Calculating sameness: Identifying early-modern image reuse outside the black box

Florian Kräutli, Daan Lockhorst, and Matteo Valleriani
Max Planck Institute for the History of Science, Berlin, Germany

Correspondence:
Florian Kräutli, Max Planck Institute for the History of Science, Boltzmannstr. 22, 14195 Berlin, Germany.
E-mail:
fkraeutli@mpiwg-berlin.mpg.de

Abstract

In this article, we present a method for identifying image reuse in a corpus of 358 books printed between the 15th and 17th century. The approach is based on image hashing, an established method for finding near duplicates of images. Our historical interpretation of the method's result produces two important insights hinting at a radical material and epistemological change taking place around 1530. We then evaluate the image hash approach against a method that employs a neural network for image recognition.

1 Introduction

Within the Sphere project we explore the dissemination andtransformation of scientific knowledge across Europe based on the edition history of a singular text on cosmology: the *Tractatus de Sphaera* by Johannes de Sacrobosco. This 13th-century treatise describes the spheres of the universe according to the geocentric worldview. Up until the 17th century, it has been repeatedly published as part of university textbooks. In these, the treatise is included in original, commented, or translated form, and accompanied by other texts that were seen as relevant for the study of cosmology from disciplines such as medicine, astronomy, or mathematics (Valleriani, 2017). As many of these textbooks were part of the mandatory curriculum at European universities, we regard their contents as representative for the scientific knowledge that was being taught and seen as relevant at the time of publication of the books. We assembled a corpus of 358 books that contain or directly comment on the treatise, starting with the earliest printed edition published in 1472 up until 1650 when the relevance of the text declined rapidly. We extract several markers from the individual books that form the material evidence of our research. In addition to bibliographic data such as publishers, printers, date, and place of publication, etc., we identified for every book the content structure: which texts it contains and whether the texts are commented or translated versions of existing texts. In doing so, we cannot only identify how the content of the books changed and—by extension—how certain disciplines gained and lost importance, but also which publishers might be responsible for certain changes.

2 Visuals as Indicators of Scientific Evolution

In addition to the texts, the books in our corpus contain various types of visuals as follows: diagrams, illustrations, decorative elements, initials, printer marks, and frontispieces. In the same way as texts, these visuals can offer insights into the kind of knowledge that is being distributed. Many images reappear throughout the publication history of the corpus. By identifying and analysing recurring images, we can evaluate

doi:10.1093/llc/fqaa054

the 'success' of certain imagery. If we find similar images being used by different printers for the same subject, for example, this can be telling of one printer being influenced by another, or even indicate a physical exchange of woodblocks when the images are identical. In addition, we can identify when images are being replaced with new ones for the same subject. Producing woodblocks was a costly endeavour. The introduction of a new image therefore constitutes a significant and potentially informative change.

The reappearance of illustrations is a valid method to reconstruct not only the evolution of the visual language in science but also of the scientific content. Especially during the early modern period when the textual aspects of treatises were charged with heavy authority and therefore not easily amendable, the insertion of a new image represented an effective way to introduce novel scientific aspects. Tracing the use of scientific illustrations, moreover, does not show only the introduction of novel representations; it also allows to recognize which visual representation and visual language became obsolete over time, as specific kinds of illustrations were sometimes dismissed and replaced.

3 Method

We obtained for every book in our corpus a digitized copy in PDF format. A team of student assistants then manually annotated the visual elements on each page using the Mirador Viewer (Project Mirador, 2014). A total of 31,610 elements have been identified and classified as either Content Illustrations, Initials, Frontispieces, Printer's Marks, Title Page Illustrations, or Decorations. They are stored in RDF as annotations on the digitized pages of the books, along with the remaining metadata that we gather in the project and store according to a CIDOC-CRM data model in a Blazegraph triple store (Kräutli and Valleriani, 2018). For processing, the cropped regions containing the images are downloaded to a local machine via a IIIF API. We focus on the Content Illustrations, 21,229 in total and the majority of all visuals identified.

We seek to identify which of the illustrations appear several times in our corpus of books. In other words, we want to organize the total set of images into groups that are duplicates or near duplicates of each other. Duplicate and near-duplicate detection of images are often addressed problems (Ke *et al.*, 2004; Foo *et al.*, 2007), specifically for preventing upload of (known) image spam to social media platforms (Mehta et al., 2008).

The approach we use is an image hashing algorithm as proposed by Venkatesan et al. (2000). A hash function takes an arbitrary sized input and deterministically produces an output of a fixed size, the so-called 'digest'. For an introduction to hash functions, see Knuth (1998). In order to identify images that are not duplicates but variations of each other, a 'perceptual' image hashing algorithm is required (Zauner, 2010). It is designed to take an image as input and produce a digest that bears a deterministic relationship to the input image. We use the difference hash or dHash, algorithm (Kravetz, 2013) in an implementation for the Python programming language (Buchner, 2017). The algorithm works by scaling down and converting the input image to greyscale and produce a digest based on each pixel's difference in brightness to its neighbouring pixels. The similarity between two images can then be expressed as the difference—the Hamming distance (Hamming, 1950)—between two digests. We regard images as near duplicates if the difference between their digests is below a certain threshold and cluster the images into groups by assuming transitivity.[1] This arguably simple method works surprisingly well for our images, resulting in 66% of the images being assigned to a group. To evaluate the performance of this method, we compare it with an alternative approach employing a deep neural network. The reason we cannot evaluate the

