xCoAx 2020
Computation,
Communication,
Aesthetics & X
2020.xCoAx.org
Graz, Austria

Collaborative Vision: Livestream, Volumetric Navigation, AI Image Processing, and Algorithmic Personalisation

Provides Ng
provides.ism@gmail.com
Strelka Institute of Media, Architecture, and Design, Moscow, Russia

Keywords: Collaborative Vision, Livestream, Volumetric Navigation, AI Image Processing, Algorithmic Personalisation.

'Vision' is the faculty of perception and visualisation that helps us to acquire information about our environment as an individual. The media industry has shown how the collection and distribution of such information can impact the public realm. For instance, the 2019 Extradition Protest in Hong Kong, which propagated itself through livestreaming events on media platforms, brought about the compliance of the government by collectively disrupting physical-urban and virtual-media spaces. These media technologies have demonstrated their capacity in facilitating a 'collaborative vision' that communicates and accumulates individuated information in real-time. This gives the urgency to understand the working of these distributed media and their potential in formulating a cohesive infrastructure that will help us in reconstructing and understanding the consequences of our own events collectively. This paper summarises a research—*Current*—which utilised 4 techno-social ideas to prototype a means of 'collaborative vision'.

1. Introduction

Our vision has changed radically since the invention of machines that are able to comprehend large datasets. For instance, the idea of 'climate change' will not be present without the invention of computers: not only because of the immense use of resources necessary to propel large scale computation, but also because computers helped us in perceiving the presence of climate change through the mathematical interpretation of data. 'Any change in the technology of the description of space has always changed the way we encounter and intervene with space' (Bottazzi 2020). The democratisation of such technologies decentralises knowledge to anthropogenic phenomena. Although large scale computation used to be instruments exclusive to institutions, the advancement of media technologies facilitates a means for individuals to gather and process data collectively—a 'collaborative vision'.

The research project *Current* is a 12-minutes prototype of 'collaborative vision' (fig. 9). *Current* speculates on the convergence of four core ideas into a media infrastructure that spans across cities and borders: livestream, volumetric navigation, AI image processing, and algorithmic personalisation. *Current* experimented with a range of digital technologies that are readily available to any individuals (e.g. livestream data, machine learning, 3D environment reconstruction, ubiquitous computing, point-clouds, etc.). It developed a production pipeline (fig. 8) using distributed technologies, which provide a means for individuals to reconstruct, navigate, and understand event landscapes that are often hidden from us, such as violence in protests, the handling of trash, changes in nordic animal behaviours, etc.. This paper tabulates the ideas and technologies explored in *Current*.

This paper first presents the four core ideas, and discusses the technologies used in examining these ideas. It then introduces how these ideas and technologies can amalgamate into a production pipeline to facilitate a 'collaborative vision'. Subsequently, this paper reflects on its findings in the discussion section. Finally, it concludes by discussing the types of data that has been created, the limitations of the system, and how this might open up new spaces of discussion.

2. The Four Core Ideas

2.1. Livestream

In the contemporary contestations of algorithmically recommended content, the screen time of scrolling between livestreams has become a new form of television. The word 'television' comes from Greek τῆλε *tèle*, meaning 'far', and Latin *visio*, meaning 'vision'; thus, it is a communication device that transmits vision from afar. Livestream identifies itself from traditional television as a distributed real-time media—anyone with a digital camera and access to the internet can produce and broadcast content simultaneously.

This gives rise to an attention economy that circulates values distinct from traditional moving-image media. First, it encompasses compelling moments alongside an infinite feed of the mundane, suggesting a sense of 'truth'—an event being preserved in its entirety—to its viewers. Second, it's arbitrary timespan allows viewers to be in control—viewers can step in and out of the stream at any moment instead of having to sit in for a standardised amount of time. Third, it facilitates a participatory authorship, where the interaction between the viewer and the streamer collaboratively directs, narrates, and curates the experience.

