Lev Manovich

Visualizing Vertov

"To represent a dynamic study on a sheet of paper, we need graphic symbols of movement."

Dziga Vertov, "We: The variant of the manifesto" (1920), in *Kino-Eye: the writings of Dziga Vertov*, ed. Annette Michelson, trans. Kevin O'Brien (University of California Press, 1984), p.7.

High-resolution versions of all visualizations appearing in this text are available on Flickr for download:

http://www.flickr.com/photos/culturevis/sets/72157632441192048/with/8349174610/

This project presents visualization analysis of the films *The Eleventh Year* (1928) and *Man with a Movie Camera* (1929) by the famous Russian filmmaker Dziga Vertov. It uses experimental visualization techniques [1] that complement familiar bar charts and line graphs often found in quantitative studies of cultural artifacts. The digital copies of the films were provided by The Austrian Film Museum (Vienna).

Visualization gives us new ways to study and teach cinema, as well as other visual time-based media such as television, user-generated video, motion graphics, and computer games. The project is a part of a larger research program to develop techniques for the exploration of massive image and video collections that I have been directing at Software Studies Initiative (softwarestudies.com) since 2007 [2]. In this project, I explore how "media visualization" techniques we developed can help us see films in new ways, supplementing already well-developed methods and tools in film and media studies. Its other goal is to make a bridge between the two fields which at present are not connected: the field of digital humanities which is interested in new data visualization techniques, but does not study cinema, and quantitative film studies research which until now has used graphs in a more limited way.

The use of visualizations for the study and design of media can be traced to the work of many modern filmmakers, choreographers, architects, music composers and visual artists who created diagrams of their projects before or after they were realized. Vertov, in particular, created many diagrams to work out production, content structures and editing in his films. Sergei Eisenstein diagrammed a short sequence from *Alexander Nevsky* (1938*)* after the film was already made. Choreographer and theorist Rudolf Laban designed a diagram language for describing and analyzing dance that was later widely used by other choreographers. Austrian experimental filmmaker Peter Kubelka exhibited his 1960 *Arnulf Rainer* film consisting only of pure black and white frames as a wall installation, thus turning a film into its own visualization. [3] Numerous artists worked out compositions, color palettes and other aspects of their paintings in diagrams before executing the actual paintings.
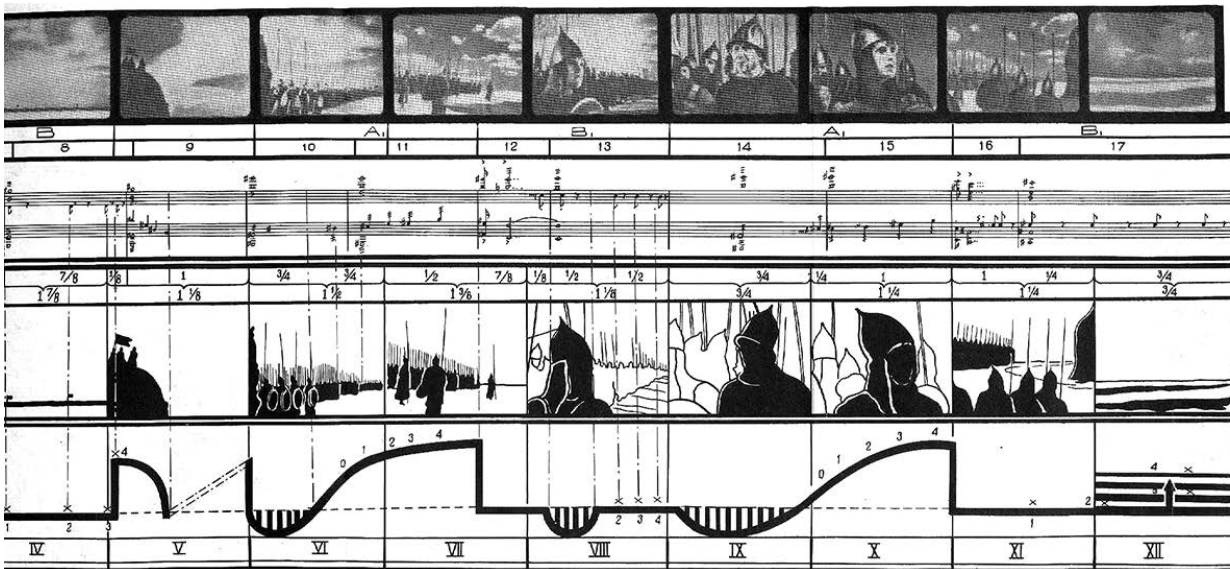

0.1. Sergei Eisenstein. A diagram of montage in a sequence from his film *Alexander Nevsky* (1938).
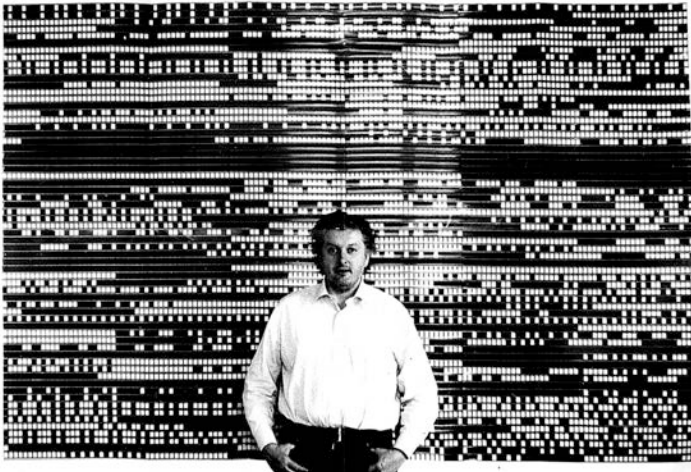
0.2. Peter Kubelka with a wall Installation of *Arnulf Rainer* (1960).

20th century commercial cinema and television relied on storyboards – the hand drawn diagrams that guided the production. More recently, many feature films started to be completely worked out in preproduction using animated 3D models (this method is called *previs*).

Other important precursors to the use of visualization for media analysis can be found in the 19th century. The development of recording technologies of photography and film motivated the work on the techniques for *automatic diagramming* of the visible world, and in particular movement. (This is particularly relevant for visualizing Vertov's films since, as I will discuss later, at the beginning of his career he identified aesthetics of cinema with movement of objects and people in space.) In the 1880s, Étienne-Jules Marey figured out how to use photography to diagram movement of humans and animals in two dimensions. (Previously, starting in the 1860s, he developed many instruments to graphically record heartbeats, respiration, movements of muscles, and other physiological functions.) In 1910-1920s Frank and Lillian Gilbreth introduced new methods for recording and studying workers movements using photography and film.
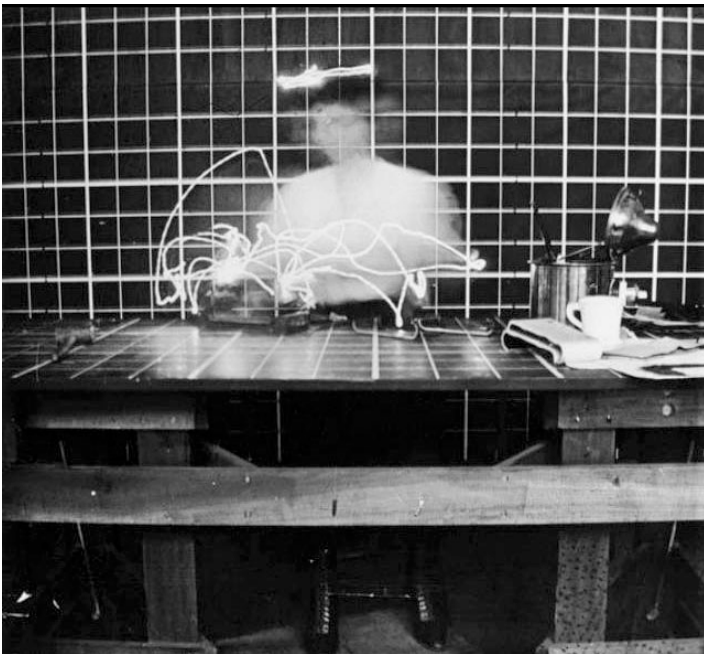
0.3. Example of the Gilbreths's motion studies using their Chronocyclegraph device.

In the 1990s, the productions of feature films and video games adopted *motion capture* to record moving bodies as 3D diagrams that are then used to animate computer generated characters. Motion capture is also commonly used to record

face movements of actors and drive animated faces (for instance, in Cameron's *Avatar*).
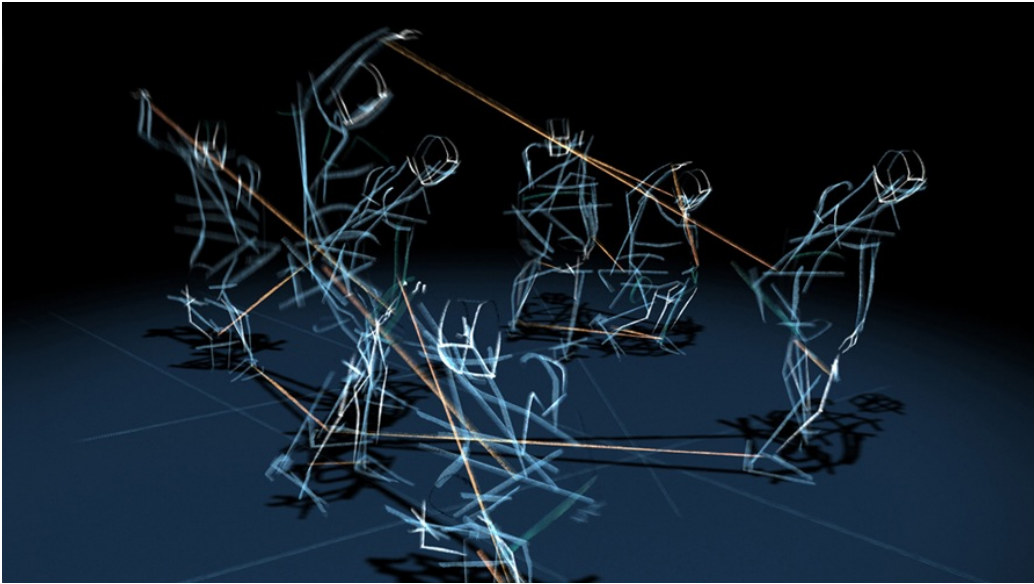
0.4. *Ghostcatching,* a digital art installation by Paul Kaiser and Shelley Eshkar uses motion capture of the dance performances by Bill T. Jones.

Computational analysis of the films combined with visualization may allow us to "reverse engineer" some of the aspects of cinema and other types of time-based media, revealing interesting patterns at any scale - from a single shot to billions of YouTube videos. Like all computational methods, these have their strengths and weaknesses. A computer does not have the same insights into the meaning and structure of a film as its director and editor. However, it can help us notice subtle patterns in editing, composition, movement, and other aspects of cinematography and narrative that maybe hard to see otherwise. Computers can also allow us to compare any number of films, helping us to understand what is typical and what is unique in the given dataset, and identify common characteristics and similar patterns. In short, while they can't easily produce the analytical diagrams such as the one created by Sergei Eisenstein (illustration above), they can do other things that will be very hard or very time consuming for a human observer.

One of Vertov's key concepts was Kino-Eye (*Kino-Glaz* in Russian), which received its best realization in *A Man with a Movie Camera* created in 1929 (many other equally radical film plans of Vertov remained unrealized). In a 1924 article titled, "The Birth of Kino-Eye," he writes:

> Kino-Eye is understood as "that which the eye does not see,"
> as the microscope and telescope of time,
> as the negative of time,
> as the possibility of seeing without limits and distances,
> as the remote control of movie cameras,
> as tele-eye,
> as X-ray,
> as "life caught unawares," etc., etc.
>
> Kino-Eye as the possibility of making the invisible visible. [4]

Today, data visualization designers often use the same phrase "making the invisible visible" to describe how visualization can reveal patterns in the data. For Vertov, this goal called for new cinematographic, editing, and logistic techniques. The visualizations presented in this project aim to reverse *kino-eye*, pointing it at Vertov's films.

The large body of research in empirical film studies already explored in detail one quantitative dimension of films – the length of the shots. [5] This is an easy dimension to quantify: one only has to count the number of frames between the

shot boundaries. The analysis of shot length patterns also makes good sense because Vertov and many other 20th century filmmakers created tables and diagrams to plan the exact shot lengths in their film sequences. [6] Barry Salt, Yuri Tsivian, and other scholars showed that working with shot lengths data can lead to important new insights in the case of single films, the works of a single director, and whole periods of film history.

One of the goals of my project is to show how other dimensions of films can also be explored using particular visualization techniques. In some cases, we use digital image processing software to measure visual properties of every film frame such as average gray scale value, contrast, number of shapes, number of edges, the relative proportions of different colors, texture properties, and so on. (We started this research in 2008; our approach is similar to that of psychologist James Cutting and his colleagues [5], but we explore a larger set of image features using techniques from computer vision.) In other cases, we don't measure or count anything. Instead, we arrange the sampled frames from a film in a single high-resolution visualizations in particular layouts. This use of *visualization without measurements, counting, or adding annotations* is the crucial aspect of my lab's approach for working with media data sets, and I hope that it can add to other approaches already used in quantitative film studies and digital humanities.

