussing a range of graphical and compositional possibilities used across in manga, we need a new kind of instrument. This fundamental limitation applies to all other visual forms developed by human beings, be they paintings, drawings, photographs, graphic designs, web pages, visual interfaces, animations, etc.

Summary: natural languages do not allow us to properly describe all visual characteristics of images, or name all their possible variations.

METHOD: DIGITAL IMAGE PROCESSING + VISUALIZATION

To address these challenges, we developed a set of methods and techniques, which together we call Cultural Analytics. The key idea of cultural analytics is the use visualization to explore large sets of images and video. These visualizations can use existing metadata (for example, publication dates or author names) and also new metadata added by researchers via annotation or coding. However, as we already discussed, adding tags or other annotations manually has serious limitations: our natural visual system can't notice subtle visual differences between a large number of images, and our natural languages do not have terms to describe all visual characteristics of images, or name their possible variations.

To overcome these limitations, our core method uses digital image processing and a new type of visualization. This section describes this method, and the next sections apply it to progressively larger numbers of images drawn from our one million manga data set. (For the description of our other visualization methods, see Manovich, Media Visualization).

The method involves two steps:

1. We use *digital image processing* to automatically measure a number of visual characteristics (i.e., features) of our images. In this process, visual qualities are mapped into numbers. (In computer science, this step is often referred to as "feature extraction.") For example, in the case of grey tones measured on 0-255 scale, black is represented as 0, white as 255, and %50 grey as 127.5. The examples of dimensions that can be measured include contrast, presence of texture and fine details, number of lines and their curvature, number and type of edges, size and positions of shapes, and so on. In the case of color images, we can also measure the colors of all pixels (hue, saturation, brightness), determine most frequently use colors, and calculate various image statistics for R, G, B color components (R, G, B) separately.

Typically, such measurements produce many numbers for each visual dimension of an image. For instance, every pixel will have its gray scale value. If we measure orientations of all lines in an image, we will end with a separate number for every line. In order to be able to compare multiple images with each other along particular dimensions, it is convenient to use the averages of the measurements on each dimension.

For example, if we are interested in grey scale values, we sum values of all pixels and divide them by the number of pixels. In the case of line orientations, we similarly add angles of all lines and divide them by the number of lines.

Besides this simple averages called mean, we can also use other types of descriptive statistics to summarize image characteristics. They include different representations of central tendency in a data (median, mode, etc.) and of data dispersion (variance, standard deviation, etc.)

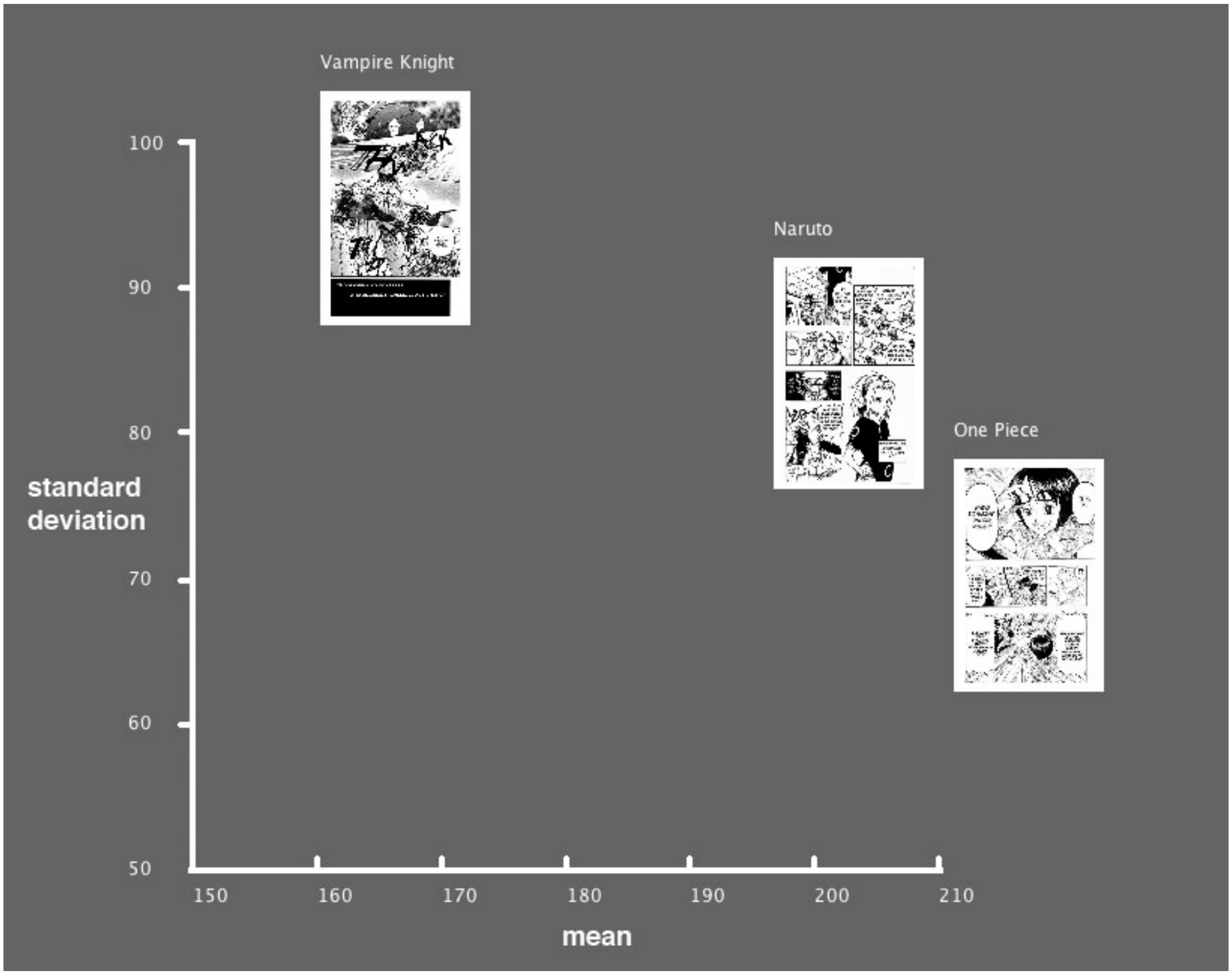
Here are the examples of such statistics calculated for the sample manga pages which already appeared in the previous section.

Left: a page from *Vampire Knight*. Mean: 164.68. Standard deviation: 102.25.
Middle: a page from *One Piece*. Mean: 214.58. Standard deviation: 75.65.
Right: a page from *Naruto*. Mean: 199.96. Standard deviation: 90.92.
Both mean and standard deviation are calculated using greyscale values of all pixels in a page.

Vampire Knight page (left) has the lowest average grayscale value because of its shaded areas, thick black contours and type, and the black area at the bottom. One Piece page (middle) has the highest average grayscale value because it uses black quite sparingly. The standard deviation values are reversed: *Vampire Knight* page has the highest, while *One Piece* page has the lowest. *Vampire Knight* page has lots of black, gray, and white, and this pushes up its standard deviation measurement. One Piece has lots of white and just a little bit of black, and this makes its standard deviation low.

2. We create 2D visualizations that position the images according to their feature values. For example, we may use horizontal dimension (X-axis) to represent grayscale mean, and vertical dimension (Y-axis) to represent grayscale standard deviation. These image features calculated by software in step 1 became image coordinates in a 2D space. In this way, *the differences between images along a particular visual dimension are translated into their positions in space* – something that human visual system can read very well.

To illustrate this, we visualize the three manga pages from the previous example using their mean and standard deviation features.



A plot of sample pages from *Vampire Knight*, *One Piece*, and *Naruto*. Mean and standard deviation of a page are used as its X and Y coordinates.

(Both mean and standard deviation are calculated using greyscale values of all pixels in a page.)

This plot uses only two very basic features. Therefore we can't expect it to capture every visual difference between these three images. Still, even with only two features, the spatial positions and the distances between the images in 2D space reflect well our sense of overall visual differences between the images: *One Piece* and *Naruto* pages are close by; *Vampire Night* page is further away.

Measuring visual features and then mapping these features into X and Y-axis allows us to separate the overall perceived difference into separate dimensions. In this plot, we use average and the spread of grayscale values (i.e., mean and standard deviation), but we can also use many other dimensions can be also used. This process is not unlike how human visual perception functions. Human visual system analyzes visual input separately in terms of different characteristics: contrast, texture, shape, color, and motion. Most psychologists and neuroscientists believe that the brain combines this information to arrive at perceptual wholes. Various theories have been proposed to explain the details of this process and to. An influential theory of attention developed by Anne Treisman and Garry Gelade suggests that different features analyzed at the early stages of perception are "binded together" into consciously experienced wholes. Another theory by L. Ward proposed the neural mechanism responsible for the binding of features which code shape, motion, color, depth and other perceptual aspects. Some of the features we measure such as Gabor filters are thought to be the exact equivalents of the features analyzed by the brain; others can be understood as being equivalent to the combination of multiple features computed by the brain.

As we discussed earlier, when a brain is confronted with a number of very similar images, or a very large number of images, it no longer can compute these differences reliably. When we measure features in image sets and visualizing images according to the feature values, we essentially augments human perception; that is, we scale up its capacity to judge visual differences.

Combination of digital image analysis and visualization makes possible to bypass the problem which haunted visual semiotics in particular, and all linguistic descriptions of the visual in general: the inability of language to adequately represent all variations which images can contain. For instance, even if we use hundreds of words for colors, images can show millions of color variations. And color is not the worst case; for other dimensions such as texture or line character, the terms provide by natural languages are much more limited.

In other words, our senses are able to register a much larger set of values on any analog dimension - loudness, pitch, grayscale, color, motion, orientation, size, etc. – than our languages have words for. This makes sense because language has developed much later evolutionary to supplement the senses. Language divides the continuous world into larger discrete categories that makes possible abstract reasoning, metaphors, and other unique capacities. It is not designed to exactly map the wealth of our sensory experience into another representational system.

So if we can't rely on a natural language to capture what our senses can register – and we can't rely on the senses because, as we discussed in the previous section, they are not able to register very subtle differences between images, or other cultural artifacts – how can we talk about visual culture and visual media?

Our approach is to use visualization as a new descriptive system. In other words, we *describe images with images*. In doing this, we are taking advantage of the ability of images to register subtle differences on any visual dimension.

Note that our method does not imply that we are getting rid of discrete categories. Rather, instead of being limited to a few provided by language, we can now define as many as we need.

For example, lets say that we want to describe grayscale levels in an image. We use software to read pixel values from an image file, and calculate an average value. This average is then used to position the image in the visualization along X or Y-axis.

Common 8-bit and 24-bit image formats such as JPEG and PNG use 256 discrete values to represent grayscale levels. This gives us 256 separate categories for grayscale values. These categories do not have distinct names. But they work - they allow us to compare multiple images in terms of their grayscale values.

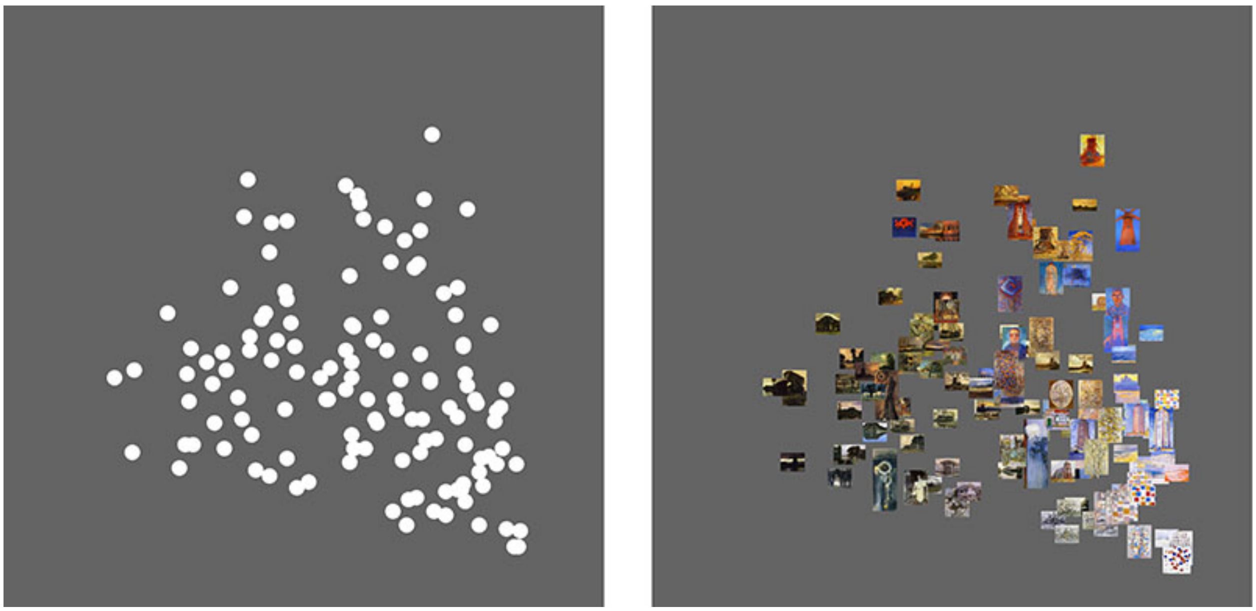
We are not limited to 256 categories – if we want, we can use 1000, 10,000, or any other number. How does this work? When we calculate the average value of all integer grayscale pixel values, we get a real number. For instance, if our image contains four pixels with grayscale values 103, 106, 121, and 112, the average of these values is 102 + 107 + 127 + 113)/4 = 109.75. If we round these values using one decimal place, we will have 256 x 10 = 2,560 distinct categories. If we keep two decimal places, we will have 25,600 distant categories. In reality, we don't need to go that high, since human perception can't even see the difference between two gray levels which are next to each other (for instance, 127 and 128) on 0-256 scale.

