Data-Driven Curated Video Catalogs: Republishing Video Footage

Gabriele Colombo
University of Amsterdam

Federica Bardelli
University of Amsterdam

ABSTRACT

This project illustrates data-driven curated video catalogs as an approach for the analysis of video footage. Having rich and diverse collections of videos as inputs, data-driven catalogs seek to identify common objects and reorganize them into thematic clusters displayed in a video format. The technique takes inspiration from two scientific practices (core sampling and light diffraction) and two publishing formats (supercuts and visual catalogs). Data-driven curated video catalogs are used to republish a collection of found footage of 2019 Venice's high water that devastated the city in an unprecedented way. Starting from an editorial selection of footage culled from YouTube, various algorithmic processes are used to demarcate and reorganize the material into thematic video series (people, boats, and birds). The resulting video catalogs support a type of visual analysis that goes beyond traditional forms of measurement while, at the same time, carrying an expressive power.

Keywords
Video footage
Video catalogs
Visual analysis
Supercuts
Computer vision

AUTHORS

Gabriele Colombo—MA in Communication Design, Politecnico di Milano. Ph.D. in Design, Politecnico di Milano. He is affiliated with DensityDesign, a research lab at the Design Department of Politecnico di Milano, and with the Department of Architecture and Arts of the Università IUAV di Venezia. From March 2019 to June 2021, he has been a researcher at the University of Amsterdam in the European research project ODYCCEUS. At Politecnico di Milano he is also a lecturer in the Communication Design Master's Degree, where he teaches Digital Methods and Communication Design. His research focuses on the design of visualizations in support of digital social research. He is a founding member of the Visual Methodologies Collective at the Amsterdam University of Applied Sciences and has a long-standing collaboration with the Digital Methods Initiative at the University of Amsterdam. Some of his latest publications include: 'Dutch Political Instagram. Junk News, Follower Ecologies and Artificial Amplification' (with C. De Gaetano; in *The Politics of Social Media Manipulation*; Amsterdam University Press, 2021) and 'Studying Digital Images in Groups: The Folder of Images' (in *Advancement in Design Research at Polimi*; Franco Angeli, 2019).

Federica Bardelli—Visual designer and visual artist (New Art Technologies), Accademia di Belle Arti di Brera. MA in Communication Design at Politecnico di Milano. She collaborated with the research department Density Design Lab. She also has taught Communication Design at Politecnico di Milano and worked with IED (Istituto Europeo di Design), the Italian Visual Arts school. From March 2019 to June 2021, she has been a researcher at the University of Amsterdam in the European research project ODYCCEUS. Her work focuses on new visual languages applied to research, specializing in information and data visualization strategies with digital methods. Currently she works at the intersection of digital and multimedia arts, deriving common languages between research and artistic practices. Some of her latest publications include: 'Confronting Bias in the Online Representation of Pregnancy' (with L. Bogers, S. Niederer, and C. De Gaetano; *Convergence: The International Journal of Research into New Media Technologies*, Vol. 26, Issue 5-6) and 'Corpi digitali - Un campionario di tecniche di produzione. Digital Bodies - An Inventory of Production Techniques' (with G. Colombo and C. De Gaetano; *Progetto Grafico*, Issue 31).

