

Visual exploration of historical image collections : an interactive approach through space and time

Evelyn Lissete Paiz Reyes

▶ **To cite this version:**

Evelyn Lissete Paiz Reyes. Visual exploration of historical image collections : an interactive approach through space and time. Geography. Université Gustave Eiffel, 2021. English. ⟨NNT : 2021UEFL2029⟩. ⟨tel-03697260⟩

HAL Id: tel-03697260

https://theses.hal.science/tel-03697260v1

Submitted on 16 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Visual Exploration of Historical Image Collections: *An Interactive Approach Through Space and Time*

Evelyn Lissete PAIZ REYES

Université Gustave Eiffel

Visual Exploration of Historical Image Collections:
An Interactive Approach Through Space and Time

Thèse de doctorat de l'Université Gustave Eiffel

École doctorale n° 532, MSTIC
Spécialité de doctorat : Sciences et Technologies de l'Information Géographique
Unité de recherche : LASTIG

Thèse présentée et soutenue à l'Institut National de l'Information Géographique et Forestière, le 08/12/2021 par

Evelyn Lissete PAIZ REYES

Composition du Jury

Livio DE LUCA Directeur de Recherche (DR), CNRS, UMR MAP	Rapporteur
Didier JOSSELIN Directeur de Recherche (DR), Univ. d'Avignon, CNRS, UMR ESPACE	Rapporteur
Paule-Annick DAVOINE Professeur (Pr.), Univ. Grenoble Alpes, LIG	Examinatrice
Gilles GESQUIÈRE Professeur (Pr.), Univ. Lumière Lyon 2, LIRIS	Président
Sidonie CHRISTOPHE Directrice de Recherche (DR), LASTIG, Univ. Gustave Eiffel, IGN-ENSG	Directrice de thèse
Mathieu BRÉDIF Ingénieur en Chef des Mines, LASTIG, Univ. Gustave Eiffel, IGN-ENSG	Encadrant

Encadrement de la thèse

Sidonie CHRISTOPHE Directrice de Recherche (DR), LASTIG, Univ. Gustave Eiffel, IGN-ENSG	Directrice de thèse
Mathieu BRÉDIF Ingénieur en Chef des Mines, LASTIG, Univ. Gustave Eiffel, IGN-ENSG	Encadrant

Visual Exploration of Historical Image Collections: *An Interactive Approach Through Space and Time*



This research work has been carried at the French National Mapping Agency (IGN), within the GeoVis team, from the Laboratory of Geographic Information Science (LASTIG). The financing comes from the French Research Agency (ANR), in the framework of ALEGORIA ANR-17-CE38-0014-01 (Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge) project (2018-2022).

The document has been written with the help of KOMA-Script and LaTeX using the kaobook class by Federico Marotta. The front and back cover are an adaptation from the design of starline/freepik.

© 2021 E. Paiz (ORCID 0000-0002-2803-9674)

For my mother.

Abstract

Visual Exploration of Historical Image Collections: *An Interactive Approach Through Space and Time*

Iconographic representations, such as **historical photos** of geographic spaces, are valuable Cultural Heritage resources enabling the description of a particular topographical area's evolution over time. With the advent of the digital era, many of these documents have been **digitized**, **spatialized**, and are now **available online**. However, it remains **difficult** for **users** to **browse** and **visualize** these images in space and time. The ANR ALEGORIA project tackles this challenge and is oriented towards photographic collections depicting the French territory from the interwar period to the present day, composed of aerial and terrestrial photographs provided by the consortium (National Archives, Nicéphore Niépce Museum, French National Mapping Agency).

This dissertation presents our proposition for a **visual exploration** of **historical image collections** using an **interactive approach** through **space** and **time**. We propose to co-visualize historical photographs inside contemporary 3D topographic models, and our **two main contributions** are: *(i)* the **extrapolation** of the **photograph's geometric distortion model** to enable users to view the 3D digital world through the photo's capturing device using an Image-Based Rendering methodology; *(ii)* the **proposition** of a set of **geographic visualization techniques** based on the **user's visual perception** to provide a 3D interactive exploration in space and time. To integrate these two contributions and provide user capacities, we implemented the **HISTOVIS prototype**. It is a web-based system providing an interactive visual exploration for extensive collections of historical images from the ALEGORIA project. It combines multiple visualization and interaction techniques, *e.g.*, heat maps, image thumbnails, viewpoint markers, etc. Three **use case scenarios** (visualization of photo collections, street-view exploration through time, and browsing non-spatialized photographs in a 3D spatial environment) were tested in a **user study** and within the project to define and evaluate our prototype usability.

Keywords — geographic visualization, geometric distortion, Image-Based Rendering, visual perception, user interaction.

Résumé

Exploration Visuelle des Collections d'Images Historiques : *Une Approche Interactive à Travers l'Espace et le Temps*

Les représentations iconographiques, telles que les **photos historiques** d'espaces géographiques, sont des ressources précieuses du Patrimoine Culturel permettant la description de l'évolution d'une zone topographique au cours du temps. Avec l'avènement de l'ère numérique, nombre de ces documents ont été **numérisés**, **spatialisés** et sont désormais **disponibles en ligne**. Cependant, il reste **difficile** pour les **utilisateurs** de **parcourir** et de **visualiser** ces images dans l'espace et dans le temps. Le projet ANR ALEGORIA relève cet enjeu et s'oriente vers des collections photographiques retraçant le territoire français de l'entre-deux-guerres à nos jours, composées photographies aériennes et terrestres fournies par le consortium (Archives Nationales, Musée Nicéphore Niépce, Institut National de l'Information Géographique et Forestière).

Cette thèse présente notre proposition pour une **exploration visuelle** des **collections d'images historiques** en utilisant une **approche interactive** à travers l'**espace** et le **temps**. Nous proposons de co-visualiser des photographies historiques à l'intérieur de modèles topographiques 3D contemporains, et nos **deux contributions principales** sont : *(i)* l'**extrapolation** du **modèle de distorsion géométrique de la photographie** pour permettre aux utilisateurs de visualiser le monde numérique 3D à travers le dispositif de capture de la photo à l'aide d'une méthodologie de Rendu Basée sur l'Image ; *(ii)* la **proposition** d'un ensemble de **techniques de visualisation géographique** basées sur la **perception visuelle de l'utilisateur** pour fournir une exploration interactive en 3D dans l'espace et dans le temps. Pour intégrer ces deux contributions et pour offrir des capacités d'interaction utilisateur, nous avons mis en place le **prototype HISTOVIS**. Il s'agit d'un système Web offrant une exploration visuelle interactive de vastes collections d'images historiques du projet ALEGORIA. Il combine plusieurs techniques de visualisation et d'interaction, par exemple, cartes thermiques, vignettes d'images, marqueurs de point de vue, etc. Trois **scénarios d'usages** (visualisation de fonds photographiques, exploration temporelle au niveau de la rue et navigation entre des photographies non spatialisées dans un environment 3D) évalués par une **étude utilisateur** et dans le cadre du projet. pour définir et évaluer son utilisabilité.

Mots-clés — visualisation géographique, distorsion géométrique, Rendu Basé sur l'Image, perception visuelle, interaction utilisateur.

Glossary

Here we introduce to the reader our proposed main definitions concerning the understanding of this thesis. These terms have been defined as a compromise among existing formal definitions.

Numbers

3D City Model: a set of 3D urban objects, *e.g.*, ground, buildings, and trees, that digitally can represent a city.

3D Model: a mathematical representation of an object that can be reproduced digitally.

A

Attention (Visual): a set of cognitive processes allowing someone to concentrate on a piece of given information while ignoring others (Goldstein 2010; Snowden, Thompson, and Troscianko 2011).

B

Bookmark (Graphic): an interactive graphical object displayed in diverse forms, *e.g.*, a text, an icon, a thumbnail, or a 3D object (Forgione et al. 2016).

C

Camera: a light-tight box with a small aperture that allows light to come inside to capture an image into a light-sensitive surface like a photographic film or digital sensor.

Clustering: a process where elements are partitioned into a set of meaningful categories (Xu and Wunsch 2015).

Co-Visualization: is the visual integration of two or more data types or rendering styles. It requires a definition of the form of representation and an interface to do the integration.

Cultural Heritage: the legacy of artifacts, sites, and practices left or inherited from former periods of time. It covers everything that a society perceives as old, meaningful, and worthy of conserving (Brumann 2015).

D

Digital Earth: a digital copy of the Earth capturing all its details (Gore 1998).

Distortion (Image): a geometric deviation from the ideal pinhole camera.

E

Exploration: examine systematically a dataset by summarizing, plotting, or conducting statical analysis to arrive at insights and hypotheses (Çöltekin, Janetzko, and Fabrikant 2018).

G

Geographic Information: the representation of an object or spatial phenomenon in a conventional form to facilitate its processing and communication: *(i)* located in a geographic space; *(ii)* at a given moment; *(iii)* related to a reference system.

Geographic Phenomenon: a set of facts and characteristics observed and located in space.

Geographic Visualization: a set of knowledge and techniques that enables the visualization and analysis of a territory or a spatialized phenomenon while interacting with geospatial data.
Graphical User Interface: allows users to interact with digital environments through visual representations.

H
Historical Photograph: a documented photograph that has cultural and historical value.

I
Immersive: the perception of being present in the real world, but through digital proxy.
Information: data with interpretation and meaning.
Interaction: graphically represent the data in a form where it is improved the way the information is associated (Çöltekin, Janetzko, and Fabrikant 2018).

P
Perception: ability to perceive and be aware using the human senses.
Phenomenon: something that can be perceived or observed by experience and is likely to be repeated or reproduced.
Photograph: a reproduction of an image that a camera has captured.
Pinhole Camera: a camera model that represents the ideal camera with a perfect perspective projection.

R
Rendering: the process of generating an image employing a computation device.

S
Saliency: visual quality of any object, particularly noticeable or prominent.
Spatial: relating or having the nature of space.
Spatio-Temporal: data represented across both space and time.
Style Representation: a way in which a 3D environment is represented, *e.g.*, its contrasts, saliency, photo-realism, or abstraction level.

T
Temporal: relating or having the quality of time.
Thumbnail: a compressed and minor version of a visual element.
Timeline: a visual and interactive graphical object displaying a set of dates in a chronological order.

U
User: a human using a tool.

V
Visibility Heat Map: a 2D map where higher color intensities depict higher numbers.
Visual Complexity: a level of intricacy within a graphic representation.
Visual Perception: the interpretation and understanding of our surroundings through the light that enters the eye (Boynton 2005; Goldstein 2010).

Visual Variable: a feature in a graphical object that can make the object visually stand out from others (Bertin 2010; MacEachren, Roth, et al. 2012; Roth 2017).

Visualization: the act and process of putting something into a material and visible form.

Notations

Notation	Description
$\mathbf{c} = (c_x, c_y)$	Center of distortion.
$\mathbf{C} = (C_{X'}, C_{Y'}, C_{Z'})$	Camera optical center.
$D(r)$	Distortion function.
$D^{-1}_{rad}(r)$	Inverse function of $D_{rad}(r)$.
$D_{rad}(r)$	Radial distortion function.
$\overline{D}_{rad}(r)$	Distortion function defined by $D_{rad}(r)$ and an extrapolation function.
$D'_{rad}(r)$	Derivative function of $D_{rad}(r)$.
f	Focal length.
H	Internal parameters of a camera.
K_i	Radial distortion coefficients.
$\mathbf{O} = (O_X, O_Y, O_Z)$	World origin.
P	Projection matrix converting a 3D point into a 2D point.
$\mathbf{p} = (p_u, p_v)$	Principal point.
$\mathbf{q} = (q_u, q_v)$	Scale factors.
R	Rotation matrix describing the orientation of the camera.
$r = \lvert\mathbf{x} - \mathbf{c}\rvert$	Radial distance from the center of distortion $\mathbf{c}$ to the ideal point $\mathbf{x}$.
$r' = \lvert\mathbf{x}' - \mathbf{c}\rvert$	Radial distance from the center of distortion $\mathbf{c}$ to the observed point $\mathbf{x}'$.
r_{ext}	Radial distance from the center of distortion $\mathbf{c}$ to the extrapolation point.
r_{img}	Image radial distance from the center of distortion $\mathbf{c}$.
r_{max}	Maximum radial distance for $D_{rad}(r)$.
s	Skew factor.
t	Time.
$\mathbf{t} = (t_X, t_Y, t_Z)$	Translation vector from camera to world origin.
$\mathbf{x} = (x, y)$	2D ideal image coordinate point.
$\mathbf{X} = (X, Y, Z)$	3D world coordinate point.
$\mathbf{x}' = (x', y')$	2D observed image coordinate point.
$\mathbf{X}' = (X', Y', Z')$	3D world coordinate point defined in the camera frame.

Abbreviations

ALEGORIA	Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge.
AN	National Archives.
ANN	Approximate Nearest Neighbor.
ANR	French National Research Agency.
AR	Augmented Reality.
BBA	Bundle Block Adjustment.
CH	Cultural Heritage.
CIM	Combier printer In Mâcon.
CRS	Coordinate Reference System.
DeepMPIs	Deep Multiplane Images.
DLT	Direct Linear Transformation.
DSM	Digital Surface Model.
DTM	Digital Terrain Model.
ENSG	French National School of Geographic Science.
GIS	Geographic Information Systems.
GLAMs	Galleries, Libraries, Archives, and Museums.
GPS	Global Positioning System.
GPU	Graphics Processing Unit.
GSD	Ground Sampling Distance.
GUI	Graphical User Interface.
HCI	Human-Computer Interaction.
HMD	Head Mounted Devices.
IBR	Image-Based Rendering.
IGN	French National Mapping Agency.
ISO	International Organisation for Standardisation.
LASTIG	Laboratory of Geographic Information Science for Sustainable Development and Smart Cities.
LIDAR	Light Detection and Ranging.
LOD	Level of Detail.
LUT	Lookup Table.
MR	Mixed Reality.
MRU	Ministry of Reconstruction and Town Planning.
NNM	Nicéphore Niépce Museum.
PBR	Physically-Based Rendering.
RGB	Red Blue Green.
SEM	Saccadic Eye Movement.
SGM	Semi Global Matching.
SIFT	Scale Invariant Feature Transform.
UML	Unified Modeling Language.

VR	Virtual Reality.
XR	Extended Reality.

Contents

Abstract/Résumé **vii**

Glossary **ix**

Notations & Abbreviations **xii**

1 Outline of Work **1**
1.1 Context: Exploration of Historical Photographs 1
1.2 Problem Statement in Geovisualization 2
1.3 Objectives of Research 3
1.4 Background and Intended Public: ALEGORIA 5
1.5 Structure of Manuscript 6

2 Background and Related Work **7**
2.1 Cultural Heritage: Photography 8
2.1.1 What is a Photograph? 9
2.1.2 Photographs and Spatial Attributes 13
2.1.3 Geographic Imagery 17
2.2 Image Exploration for Extensive Collections 20
2.2.1 Metadata Browsing 20
2.2.2 Spatial and Temporal Exploration 21
2.2.3 Use of Extended Reality 30
2.3 Image-Based Rendering 33
2.3.1 Projective Texturing 35
2.3.2 Interpolating Between Views 36
2.3.3 Artifact Correction 38
2.4 Strategies to Ease and Improve Navigation 43
2.4.1 Visual Perception 44
2.4.2 Attention 45
2.4.3 Ease of Navigation 47

3 Formalization of the Problem **52**
3.1 Problem Description 52
3.1.1 Sub-Problem 1: The Integration of Heterogeneous Spatial Information Causes Potential Visual Discomfort 55
3.1.2 Sub-Problem 2: The Navigation Through a Massive Amount of Visual and Spatio-Temporal Data Represented on a Large Scale 56
3.2 Objectives of Research 57
3.3 Proposed Approach 58
3.3.1 Approach 1: Extrapolation on the Geometric Distortion for 3D Visualization 61
3.3.2 Approach 2: Navigation Techniques to Reduce Visual Complexity 61

3.3.3 Approach 3: Co-Visualization of Historical Photos and 3D City Models . . 62

4 Extrapolation of the Geometric Distortion for 3D Visualization 63
4.1 Introduction . . . 64
4.2 Rendering with Distortion . . . 67
4.2.1 One-Pass Rendering . . . 67
4.2.2 Two-Pass Rendering . . . 68
4.3 Extrapolating Distortion Models for 3D Visualization . . . 69
4.3.1 Existent Radial Distortion Models . . . 70
4.3.2 Our Extrapolation Approach . . . 72
4.4 Results: Testing Our Extrapolated Distortion Model . . . 73
4.4.1 Description of the Datasets . . . 74
4.4.2 Test 1: Use of Different Camera Models for Projective Texturing . . . 75
4.4.3 Test 2: Variation on the Extrapolation Radius . . . 76
4.4.4 Test 3: One-Pass vs. Two-Pass Rendering . . . 77
4.5 Conclusion . . . 79

5 Navigation Techniques to Reduce Visual Complexity 80
5.1 Introduction . . . 81
5.2 User Tasks . . . 83
5.2.1 Task 1: Visualizing One or Several Images . . . 83
5.2.2 Task 2: Exploring Images in Space and Time . . . 84
5.3 Design Guidelines . . . 85
5.3.1 Design Guideline 1: How to Represent a Photograph? . . . 85
5.3.2 Design Guideline 2: How to Make Salient Certain Elements? . . . 88
5.3.3 Design Guideline 3: How to Add the Temporal Aspect? . . . 91
5.4 Results: Testing Our Design Guidelines . . . 92
5.4.1 Test 1: Bookmarks and Thumbnails for Interaction with Photographs . . . 93
5.4.2 Test 2: Image Frames, Visibility Heat Map, and Style Representation for Saliency . . . 95
5.4.3 Test 3: Timeline for Temporal Filtering . . . 98
5.5 Conclusion . . . 99

6 HISTOVIS: Co-Visualization of Historical Photos and 3D City Models 100
6.1 Presentation of the Historical Data Collections . . . 101
6.1.1 National Archives (AN) . . . 102
6.1.2 Nicéphore Niépce Museum (NNM) . . . 103
6.1.3 French National Mapping Agency (IGN) . . . 104
6.2 Modeling a Historical Photograph Inside an Interactive System . . . 107
6.2.1 Interaction States of an Image . . . 107
6.2.2 Model of a Historical Image . . . 108
6.3 Architecture and Pipeline of HISTOVIS . . . 108
6.4 Usage of HISTOVIS . . . 114
6.4.1 System and Initial Configuration . . . 114
6.4.2 User Interface and Functionalities . . . 116

6.4.3 Use Case Scenarios 124
6.5 Evaluation 130
6.5.1 User Study on Scenario 1: Visualization of Two Photo Collections 130
6.5.2 ALEGORIA Stakeholders Feedback 135
6.6 Conclusion 137

General Conclusions and Perspectives **138**

A Mock-Ups: HISTOVIS Prototype **143**

B Test Protocol: HISTOVIS Prototype **147**

C Test Results: HISTOVIS Prototype **154**

Bibliography **160**

List of Publications **174**

List of Figures **178**

List of Tables **179**

Chapter 1

Outline of Work

Preamble

The following chapter **schematizes this dissertation** entitled: **Visual Exploration of Historical Image Collections** (*An Interactive Approach Through Space and Time*).

It introduces to the reader our research context as the visual exploration of historical photographs in Section 1.1. Then, Section 1.2 describes the challenges we face in geographic visualization, Image-Based Rendering, and Human-Computer Interaction concerning this context. Section 1.3 summarizes the objectives of this thesis to solve these issues. Section 1.4 specifies our research environment (the ALEGORIA project) and the intended public. Finally, Section 1.5 displays the organization of this manuscript.

Contents

1.1 Context: Exploration of Historical Photographs **1**
1.2 Problem Statement in Geovisualization **2**
1.3 Objectives of Research **3**
1.4 Background and Intended Public: ALEGORIA **5**
1.5 Structure of Manuscript **6**

1.1 Context: Exploration of Historical Photographs

Safeguarding and exploiting Cultural Heritage (CH) documents, like **historical photographs** in the form of postcards, engravings, paintings, street-level/aerial shots, produce numerous data. The management of these records is an essential task for the study and diffusion of their content. Previously, all data handling was a handmade routine done through experienced methods[1], *e.g.*, keeping the photographs at low temperature, low humidity, and unique storage.

The **digital era** presented **new archiving practices** and allowed access to widespread information. With the growth of computational-oriented methods for CH archives, many extensive historical data collections have been managed through digital preservation treatments. It has changed the way CH libraries and archives are conceived, designed, and used. In addition to a paper version, a

[1] Many methods exist to make sure photographs are preserved. Here is an example of procedures for simple family photos:
http://www.archives.gov/preservation/family-archives

digital copy protects the content and offers new exploitation possibilities, like immediate connection and analysis (Ardissono, Kuflik, and Petrelli 2012; Baruzzo et al. 2009; Meyer et al. 2007).

Notwithstanding, many **photographic collections** are **currently exploited** in a **low-level interaction** by their central users, *e.g.*, researchers, institutions, and local communities, and their awareness is almost zero by a broader public. Overall, these photos are **scattered within different institutions**, **partially digitized**, and generally **poorly spatially documented**. In some cases, only a physical consultation may be possible, *i.e.*, *in situ*, which may be extended by remote access through a traditional online approach.

The effort to **digitalize** the **exploration** of these **visual resources** has evolved through the years. A digital photo gallery is commonly the simple solution employed, which adds meta-data, *i.e.*, annotations and descriptions, to each picture to find the desired items. A well-known example is *Google Images*. Nevertheless, an alternative approach known as **geovisualization** (short for geographic visualization) has emerged, allowing exploring and interacting with geospatial data more profoundly. It highlights the examination of the information but also the sensemaking of it. The concept of geovisualization is related to a process rather than a product; however, the term can be commonly applied to any visual element that features geospatial information, *e.g.*, images, maps, and 3D city models (Çöltekin, Griffin, et al. 2020; Çöltekin, Janetzko, and Fabrikant 2018).

Solutions like *Historypin* (Armstrong 2012; Bolter, Engberg, and MacIntyre 2013), *Photo Tourism* (Snavely, S. Seitz, and Szeliski 2006), *PhotoCloud* (Brivio et al. 2013), and *Smapshot* (Blanc, Produit, and Ingensand 2018; Produit et al. 2018), are oriented toward a geovisualization process. These tools use a **spatial** and **temporal representation** since, nowadays, attributes of images can be estimated, such as the position and orientation of the acquiring camera. These inventive and collaborative approaches facilitate locating the old photographs in a 2D or 3D topographic view, allowing users to have a higher and more **realistic level** of **interaction** with the photos.

Still, compilations of **historical photographs** are a **particular input**. These photos are likely to present: *(i)* **diversity** on the source leading to many different types and styles of pictures, *e.g.*, paintings, postcards, and old aerial/terrestrial imagery; *(ii)* **temporalities** on the photographed scene, *e.g.*, diachronism and evolution of the scene objects over time; *(iii)* conditions of **acquisition** and **digitization**, *e.g.*, illumination, the image acquiring technology, and resolution of scanning. Therefore, many existing geovisualization techniques and tools may not appropriately adapt to these historical datasets, and **novel approaches may need to be followed**.

1.2 Problem Statement in Geovisualization

Any **geovisualization approach seeking a spatial and temporal browsing method for historical photos** will face certain obstacles associated with this input data. Among them, we can mention:

Data Uncertainties: a historical image is a digitized version of an original photograph, most of the time fixed on a film (analog process) and converted into a digital version by a scanner. Consequently, the characteristics describing the camera that captured this photo may be

unknown. Unlike modern remote sensing imagery, where the data is well registered and georeferenced, historical images are only partially documented.

Sparse Samples: even if a set of photographs captures the same space, locating them spatially or reproducing a digital model of the photographed scene may not be possible. Less pictures were acquired in the past; therefore, fewer overlapping areas and different acquisition perspectives can be observed in these compilations.

Image Artifacts: the capturing devices used to acquire historical photographs may possess a diversity of effects, *e.g.*, geometric distortion, chromatic aberration, and depth of field variation, that influenced the quality of the resulting final picture. When spatially visualizing the photographs, these artifacts (any feature that appears inside an image that was not present in the original object or scene) may impact the visualized result.

Extensive Scale: pictures depicting geographic places may represent vast terrestrial scales. A spatial exploration through historical photographs has to support extensive and continuous space discovery (all possible topographic places).

Large Volumes: the amount of visual data to be explored is enormous. Many institutions involving scholars, historians, archivists, and national mapping agencies, among others, have been and still are archiving large numbers of historical images, resulting in a high volume of items.

Accordingly, if we aim to discover historical images, we will have to support the needs of this type of information. Along with it, if another type of data is included, *e.g.*, modern 3D city models, temporal information, and visual user interaction objects, **our geovisualization system must integrate all data effortlessly**. It should promote the visualization of this geospatial information so that users can interact with these representations easily and freely without considering if the information is historical or contemporary.

Research Problem

We define our general research challenge as: **how to associate the diverse type of spatio-temporal information into one geovisualization system?**

Hence trying to solve this research question related to the browsing and exploration of different and vast geospatial information may raise several difficulties, which leads to our two specific research questions (Section 3.1 of Chapter 3 will dive more deeply into these topics):

1. **How to decrease the potential visual discomfort caused by the integration of heterogeneous spatial information?**
2. **How to continuously navigate through a massive amount of visual and spatio-temporal data represented on a large scale?**

1.3 Objectives of Research

We want to enable the **discovery of historical photo collections through space and time**. A 3D city model can be used for this purpose. It is a set of 3D urban objects, *e.g.*, ground surfaces and buildings, that allow us to represent a geographic space digitally. As depicted in Figure 1.1, we

aim to browse photographic compilations inside a 3D city (and topographic) model. We want to give users an immersive feeling to **understand the representation of the environment** in which the photos are being visualized and the **spatio-temporal relationships** between images. This exploration will be adapted to the characteristics that historical datasets may present.

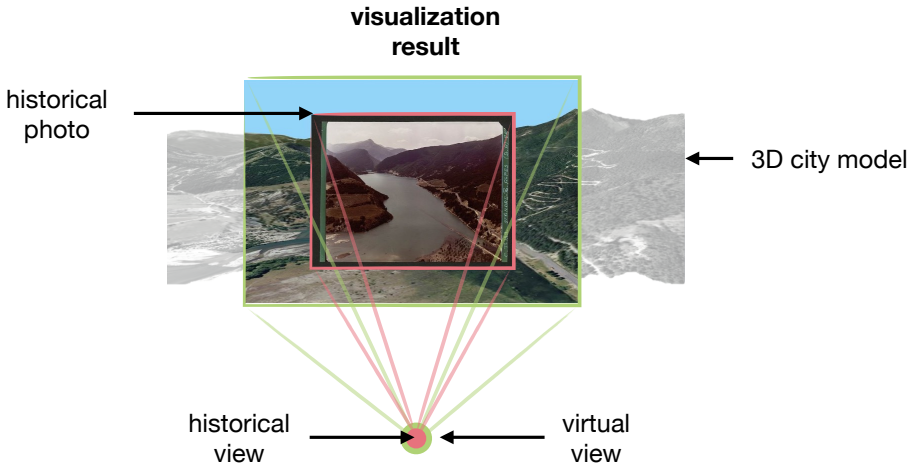


Figure 1.1: Virtual view of a 3D model from a historical photo camera location. Credits to National Archives (AN)/LAPIE photographic collection.

In this context, this dissertation involves the proposition of a new and innovative approach for a **spatio-temporal**, **immersive**, and **interactive navigation** of a 3D environment enriched with historical photographs. We want to help users examine many different images through a method that will ease the exploration process through all the displayed visual data and improve the visualized result. Section 3.2 of Chapter 3 will explain more in detail our specific goals of this research, but in summary, these are:

1. **An artifact correction on the rendering side for visualizing historical photographs inside a 3D environment.**
2. **A graphic representation for the data related to enhancing the user's navigation and interaction.**

The primary model of the study is Image-Based Rendering (IBR) because of its capacity to use imprecise or non-existent scene geometry. The first objective will be solved using this rendering process (artifact correction). We will try to answer the second goal using design guidelines to sketch an innovative navigation process. Many of these guidelines will be later implemented and tested. Our proposed approaches will be better defined in Chapter 4, Chapter 5, and Chapter 6.

Contributions

From the previously mentioned objectives and strategies to achieve these goals, the main contributions of this work are:

1. **The extrapolation of radial distortion models for rendering historical views**.

2. **The proposition of a set of techniques to support interactive visual exploration**.
3. **A framework for the co-visualization of historical photos and contemporary 3D city models**.

1.4 Background and Intended Public: ALEGORIA

Observation

Although this research is generally aimed at historical images, the final prototype will be tested using iconographic collections from particular French institutions. The prototype will also be directed at a specific group of users belonging to these establishments.

This research work has been carried at the French National Mapping Agency (IGN)[2], within the GeoVIS[3] team, from the Laboratory of Geographic Information Science for Sustainable Development and Smart Cities (LASTIG)[4]. The team's focus is the current subjects related to **geovisualization** and **interaction** with spatial data for visual spatio-temporal analytics. The financing comes from the French National Research Agency (ANR)[5], in the framework of the Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge (ALEGORIA)[6] program (2018-2022).

The ALEGORIA project aims to enhance **geographic imagery** describing the French territory at different times ranging from the interwar period to the present times. These images consist of mostly oblique and vertical aerial photographs. Terrestrial views are likewise available in less quantity. For a complete description of the image compilations and cooperating institutions of the ALEGORIA program, please refer to Section 6.1 of Chapter 6. We seek for **new techniques** to **enhance** all of the **photograph's heritage**. Their large number makes manual evaluations difficult; consequently, a digital path is essential. Primary users are interested in a cross-analysis between images and photo collections. Precisely, an **online context** (web-based application) enables all participating establishments to access their own and other institutions' photo compilations.

To allow the photographic collections to be exploited and enhanced, three Galleries, Libraries, Archives, and Museums (GLAMs) are working together:

1. The National Archives (AN)[7].
2. The Nicéphore Niépce Museum (NNM)[8].
3. The French National Mapping Agency (IGN).

As a whole, the ALEGORIA project's focus is the indexing, interlinking, and visualization of these historical collections. This dissertation centers on the latest point. The design of the system targets

[2] https://www.ign.fr
[3] https://www.umr-lastig.fr/geovis/
[4] https://www.umr-lastig.fr/
[5] https://anr.fr
[6] https://www.alegoria-project.fr
[7] https://www.archives-nationales.culture.gouv.fr
[8] https://www.museeniepce.com

archivists and **humanities researchers** as the **intended users**. However, the application may be accessible to the general public; therefore, they are also considered as final users of the desired geovisualization system.

1.5 Structure of Manuscript

From this foundation, we will now, in the coming chapters: *(i)* describe the current state of the art related to photographs and their exploration; *(ii)* formalize our research problem: the association of diverse types of spatio-temporal information; *(iii)* describe each of our proposed approaches: artifact correction and graphic representation; *(iv)* present the evaluation and results in the form of a prototype. The rest of this document is organized as follows:

Structure of the Document	
Chapter 2	identifies essential knowledge to help the reader understand the main topics (photography, exploration of geographic photo collections, Image-Based Rendering, and perception) and existing research challenges concerning this thesis purpose.
Chapter 3	presents a formalization of the problem (the association of various types of spatio-temporal data) from analyzing the research context and the existing issues related to it. We want to propose a method to help users visualize and explore extensive and diverse collections of historical photos. Therefore the following two chapters provide our distinct propositions to this general objective.
Chapter 4	examines singularly the subject of geometric distortion affecting the visualization of photographs in a 3D environment. We describe a distortion model that extends the distortion function definition to all the image domain. Therefore this model can be used even with a zoomed view.
Chapter 5	addresses the topic of user interactions considering the potential volume of the input data. We propose particular design guidelines that rely on navigation techniques that may ease the visual complexity and improve saliency on specific objects.
Chapter 6	describes in more detail the experimental input data associated with the research context of historical photos. The chapter exhibits the implementation of our propositions in the form of a web-based application. We define different use cases to show the variety of possible visualization and navigation interactions. Finally, it describes our evaluation offered to users to check if our objectives were achieved.

Chapter 2

Background and Related Work

Preamble

The following chapter provides the **general framework on this dissertation's main topics**: geospatial data in the form of photographs, the digital exploration of these photos, and how visual perception can ease the navigation through them. It presents to the reader essential notions necessary for the coming chapters related to our proposed approaches regarding the discovery of historical photographs. The purpose is to describe the current state of the art.

Section 2.1 specifies the main concepts related to photography and its evolution. Next, Section 2.2 summarizes several existing methods that allow the visualization of extensive photo collections. Thus many different visualization techniques are illustrated; Section 2.3 focuses on Image-Based Rendering. Finally, Section 2.4 shows how navigation over numerous visual data can be handled following specific strategies.

Contents

2.1 Cultural Heritage: Photography **8**
2.1.1 What is a Photograph? . . . 9
2.1.2 Photographs and Spatial Attributes . . . 13
2.1.3 Geographic Imagery . . . 17
2.2 Image Exploration for Extensive Collections **20**
2.2.1 Metadata Browsing . . . 20
2.2.2 Spatial and Temporal Exploration . . . 21
2.2.3 Use of Extended Reality . . . 30
2.3 Image-Based Rendering **33**
2.3.1 Projective Texturing . . . 35
2.3.2 Interpolating Between Views . . . 36
2.3.3 Artifact Correction . . . 38
2.4 Strategies to Ease and Improve Navigation **43**
2.4.1 Visual Perception . . . 44
2.4.2 Attention . . . 45
2.4.3 Ease of Navigation . . . 47

2.1 Cultural Heritage: Photography

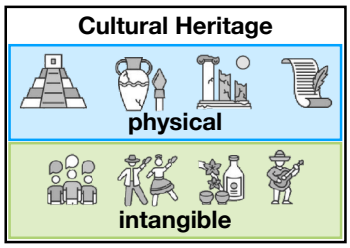


Figure 2.1: Cultural Heritage. It is divided into physical and intangible.

Heritage is defined as something that is inherited, passed down from previous generations. Hence **Cultural Heritage (CH)** can be seen as the **legacy of artifacts, sites, and practices** left or inherited from former periods (of time). It covers everything that a **society perceives as old, meaningful, and worthy of conserving** (Brumann 2015). As illustrated by Figure 2.1, the concept of CH is much broader now, considering both physical artifacts and intangible attributes that can be preserved (Willis 2014). Among them, we can mention:

Physical Artifacts: archaeological and historical objects, buildings, monuments, historical places, works of art, literature, and music.

Intangible Attributes: traditions, social customs, oral folklore, and practices grounded in aesthetic and spiritual beliefs.

The **conservation** and **protection** of CH is a crucial issue. One topic worth mentioning is the interest of the individuals of a community to preserve it. As the diagram proposed by Thurley (2005) shows (see Figure 2.2), the **heritage cycle** can give us an idea of how the past can become part of our future. By understanding Cultural Heritage, people can value it, which leads to taking care of the elements of this CH, allowing the individuals to enjoy it, bringing back the idea of trying to understand it (we come back to the start).

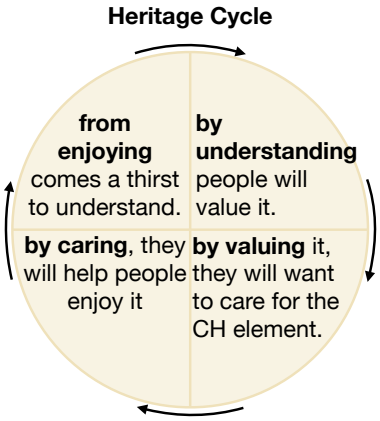


Figure 2.2: The heritage cycle (understand, value, care, and enjoy). The diagram is adapted from Thurley (2005).

On the subject of physical objects, Cultural Heritage **artifacts** are **unique** and **irreplaceable**. Minor artifacts such as sculptures, paintings, pottery, armor, and coins are primarily preserved in museums and art galleries. Monuments, buildings, and historical places are often subject to preservation orders and regulations by the government to ensure their survival for future generations (Willis 2014). However, there is an **agent impossible to overcome** that is responsible for CH destruction; **time**. Many of these physical artifacts, *e.g.*, archeological sites, may not remain in the same conditions as they are today with the passing of the years. It leads to the **destruction** and **disappearance** of historical elements, like buildings and monuments, causing irreparable damage to Cultural Heritage.

An essential tool that has helped the long-term preservation of CH is **photography**. While conditions may change over time, images help document and preserve a particular object/moment in time. Therefore, we may say that **photographs depict visual samples of past knowledge**. Cultural Heritage photography in the film-based period proved to be very successful, *e.g.*, recording events during World War II like the destruction of many buildings.[1] The advent of digital photography in the 1990s opened up an entirely new world by bringing the benefits of acquiring, manipulating, storing, and retrieving imagery with much more simplicity/efficiency (Verhoeven 2016). Overall, photography is an essential instrument used since its invention; however, **how do we define its process and result (a photograph)?**

Writing Remark

In this document, we will consider **old** and **historical** as synonyms for simplicity in writing. Please note that for something to be *historic* it has to be associated with a degree of importance, while *old* implies it has been created (or occurred) a long time ago.

2.1.1 What is a Photograph?

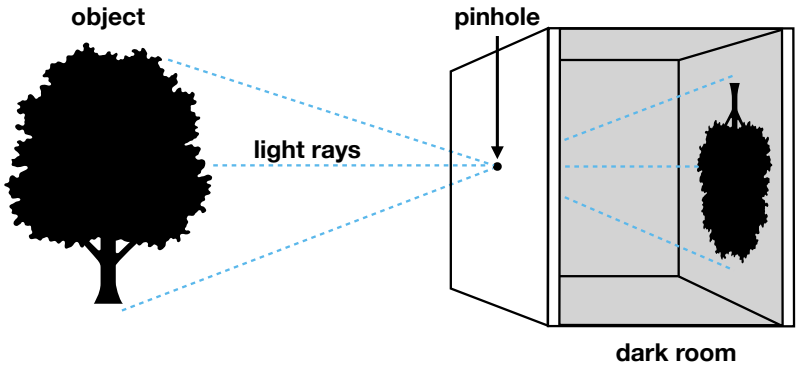


Figure 2.3: The pinhole camera. Light enters a dark room through a hole and forms an inverted image on the opposite side of the opening.

The science (and the art) of capturing light with an optical device, *i.e.*, a camera, to create an image is known as **photography**. This concept was born with the formalization of the *camera obscura* in the 11th century (Howard 1996). Its basis is the use of a darkened room with a tiny hole in the wall to allow light to pass through and form the (upside-down) image of the outside scene. Back then, no printing was possible; images were only projected onto surfaces as an aid to create an accurate drawing of real-world targets, *e.g.*, buildings, trees, and animals. As shown in Figure 2.3, this simplistic camera representation

[1] Here are some sample photographs of the Nicéphore Niépce Museum (NNM) depicting elements like soldiers, weapons, and the destruction of buildings from World War II:
https://www.open-museeniepce.com/recherche-photos?resetSearch=1&addMotcle=Deuxième%20Guerre%20Mondiale

is known as the **pinhole camera**, *i.e.*, lensless and with a single small aperture. In the 17th century, the design of cameras changed with the introduction of portable cameras. Around this time, lenses were also incorporated, which allowed focusing the light (Lefèvre 2007; Steadman 2002).

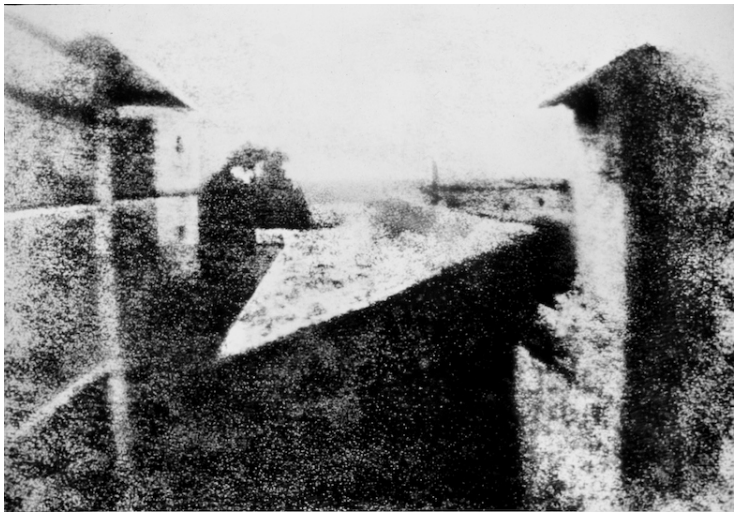


Figure 2.4: *View from the Window at Le Gras*. Credits to Nicéphore Niépce (reproduction of the original plate). The photograph is taken from Marien Warner (2006).

The notion of photography as it is known nowadays started in France in the late 1830s. The first permanent image was capture by Joseph Nicéphore Niépce. Figure 2.4 exhibits the result, where after hours of exposure, it depicts the roof of a building illuminated by the sun (Marien Warner 2006). Since then, as displayed in Figure 2.5, the basic idea of a **camera** have remained pretty much the same; a light-tight box (*i.e.*, the camera body) with a small hole (*i.e.*, the aperture) that allows light to come inside so an image can be captured in a light-sensitive surface (a photographic film or digital sensor). However, many other features may be added to improve the result, *e.g.*, the camera may employ mirrors to correct the flipped image effect (also known as a reflex camera). Among several different elements worth mentioning, the most conventional cameras have (Verhoeven 2016):

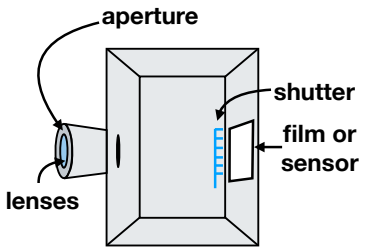


Figure 2.5: Primary components inside a camera: lenses, aperture, shutter and film/sensor.

- A **shutter** to control the exposure time (can be rolling or global). This device opens and closes, allowing light to reach the photosensitive surface only for a specific period of time. This duration is known as the **shutter speed**, measured in seconds (or fractions of a second).
- Several **lenses** to compensate for aberrations. Lenses are designed to focus the light over the (light-sensitive) surface. A perfect lens would produce a perfect image; however, this is not the case in real life. The defects (or artifacts) that the lens may present are known as aberrations, including optical

distortions and chromatic aberrations. The use of various lenses is required to correct as many artifacts as possible.

- An adjustable **aperture** to influence the amount of light and the sharpness. Overall, the aperture acts like the pupil in the human eye, allowing light to pass through the lens(es). Since working with a physical size is difficult, it is commonly measured using f-stop, which is a ratio of the lens's focal length to the diameter of the aperture.
- An **ISO** to measure the sensitivity. It is a value that expresses how sensitive is a film or a sensor to incoming light. The fundamental value standardized by the International Organisation for Standardisation (ISO) is ISO 100. Higher values of ISO mean that the photosensitive surface does not need to gather as much light to make a correct exposure.

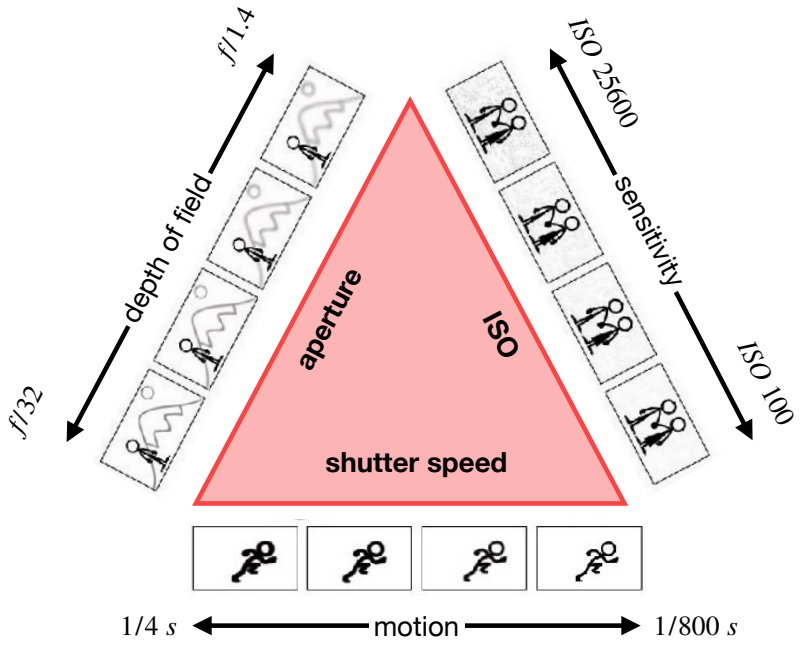


Figure 2.6: The exposure triangle: aperture, shutter speed, and ISO must be balanced in the photographic process. The diagram has been adapted from Verhoeven (2016).

The previously mentioned components can **determine** the **global appearance** of the **resulting image** in a photographic process. An example of the interrelationship between them is the triangle of exposure displayed in Figure 2.6, which showcases the connection between the aperture, shutter speed, and ISO. To create a photograph that can be seen (and perceived), the amount of light exposed to the photosensitive surface must be controlled (shutter and aperture), along with the sensitivity of the surface to the light (ISO). For example, a smaller aperture gives a deeper field depth, a higher ISO can produce noise, or a slower shutter presents a motion blur effect.

Accordingly, to our best knowledge, we define a **photograph** (or photo for simplicity) as the reproduction, *i.e.*, either fixed on a film or digital, of an image that a camera has captured on a light-sensitive material.

Observation

Note that the terms **photograph**, **photo**, **image**, and **picture** will be used in this document as synonyms for uniformity. All of them will relate to this definition even though they may be related to other applications.

From Analog to Digital Photography

Analog photography uses chemical processes to capture a picture, *e.g.*, on a rigid plate or a film. In the 1880s, George Eastman founded the Kodak company, which introduced flexible roll film instead of solid plates requiring being constantly changed (Jenkins 1975). A film uses particles of silver salts embedded in a physical substrate, like acetate or gelatin. These grains are turned dark when exposed to light, and a chemical fixer makes this change (more or less) permanent. From this development process, the photograph can be later printed.



Figure 2.7: Comparison between analog and digital photography. Film (used in analog cameras) employs disperse grains of silver, and digital sensors (related to digital cameras) utilize pixels to represent an image. The differentiating photograph has been captured in Chichén Itzá (Mexico).

The early 1990s brought an exciting revolution in photography with the advent of digital technology. Instead of utilizing grains of silver, **digital photography** uses silicon to record images as numbers, *i.e.*, an integrated circuit creates light-sensitive elements called pixels onto the silicon surface. The captured pictures are digitized and directly stored as a digital file (Dzenko 2009; Reichmann 2006). As Figure 2.7 displays, unlike the film grains, which are spaced irregularly, pixels are organized in rows and columns. Still, any analog photograph can be easily shifted into a digital format. This **digitization** process has expanded during the past decades, particularly for historical pictures captured by analog cameras.

Digitization

As modeled by the diagram in Figure 2.8, digitization **converts any original document into a digital format**. This process typically involves using either a digital camera or a scanner, creating a numerical object. One of its main advantages is increased of access and flexibility, *e.g.*, various users can visualize the material concurrently and from different locations. In general, digitization can preserve too valuable or fragile elements in comparison with regular manual handling (Astle and Muir 2002; Conway 2008, 2009).

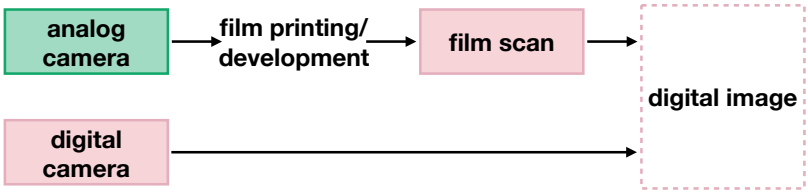


Figure 2.8: Process of digitization. When an analog camera captures an image, the resulting photo must be scanned to obtain a digital version.

This additional process (digitization) may introduce artifacts over the new numerical image, *e.g.*, geometric deformations due to the stretching of the photograph/film because of its usage or the mechanical deformities that the camera or scanner may include during the digitizing step. Figure 2.9 exhibits an example of a digitized photo; it can be seen that even if the original film was rectangular, the digitized version introduced a slight deformation, the folding of a corner (top right), and a black border.



Figure 2.9: Digitized image sample. Credits to Nicéphore Niépce Museum/Combier photo collection.

2.1.2 Photographs and Spatial Attributes

Photogrammetry is the science (art and technology) that **measures with light**, *i.e.*, estimating objects' geometric properties, *e.g.*, position, orientation, shape, and size, based on images (Collier 2009; Förstner and Wrobel 2016). Aimé Laussedat firstly introduced it on the idea of using the perspective drawings as metric documents in the 1840s (Polidori 2020). It has evolved, along with photography, from analog optical-mechanical strategies to more analytical methods based on computer-aided approaches to ultimately digital imagery. Two prevailing types exist *(i)* **aerial** (*e.g.*, taking shots from airplanes or drones), which places the camera in the air; *(ii)* **terrestrial** (*e.g.*, capturing the photos in the ground or a moving vehicle), which positions the camera on a tripod or simply is handheld (Aber, Marzolff, and Ries 2010b).

The **photogrammetric analysis** is a **sequential pipeline**, as depicted by our diagram in Figure 2.10; given a set of images (and possible background knowledge about the scene or the camera, *e.g.*, ground control points or pre-calibration values), the description of the objects in the photographed scene can be recovered. It is a passage from a 2D representation of the world captured by a camera, through the inference of the pose of that camera at the moment of the image taking, towards a 3D restitution of the lost dimension (Rupnik, Daakir, and Pierrot Deseilligny 2017). We have divided this process into three main steps:

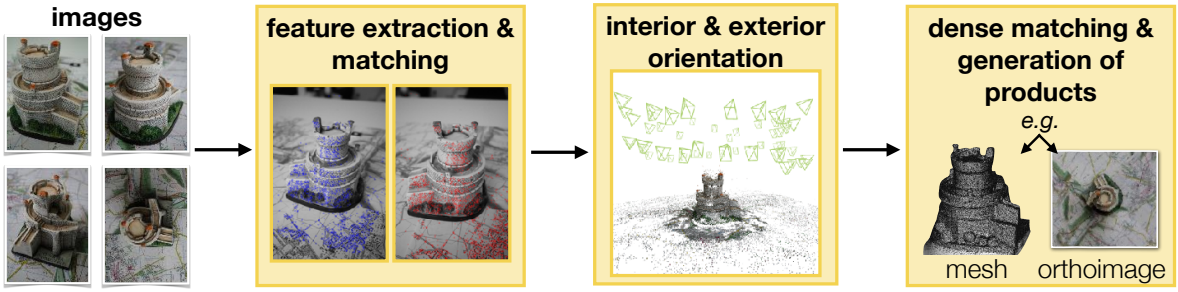


Figure 2.10: Process of photogrammetry (main steps): *(i)* feature extraction & matching; *(ii)* interior & exterior orientation; *(iii)* dense matching & generation of products. Images depicting Tvrđava Minčeta tower (Dubrovnik).

Feature Extraction & Matching: identical characteristics (*e.g.*, tie points) identified on the pictures and matched together.

Interior & Exterior Orientation: the mathematical description of the camera. It considers intrinsic parameters (internal properties, *e.g.*, focal length and lens distortion) and extrinsic parameters (position and orientation in the scene's frame).

Dense Matching & Generation of Products: constructs a dense correlation using depth maps. The result can be used to generate, for example, an orthoimage (rectified photo), a dense point cloud, or a Digital Surface Model (DSM).

Overall, photogrammetry can be used in different applications, which commonly involve: *(i)* estimation of coordinates, heights, distances, and areas; *(ii)* elaboration of topographic mapping; *(iii)* generation of digital models and orthoimages (Aber, Marzolff, and Ries 2010b; Kraus 2011). This section will focus on the global aspect of **geometric image analysis** in the photogrammetric process, particularly for camera orientation, calibration, and scene reconstruction.

Camera Geometry

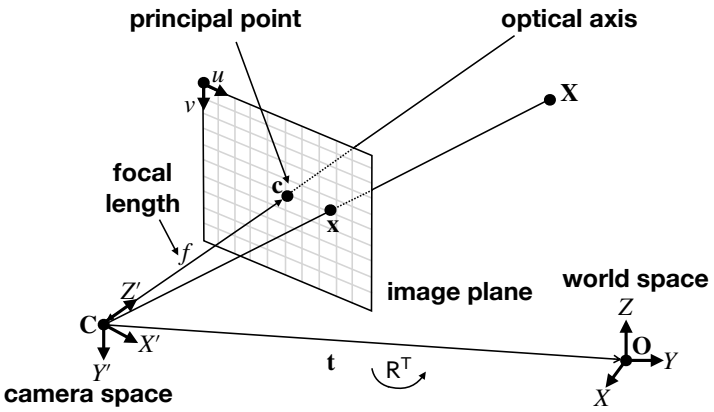


Figure 2.11: The pinhole camera model (geometric representation). Diagram adapted from Moulon (2014).

A **camera model**, as geometrically illustrated in Figure 2.11, is a mathematical description of a camera. It is defined by the camera's **intrinsic** and **extrinsic** parameters. The most used model is

known as the **pinhole camera model**, which represents a projective camera without any lens. It has a central projection, which allows all rays of light to pass through this point. The model can express a linear mapping of a 3D point to 2D using the simple form:

$$\mathbf{x} = \mathsf{H}\left(\mathsf{P}\left(\mathsf{R}\left(\mathbf{X} - \mathbf{C}\right)\right)\right) \tag{2.1}$$

where $\mathbf{x}$ is a 2D pixel coordinate of the image, $\mathbf{X}$ is the corresponding 3D world coordinate, $\mathbf{C}$ and R are the pose (position and orientation) of the camera, P is a projection from $\mathbb{R}^3 \rightarrow \mathbb{R}^2$, and H is an internal transformation related to the camera from $\mathbb{R}^2 \rightarrow \mathbb{R}^2$. Based on Equation 2.1, we define the transformation of the point $\mathbf{X}$ in $\mathbb{R}^3$ into an image point $\mathbf{x}$ in $\mathbb{R}^2$ through two operations (Förstner and Wrobel 2016; Moulon 2014):

A Change of reference from the world space to the camera space. Let $\mathbf{X}'$ be a world point defined in the frame of the camera, then:

$$\mathbf{X}' = \begin{bmatrix} \mathsf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \mathbf{X} \tag{2.2}$$

This relationship depends on 6 degrees of freedom named the extrinsic parameters:

- Three degrees represent the orientation of the camera, defined by a 3×3 rotation matrix R.
- Three degrees represent the position of the world origin $\mathbf{O}$ in the camera frame, defined by a translation vector $\mathbf{t}$. The position $\mathbf{C}$ of the camera's optical center is, therefore, $\mathbf{C} = -\mathsf{R}^\mathsf{T}\mathbf{t}$.

A projection and scaling, due to the transformation carried out by the optics and the sensor's geometry. The intrinsic parameters model it. A basic definition can have 6 degrees of freedom (note that this is not fixed, and many other degrees may be added, *e.g.*, lens distortion coefficients), defined by a 3×3 matrix K, where $\mathsf{K} = \mathsf{HP}$:

$$\mathsf{K} = \begin{bmatrix} f q_u & s & p_u \\ 0 & f q_v & p_v \\ 0 & 0 & 1 \end{bmatrix} \tag{2.3}$$

- f is the focal length, representing the distance from the optical center $\mathbf{C}$ to the image plane.
- s, q_u, and q_v are scale factors.
- $p : (p_u, p_v)$ is the principal point modeling the offset from the origin.

Camera Internal and External Orientation

The second step of the photogrammetric process (see Figure 2.10), elated to the computation of the orientation information of the camera(s), is composed of three primary estimations (Förstner and Wrobel 2016):

Internal Orientation: also known as **camera calibration**, estimates the intrinsic parameters (K) that characterize the optical, geometric, and digital properties.

Relative Orientation: describes the extrinsic parameters that model the camera's position **C** and orientation R in the 3D space using an arbitrary coordinate system.

Absolute Orientation: maps the previous computed relative orientations to a Coordinate Reference System (CRS), *e.g.*, WGS84. This conversion is identified as **georeferencing**.

The two first ones (**internal and relative orientation**) are frequently **processed simultaneously**. However, the algorithm employed will depend on the input data. Among the most known, we can mention **Direct Linear Transformation (DLT)**, which allows locating (in the world) and calibrating a camera from a **single image** (Hartley and Zisserman 2003). It requires as input at least six or more 3D points in the scene and their corresponding 2D image points. The algorithm defines an affine camera model, *i.e.*, a perfect lens with no distortion, representing only twelve degrees of freedom: six extrinsic (R and **t**) and six intrinsic parameters (f, s, k_u, k_v, c_u, c_v).

When **more input images** are available, other techniques may be used. One example is **Bundle Block Adjustment (BBA)**, which locates (exterior orientation) the camera(s) using image correspondences (Triggs et al. 2000). In a self-calibration configuration, the interior orientation of the camera(s) can be additionally estimated. **Scale Invariant Feature Transform (SIFT)** is a popular method for obtaining these correspondences by describing local areas through feature vectors (Lowe 1999, 2004). The algorithm reduces the entire image content to a set of key points later matched between photos (making associations), *e.g.*, using an Approximate Nearest Neighbor (ANN) technique.

Scene Reconstruction



Figure 2.12: Comparison between sparse and dense point clouds. Model represents the Tvrđava Minčeta tower (Dubrovnik).

A 3D reconstruction of the photographed scene is possible when multiple stereo images are the input of a photogrammetric analysis. The **Bundle Block Adjustment (BBA)** method estimates the 3D location of the feature points in the world, resulting as an (additional) output of the algorithm a sparse point cloud of the 3D scene. This result can be later used as an input for a dense world reconstruction based on these 3D tie points and the camera(s) parameters. Among any camera orientation selected (absolute or relative), depth maps can be computed and used to create the 3D scene model utilizing all the initial input images. A popular technique employed for this densification is **Semi Global Matching (SGM)**. It searches in different directions to estimate the best depth information to be used

in the reconstruction (Hirschmuller 2005). Figure 2.12 showcases the difference between a **sparse point cloud** (the first result obtained) and a **dense point cloud** (the final output). Furthermore, this 3D reconstruction pipeline can be adjusted to Cultural Heritage purposes by following customized methods and protocols (Martin-Beaumont et al. 2013; Pierrot-Deseilligny, De Luca, and Remondino 2011), *e.g.*, for digital replicas of monuments and buildings.

2.1.3 Geographic Imagery

Geographic imagery exhibits information related to objects at or near the Earth's surface and is used to **understand, analyze, and represent the geographic spaces and dynamics** (Kupfer and Emerson 2005). It can be stored as **black and white**, **color**, or **multispectral** image data in film or digital form (DeFries 2013). In black and white photographs, distinctions are expressed in gray levels. On the other hand, color films offer an increase in interpretation capacity through other features, like hue (dominant wavelength), chroma (color strength), and value (color intensity). For a broader overview, multispectral images capture characteristics outside of the visible spectrum, *e.g.*, infrared or ultraviolet rays.

Aerial Photography

Aerial photography (*i.e.*, bird-eye views of the world) refers to **all imagery captured from an airborne craft**, including helicopters, airplanes, or drones. Its origins date to the 1850s by the French balloonist Gaspard-Félix Tournachon (Doucet 2021). It is still a broadly used method in producing topographic maps and represents a cost-effective and accessible data source. Aerial images offer a spatial resolution, *i.e.*, Ground Sampling Distance (GSD), of up to $0.01 - 50$ m per pixel depending on the capturing height and the characteristics of the sensor employed, as represented in Figure 2.13.

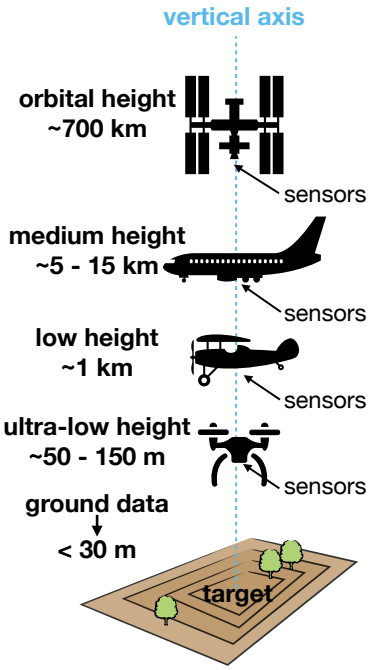


Figure 2.13: Aerial photography, diagram adapted from Aber, Marzolff, and Ries (2010a). Various instruments and platforms operated at multiple heights.

The practice of aerial photography as a geographic instrument was experimental primarily before World War II, but its use matured in the postwar (Aber, Marzolff, and Ries 2010a; Collier 2009; Wolf, Dewitt, and Wilkinson 2014a). Because of its early application dated back before satellite imagery, aerial images represent a valuable source of historical data. They provide a useful visual representation capturing fine detail, such as buildings and roads. In addition, the availability of large numbers of historical aerial photos

makes it possible to measure and analyze long-term geographical changes, *e.g.*, vegetation and landscape variations (DeFries 2013; Read, Chambers, and Torrado 2020).

As depicted by the diagram adapted from Paine and Kiser (2012) in Figure 2.14, aerial images can be classified into three different types according to the camera tilt angle (Paine and Kiser 2012; Wolf, Dewitt, and Wilkinson 2014a): *(i)* **vertical**; *(ii)* **low oblique**; *(iii)* **high oblique**. We define the camera's tilt as an angle derived by the camera's movement that rotates it along the vertical axis while fixed in a location.

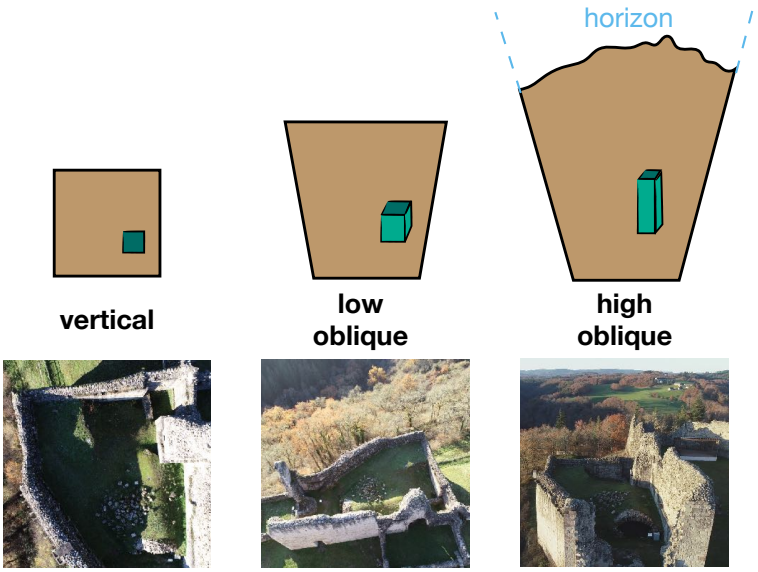


Figure 2.14: Captured ground area of aerial imagery, remodeled from Paine and Kiser (2012). **Top**: images show the relative size and shape of three different camera tilt angles. Consider the represented object as a building. **Bottom**: real-world samples taken at Château de Ventadour (Tournon-sur-Rhône).

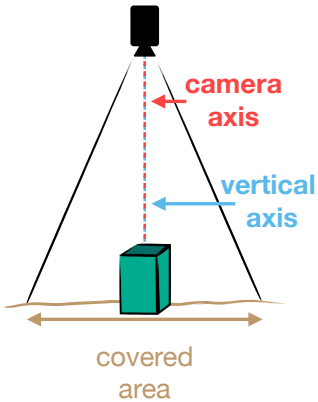


Figure 2.15: Camera orientation for vertical aerial photographs. The camera axis and vertical axis "coincide" with each other.

