Fall 2011. UCSD
ICAM 130 / VAF 149: Contemporary Computer Topics

COURSE TOPIC

Seeing cultural patterns:
Visualization, digital image processing, and big data

INFORMATION

Instructor: Lev Manovich
www.manovich.net
www.softwarestudies.com
e-mail: manovich@ucsd.edu
Office hours: Tuesday 2:15-3:15, Cafe Roma (or by appt.)

classroom: VAF 228 (Visual Arts Facility)

Monday
3:00p - 5:50p
VIS 149. section ID: 724584
ICAM 130. section ID: 724625

Monday students progress spreadsheet

Tuesday
3:30p - 6:20p
VIS 149. section ID: 724585
ICAM 130. section ID: 724626

Tuesdays students progress spreadsheet

COURSE DESCRIPTION

“The next big idea in language, history and the arts? Data.”
New York Times, November 16, 2010.

The course introduces students to the techniques of information visualization, including both classical plots and graphs, and new visualization methods for working with visual data (i.e., images and videos). We explore how these techniques can be used together with digital image processing to visualize patterns in large cultural visual data sets.

The course topics cover all the steps necessary for going from available data sources to final visualization projects:

web scraping;
organizing and cleaning data;
graphs of one, two, and multiple variables;
preparing images for visualization;
exploring image sets with ImageJ (montage, slice, average)
using image processing to extract visual features from images;
using ImagePlot to create high resolution static and animated visualizations;
interpreting and writing about visualizations of cultural data;
promoting projects through social media.

Software:
The course uses as much as possible free data analysis and visualization tools: manyeyes, Google docs, Mondrian, ImageJ, and ImagePlot (developed by Software Studies Initiative at UCSD).

We will also have access to state of the art supervisualization system at Calit2 (HIPerSpace) for presentation of final projects.

RELATED COURSES

145B: mage analysis and visualization techniques for digital humanities (Manovich, UCSD, spring 2011)

digital humanities++ (Manovich, UCSD, spring 2011)

MAT259: Visualizing Information (Legrady, UCSB)

CS 171: Visualization (Pfister, Harvard)

COURSE REQUIREMENTS AND POLICIES

Attendance

Attendance will taken every class at the beginning. You are allowed to miss one class meeting without an excuse.

Any additional absense without a proper excuse (doctor's notice) will lower your final grade by half a letter grade (for example, B becomes B-.)

Grading:

Course grading:
A 90-100%
B 80-89%
C 70-79%
D 60-69%
F 0-59%

Course participation: 10%.

6 weekly practical assignments: 60%.
(each simple assignment: 6%; each complex assignment: %12).

Assignments grading:

3 points: extra credit for excellent work.
2 points: done correctly and submitted on time.
1 point: partly done and submitted on time.
0: not correct, or not submitted.

You can submit an assignment after the due date, but you will only get 1 point.
Homework submissions will not be accepted after three days following the due date.

Final project: 30%.

Final project details:

Students will work on final projects in groups.
The final project should present analysis and visualizations of large visual data sets (at least 50,000 images in total) in the form of a short essay/ blog post (600-1200 words) with images. The essay can be delivered as a web page (.html file) or as a .doc file (Word document). The group also needs to submit all high res visualizations and the data sets used in the analysis. Please also indicate the roles and the contributions of each group member.

The students will be given a few large visual data sets already prepared by our lab to choose from. Alternatively, a group can download their own data set after consultation with the instructor. The data sets may come from visual art, graphic design, photography, fashion/street styles, feature films, animation, motion graphics, use-generated video, gameplay video recordings, web design, product design, maps, sounds, texts, or other media.

Final Projects due: Monday, December 12, 9am.

Upload your project folder into “assignments” directory on jeju.ucsd.edu

Your project folder should contain the following:

1) your short essay (.doc or HTML format).
Make sure this document contains **the names of all the students in the group**, and a clear **description of the contributions of each student**.

2) a subfolder named **Scaled_visualizations**.
Place scaled down versions of the visualizations and any other illustrations you are using in your essay in this subfolder. (If you are using HTML format for your essay, link to these images. If you are making .doc document, you can directly paste these images into the doc.)

3) a subfolder named **Full_size_visualizations**. Place full-size visualizations in this subfolder.