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

Data-Driven Curated Video Catalogs: Republishing Video Footage

Gabriele Colombo
University of Amsterdam
Faculty of Humanities
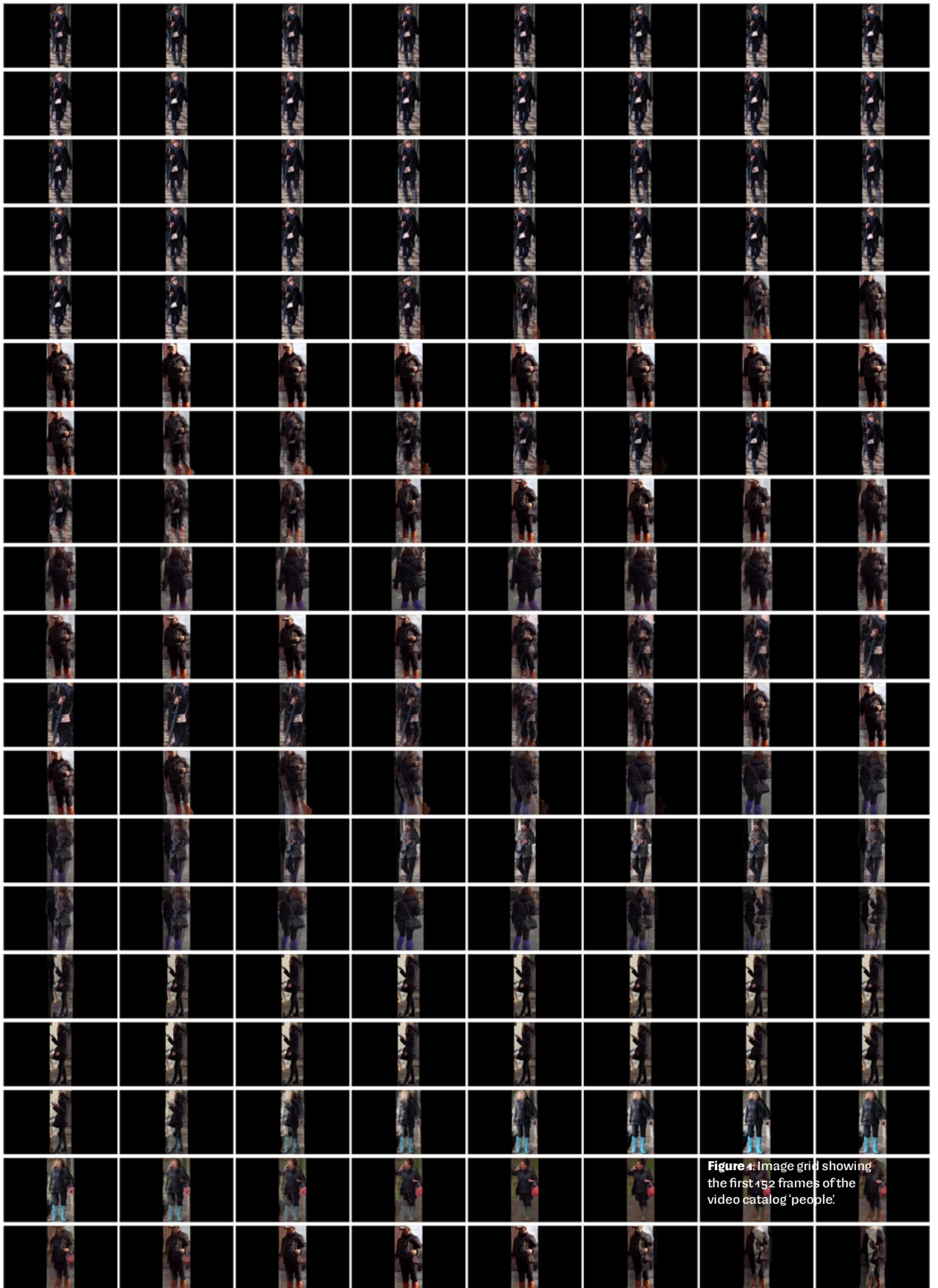Media Studies Department
Amsterdam, Netherlands
g.colombo2@uva.nl

Federica Bardelli
University of Amsterdam
Faculty of Humanities
Media Studies Department
Amsterdam, Netherlands
f.bardelli@uva.nl

REPUBLISHING VIDEO FOOTAGE FOR THE ANALYSIS: CORE SAMPLING, LIGHT DIFFRACTION, SUPERCUTS, AND VISUAL CATALOGS

How does one republish a large collection of found video footage for the analysis? The following project described applies machine vision to a catalog of YouTube videos of the 2019 Venice flood, or *acqua alta*, to create a jittery assemblage of close-ups on algorithmically detected objects and motifs. It consists of a series of data-driven curated video catalogs: starting from an editorial selection of footage culled from YouTube, various algorithmic processes are used to demarcate and consolidate the material into thematic video series (namely people, boats, and birds). The artwork offers a new lens through which the viewer can observe the amateur footage culled from YouTube, offering unexpected perspectives on the diverse experiences, objects, and themes that contributed to narrate the event. Through the juxtaposition of apparently similar images, repetition is used both as an analytical tool, in that it reveals unnoticed visual patterns, and as a rhetorical device, as it stages an inexorable feeling of instability, restlessness, and growing sense of anxiety. In the background, Venice's iconic views are barely recognizable (see Figure 1).

Figure 1: Image grid showing the first 152 frames of the video catalog 'people'.

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

The republishing technique described here nestles itself within two methodological inspirations (core sampling and light diffraction) and two publishing formats (supercuts and visual catalogs). Accordingly, in the next paragraphs each is taken in turn, followed by a detailed explanation of the process. We conclude with a series of reflections on the limitations of working with machine vision to republish video footage.

The first methodological inspiration is core sampling. In mining, geology, and archaeology, core sampling (*carotaggio* in Italian) involves taking cylindrical samples from the subsurface. Core samples are obtained by drilling with special drills into the substrate (sediments, rocks, ice) with a hollow steel tube. From extracted core drills, one can perform various observations, such as determining soil's lithological characteristics, analyzing the nature of the fluids present, and assessing the location of rock layers (Treccani Enciclopedia, n.d.). It is a sampling technique, as one isolates a portion of the material to analyze it better. This project proposes core sampling applied to a video collection, as one drills down a lens in the flux of available videos and extracts visual themes. The technique is applied to the temporal dimension of the video footage, extracting image segments of detected objects frame after frame. The goal is to isolate motifs, topics, and objects to inspect them more closely, as one would do with soil or ice samples.