In **vertical imagery**, the camera axis is pointed toward the ground as vertically as possible (see Figure 2.15). There is no tilt (*i.e.*, true vertical) or a residual, which generally depends on the system's accuracy to stabilize the camera (gimbal). The captured space may be small, but the pictures can overlap to produce more extensive areas and give additional scene information. These attributes support primary applications in cartography and photogrammetry (Paine and Kiser 2012; Ray 1960), *e.g.*, in photogrammetric acquisitions, these (vertical) images are mainly used to generate orthoimages, a scaled and even georeferenced photograph that allows more accurate and manageable measurements of objects and distances.

The camera axis is purposely tilted from the vertical in **oblique imagery**. The profile view provides the observer with more natural and comprehensive objects, *i.e.*, similar to a hill or high building view. The spatial resolution continually changes from the foreground to the background, *e.g.*, features in the back may be too

small or hidden. As illustrated in Figure 2.16, oblique photographs are known as **low oblique** when the horizon is not visible. On the contrary, if the horizon is visible, then the picture is **high oblique**. These photographs enable the recording of objects to a vertical extent. Both low and high oblique imagery are helpful for geological or archaeological investigations since it reveals topographic details of the land. These photographs can be combined with vertical views to produce 3D photorealistic models (Höhle 2013; Paine and Kiser 2012; Remondino et al. 2016), where the final output should be as close as possible to the photo and so the reality.

Low Oblique: only Earth's surface is photographed in low oblique aerial pictures since there is a slight shift in the camera. An intentional tilt is fixed of ±15° to ±30° in the camera axis from the vertical axis.

High Oblique: the degree of tilt in the camera axis for high oblique images is higher than for low oblique. This angle is around ±60° from the vertical axis. Both horizon and a large area of land are noticeable; therefore, the footprint of the photograph (extent covered) can be considered infinite compared to vertical and low oblique photos.

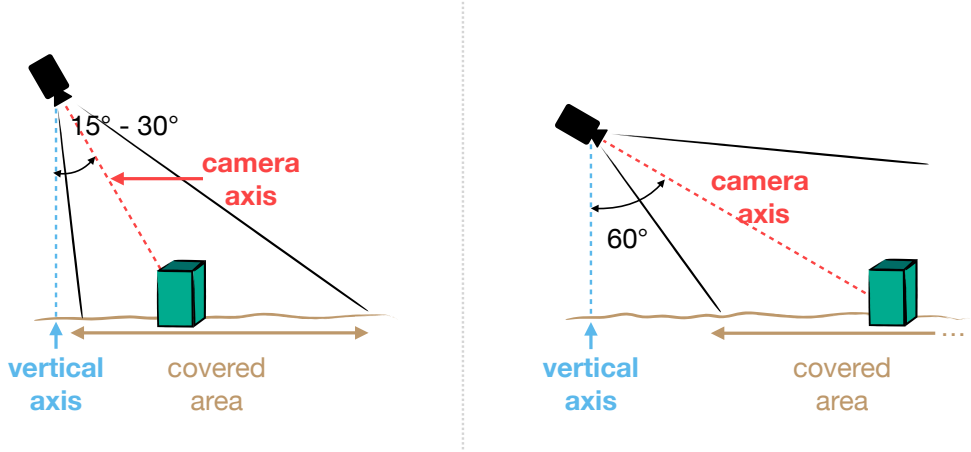


Figure 2.16: Camera orientation for oblique aerial photographs. **Left**: with a low oblique, the camera axis has a deviation of ~ 15° to ~ 30° from the vertical axis. **Right**: with a high oblique, the camera axis has a deviation of ~ 60° from the vertical axis.

Terrestrial Photography

Terrestrial imagery (*i.e.*, close-range views of the world) relates to all **photos taken from Earth's surface**. Aimé Laussedat proposed terrestrial photographs for surveying and map-making in the 1840s, but the photographic technology was inadequate. Its practical use on geographic applications started in the 1880s (Collier 2009; Polidori 2020; Wolf, Dewitt, and Wilkinson 2014a). Nowadays, terrestrial photography has become a useful tool in many areas of scientific and engineering research, *e.g.*, metrology, architecture, and archeology. One of the main reasons is the possibility of analysis of objects unavailable for direct analysis (also known as contactless measurements).

Terrestrial photographs are usually obliques or horizontals where there is a tilt of about 90° on the axis of the camera, from the vertical (see Figure 2.17). Compared with vertical imagery, a higher level of detail can be captured from objects using terrestrial imagery. However, it does not fully

cover the photographed area, *e.g.*, roof information lacks, and some objects may be occluded (Wolf, Dewitt, and Wilkinson 2014b; Wu et al. 2018).

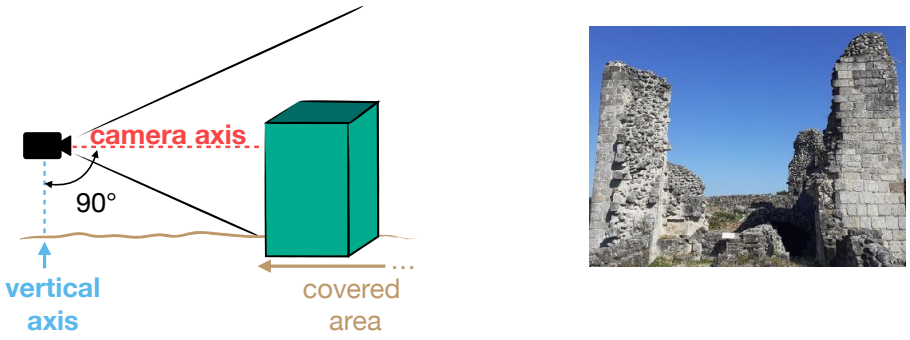


Figure 2.17: Camera orientation for terrestrial photographs. **Left**: a visual depiction showcasing the camera axis with a deviation of ~ 90° from the vertical axis. **Right**: a real-world sample taken at Château de Ventadour (Tournon-sur-Rhône).

2.2 Image Exploration for Extensive Collections

Photographs, explicitly geographic imagery, can showcase the relationships between spaces and their changes. Because images are snapshots of specific moments in time, their examination should adapt to this characteristic. From this idea, our next question goes to their discovery and analysis: **how do we browse through historical photographs representing geographical spaces?**

2.2.1 Metadata Browsing

The transition of iconographic materials from traditional to digital formats and the large and continuously expanding availability of digital content pose new challenges. A notable difficulty users experience when **browsing inside digital photographic compilations** is finding the desired content. It presupposes a comprehensive knowledge of the structure and semantics of the respective repositories and their content, *i.e.*, which is generally only grasped by experienced users.

Metadata offers **annotations**, **descriptions**, and **classifications** on an object in a collection to reduce the searching difficulties. When a concrete target criterion is formulated, users can distinguish items of interest much quicker. The most widespread and known approach connected to images is a *photo library*. As Figure 2.18 showcases, it lists and displays the photographs in a 2D grid filtered by different metadata, *e.g.*, keywords, categories, and dates. A notable example is *Google Images*[2]. In

[2] https://images.google.com

addition, the *Online Library of National Archives*[3], *Open Musée Niépce*[4], and *Médiathèque Terra*[5]are historically related samples.



Figure 2.18: Image gallery. The result is obtained when browsing the keyword "Paris". The images are presented in a 2D matrix-like form. They can be additionally filtered by theme and color. Credits to *Open Musée Niépce*.

2.2.2 Spatial and Temporal Exploration

Cartography has been the process typically employed to visualize geographic spatial (geospatial) information. Although it remains an essential communication medium for map makers, with the advent of the digital era, it has evolved and transitioned to a new approach that emerged in the 1980s and 1990s (MacEachren 1994), known as **geographic visualization** (often shortened as geovisualization or geovis). Its emphasis is on the exploration and interaction of geospatial data more profoundly. To understand it, we define these two central terms as (Çöltekin, Janetzko, and Fabrikant 2018):

Exploration systematically examine a dataset by summarizing, plotting, or conducting statical analysis to arrive at insights and hypotheses.

Interaction graphically represent the data in a form where it is improved the way the information is associated.

Therefore, the concept of geovisualization can be understood as the use of visual geospatial displays for interactively explore geographic information to generate hypotheses, develop problems, solutions and construct knowledge (Kraak 2003). As depicted by the diagram in Figure 2.19, the core framework of geovis can be seen as the one proposed by MacEachren, Gahegan, et al. (2004) based

[3] https://www.siv.archives-nationales.culture.gouv.fr
[4] http://www.open-museeniepce.com
[5] https://terra.developpement-durable.gouv.fr

on three factors: *(i)* **user**; *(ii)* **task**; *(iii)* **interaction**. Essentially, through the use of a geovisualization environment, the public (*e.g.* non-experts users) or specialists (*e.g.*, researchers) can:

1. **Explore** and discover patterns to form and model informed questions.
2. **Analyze** to confirm or reject particular hypotheses.
3. **Synthesize** to generalize the findings.
4. **Present** and communicate to the public (or to themselves) the conclusions.

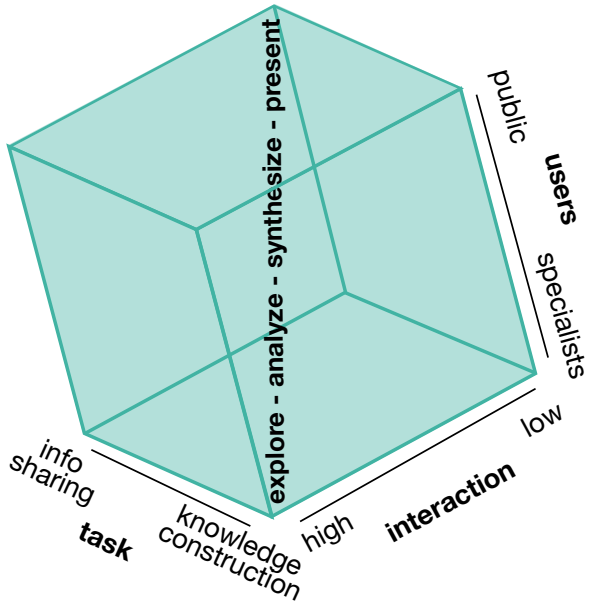


Figure 2.19: Functions for geovisualization, defined by the user, task, and level of interaction enabled in the interface. The diagram has been adapted from MacEachren, Gahegan, et al. (2004).

For instance, the framework implies that users, like specialists, use (more) interactive geovisualization in the exploratory process for knowledge construction. In contrast, these interaction requirements become lower as the process moves towards communicating the discovered information with (or by) the public (Çöltekin, Janetzko, and Fabrikant 2018). It is a distinct conception of how geographic visualization evolved from cartography.

Consequently, we may say that the **geovisualization process allows discovering any geographic information and phenomenon** through different approaches (*e.g.*, 2D maps or 3D immersive urban models) following the aforementioned framework. Overall, it aims at providing synchronized interfaces displaying temporal, spatial, or thematical points of view on the data based on specific Human-Computer Interaction (HCI) aspects. Among the most common approaches, we can mention methods employing immersive and augmented environments, aiming towards a **"Digital Earth"**, which is a digital replica of the world captured in all intricate details (Çöltekin, Griffin, et al. 2020; Gore 1998).

Two-Dimensional Spatial Exploration

An extension of simple metadata browsing is a spatial exploration strategy. Its basis falls in the **location** of the **photos** within a **geographic model**, *e.g.*, a map, through a **co-visualization** of the

data. Methods like *Pixplot*[6] and *ClustTour* (Papadopoulos et al. 2010) facilitate examining several spatially oriented images in a 2D map by clustering the pictures into small groups inside the geographic environment. *Navigae*[7], *Navilium*[8], *PhillyHistory*[9], and *Estonian Photographic Heritage Society*[10] are particular cases further adapted to a Cultural Heritage (CH) approach (see the example of *Navigae* and *Navilium* in Figure 2.20). Commonly, these techniques restrict the view to an orthogonal top view of the map, *i.e.*, aerial perspective, with a limited number of interactions. For example, they allow zooming in/out to reveal the elements contained in the clusters and display a specific image's information.

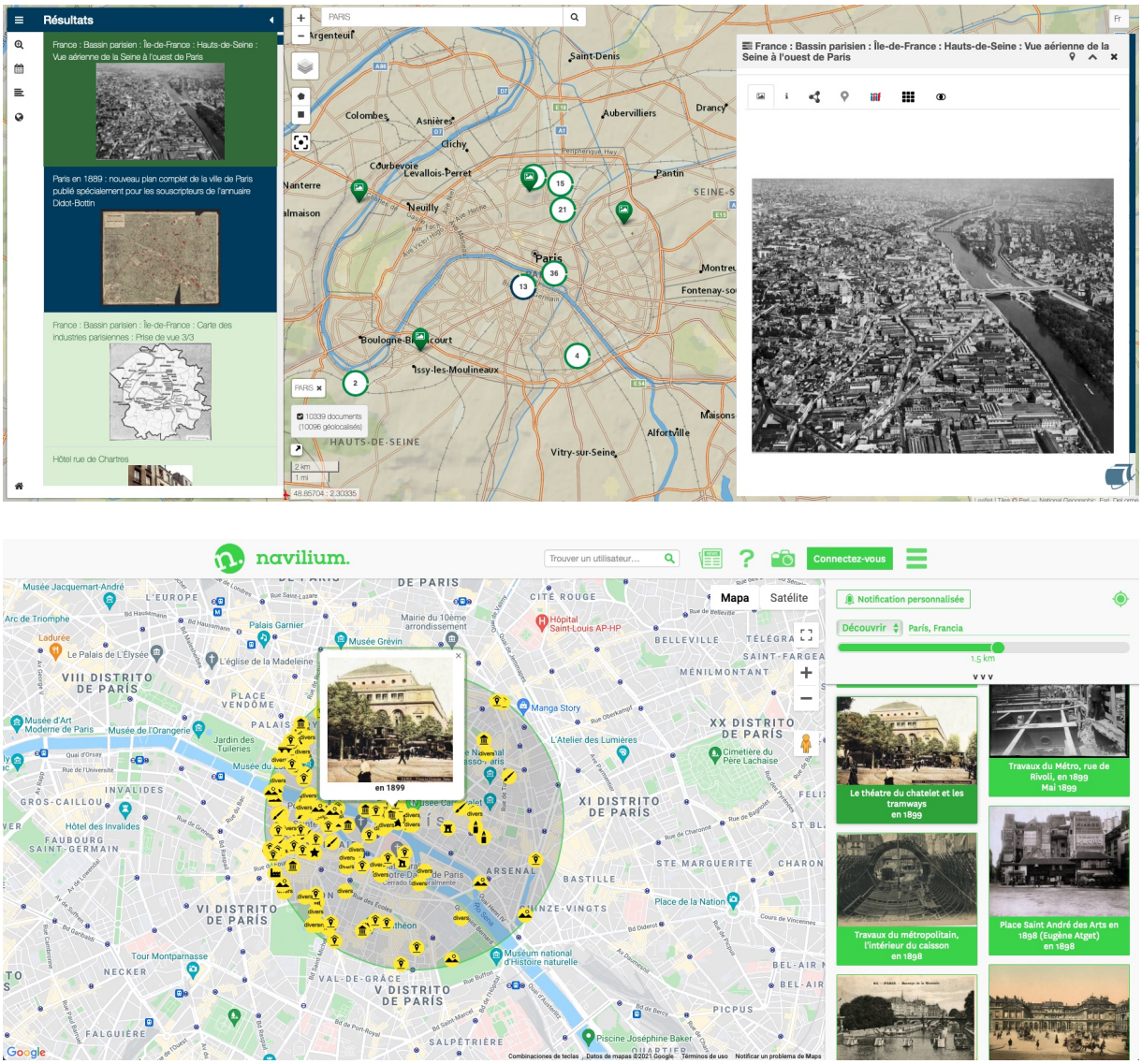


Figure 2.20: Image gallery along with a 2D map model. The keyword "Paris" is used for the search. The user can locate the capturing location of the photographs geographically through a 2D map of the region. More complex interactions are not allowed. Credits to *Navigae* (top) and *Navilium* (bottom).

[6] https://dhlab.yale.edu/projects/pixplot
[7] https://www.navigae.fr
[8] https://www.navilium.com
[9] https://www.phillyhistory.org
[10] https://ajapaik.ee

An alternative approach is taking the photographer's perspective inside the spatial environment. This viewpoint may enable users to learn about specific areas in the scene. Typical online examples are *WhatWasThere*[11] and *Historypin*[12]. Both sites are based on connecting historical photographs of landscapes, landmarks, and buildings. The applications enable users to upload and integrate their old photos on the map by locating them manually. Unfortunately, although the methods' focus is history, they use map providers containing only current material (Armstrong 2012; Bolter, Engberg, and MacIntyre 2013). Their resulting co-visualizations have considerable misalignments between images and context, as shown in Figure 2.21, with no reference to the scene's past state.



Figure 2.21: Co-visualization of a terrestrial historical image and *Google Street View*. Misalignments between scene and photo are visible. Credits to *Google Street View*; the town hall of Ivry Sur Seine and *Historypin* (top); Phil Blanco´s Conoco Station and *WhatWasThere* (bottom).

[11] http://www.whatwasthere.com
[12] https://www.historypin.org

Three-Dimensional Spatial Exploration

As explained in Section 2.1, specific properties of photographs can be automatically estimated, such as the position and orientation of the device used to acquire the photo. This data can be used later to **spatialize/georeference** the **pictures in a 3D digital environment**. To represent a geographic space digitally, 3D city models can be used. These are a collection of different 3D objects depicting elements in the real world, like the ground surface, buildings, trees, and rods. Depending on its application, their Level of Detail (LOD) can vary. For instance, they may accurately represent an object (LOD 3) or give an abstract idea with simple shapes (LOD 1). Among the most common types we can mention (Krig 2016):

- **Point Cloud:** collection of data points that depicts an object when visualized together.
- **Mesh:** set of points, edges, and faces that defines an object's shape.

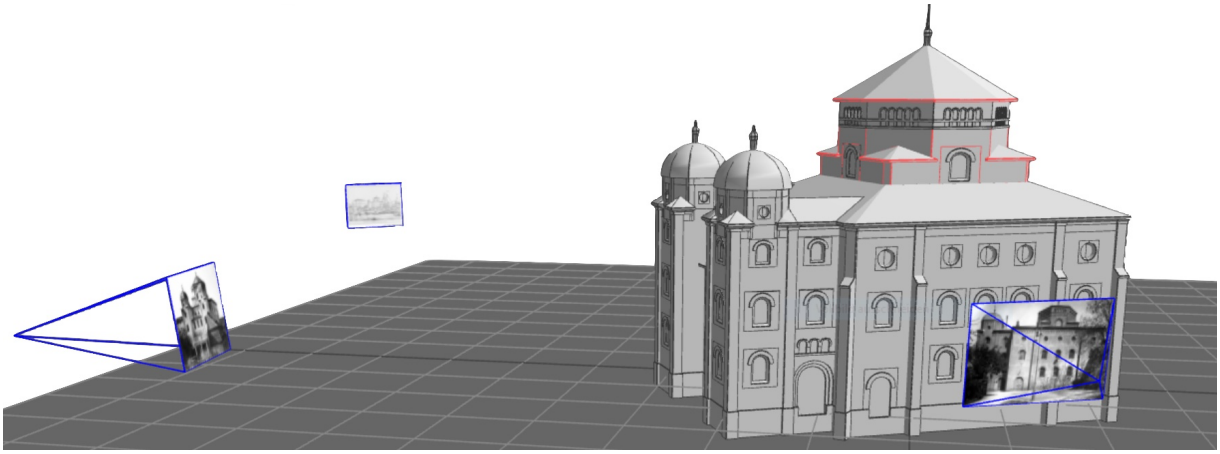


Figure 2.22: Example of a 3D model and a spatially placed image. This scene is representing Semper Synagogue in Dresden (Bruschke, Niebling, Maiwald, et al. 2017).

As presented by Figure 2.22, using these 3D models and **including the third dimension enables users to magnify the visualization of the photographs and their described objects**. The viewers are no longer limited to a specific 2D view, but navigation inside the scene is possible by taking either the position and orientation of the photographer or custom angles and visualizations (Bruschke, Niebling, Maiwald, et al. 2017).

Photo Tourism[13] is one of the first strategies to introduce the exploration of extensive unstructured photo collections in 3D (Snavely, S. Seitz, and Szeliski 2006). The method detects feature points on images and matches them with features in others (photogrammetric analysis). All photographs are visualized from the reconstructed sparse 3D point cloud and camera poses, as depicted in Figure 2.23. The user can navigate through the scene or jump from one image to another. This approach has been extended through the years by proposing optimal paths between images (Snavely, Garg, et al. 2008), artifact and occluder removals (Garg et al. 2009), and illumination correction variation (Wehrwein, Bala, and Snavely 2015). However, the success of these strategies relies on the use of a large number of photos (generally obtained online), which frequently is not possible in a historical context due to the limited amount of samples.

[13] http://phototour.cs.washington.edu/applet/index.html

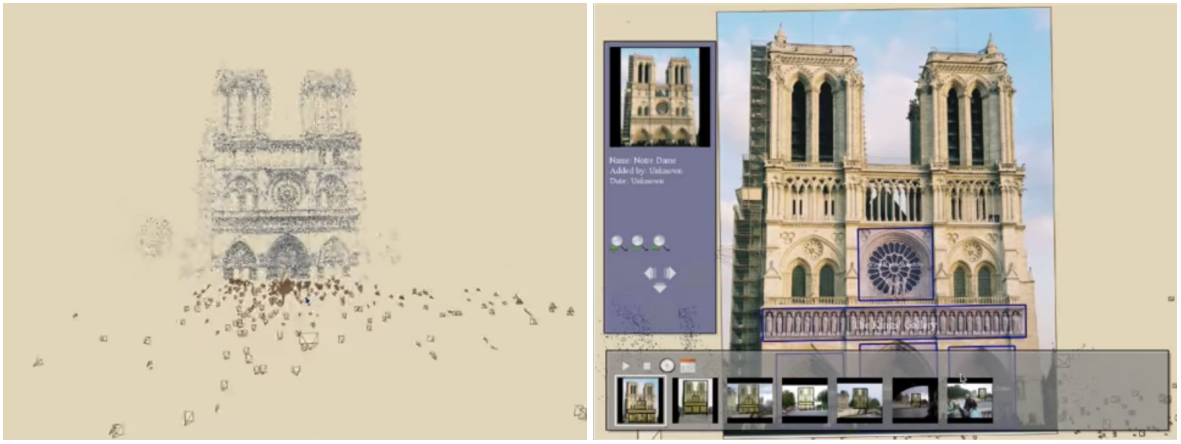


Figure 2.23: Co-visualization of a terrestrial image from Notre-Dame (Paris) and the reconstructed 3D scene. **Left**: the estimated camera locations along with the sparse point cloud. **Right**: the interface is showing the visualization of the street-view image along with all possible user interactions. Credits to *Photo Tourism* (Snavely, S. Seitz, and Szeliski 2006).

An additional example where the reconstruction of the scene depends on the number of images used is *PhotoCloud*[14]. The desktop application enables a co-visualization of 2D images and a 3D model as seen in Figure 2.24 (Brivio et al. 2013). In a pre-processing step, a 3D mesh and the camera orientations are computed from the input photographs. In the resulting visualization, images are projected in the 3D environment, providing an interactive real-time exploration of the images. As an optimization, the method computes semantic distances to cluster the images, splits the original 3D model into blocks at different resolutions, and avoids mixing images, *i.e.*, does not project more than one image at a time. Nonetheless, the final 3D environment may present holes where the reconstruction process was unsuccessful.

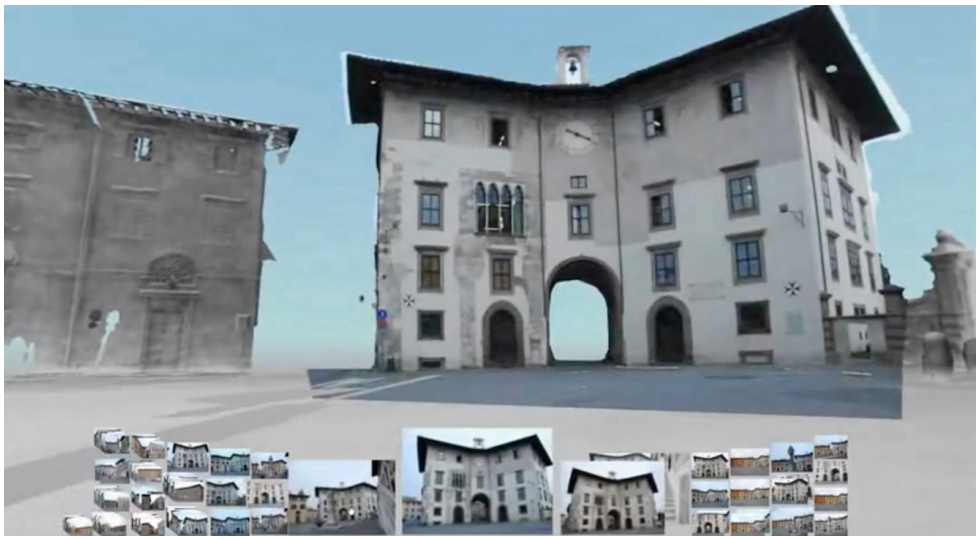


Figure 2.24: Co-visualization of a street-view photo from the Cavalieri Square. As the user navigates near the viewpoint of an image, the projected image in the view is updated. Credits to *PhotoCloud* (Brivio et al. 2013).

[14] http://vcg.isti.cnr.it/photocloud

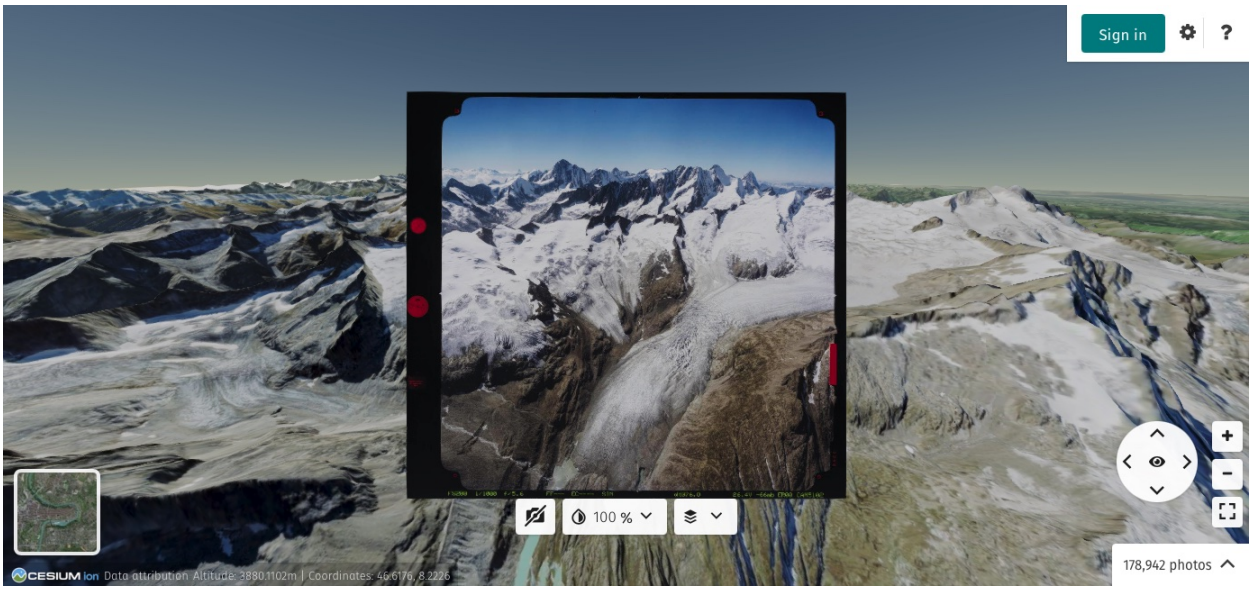


Figure 2.25: Co-visualization of an aerial picture from Swiss Air and a 3D environment. Exploration in the scene is limited to the position of the photographer's perspective. Credits to the ETH Library and *Smapshot*.

Differently, *Smapshot*[15] is a web application that uses the latest geographic data to assemble a 3D environment, as portrayed by Figure 2.25 (Blanc, Produit, and Ingensand 2018; Produit et al. 2018). It follows the same participatory method as *Historypin* for geolocating the images, *i.e.*, it allows volunteers to align the images within a 3D virtual globe. The photographs are linked using a SIFT algorithm before any user input is required for time efficiency. Instead of projecting the photos on a 3D city model during the visualization process, the online platform employs a floating quad[16] on the scene textured by the historical images. The browsing can be refined using metadata, *e.g.*, filtering by date through a time range specification. Notwithstanding, the navigation inside the 3D environment is restricted since the view can only be positioned from the photographer's perspective; therefore, the users cannot go around or behind the selected photo to explore the scene.

Spatio-Temporal Exploration

A geographic phenomenon can be considered a set of facts and characteristics observed and located in space. The term dynamic refers to the evolution of these features over time (Gautier, Davoine, and Cunty 2018). As proposed by Mennis, Peuquet, and Qian (2000)

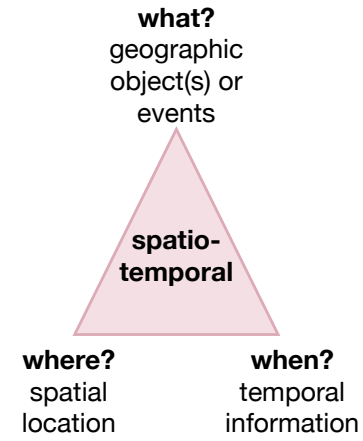


Figure 2.26: Spatio-temporal triangle adapted from Mennis, Peuquet, and Qian (2000): object(s)/event(s), location, and time.

[15] https://smapshot.heig-vd.ch

[16] A **quad** (quadrilateral) is a polygon with four edges (sides) and four vertices (corners).

with the triad framework (Figure 2.26), we can describe the concept of spatio-temporal (*i.e.*, a dynamic geographic phenomenon) through three dimensions: spatial (where?), temporal (when?), and thematic (what?). On this note, a spatial exploration of photo collections can be furthermore extended to spatio-temporal. The addition of the time variable can allow users to **visualize the evolution of the photographed area and gain a comprehensive understanding of the photograph's temporal structure**. As described by Davoine et al. (2015), several techniques have been developed to integrate the temporal dimension into geographic visualization systems, *e.g.*, maps representing time through animations or multi-view interfaces splitting the spatial and temporal content.

Photogrammar[17] is an example that illustrates the progression of a geovis application towards a spatio-temporal method (Arnold et al. 2020). Figure 2.27 shows its latest version presenting a more integrated user experience. Aside from the typical metadata search, it proposes HCI strategies like *(i)* an interactive 2D map geographically showing the spatial extent of the photos; *(ii)* a timeline capturing the temporal aspects of the collections. The method connects all the different interactive visualizations into a synchronized visualization. However, a significant limitation is retaining an acceptable visible result on all tasks, considering the restricted available screen space.

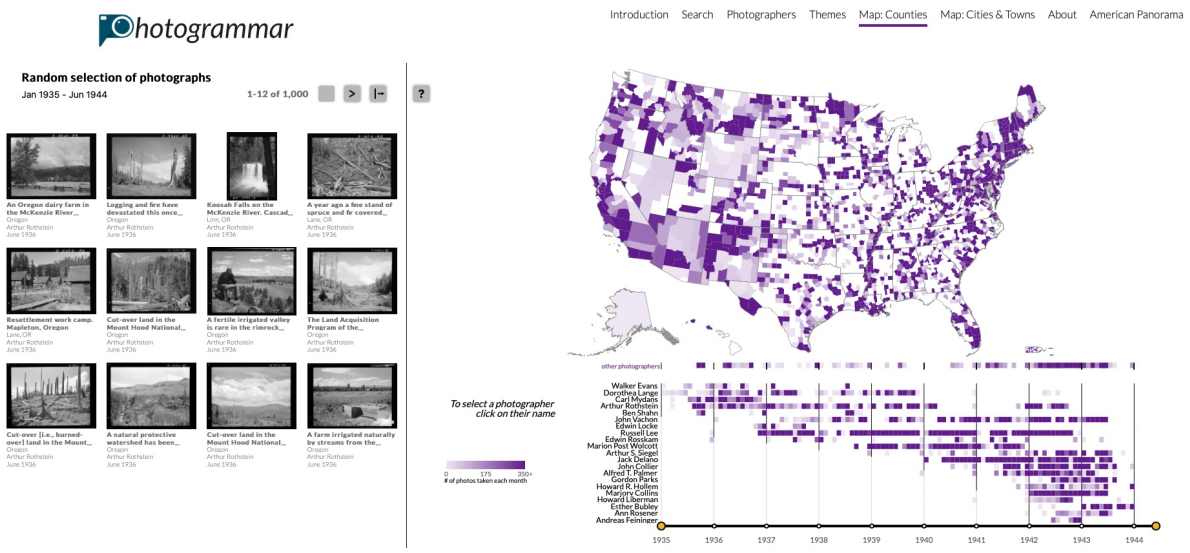


Figure 2.27: Image gallery along with a synchronized spatio-temporal view. The elements are linked together to allow the user to explore different dimensions of the collections simultaneously. Credits to *Photogrammar*.

In the three-dimensional context, *UrbanHistory4D*[18] is a web-based browser that includes temporal information in a Cultural Heritage environment (Dewitz et al. 2019; Maiwald, Bruschke, et al. 2019; Maiwald, Henze, et al. 2019). The application allows users to search within large repositories of historical images through a metadata search, along with a spatial 3D city model and a responsive time slider, as seen in Figure 2.28. It follows the same approach as *Smapshot*, where images are displayed in floating geometries textured by historical photos. In order to keep the 3D view clear, when it is possible, photographs are grouped. The timeline enables the viewer to visualize how the architecture, cityscape, and acquisition behavior have evolved through the years.

[17] https://photogrammar.org
[18] https://4dbrowser.urbanhistory4d.org

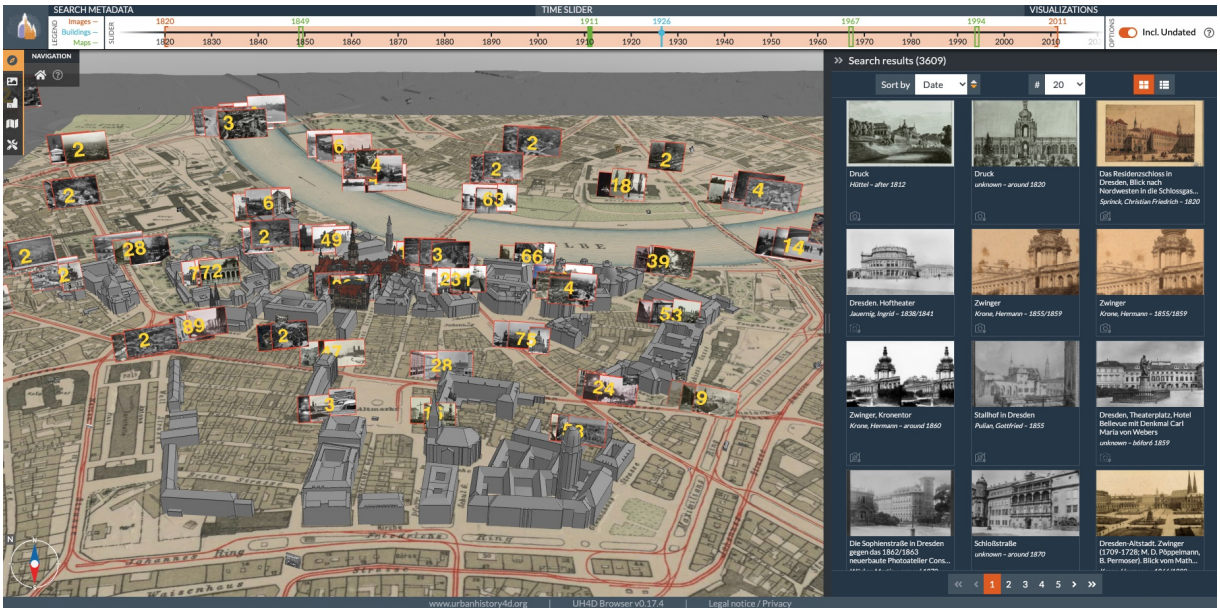


Figure 2.28: A spatio-temporal interface. A 4D browser is presenting historical pictures and 3D models of Dresden. The timeline allows the user to filter images, maps, and buildings. Credits to *UrbanHistory4D*.

Cities are lasting and thus characterized by the inertia of social structures. As Jaillot, Servigne, and Gesquière (2020) points out, 4D time-varying environments can enhance a historical and chronological evaluation of an urban exhibition. Its basis lies in the segmentation of the 3D city model through spatial, temporal, and co-visibility cues. Instead of an inactive environment, the result is a 4D spatio-temporal object, *i.e.*, a 3D model with time as an additional dimension, that changes appearance and creates a chronicle of the scene over time. For instance, as Figure 2.29 depicts, this urban model can be used inside an immersive photo explorer system to visualize historical pictures through time (Jaillot, Rigolle, et al. 2021).

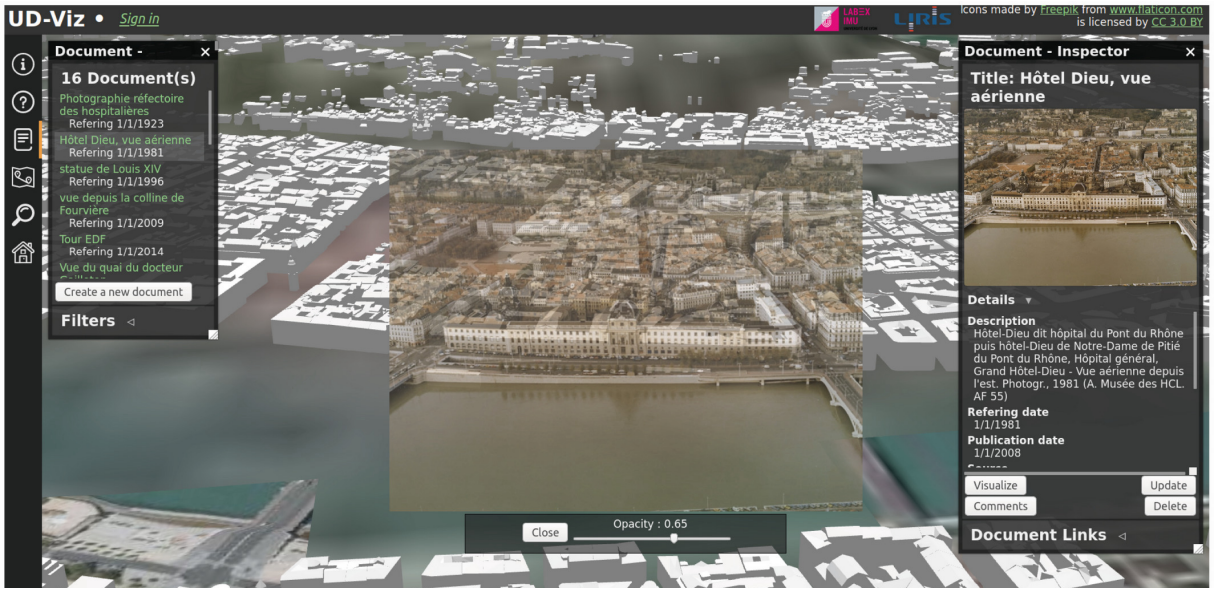


Figure 2.29: Co-visualization of a historical photograph from Hôtel-Dieu (Lyon) and a time-evolving 3D city model (Jaillot, Rigolle, et al. 2021).

4D Cities[19] is another example of image exploration inside time-varying 4D models (Schindler and Dellaert 2012). The system is built-in earlier work that automatically infers the temporal order of images and reconstructs the city model (Schindler and Dellaert 2010; Schindler, Dellaert, and Kang 2007). The resulting visualization reveals information about locations, dates, and contents of historical images. Buildings rise/fall, and images flicker in and out to reflect the changes over time, as displayed in Figure 2.30.

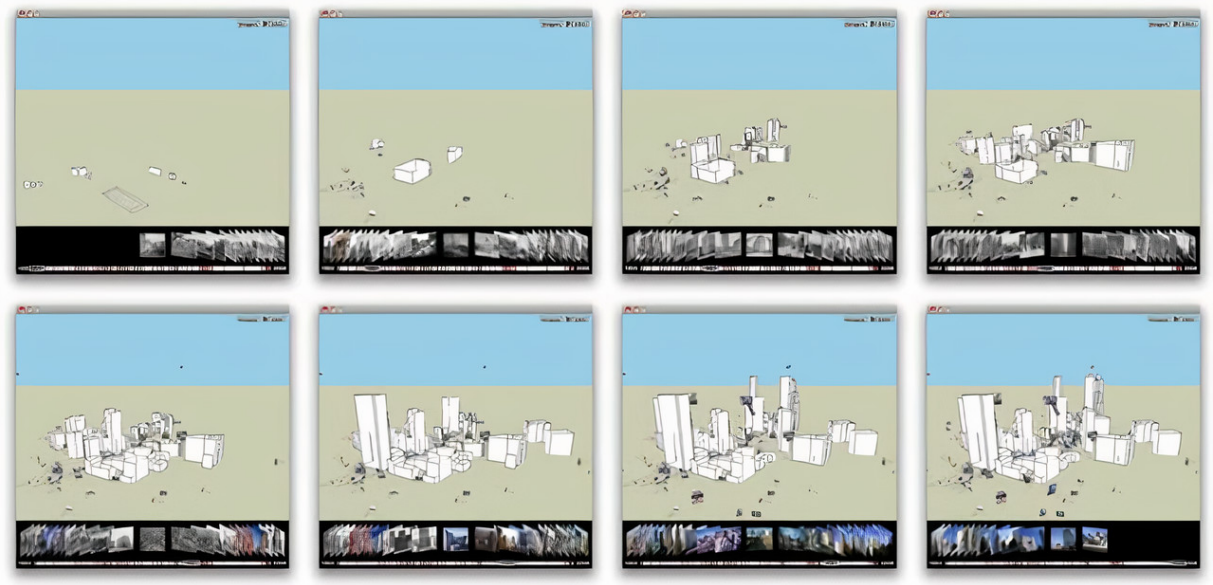


Figure 2.30: Time-varying 3D model of the city of Atlanta depicting 1864 to 2008. Credits to *4D Cities* (Schindler and Dellaert 2012).

Overall, 4D models can be used to enhance photographs furthermore. *Locus Imaginis* follows this idea by allowing visitors of cultural monuments to manually align their pictures to a digital model. The application then enriches these pictures by projecting them to the 3D model and automatically adding the corresponding 2D semantic annotation. Other visitors can later query the semantically-enriched data and visualize the result inside a 4D environment (Stefani et al. 2013).

2.2.3 Use of Extended Reality

Nowadays, the versatile evolution of smartphones, tablets, and Head Mounted Devices (HMD) has increased **Extended Reality (XR)** usage in geovisualization. It covers all the various **technologies capable of enhancing the user's senses** (Çöltekin, Griffin, et al. 2020). In the case of Virtual Reality (VR), it implies a complete immersion experience that shuts down the physical world. In contrast to traditional 3D visualization, immersive VR environments produce novel forms of spatial interaction with the presented data. For instance, a web application like *UrbanHistory4D* can be extended to XR, *e.g.*, a location-dependent historical visualization may project the photographs into a printed 3D model using a mobile device, or a fully immersive environment can challenge users to find multiple photos' specific locations and orientations as seen in Figure 2.31 (Dewitz et al. 2019; Maiwald, Bruschke, et al. 2019; Maiwald, Henze, et al. 2019).

[19] http://4d-cities.cc.gatech.edu/atlanta/



Figure 2.31: Virtual Reality environment for the exploration of photo collections. Users are challenged to discover the location and orientation of historical images (Maiwald, Henze, et al. 2019).

A technique such as *VaRt-DataExplorer* allows meaningful exploration of historical objects found typically in museums (Koebel, Agotai, and Çöltekin 2020), as presented in Figure 2.32. It combines two layers of visualization *(i)* a bird's eye view, which presents a global data visualization; *(ii)* a ground-level view, which allows discovery of items of interest. *Virtual Journey*[20] is another example that promotes Cultural Heritage. It allows viewers to visit the Strasbourg Cathedral in France virtually. Still, a particular challenge with historical image collections is the information's incompleteness, imprecision, and heterogeneity. It means that the data is often in a semi-structured format, while many exploration techniques (including VR) require harmonized datasets.



Figure 2.32: Virtual Reality system displaying to the user an overview of an entire museum collection through an aerial view. Credits to *VaRt-DataExplorer* (Koebel, Agotai, and Çöltekin 2020).

[20] http://www.voyageenimmersion.com

Photogrammetric techniques and GPS positioning facilitate **locating users** in a **specific point of view** to **align actual content** with a **virtual** one. An XR case is Augmented Reality (AR) which provides additional information about the real world by creating unreal or simulated objects inside a live view. Mixed Reality (MR) magnifies this by allowing interaction with these elements. Combined with a situated visualization, *i.e.*, a view established in a meaningful place, these techniques can improve the viewers' understanding of historical image data (Lobo and Christophe 2020; White and Feiner 2009). Users are capable of experiencing stories, solving tasks, and traveling to unexpected places and times. In addition, it provides an increasing spatial and temporal awareness of the user's surroundings.



Figure 2.33: Mixed Reality application presenting historical portals through the viewer mobile. **Left:** portal at a distance. **Right:** near the portal. Credits to *SBB Stories* (Pulver et al. 2020).

CitiViewAR (Lee et al. 2012) and *SBB Stories* (Pulver et al. 2020) are two examples of mobile AR/MR following a situated visualization approach. As shown in Figure 2.33, their immersive location-based storytelling empowers users to visualize various narratives in the past based on different 3D elements and historical photos in the real world. However, for these types of interactions, when combining historical content with the present, it is imperative to identify appropriate transition cues between the two aspects and consider this in proper design solutions.

Another example is *Aïoli*[21], which showcases the evolution of Cultural Heritage (CH) analysis by combining *in situ* (on-site, in front of the object) and *ex situ* (off-site, with a digital copy) examinations. It proposes integrating geometric, visual, and semantic features in a single platform for CH surveying. Specifically, AR is used for on-site semantic annotations. Users can upload (currently) captured images, which are stored and processed to be spatialized using a photogrammetric analysis. The AR environment displays the result containing the reconstructed point cloud and the oriented photographs (as illustrated in Figure 2.34). Users can add to this digital copy annotations, which can propagate to all the digital resources in *Aïoli* (Abergel et al. 2019, 2021).



Figure 2.34: Augmented Reality platform allowing users to add annotations in the currently viewed objects. Credits to the *Aïoli* project.

2.3 Image-Based Rendering

Rendering can be seen as the **process** of **representing** the **observed data** contained in a **scene** from a particular point of view through the **simulation** of the **way light travels** in **space**. A traditional rendering process like Physically-Based Rendering (PBR) expresses this information through the light transport equation formalized by Immel, Cohen, and Greenberg (1986) and Kajiya (1986), simultaneously. As depicted by Figure 2.35 (right), the latter describes how much light L_o is sent from a single point P on a surface into a given direction $\vec{\omega}_o$. It models how light bounces and interacts with matter to reach the eye finally.

On the contrary, as displayed in Figure 2.35 (left), **Image-Based Rendering (IBR)** focuses mainly on the light modeling rather than the scene's physical modeling. IBR intends to **simulate** the **receiving end** directly through a mathematical expression known as the **plenoptic function**. This function describes the intensity and flow of light passing through all positions and all directions in space. The complete function is 7D, defined by the observing location (V_x, V_y, V_z), the angle

[21] http://www.aioli.cloud

of incidence (ϑ, φ), the wavelength λ, and the time t (Adelson and Bergen 1991). Over the years, different simplifications have been done; the more constraints are added, the simpler the plenoptic function becomes (Chen 1995; Gortler et al. 1996; Levoy and Hanrahan 1996; McMillan and Bishop 1995; Shum and He 1999; Sloan, Cohen, and Gortler 1997; Szeliski and Shum 1997), *e.g.*, when the environment remains static, and the light condition is fixed, t and λ can be dropped, resulting in a 5D function.

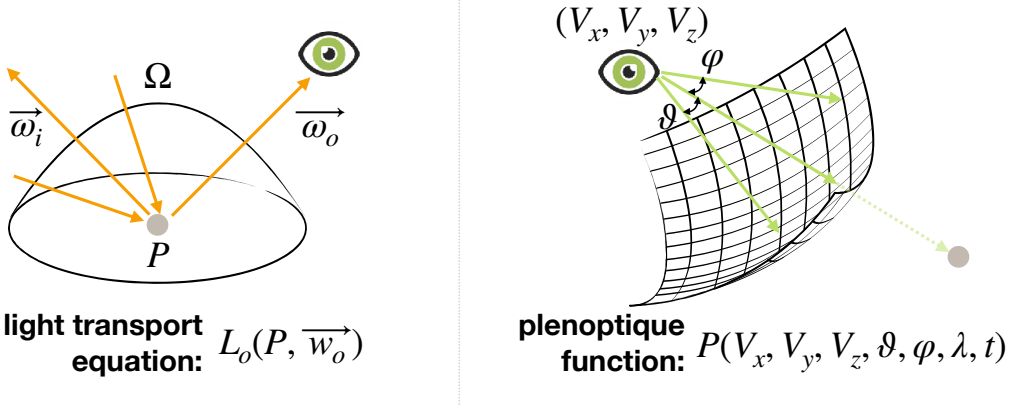


Figure 2.35: Different rendering equations. **Left:** Light transport equation. The radiance at a specific point P from a view direction w_o is defined by the outgoing light L_o (Immel, Cohen, and Greenberg 1986). **Right:** Plenoptic function. It describes all of the image information visible from a particular viewing position (McMillan and Bishop 1995).

The essence of Image-Based Rendering is **obtaining** all the **visual information** of the scene **directly through images**. It creates novel views of the environment by resampling the input images representing a discrete sample of the plenoptic function. Overall, as presented in Figure 2.36, the existing methods can be divided into three categories related to the geometric information known (from the scene): *(i)* no geometry; *(ii)* implicit geometry; *(iii)* explicit geometry (Oliveira 2002; Shum and Kang 2000). A more simplistic classification can divide all methods into two primary groups: *(i)* no geometry; *(ii)* geometry (Chang and Wang 2019). This section will focus on the latter, where at least an approximation of the geometry from the scene is known.

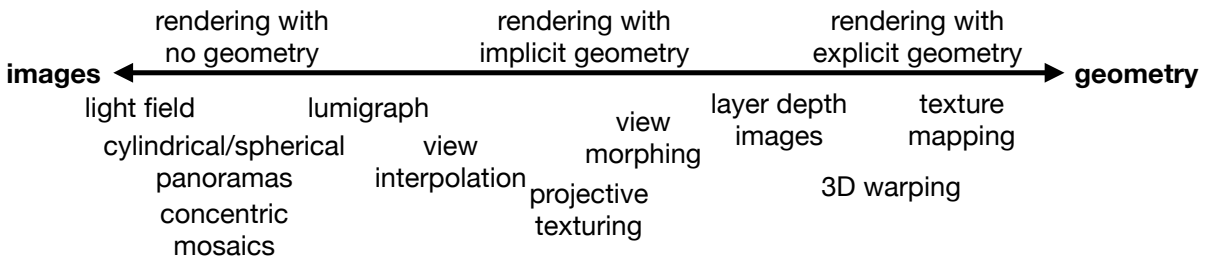


Figure 2.36: Image-Based Rendering techniques classified in the image-geometry spectrum. Diagram is adapted from Oliveira (2002) and Shum and Kang (2000).

Observation

Image-Based Rendering (IBR) is a broad subject that can be tackled from different perspectives. It does not represent a specific rendering technique but a set of methods, developed through the years, trying to **cope** with **different varieties** of **scene geometries** related to the absence, incompleteness, and inaccuracies. This chapter highlights only some specific areas of IBR, which are relevant to this research context, *i.e.*, visualizing historical photos. For a more extensive explanation, please refer to the proper reviews on the topic from Chang and Wang (2019), Oliveira (2002), and Shum and Kang (2000).

2.3.1 Projective Texturing

The concept of **projective texturing** was introduced by Segal et al. (1992), where it was used for shadows and lighting effects. However, the method can be directly applicable to IBR because it can simulate the inverse projection obtained when capturing a photograph. As shown in Figure 2.37, the approach **projects an image onto the scene from an arbitrary position in space** (Everitt 2001).

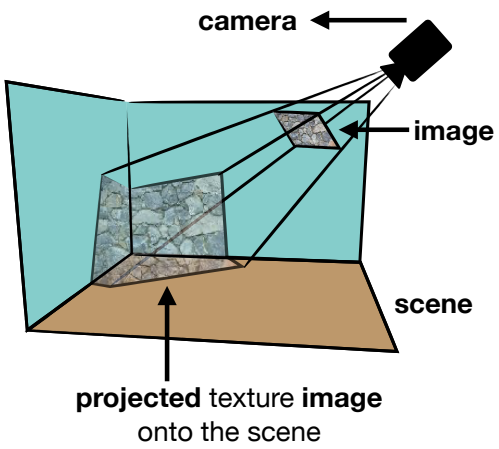


Figure 2.37: Projective texturing over a 3D scene, adapted from Segal et al. (1992).

More specifically, when the precise 3D information of a scene is known, the color of each pixel from the input photo can be projected to a specific fixed position in the 3D space. Then, according to the camera parameters of the novel point of view, the color of this space point can be projected onto the image plane of the novel rendered view. Eventually, when the view camera matches the exact parameters of the camera used to take the picture, a pixel-accurate reprojection can be visualized.

In practice, however, a projective texturing approach may fail since *(i)* not all pixels in the input picture may find the matching spatial positions; *(ii)* the spatial locations corresponding to the different pixels representing the same object may be inconsistent, *e.g.*, changes in the scene over time. Many IBR methods intend to solve the above two problems. For example, a view map can encode the best photographs to be projected for a current viewing position, *i.e.*, a multi-view projective texturing. The missing and incomplete areas on the 3D model can be filled by assigning colors based on the appearance of their surrounding surfaces (Debevec, Yu, and Borshukov 1998).



Figure 2.38: Real-time projective multi-texturing with occlusion handling using a mix of point clouds and meshes. The viewer is capable of navigating inside a scene textured only by photos (Devaux and Brédif 2016). Credits to *iTowns*.

Another approach, as exhibited in Figure 2.38, can be followed by using a hybrid rendering technique, where images are projected in a 3D scene composed of both a mesh and point cloud. Geometric details and occlusion refinement can be handled by creating a detailed depth map (Devaux and Brédif 2016). Additionally, projective texturing can be extended even to real-world environments, where a refined geometric mapping allows projections into geometrically complex, colored, and textured surfaces. Subsequently, a radiometric compensation and a multi-focal projection can improve the consistency in the depth perception and minimize artifacts, *e.g.*, blur effects (Bimber et al. 2005). Further, illumination properties (*e.g.*, normals, reflection directions, and ambient occlusions) could be considered as well to improve the registration between the projected image and the corresponding geometry (Corsini et al. 2009).

2.3.2 Interpolating Between Views

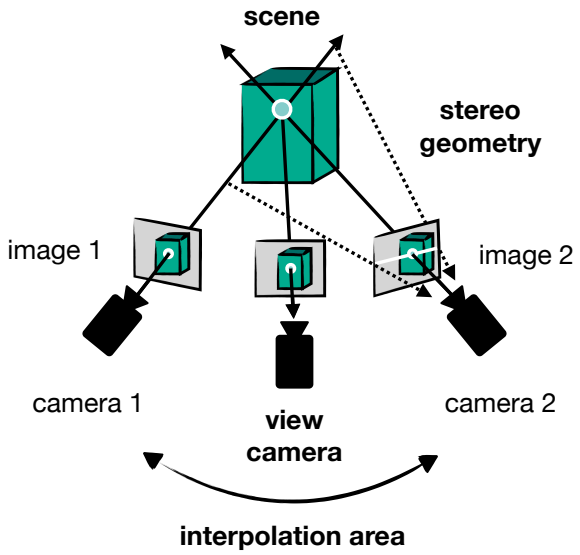


Figure 2.39: Interpolation between two images. Drawing adapted from Chen and Williams (1993).

Having described one of the primary Image-Based Rendering methods to represent photo-realistically areas of a 3D scene through the projection of input photos, we focus now on the techniques that interpolate novel views from two original photographs. View interpolation is seen as the process of **generating** a **series of synthetic** (computer-generated) **images** that, brought collectively, represent a **smooth transition** from one view of the scene to another (Manning and Dyer 1999). Essentially, these IBR techniques use image correspondence as geometrical information, as presented in Figure 2.39, where there is a correspondence of points in the scene between images one and two.

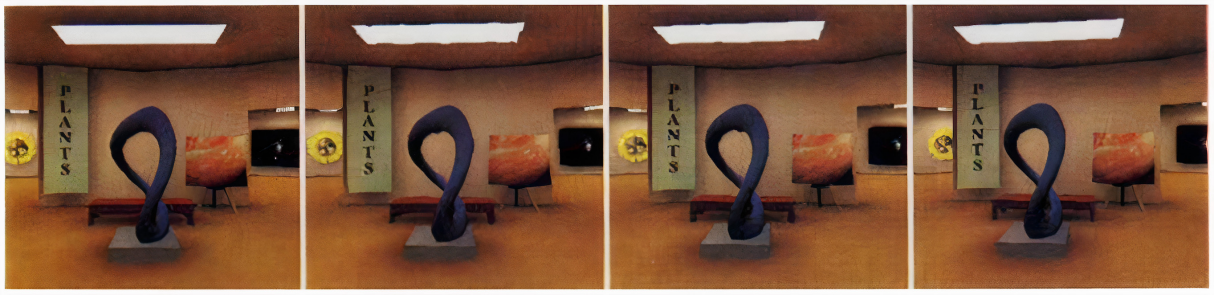


Figure 2.40: View interpolation using image correspondences. The two middle photos are interpolated ones (Chen and Williams 1993).

The **optical flow** represents the motion of the scene objects between consecutive frames caused by the relative movement between the objects and the camera. Based on it, the fundamental **view interpolation** method, interpolates the optical flows among input images. It is achieved by creating a morph map organized in a quad-tree[22] manner that contains correspondences between pixel blocks of the two views. During rendering, the colors for each block are interpolated and rendered back to front to avoid occlusion (Chen and Williams 1993). As shown in Figure 2.40, this approach is based on the observation that a series of photographs characterizing the same object are highly coherent, *i.e.*, closely spaced, from slightly different viewpoints. However, the result deteriorates when the difference between input photos is significant since the overlapping area is reduced.

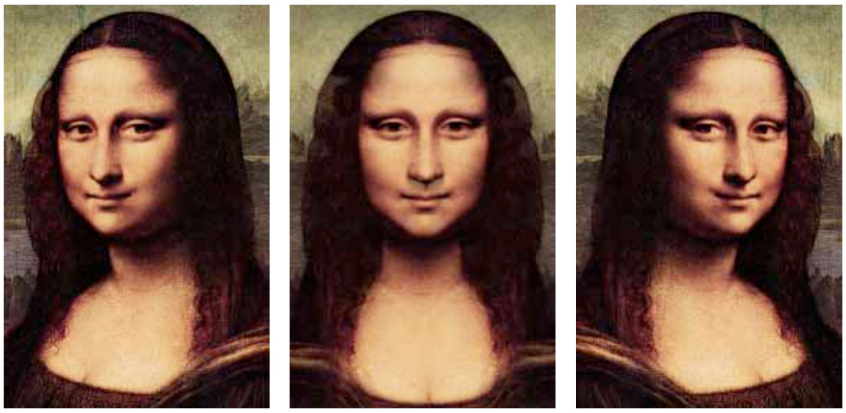


Figure 2.41: Mona Lisa View Morphing. The morphed view (center) is halfway between the original image (left) and its reflection (right) (S. Seitz and Dyer 1996).

[22] A **quad-tree** is a tree data structure in which each internal node has exactly four children.

The obstacle related to significant discrepancies can be handled by complementing the incomplete correspondences. One solution is to slowly transform specific features in one image into features in a second image, as Figure 2.41 depicts. **View morphing** follows this principle by modeling shape-preserving projections to allow a more effortless and natural transition between views (S. Seitz and Dyer 1996). The approach uses **image warping**, which is a process that digitally manipulates a picture in a way where the content is significantly distorted. Its basis is a three steps process: *(i)* pre-warping of two views from input photographs; *(ii)* computation of the morph (transformation) and interpolation of the photos; *(iii)* post-warping of each in-between image into the desired view.

Overall, **interpolation between views** does not demand a complex reconstruction of the 3D scene, nor does it depend on accurate depth maps. It is suitable for smooth transitions of viewpoints from picture to picture and has the advantage of low computational cost. Nowadays, it is a technique that is used spatially and temporally. Its applications extend to video manipulation since it allows interaction among dynamic scenes by handling (*e.g.*, freezing, slowing down, or reversing) time and changing viewpoints. (Gurdan et al. 2014; Kopf, Cohen, and Szeliski 2014; Stich et al. 2008; Zitnick et al. 2004). However, it is essential to notice that ambiguity may occur when there is a complex occlusion relationship in the scene, possibly affecting the quality of the rendering result. These methods face the challenge of dealing with occlusion, therefore.

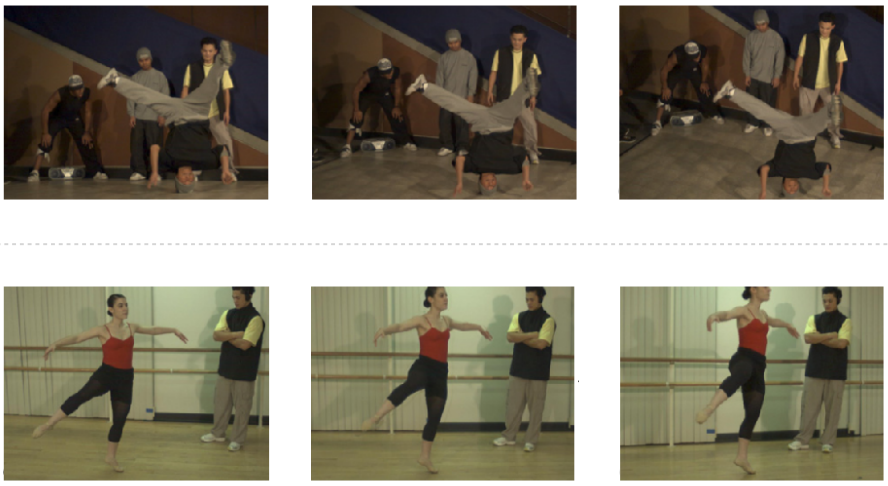


Figure 2.42: Video view interpolation for a dance and ballet studio. The interpolated view is the one in the center, while the border ones are the input photographs (Zitnick et al. 2004).

2.3.3 Artifact Correction

Rendering artifacts, *i.e.*, **any undesired or unintended alteration of the final rendered result**, may arise when inaccurate, approximate, and heterogeneous data is used. For instance, ghosting, aliasing, double images are typical challenges to all Image-Based Rendering (IBR) applications dealing with approximated 3D scene geometries. Distortion is another example observed during projective texturing, which is a deformation on an image due to many factors like the lens used and the position of the sensor in the camera. External agents like the changing illumination in the scene (when the photographs were captured) may also be considered IBR artifacts.