With more than 300 million users, the livestream media owns a virtual population the same size as the third largest country in the world (Cunningham 2019). Livestream platforms stimulate social participation by bringing vision into individuals' lives to attract both capital and data. It is a form of open-source database that represents a globalised demographic of the contemporary working class, who would regularly spend their mornings, lunch breaks, and after-work hours interacting in this virtual world as part of their daily routine—it is a collective effort in the characterising of socio-economic trends. As such, livestream is a cultural emergence to the contemporary urban demographic shift: from the domination of the modernistic nuclear family who live in the suburbs, to the domination of a mobile population of single professionals who work between metropolises. In the modernistic nuclear household, the television, which gravitates the space of the living room, represents the building block (a single modular unit) of the society. In comparison, livestream, which is the friendly companion of your studio apartment where shared screens are no longer desired except in the virtual space, represents a network of connected individuals—it is the market's response to the contemporary reinforcement of individualism.

Furthermore, livestream interfaces are generally designed with interactive components that extend its revenue model beyond television commercials: virtual chat rooms to promote engagement, virtual gifts and 'red pockets' to attract credit top-up, e-vouchers to nexus sponsors, loyalty ranking to embed competition, etc. (fig. 1). In 2016, the livestream economy in China alone surged a 20 billion Yuan revenue, which overcame Hollywood as one of the largest entertainment industries (Cunningham 2019). From the production to the consumption of its content, livestream is a distributed technology that diminishes a centralised operational cost. Its economy is based on a peculiar form of consumerism—the consumption of attention. By this very nature, branded virtual signalling becomes the contemporary currency that is based on the atomic unit of data: bits.

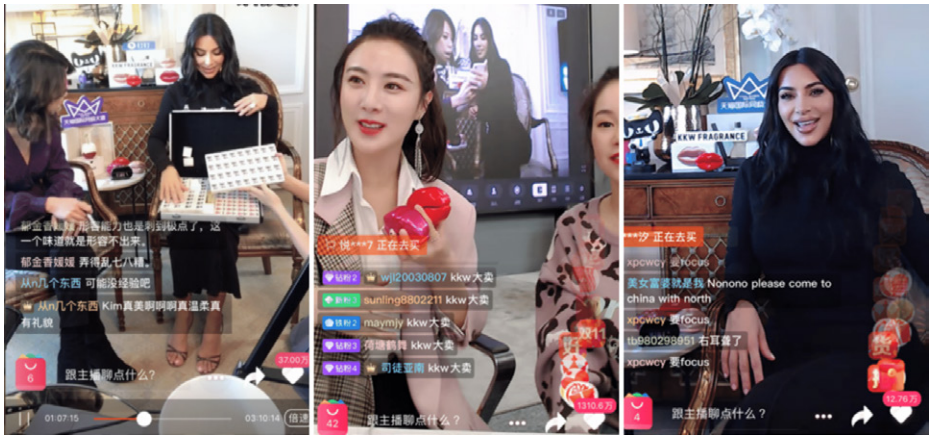

Fig. 1. Viya Huang and Kim Kardashian collaborating on livestream.
source | *Tmall.*

