Cluttering Reduction for Interactive Navigation and Visualization of Historical Images

Evelyn Paiz-Reyes,* Mathieu Brédif, Sidonie Christophe

LASTIG, Univ Gustave Eiffel, ENSG, IGN, F-94160 Saint-Mande, France
(evelyn.paiz-reyes, mathieu.bredif, sidonie.christophe)@ign.fr

* Corresponding Author

Abstract: Iconographic representations, such as historical photos of geographic spaces, are precious cultural heritage resources capable of describing a particular geographical area's evolution over time. These photographic collections may vary in size, between hundreds and thousands of items. With the advent of the digital era, many of these documents have been digitized, spatialized, and are available online. Browsing through these digital image collections represents new challenges. This paper examines the topic of historical image exploration in a virtual environment enabling the co-visualization of historical photos into a contemporary 3D scene. We address the topic of user interaction considering the potential volume of the input data. Our methodology is based on design guidelines that rely on visual perception techniques to ease visual complexity and improve saliency on specific cues. The designs are additionally implemented following an image-based rendering approach and evaluated in a group of users. Overall, these propositions may be a notable addition to creating innovative ways to visualize and discover historical images in a virtual geographic environment.

Keywords: geovisualization, historical photographs, image-based rendering, visual complexity.

1. Introduction

Photographs represent visual collections of knowledge and history. Many institutions involving historians, archivists, and national mapping agencies, among others, are archiving vast numbers of historical photographs. Examples such as the *Online Library of National Archives*[1] and *Open Musée Niépce*[2] follow a digital photo library approach to visualize their photographic compilations. This classical technique lists and displays the images in a 2D grid controlled by metadata, *e.g.*, keywords, dates, and places. Alternative strategies like *Historypin*[3] (Armstrong, 2012) and *Smapshot*[4] (Blanc et al., 2018, Produit et al., 2018) are innovative and collaborative solutions that place old photographs in a 2D or 3D realistic/topographic view. *Photo Tourism* is another example following this path, allowing to browse extensive unstructured pictures of touristic sites inside a reconstructed 3D model (Snavely et al., 2006).

Still, collections of historical images are a particular input for any **geovisualization** approach. These photos are likely to present:

Data uncertainties: unlike modern remote imagery, where the data is well registered and georeferenced, historical photographs are only partially documented. These images may only be approximately located (spatially) in a 2D/3D environment.

Diversity on the source: iconographic collections of historical photos are represented in various types and styles, *e.g.*, paintings, postcards, and old black and white or color photographs. A visualization approach for these pictures should exploit and adapt to this heterogeneity.

Large Volume: many institutions have been and are still archiving a large number of historical images. The result is a high volume of items (representing a vast amount of visual data) that need to be explored and managed digitally.

Extensive Scales: historical photos depict objects and scenes covering a large extent of geographic area. A browsing approach through these photographs needs to support extensive and continuous space discovery.

Accordingly, we aim to **discover historical photographs**, *e.g.*, postcards and old aerial/terrestrial imagery. We want to combine these historical images (2D samples of a photographed scene) with a 3D city model (constructed by current urban material) in a virtual environment to enable their co-visualization. Figure 1 (on the next page) shows a depiction of the integration of these two multi-dimensional data. Our goal involves visualizing and exploring the historical photos in this virtual environment, particularly while considering their potential volume in the contemporary 3D scene.

---

[1] https://www.siv.archives-nationales.culture.gouv.fr
[2] http://www.open-museeniepce.com
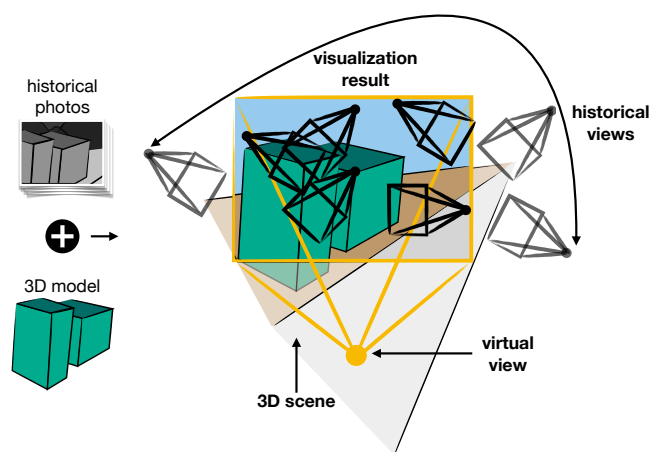[3] https://www.historypin.org
[4] https://smapshot.heig-vd.ch

Figure 1. A depiction of our visualization environment when many historical images are added in a 3D scene.