While we utilize a variety of visualization techniques, the key technique used by our method is a *scatter plot*, i.e. a 2D visual representation that uses Cartesian coordinates to display two sets of numerical values describing data. In our case, each element of a data set is an image, and the two values that determine its position in a plot are two measured visual qualities (features), such as average grayscale and standard deviation.

Equally often we use *line graphs* where the X-axis represents the dates the images were created (or their positions in a narrative sequence such as a comic book), and Y-axis represents some measured value (such as average saturation).

Along with regular scatter plots and line graphs, we also use a new visualization technique which we call *image plot.* A normal scatter plot and a line graph display the data as points and lines. An *image plot* superimposes images over data points in a graph.

The following visualizations of 127 Piet Mondrian's paintings illustrate the difference between a scatter plot and an image plot.



127 paintings by Piet Mondrian (created between 1905 and 1917) visualized as a scatter plot (left) and as image plot (right). X-axis = brightness median. Y-axis = saturation median.

(Technical details: We also use more advanced visualization techniques such as scatter plot matrix and parallel coordinates, and multivariate data analysis techniques such as PCA, cluster analysis, and so on. However, since the concepts of a multi-dimensional feature space and dimension reduction are more abstract, in this chapter all our examples are 2D visualizations where each axis corresponds to a single feature such as mean grayscale value, or metadata which was collected along with the data - such as a position of a page within the sequence of all pages in a manga title. In general, we prefer to use single features for X and Y-axis if their graph reveals interesting patterns and if their meaning is easy to explain; in contrast, it is often difficult to interpet the dimensions in a graph which uses PCA or other multivariate methods.)

While the technique we call "image plots" has already been described in a number of articles in Computer Science (see Peters, MultiMatch), it has not been made available in any graphing or visualization application. Our lab developed software to create image plots; we use this software in all our projects, and also distribute it as open source (Software Studies Initiative, ImagePlot). The software runs on regular Windows, Mac and Lunix desktops and laptops. Working with our lan, Gravity Lab at California Institute for Telecommunication and Information (Calit2), we also developed an interactive application that can generate image plots which can contain thousands of individual images in real time. The application runs on scalable tiled displays such as HiperSpace (The Highly Interactive Parallelized Display Space) which offers 35,840 x 8,000 pixels resolution (287 megapixels) on 31.8 feet wide and 7.5 feet tall display wall made from 70 30-inch monitors (Yamaoka, Manovich, Douglass, Kuester, Cultural Analytics in Large-Scale Visualization

Environments). In this article, all image plots are done with ImagePlot software; scatter plots and line graphs are done with ImagePlot and Mondrian (free data visualization software).

In the next sections of this article, we will show how our method can be used to compare image sets ranging from a few hundred to one million images.

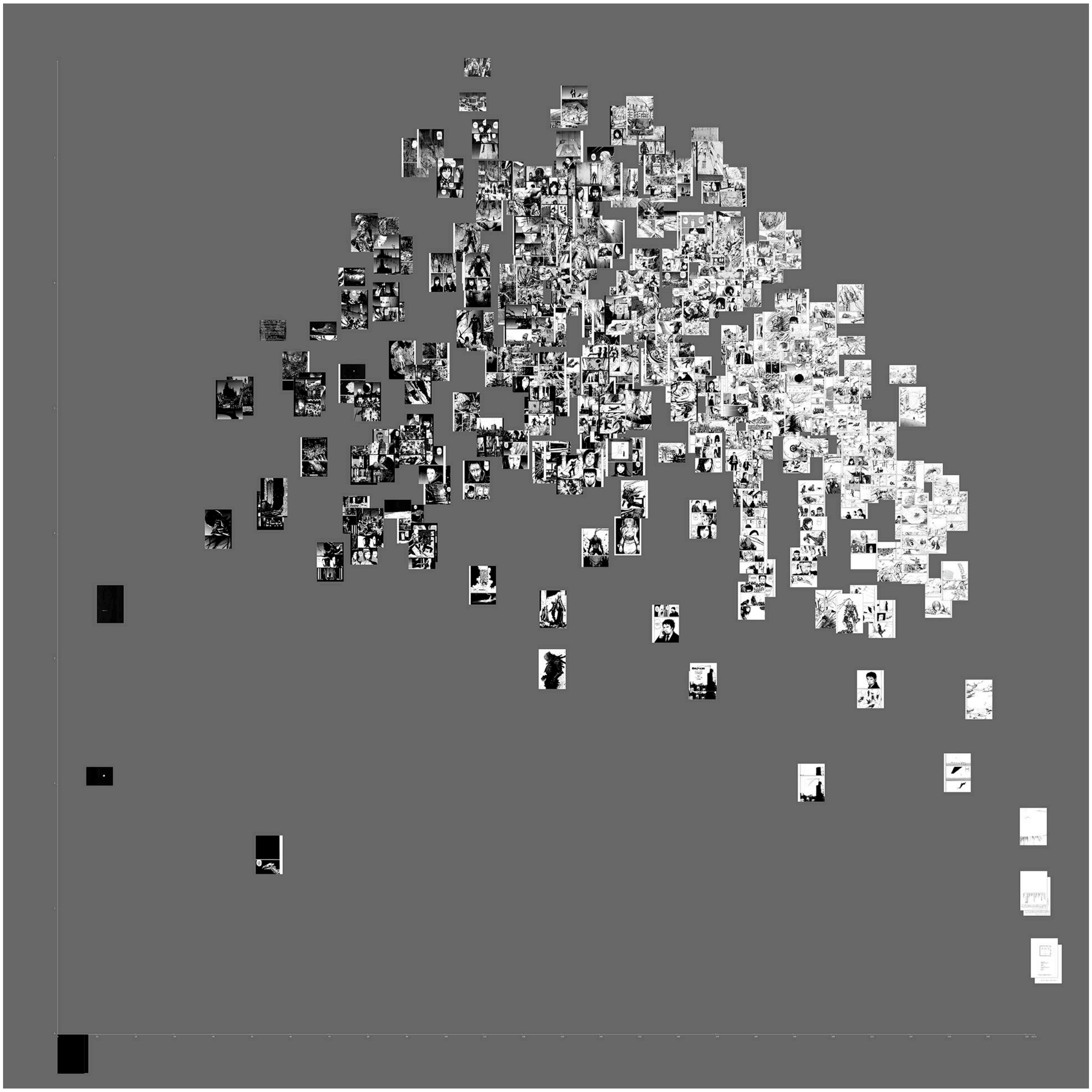
COMPARING *ABARA* AND *NOISE* (474 PAGES)

Having introduced our method – visualizing images as a scatter plot according to quantitative descriptions of their visual properties (features) measured with digital image processing techniques – lets now apply this method to our manga data set.

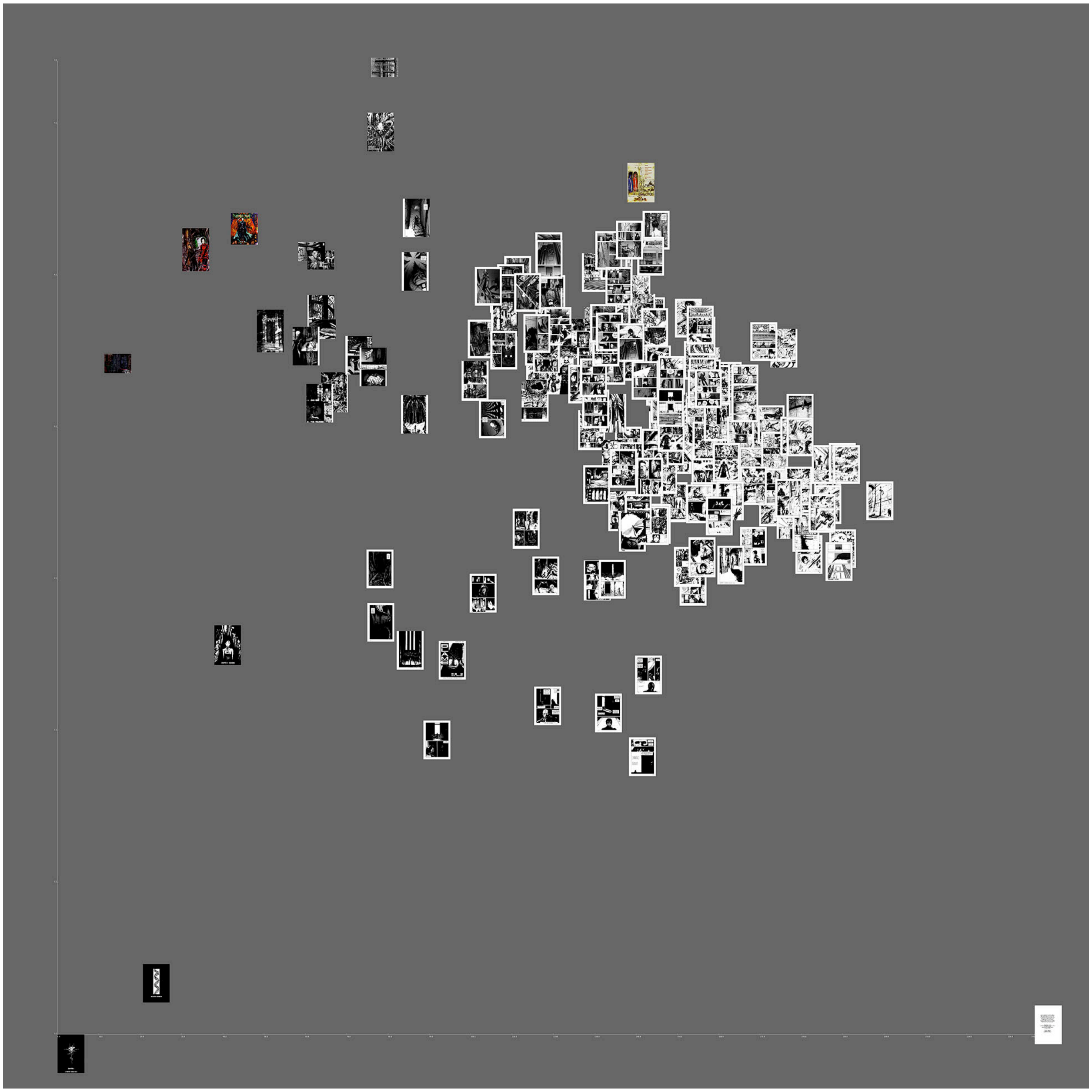
To make our visualization examples easier to follow, we will use the same two visual features in most of the exampes below. The first feature is a standard deviation of greyscale values of all pixels in an image. Standard deviation is a commonly used measure of variability. It shows how much the data is dispersed around the average. If an image has a big range of greyscale values, it will have large standard deviation. If an image employs only a few greyscale values, its standard deviation will be small.

The second feature is entropy. In information theory, the concept of entropy was developed by Claude E. Shannon in his famous 1948 paper "A Mathematical Theory of Communication". Entropy describes the degree of uncertainty in the data – i.e., how difficult or how easy it is to predict the unknown data values given the values we already know. If am image consists from a few monochrome areas, its entropy will be low. In contrast, if an image has lots of texture and details, and its colors (or greyscale values in the case of a black and white images) vary significantly from place to place, its entropy will be high.

In the following examples, we will map standard deviation values to X-axis, and entropy values to Y-axis. We will start by creating image plots of *Abara* and *Noise* pages. The first title has 291 pages; the second has 183 pages. (This count includes all pages, which were available for these titles on onemanga.com, including title and credit pages.)