Uncertainty Refinement

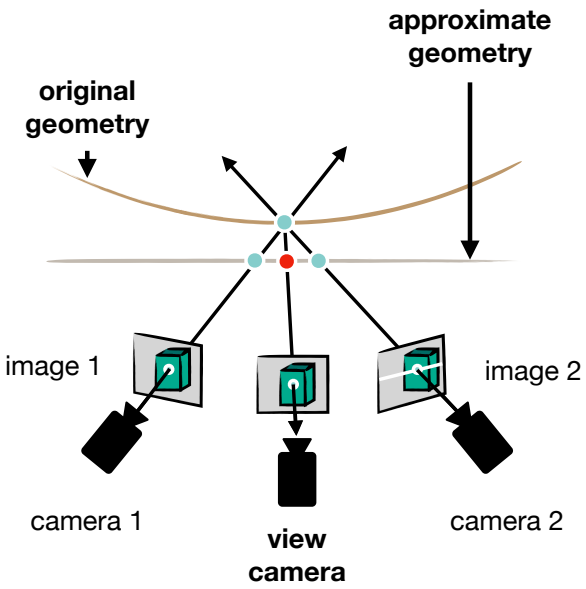


Figure 2.43: Uncertainty in multi-view projective texturing with approximate scene geometry. Figure adapted from Eisemann, De Decker, et al. (2008).

An **introduction of a filtering procedure**, *i.e.*, blurring in certain rendered areas, can improve the visualized result when a projective texturing approach is applied with inexact calibrated photos and an approximated 3D geometry (see Figure 2.43). This filtered blending can be accomplished using a low-pass filter to trade out aliasing (a stair-step look because of rasterization) with a radial blurring (Eisemann, Sellent, and Magnor 2007). An extent method, as presented in Figure 2.44, is the use of box-filtered blending on radial images. This silhouette-aware technique considers uncertain depth along a viewing ray and uncertain silhouettes that likewise should be blurred (Brédif 2014).



Figure 2.44: Silhouette-aware filtered blending (Brédif 2014).



Figure 2.45: Image transition in the Church dataset using a global geometry and an ambient point cloud (Goesele et al. 2010).

An alternative strategy is the *Floating Texture* approach, allowing **projective parameters to evolve and reduce visible artifacts** (Eisemann, De Decker, et al. 2008). It corrects image misalignments and preserves the crisp details in the appearance of the texture. Its basis is contained in two main stages *(ii)* an optical flow estimation between projected textures; *(ii)* a warping of the textures accordingly in the rendered image domain. Still, for strongly specular surfaces and badly color-calibrated cameras, the approach holds estimation problems.

As depicted by Figure 2.45, an *Ambient Point Cloud* reduces existing visual artifacts in regions where a 3D scene is uncertain (Goesele et al. 2010). The method uses a **point cloud with colored points to describe the inconsistent geometry** where the depth is uncertain. This colored point cloud, known as an ambient point cloud, is combined with a global model (either a point cloud or a mesh) to obtain the final view. As a result, the scene is rendered crisply for the reconstructed geometry during view interpolation while the unreconstructed part dissolves into the ambient point cloud. Consequently, the possible ghosting effects are reduced, and the visible cracks and holes are filled. Nevertheless, for the illusion to be consistent during the motion between images, the virtual view needs to lie on a line between the camera centers of the two transition photographs.

Geometric Distortion

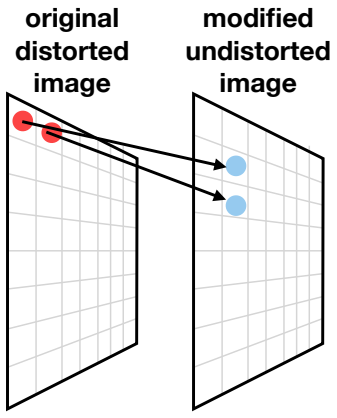


Figure 2.46: The mapping of the distorted and undistorted version of the image (Vass and Perlaki 2003).

A pinhole camera model does not hold well when Image-Based Rendering is used to render a scene textured by images from real-world cameras, *e.g.*, they may present aberrations like distortions. It may result in different artifacts, like image deformations and misalignments between the 3D model and photographs. Still, the distortion parameters may be obtained during the photogrammetric pipeline, providing more complex camera models. The use of a distortion model can modify and improve a IBR result. For example, lens distortion can be removed by **sampling, for each image position, an undistorted value using a lookup table**. As exhibited in Figure 2.46, this pixel value is calculated by mapping values between distorted and undistorted coordinates of the image to produce a distortion-free version (Vass and Perlaki 2003). Nonetheless, the resulting image may present some "empty" pixels when no input pixel is placed.

Another approach using distortion models can be followed by **transforming the points in the 3D scene**. For instance, a grid textured by an input image can be deformed, as shown in Figure 2.47, which will result in a distorted version of the original image (Keahey and Robertson 1996). This idea can be further extended to the vertices of a 3D scene. The result is the deformation of the resulting view (known as camera deformation) based on lens distortions and non-realistic projections (Spindler et al. 2006). Nevertheless, this strategy requires well-tesselated models to produce distortions of high quality, *e.g.*, the scene should be divided into many small triangles. Therefore, it is better suited for point clouds.

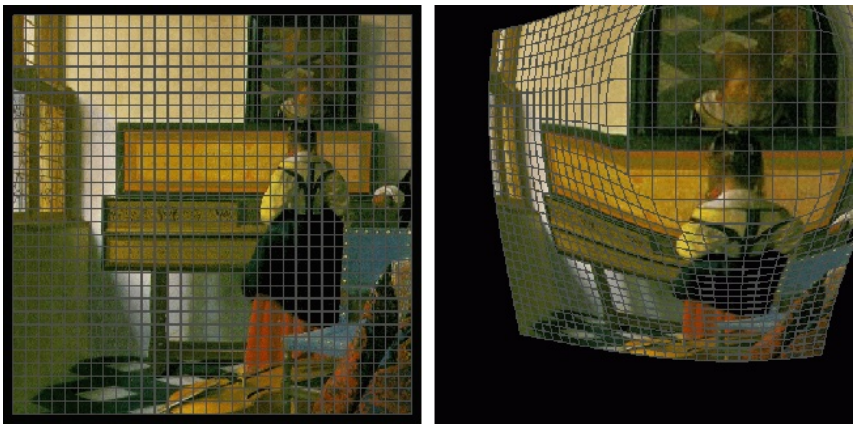


Figure 2.47: Deformed grid. **Left:** original image on a rectangular grid. **Right:** Altered grid resulting in a distorted image (Keahey and Robertson 1996).

The tessellation constrained can be removed by **amplifying the scene's geometry using predefined refinement patterns**, as shown in Figure 2.48. The basis is to dynamically refine the 3D mesh in the geometry shader (running only on the GPU). It is a three-pass algorithm: *(i)* a pattern selection; *(ii)* intermediate mesh update; *(iii)* rendering (Lorenz and Döllner 2008). The method can be applied even to non-pinhole models (with a single projection center) by approximating the projection through a set of perspective projections (Lorenz and Döllner 2009). It can operate in a dynamic scene but still requires high geometry processing that reduces the rendering speed.

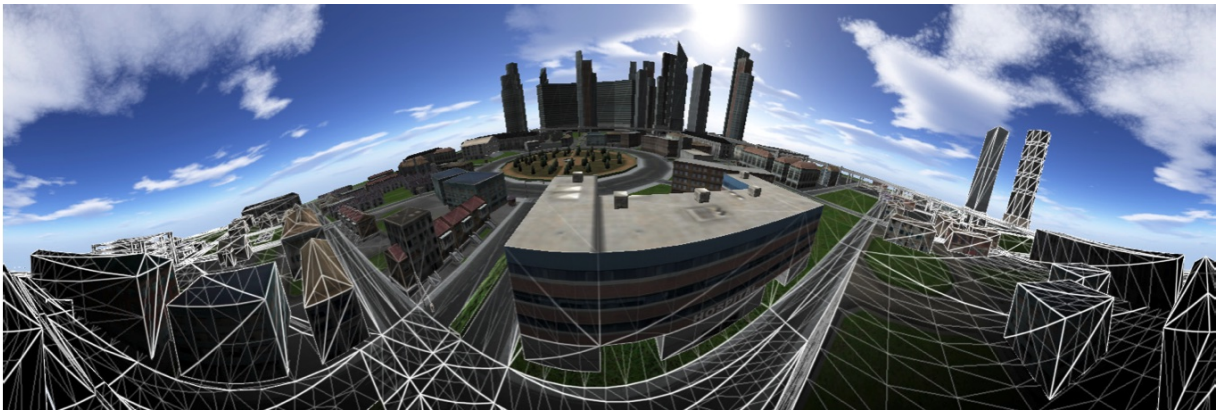


Figure 2.48: Panorama view generated using dynamic mesh refinement. The frame input mesh is delimited by thick lines, while the rendered primitives have thin lines (Lorenz and Döllner 2008).

Distortion can also be **corrected only in specific areas of the input pictures**. For example, an automatic algorithm can remove distortions from wide-angle portraits, specifically on facial regions. Given a photo, the facial regions can be detected using segmentation, and a distortion correction mesh can be generated. This mesh can be warped with the input image to locally adapt to the stereographic projection on facial regions and seamlessly unfold to the perspective projection over the background (Shih, Lai, and Liang 2019). However, a missing detection or erroneous segmentation can fail in the correction and introduce artifacts.

Radiometric Manipulation

Radiometric blending and alterations are plausible in an Image-Based Rendering environment. For instance, **a soft z-buffer can handle primary occlusion and blend pixels with similar depth values** to combine various image data. A new view can be rendered as a weighted combination of the three closest depth maps (Pulli et al. 1997). To remove more occlusion artifacts, multiples depth and color values may be used instead (Shade et al. 1998). As shown in Figure 2.49, a smooth transition between photographs with different lighting conditions can then be accomplished with the soft z-buffer and alpha normalization (Goesele et al. 2010).



Figure 2.49: Lighting changes in the Church (Italy) dataset. **Top:** transition at 0%, 50%, and 100% transition time. **Bottom:** close-up at 50% with soft z-buffer and alpha normalization (left), without soft z-buffer (middle), and without alpha normalization (right) (Goesele et al. 2010).

On the other hand, an extended work from *Photo Tourism* explores the idea of **maintaining the consistent scene appearance by modifying the color balance of the input images**, *i.e.*, color compensation (Snavely, Garg, et al. 2008). Still, this algorithm works well only with somewhat similar images. Related to histogram normalization techniques, histogram warping is a color correction method. Given an input image and a target color distribution, this approach constructs a

global color mapping function that transforms the colors of the image so it matches the target color distribution with any desired degree of accuracy (Grundland and Dodgson 2005).

Another possibility is the use of neural networks for image and style manipulation (Matzen and Snavely 2014a; Semmo, Isenberg, and Döllner 2017). Although, these methods always depend on the set of training photos to work correctly. An IBR example is another extension of *Photo Tourism*, which allows an image exploration across space and time. Given an extensive number of photographs taken at different times of the day, the system **synthesizes novel views to capture all-times-of-day scene appearance**, as presented in Figure 2.50. It is achieved by using a neural-scene representation called Deep Multiplane Images (DeepMPIs), which offers a way to represent viewpoint slices of the plenoptic function that interpolates between changes of appearances that happened over time (Li et al. 2020).



Figure 2.50: Spatial and temporal interpolation in which bot camera viewpoint and scene illumination change simultaneously (Li et al. 2020).

2.4 Strategies to Ease and Improve Navigation

A Graphical User Interface (GUI) allows humans to interact with electronic devices (*e.g.*, a computer) through visual representations. Different attributes of **visual perception** can be extensively used in any GUI design. The perception of **colors, structures, and patterns** is a fascinating topic because these are sensed exclusively through vision. Interfaces can become more accessible for users to perceive and **understand based on how humans structure visual stimuli**. Hence the understanding of visual perception can enable more compelling user interfaces (Marcus 1995; Rosenholtz, Dorai, and Freeman 2011).

Consequently, we may say that for geospatial data, the topic of visual perception related to the design of GUIs is essential since the objective of a geovisualization interface is to support the dialogue between a user and a map (or a 3D geographic environment) mediated through the computing device (Roth 2012).

Though, **what is precisely visual perception?**

2.4.1 Visual Perception

Like a camera, **the eye focuses light on the retina and creates an image** (see Figure 2.51). This light-sensitive surface (the retina) is a thin layer of photosensitive receptors that convert the light into a series of electrochemical signals transmitted to the brain. This process is call transduction, which is the transformation of one form of energy into another one (Goldstein 2010). It is achieved through two different **visual receptors: rods and cones**, which contain light-sensitive chemicals called visual pigments that react to the light and trigger the electrical signals.

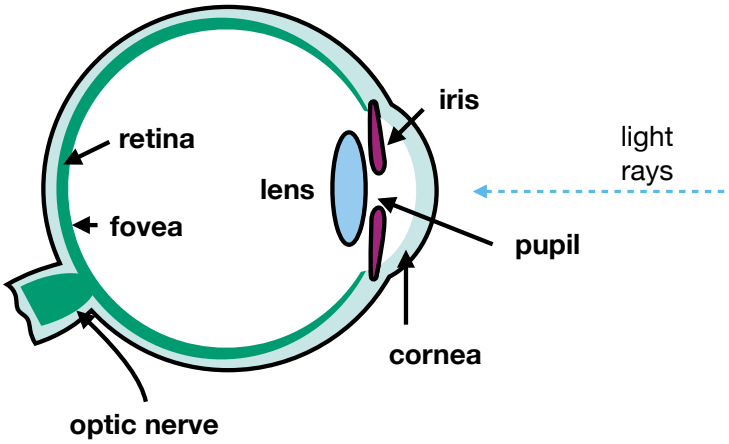


Figure 2.51: Elements inside the human eye. The light enters through the pupil and forms an image in the retina. Figure adapted from Snowden, Thompson, and Troscianko (2011).

Rods are responsible for night vision since they are designed to respond to low levels of light. They tend to be located in the peripheral area of the retina. Cones can be found in this region too but in lower quantities. They are concentrated in the small area in the retina center, known as the fovea, containing only cones. When looking directly at an object, its image falls on the fovea; therefore, **cones handle visual acuity tasks like color vision**. Color is achieved since cones come in three primary sorts, often called red, green, and blue, *i.e.*, since each one contains a photopigment that is most sensitive to the long, middle, and short wavelengths of light, respectively. Combining the signals from these three (cones) receptors allows a full range of colors (Cornsweet 2012; Snowden, Thompson, and Troscianko 2011).

Once the photoreceptors have processed the light, visual perception occurs in the brain, where the electrochemical signals flow from the ganglion cells through the optical nerve towards the cerebral cortex. These signals are passed via a network of neurons. This transmission step is crucial because **there is no perception if the signals do not reach the brain**. Overall, neural processing involves the communication between neurons to transform these signals into things we are aware of, perceiving, recognizing, and acting over the stimuli in the environment.

Therefore, we may say that **visual perception** can be seen as **interpreting the surroundings through the light that enters the eye**. This perceptual process, as illustrated by Goldstein (2010) in Figure 2.52, can be divided into a sequence of four steps that can determine the experience and reaction to any stimuli:

Stimulus: what is out in the environment, what we pay attention to, and what stimulates our receptors.

Electricity: electric signals that are created by the receptors and transmitted to the brain.

Experience and Action: the desired goal (perceive, recognize, and react to the stimuli).

Knowledge: any information that the perceiver brings to the situation.

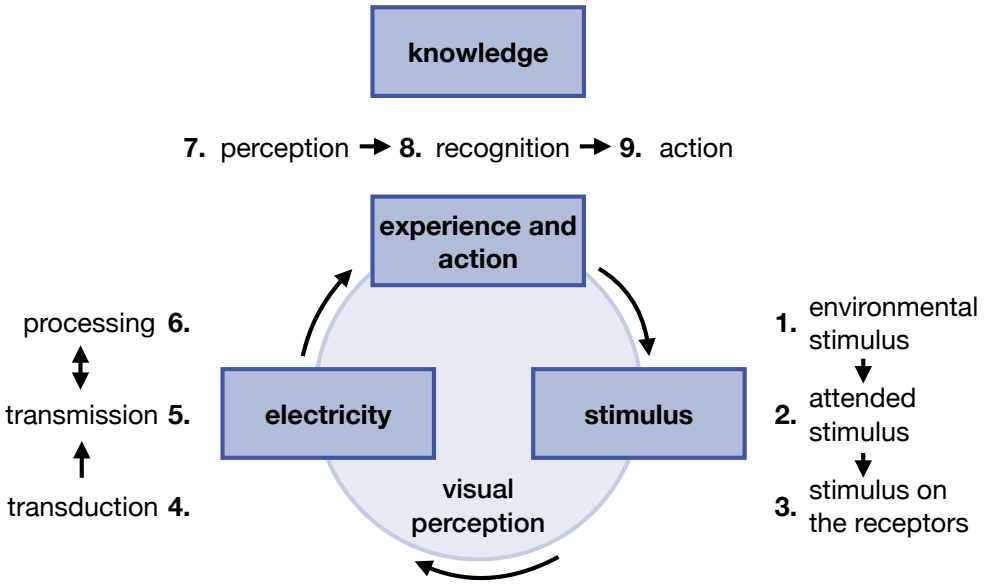


Figure 2.52: The visual perception process: stimulus, electricity, experience & action, and knowledge. The steps are arranged in a circle to highlight a dynamic and continuous change. Diagram adapted from Goldstein (2010).

2.4.2 Attention

Attention is a cognitive process that allows **concentrating on a piece of given information or external stimulus while ignoring other perceived stimuli** (Boynton 2005). It is a crucial mechanism for the brain since it provides a way to distribute the limited cognitive resources to process the vast amount of sensory input data that we receive every second. From the classical (psychological) point of view, attention can be categorized into four groups (Goldstein 2010; Snowden, Thompson, and Troscianko 2011):

Sustained: the ability to focus on one specific task for a continuous amount of time without being distracted.

Selective: the capacity to select from many stimuli and focus only on the desired one while filtering other distractions.

Alternating: the faculty to switch the focus back and forth between tasks that require different cognitive demands.

Divided: used to complete two or more tasks simultaneously.

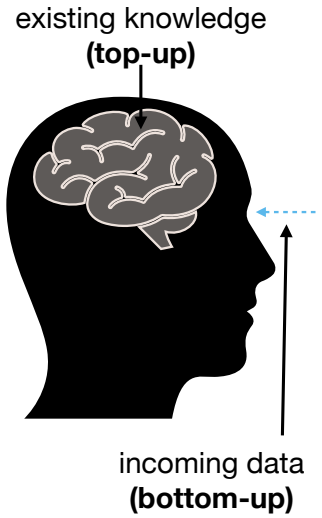


Figure 2.53: Attention is affected by the interplay between bottom-up and top-down processing.

In this work, **we are interested in selective attention** (Luck and Ford 1998), which is aimed at one stimulus, but where it is possible to jump sequentially between different ones (alternating and divided attention). As Figure 2.53 depicts, selective attention emerges from a combination of bottom-up (the stimulus) and top-down (knowledge and expectations) factors. Our brain uses Saccadic Eye Movement (SEM), a "gaze jump" from one region to another of the visual scene to achieve selective visual attention. It allows us to scan and build a mental map of our surrounding environment, as portrayed in Figure 2.54, where Land and Hayhoe (2001) shows the sequence of eye fixations of a person making a peanut butter sandwich.

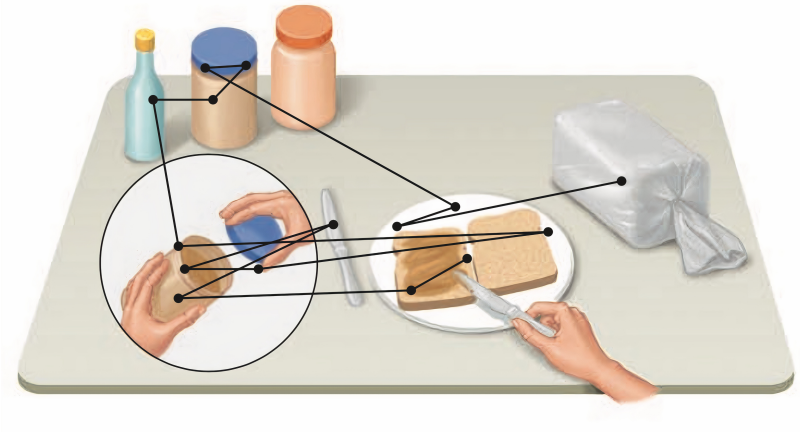


Figure 2.54: Scan path of a viewer while preparing a peanut butter sandwich. The initial fixation is on the bread. Image is taken from Land and Hayhoe (2001).

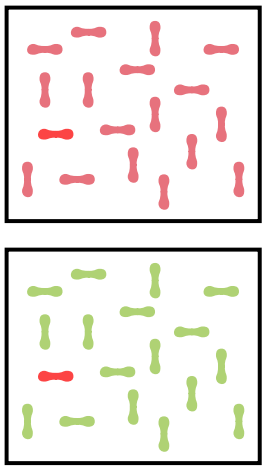


Figure 2.55: Visual salience comparison. Find the horizontal red target inside each black rectangle.

What is Saliency?

Our attention is generally drawn to salient stimuli. Visual **salience** is a **subjective and perceptual characteristic specific to objects in the surrounding environment that can instantly grab our attention** (Goldstein 2010; Itti 2007). It refers to physical elements such as color, brightness, contrast, or orientation, *e.g.*, an element could be salient because it is the only blue component or the only horizontal line, among many. Capturing the attention using salient stimuli generally depends on the stimulation patterns falling on the receptors (rods and cones), *i.e.*, bottom-up factors. However, cognitive factors (top-down) can also be influential (Treue 2003). For instance, as represented in Figure 2.55, a red target in a bunch of pink distractors would be less salient than the same red target in a bunch of green distractors. Nevertheless, if it is known that the target is a horizontal line, this information can give greater weight to all horizontal items.

Therefore, we define **salience (or saliency)** as the **visual quality** of any object, particularly **noticeable or prominent**.

2.4.3 Ease of Navigation

When visualizing geospatial datasets, *e.g.*, in a 2D map or a 3D Digital Earth, some **potential solutions** such as abstraction or saliency of certain elements may **reduce visual complexity** in the respective GUIs presented to the final users. This section introduces and evaluates existent geovisualization techniques that tackle this objective based on the user visual perception.

Representation

Firstly introduced by Bertin in 1967, a **visual variable** can be defined as **a feature of a graphical object that can visually stand out from other objects** (Roth 2017). The concept is based on semiotics, which studies signs and symbols for their use or interpretation. Visual variables remain a core notion of visualization since they can be controlled during any design process. They allow considering the graphical representation of any data at a fundamental level (Carpendale 2003; Garlandini and Fabrikant 2009). As depicted in the diagram adapted from MacEachren, Roth, et al. (2012) and Roth (2017) in Figure 2.56, the primary visual variables are:

- **Size:** the amount of space occupied by a symbol. It can be used to convey information such as quantitative or related to importance.
- **Shape:** the external form, *i.e.*, the outline of a symbol. It points features on one or many symbols.
- **Orientation:** the direction (or rotation) of the symbol.
- **Color Hue:** the dominant wavelength on the visible portion of the electromagnetic spectrum, *e.g.*, red, green, and blue. It is the visual perceptual property that corresponds to categorize a color.
- **Color Value:** the relative amount of energy emitted or reflected by the symbol. It refers to how light or dark the object shows.
- **Color Saturation:** the spectral peakedness of the symbol across the visible spectrum. It is the purity (or intensity) of the color.
- **Texture:** the coarseness of the fill pattern within the symbol.

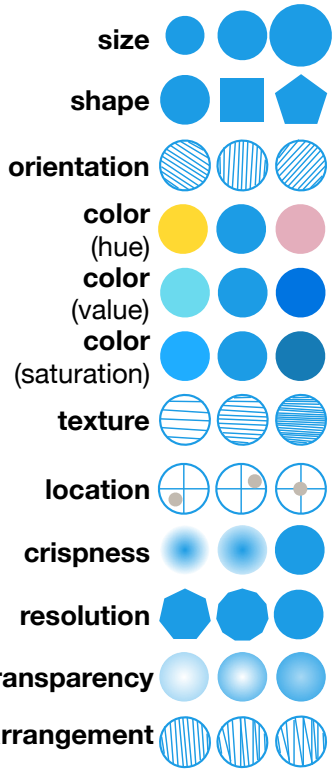


Figure 2.56: Primary visual variables. Illustration adapted from MacEachren, Roth, et al. (2012) and Roth (2017).

- **Location:** the position of the symbol related to a coordinate frame. It is the absolute location of the element.
- **Crispness:** the degree of sharpness or fuzziness of the boundary of the symbol. It can also be referred to as the depth of field in information visualization.
- **Resolution:** the spatial precision at which the symbol is displayed.
- **Transparency:** the amount of graphic blending between the symbol and the background.
- **Arrangement:** represents the layout of graphic marks. It varies from regular, perfectly aligned in a grid-like structure to irregular, randomly placed, or clustered.

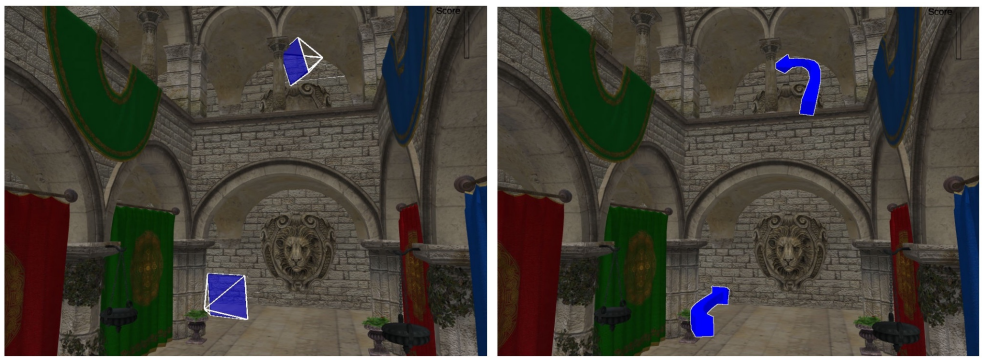


Figure 2.57: 3D bookmarks used to move to a new viewpoint in a scene. **Top:** the geometry of the camera. **Bottom:** an arrow. Photos are taken from Forgione et al. (2016).

Initially, only the first seven variables were proposed; however, the basic idea has been extended to include many other factors like digital and 3D visualization (Çöltekin, Griffin, et al. 2020), *e.g.*, to facilitate users to **recognize existing spaces and dynamics in a scene**. For instance, when browsing through photos, a higher dynamic interaction inside a 3D environment can be obtained by representing each photograph's spatial information (camera orientation and position) with **bookmarks**. As shown in Figure 2.57, these particular markers can be displayed in many manners, including a text link, a thumbnail image, or even 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 located too far away from the current view.

Large Data Volume

Spatio-temporal information is a challenging input since it represents **large amounts of data available to be visualized and explore**. Therefore, we need to stretch the capacity of existing tools to **manage such datasets and the corresponding visual complexity** they may bring. For instance, a significant difficulty users encounter when browsing inside photo collections is finding the desired content (Ardissono, Kuflik, and Petrelli 2012; Baruzzo et al. 2009). A standard approach, like the one used by *PhotoTourism* (Snavely, S. Seitz, and Szeliski 2006), *4D Cities* (Schindler and Dellaert 2012), and *Smapshot* (Blanc, Produit, and Ingensand 2018; Produit et al. 2018), 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 (see Figure 2.58). Still, as the number of pictures increases, so will the searching time.

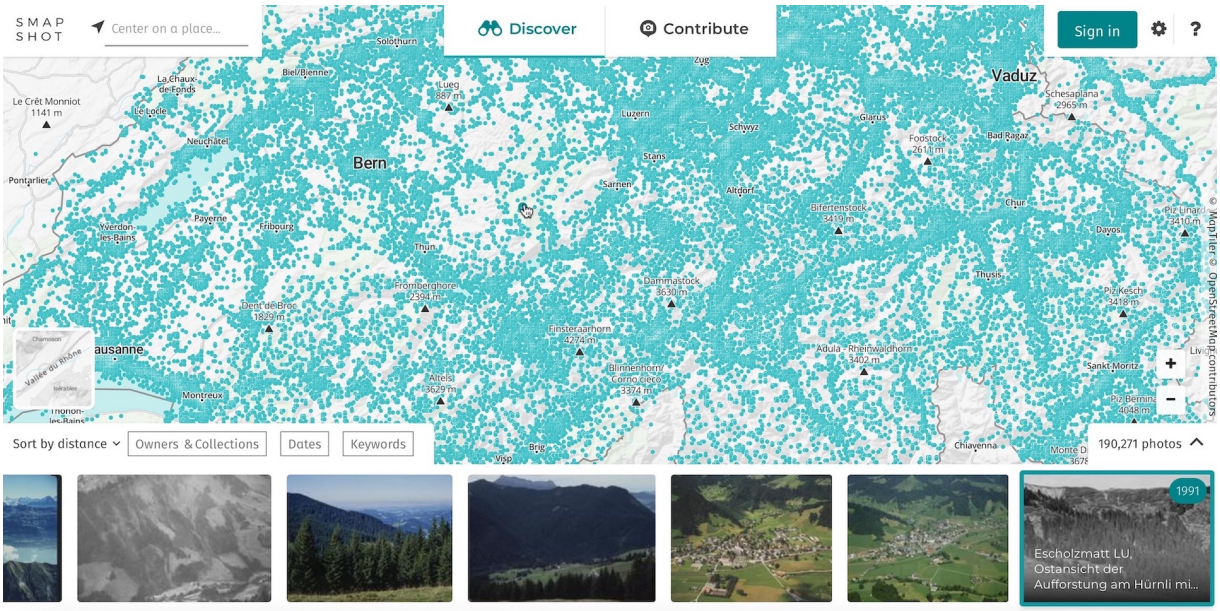


Figure 2.58: Thumbnail menu slider employed to browse over the available photographs. Credits to the ETH Library and *Smapshot*.

The **clustering of data** can be seen as an alternative solution for browsing over image compilations. It is a process where the photos are partitioned into a set of meaningful categories (Xu and Wunsch 2015). For example, in a 2D or 3D visualization like the ones presented by *Pixplot* and *UrbanHistory4D* (Dewitz et al. 2019; Maiwald, Bruschke, et al. 2019; Maiwald, Henze, et al. 2019), clustering can facilitate the exploration of several photographs in a geographical view by **grouping the images by their spatial location**. As shown in Figure 2.59, instead of visualizing the photographs all over the current view, only the groups are visible, and when selected, they can be expanded to the complete set of images.

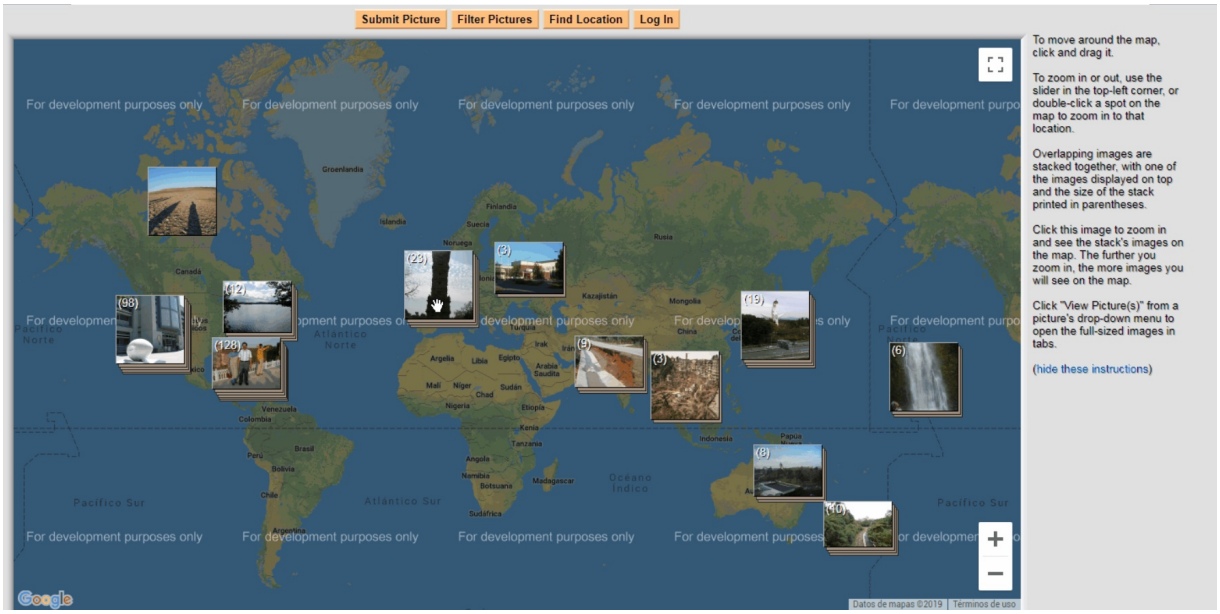


Figure 2.59: Clustering of photos in a 2D map. Credits to *Pixplot*.

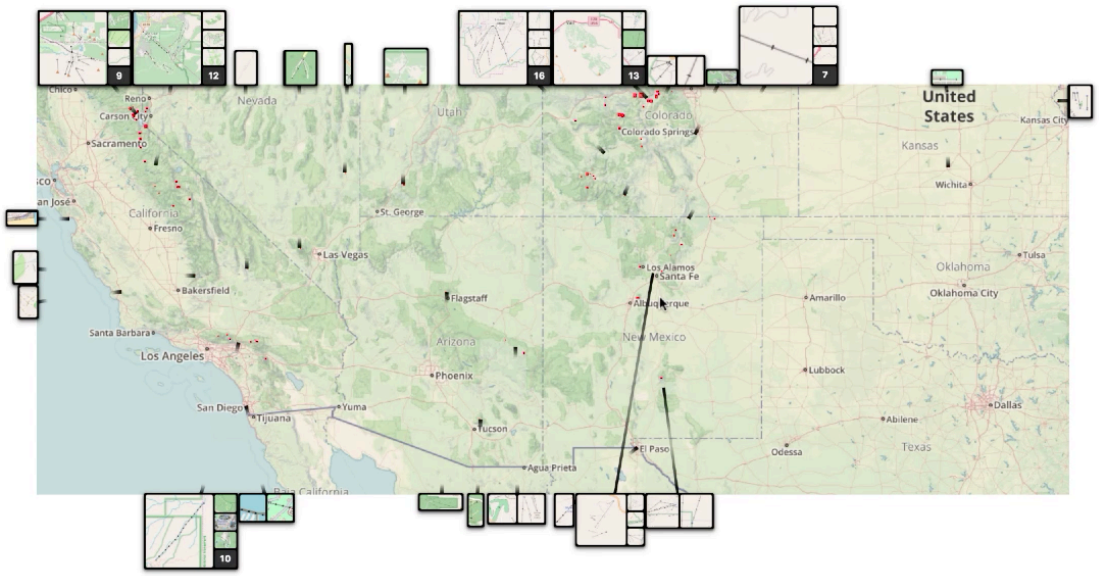


Figure 2.60: Scalable insets applied to a geographic map. Image is obtained from the presentation of Lekschas et al. (2020).

Alternatively, an extension of *PhotoTourism* (Simon, Snavely, and S. M. Seitz 2007) and *ClustTour* (Papadopoulos et al. 2010) propose a clustering procedure that can describe a scene by selecting a set of canonical views to form a scene summary. Both applications partition the photo collection into groups of related photographs based on visual features. More complex techniques involve the **computation of semantic distances to cluster the geospatial information**. The result can be placed inside a thumbnail bar employed to navigate the data using a dynamic image hierarchy where each cluster is depicted as a stack of photos, as done by *Photo Cloud* (Brivio et al. 2013). *Scalable Insets* (Lekschas et al. 2020) in Figure 2.60 presents a more creative solution that suggests clustering the images and representing them as insets (thumbnails) in a gallery form. The visual representations of these clusters evolve as the zoom level varies in the geovisualization system.

Areas of Interest

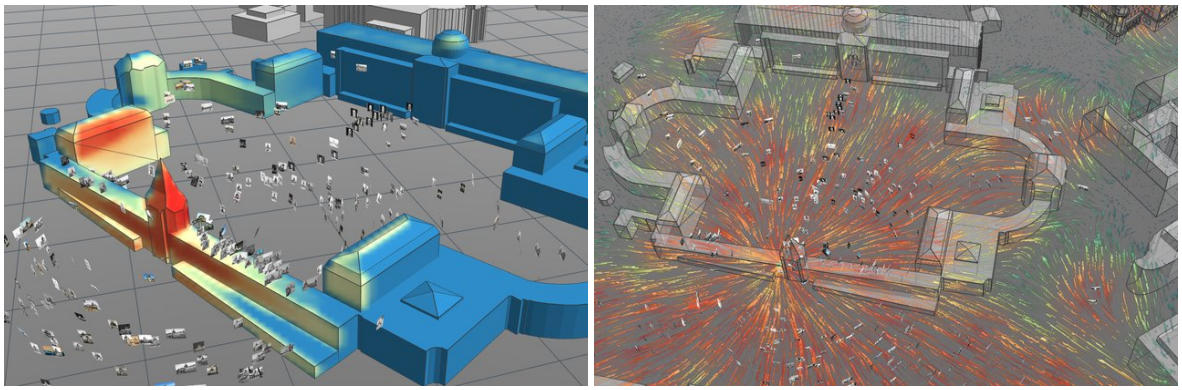


Figure 2.61: Graphical depiction of orientations from spatialized images. **Left:** Heat map on building's surface. **Right:** Particles depicting camera orientation. Images are taken from Bruschke, Niebling, and Wacker (2018).

The **spatial distribution of items in an image collection** can provide a global overview of geospatial data. This information can be described through a 2D map where higher color intensities depict higher recurrences, *i.e.*, a **visibility heat map**. For example, it can highlight the most attractive areas to visit in certain mountains (Chippendale, Zanin, and Andreatta 2008) or point towards the location of the photographer's point of view. The area on a 3D scene represented by an image can be determined by projecting it to the 3D model, leading to the computation of points of interest (Niebling et al. 2020). As Figure 2.61 exhibits, this idea can be extended using the dynamic of fluids to account also for the camera orientation information (Bruschke, Niebling, and Wacker 2018).

Chapter 3

Formalization of the Problem

Preamble

This chapter exposes the **scientific challenges** regarding the **visual exploration** of large compilations of **historical photographs inside 3D digital topographic scenes**. Based on it, we address our intended solutions tackling these subjects.

Section 3.1 formalizes the problem that we face related to the association of diverse types of spatio-temporal information. Next, Section 3.2 describes the main objective of the work to achieve a geovisualization system supporting historical photographs, along with the more specific topics dealt with within the following chapters. Finally, Section 3.3 introduces the basis of our proposed approach to solve the before stated obstacles.

Contents

3.1 Problem Description **52**
3.1.1 Sub-Problem 1: The Integration of Heterogeneous Spatial Information Causes Potential Visual Discomfort 55
3.1.2 Sub-Problem 2: The Navigation Through a Massive Amount of Visual and Spatio-Temporal Data Represented on a Large Scale 56
3.2 Objectives of Research **57**
3.3 Proposed Approach **58**
3.3.1 Approach 1: Extrapolation on the Geometric Distortion for 3D Visualization . 61
3.3.2 Approach 2: Navigation Techniques to Reduce Visual Complexity 61
3.3.3 Approach 3: Co-Visualization of Historical Photos and 3D City Models 62

3.1 Problem Description

This work aims to design and implement a **visual exploration method for extensive collections of historical images representing geographic spaces**, *e.g.*, aerial and terrestrial photos. We define our input data as follows:

1. The **historical photographs** each one depicting a snapshot of a geographical location in a specific time frame. Notice that these photos contain temporal information too. The date of acquisition may be known or at least an approximated time span. From each one, we assume that the following is known:

a) The image **metadata** summarizing and describing each photo's available information, *e.g.*, date of acquisition (or approximate), title, or photographer name.
b) The camera **orientation information** maps the coordinates in the photo to those in the digital 3D model(s). It contains:

Extrinsic: a transformation (rotation and translation) from 3D world coordinates to the 3D camera's local coordinates.

Intrinsic: a transformation from the 3D camera's coordinates into the 2D image coordinates (includes principal point and focal).

Distortion: a mathematical function that models the geometric deviation from the ideal pinhole model, defined as part of the camera's intrinsic parameters.

2. A **topographic 3D model** that enables the navigation of a virtual camera in a digital 3D environment. Different types of 3D models exist and can be used to model a digital scene. *e.g.*, a reconstructed 3D point cloud or a contemporary 3D city model containing 3D objects like the ground surface and buildings. Therefore, like the photographs, the 3D model also has temporal information. This temporal representation may differ from the photos; for instance, the model could have been created based on recent data.

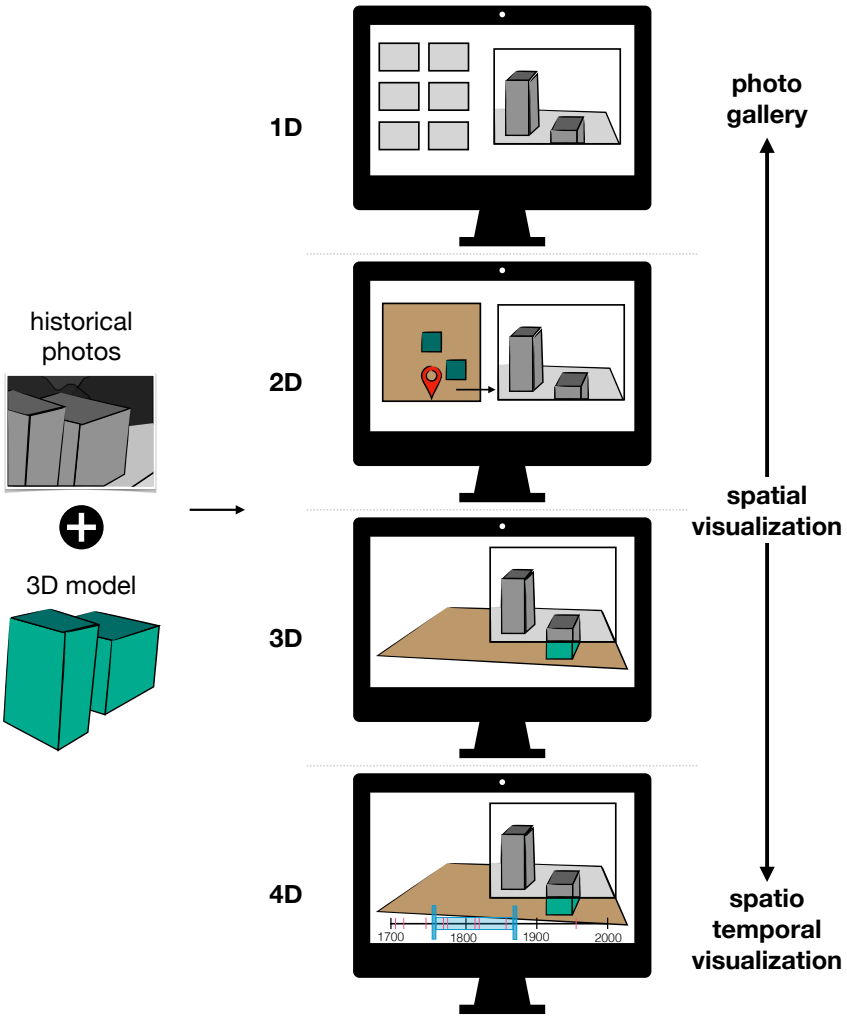


Figure 3.1: Evolution of the main photo visualization techniques. It shows the progression in the visualization methods from a primary display to more complex representations.

Many possibilities exist to **browse over compilations of photographs** representing geographical spatial information: **from a primary gallery display to more complex ones, utilizing the spatial and temporal components to facilitate locating the photos in a topographic view over time t** (see Section 2.2 of Chapter 2), as depicted in Figure 3.1. Even though simple meta-data is often the primarily used method, it does not provide any complex interactions beyond filtering. An extension of this technique is a spatial exploration strategy that can locate the user in a geographic context. However, 2D approaches may be too limiting, *e.g.*, the user may be constrained to an orthogonal top view or only to a few street-level perspectives. Therefore, the introduction of a third dimension can potentially amplify the navigation experience. Users will no longer be restricted to any specific position inside a 3D spatial model of the scene. This technique can be furthermore amplified to spatio-temporal, *i.e.*, adding the time variable t. As a result, the user will gain a comprehensive understanding of the temporal structure of the visualized data.

On this note, now lets us focus on our research context; our aim is historical photos. Nevertheless, datasets of CH archives are a challenging type of input. From a temporal point of view, the **topographic 3D model used in the visualization could be challenging to obtain or vary in time from the acquisition dates of the pictures**. For instance, traditional photogrammetric 3D reconstruction techniques are insufficient for archival imagery, such as those used by *PhotoTourism* (Snavely, S. Seitz, and Szeliski 2006) and *PhotoCloud* (Brivio et al. 2013). This complexity can be attributed to the fact that *(i)* the scanning process may result in low-quality digital images; *(ii)* a processing algorithm may not be able to match feature points due to low overlapping areas or significant changes of perspective between the photos; *(iii)* the majority of the parameters of the acquiring device(s) may be unknown. Therefore, a **3D reconstruction may be unachievable due to the incompletenes and insufficiency of the historical information**.

Still, strategies like the ones proposed by *Smapshot* (Blanc, Produit, and Ingensand 2018; Produit et al. 2018) or *UrbanHistory4D* (Dewitz et al. 2019; Maiwald, Bruschke, et al. 2019; Maiwald, Henze, et al. 2019) can provide a good visualization environment **utilizing some existent and modern geographical information to produce a contemporary 3D city model**. Even though this data may contrast in appearance with the old photos, *e.g.*, due to progressively environmental changes, it can provide enough material to present the user with an essential context of the photographed scene. These possible morphological incoherences can be addressed with a correspondence selection step (on the photogrammetric process) using the unmodified parts of the scene (De Luca et al. 2010).

However, these 3D models may not be existent in all circumstances, so this solution cannot be followed all the time. Hence trying to solve the research issue of browsing and visualizing historical image collections may raise several difficulties related to the type of input data available, which takes us to the definition of our general research problem.

General Problem

This dissertation addresses the following research question: **how to associate, visualize, and browse over heterogeneous and extensive historical information?**, which leads to our two specific research questions described as follows.

3.1.1 Sub-Problem 1: The Integration of Heterogeneous Spatial Information Causes Potential Visual Discomfort

A co-visualization visually links into a 3D environment diverse data. To be able to achieve an immersive **co-visualization** of historical photographs, we need to combine two types of data seamlessly: *(i)* 2D information in the form of **historical images**, which each represent a single snapshot of a moment; *(ii)* **3D geographic information** that is continuous in time and space. This association may result in some visualization challenges.

Let us consider **presenting the historical photos in the context of a 3D model** of the photographed scene. A virtual camera used inside a 3D environment containing the 3D model can be placed in the same position and orientation as the camera used to take a photograph. Following a projective texturing strategy, a pixel-accurate reprojection of the photo can be obtained under these conditions, displaying the photograph's exact representation. A wider zoom allows the visualization of the 3D model around the old picture of interest, which may be shown as an extra context.

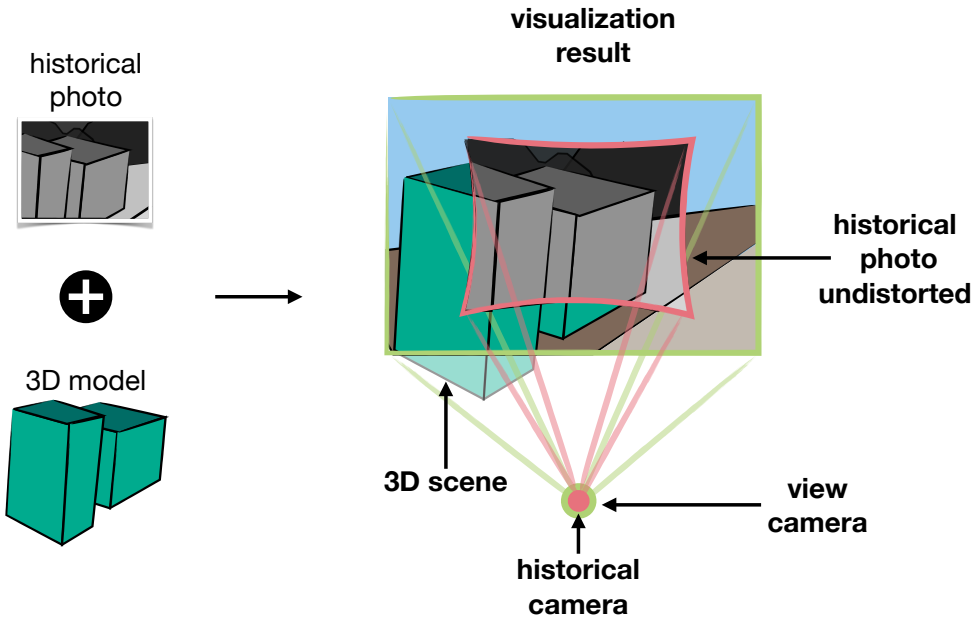


Figure 3.2: Illustration of the virtual view of a 3D environment textured by an unprojected old photograph.

The photographic acquisition process may exhibit slight to severe distortion, *i.e.*, a geometric deviation from the ideal pinhole camera model. Under the previously mentioned circumstances, as depicted by Figure 3.2, when the **distortion** attribute is **considered** during the **projective texturing** step, the visualization (through a pinhole camera model) of a 3D scene textured by a historical photo with distortion will effectively display an **undistort version of the image**. This undistorted representation will result in a non-rectangular shape, which may be **counter-intuitive or distracting for the end-user**. Historical photographs specifically should not be deformed due to their cultural value.

Sub-Problem 1

How do we keep the rectangular shape of a historical photo and still have the context information that a 3D environment provides?

3.1.2 Sub-Problem 2: The Navigation Through a Massive Amount of Visual and Spatio-Temporal Data Represented on a Large Scale

To cross-analyze different photos and collections, we need a system able of a multiple image exploration. Let us consider the **discovery** of **many photographs in our digital 3D environment**. For the user to select and interact with these photographs, *e.g.*, project an image onto the scene, get the meta-data of that picture, or move to the photographer's viewpoint, navigation support is necessary. Using a virtual camera (free to move around), a user can navigate and explore all available pictures by *(i)* taking the position and orientation of a specific camera representing a historical image; *(ii)* engaging in different angles and visualizations through the 3D model.

However, as the number of images increase, so does the amount of visual data that needs to be managed. Although **old pictures** may be sparsely sampled items, they **represent a large amount of data** (Section 1.2 of Chapter 1). Remarkably, this potential volume has to be considered within the design of a historical image exploration approach.

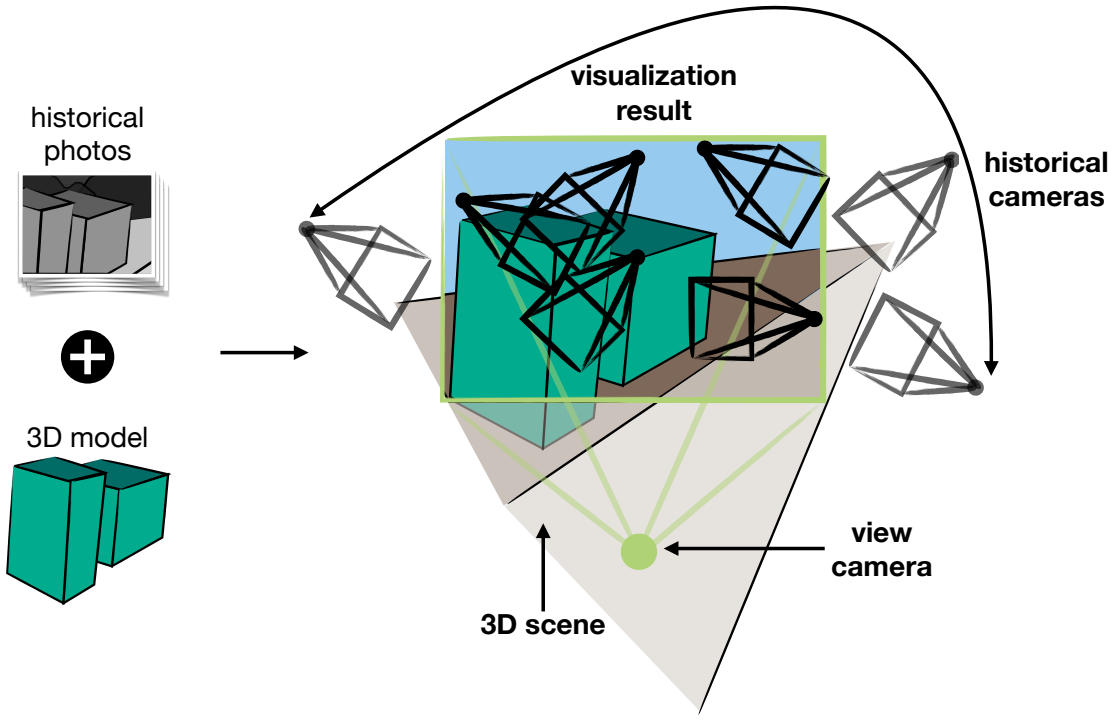


Figure 3.3: Depiction of a visualization result when many historical images are located in a 3D scene.

As illustrated in Figure 3.3, the **visual complexity** is magnified when we consider temporal information and visual interaction objects like viewpoints and thumbnails. We deal with massive

and heterogeneous information representing large scales. Therefore, we face the issue of managing all this visually intricate information collectively.

Sub-Problem 2

How do we differentiate, associate, and visualize these historical image compilations within their co-visualization with a topographic 3D model?

3.2 Objectives of Research

General Objective

We want to **enable users to navigate and visualize a 3D environment enhanced with old photos**. As visually stated by Figure 3.4, we define the **main objective** of this thesis as the proposition of a method for a **spatio-temporal**, **immersive** (in a 3D space), **interactive** (real-time response), and **continuous** (browsing spatially and temporally) **exploration** of extensive heterogeneous collections of historical images. Overall, we divide our two particular goals as *(i)* artifact correction; *(ii)* graphic representation.

Note that this thesis does not focus on calculating the photographs' spatial attributes or reconstructing the 3D scene from the overlapping images. The envisioned model is instead directed only on the **visual** and **interactive** aspects.

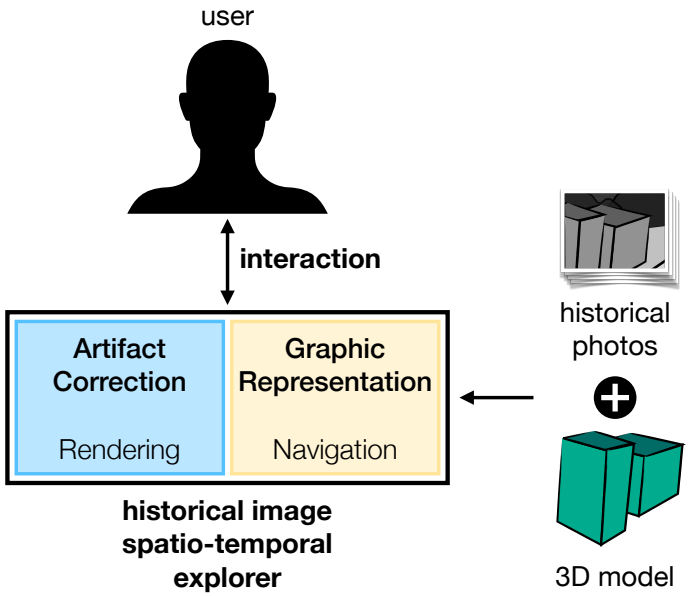


Figure 3.4: Objectives of research.

Specific Objectives

Artifact Correction: for the **potential visual discomfort** when integrating heterogeneous spatial information. Our goal is to render historical views. We want to provide a method that can be followed through the rendering process, where the photo will be visualized in its original rectangular shape and without misalignments with the 3D scene caused by the geometric distortion present on the image.

Graphic Representation: for the **reduced navigation and interaction** on massive visual and spatio-temporal data represented in large scale. Our goal is to improve every photo's saliency and reduce the visual complexity obtained when many photographs are visualized. We aim to propose a set of procedures that will facilitate and enhance navigation for users to discover many historical images individually and combined in a 3D environment.

3.3 Proposed Approach

The process of generating virtual views from digital environments is called rendering. It **simulates how light travels in space to generate an image from a virtual 3D scene** (Section 2.3 of Chapter 2). It is mainly done by placing a virtual camera in the 3D environment and **reproducing the same 3D to 2D mapping that a camera has**. The overall process can be done through the Graphics Processing Unit (GPU), which employs small programs defined as **shaders** to create the rendered image. As depicted by Figure 3.5 by Vries (2020), from a set of vertex received in the vertex shader, the geometry of the 3D scene is assembled and (possibly) modified in the geometry shader so that each pixel can be "painted" in the fragment shader.

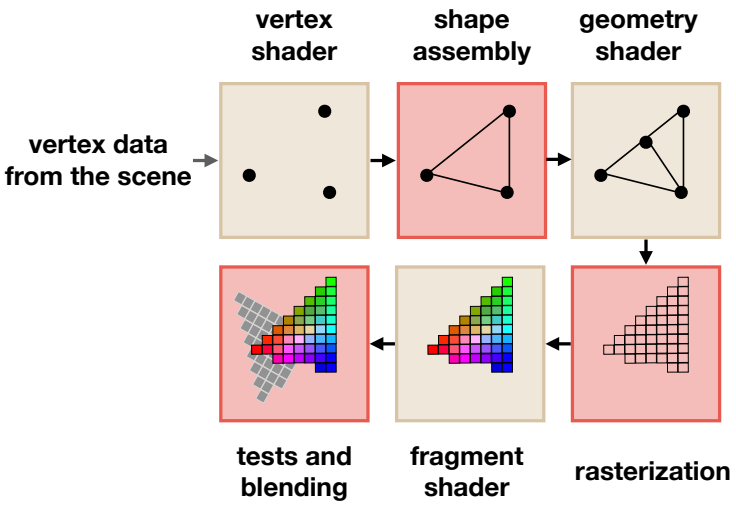


Figure 3.5: Shading process. Three shaders are applied: vertex, geometry, and fragment. The diagram is adapted from Vries (2020).

This work aims to design and implement a **visualization and navigation method for large compilations of historical images**. We decided to introduce an **exploration technique for historical images adapted to any geometric representation of the scene**, *i.e.*, considering that in some cases,

a 3D model may not exist or its (stereo image-based) reconstruction may be impracticable. Figure 3.6 depicts the different existing rendering methods based on the input information available. Since we want a visualization procedure independent of the 3D scene, we cannot use geometric model-dependent techniques, *e.g.*, Physically-Based Rendering (PBR).

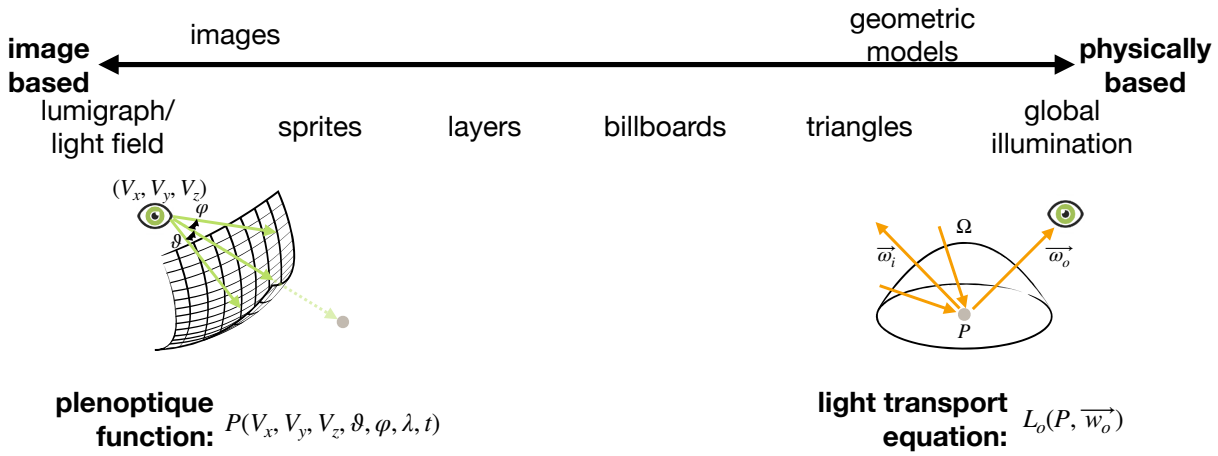


Figure 3.6: Rendering spectrum, adapted from Akenine-Möller, Haines, and Hoffman (2019). It reveals a progression in the rendering techniques that increasingly decouples itself from geometry (right to left).

Accordingly, our selected solution, **Image-Based Rendering (IBR)** (Gortler et al. 1996; Levoy and Hanrahan 1996), is more oriented to our input data, *i.e.*, images. This method can generate a result of any unknown viewpoint by interpolating through the input photographs or re-projecting their pixels. The benefit of this type of rendering is that it may cope with little or no geometrical information of the scene (Buehler et al. 2001). Furthermore, the photographs can be visualized through a rendering process that is particular for the characteristics of these old pictures.

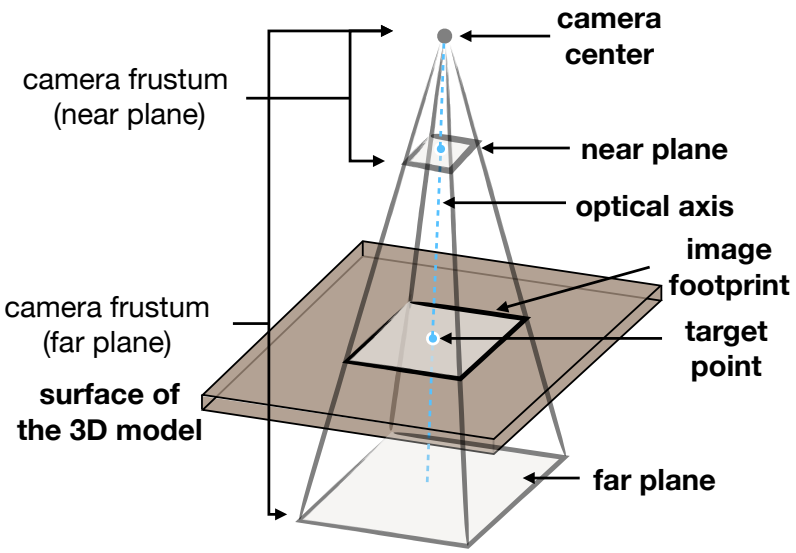


Figure 3.7: Geometrical components defined in the projection of a 2D image on the top of a 3D model.

Hence as shown in Figure 3.7, to display an image inside a 3D space, we apply the concept of projective texturing and use the geometrical information known from the camera that acquired the

picture to **project the photograph on the top of the 3D surface** of the city (or topographic) model. We define the main components as:

- The **camera center** is a 3D point representing the position of a camera in space. It is defined as the point of intersection between the lens' optical axis and the sensing plane.
- The **near plane** is located at a particular close distance to the camera, along the optical axis, where all objects in the scene closer than the plane to the camera are not rendered.
- The **far plane** is determined at a distinct far distance to the camera, along the optical axis, where all objects in the scene farther than the plane are not rendered.
- The **target point** is a 3D point depicting the camera's optical axis intersection with the 3D model.
- The **camera frustum** is a 3D volume describing the camera's geometry in the form of a pyramid truncated by the camera near or far planes.
- The **image footprint** is the portion of the city model surface containing the projection of the image content and its outline.