Our system needs to be cable of multiple image browsing inside the 3D environment to cross-analyze the different historical photographs and collections. Nevertheless, as the number of photos increases, so does the extent of visual data we need to manage. Furthermore, this complexity is magnified when we consider temporal information (*e.g.*, the date when the photograph was captured) and visual cues like the photographer's viewpoints and thumbnails of the photos. Therefore, we face the issue of **managing all this visual complex information collectively** while providing interaction capacities to the users to navigate from image to image.

We define our research challenge as: **how to continuously navigate through a massive amount of visual and spatio-temporal data represented on a large scale?** We aim to propose a procedure that will facilitate and enhance navigation when discovering many historical images in a virtual geographic environment. Our goal is to improve every photo's saliency and reduce cluttering obtained when many photographs are visualized to allow users to explore pictures through a 3D model of the photographed scene.

2. Related Work

The **geovisualization of historical photos** is a broad theme that can be tackled from different perspectives. Methods such as *Navigae*[5] and *Navilium*[6] facilitate examining several spatially oriented images in a 2D map. However, these applications lack interaction since they are restricted to an orthogonal top view. *Historypin* and *WhatWasThere*[7] may overcome this by taking the photographer's perspective inside the spatial environment. Both systems enable users to upload and integrate their old photographs on *Google Street View* (Armstrong, 2012). Unfortunately, the association of historical and current material may result in considerable misalignments between images and the current and contemporary material provided by *Google Street View*.

[5] https://www.navigae.fr
[6] https://www.navilium.com
[7] http://www.whatwasthere.com

Photo Tourism (Snavely et al., 2006), *Ambient Point Cloud* (Goesele et al., 2010), *PhotoCloud* (Brivio et al., 2013), and *Smapshot* (Blanc et al., 2018, Produit et al., 2018) are all alternative methods that enable the placement of these photos in a 3D geographical context. These techniques focus on continuous navigation and exploration through a rendering process that uses a 3D environment to display these photographs. Still, in some cases, these applications may have limitations or are not adapted to historical images, *e.g.*, *Smapshot* restricts the view to the same position as the capturing camera of the visualized picture.

This paper focuses on the **user's visual exploration**, particularly what can be perceived when many images are visualized inside a 3D topographic scene. We can define the perception process as a sequence of steps determining the experience and reaction to any stimuli (Goldstein, 2010, Snowden et al., 2011). This concept can be applied to our input data, in which we define our visual stimulus as historical photographs. Based on it, existing methods and techniques can allow us to drive attention to some parts of the 3D scene or some objects/structures/characteristics in the historical images.

The concept of **visual variables**, introduced by Bertin and Berg (2010) originally in 1967, is an essential factor derived from visual perception. These variables provide a graphical representation of any data at a fundamental level (Roth, 2017, Cöltekin et al., 2020). Aside from essential attributes like color, transparency, and sizing, other characteristics may facilitate users to recognize existing spaces in the scene.

For instance, we may achieve a higher dynamic interaction inside a 3D environment by representing any photograph capturing orientation and position with **bookmarks**. These markers can be displayed differently, including a text link, a thumbnail image, or a 3D object. Precisely, the method proposes placing the bookmark in every image's viewpoint (Forgione et al., 2016). However, these markers may not be visible for large-scale scenes when they are located too far away.

On the other hand, a **heat map** is an approach employing color as the primary visual attribute. It allows instantly identifying potential areas or objects of interest, *e.g.*, our historical photos. This technique can provide information about the photographs by showing their distributions (Bruschke et al., 2018). Still, these maps are usually used only on surfaces of buildings instead of the complete scene.