The portfolio of visualizations with commentaries below starts with three visualizations that plot the already familiar quantitative measure of films, i.e., shot lengths (1-3). It then moves to illustrate other possible strategies (4-9) using "media visualization" approach. To create media visualizations, sampled frames from a film are arranged by the software in different layouts using existing metadata (i.e., frame numbers, shot boundaries, manual annotations of shot content, etc.) and/or automatically created measurements of visual properties of images, shots and whole films. I don't include any numerical measurements of the films in my main analysis, in order to emphasize that visualization techniques can help to us explore films and identify patterns beyond mere statistics [7]. (Some of the measurements are referenced in "Details" sections).

The visualizations are organized in a sequence which starts from a "bird's eye" view of the cultural artifacts (hundreds of 20th century films) and gradually moves closer and closer – similar to how Google Earth allows you to start with the Earth view and then zoom in and eventually enter a street view. (Think also of *Powers of Ten* by Ray and Charles Eames.) We start with a landscape of thousand 20th century films where Vertov's films appear only as points (1). We then zoom into these films, with each shot represented as a bar in a bar graph (2-3). At the next "zoom level," we start seeing images (one frame representing every shot – visualizations 4 - 7). Finally we zoom into a single shot to see more details (8-9). For me, an ideal interactive application for exploring any area of visual culture would be organized in a similar way, with the visualizations showing different level of details and different aspects of the artifacts generated in real time as the user zooms in and out of the whole dataset. [8]

The presentation is an experiment. Normally an academic article consists from text with a small number of illustrations. Instead, this presentation of is a portfolio of a large number of visualizations, with text serving as the commentary. The presentation also does not advance a single argument or a concept. Instead, I progressively "zoom" into cinema, exploring alternative ways to visualize media at different zoom levels, and noting interesting observations and discoveries. We can compare its genre to that of travel writing, where the organizing principle is the writer's movements through space.

With the exception of visualizations 1.1 – 1.3, which were done in Excel, I wrote the code to create all other visualizations. The code was implemented as macros that run inside ImageJ, a popular open source image processing platform used in live sciences, astronomy, geography, and other science fields [10]. In my lab we also developed a few fully documented macros with graphical user interfaces; they are available for download along with the documentation I wrote on how to use them to visualize video and image collections [11].
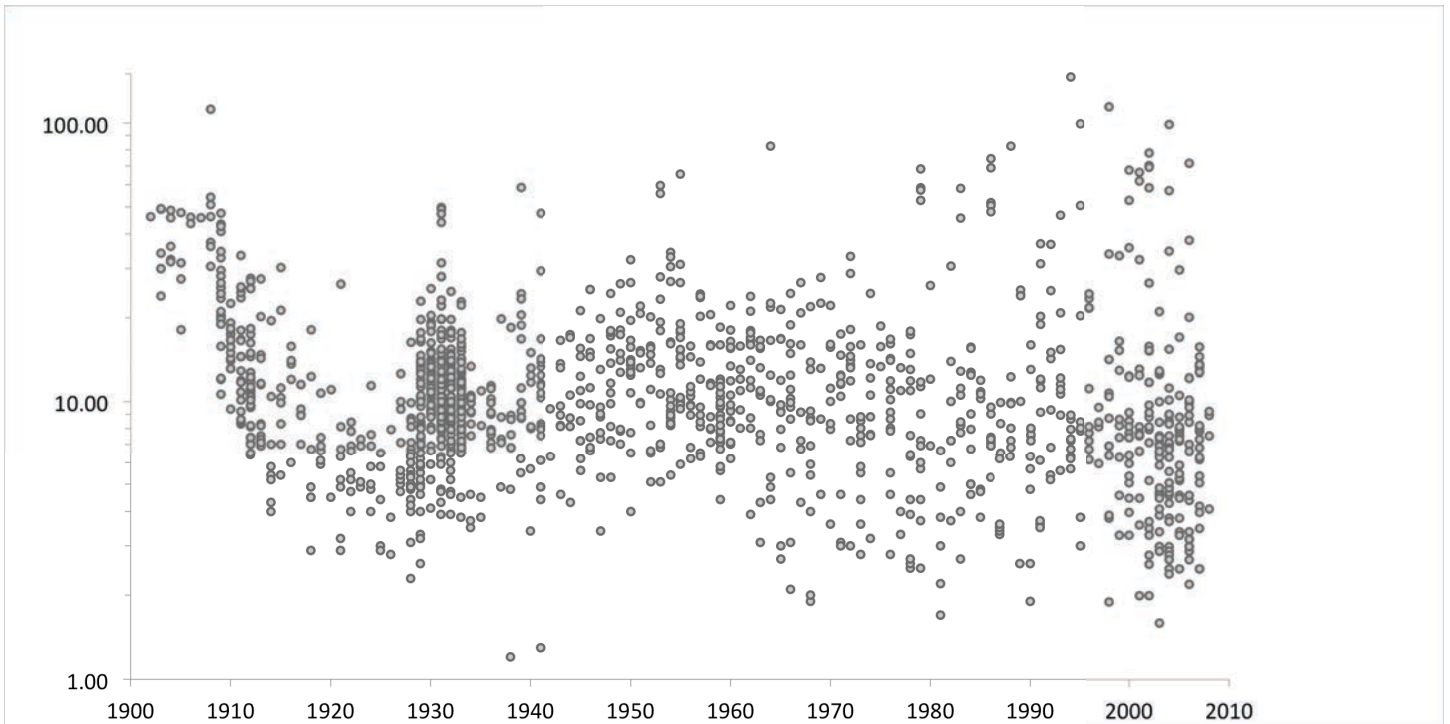
I am grateful to film researcher and Austrian Film Museum staff member Adelheid Heftberger for initiating and making possible this project in 2009, and providing detailed feedback on the work as it developed. Some of the visualizations use her manually created lists of shots in Vertov's films. While we did use shot detection software in other projects, in this case it made more sense to rely on manual annotations of the shots. This method is more accurate in recording many very short shots, which are characteristic of Vertov's films.

Some of the visualizations appeared previously as supplementary material on the DVD of two Vertov films published by Austrian Film Institute [9]; others appear here for the first time.

Notes:

1. For the detailed analysis of our visualization techniques used in this project, see Lev Manovich, Media Visualization: Visual Techniques for Exploring Large Media Collections, *Media Studies Futures*, ed. Kelly Gates (Blackwell, 2012), http://softwarestudies.com/cultural_analytics/Manovich.Media_Visualization.web.2012.v2.doc.
These techniques draw on the previous projects by digital artists and designers discussed in Lev Manovich, "What is Visualization?" *Visual Studies* (Routledge, 2011), http://lab.softwarestudies.com/2010/10/new-article-is-visualization.html.

2. See http://lab.softwarestudies.com/p/research_14.html.

3. See http://lab.softwarestudies.com/2012/11/peter-kubelkas-arnulf-rainer-film-as.html.

4. Dziga Vertov, "The Birth of Kino-Eye" (1924), in *Kino-Eye: the writings of Dziga Vertov*, ed. Annette Michelson, trans. Kevin O'Brien (University of California Press, 1984), p. 41.

5. This research started with Barry Salt, 'The Statistical Style Analysis of Motion Pictures' *Film Quarterly*, 28, 1 (1974), pp. 13-22. For current work, see the articles at http://www.cinemetrics.lv/articles.php; Yuri Tsivian, "What is cinema? An Agnostic Answer," *Critical Inquiry* 34 (Summer 2008), pp. 754-776.

Salt proposed that film analysis should consider a number of other characteristics such as shot scale, camera movement and angle of shot. However, because he and subsequent film scholars have been entering this data by hand, this limited how many films could be annotated.

More recently, James E. Cutting at Cornell University started a new series of quantitative studies that use automatic methods to measure other properties of films besides shot statistics. See James Cutting, Jordan DeLong, and Christine Nothelfer, "Attention and Hollywood Films," *Psychological Science* vol. 21, no. 3 (March 2010), pp. 432-439; Cutting, Brunick, DeLong, Iricinschi, and Candan, **"**Quicker, faster, darker: Changes in Hollywood film over 75 years", *i-Perception* (2011) volume 2, pp. 569 – 576, http://people.psych.cornell.edu/~jec7/pubs/iperception.pdf.

A pioneering project by designer Frederic Brodbeck (2011) shows how various dimensions of films can be creatively visualized: http://cinemetrics.fredericbrodbeck.de/.

6. Examples of these diagrams and tables are referenced by Yuri Tsivian in his introduction to cinemetrics site, http://www.cinemetrics.lv/.

7. The work of my lab emphasizes what we call "exploratory visualization" for working with large visual media data sets. We adopted this term from the "exploratory data analysis" approach in statistics. For the discussion of exploratory data analysis as the method in film studies, see Nick Redfern, "Exploratory data analysis and film form: The editing structure of slasher films," http://nickredfern.files.wordpress.com/2012/05/nick-redfern-the-editing-structure-of-slasher-films.pdf, 2012.

8. We prototyped such an interface in 2008: http://lab.softwarestudies.com/2008/12/cultural-analytics-hiperspace-and.html.

9. Dziga Vertov, *A Sixth Part of the World / The Eleventh Year*. DVD. Edition Filmmuseum, 2009.

10. http://rsbweb.nih.gov/ij/.

11. http://lab.softwarestudies.com/p/software-for-digital-humanities.html.

1| panorama

1.1. A scatter plot of mean shot lengths of over 1,000 films created between 1902 and 2008. Each film is represented as a small circle. X-axis – year. Y-axis – mean shot length in seconds. (Y-axis uses logarithmic scale.) Data preparation and exploration: William Huber (Software Studies Initiative), 2008. Data source: cinemetrics.lv, 2008. For some films, cinemetrics.lv database contained separate measurements for each film reel or film parts; in these cases, each reel or part are shown separately.

I will start with a panoramic view of 20th and early 21st century cinema, and then will gradually "zoom" into Vertov. In the graph above, each point represents one film. X-axis is time, and Y-axis is average shot length (ASL). The data comes from cinemetrics.lv, an important web site used by film scholars to collect information about films' shot lengths and discuss the analytical work that uses this data. The graph shows all films that were in a cinemetrics.lv database in 2008 (today it has many more entries).

Let's see what this graph can tell us about Vertov. First, I will plot separately 28 Russian films (they were in n cinemetrics.lv database when we got the data in 2008).
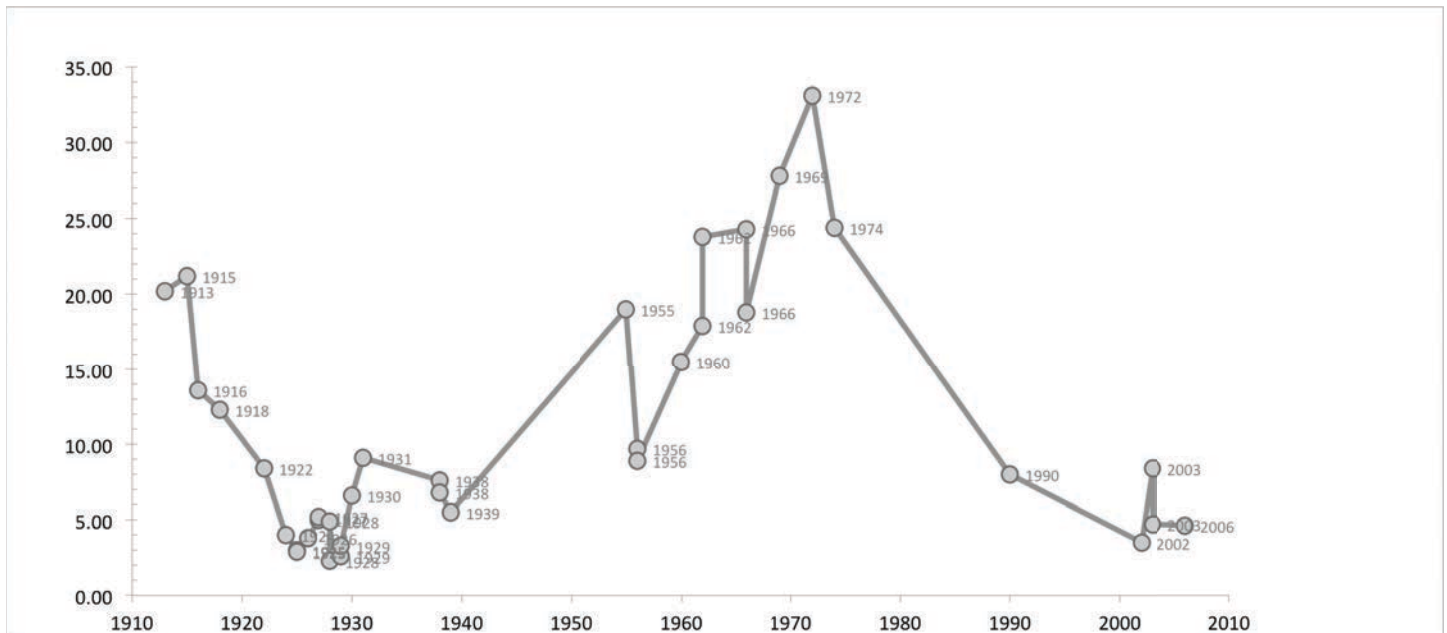
1.2. 28 Russian films contained in cinemetrics.lv database in 2008. X-axis – year. Y-axis – mean shot length.