The second inspiration is light diffraction. In optics, a diffraction grating is a component that splits and diffracts light into several colored beams traveling in different directions ("Diffraction Grating," 2021). It is an obstacle that spreads light into separate channels. Similarly, this project aims to regroup and cluster video footage by theme, using machine vision algorithms as an obstacle. As light waves diffract when encountering an obstacle, video frames are divided into multiple images, each corresponding to a particular object detected by the machine.

The third inspiration comes from the format of supercut. Supercuts is a genre of web videos popular in the early days of Internet culture, consisting of "fast-moving, detail-obsessed videos" that isolate "a recurring pop-culture trope" (Raftery, 2018). It is a form of 'metatextual montage' (Raftery, 2018): through repetition and juxtaposition, it foregrounds repetitive patterns and clichés that would otherwise go unnoticed in the wealth of massive collections of video footage. An early example compiles footage of reality-show contestants saying the phrase "I'm Not Here to Make Friends!" (Juzwiak, 2008), another one isolates and collects Hollywood characters dramatically yelling "It's showtime!" (Hanrahan, 2011). Supercuts may also fall under the category of fan art, and be specific to a movie (Maggot3560, 2006), series (Zabriskie, 2008), or actor (Noble, 2010), to which the video compilation pays homage.

While the format gained traction in the first decade of the 2000s, there are pre-internet experimentations with the technique. As a case in point,

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

Jennifer and Kevin McCoy compiled thousands of individual shots from the 1970s show Starsky & Hutch divided by theme. The artwork, titled 'Every Shot, Every Episode', breaks down the footage and sequences it into categories such as 'every zoom in', 'every car chase', 'every dead body'. The artwork foregrounds repetition in the footage, and it gives equal significance to each shot, as no narrative ties them together. More recently, the technique lives a new life as a means for political commentary and satire. VICE News (2018) has compiled a series of videos with recurring expressions by Donald Trump, such as "24 things nobody does better than Donald Trump" or "All of Trump's billions." As a form of cultural commentary, or a satirical take on politicians' recurrent tropes, supercut relies on repetition and linearity to foreground differences and similarities in video footage.

The fourth inspiration comes from the (visual) catalog as a format for collection. When collected in the same optical space, different images can gain the status of a group, and therefore they are read as a collection of associated elements. The visual catalog resembles the format of the list, which "confer(s) unity on a set of objects no matter how dissimilar among themselves" (Eco, 2009, p. 113). When collected together, disparate objects comply with 'contextual pressure', that is, they become related because they are found in the same place. No matter how dissimilar the items are, given that we can identify a 'criterion of assembly', lists are 'never incongruous' (Eco, 2009).

The catalog does not simply collect items, but it also republishes them, giving them a new connotation. An object that is part of a collection obtains a new renewed meaning, in dialogue with, alongside, in relation to, or competition with the other elements within the group (Veca, 2011). An example is 'Peripheral Landscapes: People, Gods, and Flora/Fauna' (Odell, 2015), where the artist collects decorative elements from the margins of old digitized maps. The aggregation in the same space brings to the foreground marginal items and gives them a unity that creates a dialogue. Similarly, in the 'Satellite Collections' (Odell, 2009), cut-outs of infrastructural elements (e.g., swimming pools, nuclear cooling towers, stadiums, basketball courts) are extracted from Google Earth and reorganized, decontextualized from their background, into compact aggregates. The contextual pressure provided by the visual catalog offers a point of view on a set of items that allows appreciating 'continuities and resonances' (Ahmed, 2017) across them.