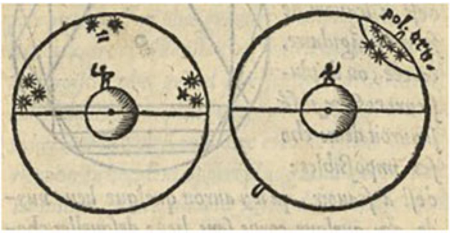


Fig. 1 Illustration appearing in a 1546 edition published by Jean Loys in Paris. Image: Biblioteca Nacional de España, CC-BY-NC-SA. Available at http://bdh-rd.bne.es/viewer.vm? id=0000000888&page=13. Database record: hdl.handle.net/21.11103/sphaera.101030

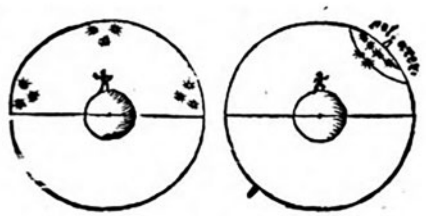


Fig. 2 Illustration appearing in a 1563 edition published by Hans Lufft in Wittenberg. Image: Bavarian State Library, NoC-NC. Available at https://reader.digitale-sammlungen.de/de/fs1/object/display/bsb11109959_00073.html. Database record: hdl.handle.net/21.11103/sphaera.100820

method by calculating an error rate is because we lack a ground truth. Although we could try to arrive at one by manually cleaning and grouping the algorithm's output, obtaining a ground truth is not a trivial endeavour in this context. Consider, for example, the two illustrations in Figs 1 and 2 that have been grouped together by the ImageHash method. Although the illustrations are evidently similar, they are not identical (most visibly in the posture of the small figures). Whether the difference between these two illustrations is significant or not depends not on the image itself but on the image's meaning in the context of the book, the research question and the specific viewpoint of a historian.

Evaluating our method against one based on a deep neural network also gives us an indication whether a more 'sophisticated' method of image analysis would yield better results. Since 2012 when the first application of a large convolutional neural network outperformed all other available methods at that time, the approach has become the de facto standard in most computer vision tasks (Krizhevsky et al., 2012). We employ a pretrained MobileNet (Howard *et al.*, 2017) neural network.[2] The network has been trained on the ImageNet database, a collection of over 14 million labelled photographs organized in more than 20,000 categories, comprising animals, people, objects, fungi, etc. MobileNet has been developed to compete in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) which uses a smaller version of the ImageNet dataset comprising only 1,000 categories. Applied to our dataset, the network outputs for every image a probability of the input image belonging to one of those categories, a vector of 1,000 activations. We are not interested in the actual classification—the assigned labels are unlikely to be useful due to how different our visuals are to the photographs in ImageNet—but we can use the probabilities in a similar way as the hash digests in the previous example. Two images that produce similar activation vectors are likely similar in visual content, too. We use Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) to project the high-dimensional activations to a two-dimensional space and visually evaluate the obtained image similarities against the groups obtained through the image hashing method.

4 Evaluation

4.1 Historical evaluation

To analyse the images in their historical context and in relation to the structural and bibliographic metadata of the books, we inserted the data and images into a visualization tool developed by Flavio Gortana and originally conceived to visualize a collection of coins (Gortana *et al.*, 2018). The web app, which is freely available on GitHub, allows us to visually inspect the entire set of images and study the identified groupings. By means of this visualization tool, we were able to identify in our corpus a radical change of habit at the beginning of the 1530s. In this period, we can trace two complementary phenomena.

First, many scientific subjects discussed since centuries in manuscripts and printed treatises were for the first time accompanied by a descriptive and explicative illustration. Visualizing the assigned groups against time as pictured in Fig. 3 makes this development evident. The groups are ordered vertically by number of images. Most image groups only appear after 1530, whereas the groups that we identified before this date cease to be published thereafter.

Second, most of the scientific subjects that were already accompanied by an explicative illustration, often since the late medieval period in the handwritten sources, were suddenly provided with a new illustration, often representing the same scientific content using novel imagery and, sometimes, introducing content-related innovations. A striking example is the illustrations demonstrating the sphericity of the