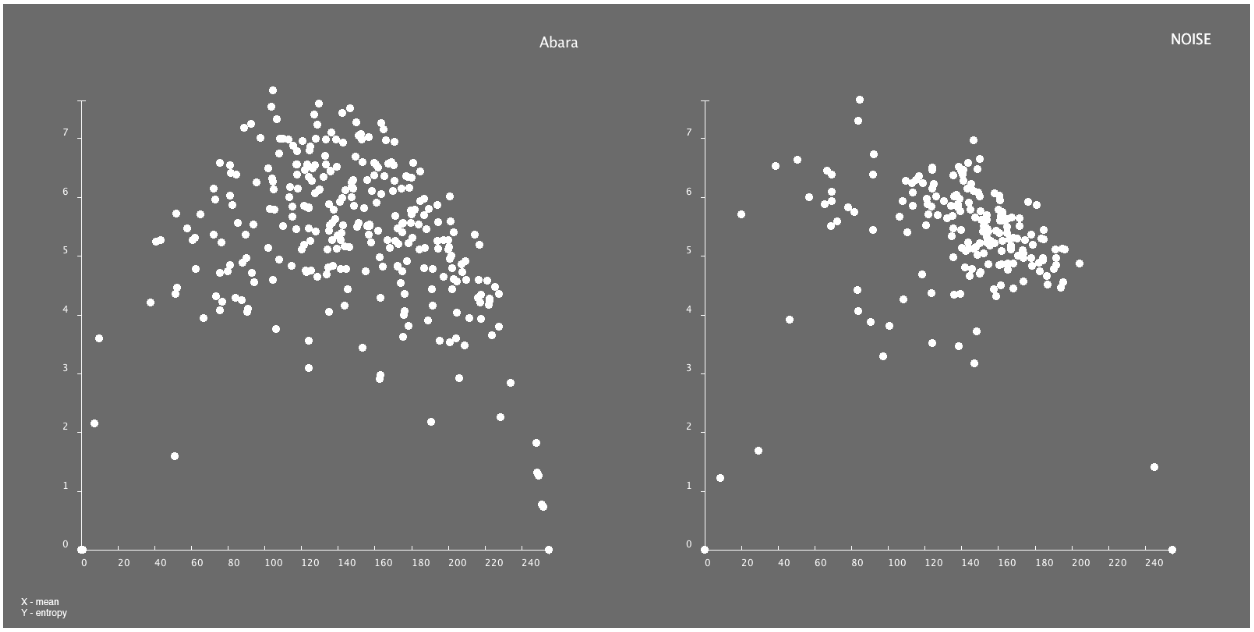
Abara pages. Artist: Tsutomu Nihei.
X-axis = standard deviation of greyscale values of all pixels in a page.
Y-axis = entropy calculated over greyscale values of all pixels in a page.



NOISE pages. Artist: Tsutomu Nihei.
X-axis = standard deviation of greyscale values of all pixels in a page.
Y-axis = entropy calculated over greyscale values of all pixels in a page.

As can be seen, standard deviation (X-axis) and entropy (Y-axis) measurements indeed correspond to perceptually meaningful visual characteristics of manga pages. The pages with low entropy values are situated in the bottom part of a plot. They consist from a small number of flat areas, with minimum detail and no texture. The pages with high entropy values situated in the top part of a plot are the opposite: they have lots of detail and texture. On a horizontal dimension, the pages that only employ a few greyscale values are on the left; the pages that have the range of values in the middle, and the pages that only have white and black are on the right.

To be able to compare the range of standard deviation and entropy values in both titles, we can plot the data using standard scatter plots, and put the two plots side by side (*Abara* is on the left, *NOISE* is on the right).



Abara pages (left) and *NOISE* pages (right). Artist: Tsutomu Nihei.
X-axis = standard deviation of greyscale values of all pixels in a page.
Y-axis = entropy calculated over greyscale values of all pixels in a page.

These visualizations of the two titles side-by-side make answering the questions we asked earlier easy. Does one title has more stylistic variety than the other? Yes, *Abara*'s style varies much more than *NOISE* style: the points on the left plot are more dispersed than the points on the right plot. Which page in each title is the most unusual stylistically? Each plot reveals a number of outliers – i.e. points that stand out from the rest. (Of course, we should keep in mind that the two measurements we are using in these plots – i.e. standard deviation and entropy - only capture some dimensions of a visual style. If we use other features, different pages may stand out as outliers.)

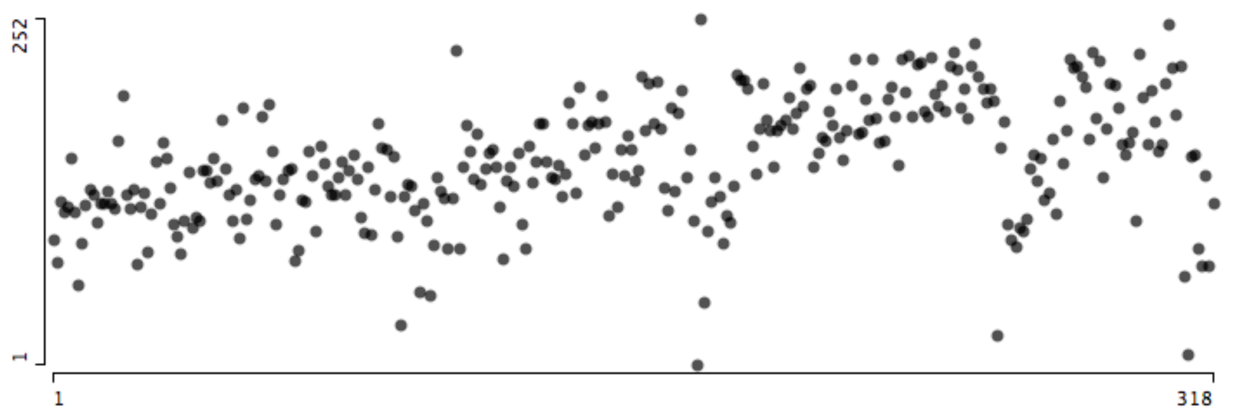
VISUALIZING TEMPORAL CHANGES IN *ABARA* AND *NOISE*

Manga is a sequential art. To understand if and how visual style in a title varies over the sequence of its chapters and individual pages, we can generate image plots where X-axis represents a page's position in a sequence, and Y-axis uses some visual feature. If we do this for two titles, we can then compare how they vary in time in relation to this feature.

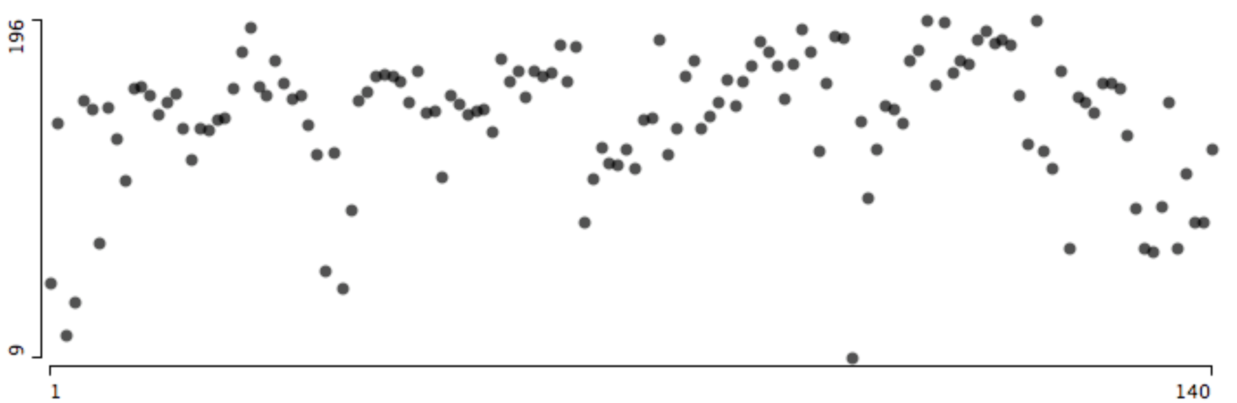
When we make image plots using this approach, they can become very long. For example, lets say we want to graph a sequence of 10,000 images and make each image 100 pixels wide. Such visualization will be 10,000 x 100 = 1,000,000 pixels wide. To address this problem, our visualizations of feature films represent each shot by a single frame. You can find examples of such plots we made for whole films on our Flickr gallery. For shorter manga series, we will use both image plots which show all pages, and line graphs that represent each page by a point. The former are easier to read; the latter are sometimes more effective in revealing patterns. For longer titles, we will use line graphs, since image plots would very long. (A *line graph* does not have to use lines to connect the data. The difference between a line graph and a scatter plot is that the former assumes that data values mapped on to X-axis are separated by the same interval, i.e. 1, 2, 3,… A scatter plot does not make this assumption).

First, we compare *Abara* (318 pages) and *Noise* (140 pages) using a simplest visual feature: a mean of grayscale values of all pixels in a page (Y-axis). The mean value indicates a relative proportion of white, black and grey areas in a page. We numbered all pages staring from the cover page and ending with the last page of the series. (We have removed extra pages inserted by fans who scanned and translated these series; we also removed "bonus" chapters.) The pages are positioned left to right using this linear sequence (Y-axis).

In Japan a manga series first published in magazines which may come out weekly, bi-weekly, or monthly, and contain new chapters for a number of series. Later, a number of already published chapters of succesful series is printed in a separate volume (*tankōbon).* The series translated into other languages are published in similar format (a number of chapters collected in a single book) Global fans who read scanlated manga on the web can also go though a sequence of chapters all at one. A visualization which arranges all pages in a short series such as *Abara* or *NOISE* assumes this reading mode.



Abara (318 consequitive pages in 11 chapters). X-aixs = page postion in the series. Y-axis = grayscale mean.

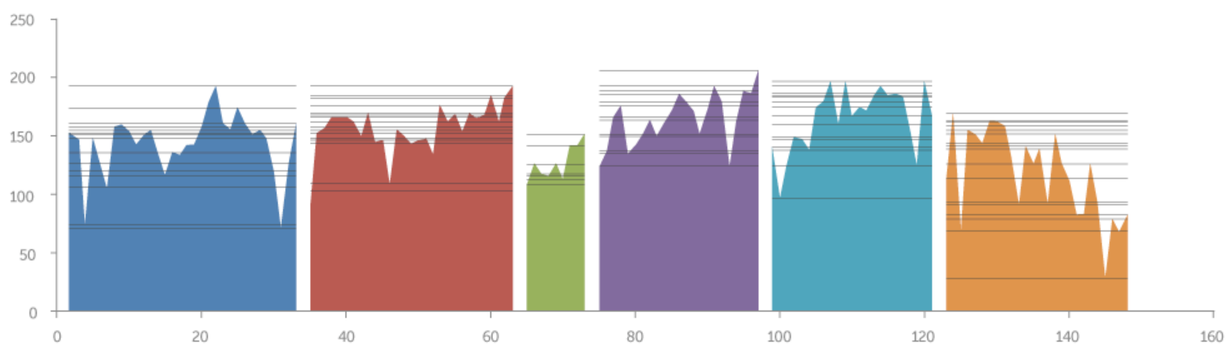


NOISE (140 consequitive pages in 6 chapters). X-aixs = page postion in the series. Y-axis = grayscale mean.

The two visualizations use the same scale for X-aixs, and this is why first graph (*Abara*, 318 pages) is twice as long as the second (*NOISE*, 140 pages). The points corresponding to individual pages are connected by lines in order to make the patterns easier to see.

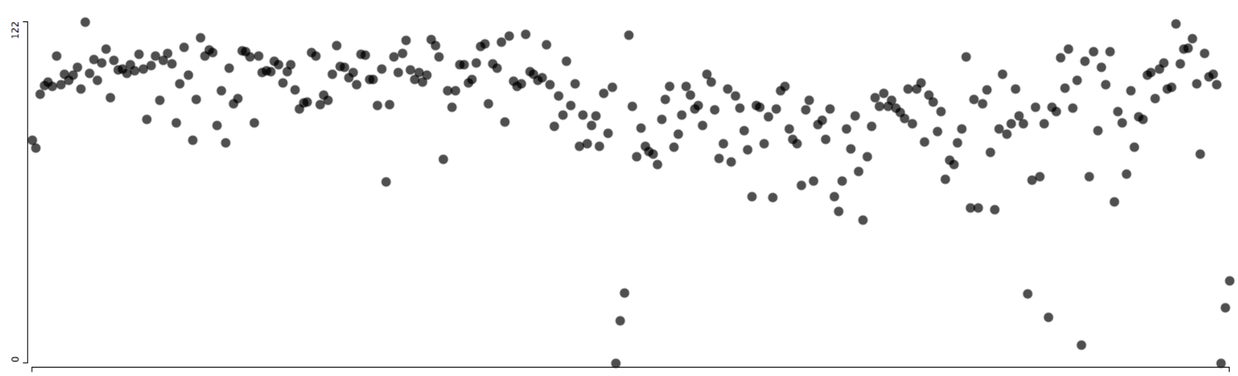
Comparing the two series created by the same arist side by side shows similarities and differences in their graphical development. *Abara*'s graph gradually goes up almost in a linear fashion until the last few pages. *NOISE* graph also slightly goes up till the middle of chapter 5, and then goes down. However, its average grayscale levels never raise as much as those of *Abara*.