This project merges two methodological perspectives (light diffraction and core sampling) with two visual formats (supercuts and visual catalogs). The result is an analytical procedure that strives to go beyond formal styles of measurement and attends to the specificity of video footage. The process is a mix of algorithmic and purely curatorial choices. In what follows, the process is outlined step-by-step and illustrated with images.
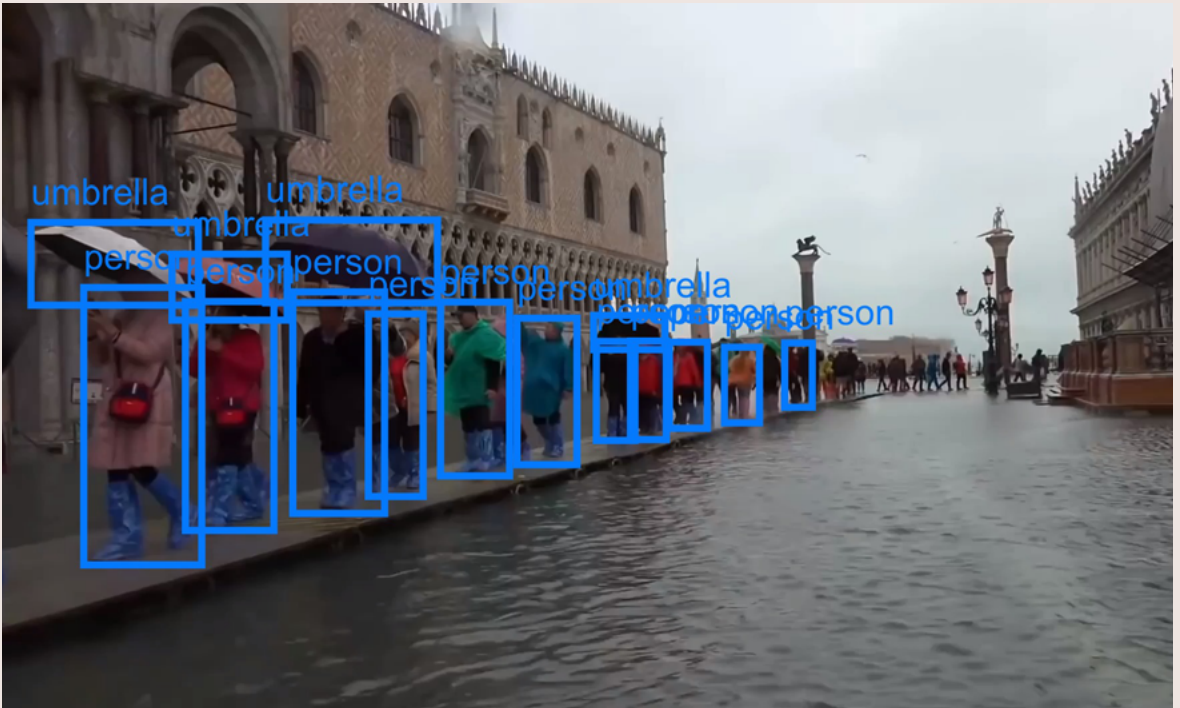
DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

Figure 2: Examples of the 25 YouTube videos used in the analysis. The curatorial selection of videos aimed at capturing a rich diversity of themes, topics, objects, and situations. Thus, the selected videos are not representative of the overall dataset but rather the results of a curatorial decision that privileged diversity over representativeness.

DATA-DRIVEN CURATED VIDEO CATALOGS: PROCESS STEP-BY-STEP

1. Data design. The process starts with a curatorial choice of videos from YouTube. The goal is to have a rich and diverse enough surface for the sampling phase. In the data selection, we aim at diversity in formats (from professional news reports to smartphone videos) and situations (Marres, 2020): tourists strolling in the water, home toilets overflowing with water, a man happily swimming in a flooded Piazza San Marco, boats adrift helplessly trying to reach the dock. The result is 25 videos, for a total running time of roughly 50 minutes (Figure 2).
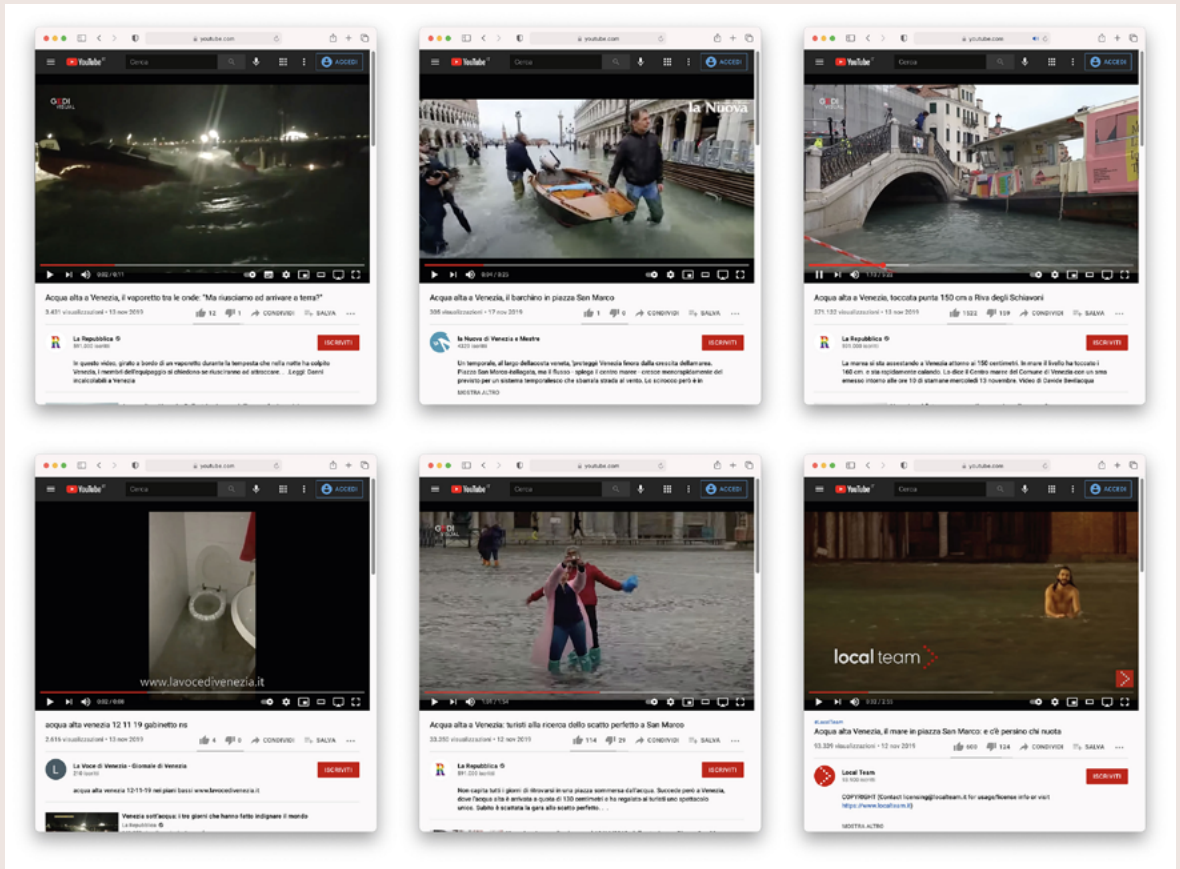
2. Object detection. Each video is run through an object detection algorithm (Bochkovskiy et al., 2020), which recognizes a predefined set of objects in the video footage and returns a cropped image for each object, labeled accordingly. Each frame is extracted, saved in a folder corresponding to a detected entity (e.g., person, car, handbag, boat, umbrella). The process is performed with the software RunwayML, a catalog of machine learning models, including image synthesis, object detection, and text generation (Figure 3).