A notable difficulty users experience when **browsing** inside image compilations is finding the desired content (Baruzzo et al., 2009, Ardissono et al., 2012). A standard approach uses a thumbnail menu slider where reduced-size versions of the images are displayed in a list that allows the user to search for the desired photographs. Still, as the number of pictures increases, so does the searching time.

The **clustering** of the photographs can be seen as an alternative solution for browsing over these image

compilations. It is a process where the photos are partitioned into a set of meaningful categories (Xu and Wunsch, 2015). For instance, in a 2D visualization, clustering can facilitate the exploration of several photographs in a simple 2D map by grouping the images by their spatial location. Instead of visualizing the photographs all over the map, only the groups are visible, and when selected, they can be expanded to the complete set of images (Papadopoulos et al., 2010).

More complex techniques in 3D involve the computation of semantic distances to cluster the images. The result can be placed inside a thumbnail bar employed to navigate the data using a dynamic image hierarchy (Brivio et al., 2013). A more creative solution suggests clustering the images and representing them as thumbnails in a **gallery form** (Lekschas et al., 2020). Alternatively, a clustering procedure can also best describe a scene by selecting a set of canonical views to form a scene summary. It partitions the photo collection into groups of related images based on visual features (Simon et al., 2007).

3. Design Guidelines

Our work aims to facilitate virtual exploration through various photo collections by reducing the visual complexity generated when these images are placed inside a 3D urban model. We have selected a set of existing methods that, combined, can help us achieve this goal. These are geovisualization techniques used to drive attention to data and some aspects of it, *e.g.*, semantics, spatial, and temporal aspects. This section presents the **design guidelines** we have created for each one of these chosen approaches.
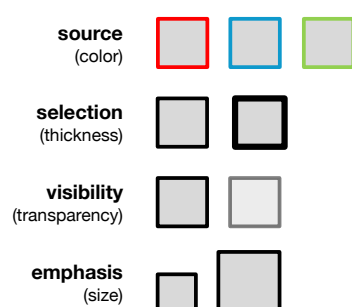
3.1 Visual Variables

Figure 2. Our selected visual variables and their usage.

A clear **visual distinction of each photograph and its photographic collections** can be presented to the user. Visual variables are used to achieve it. We selected color (hue), thickness, transparency, and size, to distinguish one image from the others. As Figure 2 denotes, the color attribute can differentiate among various image collections (sources). The thickness discriminates between selected and non-selected photos. The transparency shows images outside the field of view, and the change of size pops up an element in the scene.

3.2 Bookmarks

We adopt bookmarks to **showcase the different points of view** from the photographs. The user can interact with them inside the 3D environment, and when double-clicked, a "fly-to" animation moves the view from the current position to the selected image viewpoint. As depicted in Figure 3, we consider using the historical camera frustum (*i.e.*, a pyramid) as the visual representation of our bookmarks. The top point (apex of the pyramid) corresponds to the camera's optical center and the base to the image plane. We extend this pyramid when the user hovers over the bookmark to show the image's projection on top of the 3D city model.
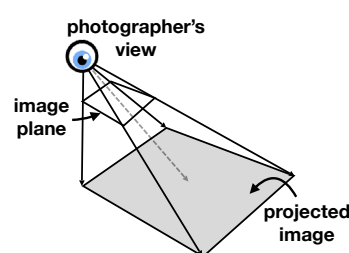

Figure 3. Our design for a bookmark. The gray area represents an image.