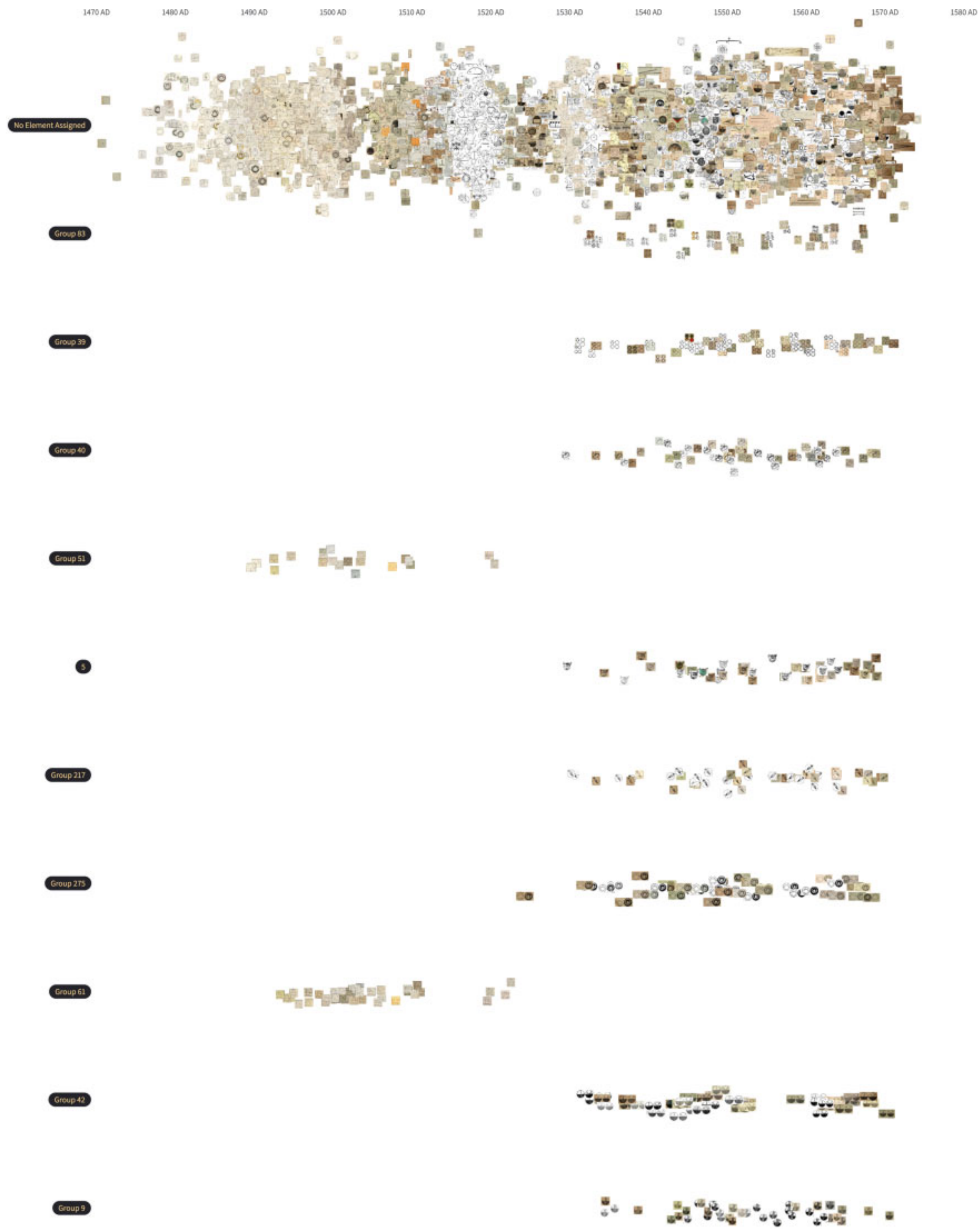
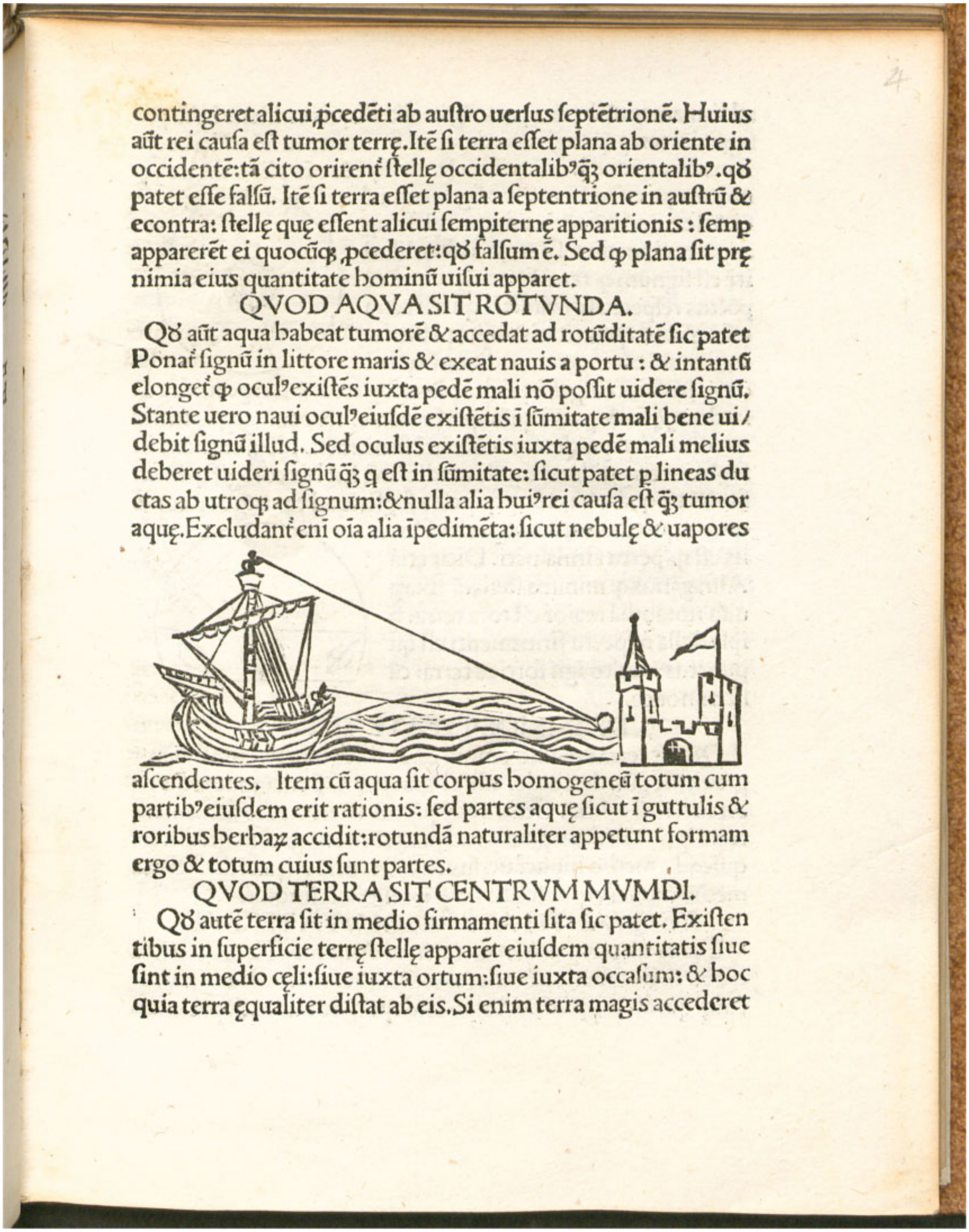