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

Figure 3: The output of the object detection algorithm. The tool labels objects in each frame and draws a bounding box around them. In a subsequent step, bounding boxes are used to crop individual images for each labeled object.

3. Preparing the folders of images. As with most visual analyses, the procedure takes its point of departure from one or more folders of images (Colombo, 2019). In each video, a set of diverse objects is detected. We focus on most found categories across videos (i.e., categories found in all the input videos): person, boat, and bird. Cropped images for each video are then grouped in three folders, one for each category (Figure 4).
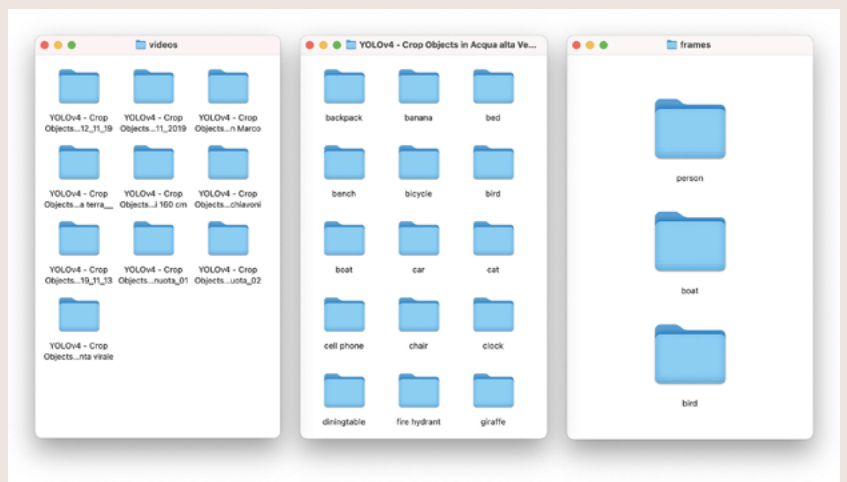
Figure 4: Creating folders of images. From left to right: the software outputs one folder per video; in each video folder, there are sub-folders for each detected object; the same objects detected in each video are grouped in a unique folder, one for each theme (person, boat, bird).

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

4. Removing smaller images. We then remove smaller images (below 10 kb) since, after an empirical assessment, they are the ones most likely to include incorrectly detected objects (Figure 5).
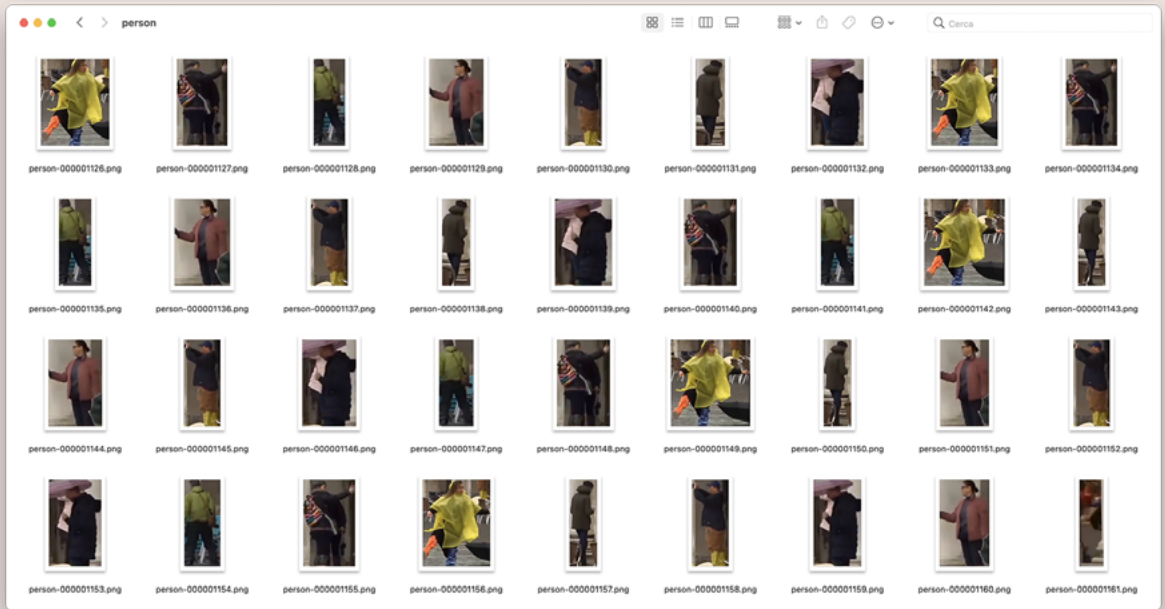
Figure 5: A folder of images with detected persons. In this step, images are collected in a folder and ordered by size. Small images are removed.