They should have exactly the same names as the scaled down versions - however, for scaled versions, add “scaled” to their names; for full versions, add “full”.

For instance:

scaled down version:

New_York_Tribune.1886_1886.Montage.rows_40_columns_40.w900_h600.scaled.jpg

full version:

New_York_Tribune.1886_1886.Montage.rows_40_columns_40.w12000_h8000.full.jpg

4) a subfolder named **Data**. If you did any measurements of the images, or added new metadata and used them in visualizations, put these measurement files in this subfolder. Make sure that the data files have descriptive names, and also a label row with clear descriptions of all columns.

5) if you downloaded new data (Newspaper group) and your images are on our lab iMac, include a text file indicating the start and end date for each newspaper (or its part) are are using, the name of the folder containing the pages, and the newspaper information provided by The Library of Congress.

why (why this is relevant), what (data sets), how (visualization techniques), and findings (interpetations of the results).

Academic Integrity

Integrity of scholarship is essential for an academic community. The University expects that both faculty and students will honor this principle and in so doing protect the validity of University intellectual work. For students, this means that all academic work will be done by the individual to whom it is assigned, without unauthorized aid of any kind.

In this course, we expect that all assigned practical assignment are done by students individually. Group projects, on the other hand, are designed to be done collaboratively.

Students with Disabilities

Students requesting accommodations and services due to a disability for this course need to provide a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities (OSD), prior to eligibility for requests. Receipt of AFAs in advance is necessary for appropriate planning for the provision of reasonable accommodations. OSD Academic Liaisons also need to receive current AFA letters.

For additional information, contact the Office for Students with Disabilities:
858.534.4382 (V)
858.534.9709 (TTY) - Reserved for people who are deaf or hard of hearing
osd@ucsd.edu
http://disabilities.ucsd.edu

REQUIRED READINGS

all required readings will be available online at no charge

COURSE SCHEDULE

class 1:
course introduction and overview of topics

homework for class 2:

Practical assignment (1): (%6 of the final grade):

add your information to the class list (skills, interests, etc).
(You will receive email containing a form. Fill out the form and click “submit.”
I sent this on 9/29, 11pm to your UCSD email addresses.)

Read:

1) Manuel Lima. Outburst of Visualization (2010). in Lev Manovich, Jeremy Douglass, William Huber. *Mapping Time*. gallery@calit2, forthcoming 2011.

2) In 500 Billion Words, New Window on Culture. NYT, 12/16/2010.

visit the following websites:

http://ngrams.googlelabs.com
Digging Into Data 2001 Competition (read introductory paragraph)
infosthetics.com
visualcomplexity.com

---

class 2:
organizing and cleaning data; web scraping

practical assignment 1 due

software:
Excel
Google spreadsheets
SiteSucker
(optional: Google Refine, Fusion Table)

do in class - practical:

give students access code to VAF 228
collect gmail addresses
add students gmail addresses and link <students progress Google spredsheet>
add students gmail addresses to practical assignment 2 spreadsheet

class topics:

common data formats for visualization:
text document, network data format, table (XML, spatial data, database, etc.)

common formats for storing data tables:
spreadsheet, tab-delimited text file, comma separated text file,

converting between table file formats.

data types: integers, floating point numbers, strings

working with spreadsheet data using functions

using functions in Google spreadsheets to add external information
example: using city names to get coordinates

converting web page tables into Google spreadsheets

example:

a list of van Gogh paintings from Wikipedia page
cleaned data using Google spreadsheets functions
the titles (with periods added) data uploaded to manyeyes

optional topics:
web scraping with SiteSucker
Scraping for Journalism: A Guide for Collecting Data
Data Driven journalism (Wikipedia)

discuss:
the rise of information visualization (Manuel Lima).

homework for class 3:

download Mondrian software to your laptop

create an account on manyeyes

download Dropbox software to your laptop, and create account

practical assignment (2): (%6 of the final grade)

1)
Monday students:
download selected chapters of Vampire Kinight and Naruto assigned to you

Mondays course assignments

Tuesdays students:
download selected chapters of Naruto assigned to you

Tuesday course assignments

To download using Chrome browser:
File -> Save Page as..
Select format: Web page, complete

2) name the directories for each chapter according to the template in the

assignment spreadsheet.

3) delete all web files from each directory - leave only pages (i.e., jpeg images).