Fig. 3 Visualizing the image groups on a timeline using Coins (Gortana *et al.*, 2018) reveals a change in image production around 1530


4

contingeret alicui ꝓcedẽti ab auſtro uerſus ſeptẽtrionẽ. Huius
aũt rei cauſa eſt tumor terrę. Itẽ ſi terra eſſet plana ab oriente in
occidentẽ: tã cito orirent ſtellę occidentalib' q̃ʒ orientalib'. qđ
patet eſſe falſũ. Itẽ ſi terra eſſet plana a ſeptentrione in auſtrũ &
econtra: ſtellę quę eſſent alicui ſempiternę apparitionis: ſemp
apparerẽt ei quocũqʒ ꝓcederet: qđ falſum ẽ. Sed ꝙ plana ſit prę
nimia eius quantitate hominũ uiſui apparet.

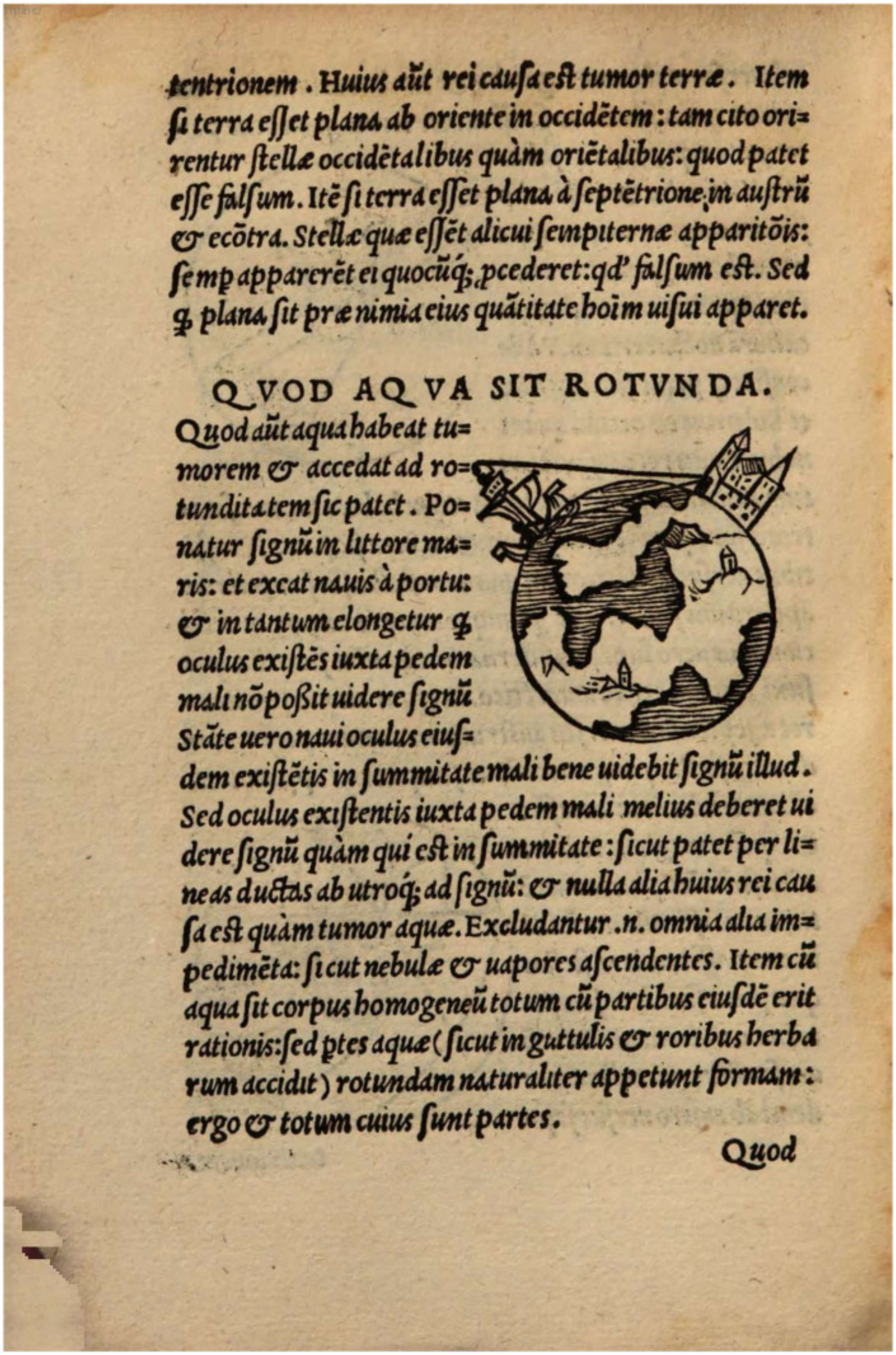
QVOD AQVA SIT ROTVNDA.