The dramatic curve reflects the turbulent turns of the Russian history, the changes in cultural policies of the State, the film language, and what films entered film history. The super-fast editing of the montage films of the 1920s gives way to slow classical narration of the 1950s-1970s. (In this context, even slower films of Tarkovsky such as *Solaris* 1972 appear not the exception but the continuation of this trend). After the beginning of Perestroika (1986-), Western cultural influence leads to very fast paced films, as seen in the last part of the graph.

The lowest points on the graph belong to the famous films of Russian "montage" filmmakers created in the 1920s: Dziga Vertov, Sergei Eisenstein, Aleksandr Dovzhenko, Yakov Protazanov. If cinemetrics.lv instead included the films popular with Russian audiences at the time – comedies, melodramas and adventure films produced both in the West and in Russia – this set of points would disappear, and the curve will look less dramatic.

At the other extreme, some of the slowest films in the cinemetrics.lv database are also by a Russian director: Andrei Tarkovsky. In contrast to the 1920s films by Vertov and other Russian directors which favored sequences of very shot shots in order to communicate particular meanings, Tarkovsky's mature films consists from shots that may lasts a few minutes – a way to give the control to the spectators who are now free to navigate the space in the frame. (The longest average shot measurement not shown in the graph also belongs to a Russian film - *Russian Arc* by Sokurov, 2002. This feature length film was the first in the history of world cinema to be shot as a single take.)

Now let us look at the classical "montage" films of Vertov and Eisenstein against all other films from the first third of the 20th century. The films by Vertov and Eisenstein shown in the graph below as rectangles are *Kino-Eye* (1924), *Strike* (1925), *Battleship Potemkin* (1925), *October* (1928), *The Eleventh Year* (1928) and *Man with a Movie Camera* (1929).
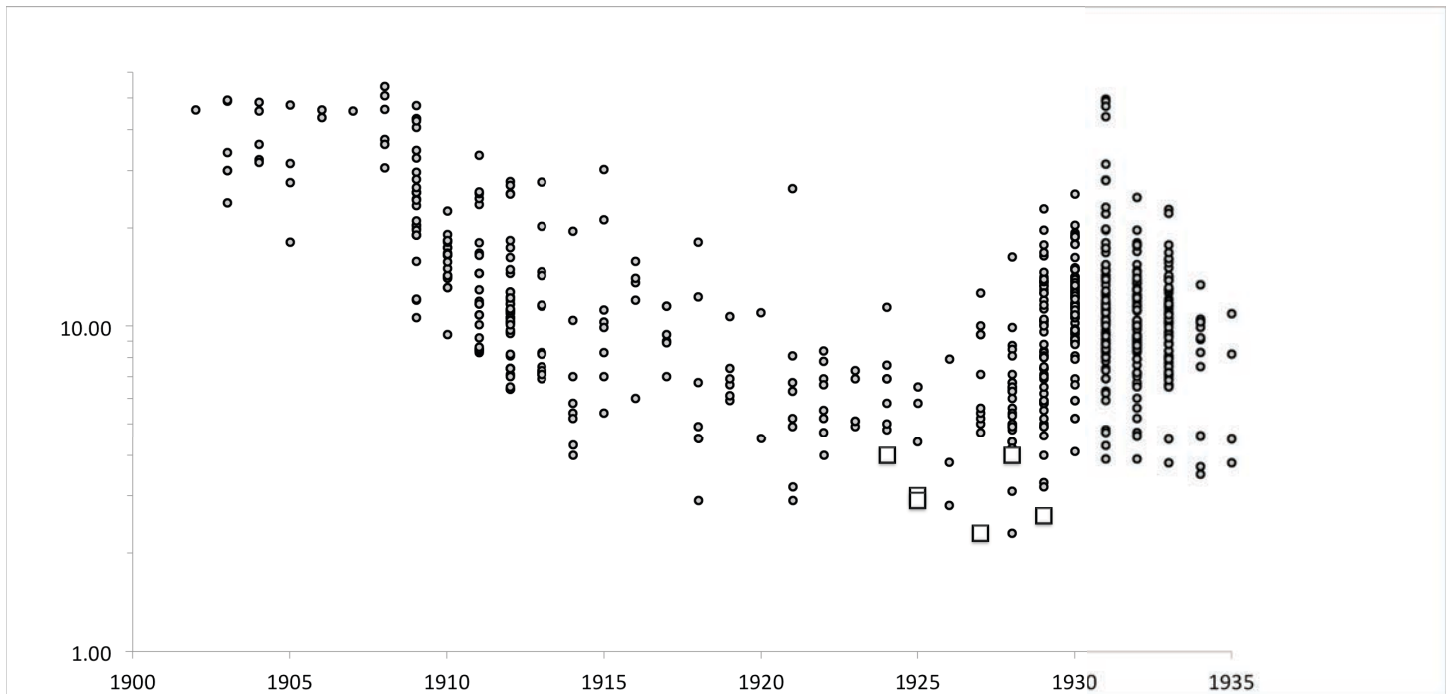
1.3. The films by Vertov and Eisenstein from 124-1930 (rectangles) against all films in cinemetrics database from 1900-1925. X-axis – year. Y-axis – mean shot length.

As we may expect, these films correspond to some of the lowest points in the graphs (i.e., they have the lowest ASL; see "Details" below for the numbers). In particular, *Man with a Movie Camera* has the smallest *median shot length* among all films made in the first part of the 20th century among all films in cinemetrics.lv. It is also the "fastest" among all other famous films directed by other Russian filmmakers of the montage school.

But something else in the graph is unexpected. "Russian montage" directors strongly objected to the cinema as it was practiced by others, claiming that it was merely filmed theatre and literature. Instead, they saw the juxtaposition of shots as the organizing principle of their cinema. To realize this idea in practice, they often relied on very short shots that were stills, almost stills, or featured a single simple action. Each such shot was to act as a basic atomic and further indivisible element; the sequence with the predetermined meaning and emotional effect was to be constructed from these elements, like a wall from bricks. If we view these most representative of montage theory sequences in films by Eisenstein and Vertov and compare them to the popular narrative films of the 1920s, the differences are immediately clear.

The graph tells a different story. The short ASL of most classical Russian montage films continues the trend from long to short ASL that starts at the end of 1900s. In the 1910s, cinema gradually shifted from being a simulation of theatre to a new language based on cutting between changing points of view. As a result, the shot lengths decreased substantially. This development can be clearly seen in the graph. Within this larger context, the Russian montage films are no longer form an opposition. Instead, they can be seen as further continuing the overall trend that started much earlier.

If we were to only consider basic shot length statistics instead of visualizing all data, the result would be different. The average of separate ASL measurements in six films by Vertov and Eisenstein is 3.13 seconds; this is almost three times less than 7.91 seconds average for all other films from 1921-1930 in cinemetrics database.

How do we understand our finding that Russian montage films fit in with other films as seen on ASL graph? It does not mean that montage film languages of Russian filmmakers were not different from other film languages of the 1920s –they were (or at least, the parts of the films structured according to montage principles). However, we may need to measure and visualize other information about the shots (rather than only shot lengths) to fully reveal this difference.

(Thanks to Yuri Tsivian for providing access to the Cinemetrics database.)

Details:

The films by Vertov and Eisenstein shown in graph 1:

Year	Mean shot length (ASL)	English Title	Director
1924	4.0	Kino-Eye	Dziga Vertov
1925	3.0	Strike	Sergei M. Eisenstein
1925	2.9	Battleship Potemkin	Sergei M. Eisenstein
1927	2.3	October	Grigori Aleksandrov, Sergei M. Eisenstein
1929	2.6	Man with a Movie Camera	Dziga Vertov
1928	4.0	The Eleventh Year	Dziga Vertov

ASL shows the mean shot length in seconds. Instead of the mean, we can alternatively use the *medium* shot length. Median is a representation of average tendency in a data set that is less sensitive to outliers (in the case of the two Vertov films, a few unusually long shots are such outliers.)

If we use median instead of mean, Vertov's film is "faster" than all other famous films of his fellow Russian montage directors:

Year	Median shot length	English Title	Director
1925	2.3	Strike	Sergei M. Eisenstein
1926	2.0	Mother	Pudovkin
1927	2.0	October	Grigori Aleksandrov, Sergei M. Eisenstein
1929	1.6	Man with a Movie Camera	Dziga Vertov
1928	2.8	The Eleventh Year	Dziga Vertov

The measurements for Vertov's films use manual breakdown of shots by Adelheid Heftberger who recorded the information using Anvil software and used the archival prints of the films at Austrian Film Museum.

The measurements for other films are from cinemetrics.lv. Due the measurement method used by people who submit the data to this site, they may not be completely accurate.

2| comparing shot lengths in two films
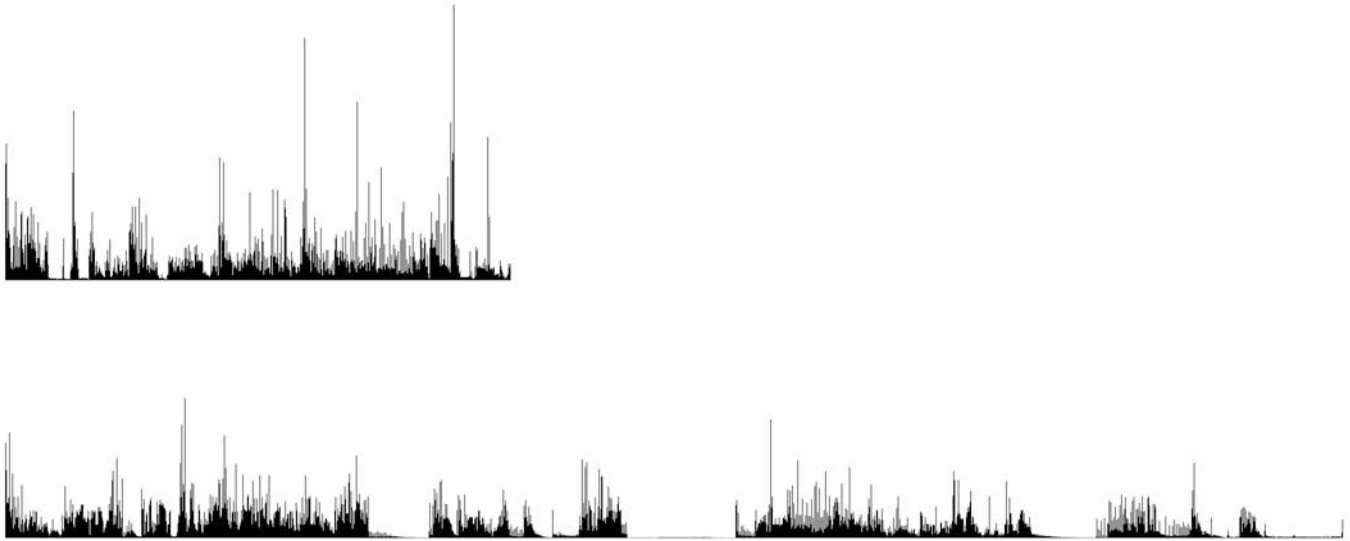
2.1. Shot lengths in *The Eleventh Year* (first image) and *A Man with a Movie Camera* (second image). Each bar represents one shot in a film; the height of the bar represents the length of a shot. The difference in graph lengths reflects the differences in numbers of shots between the two films (*A Man with a Movie Camera* contains almost three times as many shots as *The Eleventh Year*)*.*