4) delete all color pages, pages created by scanlators and any other pages which do not belong to the narrative.

5) create a tab delimited text file (.txt) containing data about the downloaded files.

The data file should have 4 columns:
manga title (string), chapter number (int), chapter directory (string), image filename (string).

Example:

Vampire Kinight	1	Vampire_Knight_ch_1	1.jpg
Vampire Kinight	1	Vampire_Knight_ch_1	2.jpg
Vampire Kinight	1	Vampire_Knight_ch_1	3.jpg
...			
Vampire Kinight	2	Vampire_Knight_ch_2	1.jpg
Vampire Kinight	2	Vampire_Knight_ch_2	2.jpg

...

Do not create header row (i.e. row with labels), because we will be combining all data files into one.

The data file should be named according to this template:

MangaTitle_ch_FirstChapterNum_SecondChapterNum.txt

Examples:

Vampire_Knight_ch_1_5.txt

Naruto_ch_400_410.txt

6) copy the directories with manga pages and the data file to <manga> folder inside Dropbox (I sent invites to your gmail accounts to this folder).

read:

1) Jeremy Douglass, William Huber, Lev Manovich. Understanding scanlation: how to read one million fan-translated manga pages. Image and Narrative

(Brussels, 2011).

2) explanation of basic graph types (from manyeyes):

bar chart
line graph
scatter plot

optional: some graph types specific to manyeyes:
phrase net
network diagram
world map

3) selected projects from datavisualization.ch:

How We Visualized 23 Years of Geo Bee Contests
Visualizing The World's Well-Being
How We Visualized America's Food and Drink Spending
Visualizing The Health Care Reform

---

class 3:
visualizing data: one, two and multiple variables

practical assignment 2 due

software:
Mondrian,
(optional: manyeyes, Google spreadsheet, TextWrangler, Tableau)

class topics:

visualizing one variable (bar chart, histogram)

visualizing time series (line graph)

visualizing two variables (scatter plot)

exploring relations between multiple variables (linked views)

visualizing multiple variables (parallel coordinates, scatter plot matrix, PCA) visualizing tags data

Homework for class 4:

practical assignment (3) (%12 of the final grade):

DUE DATE: OCT 24 (Monday class) or OCT 25 (Tuesday class)

Create an infographic poster which uses tags data for all images from two Flickr groups: Art Now (appr. 170,000 images) and Graphic Design (170,000 images).

Examples of good infographic posters:

type “inforgraphics” (without quotes) into search box of infosthetics.com

check New York Times infographics (view by “rating” to see the best ones)

Assignment details:

Flickr groups tags are stored in tab-delimited text files. The files are in Dropbox in manga > DATA_FILES_ASSINGNMENT_3 folder:

Flickr.ArtNow.image_tags.txt
Flickr.Graphic_Design.image_tags.txt

Optional:
you can also use two data fies that contain titles of all images:

Flickr.ArtNow.image_titles.txt
Flickr.Graphic_Design.image_titles.txt

The poster should contain **multiple graphs of the data**; it can contain text (which can be selected tags from the data set, descriptions of the graphs, interpetations of patterns, general notes, poster title, etc.), and other graphic elements.

It should not contain any raster (bit-mapped) images.

You need to **indicate what data is being represented in each graph, and the units of data**. You can do this by adding a label to each graphical element of the graph, or by adding labels to axis lines, or by adding a text label. if a graph is not self-explanatory, you need to provide a description explaining it.

You can design your poster using any software you like (Photoshop, Illustrator, etc.)

You don't need to use all the tags data for your posters - however if you do use complete data you will be given extra credit.

Your finished design should follow these specifications:

dimensions: 8.5 inches x 11 inches.

resolution: 300 dpi.

file formats: save your finished design as JPEG (quality - high), and as PDF.

How to submit: email your JPEG and PDF files (as attachments) to manovich@ucsd.edu**.**
Put "Flickr groups poster" in the subject of the email.
Put your last name, first name, and "Monday" or "Tuesday" (depending on the day your course is meeting) in the body of the email.

Please follow these specifications exactly if you want to get credit for your submission.