To help us understand the graphical patterns in each chapter, graphic designer Ong Kian Peng (Multimodal Analysis Lab, Singapore National University) visualized the sequence of *NOISE* pages as a bar chart. Each chapter is marked in its own color; the chapters are also separated by horizontal intervals. The graph also omits covers the chapters covers to focus on the patterns in the regular narrative pages.

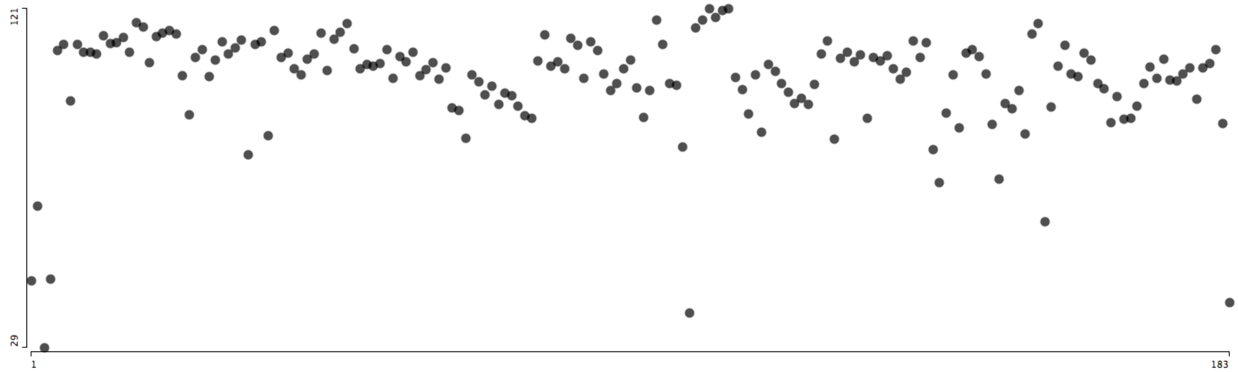


NOISE. X-aixs = page postion in the series. Y-axis = grayscale mean. Each chapter is shown in a separate color. Chapters are separate by white gaps; the title pages are omitted.

The next two visualizations compare A*bama* and *NOISE* stylistic development on low detail/texture – high detail/texture dimension as represented by entropy feature. (Here the plot of NOISE is scaled to the same width as the plot of Abara to make patterns more visible.)



Abara pages.
X-axis = page sequential position in the title (left to right).
Y-axis = entropy calculated over greyscale values of all pixels in a page.



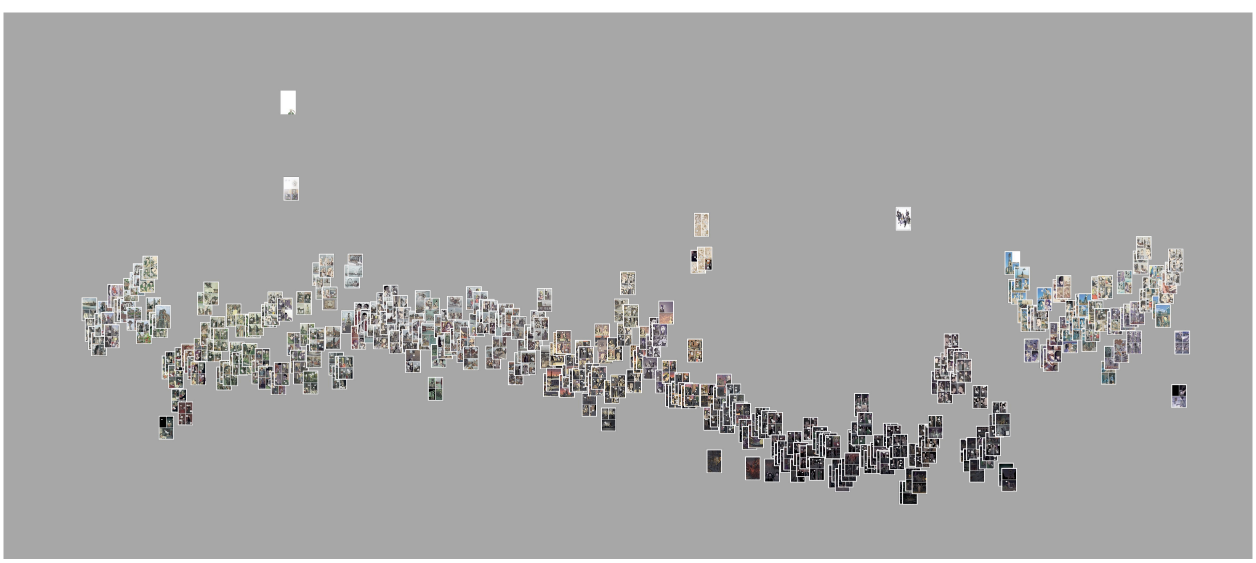
NOISE pages.
X-axis = page sequential position in the title (left to right).
Y-axis = entropy calculated over greyscale values of all pixels in a page.

In both titles, entropy values change over time. The pattern of these changes in *NOISE* can be described as a linear slight shift downward. The pattern in *Abara* is more complex. It can be characterized as a horizontal line, followed by a curve which first goes down and then goes up.

The two temporal patterns also have an interesting structural similarity. In each graph, the range between the top and bottom points (i.e.. the difference between their values on Y-axis) gradually increases. This means that stylistically the pages are at first are pretty similar, but become more varied over time. (Again, keep in mind that we are describing only one stylistic dimension.)

To illustrate interesting temporal patterns that can be revealed using such line graphs of feature values, we will go outside of our one million manga pages for the next example. We will look at the webcomic *Freakangels.* Our data set consists from 342 consecutive pages published over 15 months (2008-2009) in six-page chapters appearing weekly. Like in previous graphs, we use Y-axis to represent a single visual feature, and reserve X axis for pages publication order.

In the following visualization, Y position of each page is determined by the mean (average) of the greyscale values of all pixels in the page.



342 consecutive pages of the web comic *Freakangels* (2008-2009).
X-axis = page publication order.
Y-axis = mean of greyscale values of all pixels in a page. Greyscale value = (R value + G value + B value) / 3.

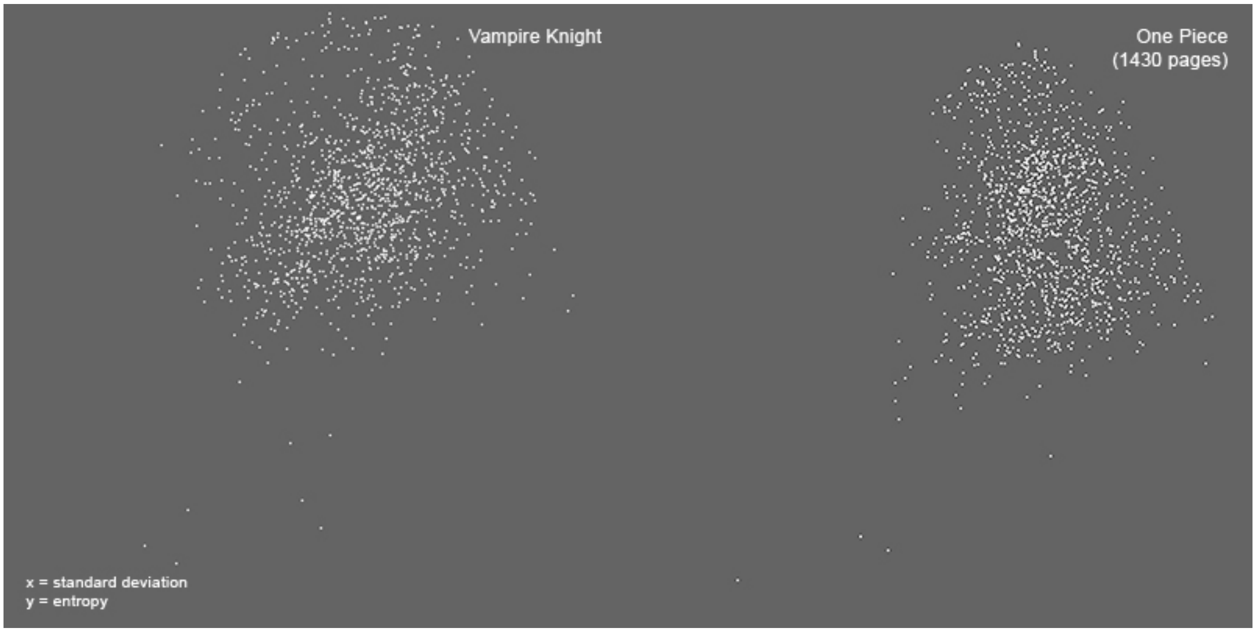
A close-up of visualizations of the web comic *Freakangels.*

Despite the weekly intervals that separate the 6-page episodes of *Freakangels*, visualization shows that visual values of the pages are remarkably consistent. For the larger part of the publication period, the changes in average grayscale values follow a smooth curve (the same applies to hue and saturation if we plot them). While the overall change from light to dark images corresponds to the development of the story from day to night, the fact that the grayscale values shifts very gradually and systematically over many months is a genuine discovery. Visualization reveals this unexpected pattern and allows us to see the exact shape of the curve.

COMPARING *VAMPIRE KNIGHT* AND *ONE PIECE SAMPLES* (2,744 PAGES)

Abara and *Noise* titles are quite short: 291 pages and 183 pages, respectively. How does our method scale for longer manga series such as *Vampire Knight* (57 chapters; 1423 pages) and *One Piece* (563 chapters; 9745 pages)?

Vampire Knight publication started in January 2005; *One Piece* started much in August 4, 1997. This explains the differences in the numbers of chapters and total pages in our download. To make a comparison more meaningful, we will only use only a part of *One Piece* data set: 481-563 chapters that contain 1321 pages. As in the earlier example, we will visualize these two sets of pages according to standard deviation (X-axis) and entropy (Y-axis). We graph the data using scatter plots; each page is represented by a point.



1423 *Vampire Knight* pages (left) and 1321 *One Piece* pages (right). X-axis = standard deviation. Y-axis = entropy.