Next, I will zoom further into the "data landscape" of hundreds of films to examine more closely two films by Vertov. Plotting the lengths of all shots reveals a number of interesting differences between them, beyond the most obvious observation that *The Eleventh Year* has a larger number of longer shots than *A Man with a Movie Camera.* The latter more "theoretical" film (i.e., it is a more systematic "visualization" of Vertov's theories) is also more tightly structured. Vertov establishes particular temporal patters of shot lengths. Each of these patterns is repeated a number of times throughout the film. The following graph shows examples of three such patterns (identified by numbers) in one part of the film.
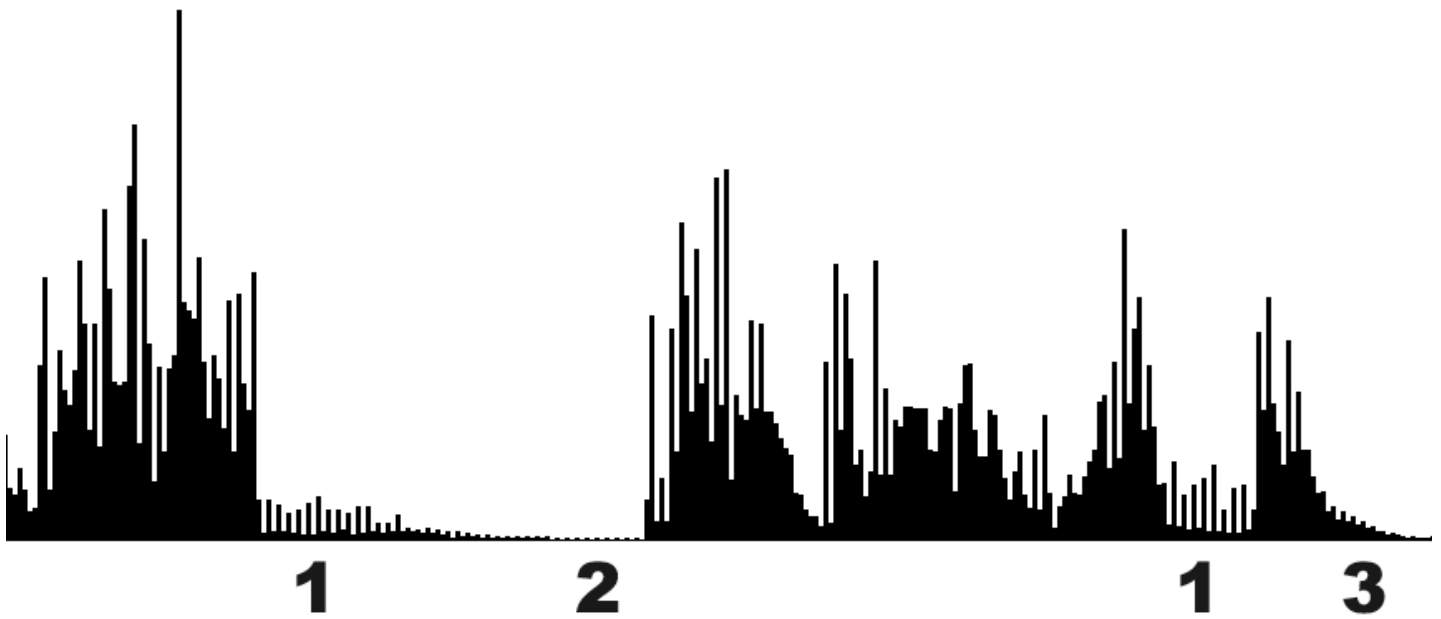
2.2. A close-up of *A Man with a Movie Camera* visualization showing examples of three shot lengths patterns:

1) A long sequence of extremely short shots of the same length.
2) Long and short shots alternating back and forth.
3) Longer and shorter shots alternating, with their lengths gradually decreasing at the same speed.

Perhaps the most interesting discovery is that such systematic patterning is responsible for a relatively small part of *A Man with a Movie Camera*, with the shot lengths in the rest of the film varying without any apparent system. In *The Eleventh Year*, the parts with systematically varying short lengths being even smaller. (Time series analysis, a statistical technique for analyzing temporal data, confirms the observation that *The Eleventh Year* overall has less structure in shot lengths than *A Man with a Movie Camera.)*

This discovery does make sense. Watching Vertov films from the 1920s and early 1930s, one gets a feeling that the filmmaker always uses some clear principle to connect number of shots in a sequence (as we can expect for films of Russian montage school). However, these principles vary throughout the film, and they can also be combined in a single sequence (Eisenstein's montage theories identify a number of these principles). Using a number of consecutive shots with the same length, or varying these lengths according to a system is only one such organizational principle among others. When we visualize a film such as *The Eleventh Year* along only a single dimension where these principles operate (such as shot lengths), all other parts of the films organized according to other principles may appear random, since the visualization can't show patterns on these other dimensions. (Visualization 6 below shows an example of one such dimension: the amount of visual change in every shot).

3| alternative visualizations of shots length
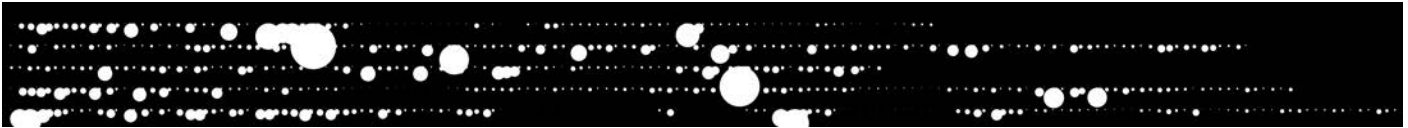
3.1. An alternative visualization of shot lengths in *The Eleventh Year*. Each shot is shown as a circle, with the size representing the length of the shot. Each film reel occupies one row.
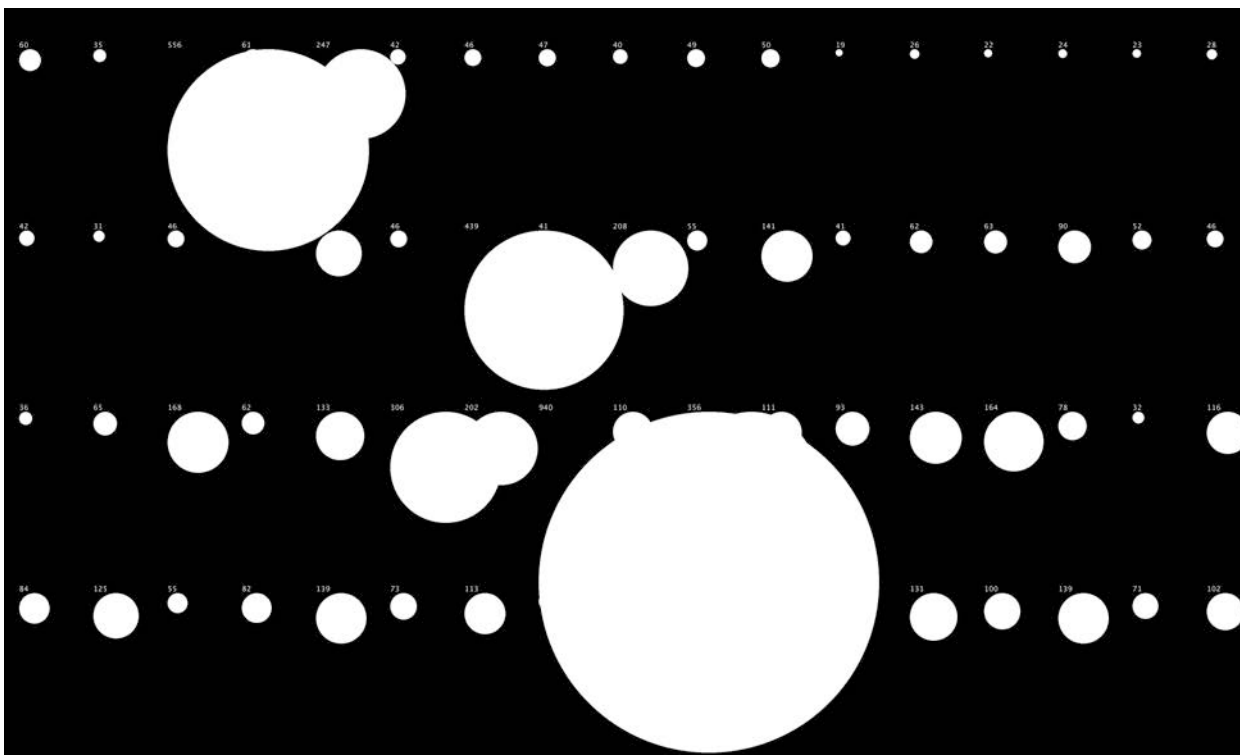
3.2 A close-up of the visualization. A number indicating the length of each shot (in frames) appears on the left above a circle representing this shot.

Besides a bar graph used in the previous visualization, we can also experiment with many other graphical techniques to show how data changes over time. For example, the visualization on this page uses circles to represent the shot length in *The Eleventh Year*. It dramatizes one particular characteristic that distinguishes this film from *A Man with a Movie Camera* – the presence of a number of very long shots (large circles in the visualizations). These long shots contradict our normal assumption that the films of Russian montage school consist from only short shots – created by the repeated use of particular famous sequences (such as Odessa steps in Eisestein's *Strike)* to represent this film movement in film textbooks and other educational materials.

This visualization compares patterns in shot lengths over time in *The Eleventh Year* (top) and *Man with a Movie Camera* (bottom) using rectangles instead of circles:

3.3. Visualization of shot lengths in *The Eleventh Year* (top) and *Man with a Movie Camera (bottom).* Each shot is shown as a rectangle, with the size representing the length of the shot.

4| 654 shots in a single visualization

4.1. Each of 654 shots in *The Eleventh Year* (Dziga Vertov, 1928) is represented by its second frame. The frames are organized left to right and top to bottom, following the order of shots in the film.

This visualization turns the film into a single image. It uses a semantically and visually important segmentation of the film – the shots sequence. Each shot is represented by its second frame. (I used a second as opposed to a first frame because the latter often are very dark, due the limitations of film printing technology available to Vertov and his wife and film editor Svilova.) We can think of this visualization as a "reverse engineered" imaginary storyboard of the film – a reconstructed plan of its cinematography, editing, and narrative.

4.2. A close-up of the visualization.

Such a close-up functions as a new "montage" of Vertov's original montage in the film. Because of the close-up proportion, it does not show a single sequence of shots. Instead, we see one sequence of 6 shots (top row), then jump to another sequence of 6 shots (second row), and so on, nine times. Each sequence is separate from the next by the same number of shots (not visible in a visualization).

I have originally included this close-up simply to help the readers see better the images in the full visualization (previous page). Later I realized that its montage-like structure also helps to better understand various patterns in the film editing. For example, we notice the repetitions of the shots of different people looking up (presumably towards the Communist future.) While this pattern already stands out in the full visualization, the close-up shows additional patterns. The shots of people looking up are combined with other shots in a number of different ways. The two methods revealed in the close-up are parallel montage (rows 1, 4, 5, 9), and a sequence consisting only from these shots (row 8). The repetition may show shots of the same person (rows 1, 4, 9) or different people in the same position (rows 5, 8). Of course, a careful study of

the film using standard digital video tools such as QuickTime or Premiere will also reveal these patterns – however it will take some time, while a visualization shows them instantly, and it also allows us to compare them.

It is also easy to notice that Vertov repeats a few shots of the people looking up in the same direction (i.e., to the right, to the left, or directly towards the center) before switching the direction.

To better see this pattern of repetition and reversal across the whole film, I selected all shots in a film that show close-ups of faces, and visualized in the sequence of their appearance in the film (once again, each shot is represented by its second frames).

4.3. 72 shots showing close-ups of faces in *The Eleventh Year*, arranged in order of their appearance in film (left to right, top to bottom). The gray levels of some frames were adjusted to reveal more detail.

Usually, a set of shots of one or a few people looking at the same direction belongs to a single sequence. These shots establish a structure of repetitions juxtaposed with a set of different shots montaged in between the shots of faces. However, visualization also shows a meta-pattern: across the whole film, a sequence of shots with faces looking in one direction is followed later by a sequence of shots with the opposite direction; next sequence reverses the direction again, and so on.

We can also combine the representation of shots by their selected frames with a bar graph representing some property of these shots. In the following visualization of *The Eleventh Year*, every shot of film is represented by its second frame; the length of a bar below the frame represents the shot duration.

Each row corresponds to one reel of the film (bottom to top). During film projection in the 1920s, there was a break between each reel. Vertov took this into account in editing the film - this is why it's more meaningful to graph each reel separately as opposed to putting them all in a single graph.

4.4. Visualization of *The Eleventh Year*.
Each row corresponds to one reel of the film (bottom to top).
Every shot of film is represented by its second frame.
Shot numbers appear below each frame (right corner).
The height of a bar below each frame corresponds to shot length (1 frame = 1 pixel).
The number below a bar shows shot length in frames.

The following two close-ups of the visualization 4.4 illustrate different editing patterns in the film. One (top) is a sequence of shots all with the same length (top) – starting with shot 590. Another (bottom) is the sequence alternating between the titles shots with the same length (13, 15, 17, 19 frames) and the shots with decreasing length (14, 16, 18). (Each close-up shows only a part of the longer sequence).

4.4. Two close-ups from the beginning (bottom) and end of the film (top).

5| comparing start and end of very shot

5.1. Visualization of *The Eleventh Year* comparing the frames from the beginning and the end of every shot.
Each column represents one shot in the film using its second from the beginning frame (top row) and second from the end frame (bottom row). The shots are organized left to right following their order in the film.
First image: The complete visualization of the whole film.
Second image: A close-ups of the complete visualization.

See full resolution visualization (60,000 pixels wide) on Flickr:
http://www.flickr.com/photos/culturevis/5674891152/in/set-72157623326872241

5.2: Further close-up showing beginning and end of seven consecutive shots.

5.3. Another close-ups showing beginning and end of another sequence of seven consecutive shots.

To create this visualization, I wrote a program that selected the frames from the beginning and end of every shot in *The Eleventh Year* and placed them together in the order of shots. For each shot, a frame from its beginning is on the top, and the frame from the end is bellow.

"Vertov" is a neologism invented by the film director who adapted it as his last name early in his career. It comes from the Russian verb *vertet*, which means, "to rotate." "Vertov" may refer to the basic manual motion required in filming in the 1920s – rotating the handle of a camera. It may also refer to the dynamism of film language developed by Vertov who, along with a number of other Russian and European filmmakers, designers and photographs working in that decade, wanted to "defamiliarize" reality by using dynamic diagonal compositions and shooting from unusual points of view. However, this visualization suggests a very different picture of Vertov. Almost every shot of *The Eleventh Year* starts and ends with practically the same composition and subject. In other words, the shots are largely static.