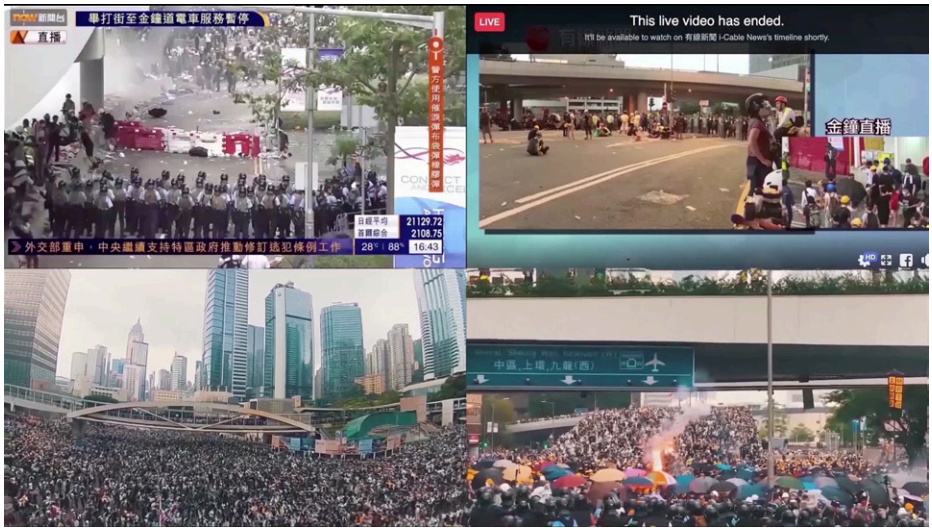

Fig. 2. Underground livestream released by protesters during the 'Hong Kong Extradition Protests' in 2019.
source | *Github.*

In our digital age, where every image can be readily copied and altered, the sense of 'truth' becomes problematised; the value of 'truth' becomes our currency. The virtual signalling of livestreams accomplishes a socially defined realism, albeit this may not always be a conscious thought process, it influences our collective intuition and reasoning. It is not only accomplished through the content that is being streamed, but also its visual cues. 'In principle, any image property can be a cue, including colour, texture, local parts, overall shape, as well as learned features' (Hertzmann 2019). Visual cues contribute to whether the vision being presented 'seems real' by providing supplementary information to its viewers. For instance, although most livestreams are quite shaky and pixelated, they imply 'real' people filming 'real' events; alternatively, if a footage is in static motion with high resolution, it signals a sense of 'curation' or 'being staged' from a well-resourced entity. As such, livestream is not only an open-sourced database, but also a form of expanded cinema, where the 'back of stage' is perceptible to its audience—its graphical aesthetic has intrinsic social value in our present-day technocracy. *Current* expanded on the definition of livestream beyond social media to animal cams, autonomous car visions, etc.. These sources represent the visions of alternative intelligence. In this respect, livestream may provide a prospective means to quantify, predict, and generate a 'collaborative vision'.

2.2. Volumetric Navigation

Volumetric navigation (VN) is a technique developed for large-scale traffic control systems, in which 'all the vehicles share information in order to be part of a collaborative navigation network' (Olmedo 2012). VN synthesises multiple sources of data, which includes positioning, navigation, volume, and time information, to simulation a subject(s) and its surrounding environment with the intention of deducing its plausible relative movements. VN is a collaborative context-reconstruction that gives multidimensional vision to an immediate moment for preemptive actions. The idea of 'volumetric' problematises the contestation between different perspectives in forming a coherent 'collaborative vision'; conversely, this gives the potential to authenticate events and truth. In 'Current', this is where livestream and VN meets: viewers will be able to volumetrically navigate events in real-time. *Current* experimented with extracting 3D information with photogrammetry frameworks from different livestream sources, including autonomous car vision, NASA, drones, animal cams, and surveillance cameras. These sources are selected specifically to examine and demonstrate the idiosyncrasies generated by different visual cues, which will reveal to its viewers how data is being collected, structured, and projected volumetrically - volumetric data analytics.

Autonomous car vision are often composed of dolly shots, which are more effective for image-matching procedural processing to deduce the depth of space by comparing consecutive frames, but it will only be able to reconstruct spaces that are perpendicular and immediate to the camera's movement. NASA often livestreams its expeditions from omnidirectional (360°) cameras mounted on mobile robots (e.g. the Mars 2020 Rover), which produces truck shots. This yields spherical interpretations of the robots' surroundings, yet often lacks information on what is behind other objects, generating a peculiar aesthetic of voids and shadows around the scene that discloses information of the relative position of the camera (fig. 3).