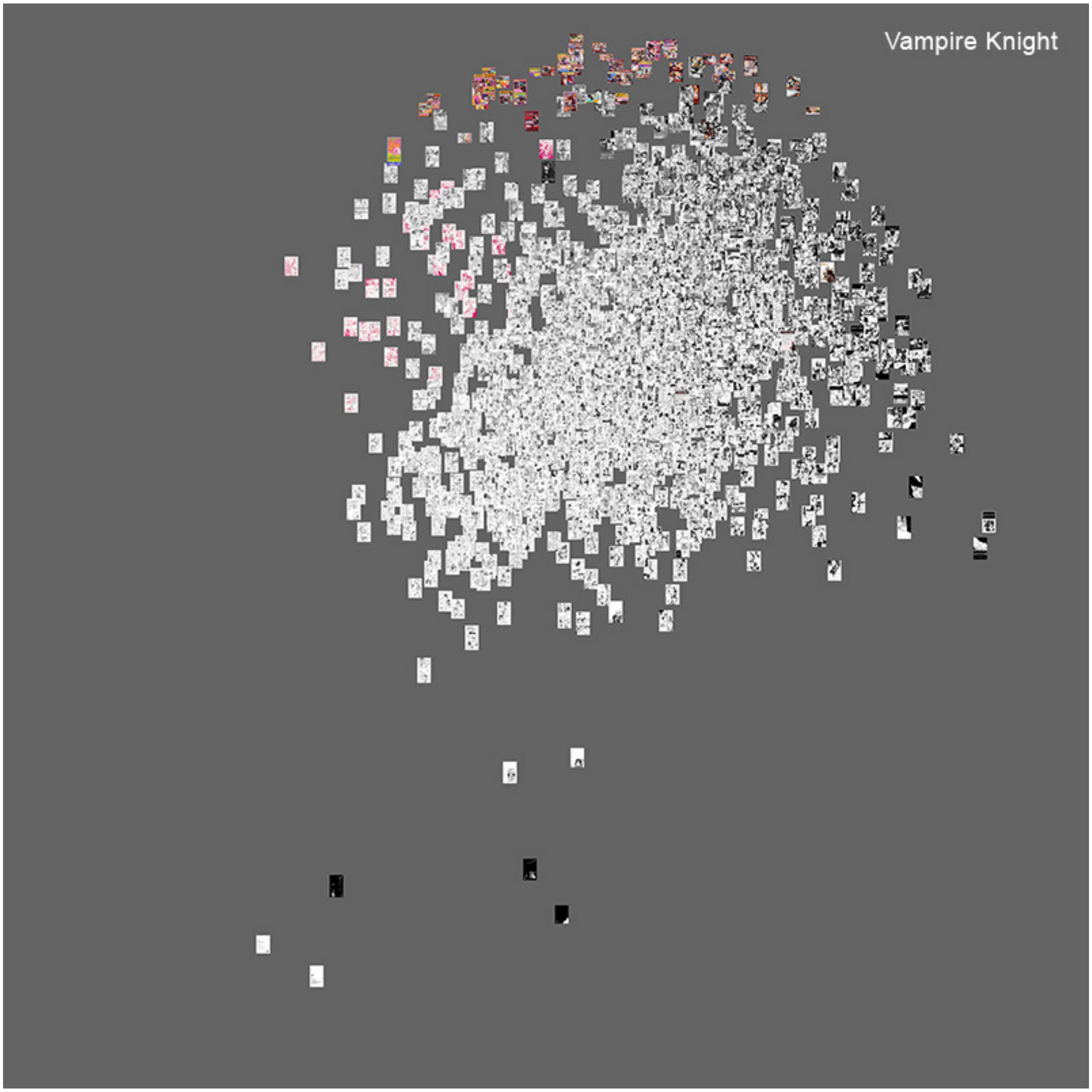
In each graph, X-axis and Y-axis start and end values are set as follows:
Std: min=0.000000, max=126.602400 (largest possible value which can be observed).
Entropy: min=-0.000000, max=7.962000 (largest possible value which can be observed).

We can notice that the center of the points in the left graph is higher than the center of the points in the right graph. Recall that Y-axis corresponds to low texture/detail – high texture/detail dimension. This confirms what we can see in the two sample pages: *Vampire Knight* page has more shading and detail than *One Piece* page. However, since each set of points also extends significantly along Y-axis, it is clear that we were lucky in our choice. We could have easily selected different pages that would lead us to an opposite idea about the graphic difference between the two sets.

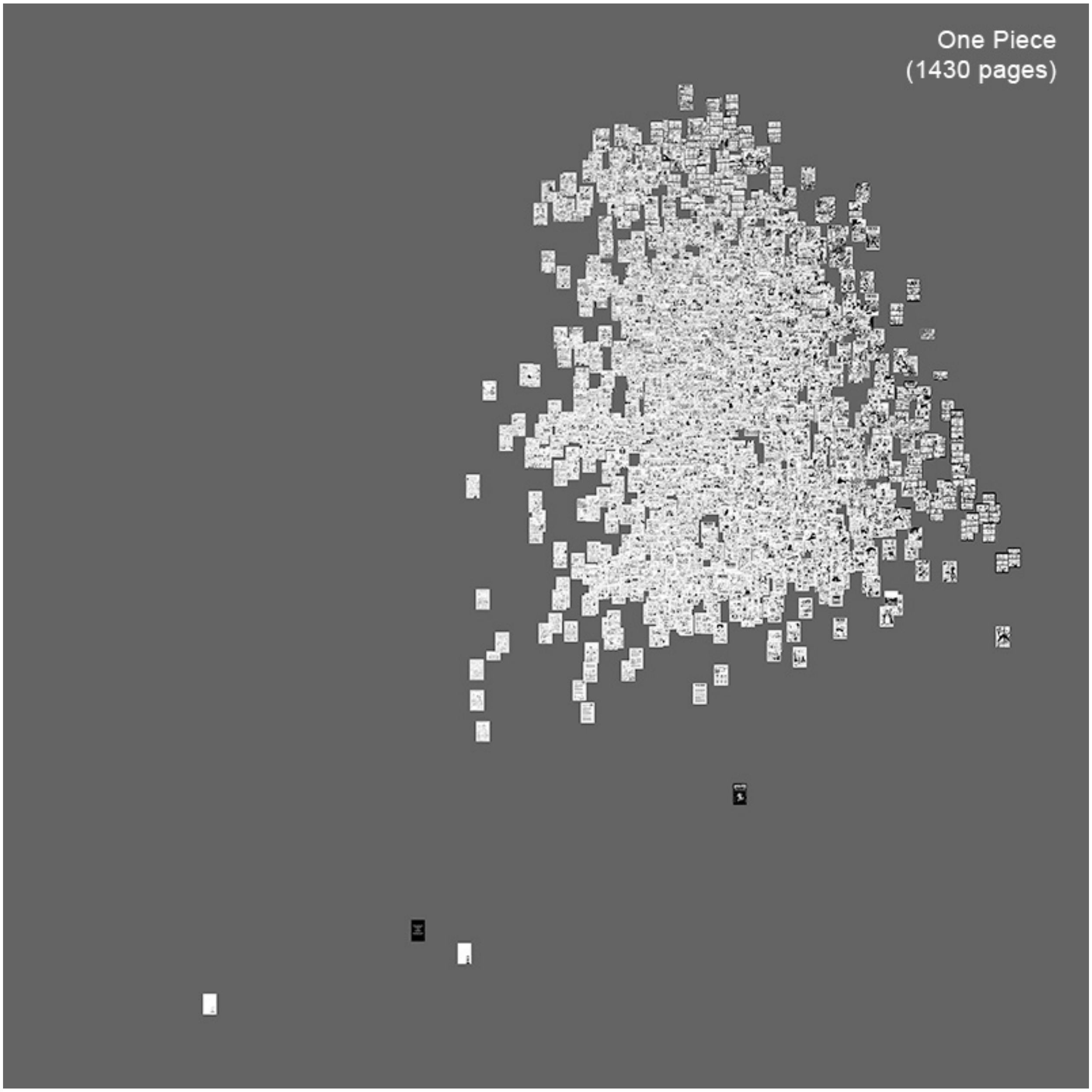
To confirm our observation about the vertical difference between the centers of the two sets of points, we calculate the actual averages of the data values projected onto Y-axis (entropy).

Vampire Knight: mean of entropy measurements of 1423 pages: 5.1.
One Piece: mean of entropy measurements of 1321 pages: 5.6.
(The averages are rounded to one digit).

To put this difference in perspective, it is useful to know the maximum possible value of entropy measurement of a single image is 7.962. This means that this value represents %6.4 of the total possible range.



1423 *Vampire Knight* pages. X-axis = standard deviation. Y-axis = entropy.



1430 *One Piece* pages. X-axis = standard deviation. Y-axis = entropy.

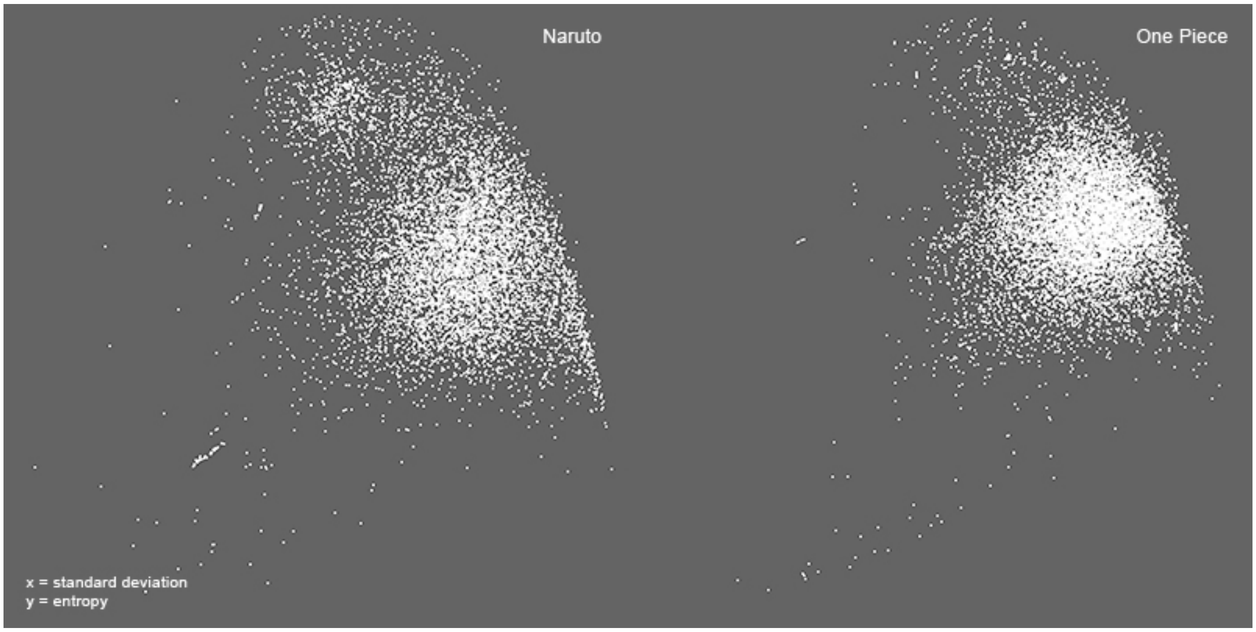
COMPARING *NARUTO* AND *ONE PIECE* (17,782 PAGES)

When we put side by side two sample pages from *Naruto* and *One Piece* manga series, we noticed that the differences in their graphical styles were more subtle than between the sample pages from *Vampire Knight* and *One Piece*. Can we better understand these more subtle differences using our method?

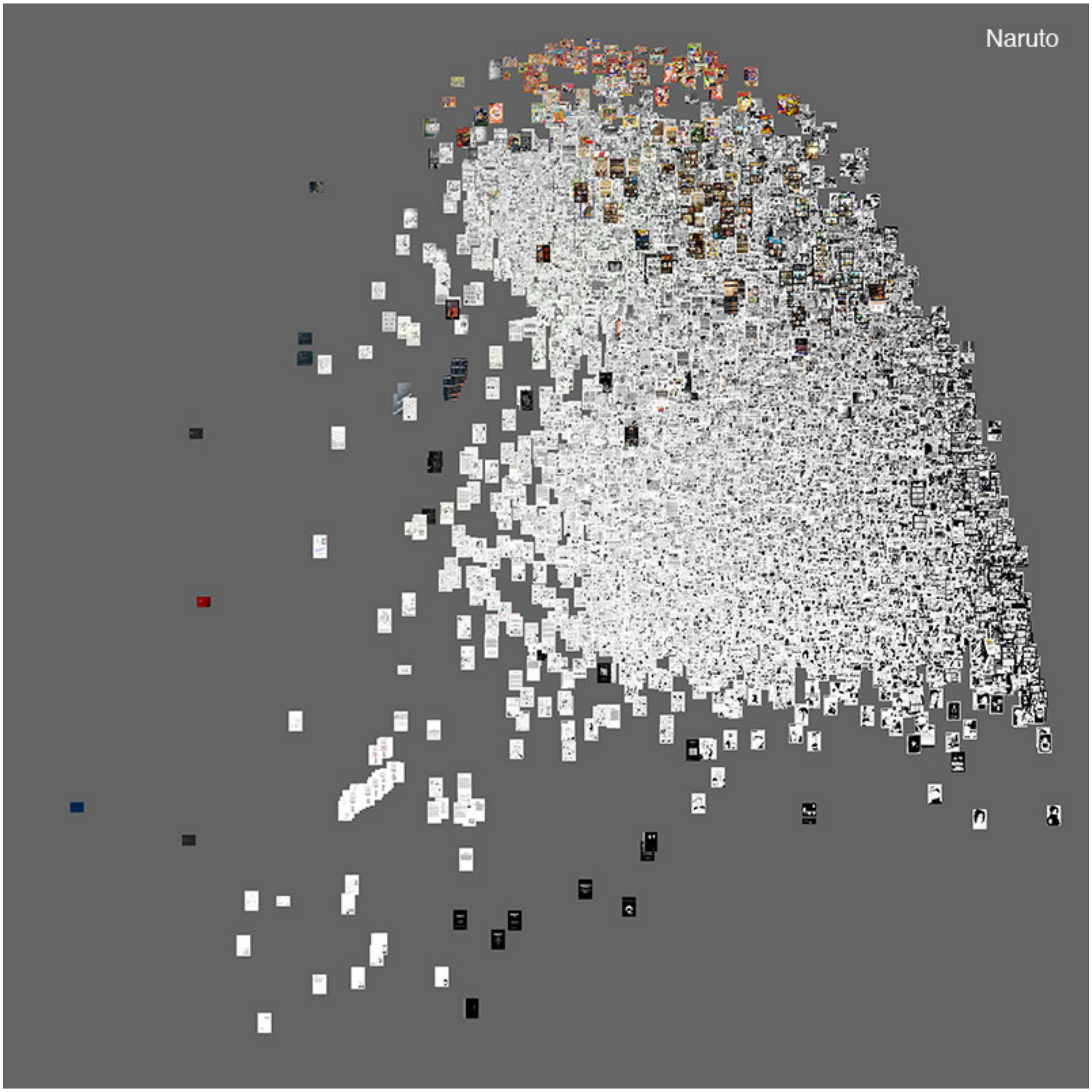
It is also interesting to compare *Naruto* (1999 -) and *One Piece (1997-)* because they among most popular manga series among global readers*.* These titles are rated no. 1 and no. 3 among OneManga.com global readers (One Manga). When we downloaded the pages from OneManga.com in the Fall 2009, the first series was published continuously for 10 years, and the second was published for 12 years. Accordingly, our download contained 8037 *Naruto* pages and 9745 *One*

Piece pages.