Grading criteria:
1) Does your poster presents interesting and unexpected patterns and relationships in the data?
2) Do these patterns and relationships are communicated visually in an easy to understand and expressive way?
3) Did you provide clear explanations of your graphs (via labels and/or text descriptions)?
4) Are you using all the data, or only small parts? (if you use complete data sets, you will be given more points.)

class 4: no class meeting - work on assigned homework

Homework for class 5:

download ImagePlot.zip

go though ImagePlot documentation (including all tutorials)

Read:

1) Lev Manovich, Jeremy Douglass, Tara Zepel, William Huber.
"How to Compare One Million Images?" article draft, 8/2011.

2) Lev Manovich. Style Space (parts 1 - 4), published on softwarestudies.com, 8/-102011.

original class plan:
preparing images for visualization; exploring image collections with ImageJ (montage, slice, average); sampling methods for media visualization

software:
ImageJ, Photoshop, Automator

read, see, discuss:
visualizing media in modern and contemporary art
“Visualizing Large Media Collections for Humanities Research” (Manovich)
articles about manga’s visual language

class 5:
introduction to image processing;
(review) preparing images for visualization using ImageJ (scale); extracting features from images; media visualization with ImagePlot.

ASSIGNMENT 3 DUE

Homework for class 6:

practical assignment 4 (%12 of final grade):

Create visualizations revealing patterns in van Gogh's paintings

Assignment details:

776 images of van Gogh paintings and the data file are distributed with ImagePlot software.

At least **one** of your visualization should be done with ImagePlot; it should show images of the paintings (you dont need to include all the images).
You can use ImagePlot or any other software for your other visualizations.

You don't need to put all our visualizations in a single poster - just submit them as separate files. However, feel free to combine some of your visualizations into a single image, if this makes it easier to compare them. Here is an one example - histograms of median brightness of van Gogh's early, Paris and Arles paintings. Here is another example - imageplots of van Gogh's and Gauguin paintings.

File formats:

Use JPEG file format at "good" or "best" quality for high resolution visualizations containing images.

Use PNG file format for visualizations containing vector data (i.e., histograms, scatter plots, bar charts, etc.).

Label each visualization using this template:

class meeting day.Your last name.van_Gogh.software used.description of the visualization. file extension

Examples:

Monday.Manovich.van_Gogh.ImagePlot.images.X_yearmonth.Y_median_brightness.jpg

Monday.Manovich.van_Gogh.ImagePlot.points.X_median_brightness.Y_median_saturation.png

Monday.Manovich.van_Gogh.ImagePlot.points.X_median_brightness.Y_median_saturation.png

Tuesday.Manovich.van_Gogh.Mondrian.histogram.median_brightness.10_bins.png

Descriptions:

Submit a single file in .txt or .doc format describing each of your visualizations. The descriptions should explain the data used and what visualization shows (for example, for scatteplot, explain what features/metadata are used for for X-axis and Y-axis.) They should also contain brief (no more than one paragraph) discussions of the visualization.

For examples, see the descriptions of the visualizations in this Flickr set (these visualizations were shown at Mapping Time exhibition at gallery@calit, Fall 2010).

Examples:

Here are a few possible directions you can take - note that they are provided as examples only, so feel free to explore other ideas:

1) Add metadata describing the content and/or form to van Gogh's data file. For example, you can tag each painting using a set of genre tages: portrait, landscape, city view, still-life, composition (for paintings which were not done from life). Or, you can label visual properties of the paintings which are not captured in features extracted by ImageMeasure.txt - for instance, you can tag paintings compositions: diagonal, vertical/horizontal, etc. Or you can tag the "energy" you feel a painting creates on some scale, such as: 1 - very quite; 2 - quite; 3 - some energy; 4 - a lot of energy; 5 - extreme energy.

You can then use this additional metadata in your visualizations. For instance, you can explore temporal patterns (X-axis: yearmonth; Y-axis: median brightness; etc.) separately for landscapes, still-lives, portraits, etc.

2) Then we think about van Gogh art, we typically refer to only a small number of his famous paintings. The same holds for various web sites about van Gogh - from Wikipedia article on van Gogh to Vincent van Gogh Museum in Amsterdam.

You can graph the patterns defined by these famous paintings and compare them to patterns in our complete set of 776 van Gogh paintings. How does our understanding of van Gogh changes if we look at the majority of his paintings as opposed to just a few?