For a more immediate identification, we add a generic icon at the top point of the frustum. Because the bookmarks are placed inside a 3D scene, these objects must be mapped to a 2D representation to fit the screen space. However, when they are placed too far from the current virtual view, this mapping may lose the bookmarks visually. To contrast this, we remove the depth factor when bookmarks are mapped from the 3D space to the 2D screen space. The result is an enhanced icon and remains visible even when markers are too far from the current view.
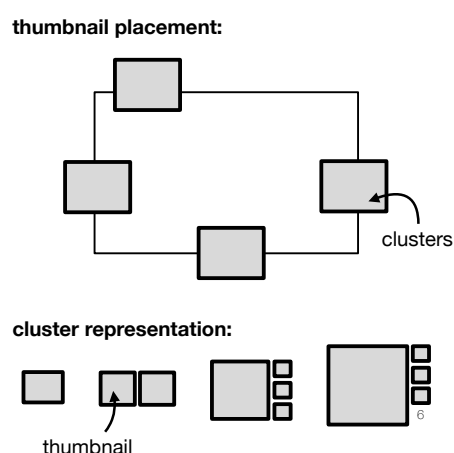
3.3 Gallery/Clustering

Figure 4. Our proposed clustered thumbnail placement and representation.

To **simplify the display of a set of historical images**, we selected a clustering approach, as presented in Figure

4. We use the alternative gallery/clustering visualization method, so the photographs are more visible, and it is possible to select a specific image quickly. To reduce the cluttering inside the scene, we choose all boundaries/sides of the visualization rather than a single 1D row or column of images as a better alternative to place the thumbnails. It presents all the thumbnails at once and avoids the need for the user to go inside the 3D scene and search for the desired image through all the bookmarks.

We use a ranking function to choose the most interesting historical photographs for the current view based on the distance between the view position and the historical image footprints. A hierarchical clustering step groups the photographs into smaller sets. The photo clusters are placed on the border of the view, as close as possible to the projections of the represented images. This process is re-computed as the user moves the 3D view (navigating through the urban model), resulting in the movement of the clusters around the view to adapt to their new computed positions.

3.4 Heat Map

We take advantage of a heat map to **identify possible areas in the scene with large numbers of photos**. We delimit its extension with the footprints of the projected pictures. A "fly-to" animation is applied to move the view to a position where the whole range of the footprints is visible and covered. It allows the viewer to glance at what areas these images describe, as shown in Figure 5.
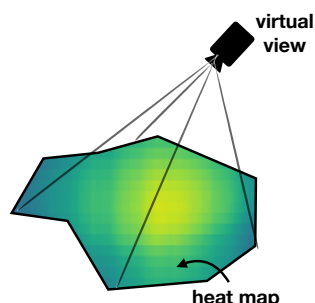

Figure 5. Our designated way (a heat map) to identify areas of interest.

3.5 Context Update

Because the 3D scene may present a context in which the historical image is visualized, we consider its design. Instead of imposing a specific style on this **3D environment**, we **update and change it as needed**. For instance, an abstract representation may allow showcasing more in-depth historical photos. At the same time, the use of an orthoimage may reveal the area's evolution from the acquisition period to now. To this, we add the possibility to change its radiometry (colors) by employing predefined lookup tables. As displayed in Figure 6, this allows the user to compare or homogenize the photographs with their surrounding 3D scene.
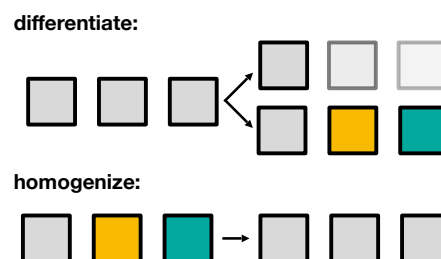

Figure 6. Our two approaches for the context: differentiate or homogenize.

3.6 Temporal Selection

The information known about historical photographs may be uncertain. It is the case regarding their temporal information. For instance, a specific acquisition date may not be available, but instead, only a time range may be known, *i.e.*, uncertainty in the dates. We adjust our temporal selection to this constraint, as is exhibited in Figure 7. We use a time axis where a time range can be selected, but instead of representing each image by a point, line, or thumbnail in the axis, we use a histogram representation. It allows us to adapt to the possible uncertainties that the photographs' temporal information may have. Each bar in the histogram denotes the sum of the photos belonging to a specific time range. Users can **control the temporal parameters** through the selected range (green bar in our diagram) and display only the photos belonging to it.
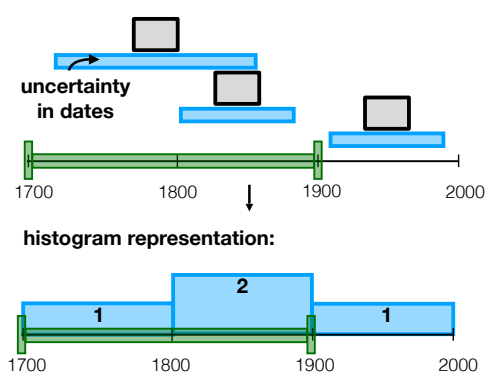

Figure 7. Our time axis, which employs a histogram to overcome the uncertainty in dates. The green bars show the selected time range.