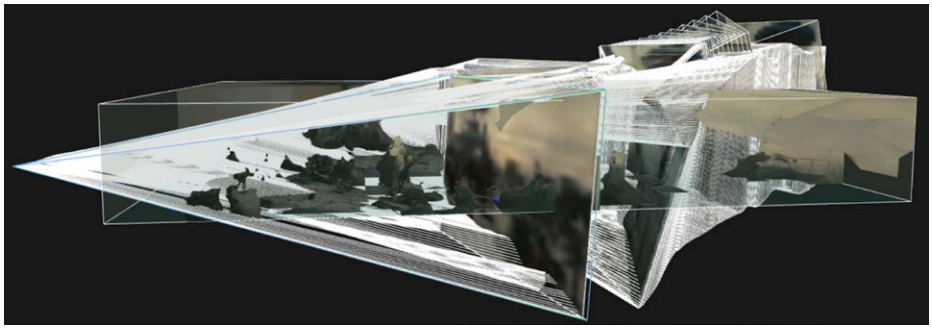

Fig. 3. Volumetric reconstruction that shows the peculiar aesthetic of voids and shadows around the scene, which discloses information of the relative position of the camera.
source | author.

This 'bug' can find its potential application in visual odometry, especially on foreign planets. Drones that were flying in circular direction produces relatively coherent reconstruction of single-body targets from aerial perspectives; while mono-directional flying produces quadrilateral scenes where objects are texture mapped all on one side (fig. 4).

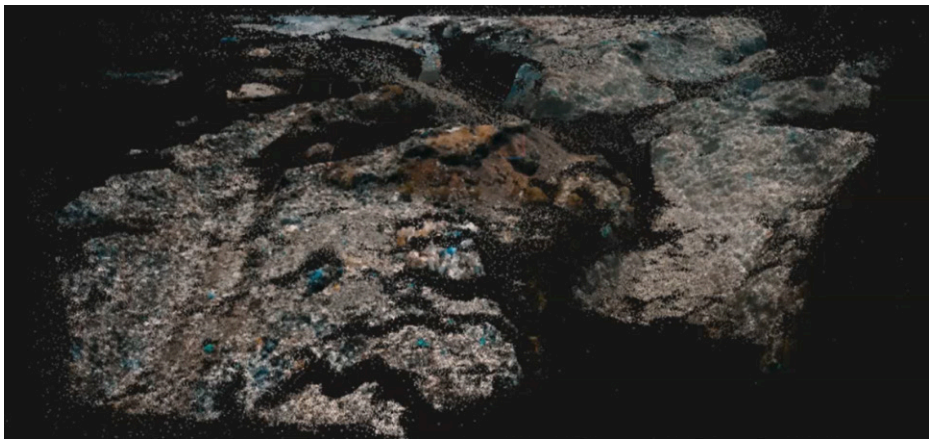

Fig. 4. Volumetric reconstruction from a drone flying from west to east over a landfill, resulting in texture mapping only on west-facing surfaces.
source | author.

This has potential application in forensic sciences by volumetrically reconstructing the path of subject(s) within a space at a given amount of time. These techniques, together with AI image processing, can help to democratise volumetric data analytics.

2.3. AI Image Processing

Today, many leading digital enterprises (e.g. Nvidia, Intel, etc.), which recognised the profound relationship between vision and intelligence, feed AI with images to conceive machine vision. *Current* experimented with two AI image processing neural networks (NN)—Autoencoders and Generative Adversarial Networks (GANs)—that are able to infer through self-organisational pair-work. In pedagogy, pair-work is 'learners working in pairs'; this allows learners to 'compare answers and clarify problems together' (British Council n.d.). In Autoencoders, an NN learns representations from high dimensional datasets to encode information, while the other NN learns to reconstruct the complexity of the dataset by decoding the representation. The two NN 'compare answers' with each other in an iterative manner. In GANs, the two NN compete against each other: one NN tries to generate synthetic data from input, while the other NN tries to identify the 'fake' from the 'real'. The competition continues until both NN have learnt to do their best in the game.

Autoencoders have been utilised in areas such as retrieving 3D information from a single 2D image, which is one of the earliest goals of AI research: mimicking human visual systems to achieve a full scene understanding (Papert 1966). On the other hand, Autoencoders have contributed to the social phenomena 'DeepFakes', where the NNs learn to swap human faces (e.g. put celebrities in pornographies, etc.). *Current* demonstrates both the bright and dark side of this technology, which influences the course of reality through synthesising images. In one scene, *Current* tried using Autoencoders to fill in voids between discrete data to present viewers with a more coherent environment reconstruction. In another scene, Autoencoders were used to swap the faces of Donald Trump and Chairman Xi in their

national speeches to demonstrate how these images may impact the public realm when streamed (fig. 5).