Going back to the actual film and studying these shots further, we find that some of them are indeed completely static – such as the close-ups of faces looking in various directions without moving. Other shots employ a static camera, which frames some movement – such as working machines, or workers at work – confined to the same area of space. Here we may recall that a number of shots in Vertov's most famous film *Man with A Movie Camera* (1929) were specifically designed as opposites: shooting from a moving car meant that the subjects were constantly crossing the camera view. But even in this most experimental of Vertov's film, such shots constitute a very small part of a film.

6| average amount of visual change in each shot
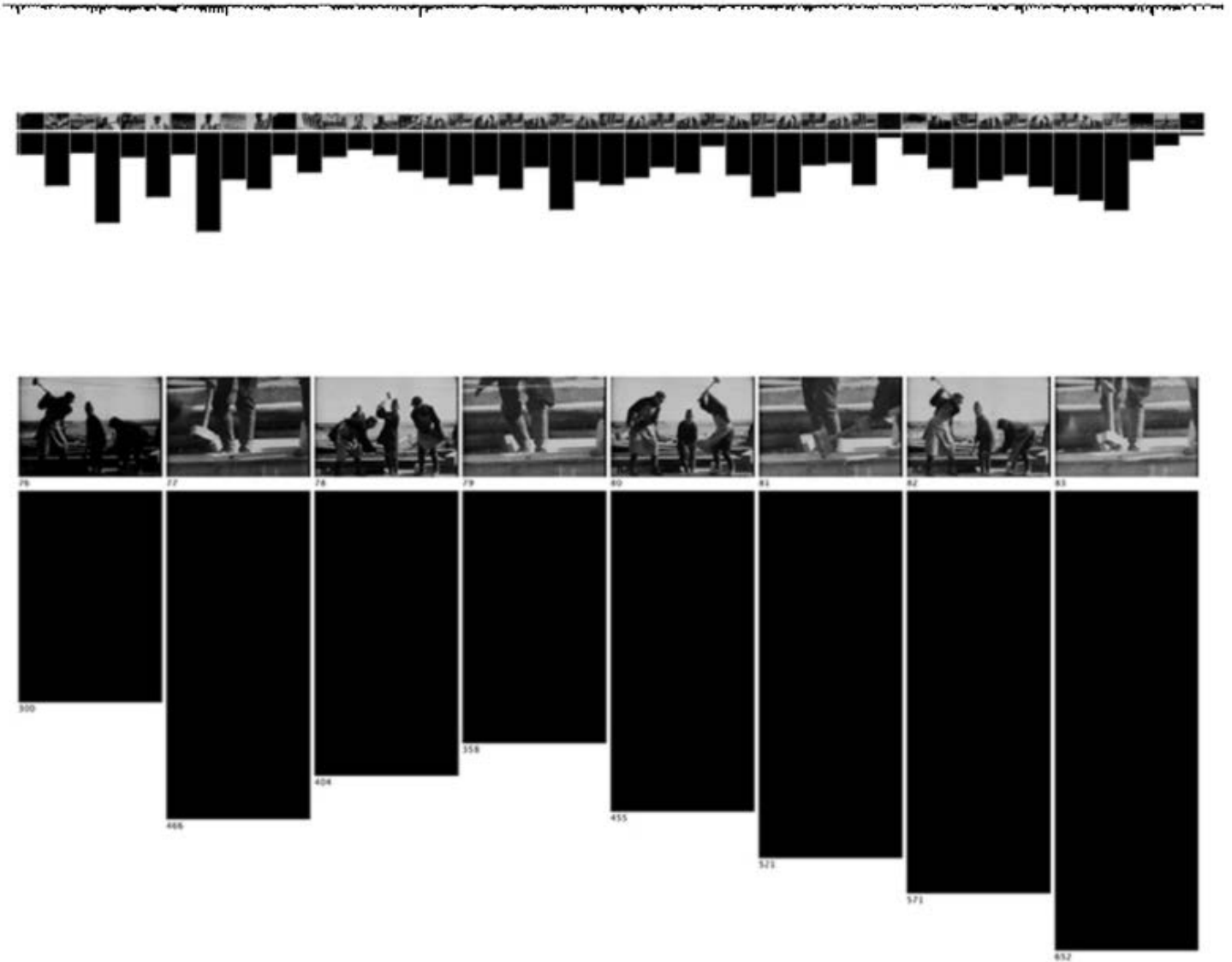
6.1. Average amount of visual change in every shot in *The Eleventh Year*.
Each bar represents one shot. The length of a bar corresponds to the average amount of visual change in the shot (details on how this was calculated can be found in "method" below). The second frame of a shot is placed above the bar.

First image: the complete visualization of the whole film.
Second image: a close-up of the visualization.
Third image: a tighter close-up showing the pattern of gradual changes in shots' average amount of visual change.

See Full resolution visualization (60,000 pixels wide) on Flickr:
http://www.flickr.com/photos/culturevis/4117658480/in/set-72157632441192048

This visualization uses a very simple algorithm (described below) to calculate the average amount of visual change in each shot in Vertov's film. Each column corresponds to one shot. The second (from the beginning) frame of a shot is above; the bar representing the measurement of the average amount of visual change in this shot is below. For a shot where little changes, the bar will be short; a shot with lots of changes (be they movements of a camera, or the subjects, or both) has a long bar. A static shot of two people talking is the example of the former; a shot filmed from a rapidly moving vehicle is the example of the latter.

Visualization of these measurements reveals both the patterns that we may expect, as well as the patterns that are quite surprising. The close-up from the visualization of the complete film shown below (6.2.) illustrates a pattern that we may anticipate in a cross-cutting sequence: two short shots alternating back and forth, each with its own level of visual activity.
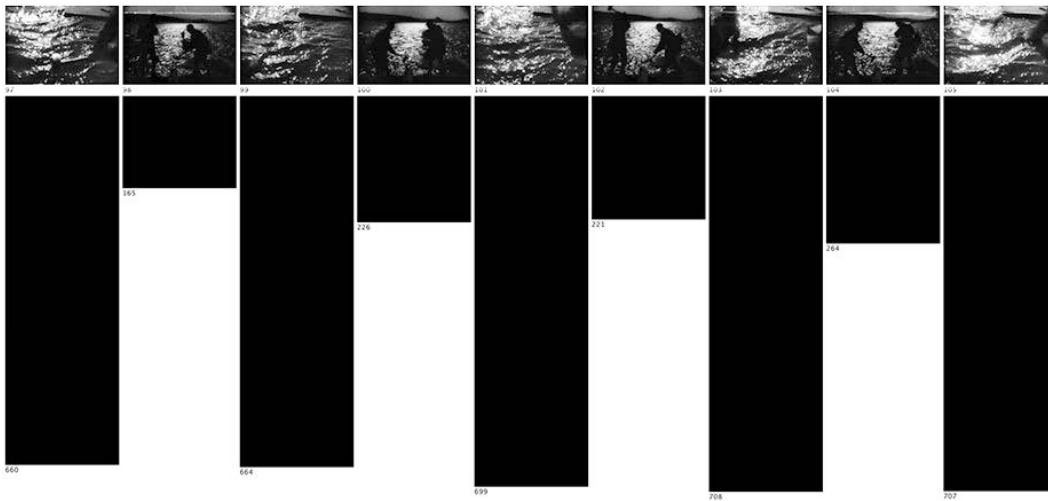
6.2.

This next close-up demonstrates a different pattern that seems to completely contradict our expectations of montage film: the average amount of visual change in each shot at first gradually decreases and then gradually increases. This pattern of gradual increase / decrease in the amount of activity occurs a number of times throughout the film.

6.3.

Vertov and a number of other Russian filmmakers (Lev Kuleshov, Sergei Eisenstein, Vsevolod Pudovkin) advocated *montage* as the key organizational principle of cinema. While they proposed a number of alternative theories of montage, common to these theories and their films was the idea of a collision between shots – i.e., generation of meaning and emotional effects through the juxtaposition rather than continuity. (In contrast, in normal film editing the progression of shots serves the primary purpose of advancing the narrative.)

However, as the example above shows, the opposition and continuity are not always enemies. In this example, the alternating close-ups and medium shots oppose each other graphically. At the same time, the amount of visual change in each shot gradually decreases and then increases over time.

Were Vertov and his collaborators Mikhail Kaufman and Yelizaveta Svilova were aware of this subtle pattern? The fact that the shots in a sequence follow a pattern is not surprising. The Russian montage theorists advocated that the shots should be arranged in a sequence following some system (for example, Eisenstein distinguished between "metric montage," "rhythmic montage," "intellectual montage," etc.). However, since they did not have computational tools, they could not analyze precisely all visual changes from frame to frame, or from shot to shot, and thus systematically plan subtle patterns of change on such dimensions such as "the average amount of visual change per shot" used in our visualization. This would not prevent Vertov and his collaborators to create such patterns "by hand" – even though they could not be graphed and named until now.

Details:

In this Vertov film, there is no correlation between shots lengths and average amount of visual change, as measured by our method described below (correlation = -0.06). While in general lots of things can result in "visual change" (think of 20th century experimental fims or contemporary motion graphics), in Vertov's films visual changes between frames are due to the movements (objects, camera, or both together). If we use this substitution, we can state that the amounts of movement shots and shot lengths do not have any connection to each other. (This is comparable to the general finding of Cutting and his collaborators for pre WWWII films they studied [1]).

Time series analysis (autocorrelation and partial autocorrelation calculated using http://www.wessa.net/rwasp_autocorrelation.wasp) of the shot lengths data and the shots movements data both show strong structures (i.e., the opposite of randomness). This means that both values change systematically in significant portion of the films.

Comparatively, shot lengths data has more structure than average amount of movement in shots data (as I measured it – see below). This makes sense: while Vertov could plan the exact length of every shot, he did not have the way to do the

same for the movement. Examining the graphs visually, we also see that the proportion of the film that has systematically varying shot lengths is larger than the part that has systematic movement patterns.

The method for measuring average amount of visual change in a shot:

To calculate the average amount of visual change in a shot, I implemented in software the following method. Each two subsequent frames are subtracted from each other, pixel by pixel. Next, the program calculates the mean (average) of all gray scale values in the difference image. Then the mean values for all difference images are then added, and the total is divided by the number of frames in the shot.

The following illustration shows difference images for two sets of consecutive frames:
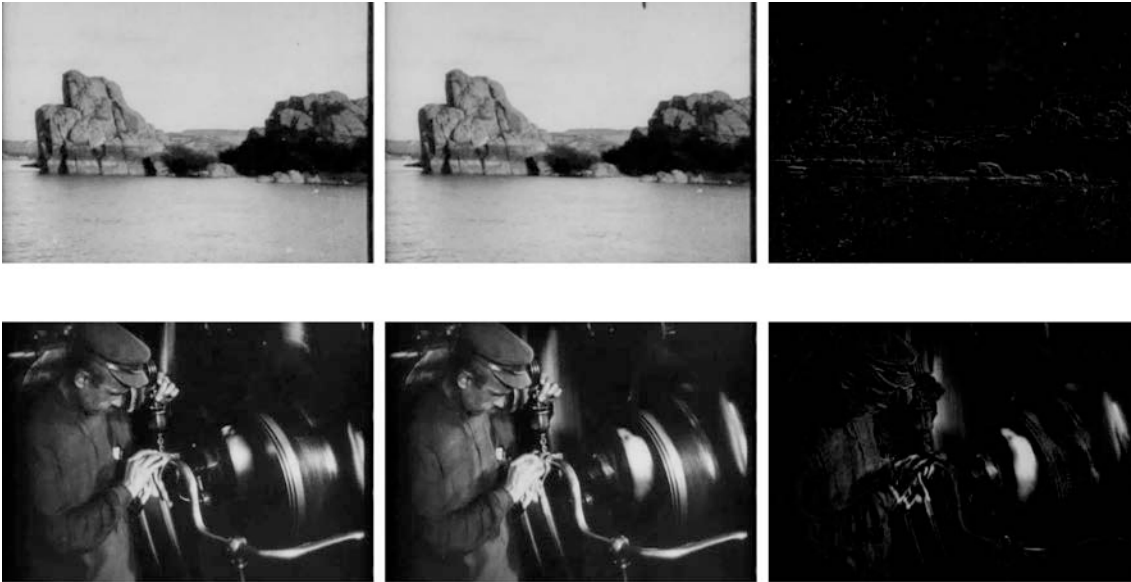
6.4. Examples of difference images. In each row, the first and second rows shows subsequent frames from *The Eleventh Year*; the third column shows their difference image.

Top row: two frames from a shot filmed from a slowly moving ship. There is a little change between the frames, and the corresponding difference image contains only a small number of non-black pixels.

Bottom row: two frames from a shot showing a worker operating machinery. The parts of the frames that change correspond to the non-black pixels in the difference image. As can be seen, in this example, there is significantly more change from frame to frame.