5. Measuring similarity across images. A similarity detection algorithm (Clarifai, 2017) is used to obtain a similarity score for each image in the collection, compared to a starting image. The technology is used for 'visual search' since one inputs an image and retrieves similar ones within a database. Here, it is used to rank images from 'more similar' to 'less similar' (compared to a starting image) (Figure 6). The starting image is the one that the object detection algorithm has labeled with the highest confidence score, chosen to represent each category (boat, person, bird).

6. Ordering images. Each image (within a thematic folder) is renumbered based on its similarity score (from high values to low ones).

7. Video Editing. For each category (person, boat, bird), images are loaded into a video editing software (such as Adobe After Effects), one image per frame, in order of similarity. The composition is stretched in time and remapped, and the software interpolates each neighboring frame with newly generated images. A morphing effect is added to smooth the transition from one frame to the other. In addition, the morphing effect mimics the floating movement of the water, and it is used as a rhetorical device to create a feeling of instability and a growing sense of anxiety. One may choose another transition technique without
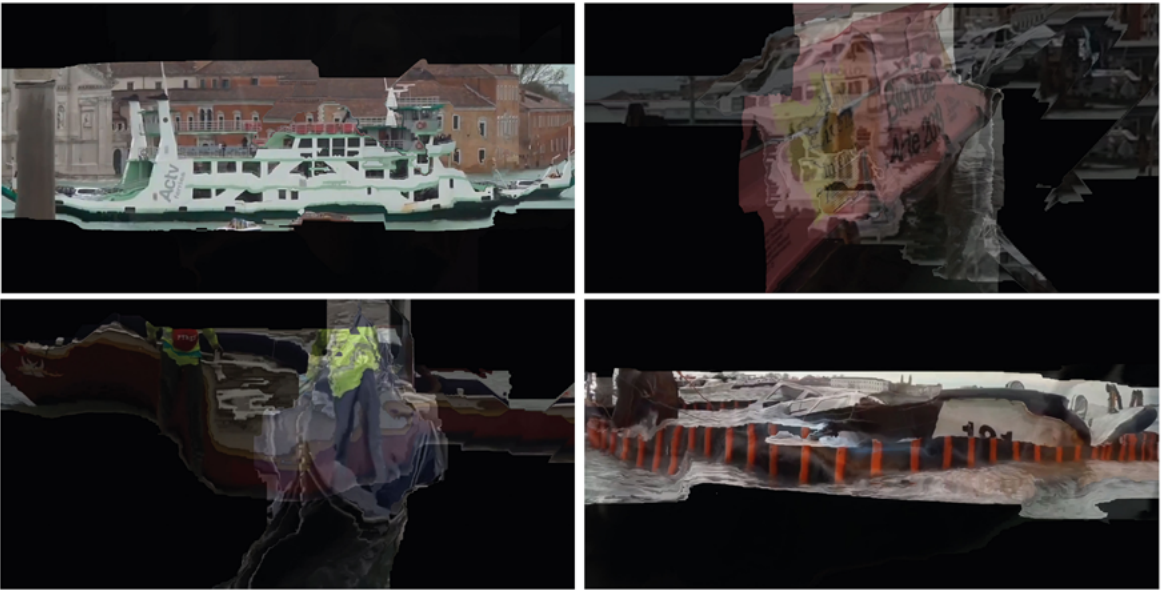
DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

Figure 6: A bird-eye view of image similarity in our image collection. To overview the full dataset before measuring image similarity, we employed the software ImageSorter (https://visual-computing. com/project/imagesorter/) to organize images according to their color. The view clusters similar images together, allowing the distant reading (Moretti, 2013) of the collection.