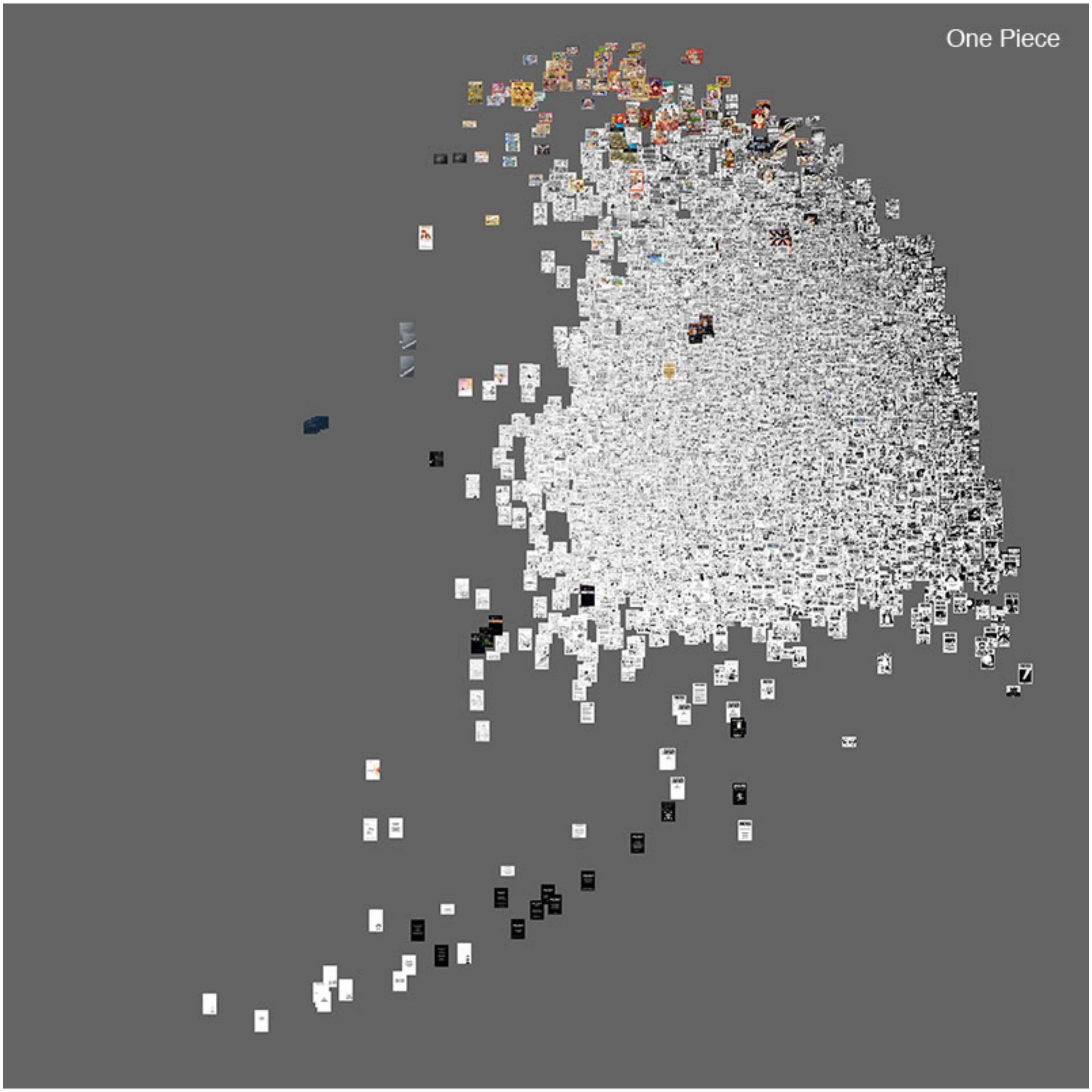
The following plots compare these two sets of pages.



Left: 8037 *Naruto* pages. Right: 9745 *One Piece* pages. Each page is represented by a point.
X-axis = standard deviation of greyscale values of all pixels in a page.
Y-axis = entropy calculated over greyscale values of all pixels in a page.



8037 *Naruto* pages. X-axis = standard deviation. Y-axis = entropy.



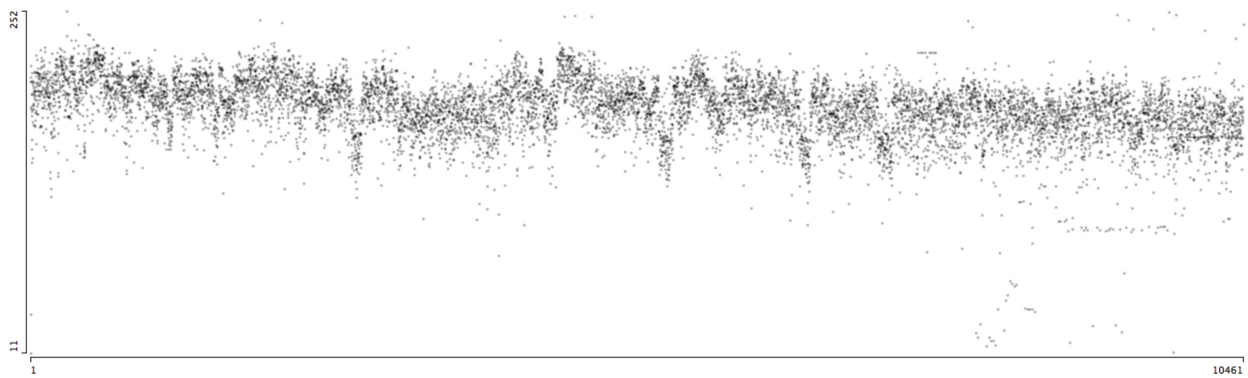
9745 *One Piece* pages. X-axis = standard deviation. Y-axis = entropy.

Projecting large number of pages from the two series into the same coordinate space helps us to better understand the similarities and the differences between their graphical styles. The visualizations show that along the two visual dimensions usedg, the distinctions between the languages of the two series are quantitative rather than qualitative. That is, the “point cloud” of *Naruto* pages significantly overlaps with the “point cloud” of *One Piece* pages both along and vertical axis.

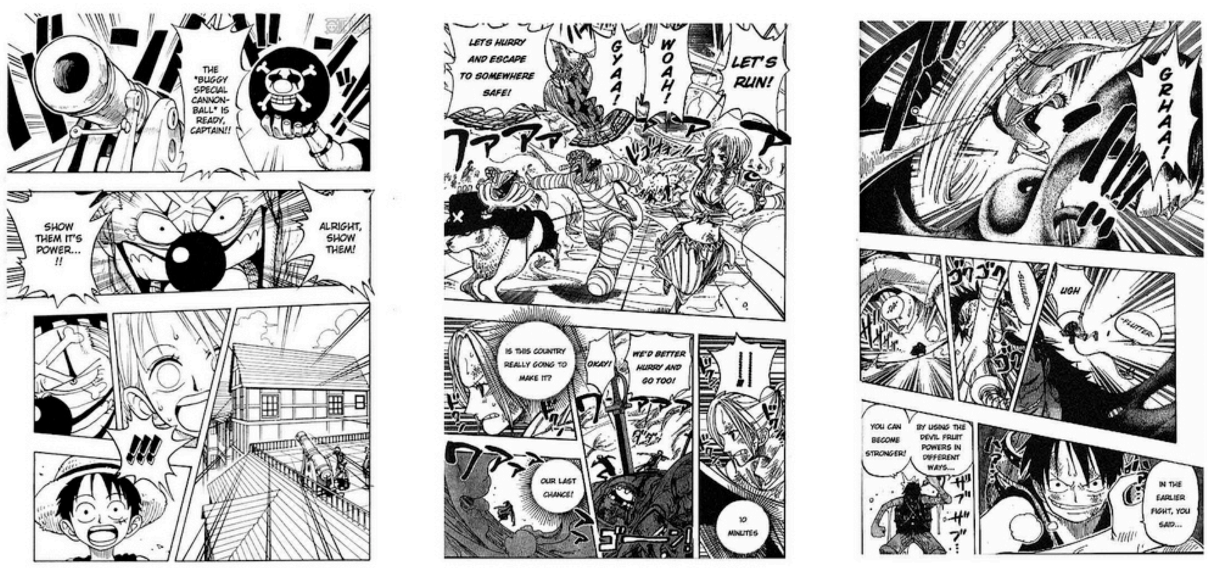
At the same, the differences between them are larger than a casual examination of only two pages implies. The visualizations reveal that both series cover a large range of graphical possibilities: from simple black and white pages with minimal detail and texture (lower part of each visualization) to the highly detailed and textured (top part). But the center of *One Piece* point cloud is slightly higher than the center of *Naruto* point cloud. This means that *One Piece* has more pages that have more textures and details that *One Piece* pages.

Visualizations also reveal the significant differences in the graphical variability between the two series. *Naruto*'s "point cloud" is much more disperse than *One Piece* "point cloud" both on horizontal and vertical dimensions. This indicates that *Naruto's* visual language is more diverse than the visual language of *One Piece*. (We already saw a similar difference when we compared *Abara* and *NOISE* – but now we are seeing this in a much larger data set.)

We can also examine the stylistic development of these long series over the time of publication in the same way used for much shorter *Abara* and *Noise.* The following graph plots 9745 *One Piece* pages left to right in the order of publication; the vertical position is determined by page grayscale mean. Below the graph are the three sample pages from which we already refered to earlier.



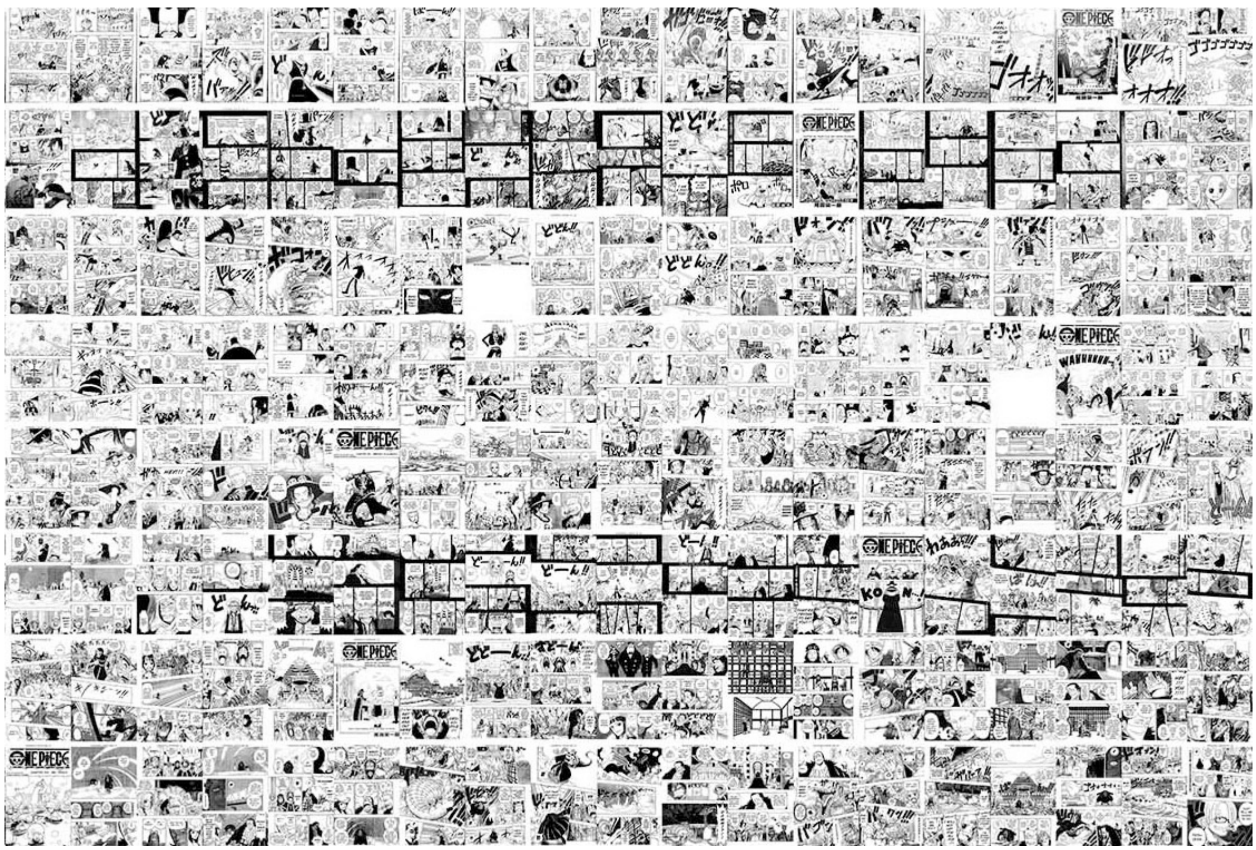
9745 *One Piece* pages (562 chapters).
X-axis = page position in publication order (left to right).
Y-axis = mean of greyscale values of all pixels in a page.