I believe that our method for measuring visual change is comparable to the one used by Cutting and his collaborators in their series of studies of 160 Hollywood films [1] – at least in the case of two Vertov's films. In their method, they calculate median of correlations between next-adjacent frames (1 and 3, 2 and 4, etc.) In my method, I calculate the mean of the gray scale difference image between each two subsequent frames (1 and 2, 2 and 3, etc.) To test if the two methods are comparable, I processed a number of shots in *The Eleventh Year* to calculate both frame correlations, and gray scale averages of the difference images, and plotted the corresponding graphs. I have also tested other measurements: number of pixels which change from frame to frame, gray scale median of difference images, and also the sum of the values of all pixels in a difference image (visualizations in section 9 below uses the last method.) In all cases, the results were similar.

To calculate correlations between subsequent frames, I used ImageJ with ImageCorrelationJ plugin [2].

Notes:

1. Cutting, Brunick, DeLong, Iricinschi, and Candan, **"**Quicker, faster, darker: Changes in Hollywood film over 75 years", *i-Perception* (2011) volume 2, pp. 569 – 576, http://people.psych.cornell.edu/~jec7/pubs/iperception.pdf. The details of the

method are described in Cutting, DeLong, Brunick, 2011b "Visual activity in Hollywood film: 1935 to 2005 and beyond," P*sychology of Aesthetics, Creativity, and the Arts* 5 (2011), pp. 115–125.

2. http://www.gcsca.net/IJ/ImageCorrelationJ.html.

# 7| frames sorted by their visual properties
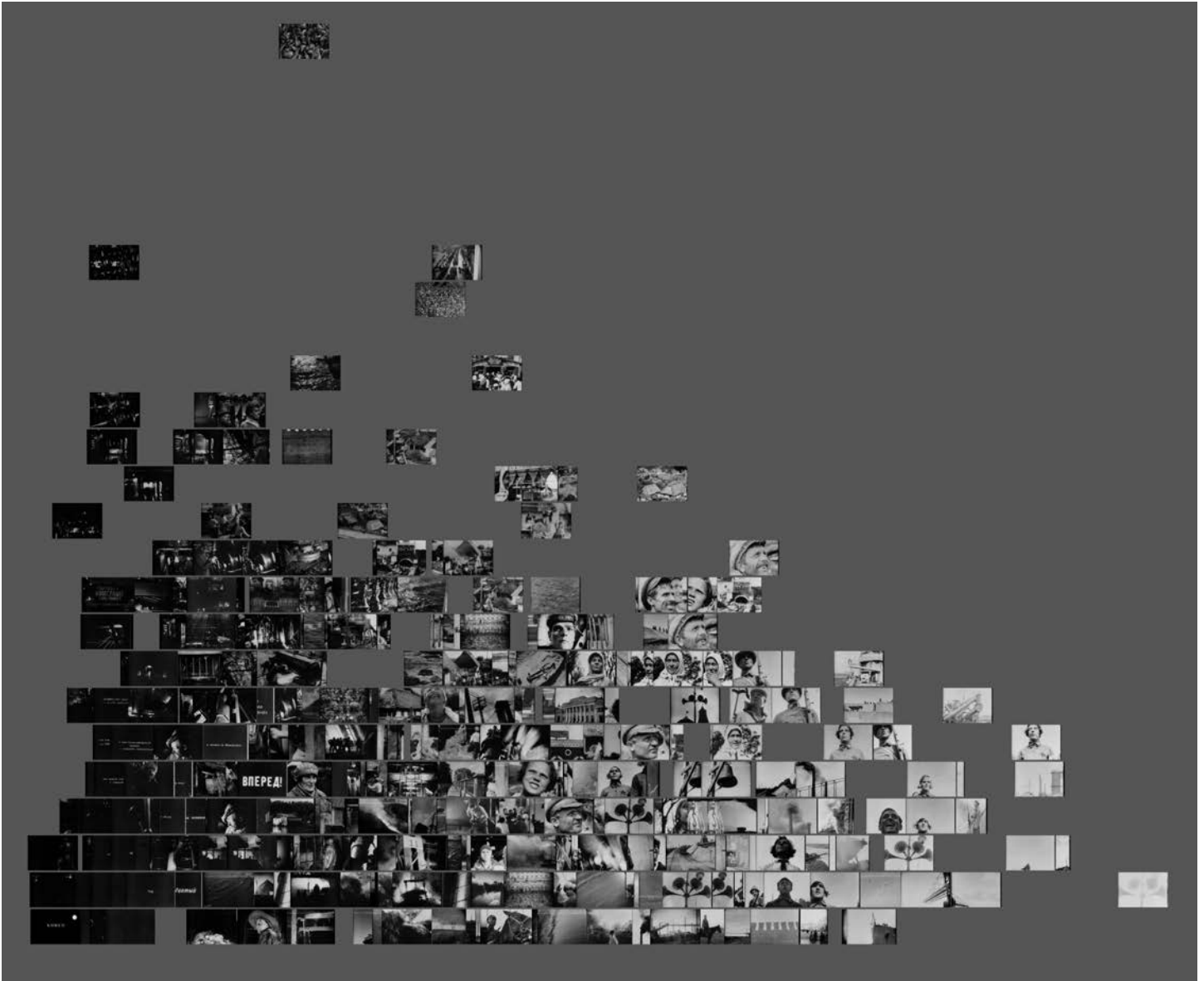
7.1. Visualization of *The Eleventh Year* showing second frame of every shot. The frames are sorted their visual properties. X-axis – mean (average) gray scale value of a frame. Y-axis – number of shapes in a frame. Note that because of the overlap, not every one of 654 frames is visible.

In this visualization, the frames representing 654 shots in *The Eleventh Year* positioned on two dimensions according to their visual characteristics automatically measured by software. Many different combinations visual characteristics can be used, each creating a different "map" of a film. Here the frames are positioned according to their average gray scale values (X-axis), and the number of shapes (Y-axis). The dark frames are on the left, while the light frames are on the right. The frames that contain only a few shapes are on the bottom, while frames with many objects are on the top.

In terms of their gray scale values (X-axis), the shots in the film are divided approximately evenly between very dark, medium, and very light gray tones. The opposition between large proportions of very dark and very light shots is specific to *The Eleventh Year*. The former are outside shots, with the sky occupying the larger part of a shot. (As I already discussed earlier, many shots show people looking up. It is not hard to understand the symbolism of such shots – people look towards the Communist future). The dark shots represent industrialization, showing people operating machinery and steelmaking.

In contrast, *A Man with a Movie Camera* takes place in a city, with time covering a whole day from morning to evening. Accordingly, the distribution of its shots is more even, with every gray tone being represented equally.
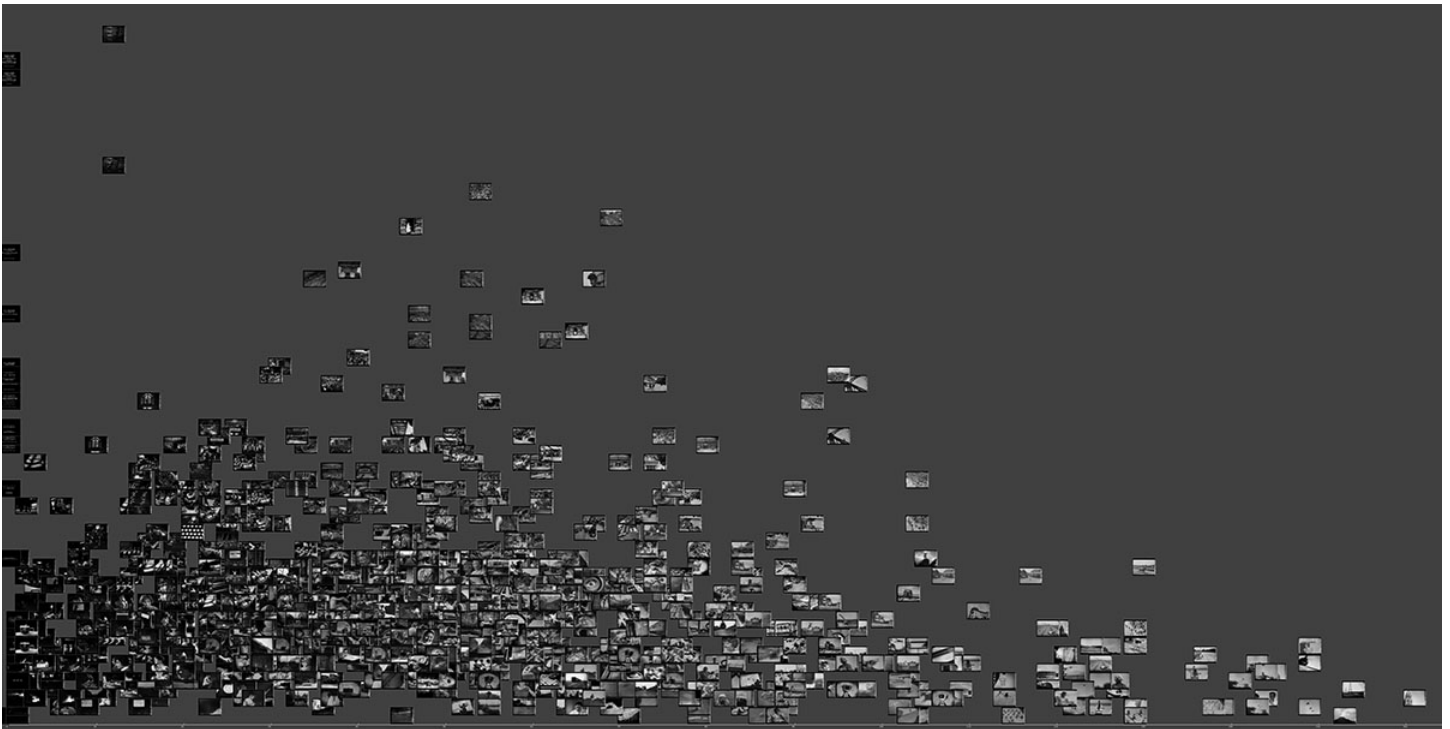
7.2. Every 100th frame of *A Man with a Movie Camera*. The frames are arranged by mean brightness (X-axis) and number of shapes (Y-axis).

7.3. A close-up of *The Eleventh Year* visualization: shots with high average gray scale value and low shape count.
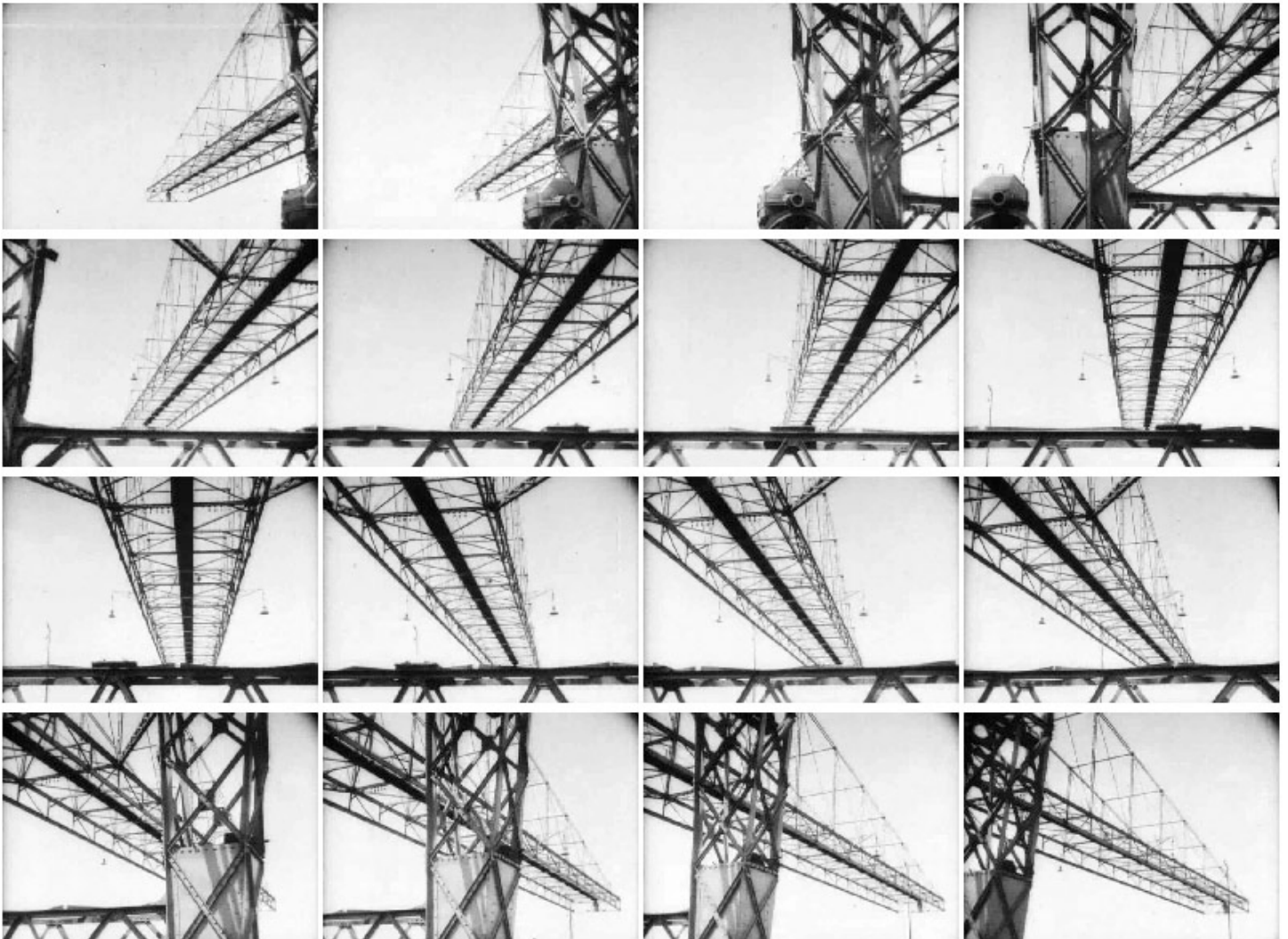

7.4. Another close-up of *A Man with a Movie Camera* visualization: shots with medium gray tones.

8| frame by frame: anatomy of a shot

This section analyzes a shot from *The Eleventh Year* – see the video clip here:
http://www.youtube.com/watch?v=_0bE9suAIDQ&feature=share&list=UUQYMh3afEnBHnwcJ8y-e9Mg

The illustration below shows sampled frames from this shot:

8.1. A shot from *The Eleventh Year* consisting from 167 frames. This montage every 21st frame; the frames are arranged left to right, top to bottom.