3) Find descriptions of van Gogh styles in different periods of his life, his works in different genre, etc. in scholarly or popular books and articles about him. Test these descriptions by visualizing all or parts of van Gogh data set. Discuss if your analysis supports the descriptions, proves them wrong, or requires their adjustements.

Example: the web site of Vincent van Gogh Museum in Amsterdam describes the changes in van Gogh after his move to Paris in the Spring 1886 as follows:

“As he begins to formulate his own artistic idiom, he progresses through the styles and subjects of the Impressionists. His palette becomes brighter, his brushwork more broken.”

However, a visualization of average brightness of van Gogh paintings over time reveals that brightness starts gradually increasing much earlier - from approximately 8/1885 (van Gogh moves to Paris in 3/1886).

How to submit:

Zip your visualization files and description file and email this .zip archive to manovich@ucsd.edu.

Put “van Gogh visualizations” in the subject of the email.

Put your last name, first name, and “Monday” or “Tuesday” (depending on the day your course is meeting) in the body of the email.

Read:

1) Lev Manovich. "Media Visualization: Visual Techniques for Exploring Large Media Collections." In Media Studies Futures, ed. Kelly Gates. Blackwell,

forthcoming 2012.

2) Cohn, Neil. 2011. A Different Kind of Cultural Frame: An Analysis of Panels in American Comics and Japanese Manga. Image [&] Narrative 12 (1):120-134.

3) Valérie Cools. The Phenomenology of Contemporary Mainstream Manga. Image [&] Narrative 12 (1):63-82.

original class plan:
image processing: extracting features from images / visualization using ImagePlot

software:
ImageJ, ImagePlot

discuss:
"How to Compare One Million Images?" (Manovich, Douglass, Zepel, Huber)
"Style Space" (Manovich)
articles about manga's visual language

---

class 6:
exploring image collections using montage, slice, and average techniques;
readings discussion: analysis of manga visual language

ASSIGNMENT 4 DUE

software:
ImageJ:

class topics:

working with stacks in ImageJ (importing file sequence, virtual stacks)

visualizing image sequences: montage, slice, average;

sampling methods for media visualization.

WORKING WITH IMAGE SEQUENCES IN IMAGEJ

An “image sequence” is a set of images which are:
1) Located in a single directory.
2) Have sequential names, i.e., 001.jpg, 002.jpg, etc.
3) Have the same size.

ImageJ uses “stack” format to work with image sequences.
Documentation for ImageJ commands in Stacks menu

To import a sequence of images into ImageJ:

```
File > Import > Image Sequence..
```

Note 1: you can only import an image sequence if all images have the same size.

Note 2: depending on the size of an image sequence, you may have to use “virtual stack” option to import it.

After you imported an image sequence into ImageJ as a stack, you can use a number of commands to visualize it:

To create montage visualization:

```
Image > Stacks > Make Montage..
```

To create slice visualization:

```
Image > Stacks > Orthogonal Views..
```

To create image average visualization:

```
Image > Stacks > Z-project..
```

Visualizing an image sequence with ImageJ commands (images do not have the same size)

(While operations described below can be also carried out using various free software tools (Unix scripting, ffmpeg, ImageMagic, files rename utilities, Photoshop scripting, Applescript, etc.), here we show how to do them using ImageJ built-in commands and our ImageJ macros.)

(The following descriptions assume that you are using Vampire Kinight manga pages set, but the same techniques can be also used on any other image set.)

ImagePlot macro does not require the input images to be the same size. The images also do not have to be inside a single folder - they can be in multiple folders across your hard drive, and as long as you specify their filepaths in the data files, you can use all of them in a single visualization.

However, ImageJ build-in commands which operate on image sets are not as flexible. For instance, if you want to use ImageJ `Process > Batch > Measure...` command to measure features of images located in multiple folders (such as our Vampire Kinight pages), you first need to copy them all into a single folder. (Same applies to Naruto pages set.)

Additionally, since Vampire Kinight pages we downloaded do not have the same size, you can't directly import them in imageJ as a stack to make a montage and a slice visualizations - you have to first make them the same size. (The same applies to Naruto pages set.)