Sample pages from *One Piece* manga series drawn from the 5th, 200th, and 400th chapters.

Since we now dealing with thousands of pages in 562 weekly chapters published over 12 years, we can discuss temporal patterns at number of scales. On the scale of years, *One Piece* mean values gradually drift over the whole time period. Within this overall almost linear pattern, we see periodic raises and falls that reverse direction anywhere between 7 and 13 months. Thus, we get the answer to the question we asked earlier when we compared three sample pages drawn from the 5th, 200th, and 400th chapters – *how* does the visual language of *One Piece* changes over time?

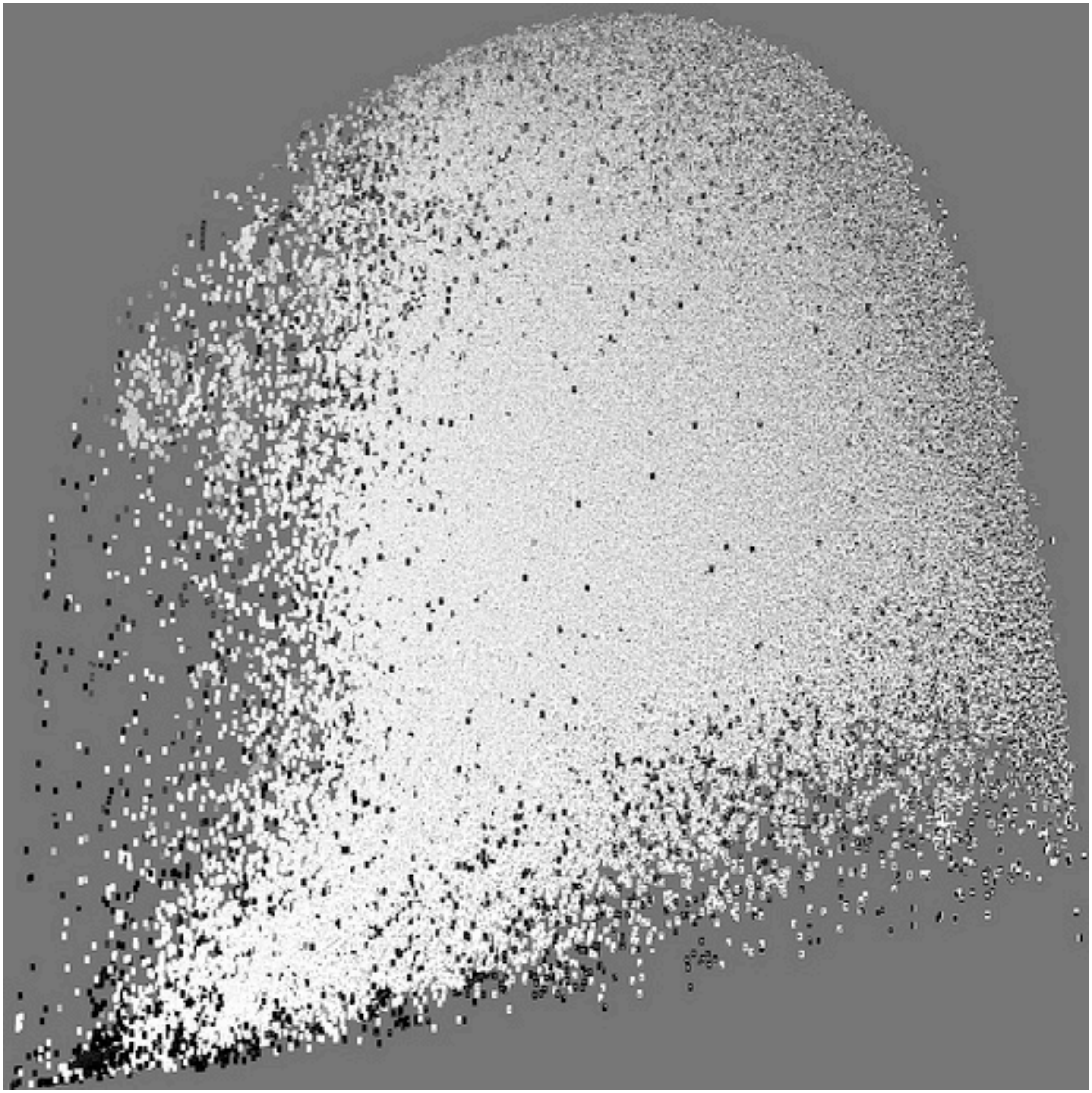
The visualization also reveals the patterns may not be visible if we only use a small number of pages. The three pages we look at earlier missed the periodic drops in grayscale values we can see when we plot all pages. The dips correspond to the flashback parts of the narrative which place the panels over black background, as can be seen in the close-up of the earlier montage of all *One Piece* pages.

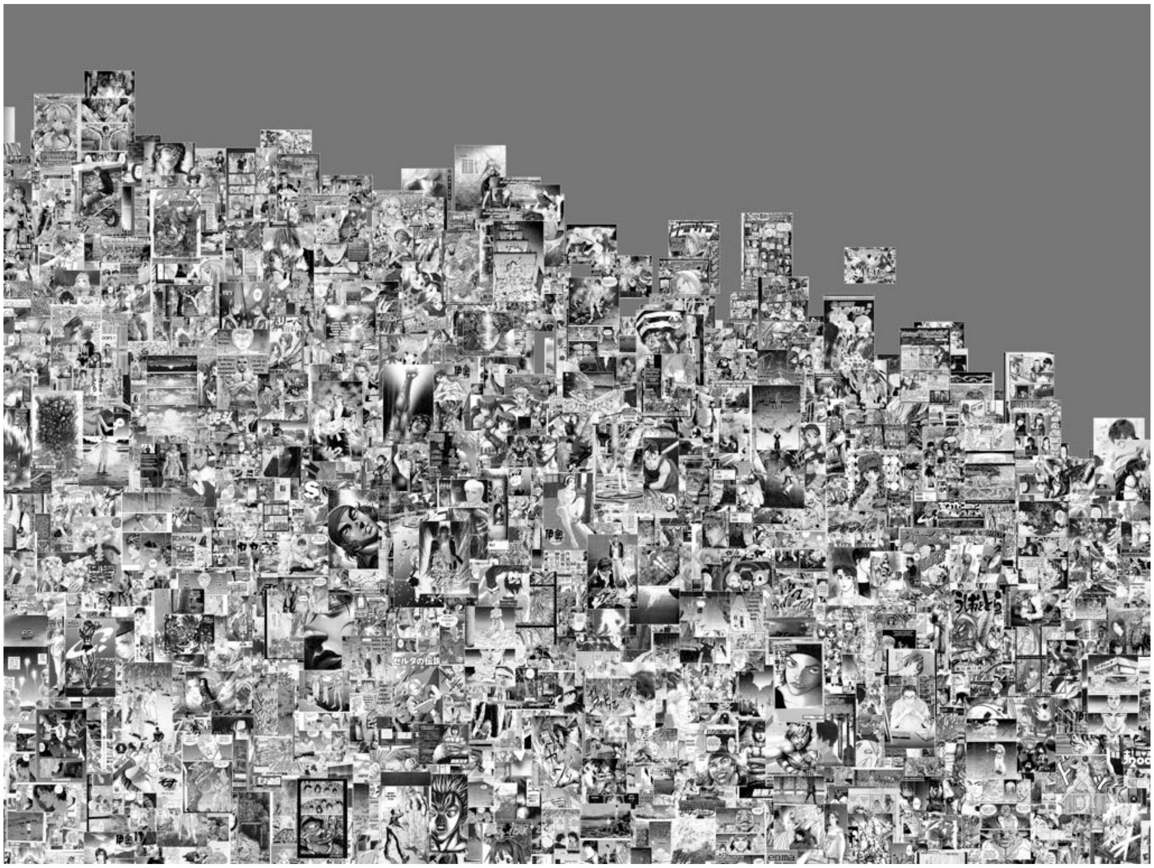


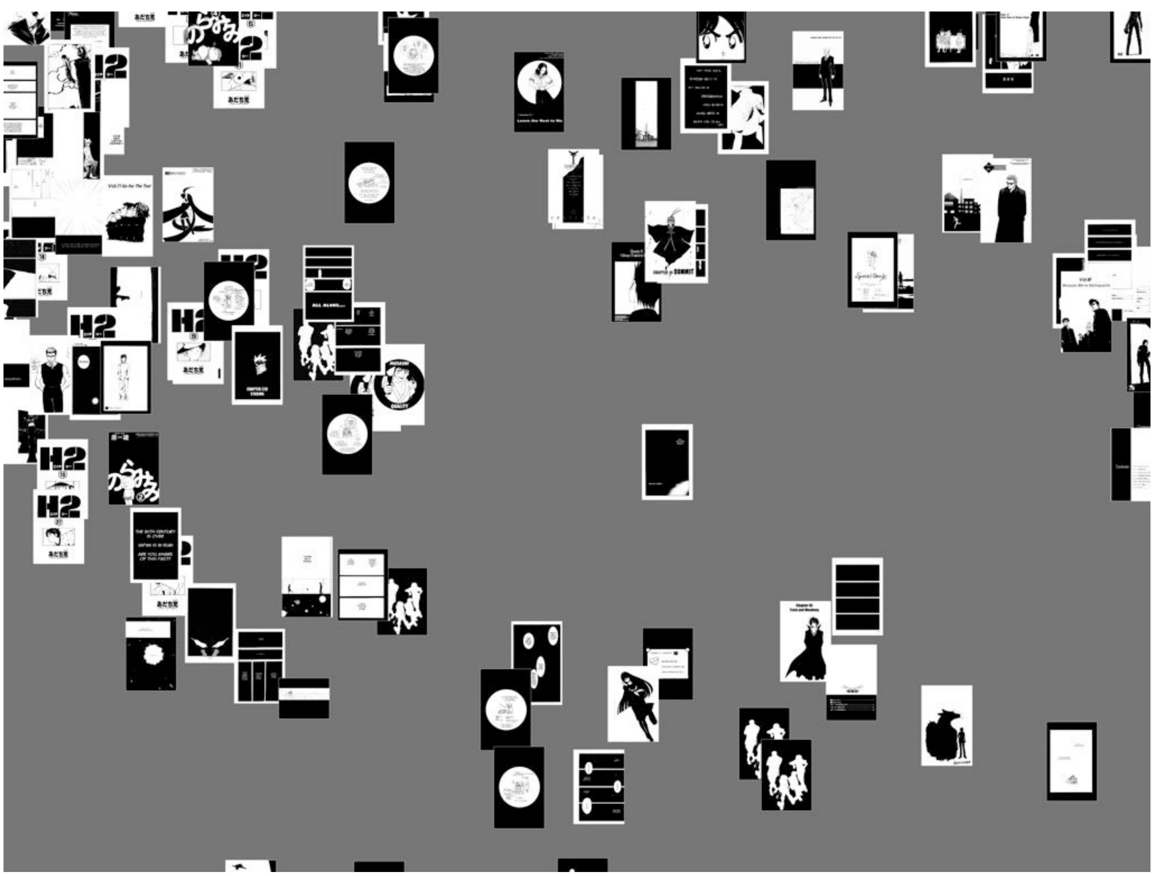
A close-up of the complete set of *One Piece* pages arranged in a grid in publication order (left to right, top to bottom).