1 The project is part of the 'Aquagranda in Venice' online exhibition (www.aquagrandain-venice.it/). The video work, titled 'The Old Is Dying and the New Can't Be Born', can be accessed at www.aquagrandainvenice.it/ nuxt/artworks/the-old-is-dying/ content. The authors of the project are Federica Bardelli, Gabriele Colombo, and Marc Tuters.

affecting the analytical purpose of the video. The same settings are used for the three videos, regardless of their initial length (Figure 7).
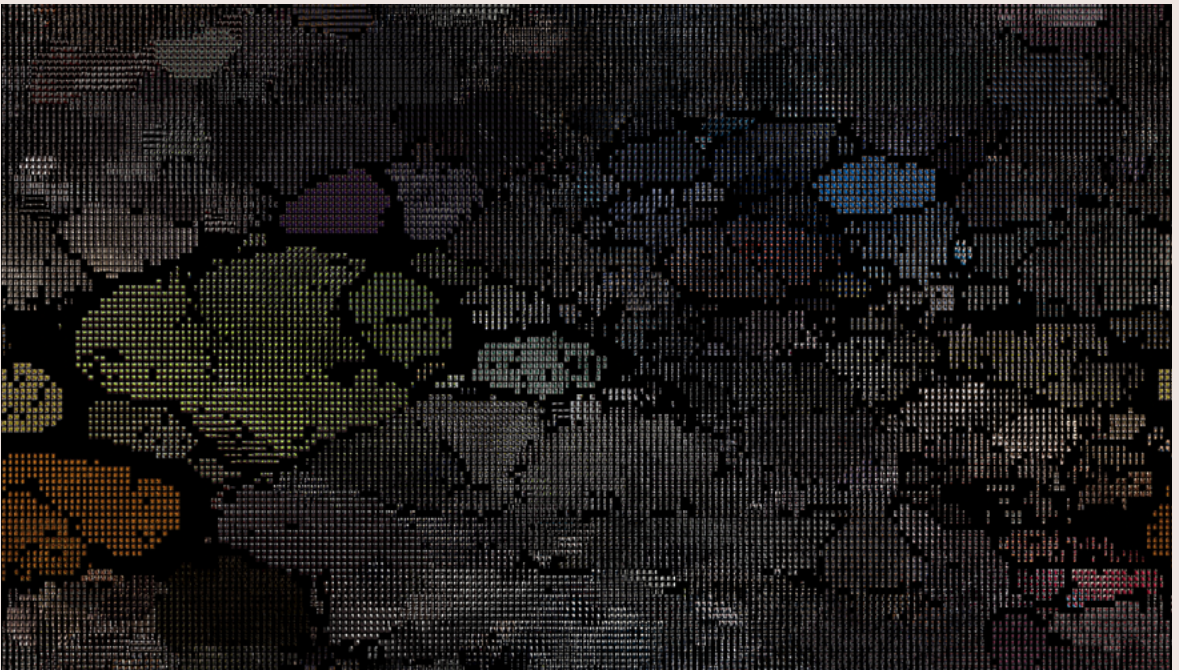
8. Sound editing. The audio design mimics the process followed for video editing. Audio tracks are extracted from the initial footage and then added to the three video catalogs, overlapped, stretched to the length of the video, and modified with reverberation.

9. Staging. The three video catalogs (person, boat, bird) are displayed online[1] in a real-time mode (mocking live webcams): always on, no controls, albeit one can smoothly switch across the three videos. The videos allow for a more cinematic form of visual analysis, going beyond measurements and analytical observations. The user can sit back and watch the images repeating one after another in a type of fruition that recalls the "filmic ways of seeing images" (Sze, 2018) of today's visual culture. Compared to a 'bird-eye' view of a collection of frames (see Figure 8), which highlights the detection of patterns and quantities by scale, the video catalog promotes a more granular and narrative-driven type of analysis.

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

Figure 7: Frames of the final video catalogs showing the morphing effect used to smooth the transition between images.

Figure 8: A bird-eye view of all frames of the category 'person', clustered by color similarity with the software ImageSorter (https://visuual-computing.com/project/imagesorter/).

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

COMPUTER VISION AND VIDEO FOOTAGE: LIMITATIONS

Above, we discussed the data-driven video catalog as an approach for the analytical republishing of video footage. Having rich and diverse collections of videos as input, data-driven catalogs strive to identify common objects and reorganize them into thematic clusters, displayed in a video format. It is a form of analytical display that aims going beyond traditional measuring approaches and relies on selection and repetition to foreground unnoticed patterns. It builds on two methodological inspirations (core sampling and light diffraction), and it repurposes two visual formats (supercuts and visual catalogs). The technique makes use of a mix of curatorial and algorithmic processes, particularly computer vision.

As a way of conclusion, it is relevant to glide over two methodological limitations connected to object detection and visual similarity analysis. They are of interest in general for human-machine co-creation processes, particularly for the visual analysis of large collections of videos. Firstly, detecting objects with a pre-trained algorithm results in imprecise results, mainly, when dealing with small elements (in terms of pixels). We found several errors in the dataset, such as persons tagged as elephants, bears, or backpacks, notably in video footage with low resolution. A way forward to overcome this limitation may have to entail training a custom object detection algorithm tailored for the footage at hand.