To copy files in multiple folders into a single new folder:

1) Edit `ImageSizeReport.txt` macro:

a) Change `save_dir_flag = 0;` to `save_dir_flag = 2;`
(this instructs the macro to copy images into a single new folder)

b) You may also need to change the number in this line:

```
new_name = substring(path, 84);
```

This line tells ImageJ to take the complete file path of each image, delete the first N characters (default N=84) and use remaining part as the new name.
Here is the example file path on my computer:

```
/Users/manovich/vampire_knight_chapters_example/Vampire_Knight_ch_1_5/V
ampire_Knight_ch_1/1.jpg
```

Lets's say I want the copy of this image to have this name:

```
Vampire_Knight_ch_1_1.jpg
```

To arrive at his name, the macro needs to cut first 69 characters, so I would change the line in the macro to this:

```
new_name = substring(path, 69);
```

Note that this line appears twice in the macro, so make the change in both places.

2) Run the macro:
It will first prompt you to select a folder which contains the relevant folders with files (for instance, all Vampire Kinight chapters), and then a folder where all files should be copied (with new names).

To make all images the same size using padding or cropping:

1) edit ImageSizeReport.txt macro:

a) change `save_dir_flag = 0;` to `save_dir_flag = 2;`
(this instructs the macro to copy images into a single new folder)

b) change `pad_images_flag = 0;` to `pad_images_flag = 1;`

c) change the numbers (width and height of new images) in this line:

```
run("Canvas Size...", "width=750 height=1076 position=Top-Left");
```

d) if you also want to control the exact naming of new files, or if you are getting an error, edit this line following the instuctions above:

```
new_name = substring(path, 84);
```

2) Run the macro:
It will first prompt you to select a folder which contains the relevant folders with files (for instance, all Vampire Kinight chapters), and then a folder where all files should be copied (with new names). While copying the files, the macro will also pad (or crop, depending on the values) all images using `width and height` settings.)

Note 1: if you want to create a media visualization (montage, slice, average) or an image sequence and you are using padding to add extra space to all images (or you are cropping them), this may effect the patterns revealed by a visualization - and also result in an inferior representation of the content. You may want to experiment with a few different approaches before your produce your final visualizations of the manga pages.

To measure basic features of gray scale images located in a single folder:

Currently `ImageMeasure.txt` macro can only handle color (RGB, 24 bit) images, but you can use ImageJ command to extract some features from all gray scale image inside a folder. Follow these steps:

1) Select the features you want to measure (such as mean and standard deviation) using `Analyze > Set Measurements...`

2) `Process > Batch > Measure..`
select the folder containing all images you want to measure
the measurements will be printed in Results window; you can then save the contents of using `File > Save as...`

To rescale all images in a folder:

```
Process > Batch > Convert
```

Using ImageMontage.txt macro

I wrote `ImageMontage.txt` macro which enchances ImageJ built-in command in two ways:

1) you can make montages of image sequences where images do not have the same size.

When you run the macro, it will prompt you for a text file which it will use to load images and place them into the montage. This file should contain images filenames (one filename per row.) For example:

```
0001.jpg
0002.jpg
0003.jpg
...
```

Before running the macro, change the following variables inside macro file:

```
filepath = 0;

montage_width = 5000; // set montage width

montage_height = 5000; // set montage height
```

```
montage_RGB = 0; // if 0, make 8-bit grey scale montage; if 1,
make RGB montage

height_row = 100; // set the height of each montage row
```

The macro will scale all images to `height_row` value and place them into the canvas, left to right, top to bottom.

Depending on how many images are given as input and the settings above, the resulting montage may have extra space left - or it may not contain all images - in this case just change the parameters.

2) The macro can also make a montage from images located in multiple directories. To use this option:

1) change this line: `filepath = 1;`

2) create a text file where each line contains a file path to an image you want to include in the montage.

3) run the macro and use this file when the macro prompts you.

For example, here is how the file for Vampire Knight pages on my computer looks:

/Users/manovich/manga/Vampire_Knight_ch_1_5/Vampire_Knight_ch_1/0001.jpg
/Users/manovich/manga/Vampire_Knight_ch_1_5/Vampire_Knight_ch_1/0002.jpg
/Users/manovich/manga/Vampire_Knight_ch_1_5/Vampire_Knight_ch_1/0003.jpg
…

You can create such file using Unix commands. For example, the file above was created using this command sequence:

```
ls */*/* > list.txt
```

Homework for class 8: due Nov 14 (Monday class) or Nov 15 (Tuesday class):

Practical assignment 5 (%12 of final grade):

Explore patterns in Naruto and Vampire Knight manga using visualization techniques learned in class.