VISUALIZING COMPLETE MANGA IMAGE SET (1,074,790 PAGES).

We can now finally provide an answer to the question in the chapter's title: how to see one million images? Using the same measurements and axes assignments (X-axis = standard deviation, Y-axis – entropy) as we did in the plots of individual titles and series, we visualize our complete set of one million pages. (Of course we can also organize this image set in many other ways using many other measurements – this is just one possible mapping.)







One million manga pages organized according to selected visual characteristics.
X-axis = standard deviation of greyscale values of all pixels in a page.
Y-axis = entropy calculated over greyscale values of all pixels in a page.

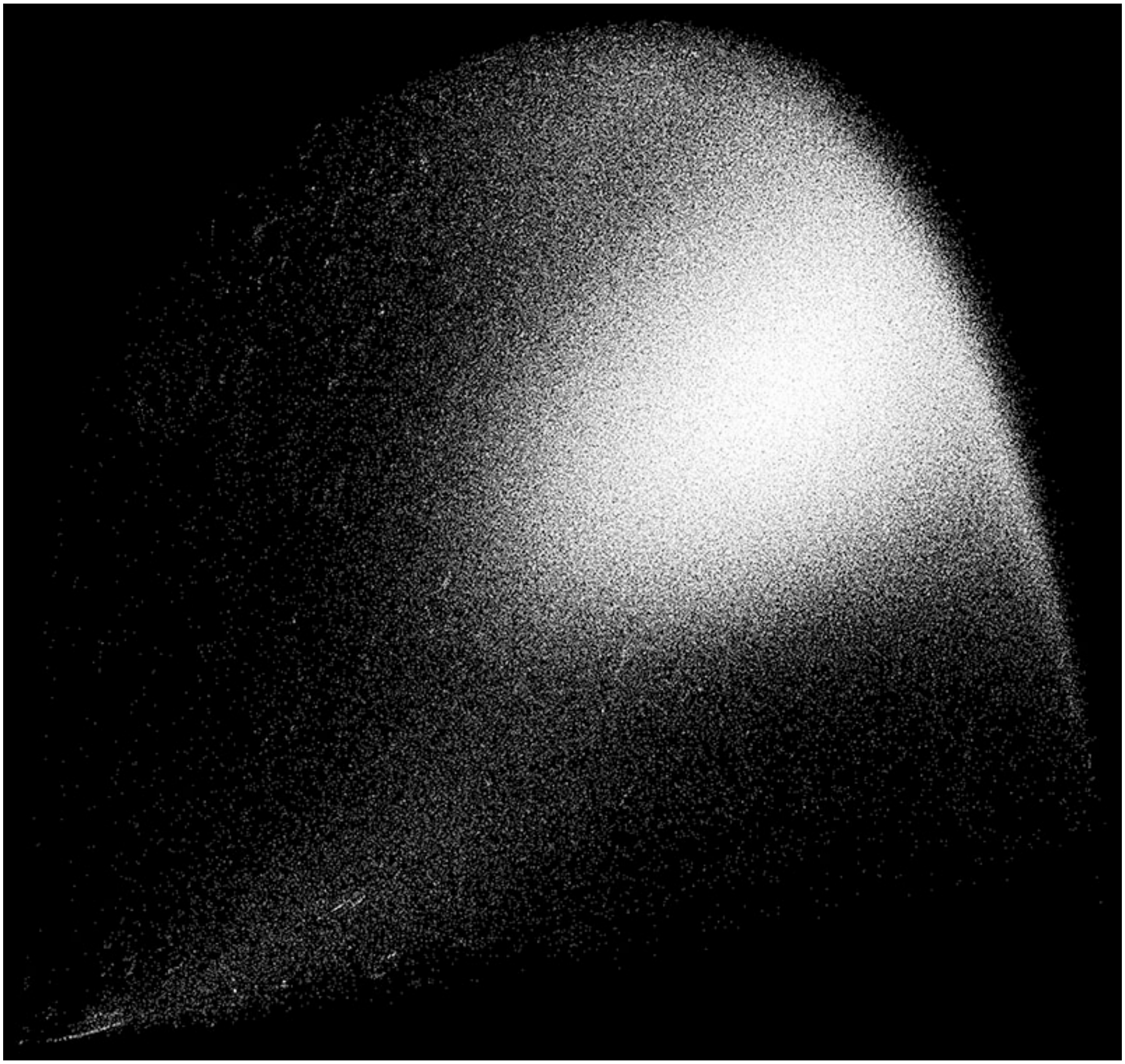
Top image: complete visualization.
Middle image close-up of the top part.
Bottom image: close-up of the bottom left corner..

Notes:
1) Some of the pages - such as all covers - are in color. In order to be able to fit all image into a single large image (the original is 44,000x44,000 pixels - scaled to 10,000x10,000 for posting to Flickr), we rendered everything in graycale.
2) Because pages are rendered on top of each other, you don't actually see one million of distinct pages – rather, the visualization shows a distribution of all pages with typical examples appearing on the top.

One million pages cover the space of graphical possibilities more broadly and with more density than *Naruto* or *One Piece* alone. In between the four graphical extremes corresponding to the left, right, top, and bottom edges of the pages "cloud," we find every possible intermediate graphic variation. This suggests that manga's graphic language should be understood as a continuous variable.

This, in turn, suggests that the very concept of *style* as it is normally used may become problematic then we consider very large cultural data sets. The concept assumes that we can partition a set of works into a small number of discrete categories. However, if we find a very large set of variations with very small differences between them (such as in this case of one million manga pages), it is no longer possible to use this concept. Instead, it is better to use visualization and mathematical descriptions to characterize the space of possible and realized variations.

To better understand the distribution of our data set within the space of all graphical possibilities, we can render the data from the last visualization using points. Such a scatter plot is not as easy to read as an image plot, however it is better in showing the shape of pages distribution. The plot shows that the distribution follows Bell-curve like pattern: single dense clusters with gradual fall off to the sides. The parts of the plot which remains black represent the graphical possibilities not realized in our manga sample: images which are almost completely white (lower right corner), and images which have large areas of black and small areas of white (left third of the plot).



One million manga pages rendered as points.
X-axis = standard deviation of greyscale values of all pixels in a page.
Y-axis = entropy calculated over greyscale values of all pixels in a page.

The fact that digital image processing and visualization of one million manga pages data set make us question the very basic concept of humanities and cultural criticism is at least as important as any particular discoveries we can make about this data set. It illustrates how computational analysis of massive cultural data sets has a potential to transform our theoretical and methodological paradigms for studying culture.

DEFAMILIARISATION WITH COMPUTERS

Alexander Rodchenko. Pine Trees in Pushkin Park. 1927. (Gelatin silver print.)

Our methodology relies on standard techniques of digital image analysis which started to be developed already in the second part of the 1950s and are now everywhere – in digital cameras, image editing software such as Photoshop, automated factories, medical imaging and all science fields which use images as sources of data (from astronomy to biology). However, when we adopt these techniques as tools for the cultural research, we should be clear about how they analyze images and what does it mean in general to see through computer "eyes." Since this chapter is focused on motivating and explaining our method in general terms, we would only make one observation. When we look at images normally, we experience all their visual dimensions at once. When we separate these dimensions using digital image analysis and visualization, we break this gestalt experience. Being able to examine a set of images along a singular visual dimension is a powerful mechanism of defamiliarization ("otstranenie") – a device for seeing what we could have not noticed previously. If avant-garde photographers, designers and filmmakers of the 1920 such as Rodchenko, Moholy-Nagi, Eistenstein, and Vertov were defamiliarizing the standard perception of visible reality using diagonal framing and unusual points of view, now we can use software to defamaliarize our perceptions of visual and media cultures.

ACKNOWLEDGMENTS

Software Studies Initiative research reported in this article was made possible by the generous support provided by California Institute for Telecommunication and Information (Calit2), Center for Research in Computing and the Arts at UCSD (CRCA), and Chancellor Office of University of California, San Diego (UCSD). The development of the custom digital image analysis software and the processing of manga image set on supercomputers at the National Energy Research Scientific Computing Center (NERSC) was funded by 2009 HEN Humanities High Performance Computing Award “Visualizing Patterns in Databases of Cultural Images and Video.” The development of ImagePlot software was funded by NEH 2010-2011 Digital Startup level II grant “Interactive Visualization of Image Collections for Humanities Research.”

CREDITS

The visualizations and data analysis and visualization tools developed by Software Studies Initiative are the result of systematic collaborations between the key lab members: Lev Manovich, Jeremy Douglass, William Huber, and Tara Zepel. HyperSpace software for interactive media visualization was developed by So Yamaoka (Gravity lab, Calit2). Batch image processing tools were developed by Sunsern Cheamanunkul (PhD student, Computer Science and Engineering Department, UCSD).

REFERENCES

Barry, Salt. “The Statistical Style Analysis of Motion Pictures.” 1974. Film Quarterly , 28, 1: 13-22.

Barry, Salt. Film Style and Technology: History and Analysis. London: Starword, 1992.

Cha, Meeyoung, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. “I Tube, You Tube, Everybody Tubes: Analyzing the World’s Largest User Generated Content Video System.” 2007 ACM Internet Measurement Conference. Web. 17 July, 2011. <http://an.kaist.ac.kr/traces/papers/imc131-cha.pdf>.

Cinemetrics.lv. Barry Salt’s Database. Web. August 11, 2011. <http://www.cinemetrics.lv/satltdb.php - cm>

Crandall, David J., Lars Backstrom, Daniel Huttenlocher, Jon Kleinberg. Mapping the world's photos. 18th international conference on World wide web, 2009. Web. July 17, 2011. <http://www.cs.cornell.edu/~dph/papers/photomap-www09.pdf>.

Douglass, Jeremy, William Huber, Lev Manovich. "Understanding scanlation: how to read one million fan-translated manga pages." Image and Narrative, 12, 1 (2011), 190–228. Brussels. Web. <http://lab.softwarestudies.com/2010/11/one-million-manga-pages.html>.

Gabor filter. “Gabor Filter.” Wikipedia. Web. August 11, 2011. <http://en.wikipedia.org/wiki/Gabor_filter>.

Herring, Susan, Lois Ann Scheidt, Inna Kouper, and Elijah Wright. A Longitudinal Content Analysis of Weblogs: 2003-2004. Blogging, Citizenship and the Future of Media. London: Routledge, 2006.

IMDb. IMDb Database Statistics. Web. July 17, 2011. <http://www.imdb.com/stats>.

Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. "What is Twitter, a Social Network or a News Media?" The 19th international conference on World Wide Web, 2010. July 17, 2011. Web. <an.kaist.ac.kr/~haewoon/papers/2010-www-twitter.pdf>.

Software Studies Initiative. 2011. ImagePlot. Open source software for visualization of large image and video collections. Web. <http://lab.softwarestudies.com/p/imageplot.html>

McLoud, Scott. Understanding Comics: The Invisible Art. Kitchen Sink Press, 1993.

Manovich, Lev. "Cultural Analytics: Visualing Cultural Patterns in the Era of 'More Media'". Domus, Spring 2009. Web. <http://lab.softwarestudies.com/2009/06/publications.html#CA_Publications>.

Manovich, Lev. 2010. "Media Visualization: Visual Techniques for Exploring Large Media Collections." Media Studies Futures, ed. Kelly Gates. Blackwell, forthcoming 2012. July 17, 2011. Web. < http://manovich.net/DOCS/media_visualization.2011.pdf>.

Moretti, Franco. Conjectures on World Literature. New Left Review 1 (2000), 55–67. July 17, 2011. Web. < http://www.newleftreview.org/A2094>.

One Manga. onemanga.com top 50 manga list. August, 12, 2011. <http://www.onemanga.com/directory/top/>.

Peters, Carol, ed. MultiMatch: Technology-enhanced Learning and Access to Cultural Heritage. D1.1.3 – State of the Art report section. 2006. July 17, 2011. Web. <http://www.multimatch.eu/docs/publicdels/D1.1.2.pdf>

Treisman, Anne and Garry Gelade (1980). "A feature-integration theory of attention." Cognitive Psychology, Vol. 12, No. 1, pp. 97-136.

Ward, L.M. (December 2003). "Synchronous neural oscillations and cognitive processes". Trends Cogn. Sci. **7** (12): 553–9.

Williams, D., N. Martins, M. Consalvo and J. Ivory. 2009. "The Virtual Census: Representations of Gender, Race and Age in Video Games." New Media & Society. 11, p. 815-834.

Yamaoka, S., L. Manovich, J. Douglass, F. Kuester. 2011. "Cultural Analytics in Large-Scale Visualization Environments." Submitted to Computer Graphics Forum.