Within this context, in our Image-Based Rendering approach, we define two cameras to visualize a photograph in the 3D model. As Figure 3.8 shows, a camera designated as the **historical camera** models the camera used to acquire the picture. The visualization through this camera will be as looking directly at the image that the camera describes. This camera may or may not have a distortion model that describes the geometric distortion of the photographic image acquisition process. Simultaneously, a second camera is responsible for the navigation through the 3D environment and is specified as the **view camera**. The camera is utilized during the rendering process and allows all the movements within the scene, *e.g.*, zooming, translating, and rotating. Consequently, as the number of photos increase, it will also do the number of historical cameras.

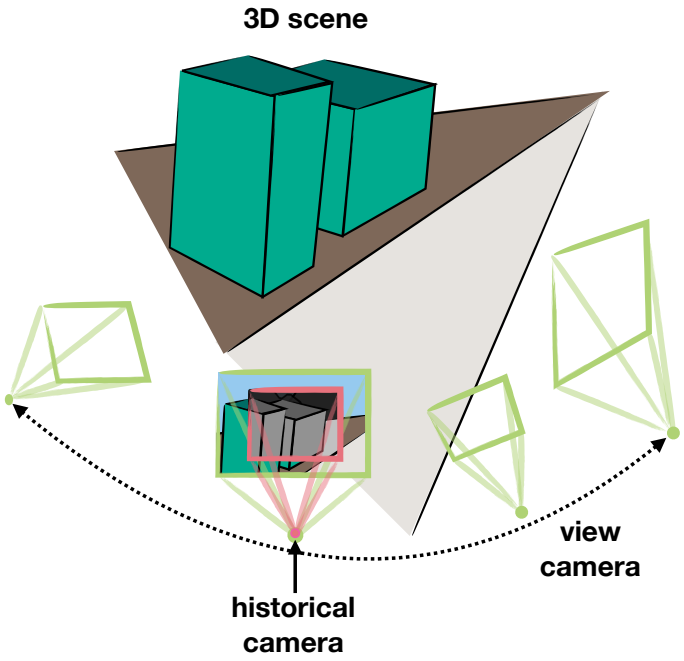


Figure 3.8: A historical camera has a fixed viewpoint. A view camera is capable of changing its viewpoint, allowing it to navigate through the 3D scene.

3.3.1 Approach 1: Extrapolation on the Geometric Distortion for 3D Visualization

While the view camera is free to navigate between different viewpoints, we are firstly interested in the case where the view camera matches the position of a historical camera. The projective texturing approach will allow visualizing the image precisely on top of the 3D model(s). The view camera can be zoomed out to provide the user with additional context through the digital 3D environment of the photographed scene.

Suppose that the distortion parameters are known from the historical camera, then the **distortion model can be used during the projective procedure of the photo**. Our approach is not directed towards calculating the distortion model of the historical image, but its objective is to improve the user experience when this information is known. Therefore, we propose to visualize the scene through the historical view by modeling the view camera with the same distortion model as the historical camera. In this case, our **rendering system** needs to consider and use a **distortion model** that can principally **correct misalignments between photo and scene** (by undistorting the image) and **display** as a final result an **unaltered rectangular version of the photograph with a zoomed-out view** (by distorting the rendered view).

The existing methods for handling distortion may work correctly when the distortion function is defined in the image domain of the view camera. Since we are working with distortion functions that have been estimated by fitting the parameters to observations of point correspondences, the function's result is subject to overfit, oscillations, or may not be adequately defined outside of the domain of the photo. Consequently, in the zoomed-out view, this function may not behave as expected, *e.g.*, not be continuous. In these circumstances, we may need to **redefine** our **distortion model** and **extend** it, which in our case is by using an **extrapolation** approach. Chapter 4 shows and explains how the new distortion model is defined and in which conditions it is used.

3.3.2 Approach 2: Navigation Techniques to Reduce Visual Complexity

We now examine the case where multiple images can be discovered on the 3D scene. The view camera can loosely travel through the whole environment and interact with visual elements to showcase the images. We allow the user to **control the view camera, browse, and select which images and collections will be displayed**. However, for this to be achieved, we need to provide a concrete exploration engine.

Therefore, we define two main user expectations for our objective system as: *(i)* visualizing one or several images; *(ii)* exploring images in space and time. Based on these required user tasks, **we design and propose our solution in the form of design guidelines**. We define the six main properties considered in our designs as:

- Recognize areas of interest.
- Saliently see the images.
- Reduce the cluttering through the scene.
- Easily recognize the different viewpoint from the historical cameras.

- Manipulate the styles and colors of the scene.
- Include the temporal element.

We collectively combine techniques that can ease and enhance navigation through the images and answer the user requirements. We aim to provide with these techniques a path for a proper navigation and interaction support for many pictures and different image collections to cross-analyze this input data. Chapter 5 furthermore illustrates these design choices and our final propositions.

3.3.3 Approach 3: Co-Visualization of Historical Photos and 3D City Models

In order to **validate our methods**, **we implemented and combined the previously stated propositions into the HISTOVIS prototype**. This technical strategy is a web-based implementation orientated to check the competence of our solutions. Firstly, we designed a functional architecture that allows the visualization of a historical image. We extended it later to support a complete discovery of collections of historical photographs. We used this prototype for our final evaluation process, where we tested:

1. The user expectations and interactions through the definition of a primary use case scenario: visualizing historical images from two photo collections.
2. The ALEGORIA project partners' observations and feedback for an iterative design during this dissertation.

A detailed description of our implementation, evaluation, and results can be found in Chapter 6.

Chapter 4

Extrapolation of the Geometric Distortion for 3D Visualization

Preamble

The following chapter describes our first research issue: **the integration of heterogeneous spatial information causes potential visual discomfort.** We tackle it by defining (in the 3D environment) a virtual camera with a camera model different from the pinhole. We consider the geometric distortion of each historical image's camera during its acquisition process and use it to model this virtual camera. The content of this chapter is mainly based on our published work: Evelyn Paiz, Mathieu Brédif, and Sidonie Christophe (2020). 'Geometric Distortion of Historical Images for 3D Visualization'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-2-2020. DOI: 10.5194/isprs-annals-V-2-2020-649-2020.

Section 4.1 introduces the non-linear effects present on an image acquisition process and how this can affect the visualization of 2D photos inside a 3D environment. Section 4.2 describes two possible rendering methods that can be followed depending on the geometry of the 3D model. Next, Section 4.3 presents the definition of an extended distortion model that can be used to model the virtual camera inside the 3D scene. Section 4.4 presents our initial tests and results. Finally, Section 4.5 concludes with an overview of our proposed approach.

Contents

4.1 Introduction **64**
4.2 Rendering with Distortion **67**
4.2.1 One-Pass Rendering 67
4.2.2 Two-Pass Rendering 68
4.3 Extrapolating Distortion Models for 3D Visualization **69**
4.3.1 Existent Radial Distortion Models 70
4.3.2 Our Extrapolation Approach 72
4.4 Results: Testing Our Extrapolated Distortion Model **73**
4.4.1 Description of the Datasets 74
4.4.2 Test 1: Use of Different Camera Models for Projective Texturing 75
4.4.3 Test 2: Variation on the Extrapolation Radius 76
4.4.4 Test 3: One-Pass vs. Two-Pass Rendering 77
4.5 Conclusion **79**

4.1 Introduction

A **camera model** is defined as a **mathematical abstraction of a real-world camera** (Section 2.1 of Chapter 2). As represented by the diagram proposed by Förstner and Wrobel (2016) in Figure 4.1, there are many different ways of modeling a camera, and from one camera model to another, the number and type of intrinsic parameters known are the ones that differ. This diversity will determine how the 3D space is mapped to the 2D image. For instance, a *perspective camera* model has a planar sensor. An invariance of straight lines distinguishes it. On the other hand, the *ideal camera* is characterized by the focal length, which is the only parameter of the interior orientation. In contrast, the *normalized camera* has a focal length $f = 1$ and a rotation matrix $\mathsf{R} = \mathsf{I}_3$, where I is the identity. Its coordinate system is centered at the principal point and is parallel to the 3D scene coordinate system.

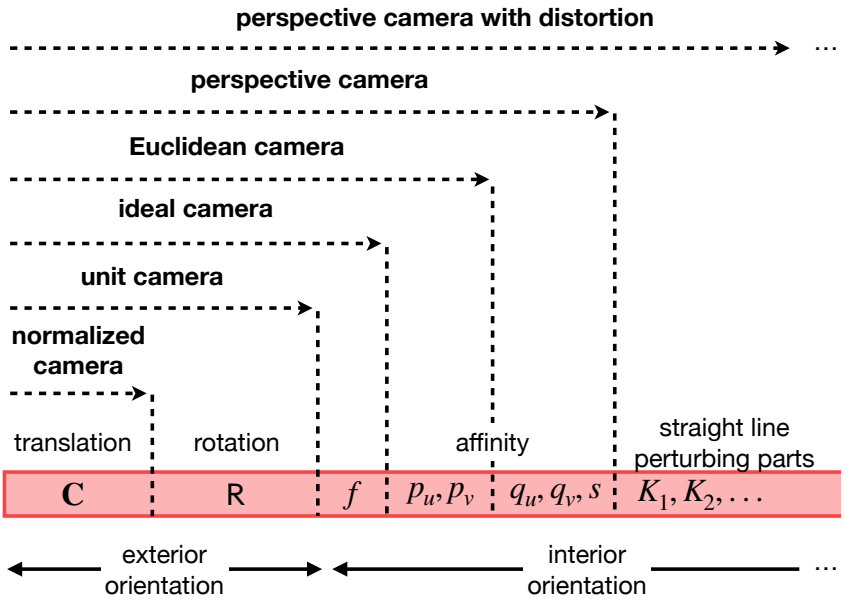


Figure 4.1: Overview of different camera models used to map 3D space points into a 2D image. Diagram is adapted from Förstner and Wrobel (2016).

Until now (in Section 2.1), we have assumed a "perfect" perspective camera model by only applying a projective transformation and exploiting the basic concept of projective geometry. However, **real-world cameras possess various effects (aberrations) that may alter the final image**. The causes for these are many and diverse, but among them we may mention (Kraus 2011; Paine and Kiser 2012):

Lens distortion occurs due to unequal bending of light. For example, the rays passing through the lens may tend to bend more near the edges than those close to the center. The effect is a curved deformity that becomes more severe near the edges of the photograph.

Nonplanarity of the photosensitive surface causes the lens and the image plane not to be parallel, *i.e.*, an imperfect production process of the camera. The result is a photo that seems to be tilted and stretched.

Atmospheric refraction of the light may cause the rays to curve. The index of refraction varies with the weather conditions. This effect is negligible for short distances but may reach a substantial error for considerably high flying heights.



Figure 4.2: Aerial photograph with a strong distortion and its rectified version after the distortion has been removed.

These **errors** are commonly known as **non-linear** because they have the property of **not conserving straight lines** (Förstner and Wrobel 2016). Figure 4.2 depicts this effect, where the original picture shows substantial deviations from a perspective mapping. Therefore, even though we may sufficiently simplify a real-world camera with a pinhole camera model, *i.e.*, perspective, it will not consider all the previously mentioned aberrations. To compensate for it, we will need to **add a last step to the 3D to 2D mapping** that will cover these non-linear effects. It will be a location-dependent shift in the image coordinate system ($D : \mathbb{R}^2 \rightarrow \mathbb{R}^2$), *i.e.*, an individual variation for each pixel described as follows:

$$\mathbf{x}' = \mathbf{x} + D(\mathbf{x} - \mathbf{c}) \tag{4.1}$$

where, as represented by Figure 4.3, $\mathbf{x}$ is the ideal image point, $\mathbf{c}$ is the distortion center, $\mathbf{x}'$ is the distorted image point, and D is the distortion function, *i.e.*, corrects the non-linear errors. Explicitly, we will establish the undistorted perspective point as $\mathbf{x}$ and the observed, distorted point as $\mathbf{x}'$.

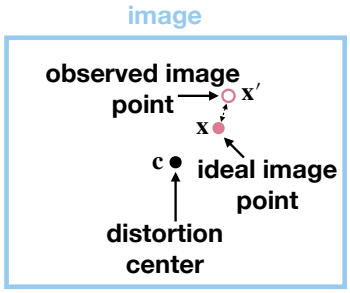


Figure 4.3: Depiction of an ideal image point and its distorted variant.

In reality, there are many reasons for these non-linear errors, and sometimes it can be hard to describe all of them. Therefore, we may just describe the phenomenon by assuming a model and fitting parameters to it.

We define this non-linear model as the **distortion model**, while we call the camera model that is employing it a **camera model with distortion**.

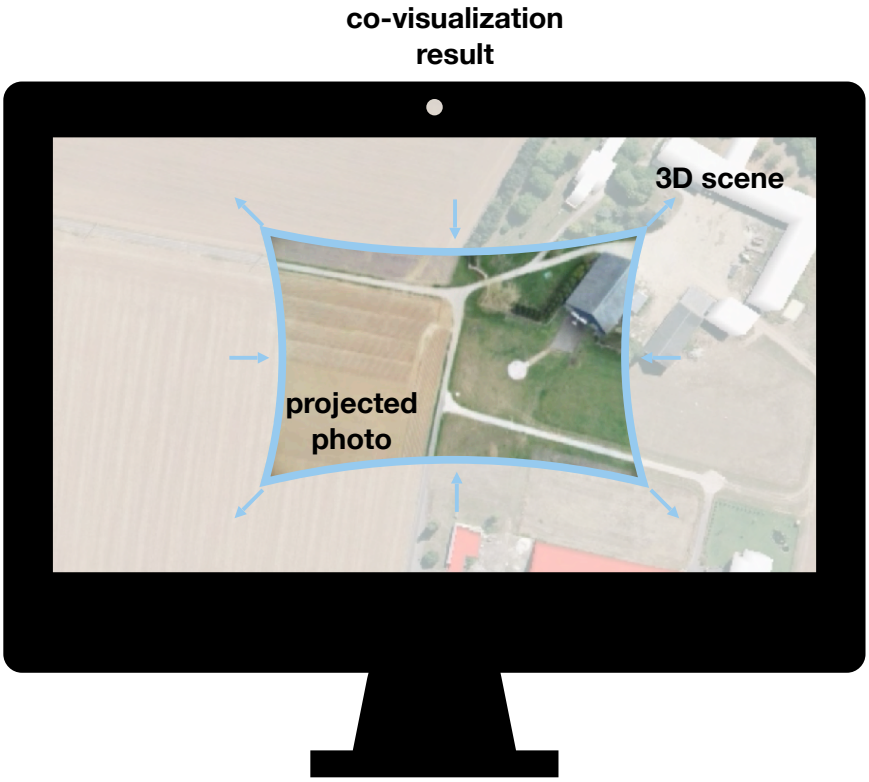


Figure 4.4: 3D city model and projected 2D image. The view camera uses a pinhole camera model (without distortion) and the historical camera a camera model with distortion.

On this note, as Figure 4.4 displays, we can now **examine singularly the topic of geometric distortion regarding the co-visualization of photographs within a 3D model of the photographed scene**, *e.g.*, using a point cloud. Let us consider first the presentation of a single 2D photo in the context of the 3D scene. We model two types of cameras for this co-visualization (as explained in Section 3.3 of Chapter 3):

- A **view camera** is responsible for synthesizing/rendering the image to be presented to the user. We model it through a pinhole camera model.
- A **historical camera** uses an estimated model of the camera used to acquire the photograph. Since it depicts a real-world camera, generally (when the information is known), we model it through a camera model with distortion.

Accordingly, assuming that the historical camera orientation is known, exterior and interior (with a distortion model), we may project the picture on this virtual 3D world using the 3D to 2D mapping from the camera model. A pixel-accurate reprojection will be visualized when the view camera is placed at the same position and orientation as the historical camera. A wider zoom will allow us to visualize the 3D model around the photograph of interest, which may be shown as extra context. Within this case, as Figure 4.4 shows, a visualization of a 3D scene textured by an unprojected historical image with distortion, rendered through a pinhole camera virtually placed where the historical photograph has been taken, will effectively undistort the image. Thus, the **undistorted representation of a rectangular historical image will be non-rectangular due to the rectification process**, which may be counter-intuitive or distracting for the end-user.

Objective and Proposed Approach

Our purpose is to decrease the possible visual distress user may have when visualizing a photograph inside a 3D environment. Therefore, **we propose using the distortion model defined in the historical camera to extend the camera model of the view camera**. Instead of observing only an undistorted photograph, the result will be a distorted view that will counteract the rectification applied to the image. It will allow users to visualize the photograph's original unaltered rectangular version instead of the rectified variant.

4.2 Rendering with Distortion

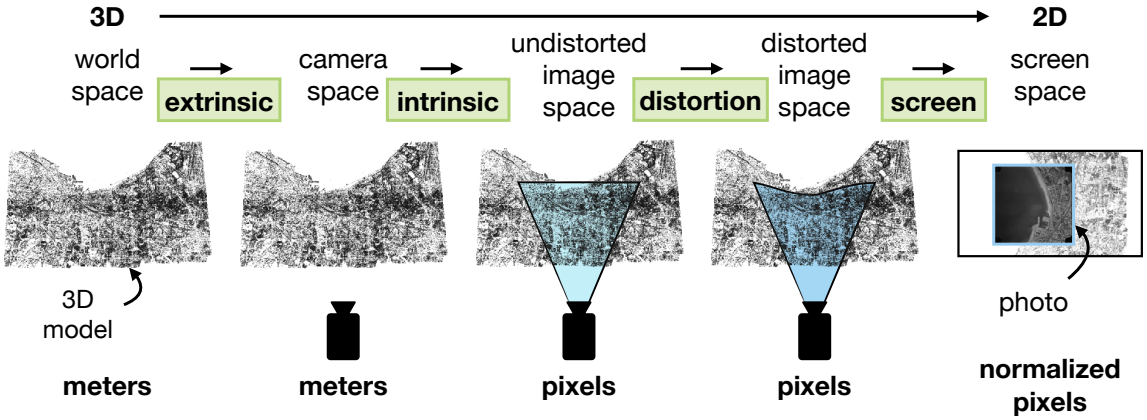


Figure 4.5: Rendering with distortion. Our desired mapping from 3D to 2D for the view camera. It follows a camera model with distortion.

Rendering allows us to generate an image digitally from a virtual scene composed of 3D models. It is essentially done by putting a virtual camera in the digital environment and reproducing the same 3D to 2D mapping that the camera has. Our objective mapping for this camera (the view camera), *i.e.*, one that considers the non-linear errors, is exposed in Figure 4.5.

Accordingly, if we consider using a distortion model in the view camera, an implementation pipeline like Lorenz and Döllner (2008, 2009) proposed can apply a piecewise projective approximation of any distortion function. Nonetheless, its usage is limited because of the requirement of geometry shaders, *i.e.*, not found in a web graphics application. **We explore the other existing approaches to be able to use a distortion model applied to the view camera in an online context**. We define two pipelines (which will be tested later with our proposed distortion model), as is sketched in Figure 4.6 and Figure 4.7, depending on each category of a 3D model used inside the 3D environment: *(i)* point cloud; *(ii)* mesh.

4.2.1 One-Pass Rendering

A **point cloud** is a collection of data points in a 3D space sampled from an object(s) (Krig 2016), *e.g.*, the photographed scene of an image we want to visualize. To model the projection of the points

of this type of scene towards the resulting view (visualized through the view camera), the most straightforward approach can be a **one-pass rendering method** as the one used by Keahey and Robertson (1996) and Spindler et al. (2006). We divide it as follows:

- **Distortion on the Historical Camera:** the distortion model from the historical camera is used when sampling the image in the fragment shader.
- **Distortion on the View Camera:** the distortion model for the view camera can be applied to each 3D vertex of the point cloud when computing their projections on the screen space in the vertex shader.

The result will be the rectification (undistortion) of the image in the fragment shader, which is counteracted by the distortion of the final view in the vertex shader, *i.e.*, resulting in the visualization of the original rectangular image. This pipeline is sketched in Figure 4.6.

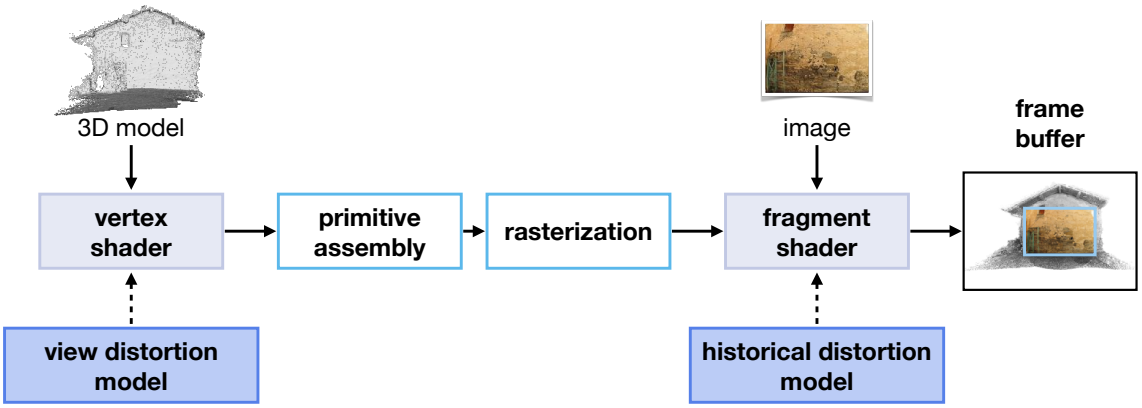


Figure 4.6: One-pass rendering, where the distortion is applied to the vertices of the scene and the reprojection of the image.

4.2.2 Two-Pass Rendering

A **mesh** is a **structural model containing polygons** (Krig 2016), commonly triangles. The one-pass rendering approach for these models is unsuitable since vertices (of the polygons) would get distorted but not their edges. Hence for these 3D models, we apply a two-pass rendering technique using the approach acknowledged in computer graphics as render-to-texture (Szirmay-Kalos, Szécsi, and Sbert 2008). It is defined as:

A First Pass rectifies the photo and removes its distortion.

- **Distortion on the Historical Camera:** the distortion model from the historical camera is used in the reprojection of the image, as explained in the first pass.
- **View Camera:** the scene is rendered with a pinhole camera model; therefore, no distortion model is used in the view camera during this first pass.

A Second Pass distorts the view.

- **Distortion on the View Camera:** the full view of the scene is saved on the first pass as an input for the second pass, where we will inversely distort it. We use the distortion model defined for the view camera to achieve it.

Following this approach, the historical photograph is undistorted in the first pass, and the image of the scene is distorted in the second pass. The pipeline is illustrated in Figure 4.7.

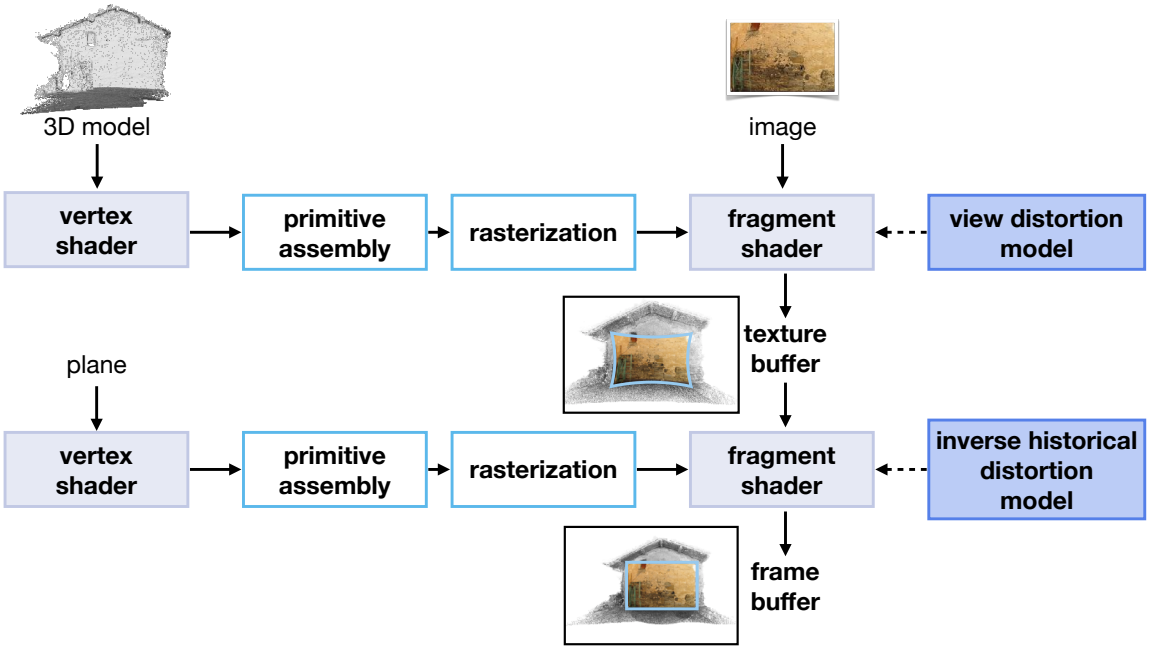


Figure 4.7: Two-pass rendering, where in the first pass, the image is undistorted. In a second pass, the inverse distortion is employed to distort the whole view.

4.3 Extrapolating Distortion Models for 3D Visualization

For a given pixel point **x** on a picture, its distortion can be seen as the **geometric deviation between its actual and ideal coordinates** (see Equation 4.1). Many distortion models have been developed to encode this divergence of the pixel projections from the ideal pinhole camera model, *e.g.*, radial, tangential, and fish-eye, depending on the non-linear errors present when the photograph was acquired.

Regarding our context, we desire to utilize the distortion model defined in the historical camera for the view camera. However, **the view camera will need to be zoomed out to allow the user to observe the 3D scene around the current visualized photograph**. Thus, the distortion model used in the view camera will need to **extend the definition of the distortion model of the historical camera**, *i.e.*, it needs to be defined outside of the image, where the 3D scene is located. Indeed, distortion models are estimated by fitting their parameters to the observations of point correspondences in the image domain only. Therefore, a mathematical distortion model may be subject to overfit, oscillations, or even not be defined adequately outside of the image domain.

This section **describes our proposition for extrapolating distortion models**. We aim for our mathematical model to use an increasing function, which is continuous in all the image domain. Therefore, as showcased by Figure 4.8, we extend a distortion model by limiting the domain of definition of the original function. Then, we determine a new function outside of this domain by estimating each new value based on the original function. We define this process as an **extrapolation of the distortion function**.

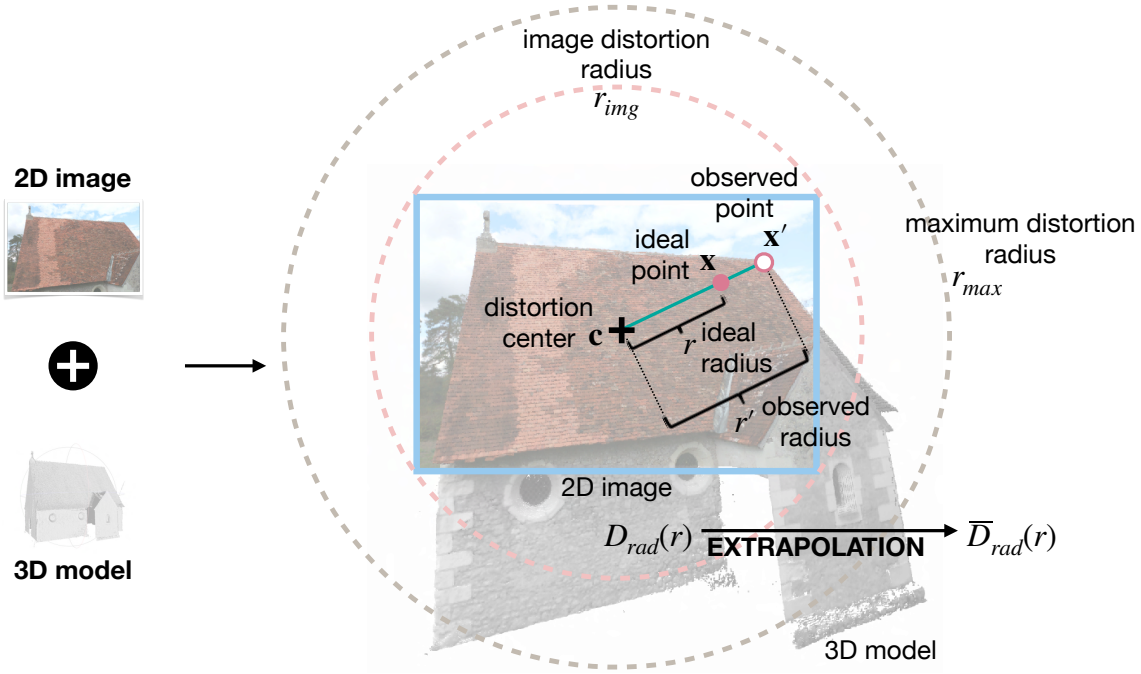


Figure 4.8: Extrapolating distortion models for 3D visualization.

4.3.1 Existent Radial Distortion Models

The simplest and most commonly used distortion model in moderate distortions is arguably the **radial distortion model**, which will be the one that we will consider in this work. Geometrically, it is an **inward or outward displacement of the pixel point x**. This movement is only performed along a radial direction from the distortion center **c**. It is represented as:

$$\mathbf{x}'_{\mathbf{rad}} = \mathbf{x} + D_{rad}(r)(\mathbf{x} - \mathbf{c}) \tag{4.2}$$

where r is the radial distance in pixels from $\mathbf{x}$ to the distortion center $\mathbf{c}$, denoted as $r^2 = |\mathbf{x} - \mathbf{c}|^2 = (x - c_x)^2 + (y - c_y)^2$ and D_{rad} is the radial distortion function specified as $D_{rad} : I \rightarrow J$, with $I = [0, r_{max}]$ and $I, J \subset \mathbb{R}^+$. We designate r_{max} as the smallest positive root of the derivative $D'_{rad}(r)$.

From this radial distortion model definition, we now describe two widely used radial distortion models: **radial polynomial** and **radial-fish eye** (with more significant non-linear errors). Since our proposed approach only supports radial distortions, tangential distortion terms are neglected in the definitions of the distortion functions.

Radial Polynomial Distortion Function

A polynomial series can express the radial distortion we have defined in Equation 4.2, as follows (Fraser 1997):

$$D_{polyrad}(r) = P(r^2) = K_1 r^2 + K_2 r^4 + K r^6 + \ldots \tag{4.3}$$

where $P(X) = K_1 X + K_2 X^2 + K_3 X^3 + \ldots$ is a polynomial utilizing the radial coefficients K_i. Then the relationship between an ideal (undistorted) image point $\mathbf{x}$ and its observed (distorted) image point $\mathbf{x}'$ is:

$$\mathbf{x}'_{\mathbf{polyrad}} = \mathbf{x} + P(r^2)(\mathbf{x} - \mathbf{c}) \tag{4.4}$$

Radial Fish-Eye Distortion Function

With a fish-eye, there is an opening of almost 180°. Generally, many lenses are employed to obtain such a large field of view, resulting in the light rays not intersecting in a single point but meeting in a small area. Hence, fish-eye lenses do not effectuate a central projection, *i.e.*, all rays do not pass through the same point. However, if we assume that the distance to the scene is not too small, we can model it with a central projection with sufficient accuracy (Förstner and Wrobel 2016). Here we adopt the polynomial fish-eye model in *MicMac* (Rupnik, Daakir, and Pierrot Deseilligny 2017) defining $\mathbf{x}'$ as:

$$\mathbf{x}'_{\mathbf{fish}} = \mathbf{c} + f\Bigg(\underbrace{(1 + K_1\rho^2 + K_2\rho^4 + K_3\rho^6 + \ldots)\begin{pmatrix} a \\ b \end{pmatrix}}_{\text{radial}} + \underbrace{\begin{pmatrix} P_1(\rho^2 + 2a^2) + 2P_2ab \\ 2P_1ab + P_2(\rho^2 + 2b^2) \end{pmatrix} + \begin{pmatrix} l_1a + l_2b \\ l_2a \end{pmatrix}}_{\text{tangential}}\Bigg) \tag{4.5}$$

where f is the focal length of the camera, $\begin{pmatrix} a \\ b \end{pmatrix} = \frac{\rho}{r}(\mathbf{x} - \mathbf{c})$, and $\rho = \arctan\left(\frac{r}{f}\right)$. K_i are the radial coefficients, P_i the tangential coefficients, and l_i polynomial coefficients. We are interested only in the radial part of Equation 4.5; consequently, the radial fish-eye distortion function $D_{fishrad}$ can be represented as:

$$D_{fishrad}(r) = \frac{\rho f}{r}\Big(1 + P(\rho^2)\Big) - 1 \tag{4.6}$$

Then for this radial fish-eye function $D_{fishrad}$ the relationship between an ideal (undistorted) image point $\mathbf{x}$ and its observed (distorted) image point $\mathbf{x}'$ is:

$$\mathbf{x}'_{\mathbf{fishrad}} = \mathbf{x} + \left(\frac{\rho f}{r}\Big(1 + P(\rho^2)\Big) - 1\right)(\mathbf{x} - \mathbf{c}) \tag{4.7}$$

Inverse Radial Distortion Function

When the objective is to **determine the radial distance r of an undistorted point x**, the goal is to find an **inverse transformation** for D_{rad} (Equation 4.3 and Equation 4.6). For values $r \in [0, r_{max}]$, an analytical solution is difficult to be calculated. From an optimization point of view, an iterative

approach can be introduced. Using Newton–Raphson's method, with a quadratic convergence rate (Benligiray and Topal 2015), an approximation of r can be obtained through the iteration:

if $r \in [0, r_{max}]$,

$$r^{(n+1)} = r^{(n)} + \frac{r' - D_{rad}(r^{(n)})}{D'_{rad}(r^{(n)})} \tag{4.8}$$

where $r' = D_{rad}(r)$.

We define $r^{(0)} = r' \cdot \frac{r_{max}}{D_{rad}(r_{max})}$ as the initial guess, guaranteeing that $r^{(0)}$ starts in the definition domain $[0, r_{max}]$ of $D_{rad}(r)$. Iteratively, this can be refined until the solution of r converges. Note that for values $r > r_{max}$, this method cannot guarantee convergence since the function may not be properly defined, *e.g.*, not be continuous or non-decreasing.

4.3.2 Our Extrapolation Approach

Both radial functions D_{rad} in Equation 4.3 and Equation 4.6 are defined for all values of r. Nevertheless, they have been estimated using observations only on the image domain $[0, r_{img}]$, where r_{img} represents the complete image's radius. Hence, values on $r > r_{img}$ are likely subject to overfitting and oscillations. **We are interested only in a bijective distortion transformation, where $\mathbf{D_{rad}}$ is a non-decreasing continuous function**. We want the distortion function to be invertible, *i.e.*, $D^{-1}_{rad}(r')$ to be defined. To achieve this, we limit the domain of D_{rad} to a range of $[0, r_{ext}]$, with $r_{ext} < r_{max}$.

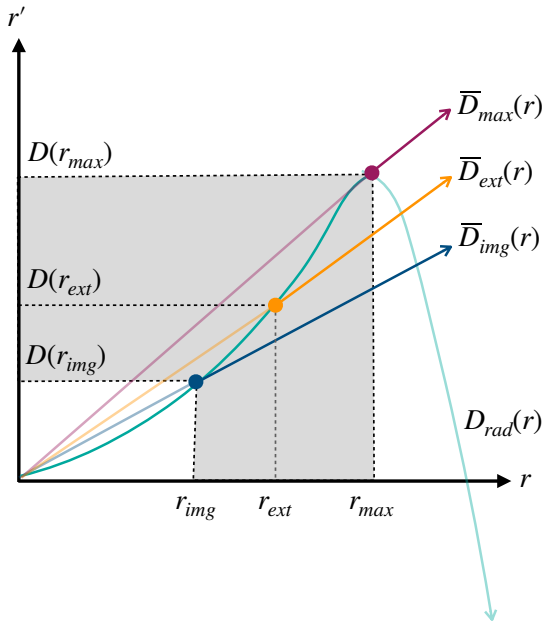


Figure 4.9: Sample behavior of the distortion model.

Consequently, for values $r > r_{ext}$, we **introduce a linear extrapolation definition**. We denote the starting point of extrapolation as $r_{ext} \in [0, r_{max}]$, where a complete distortion correction can be achieved when $r_{ext} \geq r_{img}$. Figure 4.9 presents an example of the behavior of the distortion function D_{rad}. The functions $\overline{D}_{img}$, $\overline{D}_{ext}$, and $\overline{D}_{max}$ are three possible extrapolation candidates for D_{rad}, depending on the value where the extrapolation starts (respectively r_{img}, r_{ext} and r_{max}). Accordingly, we choose $\overline{D}_{ext}(r)$ and extend the definition of the distortion function as $\overline{D}_{rad}$, where:

$$\overline{D}_{rad}(r) = \begin{cases} D_{rad}(r) & \text{if } r \in [0, r_{ext}] \\ D_{rad}(r_{ext}) & \text{if } r > r_{ext} \end{cases} \tag{4.9}$$

Therefore, if $r > r_{ext}$:

$$\mathbf{x}'_{\mathbf{rad}} = \mathbf{x} + D_{rad}(r_{ext})(\mathbf{x} - \mathbf{c}) \tag{4.10}$$

$$\begin{aligned} \mathbf{x}'_{\mathbf{polyrad}} &= \mathbf{x} + D_{polyrad}(r_{ext})(\mathbf{x} - \mathbf{c}) \\ &= \mathbf{x} + P(r_{ext}^2)(\mathbf{x} - \mathbf{c}) \end{aligned} \tag{4.11}$$

$$\begin{aligned} \mathbf{x}'_{\mathbf{fishrad}} &= \mathbf{x} + D_{fishrad}(r_{ext})(\mathbf{x} - \mathbf{c}) \\ &= \mathbf{x} + \left(\frac{\rho_{ext} f}{r_{ext}} \left(1 + P(\rho_{ext}^2) \right) - 1 \right)(\mathbf{x} - \mathbf{c}) \end{aligned} \tag{4.12}$$

with $\rho_{ext} = \arctan\left(\frac{r_{ext}}{f}\right)$.

The definition of Equation 4.9 produces an affine transform in the extrapolated domain $r > r_{ext}$. The camera behaves as a pinhole camera with modified intrinsic parameters. The focal length f and skew factor s are multiplied by $(1 + D_{rad}(r_{ext}))$ and the principal point $\mathbf{p}$ is translated by $D_{rad}(r_{ext})(\mathbf{x} - \mathbf{c})$.

4.4 Results: Testing Our Extrapolated Distortion Model

To evaluate our proposed extrapolation of the radial distortion functions, we **implemented a web application**[1] using the *three.js*[2] WebGL rendering library. In this section, we present our results from **testing** the following **three main aspects**: the visualization through a historical view, the variation of the extrapolation radius, and the comparison of one-pass rendering and two-pass rendering.

[1] https://github.com/epaizreyes/photogrammetric-camera
[2] https://threejs.org

Evaluations

- **Test 1:** visualize the difference (for both historical and view cameras) between using two camera models: a simple pinhole and a pinhole considering the distortion model.
- **Test 2:** evaluate the effect of varying the extrapolation radius on the resulting view. Additionally, we will consider it too for the transitions (view interpolation) between images.
- **Test 3:** compare (using our distortion model) a one-pass and two-pass rendering for point clouds and meshes.

As we mentioned in Section 3.3 of Chapter 3, we require as input: *(i)* a 2D **photograph**; *(i)* the camera's photo **orientation data** (intrinsics, extrinsics, and distortion model); *(iii)* the **3D model** of the photographed scene. It is essential to note that the accuracy of the 3D scene is not relevant to our method. When a single photograph is projected onto the scene, and the view camera is placed at the historical camera position, the scene information is unnecessary. Any geometry that fills the viewport can be used equally.

4.4.1 Description of the Datasets

Table 4.1 shows a summary of our selected experimental datasets. We choose two modeled with radial polynomial distortion, while two others with a radial fish-eye distortion. We followed a photogrammetric process using *MicMac*[3] to spatialize each photograph and generate the respective 3D model (a point cloud). We designate each photo collection along with its characteristics as:

1. **Maurepas**[4]**:** sample dataset produced by IGN/ENSG in the Domaine National de Chambord. It is a modern acquisition of terrestrial close-range images. It presents a notable distortion, but it can be modeled through a radial polynomial function.
2. **Frejus**[5]**:** photographs from a past IGN aerial acquisition in 1966 of the port town Fréjus. The photos are vertical views and have been obtained from *Remonter le Temp*. Since they have been acquired under controlled conditions, it presents almost no distortion. We modeled it through a radial polynomial function.
3. **Saint-Martin**[6]**:** a terrestrial photographic dataset offered by *MicMac*, depicting the street of Saint-Martin. It is a contemporary acquisition. The photos were taken with two different kinds of fish-eye lenses. The result is a set of images with significant distortion and another set with a much larger one. We modeled all the photographs of this dataset through two different radial fish-eye functions.
4. **Viabon:** a subset of images from a recent photogrammetric collection of vertical aerial photographs taken at Aerodrome Mellay Viabon. Although the photographs were not taken with a fish-eye lens, they present significant distortions; therefore, we modeled these images with a radial fish-eye function.

[3] https://github.com/micmacIGN
[4] http://dias.ensg.eu/Chambord/
[5] https://remonterletemps.ign.fr/telecharger/
[6] http://micmac.ensg.eu/data/

Table 4.1: Overall description of datasets used for testing our approaches. Credits to IGN/ENSG.

Name	Point of View	Date	Distortion Model	No. Images	Example
Maurepas	terrestrial	2015	radial polynomial	30	
Frejus	aerial	1966	radial polynomial	40	
Saint-Martin	terrestrial	2012	radial fish-eye	29	
Viabon	aerial	2020	radial fish-eye	30	

4.4.2 Test 1: Use of Different Camera Models for Projective Texturing

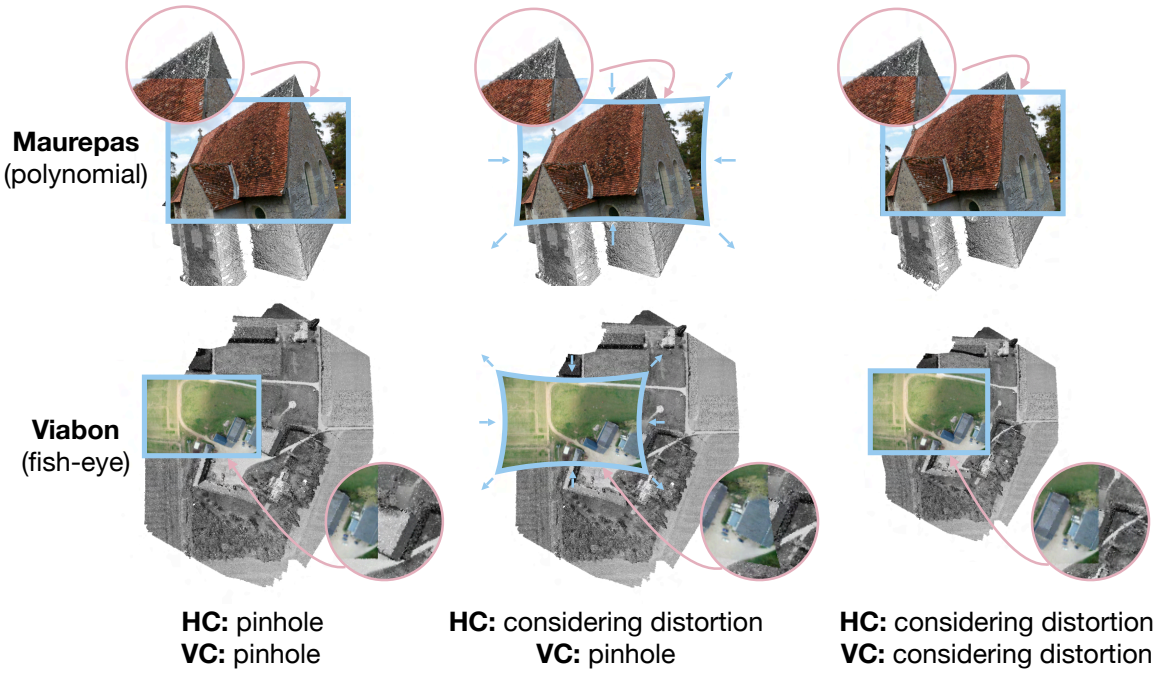


Figure 4.10: Use of different camera models for projective texturing. Consider HC as the historical camera and VC as the view camera. For the camera models, the pinhole is modeled without distortion, and considering distortion is modeled with either a polynomial or fish-eye radial distortion.

We examined the difference between using two types of camera models for both historical and view cameras: *(i)* pinhole (or perspective); *(ii)* perspective with distortion. A sample of our results from the Maurepas and Viabon datasets is shown in Figure 4.10. Our findings are:

1. The **historical and view cameras are modeled through a pinhole camera model** when we do not consider the distortion model in either camera (historical distortion is neglected).

Nonetheless, misalignments between the picture and the scene are noticeable, especially when the non-linear errors are high, like the fish-eye lenses.

2. The previous misalignments can be corrected if the **distortion model is considered for the historical camera**. Instead of modeling the historical camera only through a linear transformation, the introduction of the distortion can fix the non-linear errors. However, the result is a projected photo that is undistorted and deformed.
3. When using our proposed method, **the distortion model is applied to both historical and view cameras**. Hence the 3D world is viewed through a distorted view. This technique presents the unaltered rectangular version of the photograph without possible image misalignments or deformations. Notice that the distortion on the view will be much more noticeable for high non-linear effects, as is the case for fish-eye lenses.

4.4.3 Test 2: Variation on the Extrapolation Radius

As mentioned in Section 4.3, the starting extrapolation radius r_{ext} can vary in a range $[0, r_{max}]$. Figure 4.11 shows for all our datasets how this variation can affect the resulting visualization. When $r_{ext} = 0$, a perspective view is achieved with no distortion; therefore, the view camera is rendered using a pinhole camera model. A complete correction can be seen only when $r_{ext} \geq r_{img}$. Based on our experience with the different datasets tested, *i.e.*, with radial polynomial and radial fish-eye distortions, **the case where $\mathbf{r_{ext} = r_{img}}$ is the best configuration for visualizing any image**. It allows distortion correction only where observations were available to estimate the distortion function.

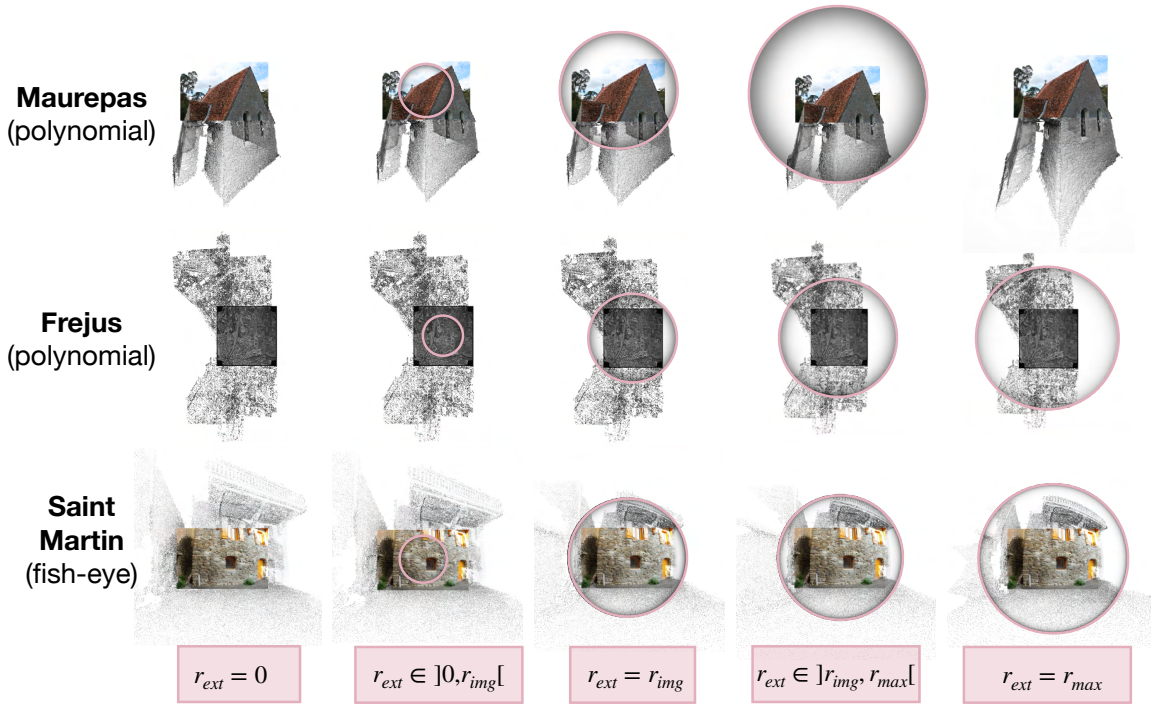


Figure 4.11: Variation on the extrapolation radius. The radius r_{ext} variates in the range of $[0, r_{max}]$. The pink circle with the black shading has radius r_{ext} and is only for illustrative purposes.

On another note, we use view interpolation to transition from the visualization of one image img^A to another image img^B. For the extrapolation radius r_{ext}, we interpolate its value during this transition. Therefore, if we assume that we are currently visualizing img^A and $r_{ext} = r_{img^A}$, then our interpolation is from r_{img^A} to r_{img^B}. However, we noticed that this abrupt change may cause a dizzy experience for any user. As depicted by Figure 4.12, we instead propose to reduce first the value of r_{ext} to 0 and achieve a perspective view. The movement from one photograph to the other can be accomplished utilizing this value. When the view camera arrives at the desired position, the value of r_{ext} can continuously increase from 0 to r_{img^B}. Hence we define the interpolation of r_{ext} in a time $t \in [0, 1]$ as:

$$r_{ext} = \begin{cases} r_{img^A} & \text{if } t = 0 \\ r_{img^A}\left(1 - \frac{t}{t_A}\right) & \text{if } t \in]0, t_A[\\ 0 & \text{if } t \in [t_A, t_B] \\ r_{img^B}\left(\frac{t - t_B}{1 - t_B}\right) & \text{if } t \in]t_B, 1[\\ r_{img^B} & \text{if } t = 1 \end{cases} \tag{4.13}$$

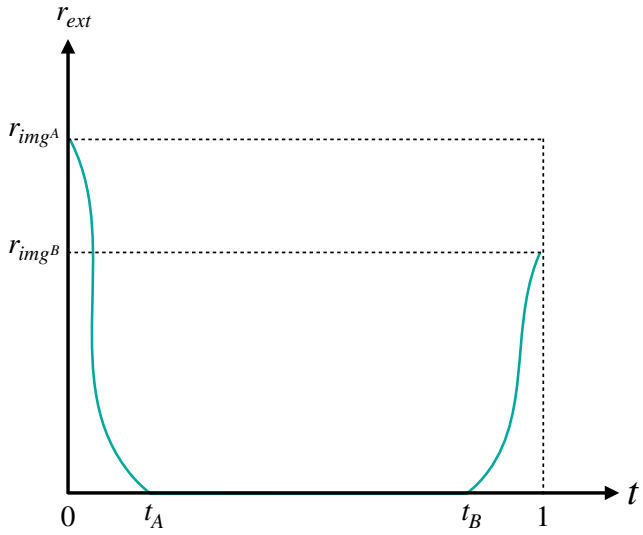


Figure 4.12: Our proposition for the variation on the extrapolation radius for view interpolation. In a time t, r_{ext} varies from r_{img^A} to r_{img^B}.

4.4.4 Test 3: One-Pass vs. Two-Pass Rendering

Even though the one-pass rendering technique presented in Section 4.2 is a more straightforward approach, it cannot be used in every 3D model. For instance, the method may fail and produce undesired deformations on 3D meshes that are not well tesselated, *i.e.*, the number of polygons that define it is small. We show a simple example in Figure 4.13, where a wireframe tessellated plane is placed in front of the viewed scene to showcase how the two rendering methods work. On the image on the left, it can be seen that even though the plane's vertices are distorted, the edges remain straight. This result is not the case for a two-pass (on the right side image), where the edges of the triangles are curved. Therefore, **for our approach we established the two-pass as the best option for meshes (badly tesselated) while the one-pass for point clouds**. Notice that the two-pass can

be used for point clouds, but a one-pass is a simpler solution, *i.e.*, has a faster computation since there is no need for inverse distortion.

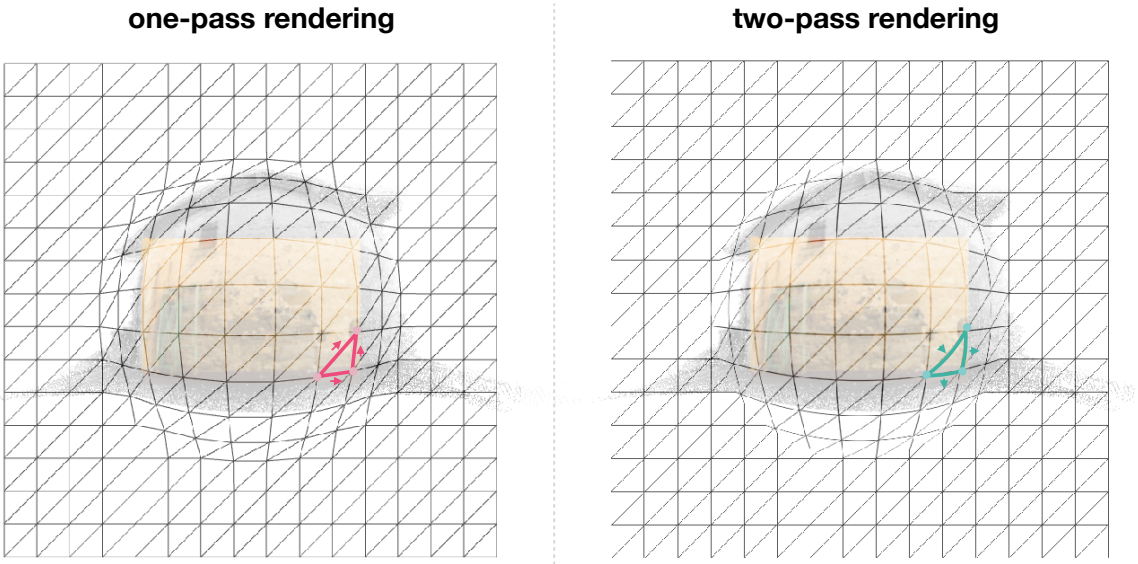


Figure 4.13: One-pass vs. two-pass Rendering. Comparison between the two rendering pipelines. A wireframe tessellated plane has been put in front of the view for illustrative purposes.

We present another comparison of the two rendering methods (one-pass and two-pass) in Figure 4.14. In the first illustration on the left, we use a point cloud, which only contains points (no edges), and consequently a one-pass rendering approach is sufficient for our distortion model. In the second sample on the right, we use the geospatial data visualization of *itowns*[7] and place the image in a 3D scene composed of meshes representing the ground and the buildings. We apply to this 3D environment a two-pass rendering strategy. As seen, when using the suitable method for the 3D models in the scene, a correct distortion can be seen and visualize in the resulting view.



Figure 4.14: Visualization of a photograph in two different 3D scene models. A one-pass rendering is used for point clouds, while a two-pass is used for meshes.

[7] http://www.itowns-project.org

4.5 Conclusion

Subsection 4.4.2 (test 1: use of different camera models for projective texturing), allowed us to visually comprehend how the use of our proposed distortion model in both view and historical cameras can preserve the rectangular shape on the projected images and avoid possible misalignments between them and the 3D models. We noticed that this was much more noticeable as the level of non-linear errors in the photographs was higher. Accordingly, let us remark that our method is only intended for the visualization of specific photographs, which show non-linear effects that cannot be modeled with a pinhole camera model. For example, the Frejus dataset was modeled with radial distortion. However, the distortion was so small that there was no highly remarkable difference between the distorted and undistorted versions of the photo. In this case, using the pinhole camera model for both view and historical cameras should be sufficient. Nevertheless, this was not the case for the Maurepas, Viabon, and Saint-Martin datasets.

Subsection 4.4.3 (test 2: variation on the extrapolation radius) enabled us to properly define an initial state for a 3D visualization environment where the users will not need to interact with the distortion parameters and only focus on the visualization of the photographs. Note that our approach only exhibits an extrapolation approach that supports a polynomial form depending on one variable (r), which is the case for the radial polynomial and radial fish-eye models. The main reason is that we search for an affine transformation in the extrapolated area, which the introduction of other distortion coefficients, like the tangential ones, do not allow.

In Subsection 4.4.4 (test 3: one-pass vs. two-pass rendering), we compared and tested our four proposed datasets and found that both one-pass and two-pass rendering applied our distortion model to the vertex and pixel coordinates as we expected. We only encountered a problem in the Saint-Martin dataset, containing a couple of photographs with substantial non-linear errors. Our two-pass rendering failed in converging and finding the correct corresponding r value in the areas close to the borders, where the distortion was more pronounced. For these types of photos, we recommend not using Newton-Raphson's method. Instead, we propose to apply only a mapping between distorted and undistorted image points, as proposed by Vass and Perlaki (2003). Another possibility could be to adopt a homography approach and model the change from distorted to undistorted positions.

Contribution

In summary, **the main contribution of in this chapter is the definition and the real-time rendering of an extrapolated distortion model given any radial distortion function**. It enables the camera in the 3D environment to use the same distortion model as the camera sensor that acquired the historical photograph, *i.e.* rendering through a historical view. Therefore, this method provides an efficient approach for the 3D visualization of historical images that present geometric distortions on scenes composed of triangular meshes and point clouds.

Chapter 5

Navigation Techniques to Reduce Visual Complexity

Preamble

This chapter details our second research issue: **the navigation through a massive amount of visual and spatio-temporal data represented on a large scale.** We address the problem by proposing a set of combined procedures and techniques to reduce the visual complexity obtained when visualizing various images inside a 3D environment. The content of this chapter is mainly based on our published work: Evelyn Paiz-Reyes, Mathieu Brédif, and Sidonie Christophe (2021). 'Cluttering Reduction for Interactive Navigation and Visualization of Historical Images'. In: *Proceedings of the ICA* 4. DOI: 10.5194/ica-proc-4-81-2021.

Section 5.1 introduces the topic of visual complexity and how it becomes an issue when navigating inside a 3D environment that handles a large volume of visual data. Section 5.2 defines the expectations of the users through two primary user tasks. Subsequent, from all existing methods to ease navigation, Section 5.3 exhibits our selected techniques explained by our design guidelines. Section 5.4 presents our first results from an implementation of each design, and Section 5.5 concludes with a summary of our proposed strategy.

Contents

5.1 Introduction **81**
5.2 User Tasks **83**
5.2.1 Task 1: Visualizing One or Several Images 83
5.2.2 Task 2: Exploring Images in Space and Time 84
5.3 Design Guidelines **85**
5.3.1 Design Guideline 1: How to Represent a Photograph? 85
5.3.2 Design Guideline 2: How to Make Salient Certain Elements? 88
5.3.3 Design Guideline 3: How to Add the Temporal Aspect? 91
5.4 Results: Testing Our Design Guidelines **92**
5.4.1 Test 1: Bookmarks and Thumbnails for Interaction with Photographs 93
5.4.2 Test 2: Image Frames, Visibility Heat Map, and Style Representation for Saliency 95
5.4.3 Test 3: Timeline for Temporal Filtering 98
5.5 Conclusion **99**

5.1 Introduction

Human vision and domain expertise are power mechanisms that, simultaneously with computational devices, make it possible to turn large volumes of heterogeneous data into interpreted information and, subsequently, into knowledge derived from integrating this information. As MacEachren and Kraak (2001) mentions, it is estimated that 80% of the digital data generated has a geospatial referencing, *e.g.*, coordinates, addresses, and postal codes. Therefore, now more than ever a support for this type of information is necessary. Geographic visualization allows users to **understand and interact with complex phenomenons through innovative scene reconstructions in virtual environments**, reaching for a "Digital Earth," as Gore (1998) exposes.

Nevertheless, current visualization platforms are not all adapted towards a geovis approach. New interface paradigms are needed to **support user interaction with the forms of representation required for geospatial data.** Since geovisualization was born from the concept of cartography, we can examine cartographic interactions. For instance, Roth (2012) adjusted 1988 Norman's stages of (inter)action model to consider an interpretation of cartographic interactions as a two-way dialogue between human and map (or a "Digital Earth" in our case), mediated by a digital computing platform, where the human and map are agents in the interaction (affecting change to the other). Consequently, as represented in Figure 5.1, we may say that three components are necessary for a digital interaction: *(i)* the human (user-centered perspective); *(ii)* the map (interface-centered perspective); *(iii)* the computing device (technology-centered perspective).

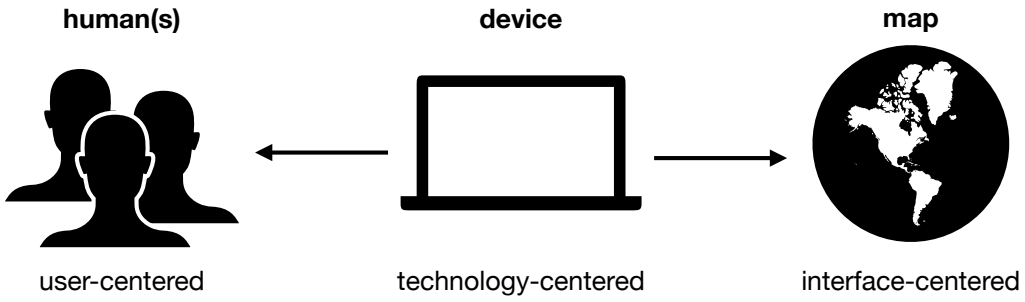


Figure 5.1: Cartographic interactions. It is represented as a dialogue between a human (top) and a map (bottom) through a computing device (center). Figure adapted from Roth (2012).

As depicted in the diagram by Çöltekin, Griffin, et al. (2020) in Figure 5.2, the elected visualization platform represents a balance between the desired visual realism and the level of immersiveness. For instance, in this work, we have selected a 3D photo-realism for an authentic and immersive representation inside an online environment. This chapter concentrates on the **user's visual exploration, particularly what can be perceived when many images are visualized inside a 3D topographic scene**. As Section 2.4 of Chapter 2 explains, the perception process is a sequence of steps determining the experience and reaction to any stimuli. 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 drive attention to some parts of the 3D scene or some objects/structures/characteristics in the historical images.

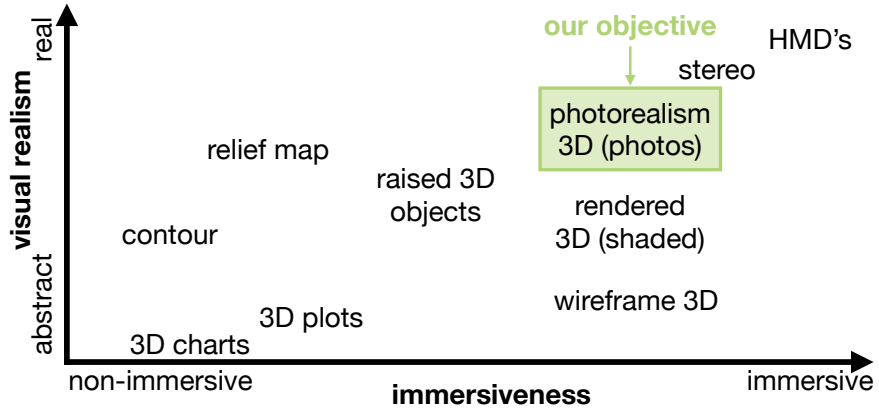


Figure 5.2: The immersiveness spectrum also analyzing the visualization design. Illustration adapted from Çöltekin, Griffin, et al. (2020).