4. Results

4.1 Implementation

We consider the presentation of all the historical images within a 3D contemporary environment. The pictures are spatially oriented and reprojected into this scene. Following an image-based rendering approach, a pixel-accurate visualization of a specific photo is obtained when a virtual view is placed in the photographer's viewpoint. The viewer is capable of navigating and

exploring all available photographs through this virtual view.

We have implemented a prototype[8] following all the techniques described previously using the three.js[9] WebGL rendering library along with the itowns[10] framework. As an input to our system, we require:

- The **historical photographs** portraying a snapshot of the geographical scene.
- The image(s) **orientation data** mapping the photo coordinates to those in the 3D scene. It contains the following information:

  Camera(s) exterior orientation: a transformation (position and orientation) from 3D world coordinates to the 3D camera's local coordinates.

  Camera(s) interior orientation: a transformation from the 3D camera's coordinates into the 2D image coordinates (including principal point and focal).

- A **3D city model** that enables the navigation of the view camera in a 3D environment. We use only the representation of building in our urban scene, generated by recent acquisitions, *i.e.*, contemporary.

When a photographic collection is loaded, the covered extent of the photographs is exposed through a heat map. The red areas have a more significant amount of images compared to the blue ones. As depicted by Figure 8, a time axis with the histogram representation allows selecting any specific time frame, where the display images will depend on this selection. To support the cross-analysis of different image collections, we utilize a specific color to represent each dataset.
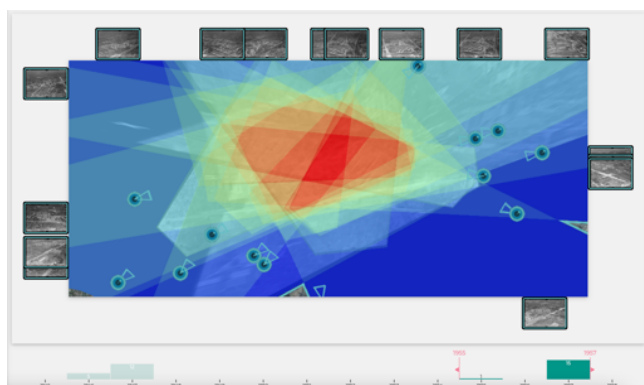

Figure 8. A heat map shows the extent of the selected photo collection where the most captured points in the scene are. The red color represents a higher amount of images than the blue one, which represents the least.

A change from the overview of a selected photo collection is possible. The user can move to any image viewpoint

---

utilizing the bookmarks or clustered thumbnails. Figure 9 displays the resulting visualization in the prototype when three historical images are projected into the scene. A juxtaposition of our 2D and 3D image selection options (bookmarks and thumbnails/clusters) is noticeable. While the bookmarks are placed inside the 3D environment, the photo clustered galleries are displayed on the view's border for fast access by the user. The number displayed in the clusters is equivalent to the amount of photos inside that cluster.
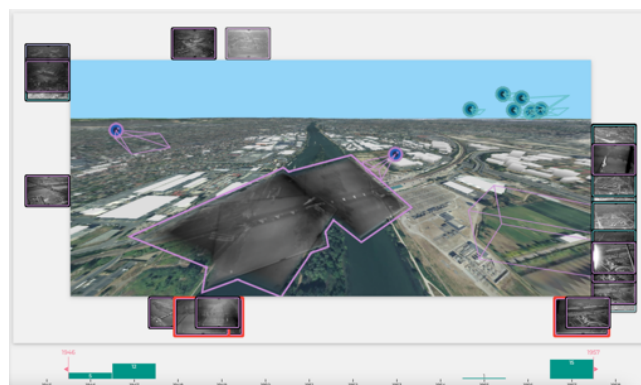

Figure 9. A visualization of multiple historical images with our proposed approach. Our prototype projects the historical images on a contemporary 3D environment.