GANs have been utilised in areas such as enhancing astronomical images and video game modding (Kincade 2019; Tang et al. 2018). *Current* experimented with GANs on two aspects of the project: up-scaling low-resolution textures and morphing data into a single viewing unit (fig. 6). Morphing, which is a digital aesthetic native to the iterative nature of AI image processing, 'adjusts shape but not colour or texture' of the data (Hertzmann 2019). In the process, GANs compose 'in-between images' that look strangely familiar to human vision, yet we cannot quite 'recognise them as anything real' (Hertzmann 2019). These 'in-between images' are a condensation of visual cues inferred by the AI, which is constrained by its inductive bias when it negotiates itself between the image topologies from one input to another. *Current* took livestreams from multiple perspectives on the same events and 'morphed' them into a 'collaborative vision' that tries to demonstrate a more encyclopaedic representation to the scenarios.

Fig. 5. A scene from *Current* where the US president Donald Trump is saying Chairman Xi's Chinese national speech. *source* | *author.*

Fig. 6. A scene from *Current* of a transitional morphing image approximating itself between rocks and trash. *source* | *author.*

2.4. Algorithmic Personalisation

Algorithmic personalisation (AP) is 'a process of gathering, storing, and analysing information' by recommendation systems (Venugopal 2009). It 'has become a standard approach to tackle the information overload problem [...as] we are still bounded by cognitive and temporal constraints.' (Perra, 2019). It is a response to the discrepancies between the amount of data different agents can process at a given point in time. Theoretically, machines can process an infinite amount of data; whereas humans often perform reduction in order to reason. The public, as a form of collective intelligence in our technocratic epoch, reach consensus depending on how many is being exposed to which content in a given period of time. Thus, AP is a communication strategy that translates between actors, and impacts the public realm by recommending the 'right' content, to the 'right' person, at the 'right' time.

AP influence our vision by presenting us with adjoining concepts. It predicts user preferences based on their viewing history, and generates an attention economy that regulates our civic lives based on data consumption. Today, AP are often operated by AI, which automate our aesthetic production and 'other cultural experiences (e.g. automatically selecting ads we see online)' (Manovich 2018). This form of AP is often employed by media platforms as part of their revenue model. For instance, Instagram offers a communal virtual space where all information is shared for free, and their profit is generated from commercial advertising through AP. As such, users are at once, both consumers and commodities. Media platforms offer us information services, the service fee is being priced not on fiat currency, but data. In capitalism, everything is quantifiable by their value and comes at a price, which ensure tradability and the circulation of value within the system. In our climate change epoch, instead of simply denouncing the data market, it may be more constructive to consider what are the kinds of data market we need to mitigate risks collectively.

Current tried to operate on a data market, where users are able to understand and operate the same set of AP tools as commercial suppliers. *Current* used recommendation systems to personalise livestream data that fits the research project. It used keyword labels (e.g. 'climate change', 'social issues', 'animal cams', etc.) to set up digital profiles with various open-sourced recommendation systems to retrieve the appropriate livestream contents amongst billions of hours of footages. The open-source quality problematises the empirical ideology of authorship in design that is entitled to a single name; at the same time, questions if AP helps us to be exposed to more choices or less.

Fig. 7. A scene from *Current* using volumetric techniques to personalise the virtual environment.

3. Result: Production Pipeline

From the experiments of *Current*, this paper formulates a production pipeline where the examined ideas and technologies come together as a media infrastructure that allows individuals to collectively gather and process data to reconstruct and understand the consequences of our own events —'collaborative vision'.