Qđ aũt aqua habeat tumorẽ & accedat ad rotũditatẽ ſic patet
Ponat ſignũ in littore maris & exeat nauis a portu: & intantũ
elonget ꝙ ocul' exiſtẽs iuxta pedẽ mali nõ poſſit uidere ſignũ.
Stante uero naui ocul' eiuſdẽ exiſtẽtis ĩ ſũmitate mali bene ui/
debit ſignũ illud. Sed oculus exiſtẽtis iuxta pedẽ mali melius
deberet uideri ſignũ q̃ʒ q eſt in ſũmitate: ſicut patet p lineas du
ctas ab utroqʒ ad ſignum: & nulla alia hui' rei cauſa eſt q̃ʒ tumor
aquę. Excludant eni oĩa alia ĩpedimẽta: ſicut nebulę & uapores
aſcendentes. Item cũ aqua ſit corpus homogeneũ totum cum
partib' eiuſdem erit rationis: ſed partes aquę ſicut ĩ guttulis &
roribus herbaꝝ accidit: rotundã naturaliter appetunt formam
ergo & totum cuius ſunt partes.

QVOD TERRA SIT CENTRVM MVMDI.

Qđ autẽ terra ſit in medio firmamenti ſita ſic patet. Exiſten
tibus in ſuperficie terrę ſtellę apparẽt eiuſdem quantitatis ſiue
ſint in medio cęli: ſiue iuxta ortum: ſiue iuxta occaſum: & hoc
quia terra ęqualiter diſtat ab eis. Si enim terra magis accederet

Fig. 4 A 1485 edition published in Venice with a visual demonstration of the sphericity of the earth. Image: Bavarian State Library, CC BY-NC-SA 4.0. Available at: http://daten.digitale-sammlungen.de/0003/bsb00036841/images/index.html?id=00036841&seite=13. Database record: hdl.handle.net/21.11103/sphaera.101123



tentrionem . Huius aũt rei causa est tumor terræ . Item
si terra esset plana ab oriente in occidẽtem : tam cito ori=
rentur stellæ occidẽtalibus quàm oriẽtalibus: quod patet
esse falsum . Itẽ si terra esset plana à septẽtrione in austrũ
& ecõtra. Stellæ quæ essẽt alicui sempiternæ apparitõis:
semp apparerẽt ei quocũq; pcederet: qd' falsum est. Sed
ꝙ plana sit præ nimia eius quãtitate hoĩm uisui apparet.

QVOD AQVA SIT ROTVNDA.

Quod aũt aqua habeat tu=
morem & accedat ad ro=
tunditatem sic patet . Po=
natur signũ in littore ma=
ris: et exeat nauis à portu:
& in tantum elongetur ꝙ
oculus existẽs iuxta pedem
mali nõ possit uidere signũ
Stãte uero naui oculus eius=
dem existẽtis in summitate mali bene uidebit signũ illud.
Sed oculus existentis iuxta pedem mali melius deberet ui
dere signũ quàm qui est in summitate : sicut patet per li=
neas ductas ab utroq; ad signũ: & nulla alia huius rei cau
sa est quàm tumor aquæ. Excludantur .n. omnia alia im=
pedimẽta: sicut nebulæ & uapores ascendentes. Item cũ
aqua sit corpus homogeneũ totum cũ partibus eiusdẽ erit
rationis: sed ptes aquæ (sicut in guttulis & roribus herba
rum accidit) rotundam naturaliter appetunt formam:
ergo & totum cuius sunt partes.

Quod

Fig. 5 A 1526 edition from Ingolstadt featuring a terraqueous globe, which represents a new understanding of the earth sphere as made up from both water and earth. Image: Bavarian State Library, NoC-NC. Available at: https://reader.digitale-sammlungen.de/de/fs1/object/display/bsb11110162_00012.html. Database record: hdl.handle.net/21.11103/sphaera.100070 earth using a ship on sea and two lines indicating that a castle on land is visible from the mast of the ship first, before becoming visible for an observer on the boat. Before 1530, the image used depicts a ship sailing on a curved sea, as visible in Fig. 4. After 1530, the illustration includes an entire world globe, a terraqueous globe (Fig. 5) representing a (new) worldview of water and landmass occupying the same sphere. Using the timeline view, we can see how the new illustration is introduced in 1530 and the use of the previous visual declining.

4.2 Evaluation of method

To compare the results of the ImageHash grouping with the MobileNet approach we look at the location of the ImageHash groups within the UMAP projection. We ingest the data into VikusViewer (Glinka *et al.*, 2017), a generic visualization tool for large image collections and visually inspect the groups obtained through image hashing, their location in the UMAP projection and the image's visual similarity.