Russian montage school privileged shot as the basic unit of cinema. However, if we look at their films as opposed to theoretical texts and manifestos, we find that the reality does not always correspond to the theory.

The Eleventh Year contains many short shots that are static (or almost static) tableau. They contain a single "burst" of information, presented on the screen just long enough for a viewer to absorb it, and then are replaced by the next burst. The film also contains longer dynamic shots that show some activity that follows a cycle. In addition to communicating semantic information (a human or a machine performing some action), such shots also communicate the larger theme of *work* central to the film. Because they are feature movement, they don't just signify "work" – instead, they are motivating the viewers to join workers they see on the screen. We can also find a third type of longer shots that do not contain

repetitions; instead, new information is communicated as the shot unfolds. (Some of these shots are filmed from a train; as the train moves forward, we see new parts of the landscape, or new objects passing by).

How can we visualize the development of a single shot? As our example, I selected a shot which exemplifies Vertov's aesthetics in a pure way: moving geometric machine forms. In his first published text "We: The variant of the manifesto" (1920), young Vertov categorically states:

> The machine makes us ashamed of man's inability to control himself.
>
> For his inability to control his movement, WE temporary exclude man as the subject of film.

This adoration of machines was typical of the aesthetic programs of many various avant-garde groups in the 1910s-1920s, including visual artists, poets, architects, photographers and graphic designers. Vertov adapts this general program of the European avant-garde to the medium of cinema. What he takes from machines is the precision and regularity of their movements. ((Vertov was apparently unaware of the time-study methods of Taylor and motion studies by the Gilbreths already developed in the 1920s. They were used these methods to standardize and rationalize workers movements, tuning them to achieve the ideal of machine-like precision. In Russia the ideas of Taylor became popular in the early 1920s.)

The shot I selected lasts 167 frames, and it contains no human beings. Instead, the camera tracks along a long crane that extends perpendicular to it. The unfolding geometry contains a horizontal part that occupies the same part of the frames; two vertical parts that rapidly cross the frames 1-50 and then again 100-167; and the long part of the crane, which is perpendicular to the camera. Because the camera pans, the position of this part continuously changes throughout the whole shot.

Here are some possible ways to visualize a shot (many others are also possible). In the first visualization below, all frames in the shot were averaged together, pixel by pixel. In such visualization, the objects that appear in the shot only briefly are no longer visible. The objects that appear in the same position for periods of time show up as dark and sharp outlines (e.g., a dark line in the bottom). What about moving objects? In this shot, the camera pans to show a long crane part that extends perpendicular to it. Because of the pan, the position of the part continuously changes throughout the shot. In the visualization, this movement is translated into the blurred triangle-like form; the faster the movement, the more blurred is its representation.
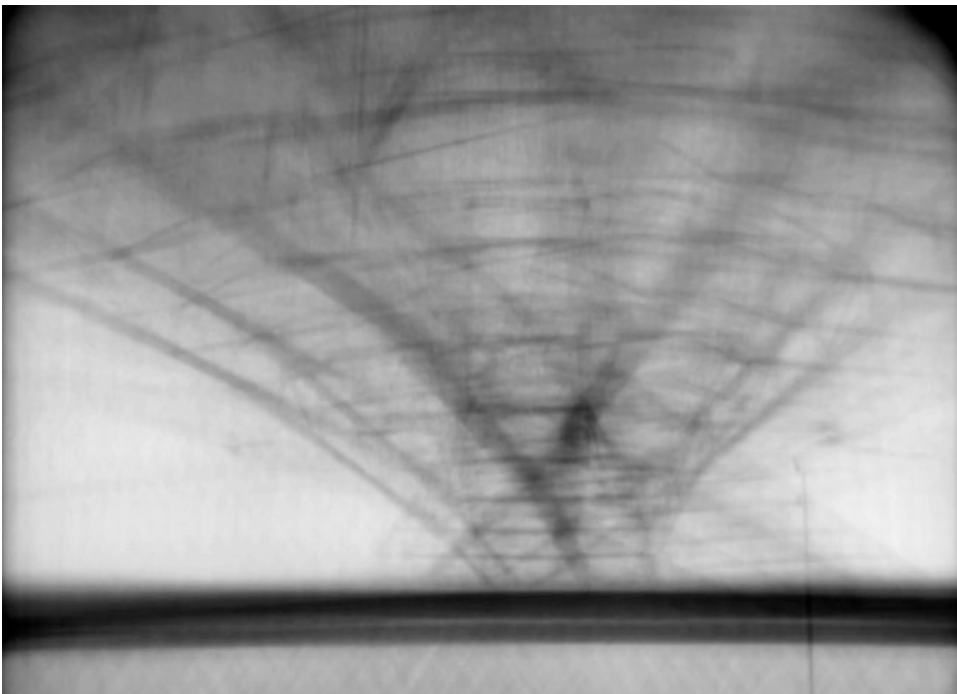
8.2. 167 frames comprising the shot averaged together. Because such composite images tend to have low contract, I have increased the range of gray levels in Photoshop.

Instead of adding all frames, we can instead add subsets of frames, generating a number of images. In the following visualization, each image is a result of adding 10 subsequent frames:
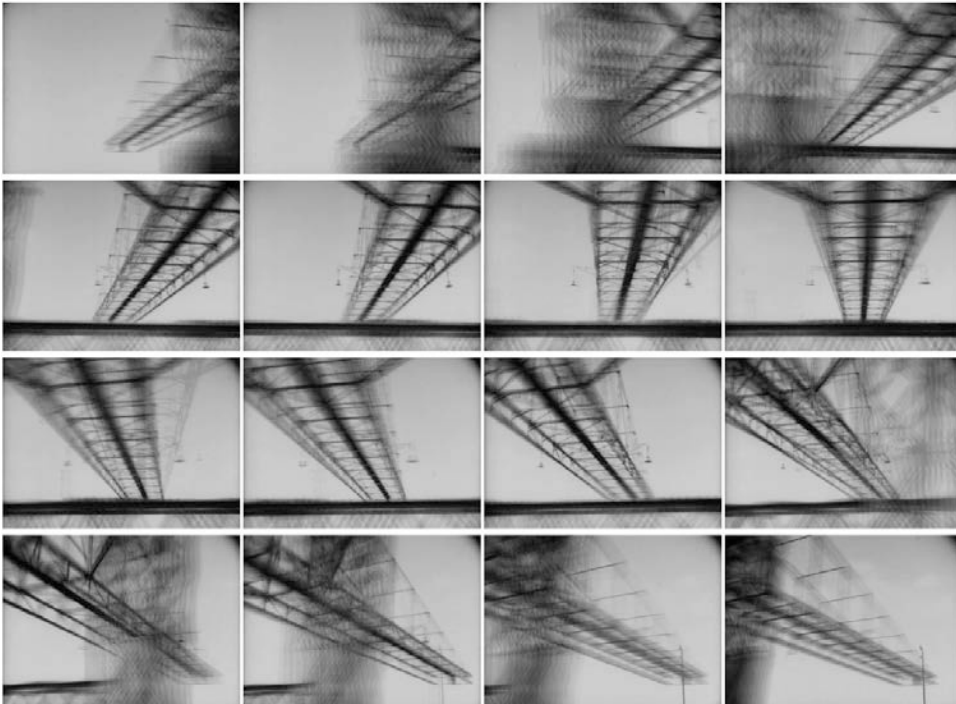
8.3. Each of 16 images shows 10 subsequent frames of the shot added together.

This visualization allows us to see more clearly the changing speed of the relative movement of the crane (I say relative because in reality the crane stands still, and the camera pans past it). The images with a lot of blur correspond to faster movement (beginning and end parts of the shot); the sharper images correspond to slow movement (middle part of the shot).

Another approach to tracking what happens inside a shot is by using graphs. Earlier (visualization 6) we graphed average amount of visual change per shot in the whole film. We can use the same method to measure and graph the changes from frame to frame in a single shot.

First, we generate difference images for each two pairs of frames. A "difference image" represents the changes between two consecutive frames. In the example below, the corresponding pixels that did not change between two frames are black; the pixels that changed have lighter values.

You can see a video of difference images for the whole shot online:
http://www.youtube.com/watch?v=zCcQq3J7OLA&list=UUQYMh3afEnBHnwcJ8y-e9Mg&index=2

The illustration below shows the three frames from the video:

8.4. Selected difference images generated by subtracting subsequent frames from the shot.
Top to bottom: difference images for frames 37-38, 77-78, 133-134.

The top and bottom images correspond to parts of the shot when the camera passes the vertical parts of the crane. Because these parts are close to the camera, their movement generates large differences between subsequent frames. The middle difference image corresponds to the part of the shot when the camera passes the part of the crane that is perpendicular to it. Because this extended part is further away from the camera, the changes in its position translate into smaller numbers of pixels changing from frame to frame.

After we generated these difference images for each two frames in a shot, we measure gray scale information in each difference image, and then plot these measurements over time. Different measurements may be used, but in general they tend to produce similar results. In section 6, we used the gray scale mean of a difference image, here we will use an alternative measure: the sum of pixel values in each an image (ImageJ software which I used to automatically do these measurements calls this Raw Integrated Density: http://imagej.nih.gov/ij/docs/menus/analyze.html - set).

The graph below plots this measurement for all 167 frames of the shot. To make the graph more understandable, I placed 9 key frames from the shot below. Every key frame was sampled at 20 frames interval, starting at frame 1.
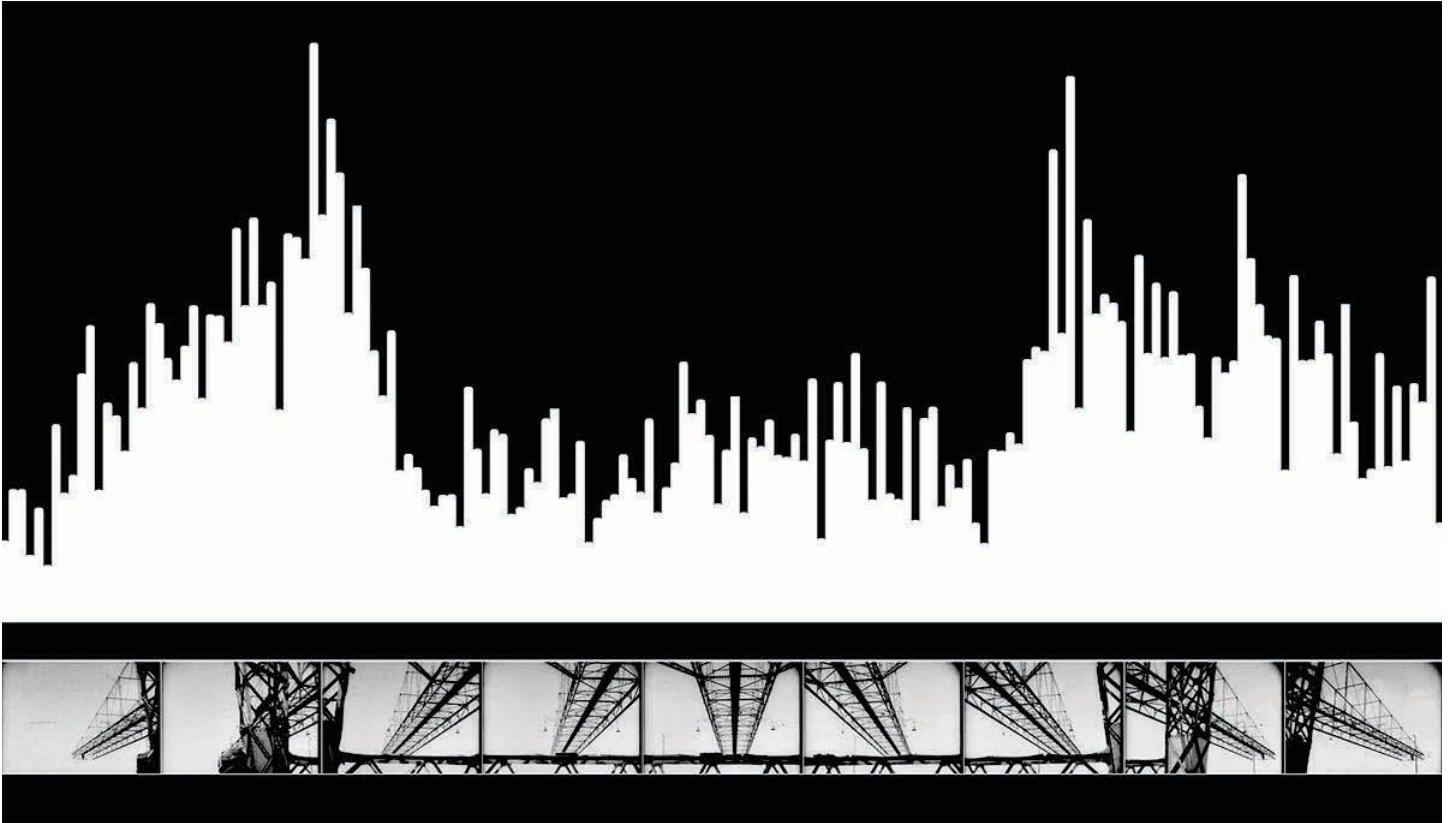
8.5. The graph of the frame differences in the shot from *The Eleventh Year.* The measure of the difference between subsequence frames used is Raw Integrated Density of a difference image. Note the occasional spikes in the graph. The reason for them is the jumps between certain frames, which is the artifacts of the original film recording. If the same shot was recorded with a contemporary camera, these spikes would disappear.