This pipeline (fig. 8) begins with individual users who produce and acquire data through livestream platforms, such as Facebook and Inke. The livestream data encompasses image data and metadata (e.g. GPS, timestamps, etc.), which can be stored and archived on any personal devices, and be extracted to AI image processing NNs for data enhancement. The learning algorithms, Autoencoders, can help to fill voids in discrete data based on the collective archived data. While GANs can assist in morphing data into a single viewing unit, and becomes an endless stream of content that can be viewed by users as requested. If the user prefers to navigate the event in 3D, the enhanced data can be delivered to volumetric reconstruction frameworks (e.g. RealityCapture, etc.) as needed. The output volumetric data can then be archived and plugged into personalisation algorithms, which would label and classify the data and deliver the recommended content retrieval to volumetric navigation engines, such as VR interface and devices, and accessed by users via media platforms. The accumulated data will be retrieved on the media platform through functions like keywords input, which helps to characterise trends collectively through a ranking system. Data on same or similar events can be used as comparative analysis or complete the void in each other's models to authenticate events. This completes the feedback loop, where the user produces and acquire further information through collaboratively characterised trends.

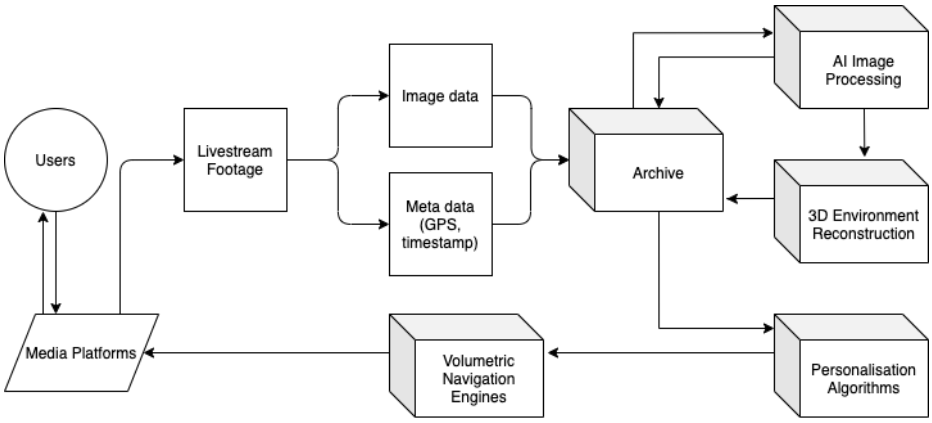

Fig. 8. Flow chart of the production pipeline with constituent components that generated a 12-minutes prototype of the concept 'collaborative vision' — *Current*.

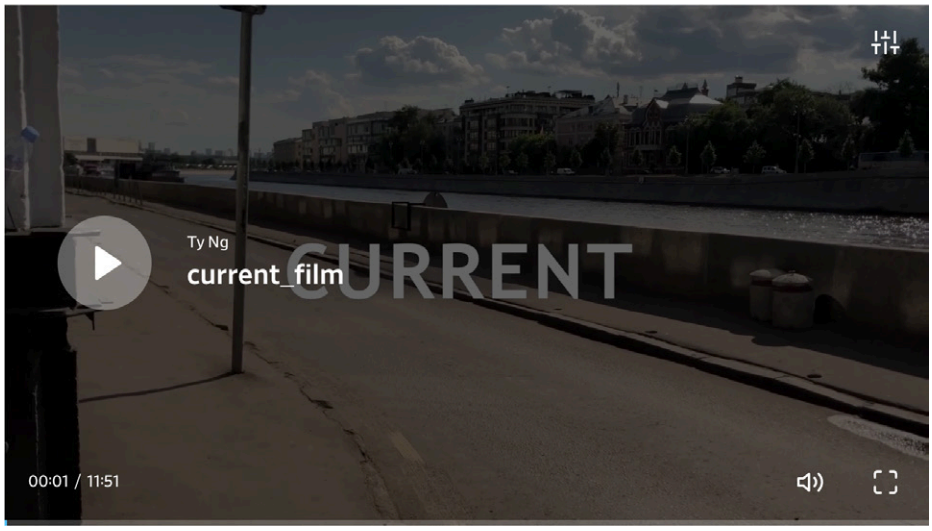

Fig. 9. A video asset: https://dai.ly/x7g1v7k password: 00000

4. Discussion

This research began as a design project that speculated on distributed technologies and slowly evolved into a 12-min prototype of volumetric media (fig. 9); which is why it has heavy reliance on media platforms. Just as Virilio (2008) indicated: the invention of any technology is also the invention of the relative accidents. Media platforms have brought convenience to the communal sharing of information, but have also brought with them the inherent accidents: ecological costs, exacerbating feedback, and data footprints.