IMPORTANT: do now work on the files inside Dropbox because if you change or accidently remove them, this will effect all other students! Before starting, please copy the relevant folders to another space on your computer.

Details:

You need to create **two required visualizations** (described below) and **the minimum of four additional visualizations** (but you welcome to do more). You also need to write a **mini-essay** what you discovered about Naruto and Vampire Knight using these visualizations.

Required visualizations:

1) Montage of all Vampire Knight pages.
The montage needs to show all 2447 pages in exactly the same order as their files names, i.e. from `Vampire_Knight_ch_1/1.jpg` to `Vampire_Knight_ch_75/40.jpg`.

2) Image plot of all Vampire Knight pages (2447 pages) organized by mean (X-axis) and standard deviation (Y-axis.)
Change the following options:
`Canvas width = 8000;`
`Canvas height = 8000.`
Leave all other options at default values (i.e. image size = 100 pixels, etc.)

Extra credit: If you also submit the same two visualizations which use all **Naruto** pages and all three visualizations are done correctly, you will get 6 extra points.

Additional visualizations:

For your own visualizations, you can use any visualization techniques. You can use all or parts of the data (i.e., all chapters or only some chapters.) You can use temporal or spatial sampling if this helps to reveal patterns. You can measure features and/or additional metadata to use in your visualizations (like in the previous assignment). You can also download selected chapters from other manga titles if you want to compare them to Naruto and Vampire Kinight manga.

What size to use for media visualizations (visualizations which show images, i.e image plots, montages and slices)?
Ideally you should make them as big as possible, but keep file size under 20 MB. (Note that unless your media visualization includes color, you should save as a grayscale image, since it will take 3 times less space than a color image.)

Mini-essay:
Along with submitting visualization files, also include a single mini-essay (.doc format) describing the interesting findings about Vampire Kinight and Naruto revealed by your visualization. Please discuss each submitted visualization in this essay (you dont need to refer to required visualizations - unless they are part of your arguments.)
Length: 400-800 words.

Below your essay, provide information about each visualization (data used, visualization technique, X-axis and Y-axis if a scatterplot or an image plot, etc.)

How to submit:
Bring your laptop with your project to class - I will explain how to submit the project.

Info about Naruto:
http://en.wikipedia.org/wiki/Naruto

“For Part II, the part of the [Naruto] manga beginning with volume 28, Kishimoto said that he attempted to not "overdo the typical manga style" by not including "too much deformation" and keeping the panel layouts to make it easy for the reader to follow the plot. Kishomoto said his drawing style changed from "the classic manga look to something a bit more realistic."[8]”

Info about Vampire Kinight:
http://en.wikipedia.org/wiki/Vampire_Knight

---

class 7:

final projects: introduction

originally planned assignment for class 8:

Read:

Cinemetrics articles by James Cutting and cinemetrics.lv group
Cinemetrics: Visualizing Movies (cinemetrics.fredericbrodbeck.de)

---

class 8:
work on final projects
readings discussion: quantitative studies of moving images;
preparing video files / extracting features from video / visualizing patterns in video data;

ASSIGNMENT 5 DUE

class topics:

review our custom imageJ macros for working with image sequences

introduction to UNIX:
basic commands;
working with directories;
generating a list of files in multiple directories

discussion of data sets for the final projects

Homework for class 9:

start working on final projects

Read:

CS articles about Twitter and MySpace

---

class 9:
work on final projects
readings discussion: big cultural data - current research

class topics:

how to select a subset of random rows from Excel spreadsheet:

add a new column to your spreadsheet;
fill it with randf ormula:

```
=rand()
```

sort spreadsheet by the column containing rand formula, and select first N rows

more UNIX commands for working with big data sets:

disk usage of directory in kilobytes:

```
du -sk subdir
```

count number of lines in a file:

```
wc -l filename
```

copy all files from a directory which match some condition
(in this examples: all jpeg images)

```
find . -name *.jpg -exec cp {} /destination/dir \;
```

creating a histogram list of items in a file:

```
sort -n input.txt | uniq -c
```

Homework for class 10:

contunue working on final projects

Read:

[Manovich, Trending article]
dana boyd article

---

class 10:
final projects progress review

presentation: big cultural and social data: current work and discussions