As showcased by Figure 10, for homogenization of the context and the projected photos, it is possible to update the style of the 3D scene, *e.g.*, to a black and white style following the same one as the displayed photographs. On the contrary, for differentiation, a more abstract representation of the environment can be used.
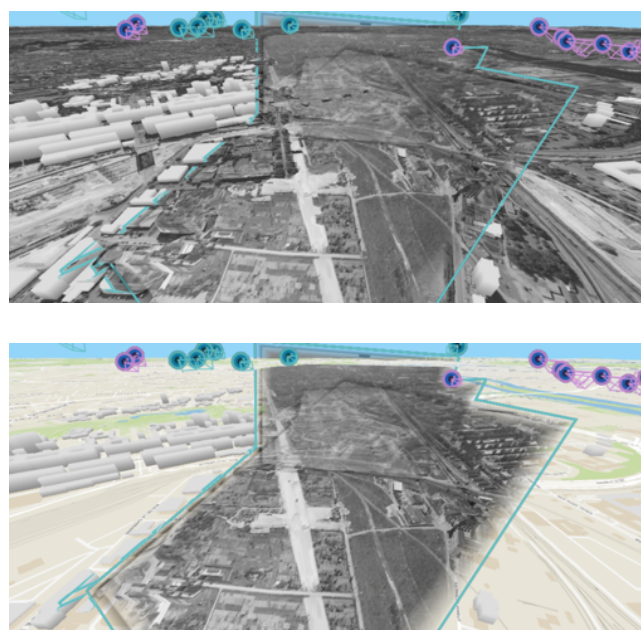

Figure 10. The effect of changing the 3D context style to homogenize(top)/differentiate (bottom) between the scene and the display photos.The border with the clusters has been removed to showcase the resulting view.

4.2 User Study

A swift user study of 10 participants has given us some initial results on the subject. This study was performed in an early prototype version, so not all our proposed methods were implemented. On it, we only tested: *(i)* visual variables; *(ii)* bookmarks; *(iii)* gallery/clustering; *(iv)* temporal selection.

The study showed that the use of visual variables and bookmarks could help the user during the 3D exploration. However, when too many bookmarks were displayed, the user noticed that the large number of rays (from each bookmark) cluttered the 3D view. A simple solution we used is to allow the option of increasing and reducing the size of the bookmark pyramid. Additionally, we give the user the option to hide/show whenever it is needed.

One of the users' main concerns was that nothing was happening in the system to let them know that a dataset had been loaded and where the data was located. To overcome this, we added the heat map extent overview as an additional design proposition. It allowed us to show where an image collection is located every time it is loaded. We connected this heat map to the timeline, allowing users to glance at the photos of a selected time range.

Finally, our primary constraint was the design of the clusters. The users expressed that they felt there was no hint of how the clusters are related to the 3D environment. Additionally, it was difficult for them to follow the change of position that the cluster suggested. Moreover, the gallery representation also caused some doubt since users wondered why one image was more significant than the others in the clusters of three or more elements.

5. Conclusions

In summary, the main contribution of this work is the proposition of a set of procedures that combined are capable of **reducing the visual complexity of a multiple historical image exploration in a 3D environment**. We define them in six main properties that our 3D browsing application has: *(i)* saliently shows the images; *(ii)* allows easy recognition of the different viewpoints from the historical capturing cameras (photographers perspective); *(iii)* reduces the cluttering through the scene; *(iv)* recognizes areas of interest; *(v)* manipulates the style and color of the scene; *(vi)* includes the temporal element. These design guidelines are combined into one prototype, a web-based implementation that checks our solutions' competence.

As future work, we would like to improve the comprehension of the clustering/gallery approach since it was revealed to be somehow confusing by the users. One of the main points mentioned was its movement around the border of the view. One possible solution is improving its continuity or changing the positioning to a static representation, *e.g.*, keeping the clusters only on one side of the view border instead of the four sides.