On this note, **we desire to improve the dialogue between users and our geovisualization environment oriented towards exploring historical images within a 3D environment composed of3D city models**. Until this point, our focus has been on the artifact correction on the system when a single photo is visualized. Aside from using the keyboard to move from one picture to another, no interaction option has been presented to the user. Let us now consider discovering many photographs inside the 3D scene and the graphical objects and representation necessary for their interaction. We want to provide proper navigation support for many pictures and different image collections to cross-analyze this input data.

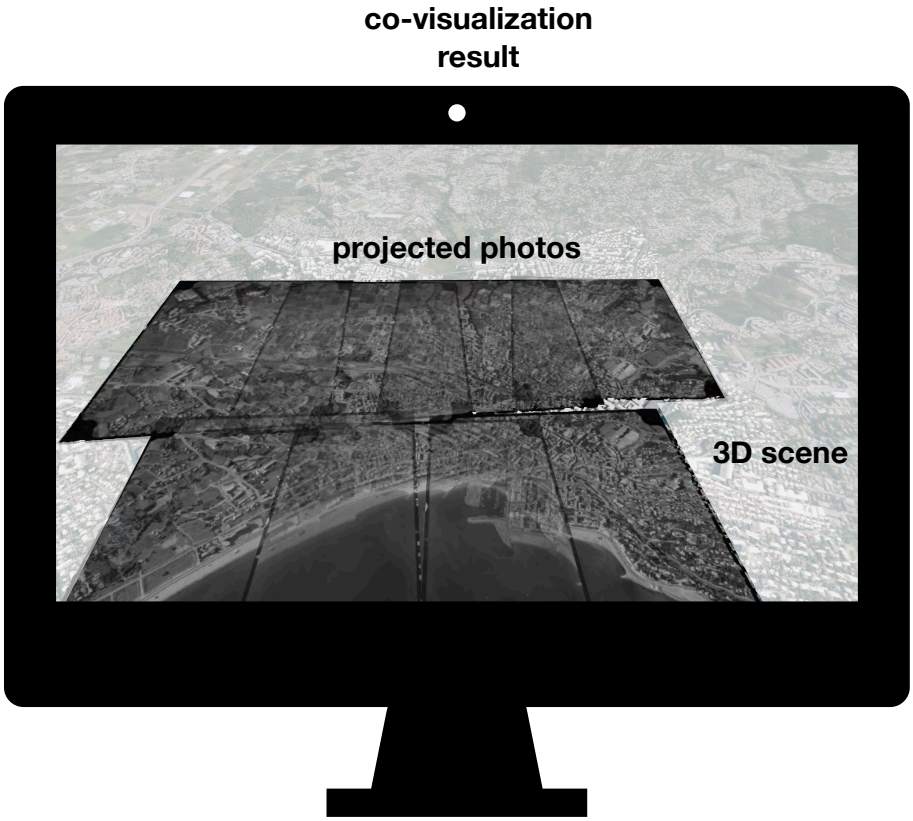


Figure 5.3: 3D city model and various projected 2D images. No interaction objects are available for the user.

We require a navigation support for the users, capable of dealing with massive and heterogeneous data representing large geographical spatial scales. As Figure 5.3 depicts, as the number of projected photos increase, so will do the extent of visual information we will need to manage. We want to add all the respective interactive elements like thumbnails and metadata, which will magnify the visual complexity of our geovis system. Therefore, we face the issue of **managing all this visually complex information collectively** while providing interaction capacities to the users to navigate from image to image.

Objective and Proposed Approach

Our goal is to **propose a set of techniques to facilitate and enhance visual exploration when users discover many historical images**. We address the topic of user interaction considering the potential volume of the input data inside a 3D scene. Accordingly, we aim to tackle this objective by selecting properties (*e.g.*, interaction, saliency, continuity) that a geovisualization system, oriented to the exploration of photographs needs. We combine this set of methods in design guidelines to test and propose the user interaction features to visually explore photos within a 3D environment.

5.2 User Tasks

We aim to facilitate virtual exploration of various photo collections by reducing the visual complexity generated when these photographs are placed inside a 3D scene. Before proposing a specific solution(s), **we will dive into the needs and expectations from the users** towards a spatio-temporal visualization system of historical photographs. Our designs target archivists, historians, and humanities researchers as the primary users. However, we wish our geovisualization application to be accessible for everyone, specifically the general public; therefore, they are also considered users of our geovis system.

We have defined primary **user requirements** in the form of two objective tasks: *(i)* visualizing one or several images in the 3D environment; *(ii)* exploring the images in space and time. We represent these graphically in Figure 5.4 and Figure 5.5 through use case diagrams using the Unified Modeling Language (UML). From the definition of this user's tasks, we can identify preliminary ideas for approaching our research challenge, tackled in Section 5.3.

5.2.1 Task 1: Visualizing One or Several Images

To cross-analyze different photos and collections, we need a system capable of providing multiple image exploration. Let us consider the discovery of many photographs inside a 3D scene composed of city models. As depicted by Figure 5.4, we define the corresponding user task as:

Goal: visualize different images inside a 3D environment.
Sub-Goals are defined as:

1. Project/unproject a photograph on top of the 3D city models.
2. Move to the photographer's viewpoint to visualize pixel-accurate a photo.
3. Read the known information of a specific image (meta-data).
4. Have an overview (*e.g.*, a list) of all available pictures that can be visualized.

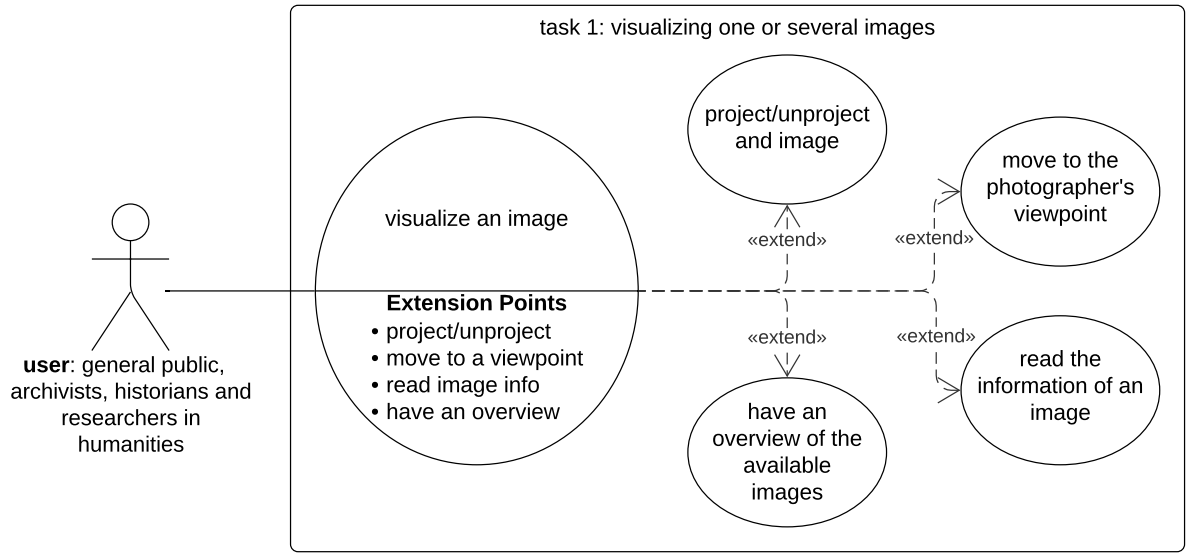


Figure 5.4: Task 1: visualizing one or several images.

5.2.2 Task 2: Exploring Images in Space and Time

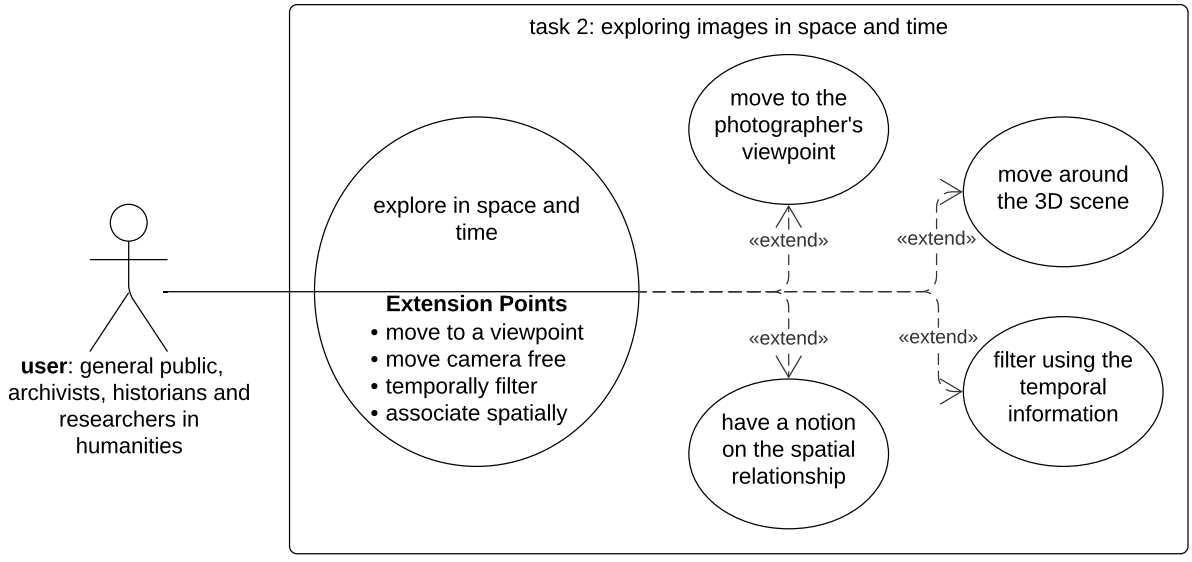


Figure 5.5: Task 2: exploring images in space and time.

To present to the user a spatio-temporal environment, we need a system capable of using and displaying the spatial and temporal data from all the photographs. Let us consider a virtual camera free to move around the 3D scene, where users can explore all available photographs in space and

time by interacting with visual elements inside the scene. As described by Figure 5.5, we define the corresponding user task as:

Goal: explore images in space and time inside a 3D environment.
Sub-Goals are defined as:

1. Take the position and orientation of many different photographer's viewpoints representing historical images.
2. Engage in different angles and visualizations through the 3D scene.
3. Filter images by their temporal information (*e.g.*, date of acquisition).
4. Have a notion of how images relate to themselves spatially.

5.3 Design Guidelines

To **achieve our goal and reach the beforehand described user expectations**, we have selected existing methods (Section 2.4 of Chapter 2) that can help us reduce the visual complexity that arises with the use of heterogeneous and large geospatial data. These are geovisualization techniques generally used to drive attention to specific data or aspects of it, *e.g.*, its semantic, spatial, or temporal features. This section presents the **design guidelines** we have created for each one of these strategies. We divided our proposition into three major categories: *(i)* 2D and 3D representation of the photos; *(ii)* saliency for selected elements and areas of the view; *(iii)* manipulation of the temporal variable (only related to the pictures).

5.3.1 Design Guideline 1: How to Represent a Photograph?

We introduce two different **objects of interaction to represent a photograph** inside a 3D space:

- **Technique 1: Bookmark** – uses the geometric idea of the historical camera to locate interactive 3D objects associated with each image inside the scene.
- **Technique 2: Thumbnail** – orders the photos and places a small-sized version of them outside the 3D view for a quicker interaction.

Technique 1: Bookmark

Definition
A virtual camera defined by its position and camera orientation information. Hence its elements are based on the camera's geometry: optical center, optical axis, image plane, and projected image plane.
Representation: a visual and interactive graphical object displayed to the user in diverse forms, *e.g.*, text, thumbnail image, or a 3D object.
Design Goal: show users the different photographs inside the 3D scene by employing the camera's information of each picture.

We adopt **bookmarks** (or viewpoint markers) inside the 3D environment to **help the user visualize and locate where each photo was captured**. Figure 5.6 shows how we represent these markers based on the geometry of each historical camera. The user can interact with them inside the 3D scene, and when double-click, a "fly-to" animation moves the view from the current position to the selected image viewpoint. This action takes the user to the perspective of the photographer when the picture was acquired.

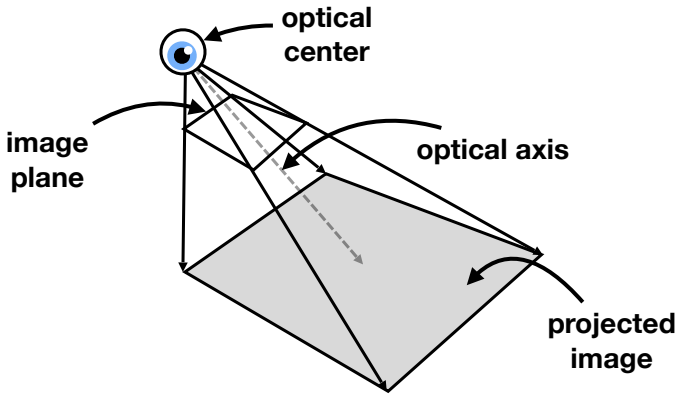


Figure 5.6: Our bookmark having: optical center, optical axis, image plane, etc.

As depicted in Figure 5.7, we designed bookmarks having the following properties:

1. The **camera frustum**, *i.e.*, a pyramid, represents our first viewpoint marker design. The top point (the pyramid's apex) corresponds to the camera's optical center, and the base is the image plane. We **show the corresponding picture in the pyramid base** to visually associate the bookmark with the photograph.
2. The extended camera frustum represents our second design. Instead of placing the respecting photo at the base of each bookmark pyramid, we project the photographs. Hence we leave the base of the pyramid transparent, and when hovering over a marker, we **extend the pyramid to show the image's projection** on top of the 3D scene.
3. We simplify our third design and employ a **ray going from the camera optical center to the center of the footprint** of the projected image in the scene. The image can also be projected over the 3D scene when hovering to connect the respective photo to the bookmark.

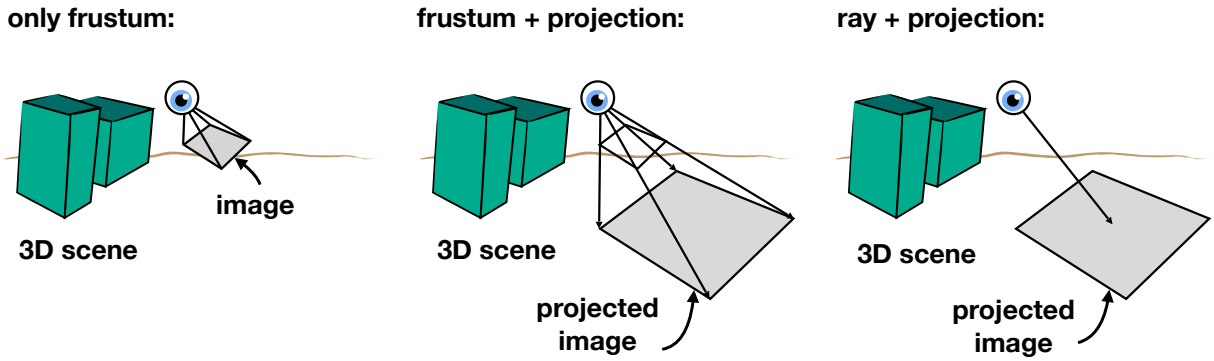


Figure 5.7: Our three designs for a bookmark: frustum, frustum & projection, and ray & projection.

Let us now consider the mapping from 3D space to 2D space that the view camera does for the objects inside the 3D view. Because the bookmarks are placed inside the scene, each will be mapped to a respective 2D representation to fit in the screen space. However, this process may discard the bookmarks placed too far from the current virtual view. To contrast this, we add a constraint to our generic icon positioned at the bookmark's location. More specifically, **we remove the depth factor used during the mapping from 3D to 2D**. Figure 5.8 shows the outcome, an enhanced icon that remains visible even when some markers are too far from the view.

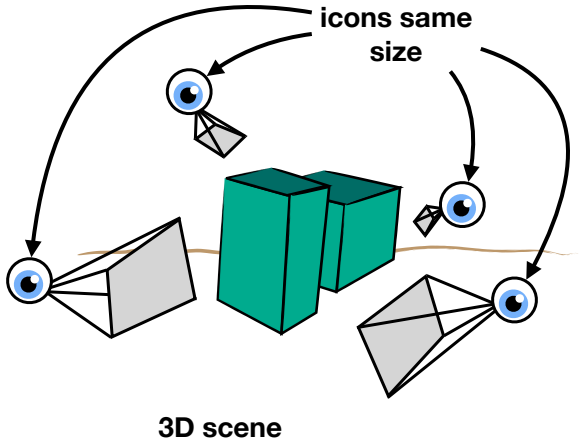


Figure 5.8: Our choice for the bookmark icons. The depth information is not considered. The icon can be visible even if it is far from the view camera.

Technique 2: Thumbnail

Definition
A compressed and minor version of a visual element.
Representation: a visual and interactive graphical object depicting a copy of a 2D image in a small-sized version.
Design Goal: present all the available photographs to the user for a quick preview or to identify the photos fast.

The use of **thumbnails** inside our exploration system allows us to **present all available photographs that can be visualized** and avoids the need for the user to go inside the 3D scene and search for the desired image through all the bookmarks. As Figure 5.9 depicts, we designed two different approaches to be used when placing these sets of thumbnails around the 3D view:

1. A one-dimensional row or column **menu slider**. It can display available images ordered, *e.g.*, depending on a ranking or priority value.
2. An alternative approach using the **boundaries/sides** of the 3D view to show all photos simultaneously. The images can be connected to the 3D scene by mapping the 3D location of the pictures to the 2D border positions.

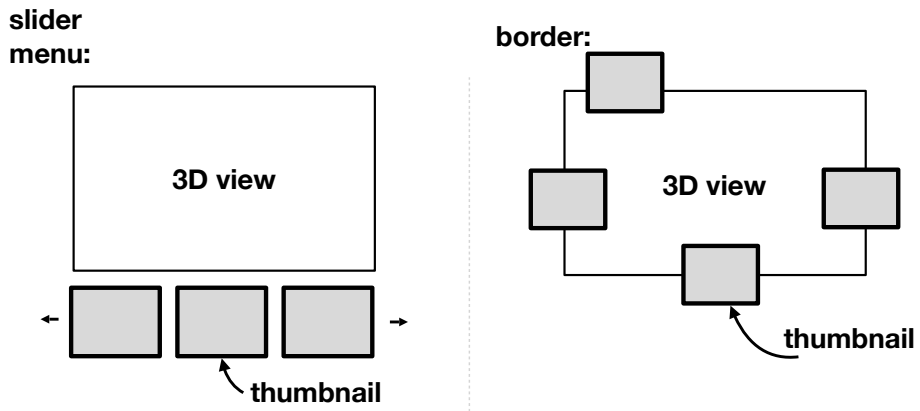


Figure 5.9: Our two designs for the placing of thumbnails around the 3D view: slider menu and border.

To simplify the display of many thumbnails, we propose a **clustering strategy**. We use a ranking function during the clustering process to choose the most interesting photographs for the current view based on the distance between the view camera position in the 3D world and the image footprints. A **hierarchical clustering step groups the photographs into smaller sets and defines the clusters**. The clustering is re-computed for the boundary placement as the user moves the view. This re-calculation results in the movement of the clusters around the view to adapt to their new computed positions.

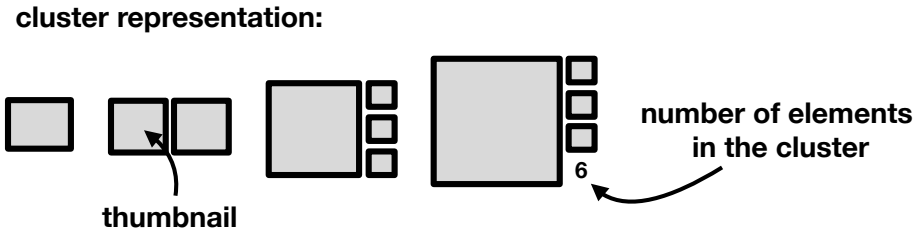


Figure 5.10: Our proposed representation for the clusters.

As presented in Figure 5.10, **to represent the clusters, we choose the gallery visualization** approach that makes the photographs more visible and allows the selection of a specific image immediately. One image is displayed as a simple thumbnail, two are shown next to the other, and for three or more, an image visualizer with a small slider allows the selection of each item.

5.3.2 Design Guideline 2: How to Make Salient Certain Elements?

We design three specific approaches to **capture the user's attention and visualize certain areas or elements inside the 3D space**:

- **Technique 3: Visual Variable** – based on visual perception and is oriented towards all the representations of the visualized photographs, *i.e.*, 2D thumbnails, 3D bookmarks, and 3D projected images.

- **Technique 4: Visibility Heat Map** – allows the user to envision areas of interest inside the scene quickly.
- **Technique 5: Style Representation** – manipulates the style representation of the 3D models to either be homogenized or differentiated from the display photographs.

Technique 3: Visual Variable

Definition

A feature in a graphical object that can make it visually stand out from other objects.

Representation: through the following properties designated on an object: size, shape, orientation, color, texture, location, crispness, resolution, transparency, and arrangement. Note that these are the primary ones, and more can be defined.

Design Goal: give the user a clear visual distinction of each visual or interactive element, *e.g.*, photographs, bookmarks, thumbnails.

We employ the notion of **visual variables** in our elements inside our geovis system to **differentiate between images and photo collections**. We selected only color (hue), thickness, transparency, and size among all the existing variables. As Figure 5.11 denotes graphically, we use them in visual and interactive elements (photographs, bookmarks, thumbnails). We define its use as follows:

- The **color** attribute distinguishes among various image collections (or sources) for a cross-reference between photos.
- The **thickness** discriminates between selected and non-selected photos. It is also designed for saliency towards certain elements, *e.g.*, projected photos with a frame (outline) around or the line depicting the 3D geometries in bookmarks.
- The **transparency** shows images outside the field of view of the view camera.
- The **size** change pops up an element in the view (either in the 3D environment or outside of it) to let the user know it is an interactive object.

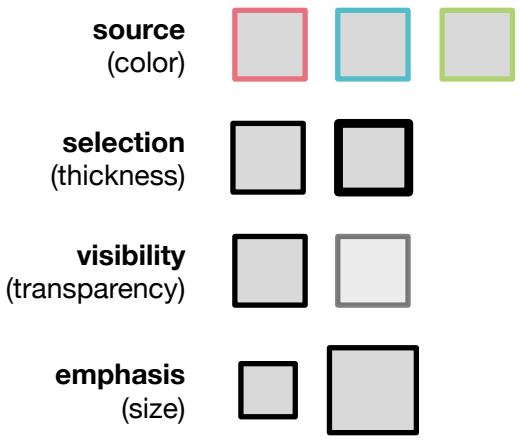


Figure 5.11: Our selected visual variables and their usage.

Technique 4: Visibility Heat Map

Definition

A graphical representation of data in which colors represent its values.

Representation: a 2D map where higher color intensities depict higher numbers.
Design Goal: exhibit within the 3D scene the areas where the user should pay attention and explore/navigate through since the photographs depict mainly those geographic places.

We take advantage of a visibility heat map to **identify possible areas in the scene with large numbers of photos**. Specifically, we consider the areas in the 3D models where the pictures are projected. We delimit its extension with the footprints of these projections. We also designed a "fly-to" animation, which is utilized to move the view to a position where the whole range of this visibility heat map is covered.

This technique allows the user to notice inside the 3D scene the most interesting geographic areas (since the images describe them), as shown in Figure 5.12. To represent the variation of color in the heat map, we define a variety of hues between Red Blue Green (RGB), where red denotes the highest number of images and blue the lowest.

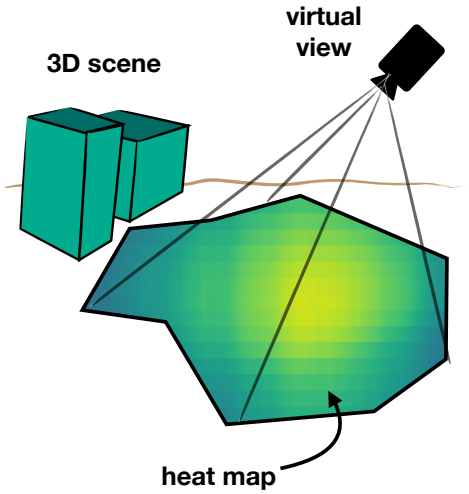


Figure 5.12: Our designated way (a visibility heat map) to identify areas of interest.

Technique 5: Style Representation

Definition

A graphical representation of geospatial data.

Representation: through the rendering of diverse styles, colors, and levels of details for the 3D models inside the scene.
Design Goal: allows the user to explore possible designs and styles for the 3D models according to the user use context, *e.g.*, homogenize it with the projected photographs.

We now consider the **style representation of the 3D models being presented to the user when the photos are visualized**. Using these 3D objects inside the digital scene can present additional user information apart from the one shown in the content of a photograph. We define it as presenting a context in which the photo is seen. Instead of imposing a specific style on this virtual scene, we update and change it as needed. For instance, an abstract representation may allow showcasing the photos' content more in-depth by removing all external noise. At the same time, the use of an orthoimage may reveal the area's evolution from the acquisition period to now.

To this style representation, we add the possibility of changing the radiometry, *i.e.*, the colors, of the 3D models and their textures by employing predefined Lookup Table (LUT) (determines the color values from a range of data values). For example, as exhibited in Figure 5.13, if we assume that the small squares represent the color scheme of our 3D scene (left side), we can change it, for instance, to the same black and white tone from a selected photo. Therefore, **according to some visual properties to preserve in our 3D environment, *e.g.*, contrasts, the saliency, photo-realism, or abstraction level, both style and radiometry manipulation can allow users to differentiate or homogenize the photographs with their surrounding 3D scene**. Additionally, we add the plain option of transparency in the projected photos as a simple solution to compare the 3D context and the photographs (right side of the image).

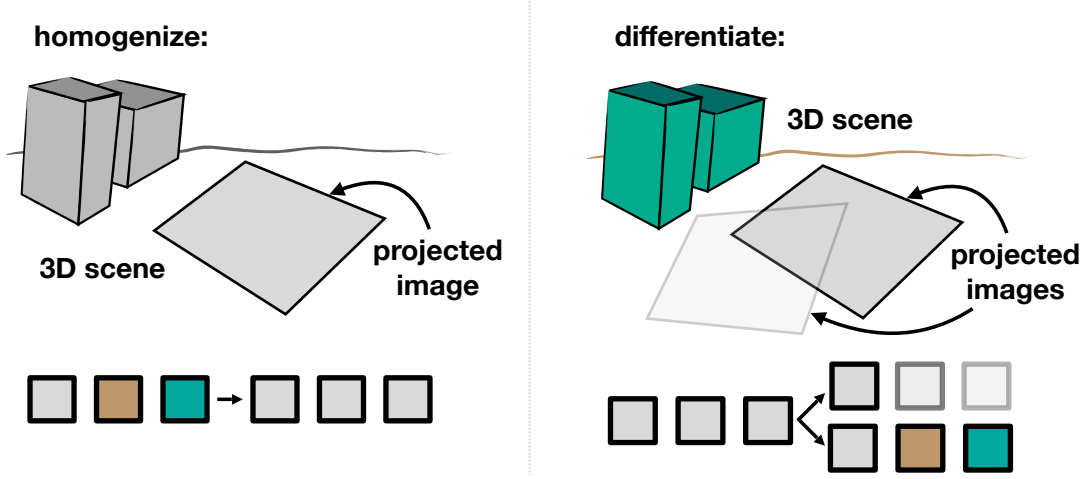


Figure 5.13: Our two approaches for the style representation of 3D models: homogenize or differentiate.

5.3.3 Design Guideline 3: How to Add the Temporal Aspect?

We would like to the **temporal information** associated with the photographs to **manipulate the set of images presented to the user**. Note that we do not explore the evolution of the 3D scene in the time since we consider it out of our research scope. However, it can be considered as a possible extension for future work.

- **Technique 6: Timeline** – designed to select a time frame in which only the photographs acquired during this period are shown.

Technique 6: Timeline

Definition
A graphical representation of a period of time.
Representation: a visual and interactive graphical object displaying a set of dates in a specific time range in chronological arrangement. It allows selecting in it a specific time frame.
Design Goal: shows only the photos acquired during the time period specify by the user.

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 5.14. 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 (blue bar in our diagram) and display only the photos belonging to it.

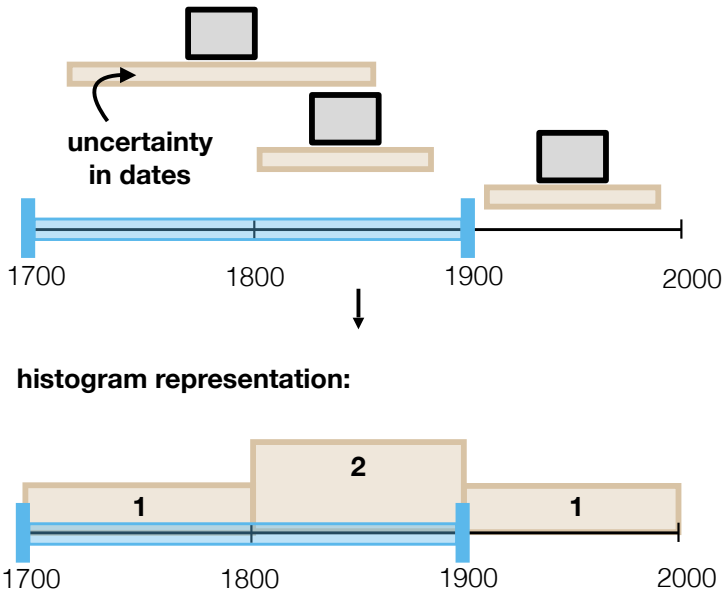


Figure 5.14: Our time axis, which employs a histogram to overcome the uncertainty in dates. The blue bars show the selected time range.

5.4 Results: Testing Our Design Guidelines

To **evaluate our design guidelines**, we implemented them in our online application using the *three.js*[1] WebGL rendering library and the *d3.js*[2] data-based library. Note that in this chapter, we

[1] https://threejs.org
[2] https://d3js.org

present each technique separately (no connection between each method), and only certain aspects of our propositions and designs were tested. The upcoming Chapter 6 will dive more in-depth into each technique's overall implementation and connection towards the historical context and the user evaluation. Therefore, in this section, we present our initial results from testing the following three design guidelines.

Evaluations

- **Test 1:** relates to our design guideline 1, "how to represent a photograph?" in Subsection 5.3.1. It compares the different styles we designed for the two interactive objects (bookmarks and thumbnails) for the photographs inside a 3D space.
- **Test 2:** linked to our design guideline 2, "how to make salient certain elements" in Subsection 5.3.2. We evaluate the use of different techniques (image frames, visibility heat maps, and style representations) for saliency inside a scene.
- **Test 3:** associates to our design guideline 2, "how to add the temporal aspect" in Subsection 5.3.2. It examines our first versions of the timeline.

Following the same image-based rendering approach, we present the images within a 3D model by projecting them onto the 3D scene. As we mentioned in Section 3.3 of Chapter 3, we require as input data: *(i)* the 2D **photos**; *(ii)* the **camera orientation** information of each photograph (intrinsics and extrinsics only. At this point, we do not use the distortion model of the camera); *(iii)* a **3D model** of the photographed scene. Additionally, we consider the **metadata** information known from each photo as an additional input, *e.g.*, title and date of acquisition. We use the same experimental datasets from Section 4.4.1 in Chapter 4 summarized in Table 4.1 for the experiments.

5.4.1 Test 1: Bookmarks and Thumbnails for Interaction with Photographs

We examined our three initial **designs for the bookmarks** representing specific points of view of each photograph loaded in the 3D environment. An example of our results from the Maurepas and Frejus datasets is shown in Figure 5.15. From this initial evaluation, we were able to remark the following:

1. When a **frustum** is used as a bookmark, and the **image plane** (pyramid's base) is **textured** with the respective **photo**, we observe potential occlusions over the 3D scene from this image plane. It is magnified as the size and number of bookmarks increase.
2. Instead of texturing the image plane of the frustum, we leave it transparent and only **extend the pyramid towards the projection of the photograph** in the 3D scene. Since the pyramid of the frustum has to be resized, we notice that depending on this new dimension, the frustum sometimes could not be fully visible in the current view because the scene is occluding it on specific areas (mainly the corners). Nevertheless, the transparent use of the image plane allows the removal of the occlusion of the 3D scene by the textured image planes.

3. Having only a **ray towards the center of the projected image** diminishes the number of visual elements in the scene. It is mainly seen when many bookmarks are placed inside the 3D environment. However, it can also be remarked that connecting the bookmark with the projected picture is more challenging when many photographs are already projected into the scene.

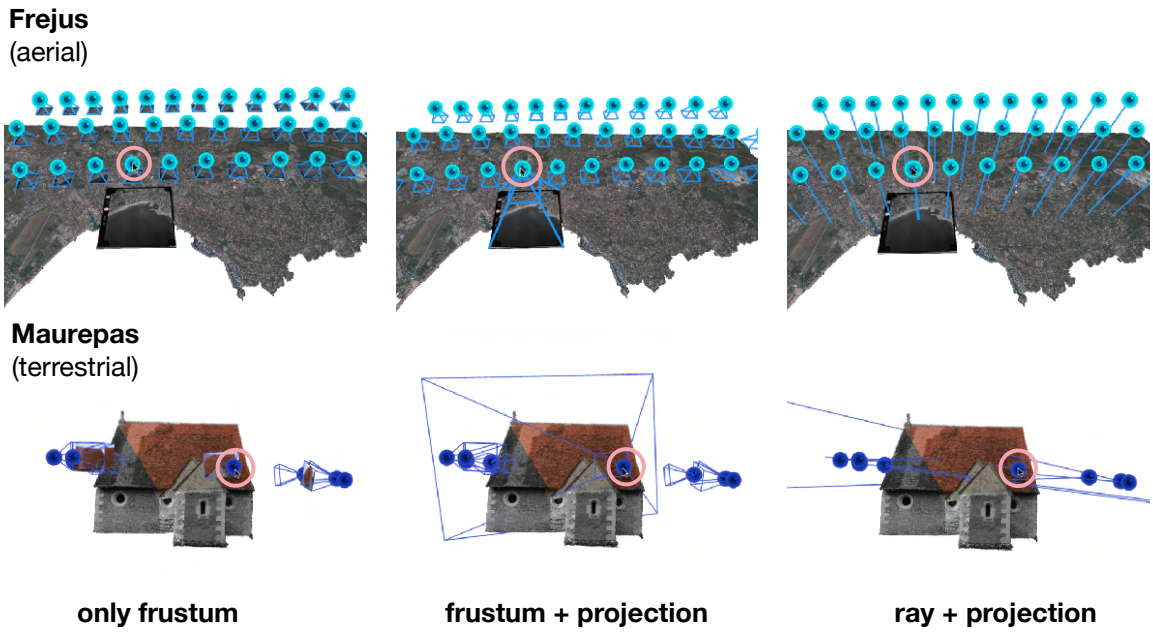


Figure 5.15: Comparison of the three designs for bookmarks representing the viewpoints of each photograph. The pink circles are only for illustration purposes to show where the cursor is located.

We also evaluated our two **designs** to provide a **menu for quick selection employing thumbnails**. Figure 5.16 displays a sample of our results for the border positioning, where we present the Saint-Martin and Viabon datasets. From our fundamental evaluation, we were able to observe the subsequent:

1. We notice that while the **slider menu** is a more straightforward solution, as the number of loaded photographs increases, so does the time it takes to search for a specific image and select it. Additionally, this implementation gives a more static visualization since it does not change as the viewer moves in the 3D environment.
2. Our **border placement** allows exploring a more dynamic approach. As we move the view in the 3D scene, the thumbnails adapt their position to match the one in 3D from the projected photos. We observed two possible constraints to this border strategy:
   - The movement of the thumbnails around the border should remain continuous to avoid the effect of elements appearing and disappearing in the scene.
   - When all the images are located (in the border) around the same position, much space is left that is not used around the 3D view. The thumbnails are occluding themselves while there is left free space in other border regions.



Figure 5.16: Thumbnail placements around the 3D scene. **Right:** a menu slider showing a list of thumbnails. **Left:** the thumbnails are spatialized; they move around the border as the view camera moves.

5.4.2 Test 2: Image Frames, Visibility Heat Map, and Style Representation for Saliency

To **differentiate between each projected photograph and the 3D scene**, we tested two **visual variables**: color (hue) and thickness, defined by an outline around the projected image (a frame). Figure 5.17 shows a sample of our results from the Frejus dataset, where we evaluated three different

cases related to the use a colored frame:

1. When **no border** is applied, the view does not provide any saliency on the pictures or a clear limit between the projected photographs and the textures or colors used on the 3D models.
2. The surrounding of **every projected image with a frame** helps distinguish each photograph from the other and the 3D space. Nonetheless, when many images are projected in the same area, finding a particular photo in the overall projection becomes visually challenging since many lines (from each picture outline) are intercepting with each other.
3. To avoid intercepting with every frame, we use the option of only the **overall border of all projections**. The result is that the outline separates the 3D space from the photos and keeps the images salient.

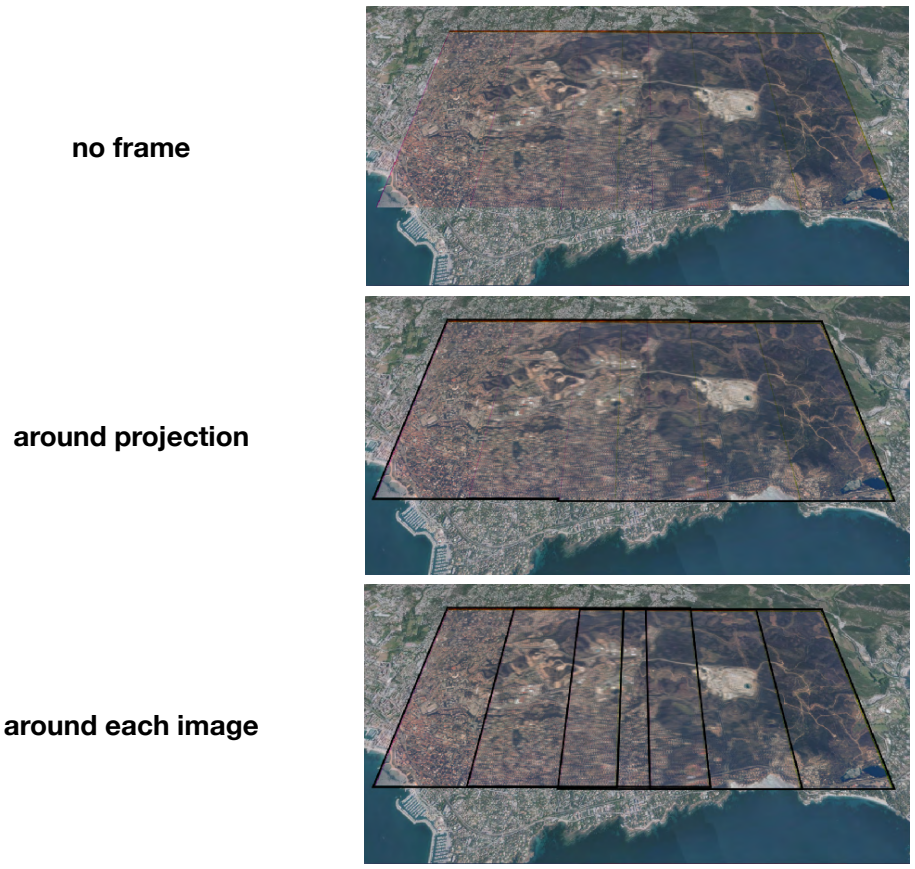


Figure 5.17: Colored frame for the projected photos. Comparison between adding no frame, a frame around the overall projection, and a frame around each image.

Another of our designs to **bring saliency in the visualization** is the use of **style representation**. As shown in Figure 5.18 with the Viabon and Frejus datasets, we examined the option of changing the radiometric values of the 3D models to be adapted to other existent Lookup Table (LUT), *e.g.*, based on black and white or sepia tones. It permits to delimit the 3D scene from the current projected photos. From our evaluation, we consider that this technique can also transfer the color scheme from a photograph to the 3D scene and homogenize both 3D models and the projected photos.

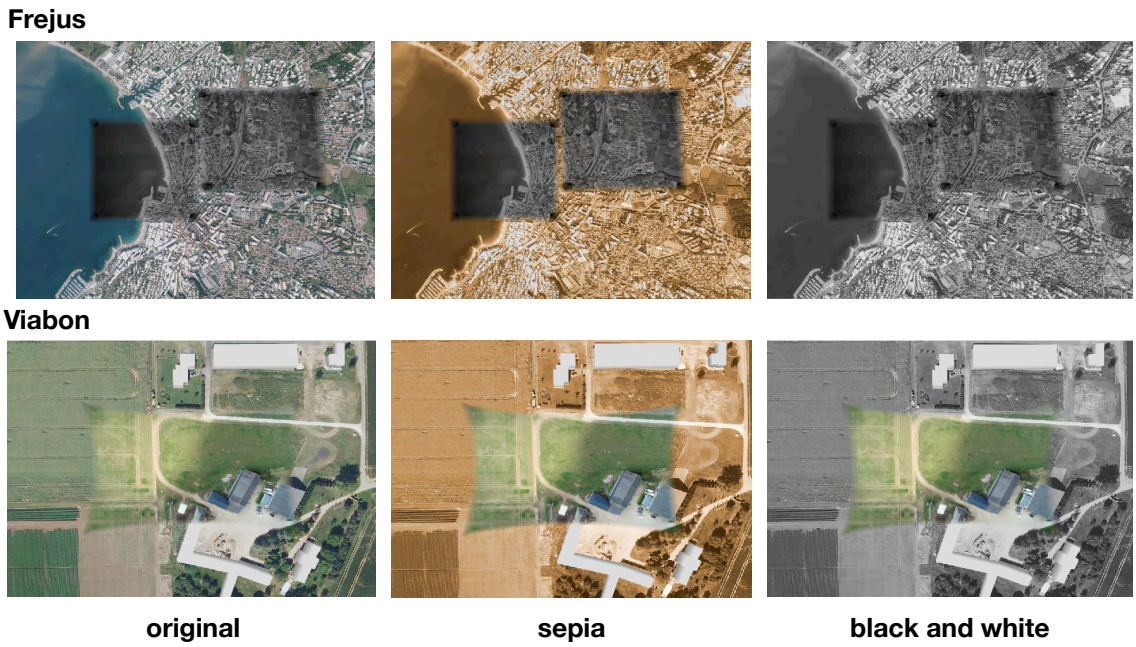


Figure 5.18: Use of style representation for the radiometry of 3D models. The radiometric changes in the scene affect the saliency of the projected images. The borders of the photos have been softened for their blending with the 3D models.

On a different note, the **visibility heat map** is able to **expose the covered extent of the photographs**. We use a basic coloration of RGB to represent the number of photographs projected at a specific point in the 3D scene. Figure 5.19 exhibits an example from the Frejus, Maurepas, and Saint-Martin datasets. We use red to represent the areas with the most significant number of photographs located in a point, while blue represents the least number of photos (equals 1). We consider that its use should be paired with the visual variable of transparency to allow users to see the 3D models under the mapped area.

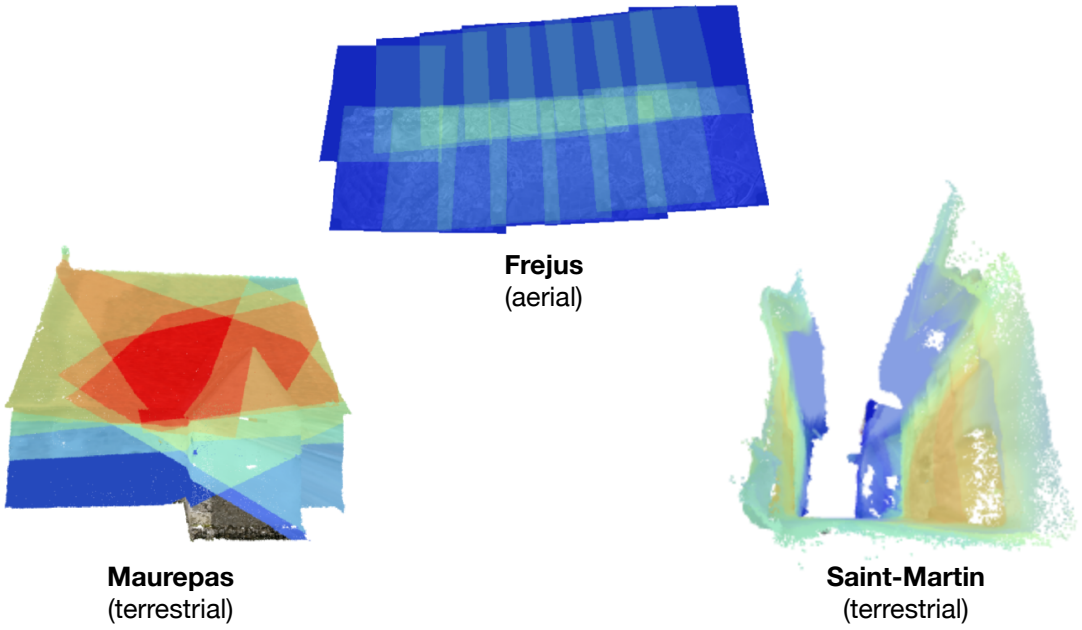


Figure 5.19: Heatmaps projected over the 3D scenes composed of point cloud geometries.

5.4.3 Test 3: Timeline for Temporal Filtering

Finally, we assess our design for the **time axis in the form of a histogram**. The time selection area represents the filtering of the images, *i.e.*, only the photos inside this defined range are available to be visualized in the 3D scene. We establish time segments, *e.g.*, a year or a month, and compute the value of that specific time segment for each bin of the histogram. To test our design, we define a random acquisition time for each image with an uncertainty of ±5 years (as depicted in Figure 5.20). For instance, one picture could be defined with a date of acquisition as 1920 and another from 1920 – 1923. Our two different evaluated representations and their outcomes are:

1. **Expected Number of Images:** We model the time interval of each image with a continuous uniform distribution. The value of each bin (from the histogram) depicts the expected number of images belonging to that specific time segment, represented by the probability in the uniform distribution function. Accordingly, the sum of all bins in the histogram is equal to the number of historical images. The constraint we found on this method is that it is challenging to be understood by non-expert users.
2. **Maximum Number of Images:** We simplify the histogram by giving each bin the maximum number of images that may have been taken during a time frame. We realized that this second approach is more user-friendly and easier to understand than the first one during our tests. However, its only constraint is that the users may think that the sum of the bins is the number of available images without considering the uncertainty on the dates.

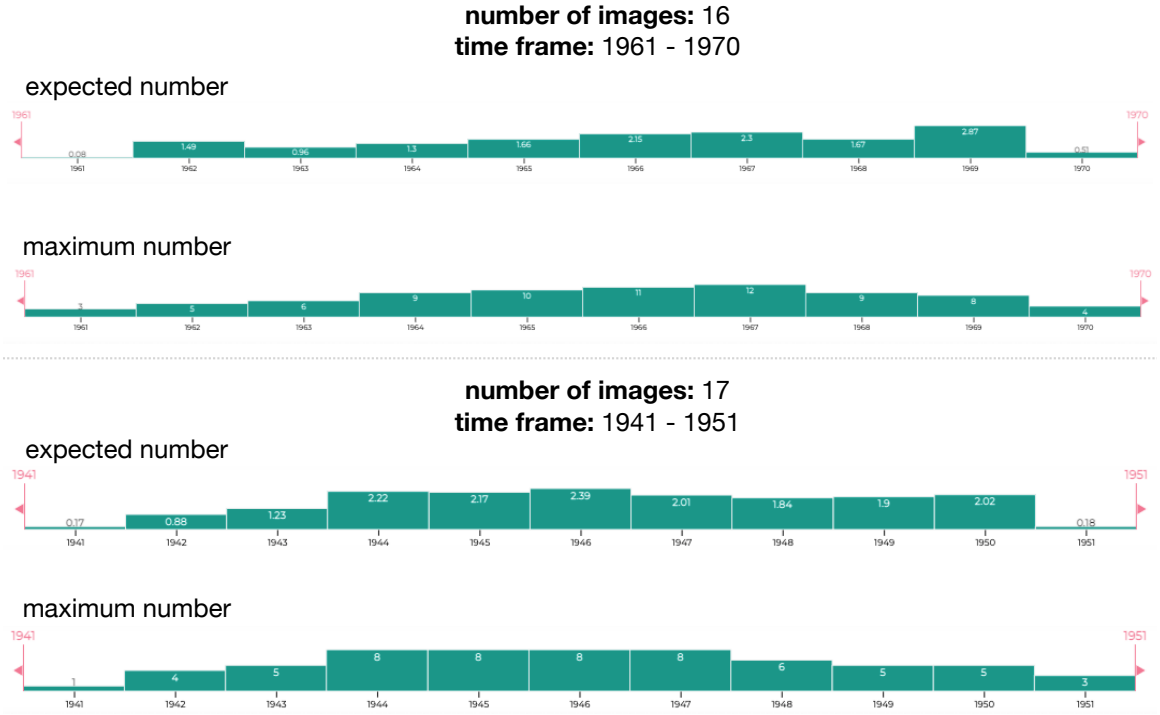


Figure 5.20: Histogram representations for timeline. Comparison between using the expected and maximum number of images for each histogram bin. Each image presents an uncertainty on the acquisition data of ±5 years.

5.5 Conclusion

Subsection 5.4.1 (test 1: bookmarks and thumbnails for interaction with photographs) enabled us to properly acknowledge that the gallery/clustering approach on the borders is a more dynamic alternative for visualizing photographs inside a 3D scene. Using the spatial information of the pictures, this approach allows the thumbnails (or clusters) to move and adapt to the current view. Therefore, it will be our selected proposal to visualize historical photographs when their orientation information is known or has been calculated. On the other hand, we choose the menu slider as a better alternative for photographs that have not been spatialized, but still, we wish to present them to the user.

Subsection 5.4.2 (test 2: image frames, visibility heat map, and style representation) allowed us to visually comprehend that the use of color did indeed made more saliently the projection of each photograph. On the other hand, the thickness of the frame is another attribute that we acknowledge since when it is too thin, the outline is not easy to visualize, and the effect is the same as if there would not be any frame around the image. However, when this thickness is too high, it clutters the 3D view with too many colored lines. During this evaluation, we only assessed the visual variables on the projected images using the color frames. However, we think its use can benefit our interaction objects (thumbnails and bookmarks). In general, we prefer to connect all these designs (for instance, the image frames, visibility heat map, and style representation) in a single visual exploration system where the user can choose the level of saliency depending on the context of use.

In Subsection 5.4.3 (test 3: timeline for temporal filtering), we compared and tested our two different representations for the histogram in the time axis. From our results, we chose the more straightforward approach (second option), where each bin value represents the number of photos that may have been taken during the time range of that histogram bin. We noticed that the continuous uniform distribution representation (for each bin) was challenging to understand because the sum of the probabilities is a floating number represented with decimal numbers, which does not correlate with the idea of a photograph denoted as an integer value.

Contribution

In summary, the main contribution of this work is **the definition of new design principles for six central and existing geovisualization techniques (bookmarks, thumbnails, visual variables, visibility heat maps, style representations, and timelines)**. We propose a structure in which these set of methods can be used through the definition of three design guidelines oriented towards a visual exploration of historical photographs: *(i)* the representation of photos in 3D environments; *(ii)* the increase of salience on certain elements; *(iii)* the addition of the temporal information of the pictures in the 3D space.

Chapter 6

HISTOVIS: Co-Visualization of Historical Photos and 3D City Models

Preamble

The following chapter introduces HISTOVIS, our implemented **prototype oriented towards the co-visualization of historical datasets and contemporary 3D city models**. We propose integrating our previous contributions from Chapter 4 and Chapter 5 to achieve an online interactive visual system capable of manipulating and exploring historical photographs. This prototype is developed in the framework of the ALEGORIA project; therefore, our designs and evaluations have been oriented towards its needs and the provided data.

Section 6.1 presents the Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge (ALEGORIA) project and the existing geospatial datasets offered by it. Section 6.3 exhibits our proposed architecture for a web application supporting the exploration of many photographs inside a 3D topographic and urban model. Next, Section 6.4 displays different use cases addressing distinct types of datasets. Finally, Section 6.5 presents the results from our user study, and Section 6.6 concludes with a summary of our findings from our tested prototype.

Contents

6.1 Presentation of the Historical Data Collections **101**
6.1.1 National Archives (AN) . . . 102
6.1.2 Nicéphore Niépce Museum (NNM) . . . 103
6.1.3 French National Mapping Agency (IGN) . . . 104
6.2 Modeling a Historical Photograph Inside an Interactive System **107**
6.2.1 Interaction States of an Image . . . 107
6.2.2 Model of a Historical Image . . . 108
6.3 Architecture and Pipeline of HISTOVIS **108**
6.4 Usage of HISTOVIS **114**
6.4.1 System and Initial Configuration . . . 114
6.4.2 User Interface and Functionalities . . . 116
6.4.3 Use Case Scenarios . . . 124
6.5 Evaluation **130**
6.5.1 User Study on Scenario 1: Visualization of Two Photo Collections . . . 130
6.5.2 ALEGORIA Stakeholders Feedback . . . 135
6.6 Conclusion **137**

6.1 Presentation of the Historical Data Collections

The French National Research Agency (ANR) has funded this study in the framework of the Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge (ALEGORIA) program (2018-2022). Its purpose is to **valorize extensive iconographic collections** depicting the French territory at different times from the interwar period to the current date. Three Galleries, Libraries, Archives, and Museums (GLAMs) are in charge of providing the photographic collections for the study: *(i)* the National Archives (AN); *(ii)* the Nicéphore Niépce Museum (NNM); *(iii)* the French National Mapping Agency (IGN). Table 6.1 exhibits a complete description of all the available datasets from all these institutions.

Table 6.1: Overall description of the historical image collections from the ALEGORIA project. These pictures are geographic shots depicting the French region. Note that for the moment, AN and NNM collections are not accessible online.

Institution	Collection	Type	Context	Point of View	Date	No. Images
AN	**Carto MRU**	image	cartographic	oblique	1948 – 1976	13 562
	Lapie		commercial	oblique	1955 – 1965	28 400
NNM	**Combier**	image	postcard	aerial and terrestrial	1949 – 1974	10 000
	Bouquet		army	vertical	1914 – 1918	49
	Bron		-	vertical and oblique	1924 – 1926	130
	Falba		-	vertical	1951 – 1960	550
	Henrad		-	oblique	1930 – 1970	1100
	Franjus		commercial	oblique	1960 – 1990	877
	Lapie			oblique	1955 – 1965	739
	Charles Gross		-	terrestrial	1975 – 2003	110
IGN	**BD ORTHO®**	image	cartographic	vertical	1945 – Now	2 800 000
	BD ALTI®	geometry			1987 – Now	-
	BD TOPO®				1998 – Now	-

The iconographic collections from the ALEGORIA project are valuable in terms of Cultural Heritage (CH). Figure 6.1 presents a set of heatmaps that visually describe the position where photographs were taken from three sample datasets (Carto MRU, Lapie, and Combier). In this section, **we explore the different participating institutions, and their corresponding provided geospatial data**.

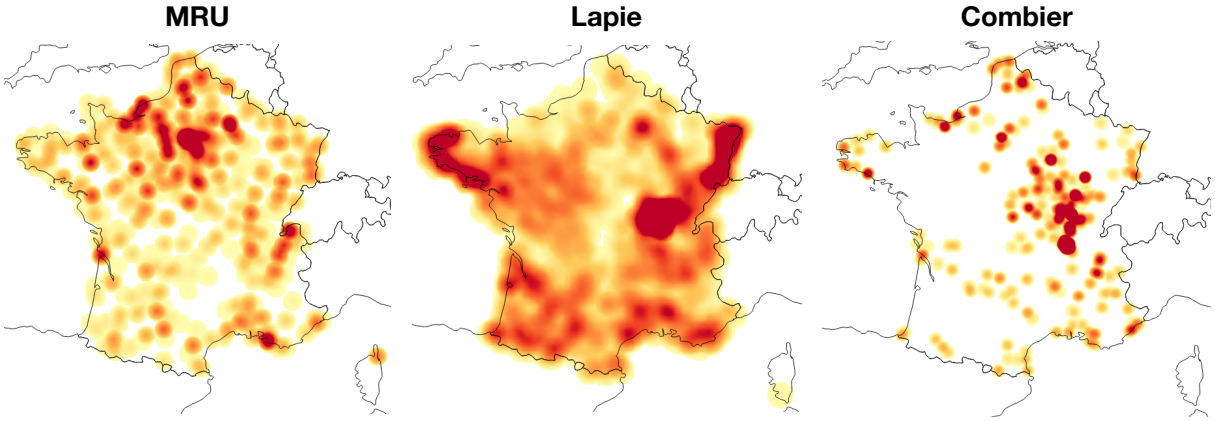


Figure 6.1: Heat maps visually describe the geographic locations where pictures from some ALEGORIA photo collections were acquired. The color red symbolizes the highest concentration. These maps were taken from the document of Lheureux et al. (2018).

Additional Information

The primary source of information used to describe the available datasets is the internal documentation from the ALEGORIA project (Gouet-Brunet et al. 2017; Lheureux et al. 2018). However, more details can also be found in the published papers related to the work done during the project (2018 – 2021) (Delavoipière, Conord, and Marshall 2020; Lecat et al. 2021).

6.1.1 National Archives (AN)

The National Archives (AN)[1] of France offers two extensive image compilations. From the map library collection of the Ministry of Reconstruction and Town Planning (MRU), 13 562 images of the first air missions between 1948 and 1976 were selected corresponding to major reconstruction projects. In the Lapie collection, 28 400 aerial views of these same reconstructed regions were chosen selected to constitute a coherent corpus, generating relevant associations of images.

Among the regions and localities represented by this corpus areas are Normandy with Le Havre shipyards, the west coast with Royan, the Mediterranean with Marseille, the north and the border towns of eastern France, the center region with Orléans, and Île-de-France. Although not affected by the reconstruction, Burgundy was added to facilitate crossings with the Nicéphore Niépce Museum collections.



Figure 6.2: MRU photo collection, Bouche-du-Rhône (Marseille), May 3, 1952, mission Nº 328, digital image polarized from negative on a flexible support, 11×13 cm, 20130290/4. Credits to National Archives.

Carto MRU Collection

Photo collection composed of aerial views that cover all of France for 30 years. The photographs were used to develop maps and produced by the MRU. More accurately, they were employed for a photogrammetric purpose to specify precise data such as the relief or volume of buildings before executing topographic plans. As the example from Figure 6.2 shows, the pictures represent various infrastructures like housing, factories, roads, natural landscapes, and historical monuments. Today, the images constitute an important historical source for studying the French territory and its evolution between 1945 and the 1960s.

[1] This institution preserves most of France's official archives. More information can be found at:
https://www.archives-nationales.culture.gouv.fr

Lapie Collection

A compilation acquired in 1972 by the AN. The planes of the Lapie company crisscrossed the French territory from the 1950s to the beginning of the following decade. Thousands of photographs were collected to produce educational documentation and postcards for tourism support (originated in a commercial context). As depicted by Figure 6.3, the collection is constituted by oblique aerial views at medium and low altitudes and includes 300 000 photos in black and white, color, negatives, and prints. For the project, color prints dating from the period of 1955 to 1965 were digitized. Additionally, a small number of black and white prints were included. The images showcase a change in France, juxtaposing traditional farms, old industries, and innovative infrastructures against accelerated urbanization.



Figure 6.3: LAPIE company, Lapie photo collection, Loyette (Ain), 1955 – 1965, mission BX 78, Nº 3K, chromogenic print, 12.7×17.7 cm, 1PH/C/1. Credits to National Archives, no commercial use, solely research and development.

6.1.2 Nicéphore Niépce Museum (NNM)

The Nicéphore Niépce Museum (NNM)[2] provides various photo collections dedicated to aerial and terrestrial imagery. One of the broadest, the Combier collection, outlines old photographs applied to the production of postcards following the end of World War II. This compilation is complementary to the Lapie samples from the National Archives. In addition, several other photographic collections kept at the NNM contain aerial and terrestrial pictures for military, tourist, educational and cartographic use, covering a broad spectrum of inter-domain photography. A total amount of 13 500 images were digitized.

Combier Collection

In 1975, the museum received the photographic archives from the company Combier printer In Mâcon (CIM). The organization was created by Jean Combier, a freelance photographer who established himself in 1922 as a postcard maker. He was the first postcard publisher to distribute aerial photographs of French municipalities without neglecting the terrestrial shots. Figure 6.4 is a sample acquisition turned into a postcard. As a result, the Combier archives provide a better understanding of the commercial, technical and



Figure 6.4: CIM company, Combier photo collection, Lozère (Gorges du Tarn), 1975. Credits to Nicéphore Niépce Museum.

[2] Dedicated to the inventor of photography Nicéphore Niépce, the museum preserves the history of this discipline. More information can be found at:
https://www.museeniepce.com

iconographic challenges of 19th century photography. The whole collection comprises more than 900 000 photos and the museum has digitized 10 000 pictures focused mainly on the regions of Saône et Loire and Ain.



Figure 6.5: Henrard photo collection, Place de la Concorde (Paris), 1930-1970. Credits to Nicéphore Niépce Museum.

Other Collections

Apart from the Combier collection, the museum has aerial and terrestrial photographs, related to military intelligence, aerial archeology, geographical inventory (cartography, geology, forest interpretation, and documentary photography) as artistic photographs (see some examples in Figure 6.5 and Figure 6.6). They constitute around 10 000 views.

Bouquet: a lot of 49 aerial photographs, taken by the French army, illustrates Jacques Humbert's conference in 1934 on the action of the Barbot en Artois division during World War I.

Bron: purchase in 1980 by the museum, an album composed of 91 plates containing 1 to 2 photographs per plate, centered around the commune of Bron in Lyon.



Figure 6.6: Charles Gros photo collection, Boucicaut district (Chalon-sur-Saône), 1989. Credits to Nicéphore Niépce Museum.

Falba: acquisition of 550 aerial views collected by Roland Falba between 1951 and 1960. The photos cover France, without precise location.

Henrard: collection of 1300 aerial views of Paris and the surrounding region.

Franjus: aerial view samples from the FRANJUS company, active from 1960 to 1990. The compilation is made up of 4559 negatives.

Lapie: a counterpart of the Lapie photo collection from the National Archives. The images are also oblique aerial views produced in a commercial context between 1955 to 1965.

Charles Gros: terrestrial acquisitions of Chalon-sur-Saône in the region of Bourgogne-Franche-Comté. The set is constituted of 110 photos between 1975 to 2003.

6.1.3 French National Mapping Agency (IGN)

The French National Mapping Agency (IGN)[a] has produced and archived much cartographic information for countless years. Aerial captures recorded since 1945 constitute the historical BD ORTHO® database of the institute. Recent digital terrestrial acquisitions are also available

[a] French public administrative institution dedicated to produce and maintain geographical data. More information can be found at:
https://www.ign.fr

through the Stéréopolis terrestrial digitization system. Various outcomes related to IGN's data are open to the public through the website *Geoportail*[3].