Figure 6 shows a screenshot of the visualization tool with the UMAP projection of the images in the centre. Numbers at the top represent the groups identified through the ImageHash algorithm. Hovering over a number highlights the corresponding group in the UMAP projection. If the highlighted images appear close together, they are classified as similar by both the ImageHash and Mobilenet approach. If they appear apart the two methods disagree and we visually inspect the classification, evaluating which of the groupings we regard as correct.

We also include images that have not been assigned a group by the ImageHash. These are highlighted in Fig. 7. Most of those images appear in the centre of the visualization, which means that they neither have been assigned a clear position in the UMAP projection. Some of them however form distinct groups at the edges of the visualization, suggesting that these are indeed groups that the

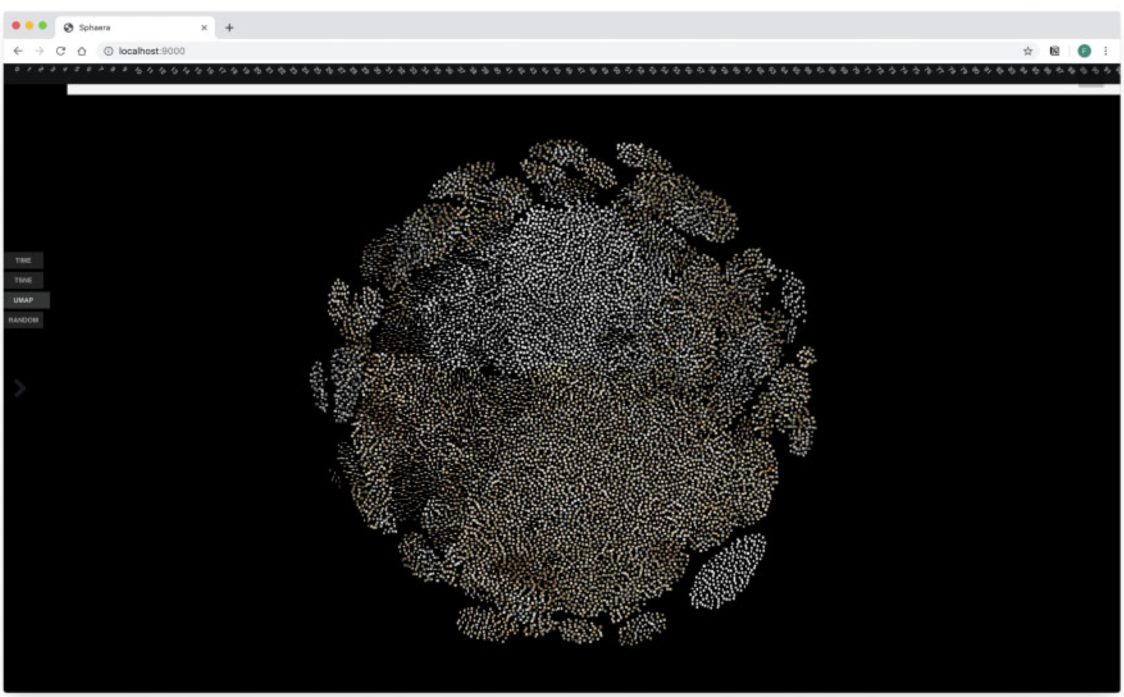


Fig. 6 Images inserted into VIKUSViewer (Glinka *et al.*, 2017) and arranged using a UMAP projection based on the ImageNet activation vectors

