

Automatic Timeline Visualization of Paintings

Tom De Smedt

St Lucas School of Arts, Antwerp

tom@organisms.be

Thomas Crombez

University of Antwerp

thomas.crombez@uantwerp.be

Lucas Nijs

St Lucas School of Arts, Antwerp

lucas.nijs@kdg.be

Keywords: art history, open data, clustering, text analytics, data visualization

Visual timelines of artists and artworks have played a crucial role in art historiography since the publication of Alfred H. Barr's catalogue *Cubism and Abstract Art* in 1936. This book, which accompanied an exhibition by the same name at the New York Museum of Modern Art, prominently featured a diagram of artistic influence on its cover. Since then, timelines have aided students and scholars of art history in identifying watershed moments, or charting the influence of artists and art movements (Grafton & Rosenberg, 2010). A notable example is the 'Great Wall' of images from Western art history by time period, constructed by British painter David Hockney in 1999. It led Hockney to conclude that the sudden rise of realism around 1420 was to be attributed to new optical techniques, notably the use of concave mirrors (Hockney, 2001).

We propose an automated method for visualizing images from art history. To this end, we have automatically collected a large-scale corpus of artworks by notable painters.

The corpus consists of over 3,500 painters mined from Wikipedia¹, including their name, age, gender, period, style, genre, bio, keywords and impact factor. This information was extracted automatically from Wikipedia articles using text mining techniques (De Smedt & Daelemans, 2012) and weak supervision. For example, the impact factor is based on the Wikipedia article length and the number of Google Search results for each painter.

The corpus has over 5,500 JPEG images (i.e., reproductions of known paintings) of reasonable quality. Each image has a star rating manually assigned by two annotators. The inter-rater agreement score (Fleiss' kappa) is +0.2, which indicates a fair agreement between annotators as to what are 'interesting' images.

¹ https://en.m.wikipedia.org/wiki/List_of_painters_by_name

Since all content on Wikipedia is licensed under the GNU Free Documentation License (GFDL), all collected texts and images are rights-free. The corpus will be released for free at DH Benelux. A companion corpus has over a 1,000 reviews of art books that mention the name of a painter in their title, mined from Amazon. Each review has been automatically annotated with a sentiment score (positive or negative) and a profile of the reviewer, including age, gender and education level².

Preliminary statistical analysis shows that:

1. most painters in the corpus have low impact while only a handful have very high impact,
2. most painters tend to outlive the average life expectancy in their time period,
3. female painters have been historically underrepresented,
4. Amazon reviews focus on painters with a high impact,
5. Amazon reviews of art books on notable painters are more often written by highly-educated adult women.

We are using the corpus to visualize a timeline of the history of painting. If such a wall piece is 10x2m, then it can display about 30% of all paintings in the corpus at 10x10cm. If more interesting paintings are displayed larger, this leaves less room for other paintings. The challenge is to automatically select a subset of paintings and fill the wall in a meaningful way: paintings should be displayed in their time period, paintings in the same style should be clustered, paintings by related authors should also be clustered, and so on.

Our current approach is as follows. Keywords for each painter are extracted from the links in the corresponding Wikipedia article. Along with time period and style, these are used as features in an unsupervised k -nearest neighbor classification model (k -NN). It will classify painters in the same period, same style, and sharing more Wikipedia links, as more similar, e.g., Pieter Bruegel the Elder ↔ Pieter Bruegel the Younger (cf. Crombez, 2012). Styles that occur more frequently in the corpus such as Impressionism and Cubism – which may or may not reflect reality – are assigned an anchor position on the wall. Painters with a high impact and paintings with a high star rating are positioned near their respective anchor. Then, similar painters are iteratively positioned near those painters, using spreading activation (Fig. 1). This means that the algorithm will start from an anchored painter and radiate outwards to find a free spot (Steele & Iliinsky, 2010). For each painter, we display the JPEG image with the most stars and the highest resolution. We use NodeBox 1 (De Smedt, 2013) to render the visualization (Fig. 2). NodeBox is free and open source³. We will release the source code of the algorithm once it has been cleaned up.

Automatic techniques for constructing visual timelines can be useful for many purposes. Firstly, they are unbiased, image selection and placement is based on majority voting instead of one opinion. Secondly,

² <https://www.textgain.com>

³ <http://www.nodebox.net>

they propose layout solutions for thousands of images in a matter of minutes, providing a starting point for new canonical insights⁴ for art historians, students, archivists, photographers, and so on.