Additionally, IGN's objective is also the production of 3D city models. These can be created using geographic vector databases, which aim to describe the topography of the land. The geographic entities are represented in elementary and identifiable objects. Their shape and location are depicted by one or more geometric primitives (point, line, or polygon), chosen according to the nature of the geographic entity. The geometric detail is a set of alphanumeric attributes that completes their description (Bian 2009; Harmon and Anderson 2003). IGN's vector database follow an open CityGML[4] data model. BD TOPO® allows the creation of 3D models of a Level of Detail (LOD) either of LOD 0 or LOD 1, where the ground and basic shape (rectangular) of buildings can be defined.

IGN BD ORTHO®

The IGN has archives of aerial captures since 1945, making possible to constitute the IGN's historical BD ORTHO® database[5] using shots taken on different dates on the same land area (see an example in Figure 6.7). Today, it represents nearly 4.7 million aerial photographs in black-and-white or color. A total of 4 million have been digitized and classified by mission. The photos are relatively well documented in time and space (area, location, cameras used, date and time of flight, resolution, height of the sun, and specific pose for the most recent acquisitions). Some of these images (2.8 million) are accessible via IGN's *Remonter le Temps*[6] portal.



Figure 6.7: Extract of IGNs BD ORTHO®. Credits to IGN: https://geoservices.ign.fr/bdortho.

IGN BD ALTI®

The IGN's BD ALTI is a raster-based database[7] describing the relief of the French territory on a medium scale through a Digital Terrain Model (DTM). Between 1987 and 2001, it was calculated

[3] https://www.geoportail.gouv.fr
[4] https://www.ogc.org/standards/citygml
[5] https://geoservices.ign.fr/bdortho
[6] https://remonterletemps.ign.fr
[7] https://geoservices.ign.fr/bdalti

through the digitization of maps and a photogrammetric process. Nowadays, it uses the IGN's RGE ALTI® database to define the shape and elevation of the ground surface on a larger scale. It is mainly obtained from aerial surveys utilizing a Light Detection and Ranging (LIDAR) method. It constitutes a DTM with a step of up to 1 m over the whole of France.

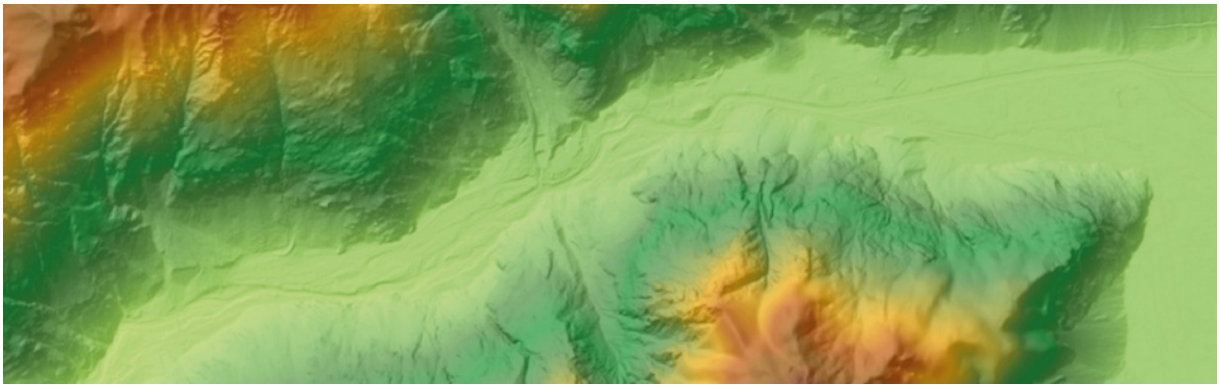


Figure 6.8: Extract of a cartographic representation of the BD ALTI®. Credits to IGN: https://geoservices.ign.fr/bdalti

IGN BD TOPO®

The IGN's BD TOPO® is a vector-based database[8] that defines the French territory's elements and infrastructures with metric precision, usable at scales ranging from 1:2 000 to 1:50 000. The geographic entities are accompanied by altimetry information and, in the case of buildings, by height. Thus, this database provides a structured and semantic description of the territory. The themes represented are the road network, the rail network, the energy transport network, the hydrographic network, buildings, tree vegetation, and toponyms. The timeliness of the data varies from 6 months for the leading road network to a maximum of 5 years for themes produced from aerial shots such as buildings, hydrography, or the electricity network. Figure 6.9 depicts a sample of the cartographic representation of the BD TOPO®.

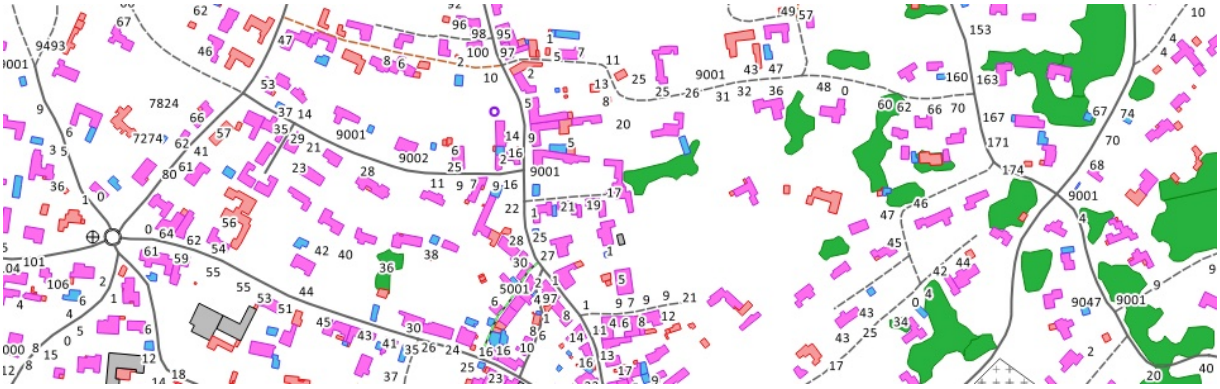


Figure 6.9: Extract of a cartographic representation of the BD TOPO®. Credits to IGN: https://geoservices.ign.fr/bdtopo

[8] https://geoservices.ign.fr/bdtopo

6.2 Modeling a Historical Photograph Inside an Interactive System

We will now present a more conceptual view of the **representation of a historical photograph inside a 3D interactive visualization system**. We illustrate first the different states in which an image can be defined inside a 3D environment through the interactions by the users. From this classification, we particularly model in this interactive system a historical image and its corresponding parameters.

6.2.1 Interaction States of an Image

The visibility and interactivity of a loaded photograph inside the system depend on the user's interactions and the different states the system assigns to the photo. We define the possible user-interactive objects inside the application as bookmarks (or viewpoint markers) and thumbnails (as were previously defined in Chapter 5). Consequently, as seen in Figure 6.10, we delimit five different **states in which an image can be specified** inside our system based on the different triggerings done by the current and prior **user interactions** (or by the system):

1. *Filtered:* the image has been filtered out by a timeline selection time frame, and the system does not allow **any interaction** for the user concerning the photo.

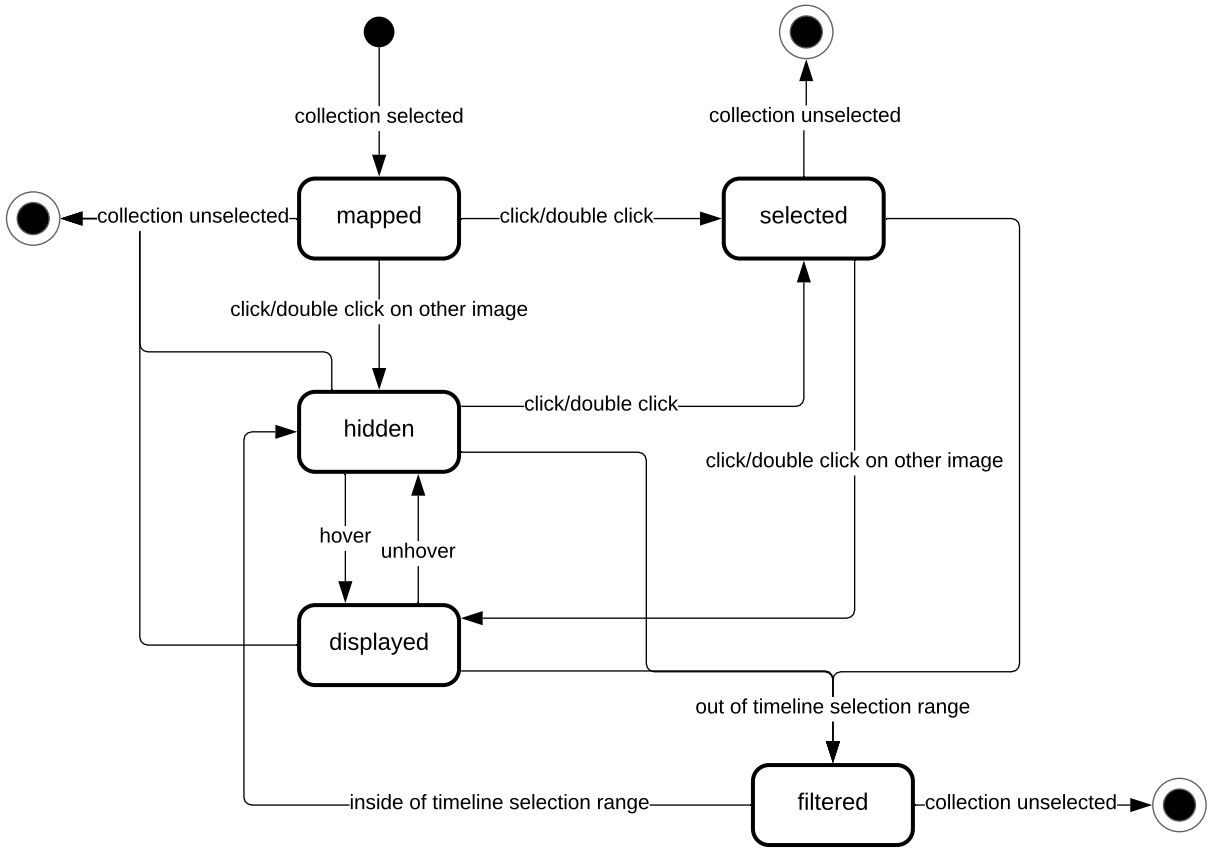


Figure 6.10: Image states and the possible triggering user interactions.

2. *Hidden:* **no image footprint** is **visible**, but interactive objects (viewpoint markers and thumbnails) can be seen in the view, and the user can interact with them to change the state of the image.
3. *Mapped:* the **image footprint** containing the **visibility heatmap** with/without an outline is shown to the user. Interactive objects can be seen in the view.
4. *Displayed:* the **image footprint** containing the **image content** with/without an outline is shown to the user. Interactive objects can be seen in the view. The system can **display more than one image**, where these can be either previously selected (click/double click) photographs or currently hover over (with the cursor) ones.
5. *Selected:* the **image footprint** containing the **image content** with/without an outline is shown to the user. Interactive objects and the photo's information can be seen in the view. The system considers only **one image selected**.

6.2.2 Model of a Historical Image

From our previous specification related to the interaction with an image inside the visualization system, we can now formally establish a historical image and its respective properties using the Unified Modeling Language (UML). Figure 6.11 showcases this definition as:

Class: Historical Image – depicts a graphical object, *i.e.*, a historical photo acquired in the real world. It is characterized by:

- *Camera:* models the optical device that captured the photograph. It is represented based on the camera's orientation information (intrinsics, extrinsics, and distortion model).
- *Texture:* the historical image loaded in the Graphics Processing Unit (GPU).
- *Date:* an approximate time frame where the photograph could have been taken.
- *State:* the current state of the image, *e.g.*, selected, hidden, or filtered.

Enumeration: Image State – determines the state in which an image is currently defined inside the system (see our previous definitions): *Filtered, Hidden, Mapped, Displayed, or Selected*.

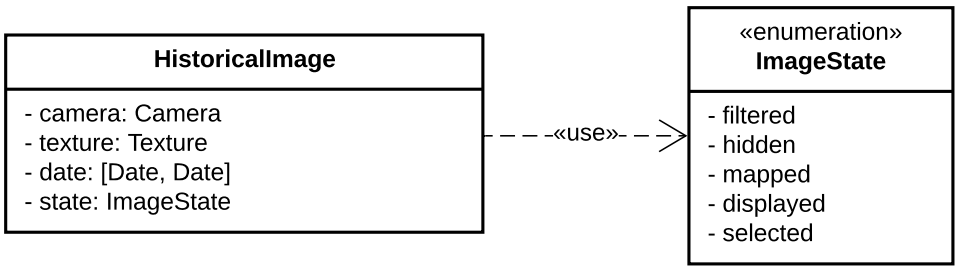


Figure 6.11: UML class diagram modeling a historical image.

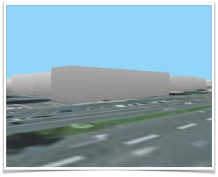
6.3 Architecture and Pipeline of HISTOVIS

For the ALEGORIA project, **we propose a method to visualize historical images inside an 3D environment composed of contemporary 3D city models**. We desire to provide the final users

with a spatio-temporal photo browsing system that allows continuous, interactive, and immersive exploration. We designed and implemented the HISTOVIS prototype with this objective. In this section, we present an overall description of the architecture and pipeline of our application.

Table 6.2 presents the definition of our **system's primary inputs**: *(i)* **2D historical photos** from either National Archives (AN), Nicéphore Niépce Museum (NNM), or French National Mapping Agency (IGN); *(ii)* a **3D city model** from IGN's BD ORTHO®, BD ALTI®, and BD TOPO®. We assume that each photograph has been spatialized and contains its camera orientation information and metadata describing it. We define the 3D model with a Level of Detail (LOD) 1, where the urban model is represented by the ground surface (through a Digital Surface Model (DSM)) that contains rectilinear buildings with a flat roof (rectangular meshes) without any particular detail.

Table 6.2: Inputs of the HISTOVIS prototype. Credits to Nicéphore Niépce Museum (NNM)/Charles Gross photographic collection and *itowns* for the example photos.

Input	Components	Temporality	Example
photograph	► **Photograph:** digitized (~10 - 100 Mpx) ► **Camera Orientation:** semi-automatic approach (≈ georeferenced with current data) ► **Metadata:** dated (~1 day - 10 years)	historical	
city model	► **3D Ground:** BD ALTI® (~1 m) ► **3D Buildings:** BD TOPO® (~1 m) ► **2D Textures:** BD ORTHO® (~10 cm)	modern	

Our approach supports the interaction between a user and the visualized historical images inside a 3D environment. We divide our **architecture into three main components** that allow a flow of communication between the user, the interface, and the system, where:

1. The available **geographic data** is organized temporally into historical (the photos) and contemporary (the city model).
2. The **system** provides support for the exploration of the historical images inside the 3D scene composed of the modern city models.
3. The **interface** enables the user to interact with the data inside the 3D environment.

Figure 6.12 shows an overall depiction of these three components combined inside our HISTOVIS prototype. Our web-based interface connects the user with a spatio-temporal visual exploration system. Based on the inputs for this system (*i.e.*, historical photos and 3D models) and the desired dynamics in the 3D environment focused on the connection of our propositions of Chapter 4 and Chapter 5, **we establish the corresponding objects to support our system**. Figure 6.13 illustrates them and their relationships through a class diagram using the Unified Modeling Language (UML). Besides the already defined *Historical Image*, among all our interacting objects, we define the main ones as: *Scene*, *Model*, *Material*, *Camera*, *Distortion Model*, *Cluster*, and *Timeline*.

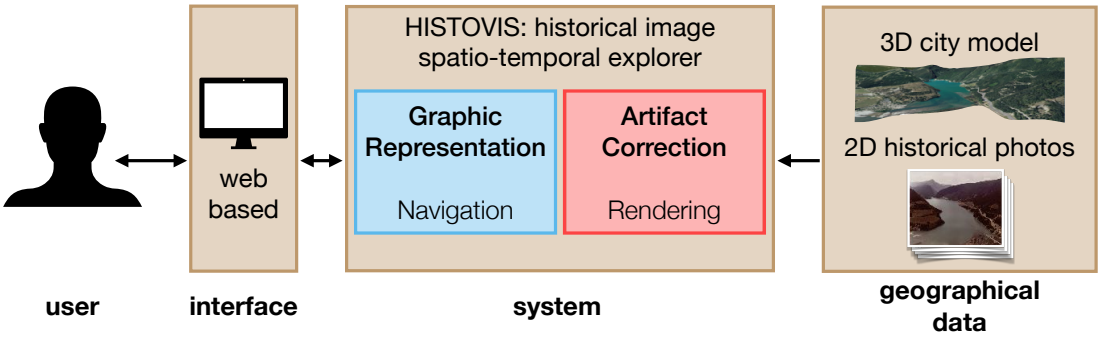


Figure 6.12: General architecture broken down into three main components: data, system, interface. Credits to National Archives (AN)/LAPIE photographic collection and *itowns* for the example images.

Class: Scene represents a 3D space that holds all digital objects. It contains:

- *Models:* all existent objects inside the scene.
- *Cameras:* optical devices capable of capturing a picture. It includes the virtual (for the resulting view) and the historical (for all photos) cameras. Notice that the geometry of these cameras (frustums) can be represented by 3D models inside the scene.

Class: Model describes geometrically any object using coordinates in a 3D space. It comprises:

- *Model Matrix:* a transformation matrix from the local coordinates of an object to coordinates in the 3D scene.
- *Geometries:* containing vertices, colors, normals, and texture mapping coordinate information of the model.
- *Materials & Styles:* properties that define the appearance of the object.

Class: Material determines the overall character and appearance of a model surface. It is computed in the Graphics Processing Unit (GPU) and is defined by:

- *Vertex Shader:* program that establishes where each coordinate of a model is located inside the 3D scene.
- *Fragment Shader:* program that "paints" each pixel rendered on the surface of the model.

Class: Camera captures images in a scene. A *Perspective Camera* follows a pinhole camera model. The *Photogrammetric Camera* is defined to be adapted to historical image data, *i.e.*, it can be represented by a camera model with distortion. It is composed by:

- *Position:* the 3D coordinates representing the location of the camera inside the scene.
- *View Matrix:* a transformation matrix from coordinates located in the scene to the local coordinates of the camera.
- *Projective Matrix:* a transformation matrix from 3D coordinates into 2D coordinates through the projection of these points.
- *Zoom:* value representing how close or further away the objects seem from the camera.
- *Near Plane:* a plane located at a particular close distance to the camera, along the optical axis, where all objects in the scene closer than the plane to the camera are not rendered.
- *Far Plane:* a plane positioned at a singular far distance to the camera, along the optical axis, where all objects in the scene farther than the plane are not rendered.
- *Marker:* a 3D volume (pyramid) truncated by the camera's near or far planes.

Class: Distortion Model a mathematical expression that corrects (in 2D image coordinates) the geometric deviation from the ideal pinhole camera model. It is defined by:

- *Boundary:* the geometric delimitation that depicts where the distortion function is no longer applied, and the extrapolation begins.
- *Distortion Function:* a function that describes the model of the distortion that a historical image has. For instance, it can be radial or fish eye.
- *Extrapolation Function:* the linear extrapolation of the distortion function.

Class: Cluster groups the historical images in a way where all objects in the same group are more similar (in some sense) to each other than those in other groups. It is defined by:

- *Tree:* the hierarchical binary tree that delineates how each image is related to the other.
- *Number of Images:* number of historical photos to be clustered.
- *Number of Clusters:* the number of desired clusters.

Class: Timeline lists the known time frames from where each historical image was capture. It is arranged in increasing order and depicted by a histogram representation through:

- *Histogram:* set of bars depicting each group of images captured in the represented period.
- *Segment:* the selected time segment to divide the time axis, e.g., weeks, months, or years.
- *Selection:* the time frame that has been selected for visualization. It is used to filter the historical images in the 3D scene.

We define a more hierarchical point of view through Figure 6.14, which displays our UML package diagram depicting high-level visibility of HISTOVIS. It divides the system's organization into the following abstract units:

Package: View deals with the 3D scene and its rendering process to generate the final view to the user. It contains the classes: *Renderer* and *Scene*.

Package: Objects describes the properties of each 3D element inside the digital scene. It contains the classes: *Model*, *Geometry*, *Material*, *Style*, and *Appearance*.

Package: Camera defines digital cameras used inside the 3D scene. Different camera models can be specified, but in HISTOVIS, we are interested in two: perspective (*i.e.*, pinhole) and perspective considering the distortion. It contains the classes: *Camera*, *Perspective Camera*, and *Photogrammetric Camera*.

Package: Distortion encapsulates the distortion model that can be used in the definition of a camera. Although the view camera only supports the *Radial* and *Radial Fish Eye* models, for the historical cameras, we add the options of *Tangential*, *Photogrammetric* (Radial + Tangential), and full *Fish Eye*. It contains the classes: *Distortion Model*, *Distortion Function*(s), *Extrapolation Function*, and *Circle*.

Package: Image describes abstract objects that represent historical photographs. It contains the classes: *Historical Image*, *Image State*, *Texture*, *Image Material*, and *Frame*.

Package: Clustering deals with the overall clustering (grouping) process of historical images. Different cluster methods exist, but HISTOVIS only supports a hierarchical approach using a binary tree. It contains the classes: *Cluster* and *Bi Nodes*.

Package: Temporal defines a time axis that allows temporal filtering of the historical images. It contains the classes: *Timeline* and *Histogram Bar*.

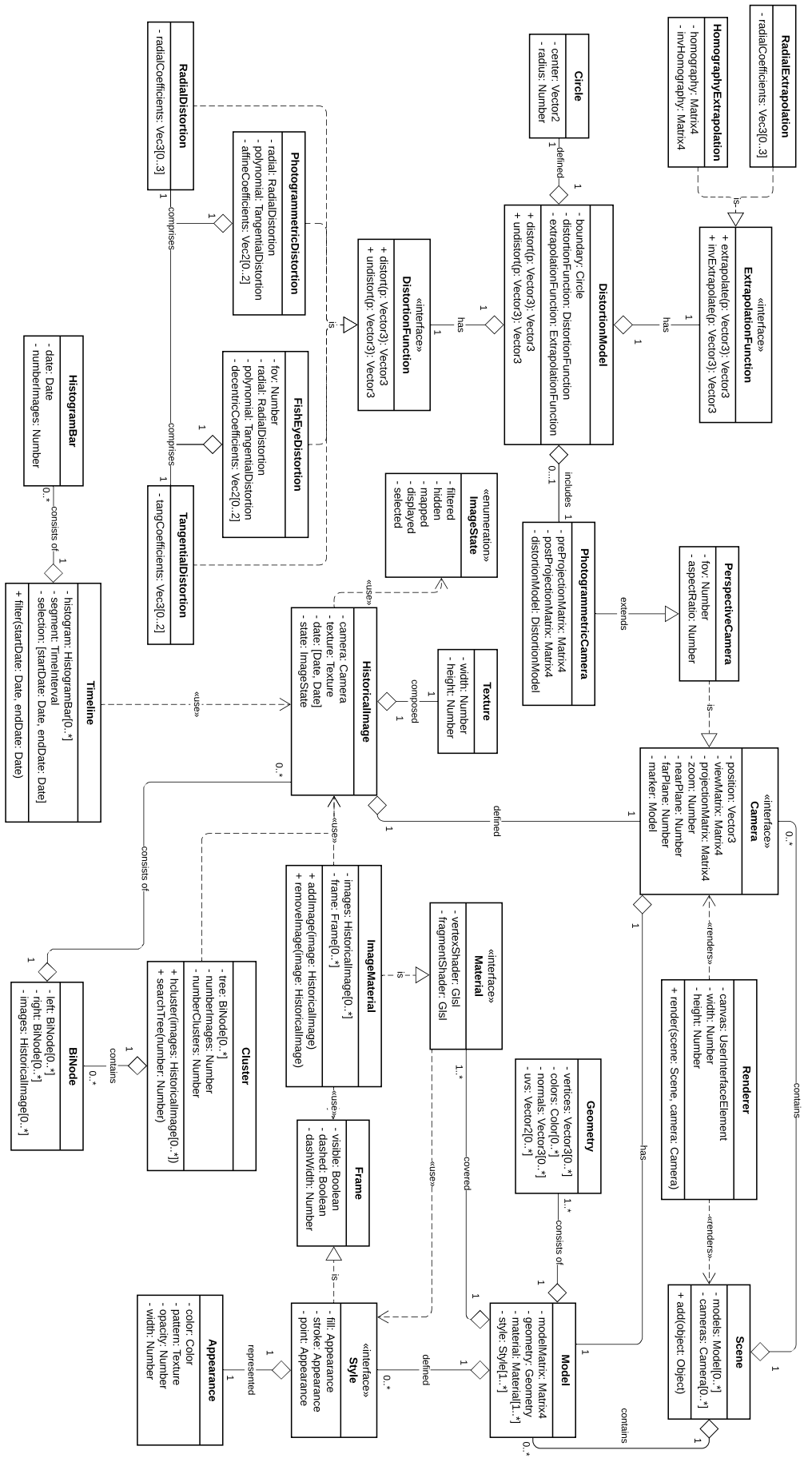
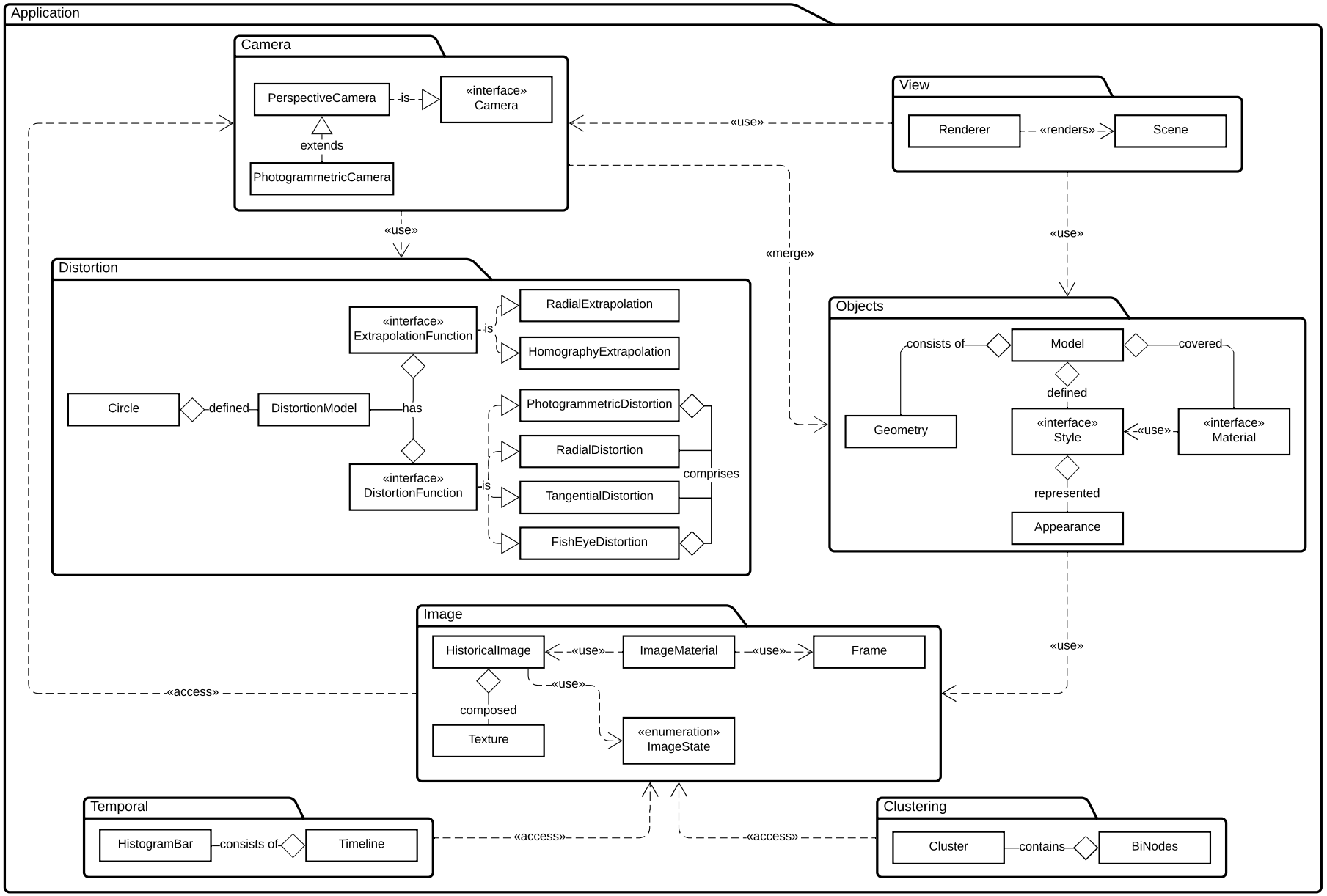


Figure 6.13: UML class diagram of HISTOVIS.

Figure 6.14: UML package diagram of HISTOVIS.



6.4 Usage of HISTOVIS

We developed HISTOVIS[9] using the *three.js*[10] WebGL rendering library, the *d3.js*[11] data-based library, and the *itowns*[12] geospatial data visualization framework based on our previously described structural design. This section will present our implemented system and its initial configuration, the corresponding user interface with diverse functionalities based on the propositions from Chapter 4 and Chapter 5, and three different use case scenarios using the ALEGORIA historical photo collections.

6.4.1 System and Initial Configuration

HISTOVIS is a web-based online prototype application accessible through any web browser on a desktop/laptop computer. When a user enters to its URL[13], a Graphical User Interface (GUI) is presented. As seen in Figure 6.15, in the center, we display our 3D environment containing the 3D model of a "Digital Earth." On the left side of this 3D view, the interface exhibits a list of the possible photo collections that can be explored inside HISTOVIS. The user can click and select (using a checkbox) the desired collection(s) to be loaded in the system. Figure 6.16 showcases this initial configuration with a UML object diagram, where it shows that HISTOVIS system's engine starts by only rendering a 3D scene containing a view camera a 3D model of the digital globe.

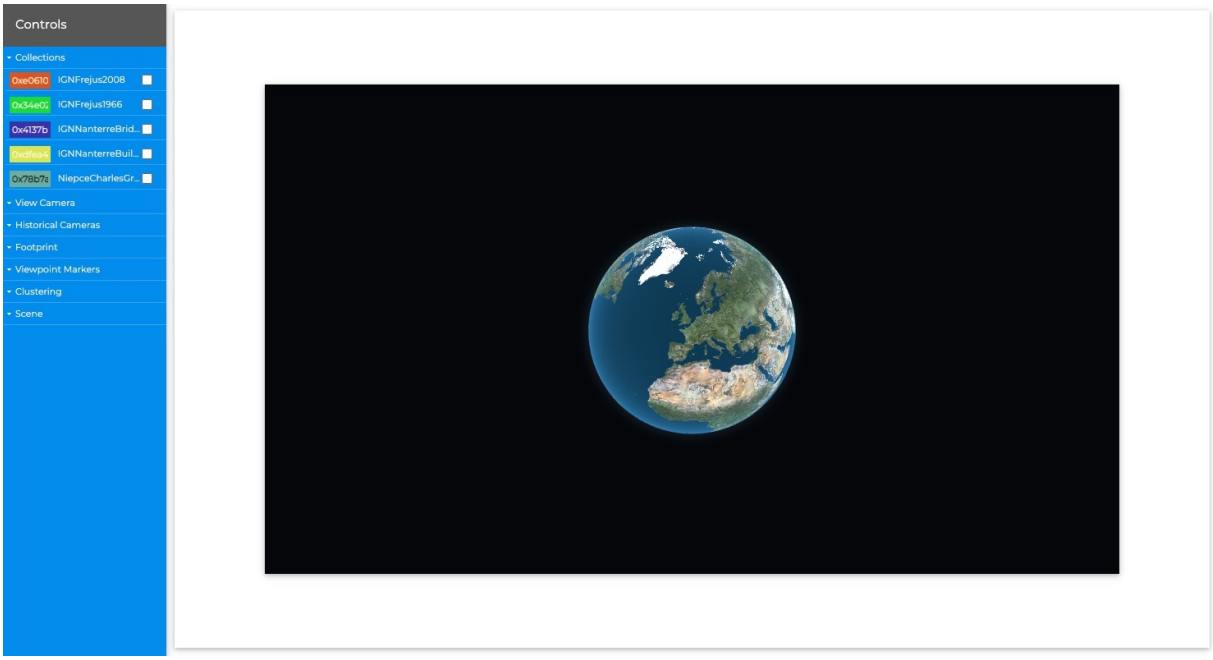


Figure 6.15: Graphical User Interface of HISTOVIS when the system has just been launched.

[9] https://github.com/epaizreyes/histovis
[10] https://threejs.org
[11] https://d3js.org
[12] http://www.itowns-project.org
[13] https://epaizreyes.github.io/histovis/examples/

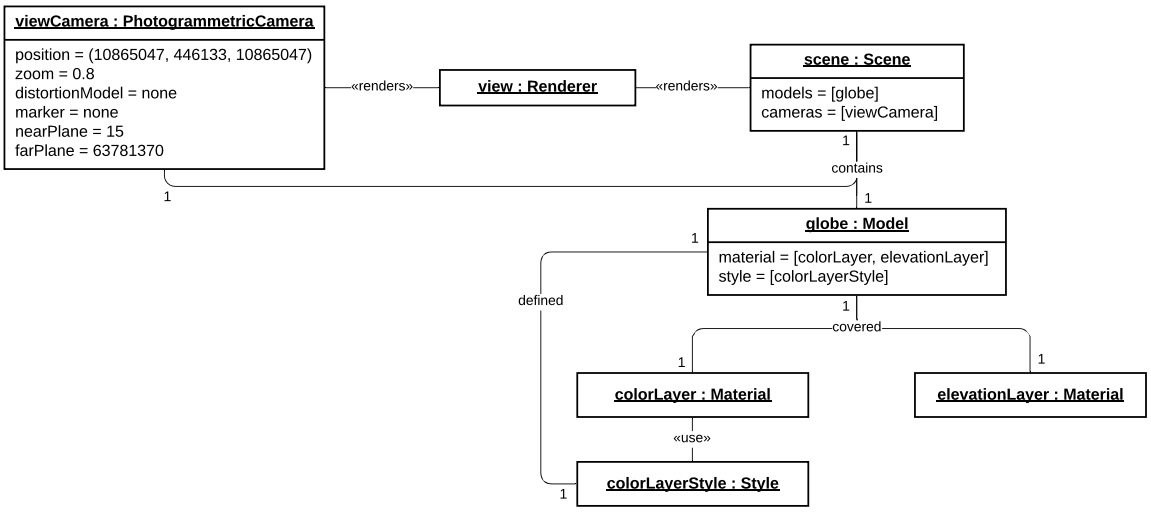


Figure 6.16: UML object diagram of HISTOVIS from the initial state. It depicts the configuration of the system's engine (presenting the 3D environment to the user in the GUI).

As a user selects and loads a photo collection(s), the internal system of HISTOVIS evolves from its initial state that only contains the view camera and the "Digital Earth" to one with a loaded dataset(s), where new objects, *e.g.*, historical images, clusters, and the timeline, are created based on each loaded photograph. Figure 6.17 illustrates the system's new engine configuration through another UML object diagram, where new objects are rendered in the scene, like historical cameras in the visible and interactive form of viewpoint markers (bookmarks).

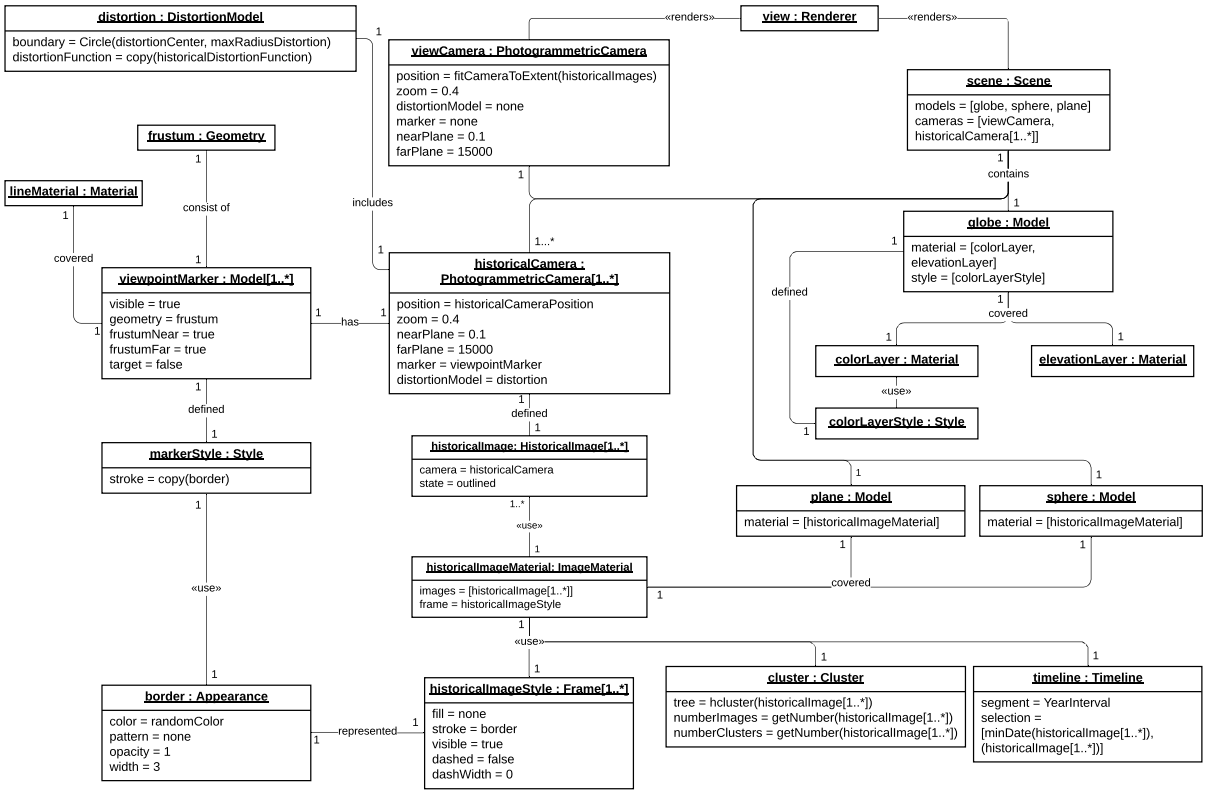


Figure 6.17: UML object diagram of HISTOVIS when a photo collection has been selected.

6.4.2 User Interface and Functionalities

Our HISTOVIS prototype is oriented towards visualizing and exploring historical images inside a contemporary 3D city model. The **user interface** of HISTOVIS **enables the interaction between the user and the system's engine**. Figure 6.18 presents our implemented interface with the different available interactive blocks: the central area (A), the border area (B), the controls menu (C), and the timeline (D).



Figure 6.18: Graphical User Interface of HISTOVIS: the 3D view (A), the thumbnails (B), the controls (C), and the timeline (D). Credits to IGN/Nanterre collection.

The central area (A) displays the overall 3D digital geographic environment. In it, the user can visualize the "Digital Earth" and its respective 3D city models (a Digital Surface Model and rectangular meshes depicting buildings) through the view camera. The 3D bookmarks representing the camera geometry of each historical image are also included inside this 3D environment. The user can adjust the view camera or interact with each bookmark using the following mouse actions:

- **View Camera:** interaction is plausible when the mouse's cursor is located in any position inside (A).
  - *Wheel:* zoom in and out of the view.
  - *Click and Drag:* camera rotation in the direction where the mouse is being dragged.
- **Bookmark:** interaction is possible only when the mouse's cursor is positioned on top of the bookmark.
  - *Hover:* projects the image temporally while the cursor is on the bookmark.
  - *Click:* projects/unprojects the image.
  - *Double Click:* projects the image and moves the view camera to the position of the historical camera represented by the bookmark.

The border area (B) presents the thumbnails (or clusters if the option selected) of the loaded photos that the timeline has not filtered out. The user can interact with them using the same mouse actions defined for the bookmarks.

The controls menu (C) enables the user to load a photo collection in the system and change the default configuration of each system's respective functionalities. We will explain these more in detail in the upcoming subsections, but overall the main configurations are:

- **Collections:** allows the user to select and load a photo collection. Note that more than one can be chosen, and a specific color represents each one (modifiable by the user).
- **View Camera & Historical Cameras:** configures the model of the view camera or all the historical cameras parameters, *e.g.*, zoom, far plane, distortion model applied.
- **Footprint:** manages the content presented in the footprints of all historical images that have been projected. It allows the user to show the overall image content, the heat map, and change/add a frame or blurriness (on the border) to these footprints.
- **ViewPoint Markers:** permits modifying the style of all bookmarks.
- **Clustering:** establishes if the system should cluster the historical images and how many clusters this data should be split in.
- **Scene:** updates the 3D city models' style and colors.

The timeline (D) illustrates the temporal information known from the loaded photo collections. Additionally, it allows a user to filter the available historical images to be visualized inside the 3D environment by date of acquisition. It supports the corresponding mouse actions:

- **Histogram Bars:** interaction is plausible when the mouse's cursor is inside the histogram.
  - *Click:* when the cursor is on a specific histogram bar, it selects only the time frame represented by that bar.
  - *Wheel:* zooms in and out of the histogram by changing the time frames that each bar represents.
- **Handles:** interaction is possible only when the mouse's cursor is positioned on top of a selected time range handle.
  - *Click and Drag:* changes the selected time frame in the timeline.

Based on the proposed methods from Chapter 4 and Chapter 5, we have designed and implemented the system's corresponding **functionalities** that **allow the interactive visual exploration of historical photos**. We divide our functionalities into three main groups: *(i)* interaction with the historical photos; *(ii)* adding saliency on the elements inside our 3D space; *(iii)* the temporal filtering of the photographs.

Functionality 1: Interacting with Historical Photos

We define our system's first functionality as the one in charge of **displaying the photographs inside the 3D city model to the user**. We establish its subcomponents as:

Viewpoint Markers: (or bookmarks) **show where a picture was taken** and the **camera's orientation**. Inside the 3D environment, they can provide an overview of what was the perspective of the photographer. In HISTOVIS, we use them as an interactive object for the user

to project an image on top of the 3D models or move the view camera to the corresponding historical camera represented by the bookmark. As previously mentioned, we allow these interactions through mouse actions (*hovering*, *click*, and *double click*). Figure 6.19 shows how the user can modify these 3D markers for the following attributes:

- **Design:** we allow the user to select the best-desired design based on the corresponding needs. For instance, a targeting ray can show a fundamental idea of where the image may be projected, an extended frustum depicts precisely where the footprint is located, and a simple symbol with no geometry only gives an abstract view.
- **Visual Features:** the general visual characteristics of the bookmarks can be updated:
  - *Color:* is defined by the color assigned to the photo collection that the viewpoint marker belongs. It can be updated, but it will change the color of all the viewpoint markers, image frames, and thumbnail frames of each photograph in that specific collection.
  - *Width:* the line width of the 3D geometry (representing the frustum or target ray) can be increased or reduced.

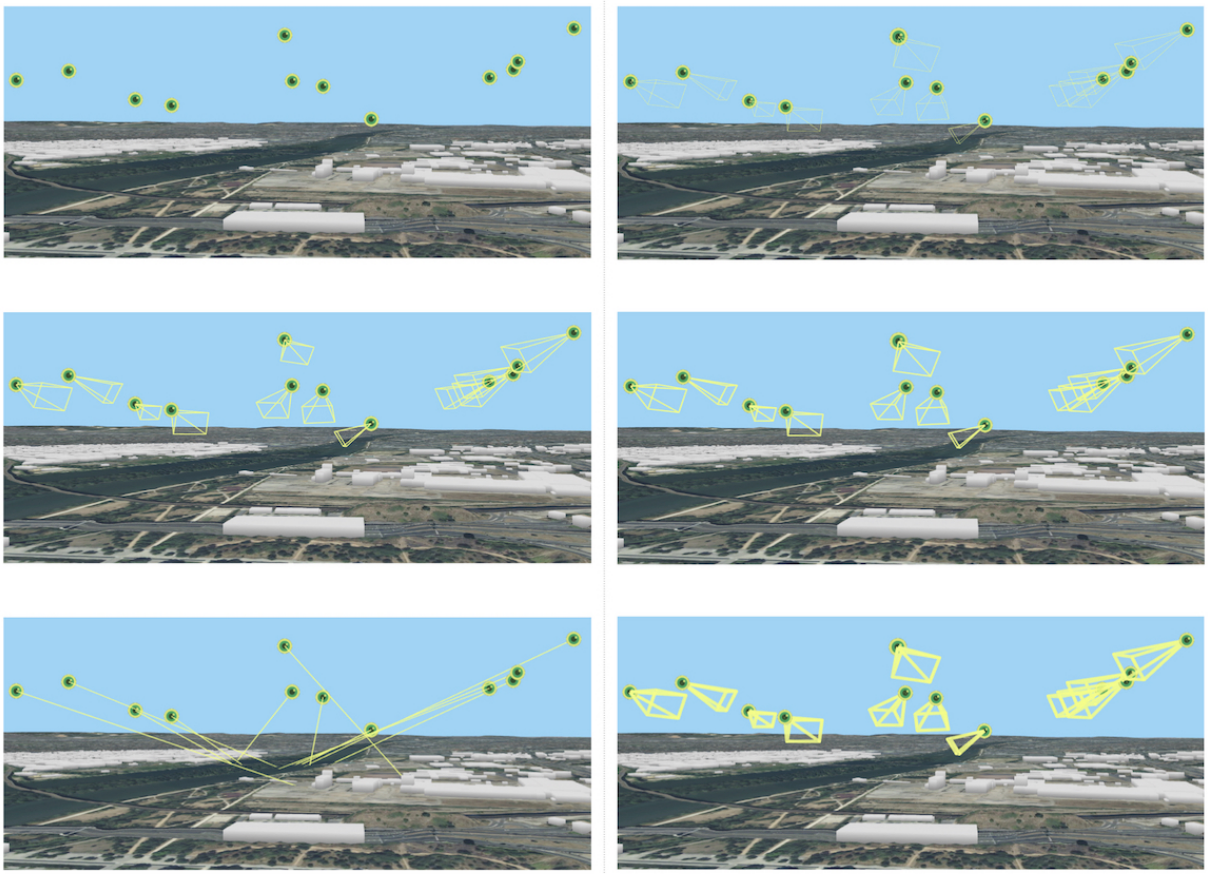


Figure 6.19: Viewpoint markers inside HISTOVIS. **Left:** three different display designs: only the icon (position), a small transparent frustum, and a ray target to the center of projection. **Right:** the width of the frustum lines are changed to a pixel rate of 1, 4, 10, respectively.

Thumbnails: are a **small representation of a photograph** that has been loaded into the system. As bookmarks, they are considered interactive objects related to the projection of historical images on top of the 3D models. We allow these interactions through the same mouse actions (*hovering*, *click*, and *double click*). To connect the corresponding viewpoint marker located in the 3D view and the thumbnail positioned outside it, we use the visual variable of size. Both of these objects change size when the user is interacting with either of them. Additionally, an outline is placed around the thumbnails with the identifying color to distinguish each thumbnail's photo collection. Figure 6.20 shows a sample placement of the thumbnails of photographs that have been spatialized; however, we divide the location of the thumbnails into two primary ways:

- **Border of the View:** all thumbnails of loaded images (that have not been filtered temporally) are displayed on the borders/sides of the 3D view. The placement of these is dependent on the movement of the view camera, *i.e.*, as this one moves, it also does the thumbnails. Additionally, to reduce the number of thumbnails, the option of clustering is available. It groups the historical photos based on the distance from the projected center points. The user can select how many clusters should be displayed. We use a gallery/clustering visualization to showcase each cluster and its items.
- **Menu Slider:** to include historical photos from other photo collections that have not been spatialized (*i.e.*, the orientation information of the camera is not known), we add the option of displaying them in a menu slider. For each photograph selected (clicked), the list of non-spatialized pictures updates based on an image-based retrieval request (only used and not implemented by this dissertation).



Figure 6.20: Thumbnails around the 3D view of HISTOVIS. Credits to NNM/Charles Gross collection.

Distortion on Photograph and View: when the distortion model of a historical camera is known, we model this camera considering it. In a basic configuration, the user will visualize the projected rectified version of each historical photo. The distortion sub-functionality allows the user to **select a particular image** and **use its distortion model in the definition of the view camera**. For the photographs in the ALEGORIA project, we do not know the distortion model of the photos. However, we include the functionality for future use when the orientation information of the cameras can be extended. As pictured in Figure 6.21 (using the Viabon dataset), we provide to the user the following features:

- **Distortion on View Camera:** decides if the system should model the view camera with the distortion model of a particular historical image. If it is not available, then the view camera is represented through a pinhole camera model.
- **Distortion on Historical Cameras:** allows selecting if the system should use a camera model that considers each photo's distortion (different for each historical camera).
- **Extrapolation Radius:** the user can configure the extrapolation radius (of the distortion model) using a factoring value of 0 to 1, where 1 represents the minimum radius surrounding the four corners of the image.

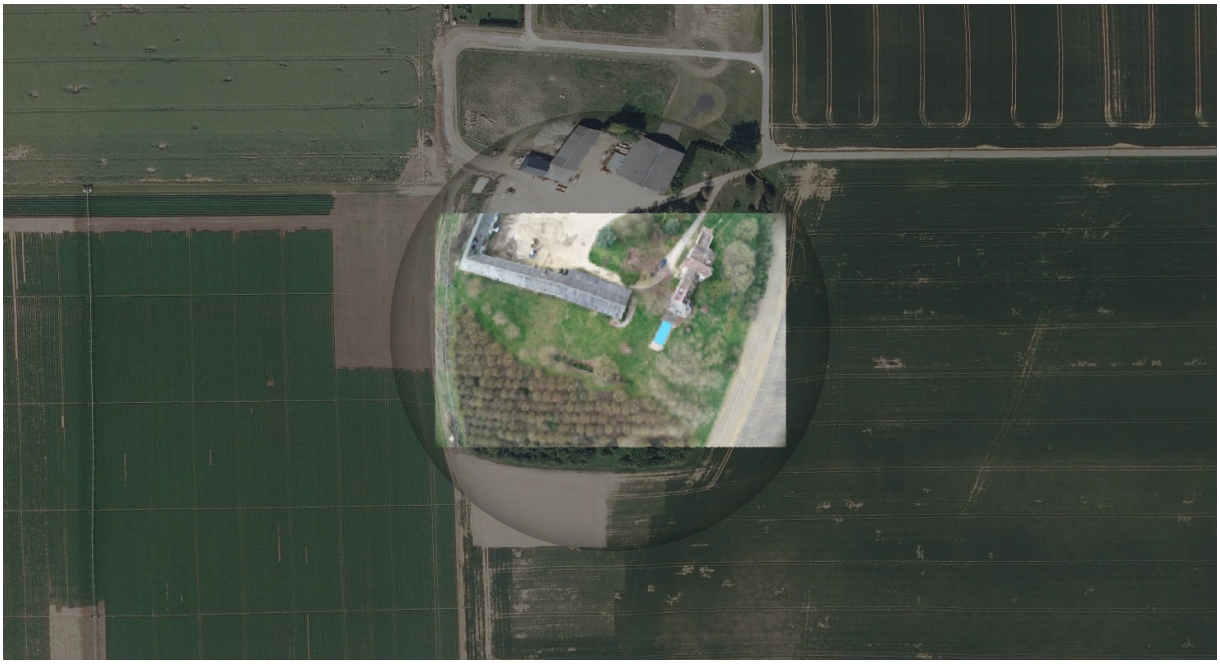


Figure 6.21: Distortion model applied on the view and historical camera inside HISTOVIS. The system automatically chooses the extrapolation radius around the photo. The black shading is only for illustration purposes.

Functionality 2: Adding Saliency on Elements Inside a 3D Space

The system's second functionality relates to the concept of **saliency towards specific elements inside our 3D digital space**. We define its subcomponents as:

Image Frames: (or outlines) accentuate where the photos are projected inside the 3D scene and **contrast the photographs with the** (possible) **background** that the 3D city models may present. Figure 6.22 displays how the user can add/remove these frames and which are the visual configurations that HISTOVIS provides:

- **Display:** the user is capable of defining if the image footprints should be outlined and how this framing should be displayed:
  - *None:* hides the frame and only displays in the footprint the content of the projected images.
  - *Around the Overall Projection:* considers the footprint of all projected photos as one and only displays an outline around this overall projection.
  - *Around Each Image:* sets a frame around each projected photograph.
- **Visual Features:** the global visual characteristics of the frame can be updated:
  - *Color:* is determined by the color assigned to the photo collection that the image belongs. It can be changed, but it will update the color of all the viewpoint markers, image frames, and thumbnail frames of each photograph in that specific collection.
  - *Thickness:* the width of the frame can be increased or reduced.

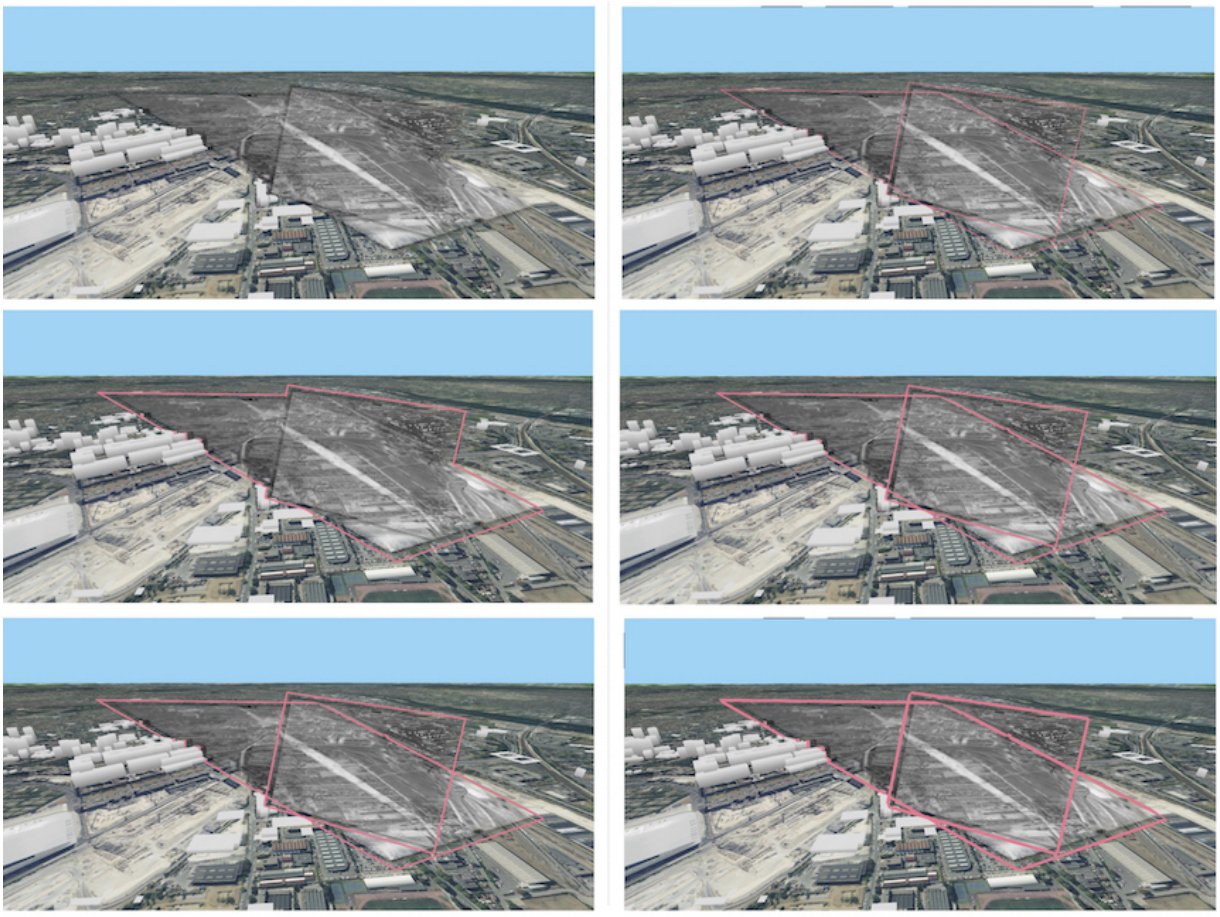


Figure 6.22: Frames around the projected photographs (blended) in HISTOVIS. **Left:** three different display designs: none, around the overall projection, and around each image. **Right:** the width of the frame line is changed to a pixel rate of 2, 5, and 10, respectively. Credits to IGN/Nanterre collection.

Visibility Heat Map: provides a **quick spatial overview of where all the historical photographs** that have been loaded into the system **are located**. In HISTOVIS, it is activated each time the number of pictures updates. For instance, when a new dataset is loaded, or the user temporally filters the photos with the timeline, the view camera moves to a position where the extent of the heat map is visible. As shown in Figure 6.23, to represent the number of photographs projected in a specific area of the 3D city model, we use an RGB (red - green-blue) color game, where red represents the maximum (all loaded that are not filtered out) amount of available images and blue the minimum (one).

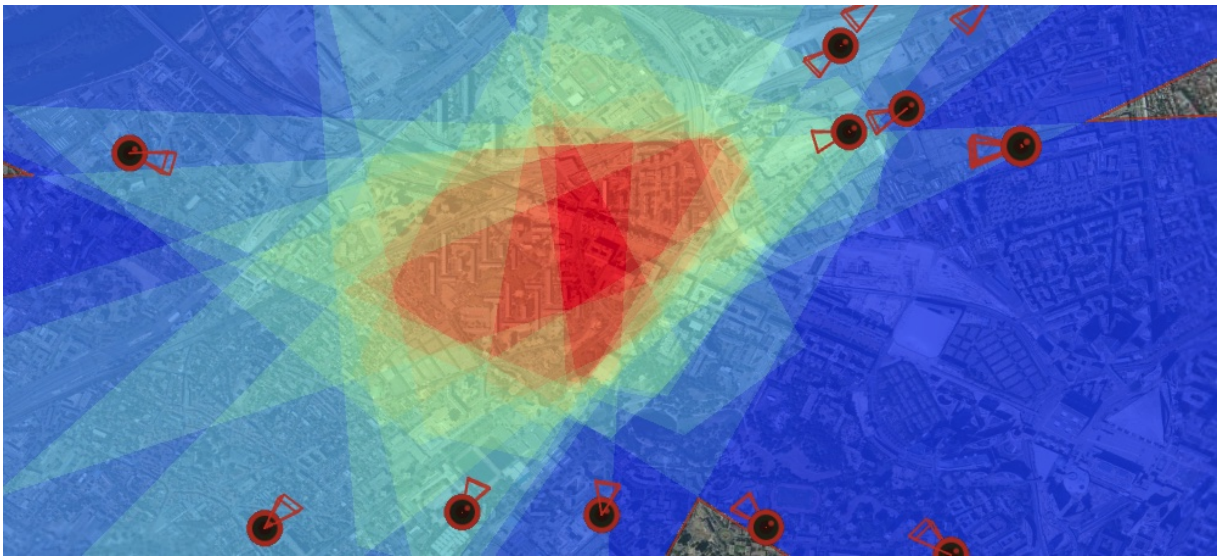


Figure 6.23: A projected heat map on top of a 3D city model inside HISTOVIS. It depicts the areas where there is more concentration of photographs.

Style Represenatino: we define the 3D city models as the context where the photographs are being visualized. As default configuration, HISTOVIS utilizes the orthoimages from IGN's BD ORTHO® to texturized the Digital Surface Model depicting the city ground. However, as illustrated by Figure 6.24, to **differentiate or homogenize the projected images** with their context, we allow the user to update two of its main characteristics:

- **Style:** the user can vary the style of the DSM as:
  - *Abstract:* no texturing on the model, only a gray color is used instead.
  - *Ortho:* uses orthoimages (recently acquired data) to texture the model.
  - *Open Street Map:* use the open street map geographic syle for texturing.
- **Radiometry:** from the currently selected style of the DSM, the user can update its color scheme as:
  - *Original:* no change, the fundamental version.
  - *Monochrome:* transforms it to a black and white version.
  - *Sepia:* converts it to a sepia version.
  - *Saturated:* makes the colors more saturated than usual.

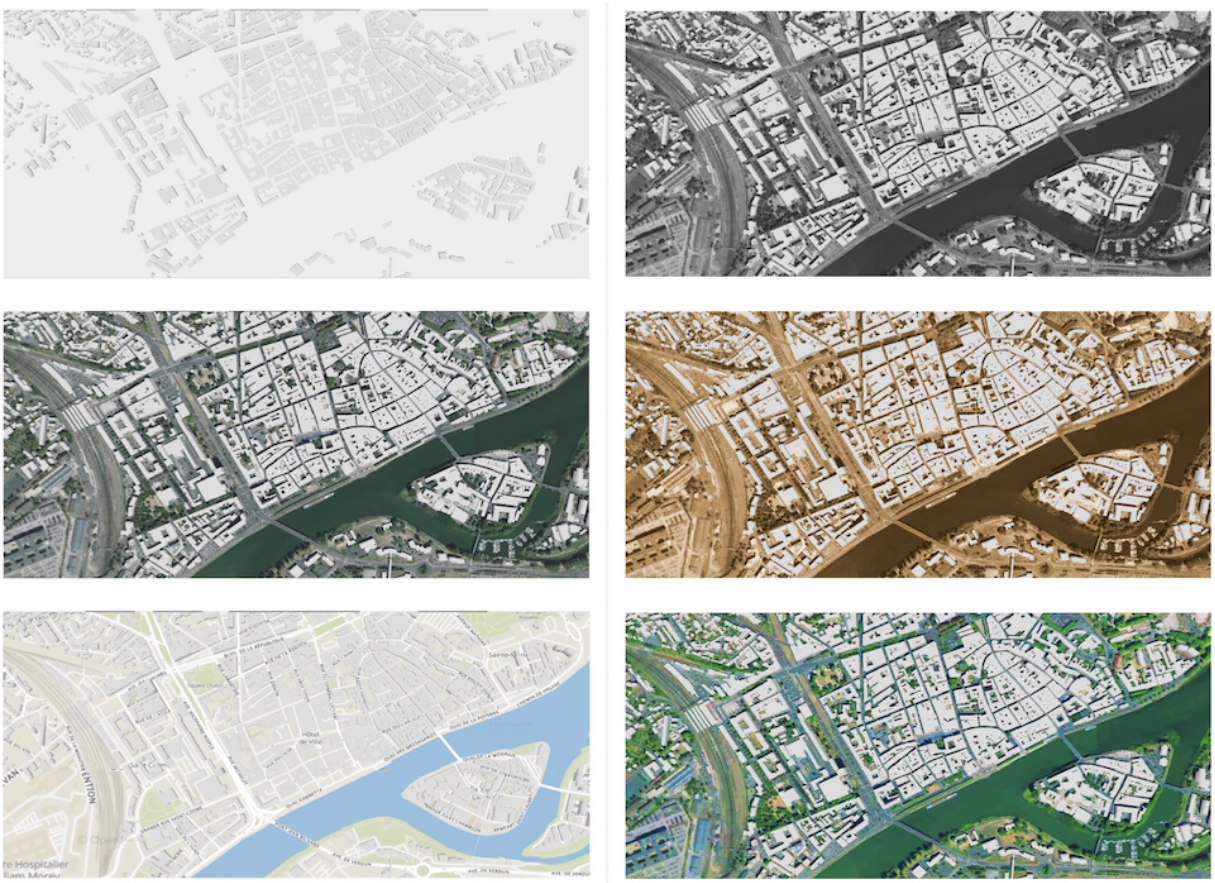


Figure 6.24: Style representation inside HISTOVIS. **Left:** three different display styles: abstract, textured with orthoimages, and textured with open street map. **Right:** three different color schemes: monochrome, sepia, and saturated.