As stated in the beginning of this paper, computation demands immense consumption of energy, albeit without which we will not be able to comprehend and communicate the changes in our climate. In the face of this problem, as opposed to simply denying computation, it is perhaps a more useful question to ask: what are the forms of computing we need in this climate emergency? This paper tries to demonstrate that distributed media, such as livestream, helps to diminish centralised operational monetary costs and has the potential to establish an information economy where individual's computing power can be put to uses beyond entertainment (e.g. volumetric data analysis). The author reflects that a technological solution is not enough to reform our present information economy, in which costs

are priced on dollars instead of carbon. The author has come to realise that any technological solution will have to be coupled with pertinent socio-economical solutions, to which this paper regrettably does not conform to cover. The reviewers of this paper have accurately commented: 'In short, the four ideas used as a strategy in the presented research are highly novice to the environment... but nowadays it seems more difficult to innocuous practices.'

As early as 1995, Nicolas Negroponte had expressed his concerns on personalised media in his book 'Being Digital'. He feared that too much positive feedback in a single direction will lead to polarisation on public's opinion and reinforces individualism. Exacerbating feedback may occur and amplify the effects of a small disturbance that leads to perturbation. 'That is, A produces more of B which in turn produces more of A' (Keesing 1981). It is worthwhile to discuss how algorithmic personalisation can become a hybrid system, and combine social, ecological, cultural, economic, and other relevant data.

Although digital communication facilitates a 'collaborative vision' to the consequences of our actions, it has also produced data footprints that have led to the inevitable discussions of surveillance. It is easy to demonstrate that digital technologies strengthen autocracy, but it takes immense effort to extract the utility of surveillance. With the present pandemic, the debate on surveillance has been opened up in more important ways. For instance, 'reconstructing infection by tracking phones can be an important tool, despite the direct confrontation with principles of libertarian anonymity' (Bratton 2020). It is perhaps more useful to ask: how are we mining data? Are we mining the right kind of data? What data should algorithmic personalisation be recommending to us? These questions are expected in the discussion of this project.

As a final point of discussion, I couldn't have put it in better words than Benjamin Bratton: 'it is a mistake to reflexively interpret all forms of sensing and modelling as "surveillance" and all forms of active governance as "social control." We need a different and more nuanced vocabulary' (Bratton 2020).

Fig. 10. Scenes of volumetric reconstruction in *Current*.

5. Conclusion

This paper is an after thought to the project *Current* (fig. 9), which focuses on delineating the slippery relationship between media and urbanism. 'Current' is a project that tries to inverse the negative value metrics of media surveillance into a utilisable communication device that spans across cities and borders—a media infrastructure that facilitates a 'collaborative vision'. *Current* goes beyond the collection of data to the structuring of it: it appropriated livestreams as competing datasets to reconstruct our event landscapes (e.g. where did the wastes go? Who exercised violence during protests?), and used a combination of techniques to facilitate a collective processing of data: volumetric navigation, AI image processing, and algorithmic personalisation. These experiments present us a method of statistical reasoning to urban phenomena, which is readily becoming our contemporary realism.

This paper contrasted the modernists' with the contemporary household to emphasise how livestream is transforming our domestic routines: from the television that keeps you awake in the middle of the night so you can watch your favourite show with your family after a standard 9-to-5 working hours, to livestream platforms, which collapse all your favourite content into an infinite stream so you can enjoy it anytime by yourself during your precarious employment. Livestream reflects our existing socio-economic structure: from having the nuclear family as the atomic unit that represents societal preferences, to a generation that is based on a compilation of many 'self'.