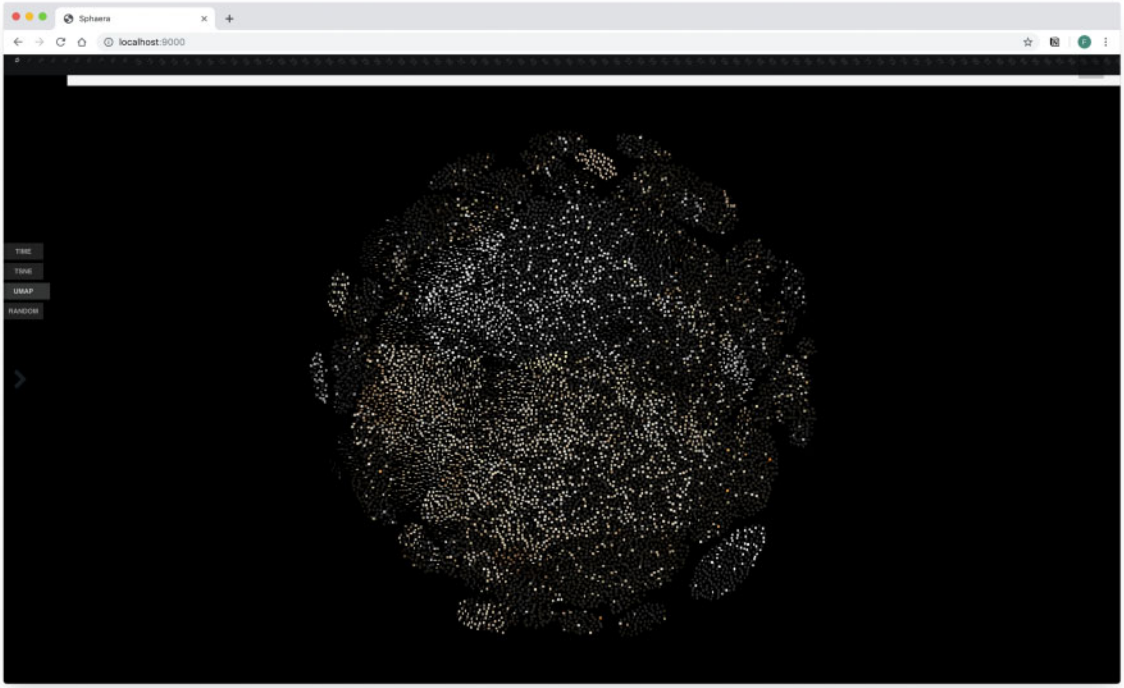


Fig. 7 Evaluating the images within VikusViewer (Glinka *et al.*, 2017). Highlighted are the images that have not been classified by the ImageHash algorithm

ImageHash method has missed. In most cases, we can attribute the 'missed' groupings to slight differences in the images that become evident upon closer inspection. The group highlighted in the top middle represents a set of star maps, each similar in layout, but slightly different in content (Fig. 8). Another set of images depicting a geometric demonstration of the circle as a perfect form has not been grouped by the ImageHash (Fig. 9). Again we can attribute this behaviour to the slight differences in the images with the individual geometric figures within the illustrations being arranged in different order. Whether these variations are considered significant depends on the individual research question.

Inspecting the individual groups obtained through the image hashing against the UMAP projection we find that the majority of them align, indicating that the groups we obtained are correct by this measure. We identify several examples where the colour of the paper or the quality of the scan has produced separate clusters of images in the UMAP projection from images that have been classified as similar by the image hash. As we are not interested in comparing paper or scan quality, we regard the image hash approach, which discards colour information altogether, as correct.[3] Another area where the methods disagree are long or tall images. Both methods require the input images to be resized to a square aspect ratio. Although this causes the image hash approach to miss commonalities within wider images, the neural network appears to be more robust in processing images in all aspect ratios.

5 Conclusion

We observed that the arguably simple method of calculating and comparing image hashes reliably identifies near-duplicate images and forms groups of recurring visuals that, in our case, lead to important

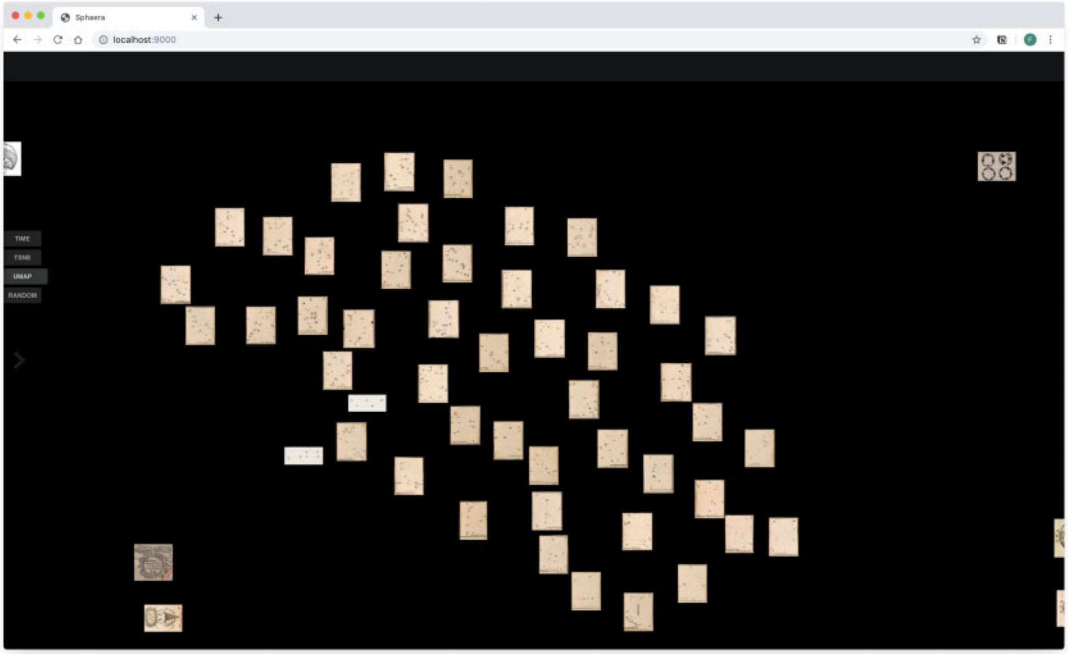


Fig. 8 A set of similar, but slightly different star maps has been grouped together by UMAP, but not by the ImageHash algorithm