Functionality 3: Filtering through Time

The last functionality of the system is defined by the **temporal information known from the historical photographs**, *i.e.*, the date of acquisition (or an approximate range of dates). We outline its main component as:

Timeline: is the **graphical representation** implemented in HISTOVIS to **depict the periods represented by the historical images**. Figure 6.25 show the two primary uses for it:

- **Uncertainty Representation:** because a specific date where each historical photo was taken may not be known, the system represents dates in time frames, *e.g.*, years, months, or weeks. It is done through a histogram representation where each bar depicts the number of photos that may have been captured during that specific time range.
- **Filtering:** when a selected time range has been fixed or changed, only the historical images acquired during this period are available for user interaction (remain projected in the scene, the bookmarks and thumbnails are visible). The heat map sub-functionality is activated to showcase this change, *i.e.*, the view camera is moved, and the extent covered by the photos is shown.



Figure 6.25: A timeline next to 3D view in HISTOVIS. Only five photos (1946) are shown from the filtered selection. Credits to IGN/Frejus collection.

6.4.3 Use Case Scenarios

In this section, we exemplify the use and application of HISTOVIS with ALEGORIA photo collections. We have chosen **three specific use case scenarios in which the spectrum of the usefulness of our proposed method** can be illustrated for: *(i)* the visualization of multiple images and collections; *(ii)* the temporal exploration of geographic elements; *(iii)* the browsing of non-spatialized photographs in 3D digital spaces.

Table 6.3: Overall description of experimental datasets for HISTOVIS prototype. Credits to NNM and IGN.

Name	Point of View	Dates	No. Images	Example
Frejus	aerial (vertical)	1966 – 2008	102	
Nanterre	aerial (oblique)	1946 – 1960	40	
Charles Gros	terrestrial	1975 – 2003	110	

Notice that the collections from ALEGORIA are only partially spatialized. As represented in Table 6.3, we have chosen (and georeferenced) only a sub-set of the photo collections that can showcase most functionalities of our prototype and agree with the ALEGORIA social research objectives (Delavoipière, Conord, and Marshall 2020; Lecat et al. 2021).

Scenario 1: Visualization of Two Photo Collections

Our first use case scenario depicts a **cross-reference analysis and visualization of two different photo collections from the same geographical region**. Figure 6.26 graphically illustrates it with the Frejus photo collection (aerial vertical shots) from 1966 (black and white) and 2008 (color).

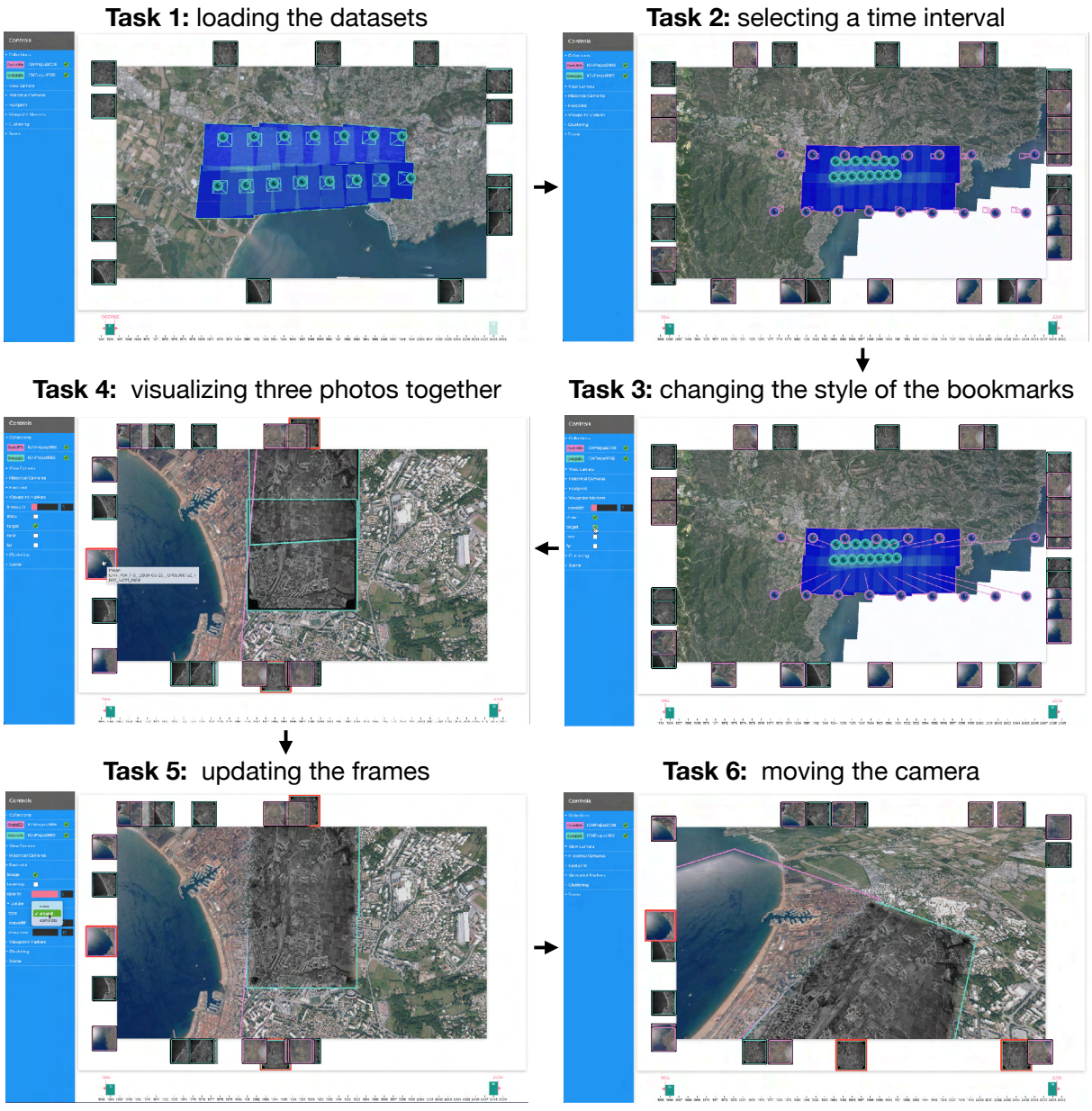


Figure 6.26: Use case scenario 1: visualization of two photo collections. Credits to IGN/Frejus collection.

We define the interactions that the user and the system have through the GUI as follows:

Task 1: Loading the Datasets – the user selects the photo collection of `IGNFrejus2008`. The system loads all the corresponding photographs and projects them on top of the 3D city models in the area depicted by these images (Frejus in Côte d'Azur). Instead of displaying the content of each image, the system shows on the projected footprint a heat map delineating the regions with the highest amount of projected photos. Finally, the system moves (with a transition) the view camera to a new position where the extent of the heat map can be visualized. The process is repeated for the photo collection of `IGNFrejus1966`. The system now filters the photos and displays only those belonging to the newly selected dataset.

Task 2: Selecting a Time Interval – the user moves the mouse cursor to the bottom part of the interface where the timeline is located. The user then defines a new "selected" time interval from 1966 – 2008 by dragging the selection handles. The system repeats the same process followed in Task 1 and displays the heat map of the two datasets.

Task 3: Changing the Style of the Bookmarks – the user moves the mouse's cursor to the left position of the interface where the controls are located. In the `Viewpoint Markers` (bookmarks) subcontrols, the user selects the target style and unselects the default options with a frustum (considering the near and far plane). The system updates all the corresponding bookmarks in the view by removing the frustums and displaying the target ray instead.

Task 4: Visualizing Three Photographs Together – the user double clicks a bookmark from the photo collection of 1966. The system removes the projected heat map and projects only the selected image. Then, it moves the view camera to the position of the historical camera of the selected photograph. The resulting visualization is a pixel-accurate projection. The user then clicks one thumbnail from 1966 and another from 2008. The system does not move the view camera but only projects the corresponding images when a thumbnail is clicked.

Task 5: Updating the Frame of the Images – the user moves the mouse's cursor to the left position of the interface where the controls are placed. In the `Footprint/border` subcontrols, the user: *(i)* reduces the sharpness of the border to 13; *(ii)* increases the line width of the frame to 5; *(iii)* changes the type of display to around the projection. The system updates the visualization with these new representations.

Task 6: Moving the Camera – the user clicks (and holds) inside the 3D view to move the camera by dragging the cursor down. The camera rotates, and the markers become visible. The system updates the new orientation of the camera and, based on it, computes the new positions of each thumbnail located on the border/sides of the view. The user then zooms out the view camera to visualize more context provided by the 3D city models.

Scenario 2: Street-View Exploration Through Time

Our second use case scenario describes a **temporal exploration process from a street-view photo collection, where the evolution of the photographed scene can be depicted**. Figure 6.27 graphically demonstrates the overall process using the Charles Gros photo collection (terrestrial views) representing the area of Chalon-sur-Saône in the region of Bourgogne-Franche-Comté from 1975 to 2003.

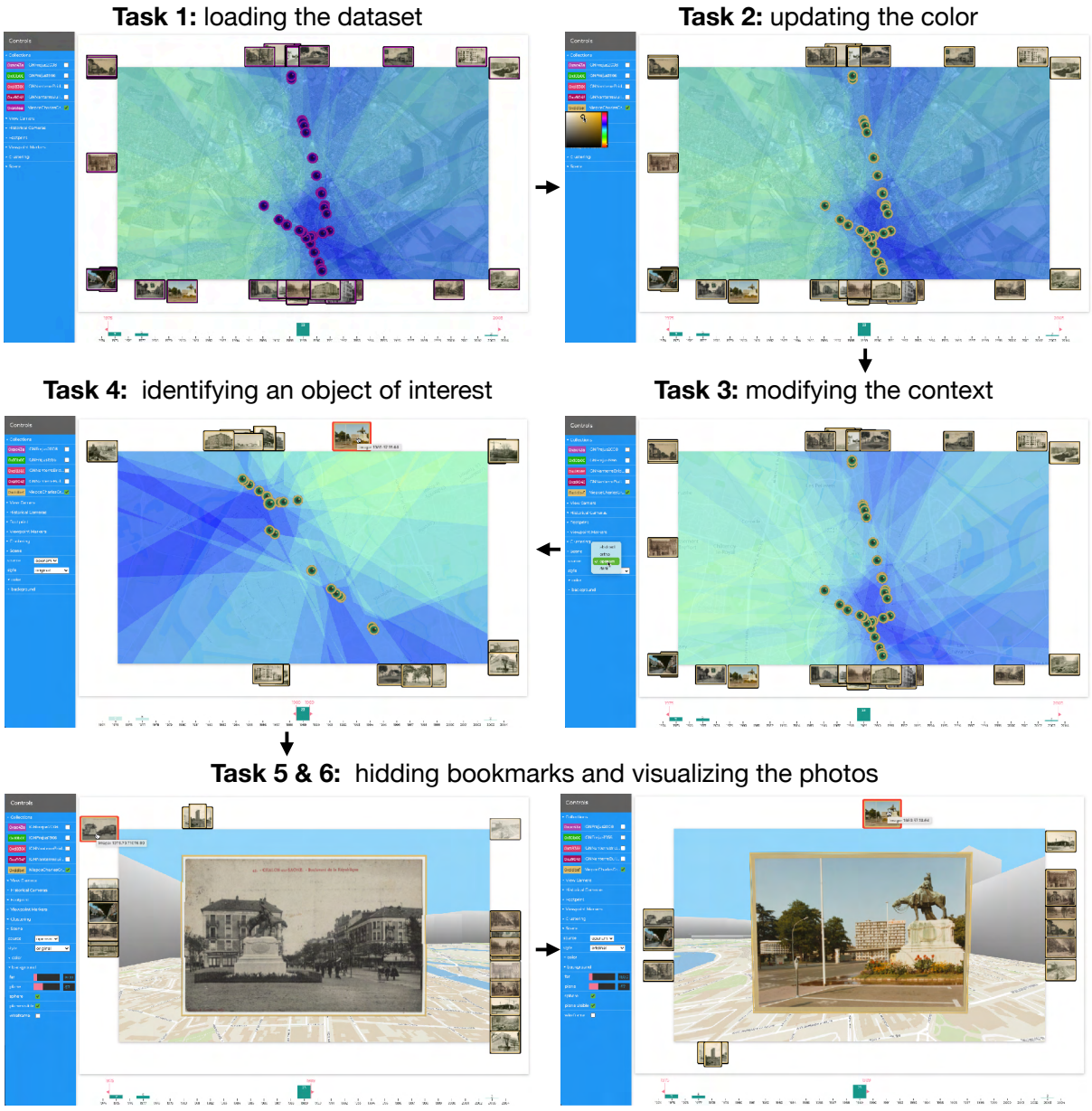


Figure 6.27: Use case scenario 2: street-view exploration through time. Credits to NNM/Charles Gros collection.

We determine the interactions that the user and the system have through the GUI as follows:

Task 1: Loading the Dataset – the user selects the photo collection of `NiepceCharlesGros`. As in Task 1 from Scenario 1, the system loads all the corresponding images. It projects on top of the 3D city models the footprint of the heat map depicting the regions with the highest amount of projected pictures. The system then moves the view camera to a new position where the extent of the heat map can be visualized.

Task 2: Updating the Color – the user does not like the color representing the selected collection. Therefore, the user moves the mouse cursor to the left position of the interface where the controls are located. Next to the name of the loaded dataset (on the left side), the user selects a yellow color instead. The system updates the viewpoint markers, the frames, and the outline of the thumbnails with this new color. The user visualizes the new color on all these elements.

Task 3: Modifying the Context – the user keeps the cursor in the left position of the interface but moves the attention to the `Scene` subcontrols. The user changes the data source used to texture the Digital Surface Model (DSM) to `Open Street Map`. The system updates the view with the new style for the DSM.

Task 4: Identifying an Object of Interest – the user moves the cursor to the bottom part of the interface where the time axis is located. The user then defines a new "selected" time interval to 1975 by dragging the selection handles. The system repeats the same process followed in Task 1 and displays the heat map of images captured in that year. The user then identifies a monument of interest in the displayed thumbnails. The process is repeated for 1977, 1989, and 2003, where in in 1989, the same monument is recognized. Finally, the user defines the "selected" time range from 1975 to 1989. One more time, the system updates the view.

Task 3: Hidding the Bookmarks – the user now decides to hide the bookmarks and moves the cursor to the `Viewpoint Markers` subcontrols. The user unselects the show option. The system hides all the icons representing these markers. The user does not see them anymore inside the 3D view.

Task 5: Visualizing Two Photographs Separately – the user double clicks the thumbnail of the photograph depicting the monument in 1975. The system removes the projected heat map and projects only the selected photograph. Then, it moves the view camera to the position of the historical camera of the selected photograph. The resulting visualization is a pixel-accurate projection. After seeing the picture, the user unselects the photograph by clicking the thumbnail. The system unprojects it. The user now double clicks the thumbnail of the photograph depicting the monument in 1989. The system projects the new image and moves the view camera to the position of its historical camera. The user observes the temporal evolution of the monument's area.

Scenario 3: Browsing Non-Spatialized Photos in a 3D Spatial Environment

Our last use case scenario demonstrates **how photographs that have been spatialized can be used to explore other photo collections with non-spatialized pictures**. Figure 6.26 graphically shows it with the Nanterre photo collection (aerial oblique acquisitions) from 1946 to 1960 (black and white). To incorporate other collections, we use the content-based image retrieval method from ALEGORIA (Gominski, Gouet-Brunet, and Liming Chen 2021; Gominski, Poreba, et al. 2019). The resulting queries return non-spatialized photographs from the Lapie, Carto MRU, Combier, and IGN BD ORTHO® collections. We determine the interactions that the user and the system have through the GUI as follows:

Task 1: Loading the Dataset – the user selects the photo collection of `INGNanterre`. As in Task 1 of Scenario 1, the system loads all the corresponding photographs and moves the view camera to a new position where the extent of the heat map can be visualized.

Task 2: Clustering the Spatialized Photos – the user decides to cluster the thumbnails, so moves the mouse cursor to the left side of the interface. In the `Clustering` subcontrols, the user selects 9 clusters. The system updates the thumbnails based on the information provided by the hierarchical clustering process. The user visualizes the clusters in a gallery representation.

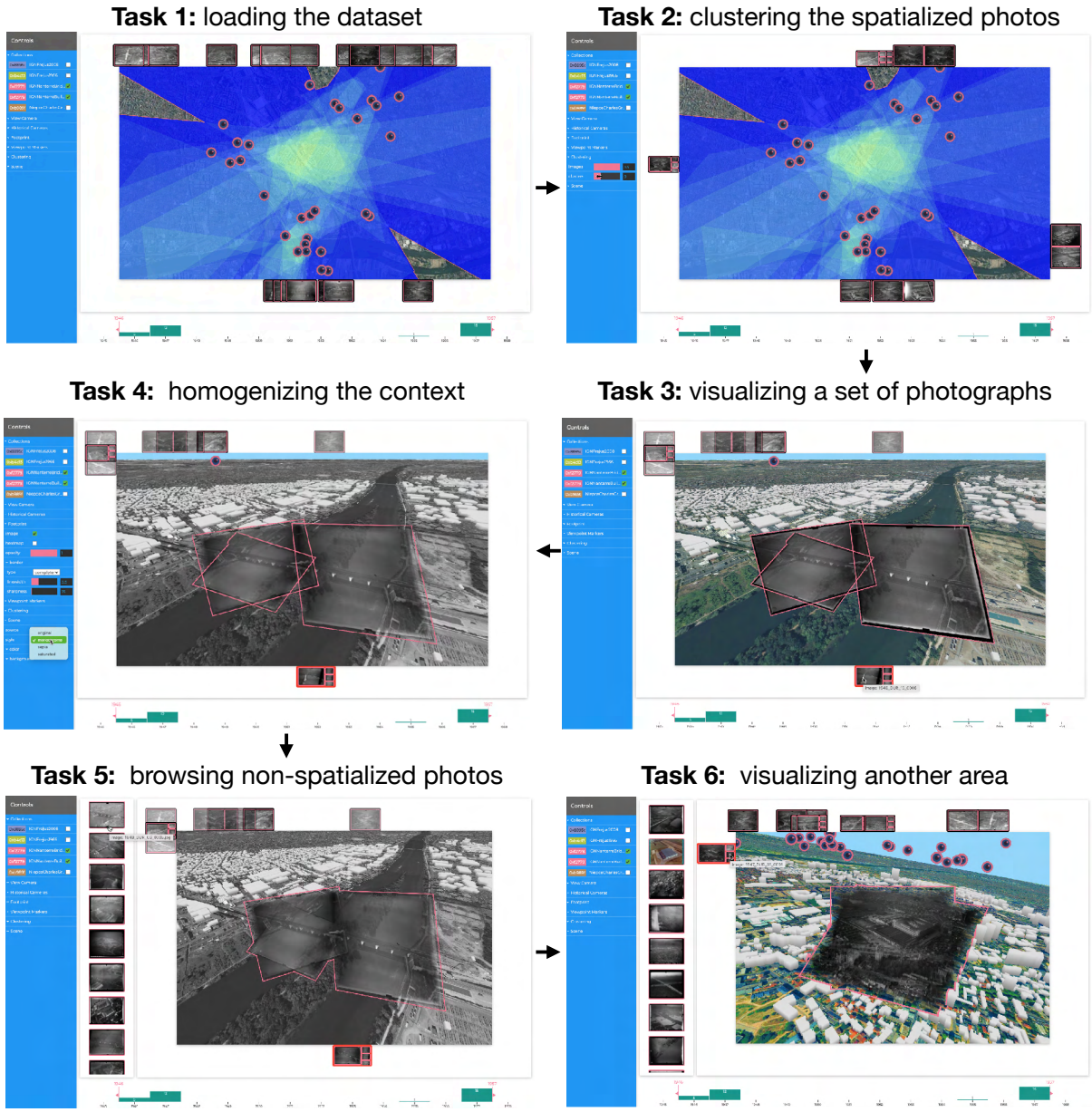


Figure 6.28: Use case scenario 3: browsing non-spatialized photos in a 3D spatial envirionment. Credits to IGN/Nanterre collection.

Task 3: Visualizing a Set of Photographs – the user double clicks a thumbnail from the photo depicting the object (a bridge) of interest. The system removes the projected heat map and projects only the selected photograph. Then, the view camera is moved to the position of the historical camera (of the photo). The user adds two more images by clicking their spatialized thumbnails.

Task 4: Homogenizing the Context – the user moves the mouse cursor to the left position of the interface (the controls). In the `Scene` subcontrols, the user selects a monochrome style to homogenize the photo's context. The system updates the Digital Surface Model. Additionally, the user goes to the `Footprint/border` subcontrols, reducing the border sharpness to 13 and placing the image frame around the projected area.

Task 5: Browsing through Non-Spatialized Photos – the user now moves the mouse cursor to the

lower corner, in to the controls, and clicks the image icon. The system requests the list of all the photographs (including other photo collections) that can present the same content as the selected images in Task 3. A lateral bar contains a menu slider with the corresponding thumbnails of this list. The user scrolls over the lists and checks each thumbnail.

Task 6: Visualizing Another Area – – the user unselects (by clicking the thumbnails) the currently displayed images and selects (by double-clicking) two photos from another cluster depicting a different area in the scene. The system moves the view camera to this region and updates the non-spatialized images. To differentiate the images with the 3D scene, the user changes the radiometry of the DSM again to saturated. The system updates the view.

6.5 Evaluation

In order to **validate our proposed HISTOVIS prototype**, we evaluated its use in two parts. Firstly, we performed a user test of our primary described use case scenario: the visualization of two different photo collections. Along with it, during its design and implementation, we had iterative feedback from the stakeholders from the ALEGORIA project every six months. This section presents our findings from the two separate evaluation outcomes to the reader. Overall, we aim to discover if our prototype is:

- Accessible to every user, but precisely: archivists and humanities researchers.
- Capable of effectively provide continuous browsing and visualization of historical photographs inside a 3D environment.
- Effective on reducing the potential visual complexity provided by the system's large and heterogeneous geospatial data displayed.

6.5.1 User Study on Scenario 1: Visualization of Two Photo Collections

Our first evaluation approach considered **testing one of our proposed use case scenarios** described in the previous Section 6.4. We selected the first scenario (1) since it depicts the most fundamental use case of HISTOVIS: the visualization of historical images within a 3D environment. More specifically, a cross-reference of two photo collections. The study was performed in an early prototype version, so not all the functionalities of HISTOVIS were implemented; only the following were assessed: *(i)* timeline; *(ii)* bookmarks; *(iii)* image frames; *(iv)* thumbnails and clustering.

We established a test protocol, and the evaluation was carried on the 3rd of February 2021. The online survey form presented to the users, along with the descriptions of the tasks, is available in Appendix B. Nine users experimented with the HISTOVIS prototype in an average time of 37 min. The evaluation was completed online using a web browser on a desktop computer. We performed the study using the Nanterre photo collection. Each user answered a total of 21 questions, of which 11 were compulsory. The evaluation results are in Appendix C; Figure 6.29, Figure 6.30, Figure 6.31, and Figure 6.32 illustrate them graphically. This section presents the analysis of these results.

Evaluation 1: Temporal Selection

Summary of the Task

The user is asked to do two temporal selections using the timeline. The first one with a broader range, *e.g.*, in years from 1945 to 1947, and the second with a smaller one, *e.g.*, from October 1945 to April 1946. For it, the user is required to zoom in and out inside the time axis.

1. Inside the temporal selection functionality, we tested the possibility of zooming in and out of the histogram. For instance, the time axis could present the time intervals in years, and the user could zoom in and reduce them to weeks. However, 44% of the test users felt that the histogram transition (zoom in/out) was discontinuous. A possible solution mentioned was to use a "pan" option instead of the zoom. Therefore, with every mouse scroll, the user could pan between zooms for perfect smoothness. On the other hand, to improve the transitions from big histogram bars to smaller ones, these can stay to relocate when zooming, and only the smaller bins could appear.
2. In relationship to the design of the timeline using a histogram representation. 78% of the test users understood it quickly. They mentioned that each bar represented the number of images per time range. However, the number label was missing; therefore, in our final prototype, we added this label.

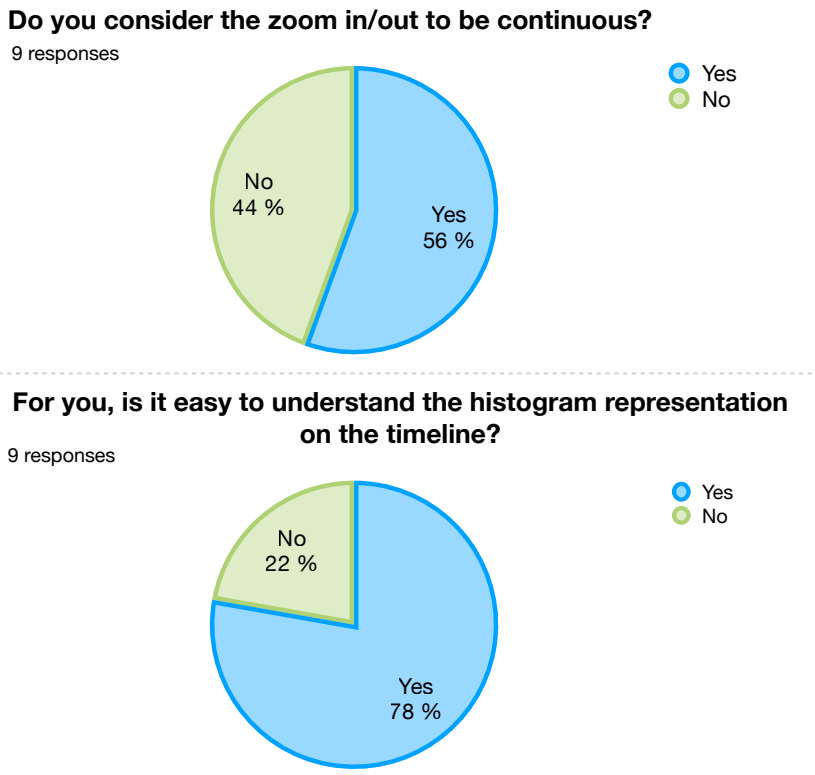


Figure 6.29: Graphical results on the temporal selection task.

Evaluation 2: Interaction with Bookmarks

Summary of the Task

The user is requested to interact inside the 3D scene with the bookmarks representing the viewpoints of loaded images in the system. It is asked to perform the following mouse actions:

- *Hover:* in and out to see a photograph projected on top of the 3D models.
- *Click:* to permanently project the picture.
- *Double Click:* to move the view camera to the viewpoint of the photo.

Additionally, the user is challenged to find the preferred style displayed for the bookmarks: target, near frustum, and far frustum.

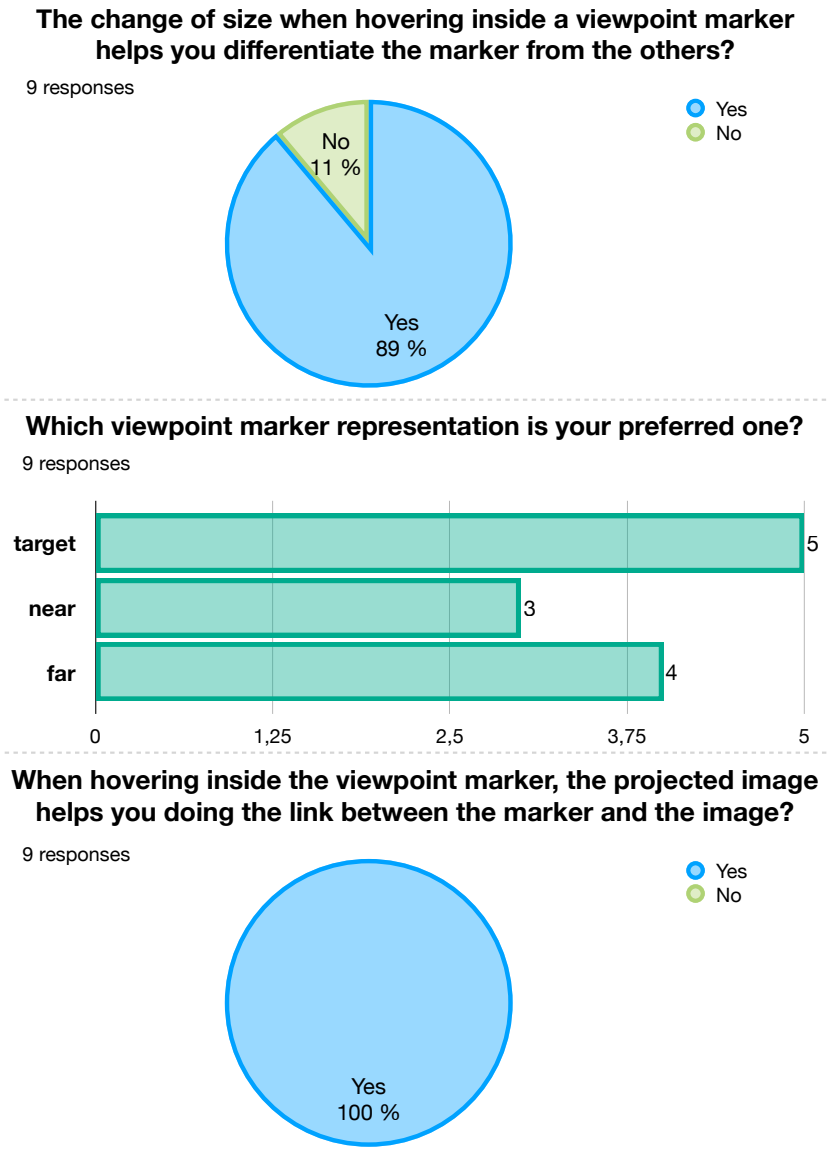


Figure 6.30: Graphical results on the bookmark interaction task.

1. One visual variable tested in the bookmarks was the changing size when the user interacts with the object. 89% of the test users consider it useful when hovering over a bookmark and identify which one is selected. Nevertheless, one issue mentioned was the fact that far bookmarks were too small and not visible. To solve it, we added a constraint in HISTOVIS to avoid bookmark icons using the scene's depth information to define its size.
2. Regarding the style of the bookmarks, 55% favored the ray target and 44% the far frustum representation. Still, for images with a terrestrial or high oblique view, we do not consider it the best option. For instance, a user mentioned that some rays were leading away into the horizon, and it became difficult to find to which marker the lines belonged. We suggest limiting its use for vertical and low oblique aerial photographs. Therefore, for terrestrial and aerial high oblique, we consider the near frustum as a more suitable option. On another note, the change of line width and opacity was requested to reduce or improve the saliency of these markers.
3. The projection of the photographs when hovering was also discussed. 100% of the test users agreed that projecting the photograph on the 3D models helped them associate a specific bookmark with the corresponding photograph.

Evaluation 3: Use of Color for Differentiation

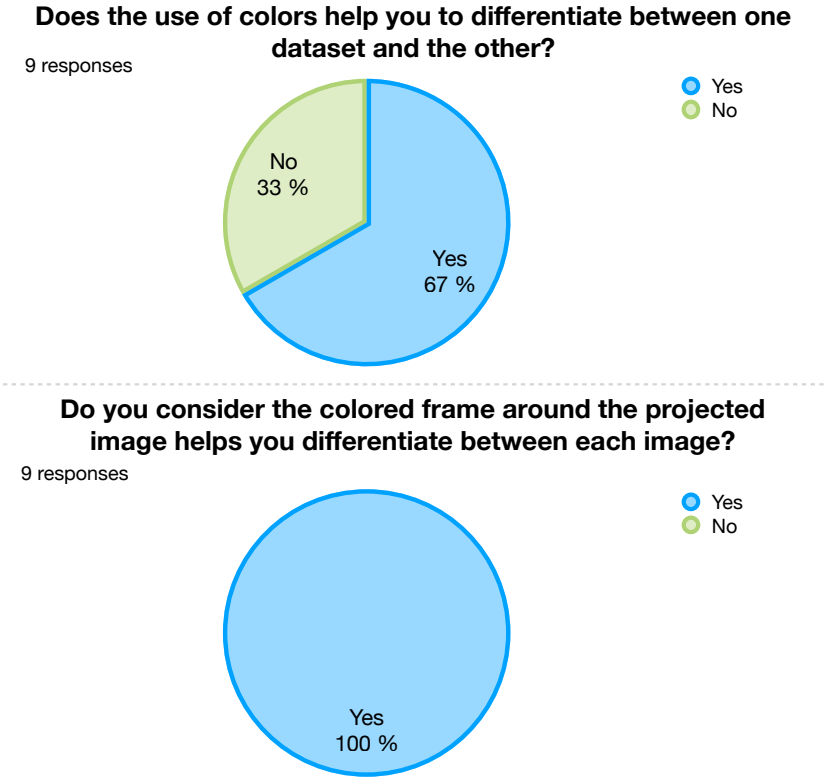


Figure 6.31: Graphical results on the color differentiation task.

Summary of the Task

The user is invited to manage the visual variable of color (hue) to differentiate between images from two loaded photo collections. It is requested to change the colors of each one and project two images in the scene (one from each collection). The evaluation is done concerning the outlines (frames) surrounding the pictures and the corresponding bookmarks.

1. 67% of the test users found that color helped differentiate between two separate loaded photo collections. Nevertheless, users noticed that in the bookmarks, the icon did not apply this color scheme. We evaluated the option and decided to add it to HISTOVIS since the icon is the most visible component of the 3D marker.
2. The use of frames proved to be a way for users to differentiate between projected photographs, as 100% of the test users agreed with the notion.
3. Concerning the loading of a new dataset, it was stated by the test users that nothing in the 3D view was happening to let them know that the photos had been loaded and their spatial location in the scene. We evaluated possible solutions and decided to implement the additional functionality of the heat map. We use it each time a collection is loaded, providing the spatial information of the data through a global view from the top.

Evaluation 4: Clustering of Spatialized Thumbnails

Summary of the Task

The user is directed to add the thumbnails of the loaded photographs around the border of the view. It is asked to perform the same mouse actions on the thumbnails as in task 2 related to the bookmarks (*Hover*, *Click*, *Double Click*). Additionally, the user has to cluster these thumbnails into smaller groups until they reach only one cluster.

1. The movement of the thumbnails around the view was rated continuous by 78% of the test users. Despite this, some users expressed a lack of fluidity. They considered that it took too long to travel from one position to the another for some thumbnails, while others had already stopped moving. To solved the issue in HISTOVIS, we reduced the timing of movement. Though, it is still an open topic since it could be tackled through different approaches. For instance, the movement could be removed, and the thumbnails could appear and disappear in the corresponding positions.
2. One matter worth mentioning is the positioning of the thumbnails around the border of the view. 67% of the test users do not think it is helpful to predict where a photograph could be projected. Overall, the users declared that it was not easy to understand the position the clustering is suggesting. One user considered it would help better associate the thumbnails to the bookmarks inside the 3D scene instead of correlating them to the projection of the photographs. Another solution that we propose is the use of leader lines when hovering over a thumbnail. It could also help the user to make the connection to the 3D view.

3. The gallery representation when clustering was the last feature evaluated. 56% of the test users did not found the design easy to understand. The primary concern was the image that is more prominent than the others. Various users mentioned that they understood it, along with the clustering and the predictions of the positions around the border only after a few tries inside the system. Therefore, we think this functionality is one of the most complex ones proposed among all.

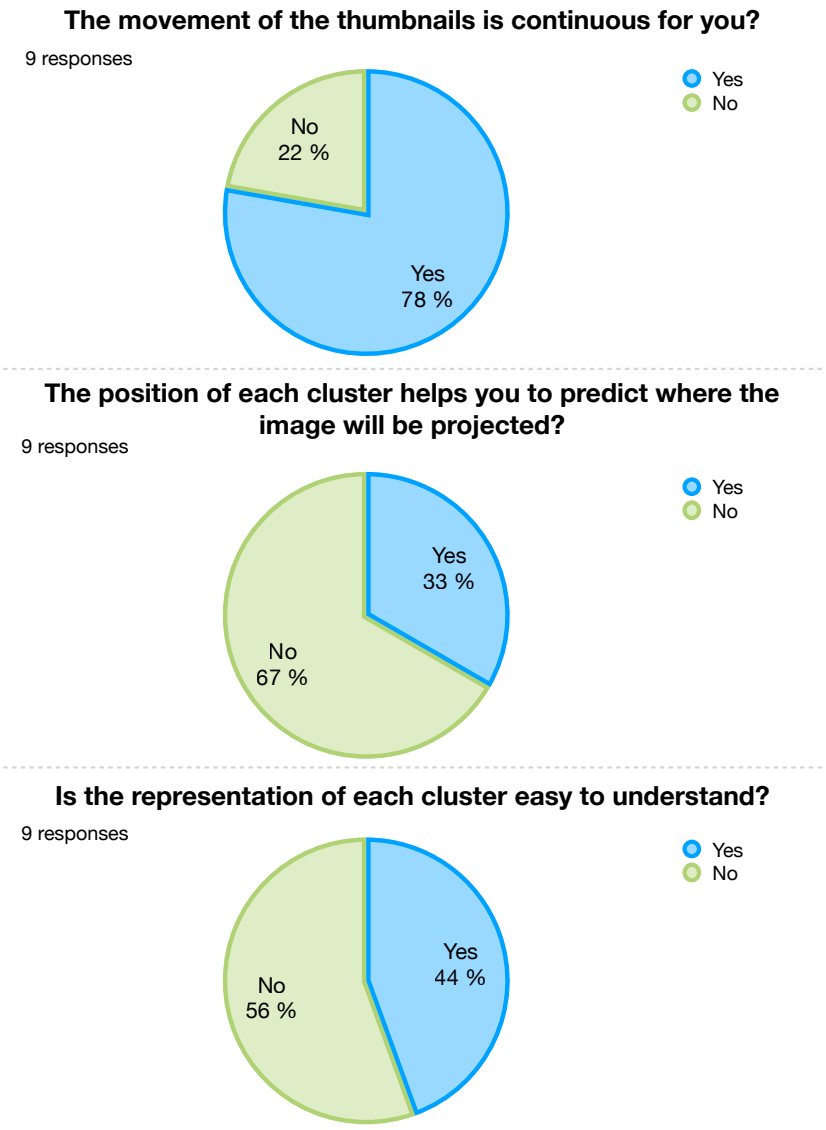


Figure 6.32: Graphical results on the clustering spatialized thumbnail task.

6.5.2 ALEGORIA Stakeholders Feedback

As a whole, the ALEGORIA project focuses on the indexing, interlinking, and visualization of the iconographic collections from the partner institutions (AN, NNM, and IGN). The work of the project aims to design two innovative online platforms:

1. Multimodal and large-scale indexing and search engine. It allows browsing by content and by metadata inside the same collection or through multiple collections.
2. The rendering engine for spatio-temporal, immersive, and interactive navigation in the 3D environment enriched with historical photographs.

This dissertation proposes a prototype (HISTOVIS) for the second objective platform (the rendering engine). To **evaluate our designs** and make sure they stayed in tune with the objectives of the project, **we attended five of the ALEGORIA consortium meetings**, which took place:

- 22nd of January 2019 (Nicéphore Niépce Museum)
- 6th – 7th of June 2019 (Centrale Lyon)
- 16th – 17th of January 2020 (Univ. Nanterre)
- 17th of December 2020 (video conference)
- 28th – 29th of June 2021 (French National Mapping Agency)

The following is the analysis of the general remarks and requests from the participating stakeholders during these gatherings:

1. The partners pointed out that the photo collections' spatial understanding (3D navigation) is only a part of the visualization feature we provide to users. It was requested to add the time variable as an additional interaction function in the system. Based on this demand, we added the functionality of the temporal selection through a chronological slider. Our first designs were oriented towards using small thumbnails or dots in the time axis to represent the photographs; nevertheless, we simplified it, in the end, applying a simplistic histogram representation.
2. The partners mentioned that primary users are interested in comparing historical photographs across time and source (photo collections). Therefore, we added to HISTOVIS the option of loading multiple datasets at a time to cross-reference collections. The system allows users to co-visualize images from different collections and compare them visually (*e.g.*, by switching transparency).
3. A clear visual distinction between loaded photo collections was asked. The visual variable of color was considered and used to provide this feature to the users. Hence in the system, we color the bookmarks entirely, outline the thumbnails, and frame the projected photographs.
4. The partners asked for the possibility of having an overview of a collection in the 3D scene. Firstly, we introduced to HISTOVIS the bookmarks to signal the position and orientation in which the photograph was capture. Later, we added the heat map functionality to show users the geographic extent depicted by the content of a photo collection.
5. The partners accepted the idea of spatialized thumbnails and clustering. They requested an extension by adding not oriented images of other photo collections to the HISTOVIS system. We established the previously defined scenario 3: browsing non-spatialized photos in a 3D spatial environment. We separated the positioning of thumbnails into two categories: *(i)* border of the view with a continuous transition dependent on the view camera for oriented ones; *(ii)* a carrousel (menu slider) for the non oriented ones but related to the currently selected photograph (even from other collections). We manage to link the two objective platforms of the ALEGORIA project (search and rendering engine) with it.

6.6 Conclusion

Subsection 6.5.1 (user study on scenario 1: the visualization of two photo collections) enabled us to observe the use of our HISTOVIS prototype by different users. We identified that indeed the system allows users to visually explore historical photographs inside contemporary 3D city models by meeting their primary expectations. Most users also noted the usefulness of the different proposed techniques for the saliency of elements inside the scene. Overall, they seemed to have fun using the prototype and were intrigued by the different design propositions.

Certain users' observations remain to be improved, primarily to support beginner users. For instance, we would like to enhance the comprehension of the thumbnail placement for spatialized photographs. Since this design does not follow the conventional approach, it can be somehow confusing for some users. One of the main points mentioned by users was its movement around the border of the view. During our experiment, several test subjects commented that they understood its use after a couple of tries inside the system. Even though an easy solution could be to change the position of the thumbnails to a static representation, *e.g.*, as we do for non-spatialized photos, we think it just needs more continuity for the thumbnails and a better association to the objects inside the 3D view.

Subsection 6.5.2 (the ALEGORIA stakeholders feedback) presents the outcomes from our iterative evaluation approach. We tried to improve our proposed functionalities and designs depending on all the input comments and feedback during the five gatherings with the ALEGORIA consortium. We observed that throughout the design of our proposition, they liked the idea of the image browser using the digital 3D environment. They responded the most positive towards the idea of the framing of the images and the use of projective texturing in the bookmarks. In addition, their request allowed us to add new functionalities; for instance, the use case scenario 3 (the browsing of non-spatialized photos in a 3D spatial environment) was designed along with them.

Contribution

In summary, this work introduces the HISTOVIS prototype, a web-based online application containing our implemented propositions from Chapter 4 and Chapter 5 oriented towards a Cultural Heritage context. It encapsulates a system capable of enhancing historical images by placing them inside a 3D environment containing contemporary city models. More specifically, it **provides to users a tool for the visual exploration of photographs using an interactive approach through space and time** by presenting the following three main functionalities: *(i)* interaction with historical images; *(ii)* addition of saliency on the elements inside the 3D space; (iii) the temporal filtering of the photographs.

General Conclusions and Perspectives

Before ending this document, we present our general conclusions and perspectives to the reader. This dissertation tackles the research issue linked to the association, visualization, and browsing over heterogeneous and extensive historical photo collections of geographic spaces, like postcards and street-level/aerial acquisitions. We have proposed a set of methods for the design of a spatio-temporal, immersive, interactive, and continuous visual exploration of historical images inside a 3D environment constituted of contemporary digital city models. In Chapter 3, we identified two objectives to be accomplished, and we proposed the corresponding solutions to achieve each one:

1. To **decrease users' visual distress when visualizing a historical photograph inside a 3D space**, we propose applying the same camera model describing the device that captured each photograph for the view camera used to render the resulting view of the 3D digital space. However, since the view camera needs to model a real-world camera, our adopted camera model includes the geometric distortions considering non-linear errors during the acquisition process. For the distortion model to be used in the modeling of the view camera, we extrapolated the distortion function to ensure it is defined outside the image domain and is invertible. To reduce the complexity in the distortion computation during the rendering process, we follow a one-pass rendering technique (applying the distortion function directly in both vertex and fragment shaders) for point cloud geometries. On the other hand, we use a two-pass rendering method (using render-to-texture) for other types of geometries, such as triangular meshes.
2. To **facilitate and enhance the user's exploration of historical images through space and time**, we propose the design of a set of techniques for the user's visualization and interaction with historical images inside a 3D environment. As the basis of our design, we take the primary users' needs related to their expectations towards a spatio-temporal exploration. To meet these users' requirements, we define a set of design guidelines where we determine and illustrate our own ideas on how to adapt existing geovisualization techniques for the visual exploration of historical images. We judge our design guidelines and their efficiency by combining them into a single proposition and iteratively evaluate these designs with users to test the consistency of each design, individually and all combined. We improve our designs (and propose new ones) over these observations.

These solutions have made it possible for us to design and propose our model for the visual exploration of historical image collections, choosing an interactive approach for the user to discover these photo collections through space and time.

Main Contributions of the Work

In this section, we present the main contributions of our work linked to *(i)* the rendering process and our proposal for an artifact correction; *(ii)* the photo navigation process and our proposal for the

graphic representations for the users; *(iii)* the overall visual exploration process and our proposal for the formalization of our methods oriented towards historical photographs visualized inside a digital 3D scene containing contemporary 3D city models.

Contribution 1: The Extrapolation of Radial Distortion Models for Rendering Historical Views

Camera calibrated distortion models are estimated only within the image domain. They may thus present extreme deformations or even not be defined outside of the image domain. However, the proposed visualization of the scene through a view camera that offers a zoomed-out version of the camera that acquired the historical photograph requires the evaluation of the distortion outside of its image domain. **Our contribution is the definition and the real-time rendering of an extrapolated distortion model given any radial distortion function.** We have proven this contribution by testing its use by rendering with view cameras modeled using the extrapolated version of different distortion models (radial polynomial and radial fish-eye functions) and compared the result with one from a pinhole camera model. Therefore, we can say that we propose a method that provides an efficient approach for the 3D visualization of historical images that present geometric distortions on scenes composed of digital models represented either by triangular meshes and point clouds.

Among the limitations of our method, we may mention that our proposition for the extrapolation on distortion models is only intended for radial functions (defined by the radius r). Accordingly, we only support radial distortions, where tangential distortion terms are neglected in the definitions of the distortion functions we extrapolate. It narrows down how the distortion presented in a photograph can be modeled inside our environment to just two options: radial polynomial or radial fish-eye. In addition, another limitation is that our approach is mainly oriented toward visualizing one image at a time. Hence our proposed strategy will only allow the user to render one historical view at a time.

Contribution 2: The Proposition of a Set of Techniques to Support Interactive Visual Exploration

We followed a design approach for the proposition of an exploratory strategy for historical photographs based on the user's primary expectations: the visualization of one or several images and the exploration of these images in space and time. Based on the principal properties needed in a 3D system to support these users' requirements for interactive visual exploration, **our contribution is the definitions of new design principles for six central and existing geovisualization techniques (bookmarks, thumbnails, visual variables, visibility heat maps, style representations, and timelines)**. We provide a structure in which these techniques could be used through the definition of three design guidelines:

1. The representation of historical photographs inside a 3D space.
2. The increase of saliency on certain elements in the interactive view.
3. The addition of the temporal information of the photos.

Contribution 3: A Framework for the Co-Visualization of Historical Photos and Contemporary 3D City Models

The HISTOVIS prototype blends our two previously mentioned contributions in a Cultural Heritage context by adopting as input the iconographic photo collections from the Advanced Linking and Exploitation of diGitized geOgRaphic Iconographic heritAge (ALEGORIA) project. **It is an implemented web-based online application with our proposed methods and designs that places historical photographs inside a digital scene composed of contemporary 3D city models**. Three use case scenarios define our view of the broad spectrum in which HISTOVIS can be used and applied:

1. The visualization of two photo collections.
2. A street-view exploration through time.
3. The browsing of non-spatialized photos in a 3D spatial environment.

To evaluate the contribution of our proposed prototype, we performed two separate evaluations: a user study with nine test subjects following a strict test protocol and an iterative presentation (approximately every six months) of the design guidelines for the stakeholders of the ALEGORIA project. This user study and the constant evaluation with the project consortium proved that combining our proposed methods, our HISTOVIS geovisualization system provides a central form of user interaction and reduces the visual complexity of multiple historical images exploration (in space and time) inside a 3D environment.

Among the possible limitations of our method, we may mention that our rendering technique considers the input data to be certain. However, since our primary input is historical images, we deal with imprecise camera calibrations, and limited scene geometries only generated from the contemporary 3D city models. These uncertainties (on relative or absolute poses, camera models, distortion models, and 3D geometries) may hinder the accurate co-registration and calibrations between the photographs and the 3D scene containing the city models. As a result, the visualization of our images inside the 3D space is affected by rendering artifacts, *e.g.*, misalignments between projected photos and 3D geometries or blurring on the blending of various photos.

On the other hand, during the user study, the central aspect pointed out by the users was the comprehension of the spatialized thumbnails around the border of the 3D view. We believe that one of our main limitations towards this technique was our placement algorithm. For our prototype, we did not consider that each thumbnail (or cluster) could potentially occlude each other. We place all directly at the respective position. Additionally, since our design does not follow the menu slider's conventional approach, it can also confuse users, mainly because there is movement inside the 3D scene and the thumbnails move simultaneously with it.

Open Perspectives for Future Work

This section presents the possible extensions related to the other research issues worth exploring concerning our proposed work. We examine different perspectives related to the topics of rendering, geovisualization, and user interactions, among others.

Perspective 1: Improving the Image-Based Rendering of Projected Historical Photos in 3D Spaces

Using Other Distortion Models and Extrapolation Functions: the support for non-radial distortion models is a potential extension to our proposed distortion extrapolation approach. Additionally, it can also be explored the idea of other extrapolation approaches that are not necessarily affine. Furthermore, Newton's method for the inverse distortion calculation can be substituted with other techniques like the one proposed by Vass and Perlaki (2003).

Dealing with the Uncertainty in the Input Data: to avoid rendering artifacts and misalignments between projected photos and the 3D geometry of the city models, we propose computing and analyzing the uncertainty on the input data during the rendering process. For instance, Castro (2021) has already extended this idea for HISTOVIS by developing a view-dependent multi-projective texturing approach using shadow mapping and sprites. However, it does not yet consider the uncertainties present, like the fact that the 3D city models are not a perfect replica even if diachronism is not considered. A solution could be the application of techniques like the ones proposed by Brédif (2014) and Goesele et al. (2010).

Perspective 2: Converging Towards Other Geovisualization Approaches

Reaching Towards Extended Reality: we propose examining the possibility of using our method in Extended Reality (XR) applications. However, we do not recommend adding the functionality related to using a view camera with a distortion model. In Augmented Reality (AR), there is no direct control over the view camera, so the (see-through) scene cannot be distorted. Additionally, in Virtual Reality (VR), there might be user optical discomfort associated with viewing through a distorting lens. Hence, rendering XR with a camera model with distortion would not be ideal, and the user discomfort of rendering should be evaluated before the approach being applied.

Transfering Styles and Radiometries Between Images and Scenes: we would like to add the option of transferring the style or appearance of a selected image directly to the 3D scene instead of having predefined ones. Beyond this, the idea could be significantly extended by using Machine or Deep Learning algorithms, *e.g.*, as it is done by Matzen and Snavely (2014b) and Semmo, Isenberg, and Döllner (2017), to transfer the radiometry and style from historical photos to the 3D city models and vice-versa.

Including Temporal 3D City Models: to treat the diachronism effect, we would like to take benefit of 4D digital models like the strategies proposed by Jaillot, Servigne, and Gesquière (2020) and Schindler and Dellaert (2012). For instance, each model could contain a valid time interval between its construction and destruction, and when a historical photograph is selected and visualized, only the existing models in the temporal time frame defined by the image should be shown in the 3D environment.

Perspective 3: Extending Our Algorithms and Designs to Continue Enhancing User Interactions

Performing Deeper User Studies: we propose to dive further and deeply into evaluating our propositions by using a more significant number of test users. We recommend explicitly checking the most controversial techniques in our study, the spatialized thumbnails, and their clustering (gallery) representation. Additionally, a subjective metric and evaluation could also be defined for the use of a distortion model in the view camera. For instance, the user could be presented with different renderings outcomes and be asked to select the less visually discomforting.

Improving the Placement Algorithm for Spatial Thumbnails: as future work, we want to improve the placement algorithm of the spatialized thumbnails around the border using, for instance, an optimization cost function as Lekschas et al. (2020), where occlusion can be avoided between each thumbnail by updating the positions based on many other factors, *e.g.*, on the current position and the potential moves to the new position.

Examining other Clustering Algorithm Structures: for spatially clustering the thumbnails, we use a hierarchical clustering approach. However, this algorithm is simple compared with many other existing ones. We propose the extension of our clustering strategy by defining more complex algorithms, like density-based or learning-based.

Increasing the System's Scalability: rather than using simple 3D city models with a LOD 1, it would be interesting to place the historical images in a 3D scene depicting more complex structures (*e.g.*, LOD 2 and LOD 3 containing the minor scale characteristics of the roofs and façades from the buildings) and using other 3D models (*e.g.*, point clouds generated from LIDAR acquisitions). Furthermore, we consider that the scalability of our system can be challenged by adding and testing much more extensive and massive historical image collections.

Appendix A

Mock-Ups: HISTOVIS Prototype

Preamble

This appendix presents the ensemble of our initial mock-ups from the HISTOVIS prototype.

Mock-Up 1: Selecting a Time Frame

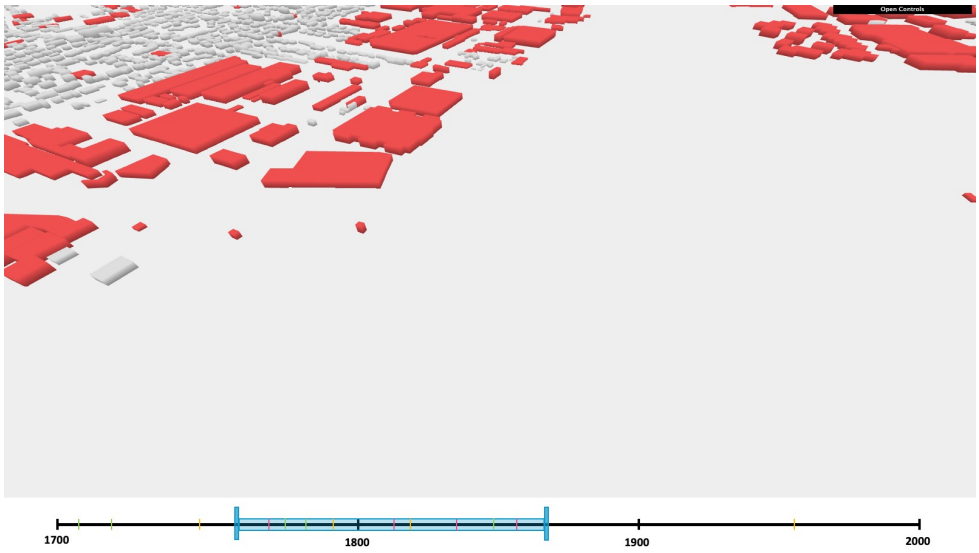


Figure A.1: Mock-up 1: selecting a time frame.

Description

User Task: the user controls the temporal parameter of the photos that are displayed.
System Task: the system performs the following:

1. Determines the best temporal extent to be showed in the time axis for all selected images.
2. Orders all images in a chronological order.
3. Obtains the input of the user related to the selected chronological range.
4. Discards the images that are outside of the chronological range.

Mock-Up 2: Interacting with Bookmarks

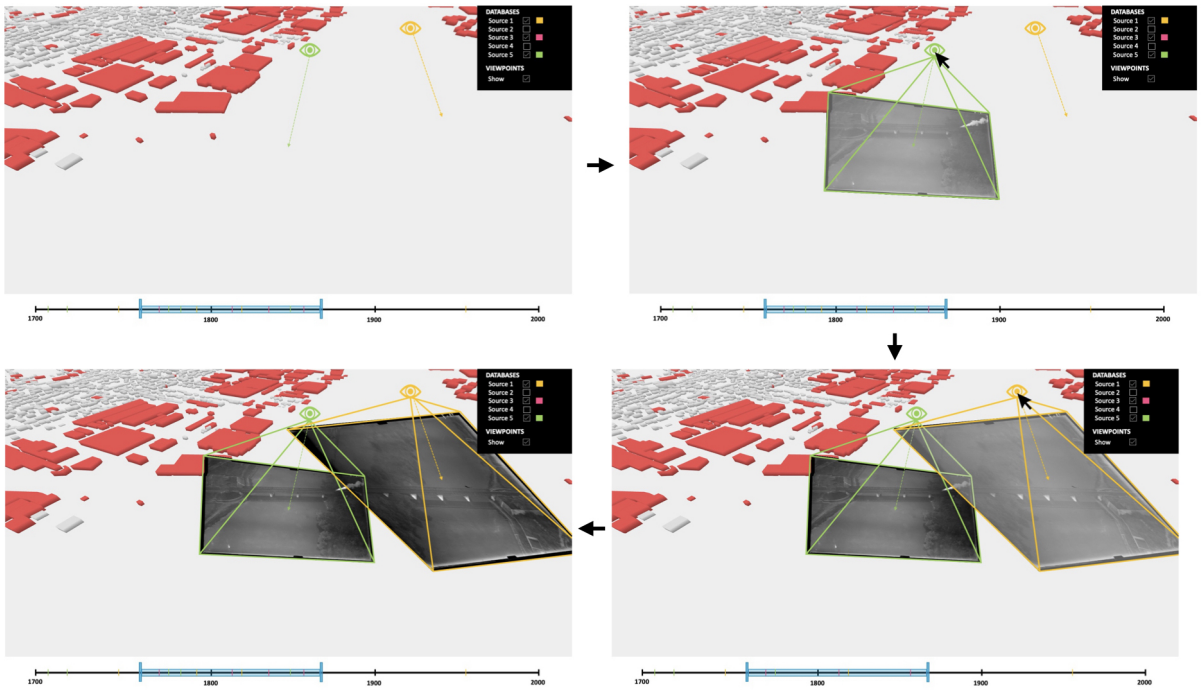


Figure A.2: Mock-up 2: interacting with bookmarks.

Description
User Task: at any viewpoint, the user visualizes the other existing viewpoints from other images. **System Task:** the system performs the following: 1. Shows a viewpoint marker (bookmark) at the position of the optical center of the historical image viewpoint. 2. When hover over the viewpoint marker: a) A ray is projected, and visualized by the user, to the projected center point. b) The image is projected on the scene with a transparent effect. c) When hover out, the image the projection of the image disappears. 3. When the viewpoint marker is clicked, the image is directly projected in the scene. 4. When the viewpoint marker is double-clicked, the virtual camera changes its position to the image viewpoint.

Mock-Up 3: Exploring through Space

Description
User Task: the user navigates and searches through images of same landmark/area/region

targeted by the photo (not from the same position).

System Task: the system performs the following:

1. For each historical image:
   a) Generates a ray (virtually) from the optical center of the viewpoint of the historical camera through the center of the image.
   b) Considers the projected center point, as the point where the generated ray and the scene intercepts.
   c) Calculates the distance of the projected center point of the current evaluated image with the projected center point of the other images.
2. Ranks in ascending order all distances.
3. Hierarchically groups the images based on the ranking of distances.
4. Renders and displays each cluster of images on the screen.

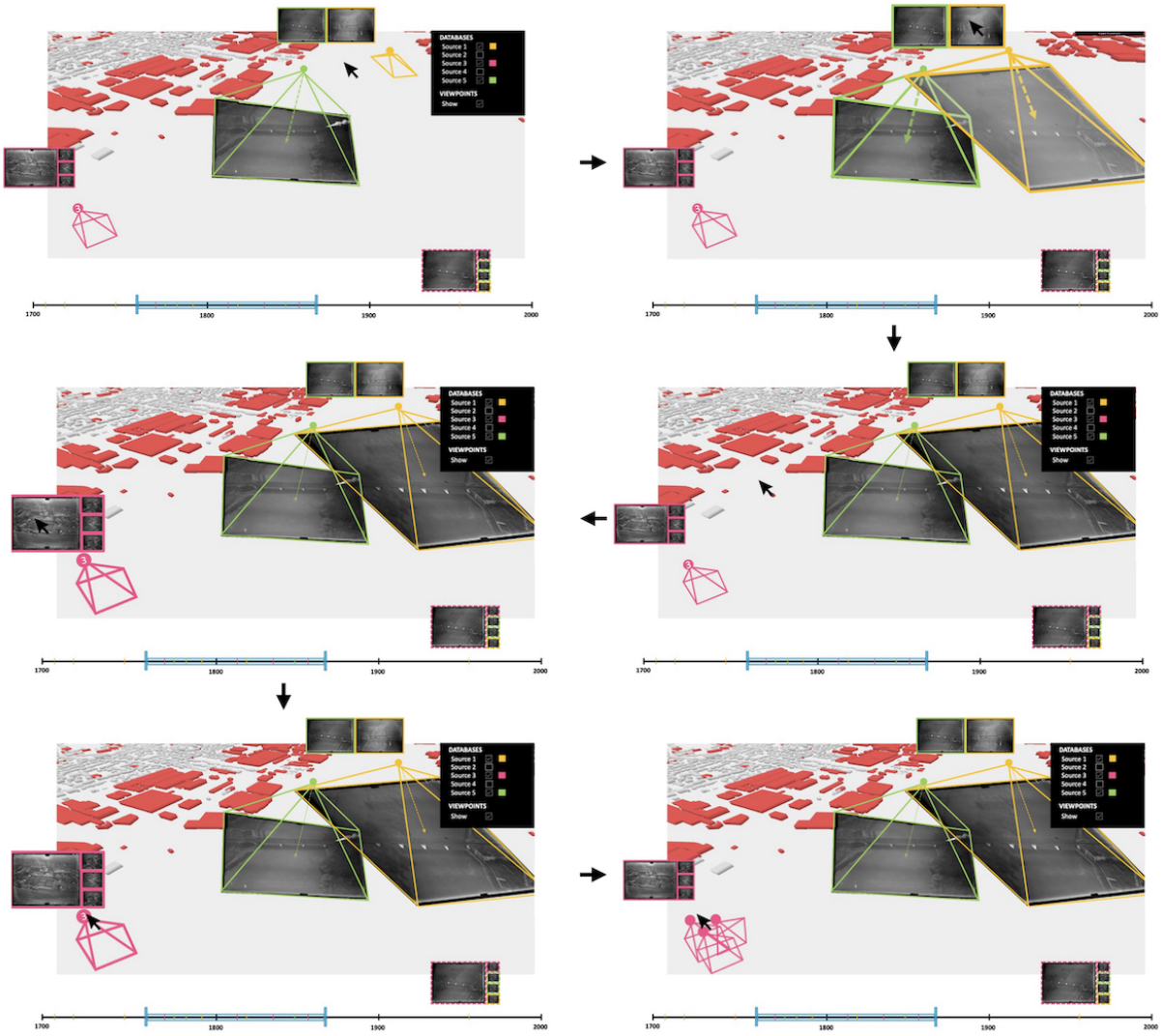


Figure A.3: Mock-up 3: exploring through space.