Secondly, when sorting images by similarity, since the visual search approach outputs a similarity score contrasted to one image, it forces linearity onto a multidimensional space and flattens the results. Algorithms for image analysis usually output high-dimension data, most commonly a vector of 4096 dimensions (Krizhevsky et al., 2017), which is then reduced into a two-dimensional space with dimensionality reduction algorithms (McInnes et al., 2020; Van der Maaten & Hinton, 2008). The result is a statistical approximation and cluster visualization where images closer to each other are more likely to have similar characteristics. Forcing a linear path crossing this space, although it serves well video output, which is linear, does not account for the complexity of the data. A possible procedure could be to measure the shortest path between two images, rather than starting from a single one. These two limitations reiterate the need for further research into human-machine collaborations to analyze collections of visual materials. ⊡

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

REFERENCES

AHMED, S. (2017, May 1). *The Effort to Transform: Intellectual Legacies of Stuart Hall*. Feministkilljoys. https://feministkilljoys.com/2017/05/01/the-effort-to-transform-intellectual-legacies-of-stuart-hall/

BOCHKOVSKIY, A., WANG, C.-Y., & LIAO, H.-Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *Association for Computing Machinery ArXiv*, arXiv:2004.10934. https://arxiv.org/abs/2004.10934

CLARIFAI. (2017, October 12). *Search Images by Visual Similarity with the Clarifai API*. https://www.clarifai.com/blog/search-images-by-visual-similarity-with-the-clarifai-api

COLOMBO, G. (2019). Studying Digital Images in Groups: The Folder of Images. In L. Rampino & I. Mariani (Eds.), *Advancements in Design Research: 11 PhD Theses on Design as We Do in Polimi* (pp. 185–195). Franco Angeli Open Access.

DIFFRACTION GRATING. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Diffraction_grating&oldid=1036649242

ECO, U. (2009). *The Infinity of Lists: An Illustrated Essay* (A. McEwen, Trans.). Rizzoli.

HANRAHAN, H. (2011). *It's showtime!* [Video]. https://www.youtube.com/watch?v=LUapZhcsdx8

JUZWIAK, R. (2008). *I'm Not Here to Make Friends!* [Video]. https://www.youtube.com/watch?v=w536Alnon24

KRIZHEVSKY, A., SUTSKEVER, I., & HINTON, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM, 60*(6), 84–90. https://doi.org/10.1145/3065386

MAGGOT3560. (2006). *Scarface F Word* [Video]. https://www.youtube.com/watch?v=Ft07LYRjFwM

MARRES, N. (2020). For a Situational Analytics: An Interpretative Methodology for the Study of Situations in Computational Settings. *Big Data & Society, 7*(2), 2053951720949571. https://doi.org/10.1177/2053951720949571

MCINNES, L., HEALY, J., & MELVILLE, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *Association for Computing Machinery ArXiv*, arXiv:1802.03426v3. https://arxiv.org/abs/1802.03426v3

NOBLE, O. (2010). *Every Arnold Scream from Every Arnold Movie* [Video]. https://www.youtube.com/watch?v=5aLR-8c11ms

ODELL, J. (2009). *Satellite Collections*. https://www.jennyodell.com/satellite.html

ODELL, J. (2015). *Peripheral Landscapes: People, Gods, and Flora/Fauna*. https://www.jennyodell.com/peripheral-landscapes.html

RAFTERY, B. (2018, August 30). I'm Not Here to Make Friends: The Rise and Fall of the Supercut Video. *Wired*. https://www.wired.com/story/supercut-video-rise-and-fall/

SZE. (2018). Sarah Sze: Infinite Generation [Interview]. (L. Neri, Interviewer). *Gagosian Quarterly*, (Winter). https://gagosian.com/quarterly/2019/10/08/interview-sarah-sze-infinite-generation/

TRECCANI ENCICLOPEDIA. (n.d.). Carotaggio. In *Treccani Enciclopedia online*. Istituto della Enciclopedia Italiana fondata da Giovanni Treccani. Retrieved November 7, 2020, from https://www.treccani.it/enciclopedia/carotaggio

VAN DER MAATEN, L., & HINTON, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research, 9*(86), 2579–2605.

VECA, A. (2011). Figure di argomentazione. In V. Bucchetti (Ed.), *Altre figure. Intorno alle figure di argomentazione* (pp. 17–61). Franco Angeli.

DISEÑA 19
AUG 2021
ARTICLE.4

GABRIELE COLOMBO
FEDERICA BARDELLI

DATA-DRIVEN CURATED VIDEO CATALOGS: REPUBLISHING VIDEO FOOTAGE

VICE NEWS. (2018). *Trump Talk: All Our Best Mashups in One Video* [Video]. https://www.youtube.com/watch?v=2i4JxWkSYzU

ZABRISKIE, C. (2008). *Previously on Lost: What?* [Video]. https://www.youtube.com/watch?v=GcatQSYRK6c