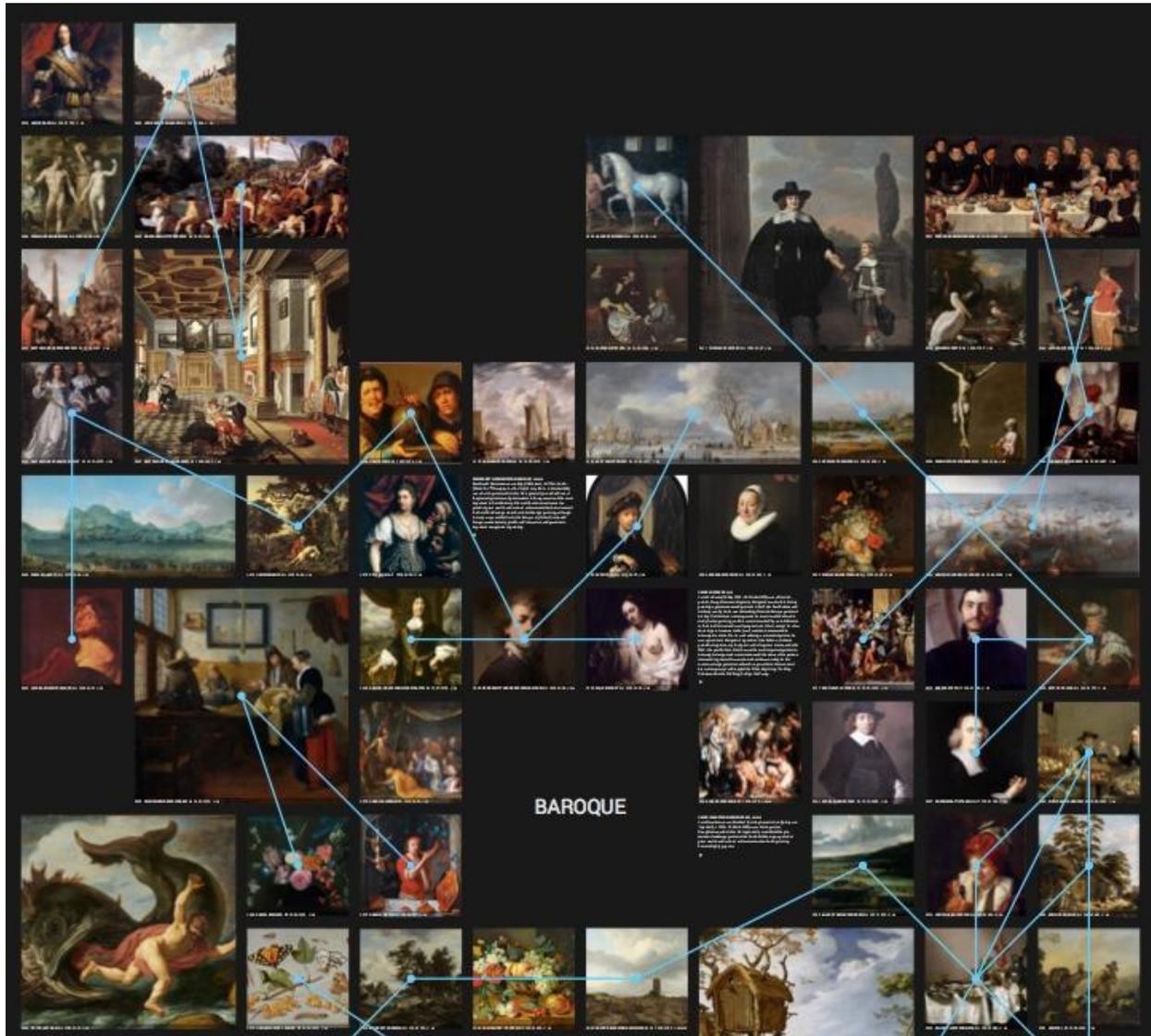


Figure 1: Paintings being placed around the BAROQUE anchor, starting with Rembrandt and Van Ruisdael

⁴ <https://www.youtube.com/watch?v=JKbFZlpNK10>



Figure 2: Timeline exhibited in the St Lucas School of Arts Showroom, Antwerp, April 6, 2016

References

- Barr, A.H. Jr. (1936). *Cubism and Abstract Art*. The Museum of Modern Art.
- Crombez, T. (2012). Mapping a Landscape of Texts. *Performance Research*, 17(3), 120-124.
- De Smedt, T., & Daelemans, W. (2012). Pattern for python. *The Journal of Machine Learning Research*, 13(1), 2063-2067.
- De Smedt, T. (2013). *Modeling Creativity: Case Studies in Python*. University Press Antwerp.
- Grafton, A., & Rosenberg, D. (2010). *Cartographies of Time: A History of the Timeline*. Princeton Architectural Press.
- Hockney, D. (2001). *Secret Knowledge: Rediscovering the Lost Techniques of the Old Masters*. Thames & Hudson.
- Steele, J., & Iliinsky, N. (2010). *Beautiful visualization: looking at data through the eyes of experts*. O'Reilly Media, Inc.