6. Acknowledgements

References

Ardissono, L., Kuflik, T. and Petrelli, D., 2012. Personalization in cultural heritage: the road travelled and the one ahead. *User Modeling and User-Adapted Interaction* 22(1), pp. 73–99.

Armstrong, N., 2012. Historypin: bringing generations together around a communal history of time and place. *Journal of Intergenerational Relationships* 10(3), pp. 294–298.

Baruzzo, A., Casoto, P., Challapalli, P., Dattolo, A., Pudota, N. and Tasso, C., 2009. Toward semantic digital libraries: Exploiting web 2.0 and semantic services in cultural heritage. *J. Digit. Inf.*

Bertin, J. and Berg, W. J., 2010. *Semiology of graphics: diagrams, networks, maps.* First edn, ESRI Press: Distributed by Ingram Publisher Services.

Blanc, N., Produit, T. and Ingensand, J., 2018. A semi-automatic tool to georeference historical landscape images. *PeerJ Preprints* 6(e27204), pp. 1–6.

Brivio, P., Benedetti, L., Tarini, M., Ponchio, F., Cignoni, P. and Scopigno, R., 2013. PhotoCloud: interactive remote exploration of joint 2d and 3d datasets. 33(2), pp. 86–96, c3.

Bruschke, J., Niebling, F. and Wacker, M., 2018. Visualization of orientations of spatial historical photographs. *Eurographics Workshop on Graphics and Cultural Heritage* pp. 189–192.

Cöltekin, A., Griffin, A. L., Slingsby, A., Robinson, A. C., Christophe, S., Rautenbach, V., Chen, M., Pettit, C. and Klippel, A., 2020. Geospatial information visualization and extended reality displays. In: *Manual of Digital Earth*, Springer Singapore, pp. 229–277.

Forgione, T., Carlier, A., Morin, G., Ooi, W. T. and Charvillat, V., 2016. Impact of 3D bookmarks on navigation and streaming in a networked virtual environment. In: *Proceedings of the 7th International Conference on Multimedia Systems*, ACM, pp. 1–10.

Goesele, M., Ackermann, J., Fuhrmann, S., Haunbold, C., Klowsky, R., Steedly, D. and Szeliski, R., 2010. Ambient point clouds for view interpolation. *ACM Transactions on Graphics.*

Goldstein, E. B., 2010. *Sensation and perception.* Eighth edn, Wadsworth Cengage Learning.

Lekschas, F., Behrisch, M., Bach, B., Kerpedjiev, P., Gehlenborg, N. and Pfister, H., 2020. Pattern-driven navigation in 2D multiscale visualizations with scalable insets. *IEEE Transactions on Visualization and Computer Graphics* 26, pp. 611–621.

Papadopoulos, S., Zigkolis, C., Kapiris, S., Kompatsiaris, Y. and Vakali, A., 2010. ClustTour: city exploration by use of hybrid photo clustering. In: *Proceedings of the international conference on Multimedia - MM '10*, ACM Press, pp. 1617–1620.

Produit, T., Blanc, N., Composto, S., Ingensand, J. and Furhoff, L., 2018. Crowdsourcing the georeferencing of historical pictures. *Proceedings of the Free and Open Source Software for Geospatial (FOSS4G) Conference* 18(6), pp. 1–8.

Roth, R. E., 2017. Visual variables. In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*, John Wiley & Sons, Ltd, pp. 1–11.

Simon, I., Snavely, N. and Seitz, S. M., 2007. Scene summarization for online image collections. In: *2007 IEEE 11th International Conference on Computer Vision*, IEEE, pp. 1–8.

Snavely, N., Seitz, S. and Szeliski, R., 2006. Photo Tourism: exploring photo collections in 3D. *ACM Transactions on Graphics* 25(3), pp. 835–846.

Snowden, R. J., Thompson, P. and Troscianko, T., 2011. *Basic vision: an introduction to visual perception*. Revised edn, Oxford University Press.

Xu, R. and Wunsch, D., 2015. *Clustering*. John Wiley & Sons.