This paper then discussed how we are reconfiguring our vision from 2D to 3D, which means adding a z-depth value to our images. We are adding information to the image world faster than we can appropriate. The aesthetic, values, and validity of this extra information, this extra dimension, foster the relative measurements to our images in the digital age. For instance, traditional cartography performs reduction on Earth's spatial information to achieve efficiency for human perception; whereas the recent geoid simulation developed by NASA (fig. 8) presents high-dimensional information that is beyond human perception—our attempt to grasp such level of information influx can only be metaphorical: 'Earth looks like a potato'.

Subsequently, this paper reflected on its findings by discussing the future of our information economy, which based on mining the past. If our media infrastructure allows every moment in every corner to be stored, processed, and recommended to you according to your digital profile, what will be the value of history? History, from Latin '*historia*', means the art of narrating past accounts as stories. What will be the future of our urban environment if every single event is preserved and archived in real-time to such accuracy that there will be no room for his-story? Will we be submerged in an absolute evident-based society, where every subject is infinitely contested?

Acknowledgements: The author would like to give thanks to her dearest friends—Eli Joteva and Alexey Yansitov (Ya Nzi)—who have spent numerous sleepless nights with her to bring 'Current' to this world. The author would like to thanks Prof. Benjamin Bratton, Prof. Mario Carpo, and Prof. Frederic Migayrou, who's discussions I enjoyed the most and have consistently inspired me. The author would like to thanks the reviewers to this paper, who have instinctively advised on the weakness of this paper and left space for much reflections. The author would also like to thanks the following people for giving help to *Current* along the way: Strelka Institute, Metahaven, Liam Young, Nathan Su, Nikita Suslov, Atrem Konevskikh, Michael Villiers, Mary Anaskina, Elizaveta Dorrer, Nicolay Boyadjiev.

References

Bottazzi, R.
2020. Digital Architecture beyond Computers: Fragments of a Cultural History of Computational Design. London: Bloomsbury Visual Arts.

Bratton, B.,
2020. 18 Lessons of Quarantine Urbanism. Strelka Mag. Available at: https://strelkamag.com/en/article/18-lessons-from-quarantine-urbanism.

British Council.
(n.d.). Pair work. https://www.teachingenglish.org.uk/article/pair-work.

Cunningham, S., Craig, D., & Lv, J.
2019. China's livestreaming industry: platforms, politics, and precarity. International Journal of Cultural Studies, 22(6), 719–736. doi: 10.1177/1367877919834942

Hertzmann, A.
2019. Aesthetics of Neural Network Art. arXiv preprint arXiv:1903.05696.

Keesing, R.,
1981. Cultural anthropology: A contemporary perspective (2nd ed.) p.149. Sydney: Holt, Rinehard & Winston, Inc.

Kincade, K.,
2019. "CosmoGAN: Training a neural network to study dark matter". Phys.org.

Manovich, Lev.
2018. AI Aesthetics. Moskau: Strelka Press.

Negroponte, N.,
1995. Being digital, New York: Alfred A. Knopf.

Olmedo, R., Cuesta, C. D. L., Nemeth, J., Aichorn, K., Cabeceira, M. L., & Andres, N.
2012. ARIADNA: A Volumetric Navigation System implementation for maritime applications. 2012 6th ESA Workshop on Satellite Navigation Technologies (Navitec 2012) & European Workshop on GNSS Signals and Signal Processing. doi: 10.1109/navitec.2012.6423073

Papert, S.,
1966. "The Summer Vision Project". MIT AI Memos (1959–2004). hdl:1721.1/6125.

Perra, N., Rocha, L.E.C.
2019 Modelling opinion dynamics in the age of algorithmic personalisation. Sci Rep 9, 7261 .

Tang, X., Qiao, Y., Loy, C., Dong, C., Liu, Y., Gu, J., Wu, S., Yu, K., Wang, X.,
2018. "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks". arXiv:1809.00219.

Venugopal K.R., Srinivasa K.G., Patnaik L.M.
2009. Algorithms for Web Personalization. In: Soft Computing for Data Mining Applications. Studies in Computational Intelligence, vol 190. Springer, Berlin, Heidelberg

Virilio, P.,
2008. The Original Accident. Polity Press.