Single measurements of visual change between pairs of frames such as a gray scale mean of a difference image, or its raw pixel density reduce all visual changes to a single number. Therefore, they don't allow us to track separate dimensions of a shot - movements of particular objects, composition, camera position, people gestures and faces, etc.

The same holds for all other visual dimensions of a shot. Unless we are dealing with an extremely "structural" shot where visual change is confined to a single parameter (think of Hans Richer's *Rhyhmus 21* or parts of Michael Snow's *Wavelength*), no single graph can capture all the changes you can see with your eyes. While we can construct many graphs that show patterns on many separate dimensions in a single shot, they may still not capture the overall gestalt that we experience in watching the shot.

However, the graphs have their own advantage: by representing changes on distinct visual dimensions with curves, they give us a visual language to talk about temporal patterns. In the graph above we see two peaks (around frame 37 and frame 137) that correspond to the moments than vertical parts of the crane pass the frame. We can also see that sweeping movement of the long perpendicular crane part (the part between the two peaks) is translated into approximately the same rate of visual change. The graph also confirms that the rate of change of the first part of the shot (0-37) is larger than in the last part (137-168). (The jumps in value from frame to frame reflect the artifacts in the film recording.)

9| **visualizing movement**

This section analyzes three shots from *Man with a Movie Camera* – see the video clip with these shots here:
http://www.youtube.com/watch?v=PV-FzvHi0lk&feature=share&list=UUQYMh3afEnBHnwcJ8y-e9Mg

The illustration below shows sampled frames from these shots.

9.1. Three consecutive shots from *Man with a Movie Camera.* The lengths of the shots are 94 frame, 115 frames, and 138 frames. The montage shows uses every 21st frame of this sequence.

In the same manifesto "We" (1920) I quoted in the beginning, young Vertov defines his cinema (still to be created) as the art of organizing movements:

Kinochestvo is the art of organizing the necessary movements of objects in space as a rhythmic artistic whole.

Cinema, as well, the art of inventing the moments of things in space in response to the demand of science.

Did the practice, which came later, fit these statements made much earlier? Looking at films such as *The Eleventh Year* and *Man with a Movie Camera,* we find that only some parts of the films use objects movements in space as the organizing principle. Other parts contain sequences of still-like shots, or shots with little or slow movement.

This impression is confirmed if we analyze the films. Adelheid Heftberger has manually annotated motion information in each shot in these films using a number of tags. In *The Eleventh Year*, she tagged only %7 of shots as having "fast motion"; she assigned "no motion" tag to %22. As we may expect from viewing the two films, in *Man with a Movie Camera,* he percentage of shots tagged as "fast motion" was much higher (%30) - but its still only represents one third of the film.

Nevertheless, Vertov's initial idea of film as an organization of "movements of objects in space" is important. We can find this technique used later in numerous 20th century films, beyond key parts of Vertov's own films. But in order to study movements more precisely and on a large scale, we need automatic techniques for movement tracking, analysis and visualization. Where do we start?

In the introduction I noted that the development of visual recording media in the 19th century went hand in hand with inventing techniques to capture movements for the analysis by Marey, the Gilbreths, and others. In the 20th century, rotoscoping – manual tracing of filmed movements of actors – became a major technique in animation industry. In the last decades of the century, more methods for automatic *motion capture* and analysis that use computers and expand on the original techniques of Marey and the Gilbreths were developed and adopted in many fields. These fields range from sports and video game production to surveillance and automatic car navigation. Given the ubiquity of inexpensive video cameras, detection, tracking and identification of movement-based activity (including behavior analysis) using video recording has became particularly large area of research in computer science.

Computer scientists have developed a number of techniques for automatic motion estimation and movement tracking, and these techniques are built into most basic digital media technologies. For example, MPEG video codecs use automatic motion estimation to compress video.
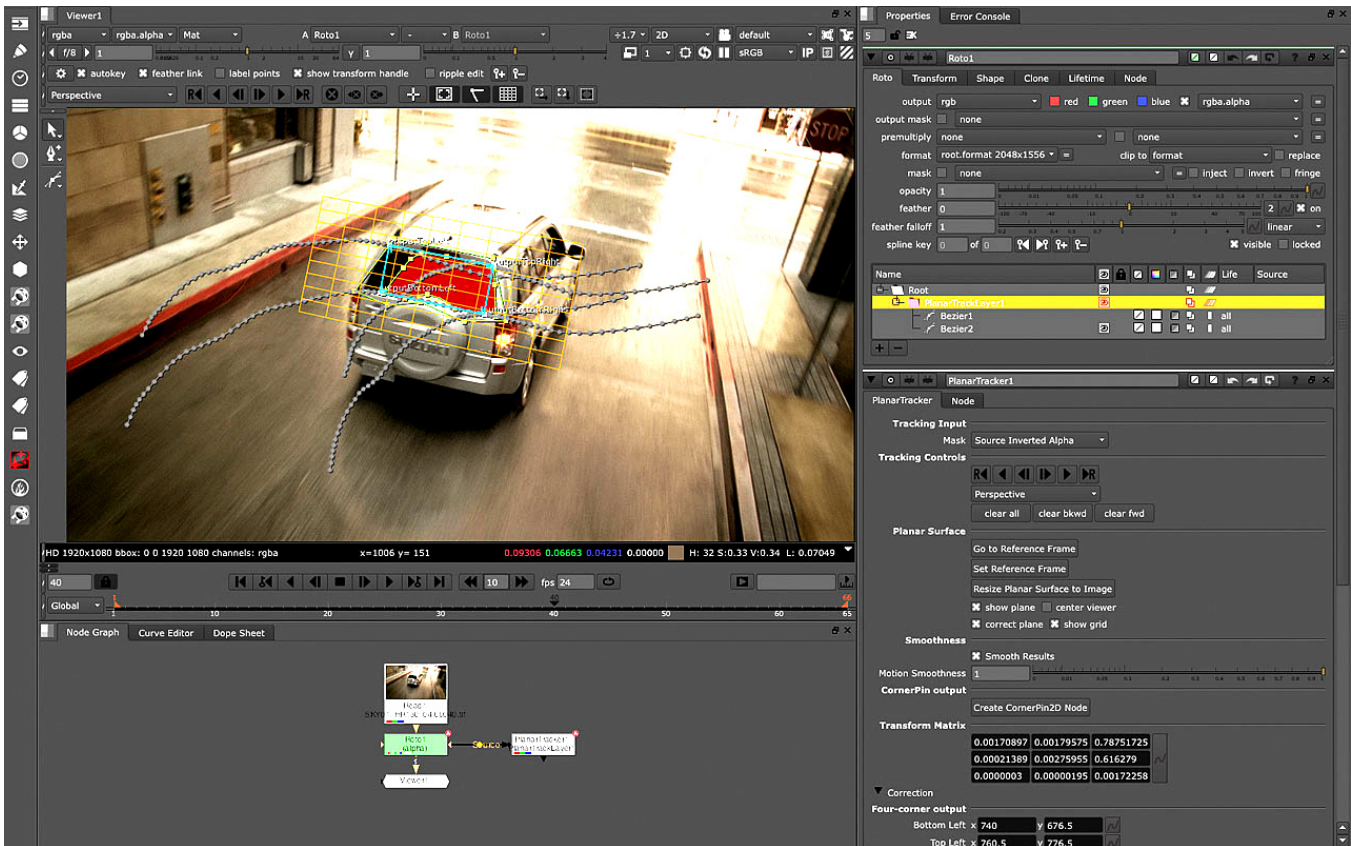
9.2. Planar tracker in popular NUKE 6.3, one of the most popular software used in cinema and TV production for motion tracking.

9.3. Typical use of motion capture in cinema and video games: the movements of an actor are captured and used to drive an animated character.

Most of the software for animation, compositing, and visual effects also offer a number of techniques that use automatic or manual methods or their combination (A comprehensive overview of tracking in film/video production is provided in [1]). All animated characters in video games and feature films today use motion and face capture of live actors. Films and commercial also rely on the tracking of movements of objects and camera in video footage in order to combine live action footage and computer graphics. In fact, *motion tracking and motion capture maybe as fundamental to early 21st century cinema and video production as the development of editing was a hundred years ago.* (In this respect, the idea of cinema as the art of moving things in space put forward by 24-year old Vertov in 1920 anticipated both the major technology of early 20th century cinema and aesthetics of many fantasy and action films made possible by these technologies. Imagine what Vertov would have done if he had access to these technologies in his time!).

Typically in professional film and video production today, tracking movements within a single shot requires some time, and the quality of the results and the time requires depends on the type of the shot. [2] However, the most challenging issue in using any method to analyze movements in films of Vertov or any other director is not technical but theoretical. Let's imagine that we managed to track every movement of every object in the complete film across hundreds or thousands of shots: how do we visualize, analyze and interpret this data? In the case of average shot lengths, the data was in one dimension (a sequence of numbers representing the lengths of shots in a sequence) – but now, we may have hundreds of numbers for every second representing movements of potentially dozens of objects in every shot.

To illustrate this conceptual challenge, I will use extremely simple "toy method" for visualizing movement that already appeared in previous section (8.2): averaging all frames in a shot together. This method works only partially but it's sufficient for an illustration. Here are averages of three shots from a clip above created using the method described in section 6.

9.4. Visualizations of movement in three shots from *Man with a Movie Camera* using frame averaging techniques. The video of the shots is shown in 9.1.

As I explained earlier, when we apply this technique to map a shot into a single image, the objects that appear in the shot only briefly or move quickly through space disappear; the objects which move more slowly for longer periods of times are translated into blurred parts of the image.

In the visualization of the first shot (9.4, left), the main movement of the tram is well represented. In the visualization of the second shot (9.4, center), the results are less successful: the arc movement of the tram is preserved, but the faster moving car almost disappears. In the visualization of the third shot (9.4, right), all we can tell is that some movements occurred where the image is blurred, but we can't say anything more.

All in all, the technique is partially successful in making visible "large" movements which have simple geometry (here, moving trams), but most of the "small" movements" (here, people on the street, or the hands of the typist) are not preserved. I take their lack of visibility in these images as a metaphor for the conceptual challenge of describing them. What would we do with the tracks of all people in the first two shots if we could obtain them? It is not easy to describe conceptually - and yet I think they are as important to these shots as the big scale movements of the trams. (Gilles Deleuze*'s* verbal descriptions of different types of cinematic movements in his *Cinema 1: The Movement Image* are both fascinating and frustrating to read it, as we witness his struggle to express linguistically the variety of the movements in cinema universe [3]).

This article illustrated only some of many other possible ways to visualize cinema – from the scale of thousands of films to a single shot. New film languages and new cinematic forms which develop beginning in the middle of the 1990s as the result of adoption of software tools and digital workflow calls for their own visualization techniques. *Motion graphics* videos that often do not have discrete objects or representational content allow for particularly interesting visualization approaches that I will describe in a future article. Because movement is even more crucial for motion graphics (as the name itself implies) than for feature films, developing adequate techniques for visualization of movement in these works should allow to reuse them with Vertov films. Ideally, we will go beyond his desire for "graphic symbols of movement" creating more precise, expressive, and rich visualization and analytical tools that will allow us see cinema in new ways.

11/2008 - 1/2013

Notes:

1. The comprehensive overview of the related concepts, technologies and their applications can be found in this set of Wikipedia articles:

http://en.wikipedia.org/wiki/Camera_tracking
http://en.wikipedia.org/wiki/Match_moving
http://en.wikipedia.org/wiki/Motion_estimation
http://en.wikipedia.org/wiki/Motion_capture

2. For a partial history of tracking in visual effects production up to 2004, see Mike Seymour, "Art of Tracking Part 1: History of Tracking," fxguide.com, 8/24/2004, http://www.fxguide.com/featured/art_of_tracking_part_1_history_of_tracking/.

3. Gilles Deleuze*, Cinema 1: The Movement Image* (Athlone Press, 1983), translation of the original 1983 publication.