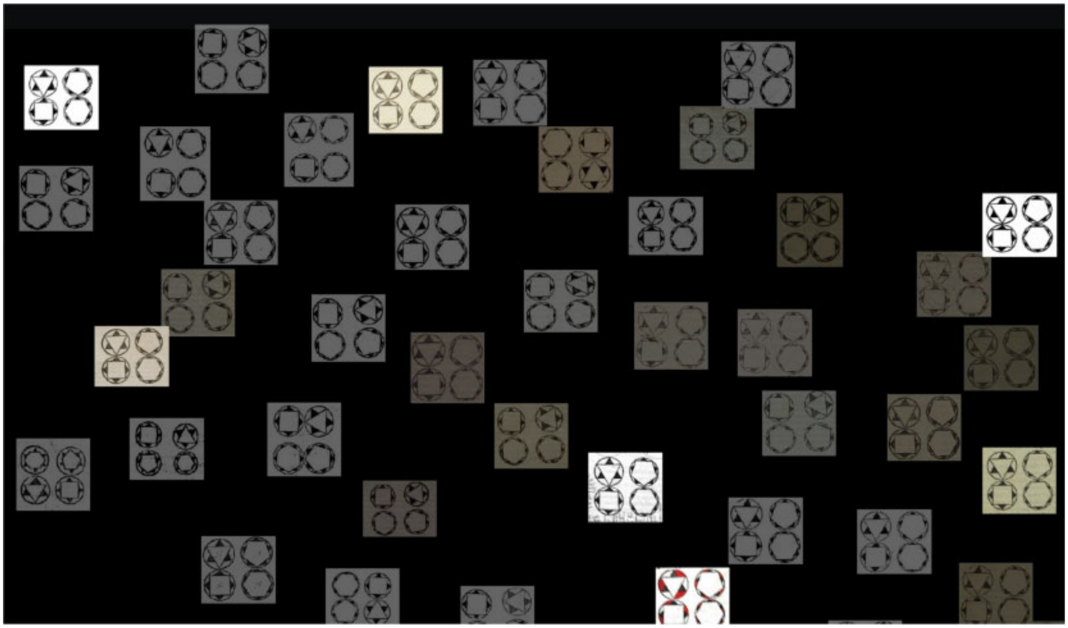


Fig. 9 Although visually similar, the images that are not highlighted are all slightly different and have therefore not been grouped using the ImageHash algorithm

new insights. The method's main limitation is its inability to group images that exhibit slight variations, but may nevertheless be regarded as 'same' or similar by a researcher. An important point to consider is the fact that, unlike the MobileNet approach or most other methods that employs machine learning, the algorithm does not need to be trained and works on any set of images. The ImageHash method has only few adjustable parameters and the algorithm works by executing a small number of reproducible steps. For historical research where sources and interpretations need to be transparent and traceable, the fact that the algorithm does not constitute a black box may be crucial.

References

Buchner, J. (2017). Imagehash. https://github.com/JohannesBuchner/imagehash (accessed 6 August 2018).

Foo, J. J., Zobel, J., Sinha, R., and Tahaghoghi, S. M. M. (2007). 'Detection of near-duplicate images for web search', In *Proceedings of the 6th ACM International Conference*, Amsterdam, the Netherlands, pp. 557–64.

Glinka, K., Pietsch, C., and Dörk, M. (2017). Past visions and reconciling views: visualizing time, texture and themes in cultural collections. *Digital Humanities Quarterly*, **11**(2).

Gortana, F., von Tenspolde, F., Guhlmann, D., and Dörk, M. (2018). Off the grid: visualizing a numismatic collection as dynamic piles and streams. *Open Library of Humanities*, **4**(2): 30.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, **29**(2): 147–60.

Howard, A. G., Zhu, M., Chen, B. *et al.* (2017). *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. a*rXiv preprint arXiv:1704.04861.

Ke Y., Sukthankar R., and Huston L. (2004). Efficient near-duplicate detection and sub-image retrieval. *Proceedings of ACM International Conference on Multimedia (MM)*, **4**(1):5.

Knuth, D. (1998). *The Art of Computer Programming, Volume 3: Sorting and Searching*. Upper Saddle River, NJ: Addison Wesley.

Kravetz, N. (2013). Kind of Like That. hackerfactor.com. http://www.hackerfactor.com/blog/index.php?/archives/529-Kind-of-Like-That.html (accessed 6 August 2018).

Kräutli, F. and Valleriani, M. (2018). CorpusTracer: a CIDOC database for tracing knowledge networks. *Digital Scholarship in the Humanities*, **33**(2): 336–46.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). *ImageNet Classification with Deep Convolutional Neural Networks.* https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf (accessed 4 November 2019).

McInnes, L., Healy, J., and Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv preprint arXiv:1802.03426.*

Mehta, B., Nangia, S., Gupta, M., and Nejdl, W. (2008). Detecting image spam using visual features and near duplicate detection. In *Proceeding of the 17th International Conference*, pp. 497–506.

Project Mirador (2014). Mirador IIIF Image Viewer. projectmirador.org. http://projectmirador.org/ (accessed 6 August 2018).

Valleriani, M. (2017). The tracts of the sphere: Knowledge restructured over a network. In: Valleriani M. (ed.), *The Structures of Practical Knwoledge.* Basel, CH: Springer International Publishing, pp. 421–73.

Venkatesan, R., Koom, S. M., Jakubowski, M., and Moulin, P. (2000). Robust image hashing. In *Proceedings of IEEE ICIP*. Vancover: IEEE.

Zauner, C. (2010). *Implementation and Benchmarking of Perceptual Image Hash Functions.* Master's thesis, Upper Austria University of Applied Sciences, Hagenberg Campus, pp.1–107.

Notes

1 For example, if Image A is similar to Image B and Image B is similar to Image C, we assume Image A is also similar to Image C.

2 https://github.com/tensorflow/models/blob/master/research/slim/nets/mobilenet_v1.md

3 This disagreement between the two approaches could likely be eliminated by converting the images to black and white before computing the MobileNet activation vectors.