Mock-Up 4: Co-Visualizing 2D Images and 3D Objects

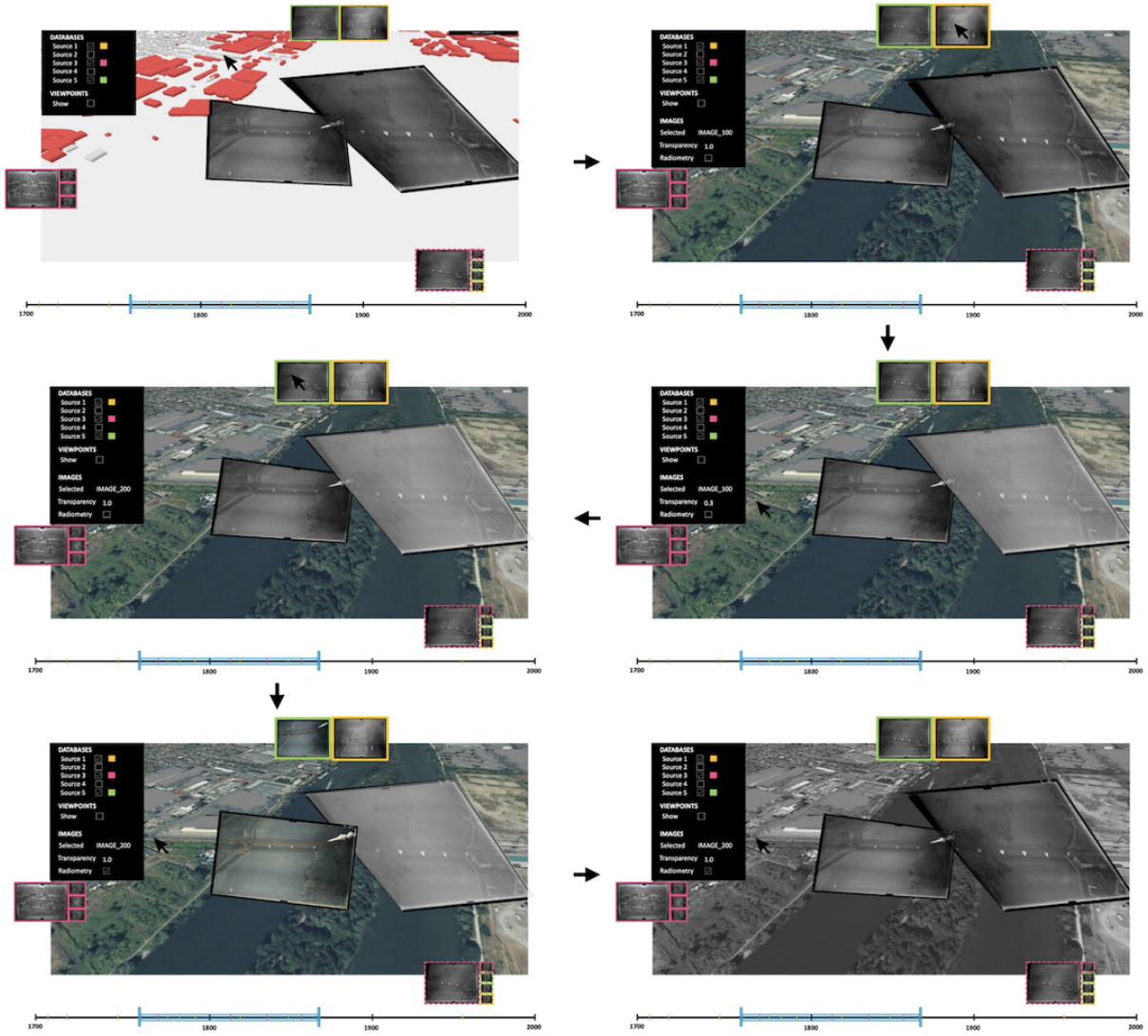


Figure A.4: Mock-up 4: co-visualizing 2D images and 3D objects.

Description
User Task: the user co-visualizes various images for comparison. **System Task:** the system performs the following: 1. Projects each selected image by the user into the scene. 2. Obtains all user inputs (e.g.: transparency, radiometry style, frame parameters, etc.) and updates the projected images with their respective attributes.

Appendix B

Test Protocol: HISTOVIS Prototype

Preamble

The following test protocol was designed and carried online through *Google Forms*[a]. In this appendix, we define the set of tasks and questions given to the users.

[a] https://forms.gle/P99NjJmaLFnUTYfQ8

Thank you for completing this user feedback form. Please follow the instructions that are presented and then answer the questions.

Duration: The maximum time you will need to complete this form is one hour. Please time yourself while completing this form and write down your time in the last comment section.

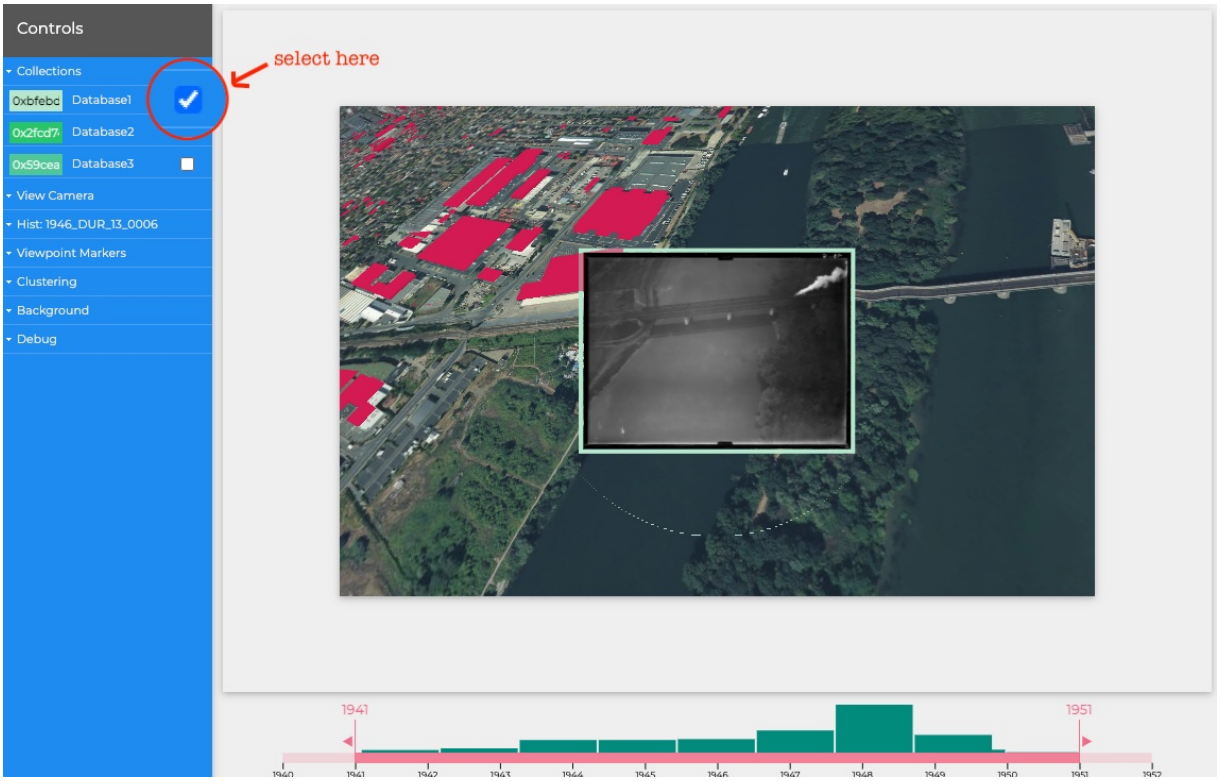


Figure B.1: Test protocol: loading a photo collection in HISTOVIS.

Initial Instructions: You can find the demo in the following URL: https://epaizreyes.github.io/histovis/examples/multiple.html. Please perform the subsequent action:

1. Open the application following the previous URL.
2. On the menu bar on the left, select the first dataset located in `Collections/Database1`.

Temporal Selection

Duration: 15 minutes.
Task Description: Focus on the down section where the timeline is located.

1. Change the selection range (for example, from 1945 to 1947).
2. Place your mouse inside the selected range and zoom in on the timeline using your mouse wheel.
3. Select a smaller range (for example, from October 1945 to April 1946). You can move horizontally in the axis by holding your click and dragging the mouse either to the left or right.

Now please answer the questions and return the range of your timeline to the initial state (where all images are selected).

Question 1: Do you consider the zoom in/out to be continuous? *

- ◯ Yes
- ◯ No

Question 2: If not, how would you improve the continuity of the zoom?

_____

Question 3: For you, is it easy to understand the histogram representation on the timeline? *

- ◯ Yes
- ◯ No

Question 4: If yes, what do you think each bin represents?

_____

Question 5: If not, which representation would you use (considering the uncertainty of the dates)?

_____

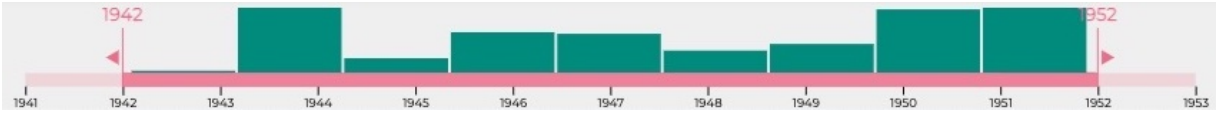


Figure B.2: Test protocol: the timeline segment in HISTOVIS.

Interaction with Bookmarks

Duration: 20 minutes.

Task Description: Move your focus to the 3D scene. Your virtual camera is located at the same position and orientation as a historical image. You should be able to see the image with a little bit of zoom. By holding your click and dragging your mouse around, you can move the virtual camera.

1. Move your camera to search for other viewpoint markers (bookmarks of other existing images). They are represented by a circle that mimics an eye.
2. Hover over a viewpoint marker to see how the image is projected into the scene. When you hover out, the image should disappear.
3. Project another image permanently by clicking on a viewpoint marker. (you should see two projected images now).
4. Move to the marker viewpoint of the previously projected image by double-clicking.
5. Un-project the first image (the one you started with) by trying to find its viewpoint marker and clicking on it.

Focus now on the style of the bookmarks. Go to the menu bar on the left, and open it in the section of `Viewpoint Markers`.

7. Mix and find your favorite representation between target, near, far (only visible when hovering over). You can select more than one.

Now please answer the questions.

Question 6: The change of size when hovering inside a viewpoint marker helps you differentiate the marker from the others? *

- ◯ Yes
- ◯ No

Question 7: If not, how would you differentiate between markers?

Question 8: Which viewpoint marker representation is your preferred one? (you can select more than one) *

- ☐ Target
- ☐ Near
- ☐ Far

Question 9: Would you choose any other representation than the available ones? *

- ◯ Yes
- ◯ No

Question 10: If yes, which one and why?

Question 11: When hovering inside the viewpoint marker, the projected image helps you doing the link between the marker and the image? *

- ◯ Yes
- ◯ No

Question 12: If not, how would you do the link?

_____

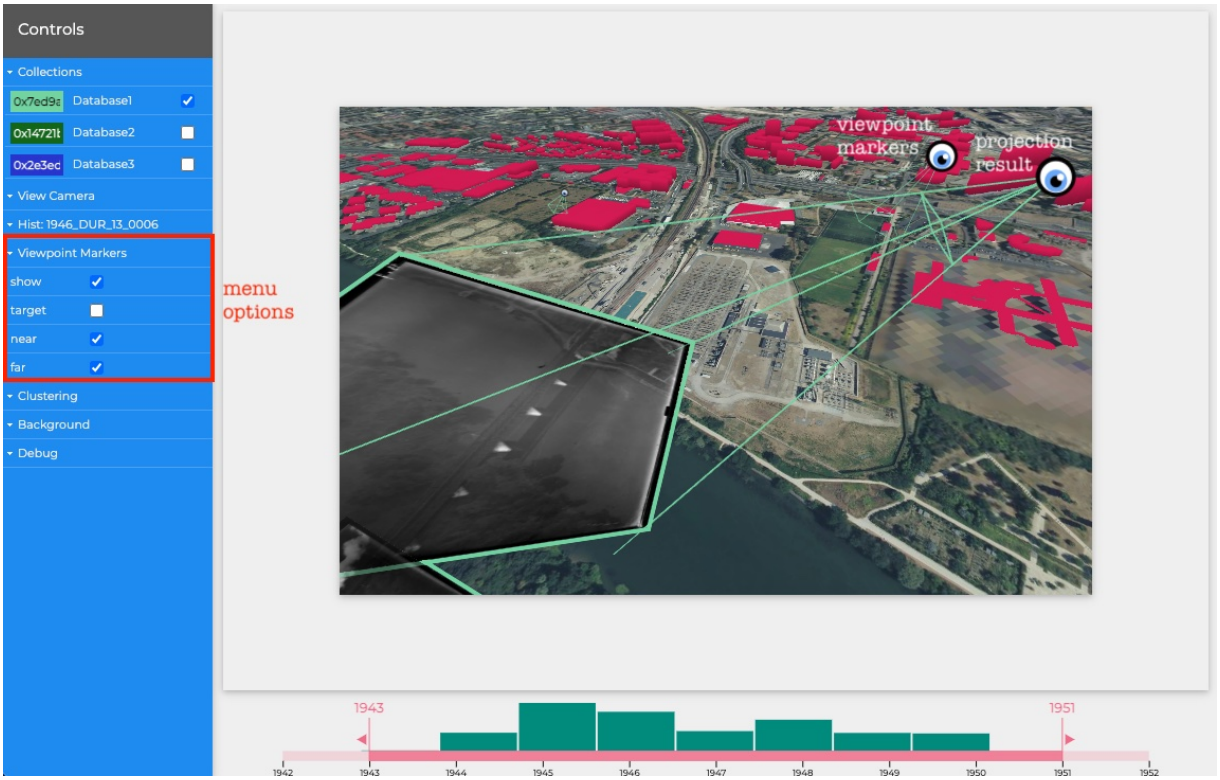


Figure B.3: Test protocol: interaction with bookmarks in HISTOVIS.

Use of Color for Differentiation

Duration: 10 minutes.

Task Description: You will work with two datasets at the same time. Go to the left menu, and select the second collection from `Collections/Dataset2`.

1. Change the color of the first dataset (choose red color).
2. Change the color of the second dataset (choose green color).
3. Project two additional images on your scene. One from the first dataset (viewpoint markers red) and one from the second (viewpoint markers green).

Now please answer the questions.

Question 13: Does the use of colors help you to differentiate between one dataset and the other? (see, for example, the viewpoint markers) *

- ◯ Yes
- ◯ No

Question 14: If not, how would you show the difference between each dataset?

Question 15: Do you consider the colored frame around the projected image helps you differentiate between each image? *

- ◯ Yes
- ◯ No

Question 16: If not, how would you differentiate between projected images?

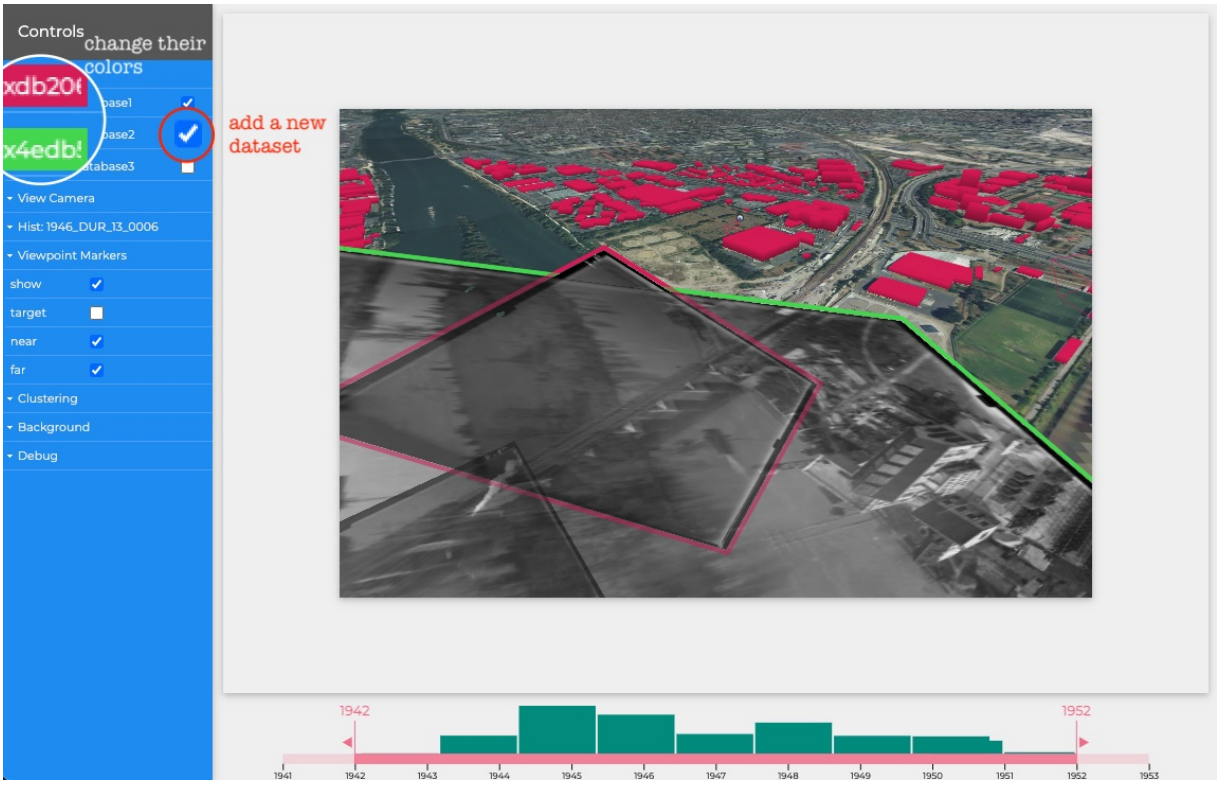


Figure B.4: Test protocol: differentiation of photo collections using different colors in HISTOVIS.

Clustering of Spatialized Thumbnails

Task Description: You need to add the thumbnails around the border of your scene. Go to the left menu, and open the section of clustering.

1. Move the images slider to select all the images (eleven images).
2. Move the clusters slider to select eleven clusters (simple thumbnails).

3. Click and hold inside the scene to move the camera around. You should see the movement of the clusters around.
4. Hover/click/double-click over some clusters to see their projections on the scene (same mouse gestures as the viewpoint markers).
5. Start reducing the number of images in the menu and see how the clusters are generated (reach the one cluster option).

Now please answer the questions.

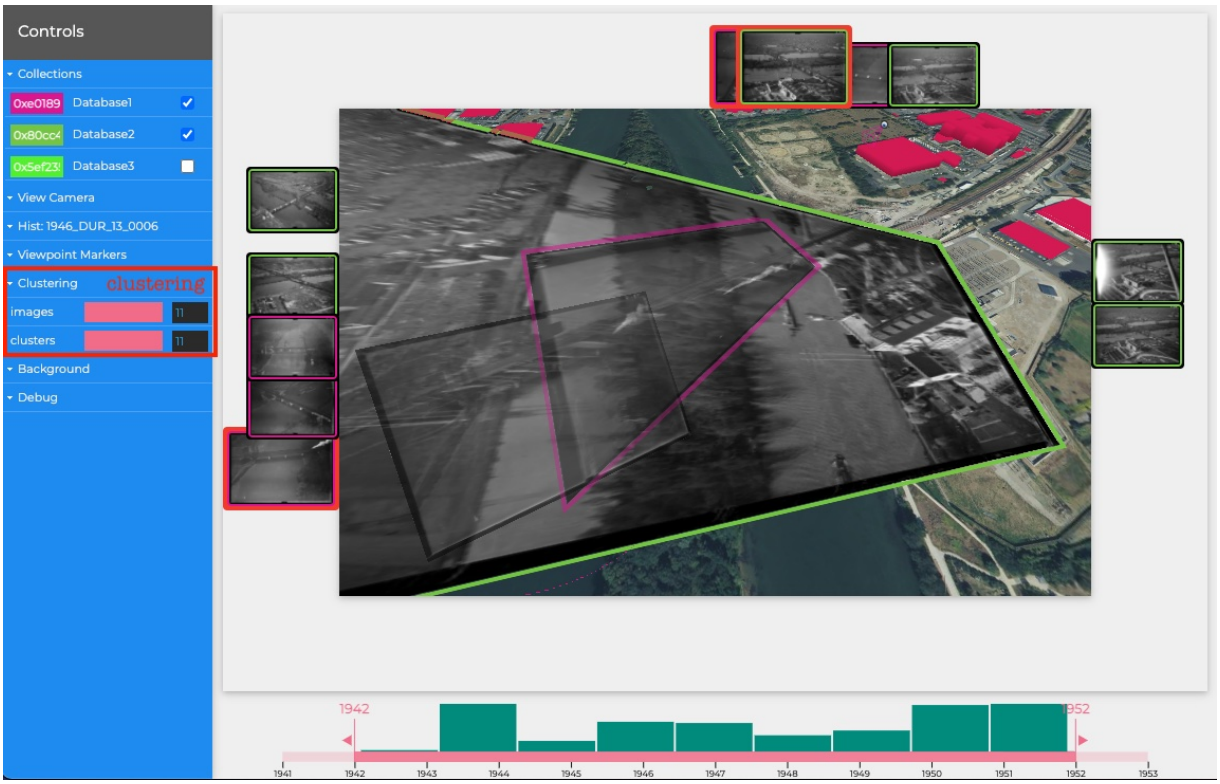


Figure B.5: Test protocol: thumbnails in HISTOVIS.

Question 17: The movement of the thumbnails is continuous for you? *

- ◯ Yes
- ◯ No

Question 18: If no, in which cases do you see a discontinuity?

Question 19: The position of each thumbnail helps you to predict where the photograph will be projected? *

- ◯ Yes
- ◯ No

Question 20: Is the representation of each cluster easy to understand? *

- ◯ Yes
- ◯ No

Question 21: If not, which representation do you think would be a better choice?

Final Comments: Thank you again for helping us in this user feedback experiment!

Do you have any additional comments?

Appendix C

Test Results: HISTOVIS Prototype

Preamble

This appendix presents the ensemble of results collected in the study performed on nine users to evaluate the HISTOVIS prototype. The analysis is given in Section 6.5 of Chapter 6. We divided these results into the following distribution:

- Table C.1 displays the timing information of each test user.
- Table C.2 exhibits the additional comments from specific participants.
- Table C.3 showcases the results related to the temporal selection through a time axis.
- Table C.4 defines the outcomes of interaction with bookmarks inside the 3D environment.
- Table C.5 describes the results associated with the differentiation of images.
- Table C.6 explains the outcomes of our thumbnail proposition for spatialized photographs.

Table C.1: Results 1: timing of test users.

User	Timestamp	Duration (min)
1	3/2/2021 15:57:07	43
2	3/2/2021 16:00:08	20
3	3/2/2021 16:10:47	30
4	3/2/2021 16:13:48	40
5	3/2/2021 16:15:29	35
6	3/2/2021 16:18:47	38
7	3/2/2021 16:18:50	40
8	3/2/2021 16:18:57	45
9	3/2/2021 16:19:20	40
	Average	37

Table C.2: Results 2: additional comments from test users.

User	Comment
1	I was expecting that the small arrow could be used to drag the slider. I find it difficult to map the cluster position to the viewpoint position. I think for me it would make more sense for me to project the viewpoint position to the border, not the image.
3	I did not understand what are the cluster representing.
6	At first, I didn't get the point of controlling the number of images and the number of clusters independently. I figured out a bit later when I saw effectively some images clustering together (which is effectively more clear than having one cluster per image).
9	A global view from top (map like) would be very useful.

Table C.3: Results 3: temporal selection task.

User	Do you consider the zoom in/out to be continuous?	If not, how would you improve the continuity of the zoom?	For you, is it easy to understand the histogram representation on the timeline?	If yes, what do you think each bin represents?	If not, which representation would you use (considering the uncertainty of the dates)?
1	No	Maybe make it match the timeline ticks.	Yes	Number of images per date.	
2	No	Instead of incrementing the zoom with every mouse scroll, "pan" between zooms for perfect smoothness.	Yes	The amount of pictures of the bridge available.	
3	Yes		Yes	The quantity of image for each period.	
4	Yes	In the scene or in the timeline? I don't see anything in the scene when I focus on a range.	Yes	What is a bar?	Which uncertainties of dates? For the moment, I have no ideas on that matter.
5	Yes		Yes	Number of photos.	
6	No	Maybe by also displaying previous scale's bins as it's difficult when small bins appears to relocate.	No		
7	Yes		Yes	Is the number of images you have within that time period.	
8	No	Time line is continuous, but the histogram transition is discontinuous.	Yes	Proportional to the number of photos for the period (absolute number of images is not displayed).	
9	Yes		No	Bin = number of images per time range unit.	The dates are cut (issue: it is on the bottom of my screen and I don't have a vertical scroll bar).

Table C.4: Results 4: interaction with bookmarks task. VPM = viewpoint marker (bookmark).

User	The change of size when hovering inside a VPM helps you differentiate the marker from the others?	If not, how would you differentiate between markers?	Which VPM representation is your preferred one? (you can select more than one)	Would you choose any other representation than the available ones?	If yes, which one and why?	When hovering inside the VPM, the projected image helps you doing the link between the marker and the image?	If not, how would you do the link?
1	Yes		Target, Near	Yes	The border of the image where it would be.	Yes	
2	Yes	Different colours.	Near, Far	No		Yes	Colour coordinate/ have a reference symbol or letter over the picture & marker.
3	Yes		Far	No		Yes	
4	Yes	Make them bigger.	Target	Yes	I would like to make the lines bigger.	Yes	
5	Yes		Target	No		Yes	
6	Yes		Near, Far	Yes	The target rays pollute the view and I'd like them maybe with more opacity.	Yes	
7	Yes		Far	No		Yes	With "target" tick on, I see some lines leading away into the horizon.
8	Yes		Target	Yes	An icon with interactions at the target, not just at the source of the image.	Yes	
9	No	The marker near is the same color than the sky: issue for PoV (=eye) in the sky. "Far" marker not visible for me.	Target	Yes	Maybe by filling the projection cone/pyramid with a semi-transparent color	Yes	

Table C.5: Results 5: use of frame and color for differentiation task.

User	Does the use of colors help you to differentiate between one dataset and the other?	If not, how would you show the difference between each dataset?	Do you consider the colored frame around the projected image helps you differentiate between each image?	If not, how would you differentiate between projected images?
1	Yes		Yes	
2	Yes		Yes	
3	Yes		Yes	
4	No	I do not see the second dataset.	Yes	
5	Yes		Yes	
6	Yes		Yes	
7	No	For the eye-marker that is far away, I don't know which color it is (because it's far and I cannot see the "near" leadlines).	Yes	
8	Yes		Yes	
9	No	Changing the color of the PoV (eye) would be a "plus".	Yes	Clear for me.

Table C.6: Results 6: thumbnails and clustering task.

User	The movement of the thumbnails is continuous for you?	If no, in which cases do you see a discontinuity?	The position of each thumbnail helps you to predict where the photograph will be projected?	Is the representation of each cluster easy to understand?	If not, which representation do you think would be a better choice?
1	Yes		No	No	I don't understand why there is an image bigger than the others.
2	Yes		Yes	Yes	
3	Yes		No	No	
4	Yes		No	Yes	
5	No	I don't understand the question.	No	No	
6	Yes		Yes	Yes	
7	Yes	Yes, but for some thubnails, it takes quite a long time for it to travel to it's positon.	Yes	Yes	I can understand that the position of the cluster is suggesting where the image is coming from, but I can't always find it.
8	Yes	Movement is almost continuous, but not fluid (staggerring).	No	No	Unless you interact with it, you have no hint what the cluster relates to in 3D.
9	No	It is slow.	No	No	Well, I understood the prediction after a few tries.

Bibliography

Here are the references in alphabetical order.

Aber, James, Irene Marzolff, and Johannes Ries (2010a). 'Introduction to Small-Format Aerial Photography'. In: *Small-Format Aerial Photography*. Amsterdam: Elsevier, pp. 1–13. ISBN: 978-0-444-53260-2. DOI: 10.1016/B978-0-444-53260-2.10001-8 (cited on page 17).

Aber, James, Irene Marzolff, and Johannes Ries (2010b). 'Photogrammetry'. In: *Small-Format Aerial Photography*. Amsterdam: Elsevier, pp. 23–39. ISBN: 978-0-444-53260-2. DOI: 10.1016/B978-0-444-53260-2.10003-1 (cited on pages 13, 14).

Abergel, Violette, Kévin Jacquot, Livio De Luca, and Philippe Veron (2019). 'Towards a SLAM-Based Augmented Reality Application for the 3D Annotation of Rock Art'. In: *International Journal of Virtual Reality* 19.2. ISSN: 1081-1451. DOI: 10.20870/IJVR.2019.19.2.2910 (cited on page 33).

Abergel, Violette, Kévin Jacquot, Livio De Luca, and Philippe Veron (2021). 'Combining On-Site and Off-Site Analysis: Towards a New Paradigm for Cultural Heritage Surveys'. In: *DISEGNARECON* 14.26. DOI: 10.20365/DISEGNARECON.26.2021.13 (cited on page 33).

Adelson, Edward and James Bergen (1991). 'The Plenoptic Function and the Elements of Early Vision'. In: *Computational Models of Visual Processing*. MIT Press, pp. 3–20 (cited on page 34).

Akenine-Möller, Tomas, Eric Haines, and Naty Hoffman (2019). 'Image-Based Effects'. In: *Real-Time Rendering*. CRC Press, pp. 439–502. ISBN: 978-1-315-36200-7 (cited on page 59).

Ardissono, Liliana, Tsvi Kuflik, and Daniela Petrelli (2012). 'Personalization in Cultural Heritage: the Road Travelled and the One Ahead'. In: *User Modeling and User-Adapted Interaction* 22.1, pp. 73–99. ISSN: 0924-1868, 1573-1391. DOI: 10.1007/s11257-011-9104-x (cited on pages 2, 48).

Armstrong, Natasha (2012). 'Historypin: Bringing Generations Together Around a Communal History of Time and Place'. In: *Journal of Intergenerational Relationships* 10.3, pp. 294–298. DOI: 10.1080/15350770.2012.697412 (cited on pages 2, 24).

Arnold, Taylor, Nathaniel Ayers, Justin Madron, Robert Nelson, and Lauren Tilton (2020). 'Visualizing a Large Spatiotemporal Collection of Historic Photography with a Generous Interface'. In: *2020 IEEE 5th Workshop on Visualization for the Digital Humanities (VIS4DH)*, pp. 30–35. DOI: 10.1109/VIS4DH51463.2020.00010 (cited on page 28).

Astle, Peter and Adrienne Muir (2002). 'Digitization and Preservation in Public Libraries and Archives'. In: *Journal of Librarianship and Information Science* 34.2, pp. 67–79. DOI: 10.1177/096100060203400202 (cited on page 12).

Baruzzo, Andrea, Paolo Casoto, Prasad Challapalli, Antonina Dattolo, Nirmala Pudota, and Carlo Tasso (2009). 'Toward Semantic Digital Libraries: Exploiting Web 2.0 and Semantic Services in Cultural Heritage'. In: *Journal of Digital Information* 10.6 (cited on pages 2, 48).

Benligiray, Burak and Cihan Topal (2015). 'Lens Distortion Rectification Using Triangulation Based Interpolation'. In: *Advances in Visual Computing*. Vol. 9475. Cham: Springer International Publishing, pp. 35–44. ISBN: 978-3-319-27863-6. DOI: 10.1007/978-3-319-27863-6_4 (cited on page 72).

Bertin, Jacques (2010). *Semiology of Graphics: Diagrams, Networks, Maps*. First Edition. Redlands, California: ESRI Press: Distributed by Ingram Publisher Services. 438 pp. ISBN: 978-1-58948-261-6 (cited on pages xi, 47).

Bian, Ling (2009). 'Spatial Data Models'. In: *International Encyclopedia of Human Geography*. Ed. by Rob Kitchin and Nigel Thrift. Oxford: Elsevier, pp. 337–344. ISBN: 978-0-08-044910-4. DOI: 10.1016/B978-008044910-4.00417-X (cited on page 105).

Bimber, Oliver, Gordon Wetzstein, Andreas Emmerling, and Christian Nitschke (2005). 'Enabling View-Dependent Stereoscopic Projection in Real Environments'. In: *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. Vienna, Austria: IEEE, pp. 14–23. ISBN: 978-0-7695-2459-7. DOI: 10.1109/ISMAR.2005.27 (cited on page 36).

Blanc, Nicolas, Timothée Produit, and Jens Ingensand (2018). 'A Semi-Automatic Tool to Georeference Historical Landscape Images'. In: *PeerJ Preprints* 6. ISSN: 2167-9843. DOI: 10.7287/peerj.preprints.27204v1 (cited on pages 2, 27, 48, 54).

Bolter, Jay David, Maria Engberg, and Blair MacIntyre (2013). 'Media Studies, Mobile Augmented Reality, and Interaction Design'. In: *Interactions* 20.1, pp. 36–45. ISSN: 1072-5520, 1558-3449. DOI: 10.1145/2405716.2405726 (cited on pages 2, 24).

Boynton, Geoffrey (2005). 'Attention and Visual Perception'. In: *Current Opinion in Neurobiology* 15.4, pp. 465–469. ISSN: 09594388. DOI: 10.1016/j.conb.2005.06.009 (cited on pages x, 45).

Brédif, Mathieu (2014). 'Projective Texturing Uncertain Geometry: Silhouette-Aware Box-Filtered Blending Using Integral Radial Images'. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3, pp. 17–23. ISSN: 2194-9050. DOI: 10.5194/isprsannals-II-3-17-2014 (cited on pages 39, 141).

Brivio, Paolo, Luca Benedetti, Marco Tarini, Federico Ponchio, Paolo Cignoni, and Roberto Scopigno (2013). 'PhotoCloud: Interactive Remote Exploration of Joint 2D and 3D Datasets'. In: *IEEE Computer Graphics and Applications* 33.2, pp. 86–96, c3. ISSN: 0272-1716. DOI: 10.1109/MCG.2012.92 (cited on pages 2, 26, 50, 54).

Brumann, Christoph (2015). 'Cultural Heritage'. In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. Second Edition. Oxford: Elsevier, pp. 414–419. ISBN: 978-0-08-097087-5. DOI: 10.1016/B978-0-08-097086-8.12185-3 (cited on pages ix, 8).

Bruschke, Jonas, Florian Niebling, Ferdinand Maiwald, Kristina Friedrichs, Markus Wacker, and Marc Erich Latoschik (2017). 'Towards Browsing Repositories of Spatially Oriented Historic Photographic Images in 3D Web Environments'. In: *Proceedings of the 22nd International Conference on 3D Web Technology - Web3D '17*. Brisbane, Queensland, Australia: ACM Press, pp. 1–6. ISBN: 978-1-4503-4955-0. DOI: 10.1145/3055624.3075947 (cited on page 25).

Bruschke, Jonas, Florian Niebling, and Markus Wacker (2018). 'Visualization of Orientations of Spatial Historical Photographs'. In: *Eurographics Workshop on Graphics and Cultural*

Heritage, pp. 189–192. ISSN: 2312-6124. DOI: 10.2312/GCH.20181359 (cited on pages 50, 51).

Buehler, Chris, Michael Bosse, Leonard McMillan, Steven Gortler, and Michael Cohen (2001). 'Unstructured Lumigraph Rendering'. In: *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '01*. ACM Press, pp. 425–432. ISBN: 978-1-58113-374-5. DOI: 10.1145/383259.383309 (cited on page 59).

Carpendale, Sheelagh (2003). 'Considering Visual Variables as a Basis for Information Visualisation'. In: *PRISM* 2001-693-16. DOI: 10.11575/PRISM/30495 (cited on page 47).

Castro, Scarlleth (2021). *Rendu Basé Image d'Images Historiques*. France: École Polytechnique (cited on page 141).

Chang, Yuan and Guo-Ping Wang (2019). 'A Review on Image-Based Rendering'. In: *Virtual Reality & Intelligent Hardware* 1.1, pp. 39–54. ISSN: 2096-5796. DOI: 10.3724/SP.J.2096-5796.2018.0004 (cited on pages 34, 35).

Chen, Shenchang Eric (1995). 'QuickTime VR: An Image-Based Approach to Virtual Environment Navigation'. In: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '95*. ACM Press, pp. 29–38. ISBN: 978-0-89791-701-8. DOI: 10.1145/218380.218395 (cited on page 34).

Chen, Shenchang Eric and Lance Williams (1993). 'View Interpolation for Image Synthesis'. In: *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '93*. ACM Press, pp. 279–288. ISBN: 978-0-89791-601-1. DOI: 10.1145/166117.166153 (cited on pages 36, 37).

Chippendale, Paul, Michele Zanin, and Claudio Andreatta (2008). 'Spatial and Temporal Attractiveness Analysis Through Geo-Referenced Photo Alignment'. In: *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium* 2, pp. II–1116–II–1119 (cited on page 51).

Collier, Peter (2009). 'Photogrammetry / Aerial Photography'. In: *International Encyclopedia of Human Geography*. Oxford: Elsevier, pp. 151–156. ISBN: 978-0-08-044910-4. DOI: 10.1016/B978-008044910-4.00059-6 (cited on pages 13, 17, 19).

Çöltekin, Arzu, Amy Griffin, Aidan Slingsby, Anthony Robinson, Sidonie Christophe, Victoria Rautenbach, Min Chen, Christopher Pettit, and Alexander Klippel (2020). 'Geospatial Information Visualization and Extended Reality Displays'. In: *Manual of Digital Earth*. Singapore: Springer, pp. 229–277. ISBN: 978-981-329-914-6. DOI: 10.1007/978-981-32-9915-3_7 (cited on pages 2, 22, 30, 48, 81, 82).

Çöltekin, Arzu, Halldór Janetzko, and Sara Fabrikant (2018). 'Geovisualization'. In: *Geographic Information Science & Technology Body of Knowledge* 2018 (Q2). ISSN: 25772848. DOI: 10.22224/gistbok/2018.2.6 (cited on pages ix, x, 2, 21, 22).

Conway, Paul (2008). 'Best Practices for Digitizing Photographs: A Network Analysis of Influences'. In: *Imaging Science & Technology*. Berne, Switzerland, pp. 94–102. DOI: 2027.42/85229 (cited on page 12).

Conway, Paul (2009). 'Building Meaning in Digitized Photographs'. In: *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1.1, p. 19. DOI: 2027.42/85222 (cited on page 12).

Cornsweet, Tom (2012). *Visual Perception*. Elsevier Science & Technology Books. ISBN: 978-0-323-14821-4. DOI: 10.1016/B978-0-12-189750-5.X5001-5 (cited on page 44).

Corsini, Massimiliano, Matteo Dellepiane, Federico Ponchio, and Roberto Scopigno (2009). 'Image-to-Geometry Registration: a Mutual Information Method exploiting Illumination-related Geometric Properties'. In: *Computer Graphics Forum* 28.7, pp. 1755–1764. ISSN: 01677055, 14678659. DOI: 10.1111/j.1467-8659.2009.01552.x (cited on page 36).

Davoine, Paule-Annick, Hélène Mathian, Cécile Saint-Marc, Jean-Yves Blaise, and Lahouari Kaddouri (2015). 'The Visual Representations of Territorial Dynamics: Retrospective and Input from New Computing Environments'. In: *Proceedings of the ICA*. Vol. 2. Rio de Janeiro, Brazil. ISBN: 978-85-88783-11-9 (cited on page 28).

De Luca, Livio, Chawee Busarayat, Chiara Stefani, Noemie Renaudin, Michel Florenzano, and Philippe Véron (2010). 'An Iconography-Based Modeling Approach for the Spatio-Temporal Analysis of Architectural Heritage'. In: *2010 Shape Modeling International Conference*. Aix-en-Provence, France: IEEE, pp. 78–89. ISBN: 978-1-4244-7259-8. DOI: 10.1109/SMI.2010.28 (cited on page 54).

Debevec, Paul, Yizhou Yu, and George Borshukov (1998). 'Efficient View-Dependent Image-Based Rendering with Projective Texture-Mapping'. In: *Rendering Techniques '98*. Vienna: Springer, pp. 105–116. ISBN: 978-3-7091-6453-2. DOI: 10.1007/978-3-7091-6453-2_10 (cited on page 35).

DeFries, Ruth (2013). 'Remote Sensing and Image Processing'. In: *Encyclopedia of Biodiversity*. Second Edition. Waltham: Academic Press, pp. 389–399. ISBN: 978-0-12-384720-1. DOI: 10.1016/B978-0-12-384719-5.00383-X (cited on pages 17, 18).

Delavoipière, Laetitia, Sylvaine Conord, and Anaïs Marshall (2020). 'La Méthode de la Photo-Interview à Partir de la Photographie Aérienne: le Cas d'un Bidonville à Nanterre dans les Années 1960'. In: *EchoGéo* 54. ISSN: 1963-1197. DOI: 10.4000/echogeo.20672 (cited on pages 102, 125).

Devaux, Alexandre and Mathieu Brédif (2016). 'Realtime Projective Multi-Texturing of Pointclouds and Meshes for a Realistic Street-View Web Navigation'. In: *Proceedings of the 21st International Conference on Web3D Technology - Web3D '16*. Anaheim, California: ACM Press, pp. 105–108. ISBN: 978-1-4503-4428-9. DOI: 10.1145/2945292.2945311 (cited on page 36).

Dewitz, Leyla, Cindy Kröber, Heike Messemer, Ferdinand Maiwald, Sander Münster, Jonas Bruschke, and Florian Niebling (2019). 'Historical Photos and Visualizations: Potential for Research'. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W15, pp. 405–412. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLII-2-W15-405-2019 (cited on pages 28, 30, 49, 54).

Doucet, Emily (2021). '"The Idea Was in the Air": Nadar's Aerial Media'. In: *Grey Room* 83, pp. 112–137. ISSN: 1526-3819. DOI: 10.1162/grey_a_00322 (cited on page 17).

Dzenko, Corey (2009). 'Analog to Digital: The Indexical Function of Photographic Images'. In: *Afterimage* 37.3, pp. 19–23. DOI: 10.1525/aft.2009.37.2.19 (cited on page 12).

Eisemann, Martin, Bert De Decker, Marcus Magnor, Philippe Bekaert, Edilson de Aguiar, Naveed Ahmed, Christian Theobalt, and Anita Sellent (2008). 'Floating Textures'. In: *Computer Graphics Forum* 27.2, pp. 409–418. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/j.1467-8659.2008.01138.x (cited on pages 39, 40).

Eisemann, Martin, Anita Sellent, and Marcus Magnor (2007). 'Filtered Blending: A New, Minimal Reconstruction Filter for Ghosting-Free Projective Texturing with Multiple

Images'. In: *Proc. Vision, Modeling and Visualization (VMV)*. Eurographics, pp. 119–126 (cited on page 39).

Everitt, Cass (2001). 'Projective Texture Mapping'. In: *White Paper, Nvidia Corporation* 4.3, p. 7 (cited on page 35).

Forgione, Thomas, Axel Carlier, Géraldine Morin, Wei Tsang Ooi, and Vincent Charvillat (2016). 'Impact of 3D Bookmarks on Navigation and Streaming in a Networked Virtual Environment'. In: *Proceedings of the 7th International Conference on Multimedia Systems*. Klagenfurt Austria: ACM, pp. 1–10. ISBN: 978-1-4503-4297-1. DOI: 10.1145/2910017.2910607 (cited on pages ix, 48).

Förstner, Wolfgang and Bernhard Wrobel (2016). *Photogrammetric Computer Vision: Statistics, Geometry, Orientation and Reconstruction*. First Edition. Springer Publishing Company, Incorporated. ISBN: 3-319-11549-9 (cited on pages 13, 15, 64, 65, 71).

Fraser, Clive (1997). 'Digital Camera Self-Calibration'. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 52.4, pp. 149–159. ISSN: 0924-2716. DOI: 10.1016/S0924-2716(97)00005-1 (cited on page 71).

Garg, Rahul, Hao Du, Steven Seitz, and Noah Snavely (2009). 'The Dimensionality of Scene Appearance'. In: *IEEE 12th International Conference on Computer Vision*. Kyoto, pp. 1917–1924. ISBN: 978-1-4244-4420-5. DOI: 10.1109/ICCV.2009.5459424 (cited on page 25).

Garlandini, Simone and Sara Fabrikant (2009). 'Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization'. In: *Spatial Information Theory*. Vol. 5756. Berlin, Heidelberg: Springer, pp. 195–211. ISBN: 978-3-642-03832-7. DOI: 10.1007/978-3-642-03832-7_12 (cited on page 47).

Gautier, Jacques, Paule-Annick Davoine, and Claire Cunty (2018). 'Visual Search of Cyclic Spatio-Temporal Events'. In: *Proceedings of the ICA*. Vol. 1. Washington, United States, pp. 44–51. DOI: 10.5194/ica-proc-1-44-2018 (cited on page 27).

Goesele, Michael, Jens Ackermann, Simon Fuhrmann, Carsten Haunbold, Ronny Klowsky, Drew Steedly, and Rick Szeliski (2010). 'Ambient Point Clouds for View Interpolation'. In: *ACM Transactions on Graphics* 29 (cited on pages 40, 42, 141).

Goldstein, Bruce (2010). *Sensation and Perception*. Eigth Edition. Belmont, California: Wadsworth Cengage Learning. ISBN: 978-0-495-60442-6 (cited on pages ix, x, 44–46).

Gominski, Dimitri, Valérie Gouet-Brunet, and Liming Liming Chen (2021). 'Connecting Images through Sources: Exploring Low-Data, Heterogeneous Instance Retrieval'. In: *Remote Sensing* 13.16. ISSN: 2072-4292. DOI: 10.3390/rs13163080 (cited on page 128).

Gominski, Dimitri, Martyna Poreba, Valérie Gouet-Brunet, and Liming Chen (2019). 'Challenging Deep Image Descriptors for Retrieval in Heterogeneous Iconographic Collections'. In: *Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents - SUMAC '19*. Nice, France: ACM Press, pp. 31–38. ISBN: 978-1-4503-6910-7. DOI: 10.1145/3347317.3357246 (cited on page 128).

Gore, Al (1998). 'The Digital Earth'. In: *Australian Surveyor* 43.2, pp. 89–91. ISSN: 0005-0326. DOI: 10.1080/00050348.1998.10558728 (cited on pages ix, 22, 81).

Gortler, Steven, Radek Grzeszczuk, Richard Szeliski, and Michael Cohen (1996). 'The Lumigraph'. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and*

Interactive Techniques - SIGGRAPH '96. New York, NY, USA: ACM Press, pp. 43–54. ISBN: 978-0-89791-746-9. DOI: 10.1145/237170.237200 (cited on pages 34, 59).

Gouet-Brunet, Valérie et al. (2017). *ALEGORIA Internal Report: Description of the Project*. France: IGN, p. 34 (cited on page 102).

Grundland, Mark and Neil Dodgson (2005). 'Color Histogram Specification by Histogram Warping'. In: Electronic Imaging 2005. San Jose, CA, p. 610. DOI: 10.1117/12.596953 (cited on page 43).

Gurdan, Tobias, Martin R. Oswald, Daniel Gurdan, and Daniel Cremers (2014). 'Spatial and Temporal Interpolation of Multi-View Image Sequences'. In: *Pattern Recognition*. Vol. 8753. Cham: Springer International Publishing, pp. 305–316. ISBN: 978-3-319-11752-2. DOI: 10.1007/978-3-319-11752-2_24 (cited on page 38).

Harmon, John E. and Steven J. Anderson (2003). 'Designing Spatial Data'. In: *The Design and Implementation of Geographic Information Systems*. Science/Geography. Wiley, pp. 69–78. ISBN: 978-0-471-20488-6 (cited on page 105).

Hartley, Richard and Andrew Zisserman (2003). 'Estimation: 2D Projective Transformations'. In: *Multiple View Geometry in Computer Vision*. Second Edition. Cambridge University Press. ISBN: 978-0-521-54051-3. DOI: 10.1017/CBO9780511811685 (cited on page 16).

Hirschmuller, Heiko (2005). 'Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information'. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 2, pp. 807–814. DOI: 10.1109/CVPR.2005.56 (cited on page 17).

Höhle, Joachim (2013). 'Oblique Aerial Images and Their Use in Cultural Heritage Documentation'. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-5/W2, pp. 349–354. ISSN: 2194-9034. DOI: 10.5194/isprsarchives-XL-5-W2-349-2013 (cited on page 19).

Howard, Ian (1996). 'Alhazen's Neglected Discoveries of Visual Phenomena'. In: *Perception* 25.10, pp. 1203–1217. DOI: 10.1068/p251203 (cited on page 9).

Immel, David, Michael Cohen, and Donald Greenberg (1986). 'A Radiosity Method for Non-Diffuse Environments'. In: *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '86*. ACM Press, pp. 133–142. ISBN: 978-0-89791-196-2. DOI: 10.1145/15922.15901 (cited on pages 33, 34).

Itti, Laurent (2007). 'Visual Salience'. In: *Scholarpedia* 2.9, p. 3327. DOI: 10.4249/scholarpedia.3327 (cited on page 46).

Jaillot, Vincent, Valentin Rigolle, Sylvie Servigne, John Samuel, and Gilles Gesquière (2021). 'Integrating Multimedia Documents and Time-Evolving 3D City Models for Web Visualization and Navigation'. In: *Transactions in GIS* 25.3, pp. 1419–1438. ISSN: 1361-1682. DOI: 10.1111/tgis.12734 (cited on page 29).

Jaillot, Vincent, Sylvie Servigne, and Gilles Gesquière (2020). 'Delivering Time-Evolving 3D City Models for Web Visualization'. In: *International Journal of Geographical Information Science* 34.10, pp. 2030–2052. ISSN: 1365-8816, 1362-3087. DOI: 10.1080/13658816.2020.1749637 (cited on pages 29, 141).

Jenkins, Reese (1975). 'Technology and the Market: George Eastman and the Origins of Mass Amateur Photography'. In: *Technology and Culture* 16.1, pp. 1–19. ISSN: 0040165X, 10973729 (cited on page 12).

Kajiya, James (1986). 'The Rendering Equation'. In: *Proceedings of the 13th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '86*. ACM Press, pp. 143–150. ISBN: 0897911962. DOI: 10.1145/15922.15902 (cited on page 33).

Keahey, Alan and Edward Robertson (1996). 'Non-Linear Image Magnification'. In: *Proceedings IEEE Symposium on Information Visualization '96*, pp. 38–45 (cited on pages 41, 68).

Koebel, Kathrin, Doris Agotai, and Arzu Çöltekin (2020). 'Exploring Cultural Heritage Collections in Immersive Analytics: Challenges, Benefits, and a Case Study Using Virtual Reality'. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B4-2020, pp. 599–606. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLIII-B4-2020-599-2020 (cited on page 31).

Kopf, Johannes, Michael Cohen, and Richard Szeliski (2014). 'First-Person Hyper-Lapse Videos'. In: *ACM Transactions on Graphics* 33.4, pp. 1–10. ISSN: 0730-0301, 1557-7368. DOI: 10.1145/2601097.2601195 (cited on page 38).

Kraak, Menno-Jan (2003). 'Geovisualization Illustrated'. In: *Challenges in Geospatial Analysis and Visualization* 57.5, pp. 390–399. ISSN: 0924-2716. DOI: 10.1016/S0924-2716(02)00167-3 (cited on page 21).

Kraus, Karl (2011). *Photogrammetry*. Second Edition. De Gruyter. ISBN: 978-3-11-089287-1. DOI: 10.1515/9783110892871 (cited on pages 14, 64).

Krig, Scott (2016). 'Image Capture and Representation'. In: *Computer Vision Metrics*. Textbook Edition. Springer International Publishing. ISBN: 978-3-319-33762-3 (cited on pages 25, 67, 68).

Kupfer, John and Charles Emerson (2005). 'Remote Sensing'. In: *Encyclopedia of Social Measurement*. New York: Elsevier, pp. 377–383. ISBN: 978-0-12-369398-3. DOI: 10.1016/B0-12-369398-5/00340-6 (cited on page 17).

Land, Michael F. and Mary Hayhoe (2001). 'In What Ways do Eye Movements Contribute to Everyday Activities?' In: *Vision Research* 41.25, pp. 3559–3565. ISSN: 00426989. DOI: 10.1016/S0042-6989(01)00102-X (cited on page 46).

Lecat, Paul, Émile Blettery, Lætitia Delavoipière, Frédéric Saly-Giocanti, Sylvaine Conord, Valérie Gouet-Brunet, Alexandre Devaux, Mathieu Brédif, and Frédéric Moret (2021). 'Restituer les Bidonvilles de Nanterre : l'Apport d'un Outil de Visualisation 3D à un Projet de Sciences Sociales'. In: *Humanités Numériques* 3. ISSN: 2736-2337. DOI: 10.4000/revuehn.1946 (cited on pages 102, 125).

Lee, Gun A., Andreas Dunser, Seungwon Kim, and Mark Billinghurst (2012). 'CityViewAR: A Mobile Outdoor AR Application for City Visualization'. In: *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*. Atlanta, GA, USA: IEEE, pp. 57–64. ISBN: 978-1-4673-4665-8. DOI: 10.1109/ISMAR-AMH.2012.6483989 (cited on page 32).

Lefèvre, Wolfgang (2007). *Inside the Camera Obscura: Optics and Art Under the Spell of the Projected Image*. Berlin: Max-Planck Institute for the History of Science (cited on page 10).

Lekschas, Fritz, Michael Behrisch, Benjamin Bach, Peter Kerpedjiev, Nils Gehlenborg, and Hanspeter Pfister (2020). 'Pattern-Driven Navigation in 2D Multiscale Visualizations with Scalable Insets'. In: *IEEE Transactions on Visualization and Computer Graphics* 26, pp. 611–621 (cited on pages 50, 142).

Levoy, Marc and Pat Hanrahan (1996). 'Light Field Rendering'. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '96*. ACM Press, pp. 31–42. ISBN: 978-0-89791-746-9. DOI: 10.1145/237170.237199 (cited on pages 34, 59).

Lheureux, Rosine, Solange Bidou, Sandrine Bula, Florence Clavaud, Marie-Eve Bouillon, and Alice Marsal (2018). *ALEGORIA Internal Report: Description of the Photographic Collections Processed*. France: National Archives, p. 30 (cited on pages 101, 102).

Li, Zhengqi, Wenqi Xian, Abe Davis, and Noah Snavely (2020). 'Crowdsampling the plenoptic function'. In: *European Conference on Computer Vision*. Springer, pp. 178–196. DOI: 10.1007/978-3-030-58452-8_11 (cited on page 43).

Lobo, Maria and Sidonie Christophe (2020). 'Opportunities and Challenges for Augmented Reality Situated Geographical Visualization'. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* V-4-2020, pp. 163–170. ISSN: 2194-9050. DOI: 10.5194/isprs-annals-V-4-2020-163-2020 (cited on page 32).

Lorenz, Haik and Jürgen Döllner (2008). 'Dynamic Mesh Refinement on GPU Using Geometry Shaders'. In: *Proceedings of the 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*. Plzeñ, Czech Republic, pp. 97–104 (cited on pages 41, 67).

Lorenz, Haik and Jürgen Döllner (2009). 'Real-time Piecewise Perspective Projections'. In: *Proceedings of the Fourth International Conference on Computer Graphics Theory and Applications*. Lisboa, Portugal, pp. 147–155. DOI: 10.5220/0001786401470155 (cited on pages 41, 67).

Lowe, David (1999). 'Object Recognition from Local Scale-Invariant Features'. In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Kerkyra, Greece: IEEE, 1150–1157 vol.2. ISBN: 978-0-7695-0164-2. DOI: 10.1109/ICCV.1999.790410 (cited on page 16).

Lowe, David (Nov. 2004). 'Distinctive Image Features from Scale-Invariant Keypoints'. In: *International Journal of Computer Vision* 60.2, pp. 91–110. ISSN: 0920-5691. DOI: 10.1023/B:VISI.0000029664.99615.94 (cited on page 16).

Luck, S. J. and M. A. Ford (1998). 'On the Role of Selective Attention in Visual Perception'. In: *Proceedings of the National Academy of Sciences* 95.3, pp. 825–830. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.95.3.825 (cited on page 46).

MacEachren, Alan (1994). 'Visualization in Modern Cartography: Setting the Agenda'. In: *Modern Cartography Series*. Vol. 2. Academic Press, pp. 1–12. DOI: 10.1016/B978-0-08-042415-6.50008-9 (cited on page 21).

MacEachren, Alan, Mark Gahegan, William Pike, Isaac Brewer, Guoray Cai, Eugene Lengerich, and Frank Hardistry (2004). 'Geovisualization for Knowledge Construction and Decision Support'. In: *IEEE Computer Graphics and Applications* 24.1, pp. 13–17. ISSN: 0272-1716. DOI: 10.1109/MCG.2004.1255801 (cited on pages 21, 22).

MacEachren, Alan and Menno-Jan Kraak (2001). 'Research Challenges in Geovisualization'. In: *Cartography and Geographic Information Science* 28.1, pp. 3–12. ISSN: 1523-0406, 1545-0465. DOI: 10.1559/152304001782173970 (cited on page 81).

MacEachren, Alan, Robert Roth, James O'Brien, Bonan Li, Derek Swingley, and Mark Gahegan (2012). 'Visual Semiotics & Uncertainty Visualization: An Empirical Study'.

In: *IEEE Transactions on Visualization and Computer Graphics* 18.12, pp. 2496–2505. ISSN: 1077-2626. DOI: 10.1109/TVCG.2012.279 (cited on pages xi, 47).

Maiwald, Ferdinand, Jonas Bruschke, Christoph Lehmann, and Florian Niebling (2019). 'A 4D Information System for the Exploration of Multitemporal Images and Maps using Photogrammetry, Web Technologies and VR/AR'. In: *Virtual Archaeology Review* 10.21, p. 1. ISSN: 1989-9947. DOI: 10.4995/var.2019.11867 (cited on pages 28, 30, 49, 54).

Maiwald, Ferdinand, Frank Henze, Jonas Bruschke, and Florian Niebling (2019). 'Geo-Information Technologies for a Multimodal Access on Hisotrical Photographs and Maps for Research and Communication in Urban History'. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-2/W11, pp. 763–769. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLII-2-W11-763-2019 (cited on pages 28, 30, 31, 49, 54).

Manning, Russell and Charles Dyer (1999). 'Interpolating View and Scene Motion by Dynamic View Morphing'. In: *1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Vol. 1, pp. 388–394. DOI: 10.1109/CVPR.1999.786968 (cited on page 37).

Marcus, Aaron (1995). 'Principles of Effective Visual Communication for Graphical User Interface Design'. In: *Readings in Human–Computer Interaction*. Morgan Kaufmann, pp. 425–441. ISBN: 978-0-08-051574-8. DOI: 10.1016/B978-0-08-051574-8.50044-3 (cited on page 43).

Marien Warner, Mary (2006). 'The Origins of Photography (to 1839)'. In: *Photography: A Cultural History*. Second Edition. Laurence King. ISBN: 9781856694933 (cited on page 10).

Martin-Beaumont, Nicolas, Nicolas Nony, Benoit Deshayes, Marc Pierrot-Deseilligny, and Livio De Luca (2013). 'Photographer-Friendly Work-Flows for Image-Based Modelling of Heritage Artefacts'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-5/W2, pp. 421–424. ISSN: 2194-9034. DOI: 10.5194/isprsarchives-XL-5-W2-421-2013 (cited on page 17).

Matzen, Kevin and Noah Snavely (2014a). 'Exploiting Deep Generative Prior for Versatile Image Restoration and Manipulation'. In: *Computer Vision – ECCV 2014*. Vol. 8695. Glasgow, UK: Springer International Publishing, pp. 615–630. ISBN: 978-3-319-10584-0. DOI: 10.1007/978-3-319-10584-0_40 (cited on page 43).

Matzen, Kevin and Noah Snavely (2014b). 'Scene Chronology'. In: *Computer Vision – ECCV 2014*. Vol. 8695. Cham: Springer International Publishing, pp. 615–630. ISBN: 978-3-319-10584-0. DOI: 10.1007/978-3-319-10584-0_40 (cited on page 141).

McMillan, Leonard and Gary Bishop (1995). 'Plenoptic Modeling: An Image-Based Rendering System'. In: *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '95*. ACM Press, pp. 39–46. ISBN: 978-0-89791-701-8. DOI: 10.1145/218380.218398 (cited on page 34).

Mennis, Jeremy, Donna Peuquet, and Liujian Qian (2000). 'A Conceptual Framework for Incorporating Cognitive Principles into Geographical Database Representation'. In: *International Journal of Geographical Information Science* 14.6, pp. 501–520. ISSN: 1365-8816, 1362-3087. DOI: 10.1080/136588100415710 (cited on page 27).

Meyer, Élise, Pierre Grussenmeyer, Jean-Pierre Perrin, Anne Durand, and Pierre Drap (2007). 'A Web Information System for the Management and the Dissemination of

Cultural Heritage Data'. In: *Journal of Cultural Heritage* 8.4, pp. 396–411. ISSN: 12962074. DOI: 10.1016/j.culher.2007.07.003 (cited on page 2).

Moulon, Pierre (2014). 'Positionnement Robuste et Précis de Réseaux d'Images'. Issue: 2014PEST1033. PhD thesis. Université Paris-Est (cited on pages 14, 15).

Niebling, Florian, Jonas Bruschke, Heike Messemer, Markus Wacker, and Sebastian von Mammen (2020). 'Analyzing Spatial Distribution of Photographs in Cultural Heritage Applications'. In: *Visual Computing for Cultural Heritage*. Cham: Springer, pp. 391–408. ISBN: 978-3-030-37190-6. DOI: 10.1007/978-3-030-37191-3_20 (cited on page 51).

Norman, Donald (2002). *The Design of Everyday Things*. Tenth Edition. USA: Basic Books, Inc. ISBN: 978-0-465-06710-7 (cited on page 81).

Oliveira, Manuel M. (2002). 'Image-Based Modeling and Rendering Techniques: A Survey'. In: *RITA* 9.2, pp. 37–66 (cited on pages 34, 35).

Paine, David and James Kiser (2012). 'Geometry of a Vertical Aerial Photograph'. In: *Aerial Photography and Image Interpretation*. Third Edition. New Jersey: John Wiley & Sons, Ltd, pp. 25–43. ISBN: 978-1-118-11099-7. DOI: 10.1002/9781118110997.ch2 (cited on pages 18, 19, 64).

Paiz, Evelyn, Mathieu Brédif, and Sidonie Christophe (2020). 'Geometric Distortion of Historical Images for 3D Visualization'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-2-2020, pp. 649–655. ISSN: 2194-9050. DOI: 10.5194/isprs-annals-V-2-2020-649-2020 (cited on page 63).

Paiz-Reyes, Evelyn, Mathieu Brédif, and Sidonie Christophe (2021). 'Cluttering Reduction for Interactive Navigation and Visualization of Historical Images'. In: *Proceedings of the ICA* 4, pp. 81–87. ISSN: 2570-2092. DOI: 10.5194/ica-proc-4-81-2021 (cited on page 80).

Papadopoulos, Symeon, Christos Zigkolis, Stefanos Kapiris, Yiannis Kompatsiaris, and Athena Vakali (2010). 'ClustTour: City Exploration by Use of Hybrid Photo Clustering'. In: *Proceedings of the International Conference on Multimedia - MM '10*. Firenze, Italy: ACM Press, p. 1617. ISBN: 978-1-60558-933-6. DOI: 10.1145/1873951.1874302 (cited on pages 23, 50).

Pierrot-Deseilligny, Marc, Livio De Luca, and Fabio Remondino (2011). 'Automated Image-Based Procedures for Accurate Artifacts 3D Modeling and Orthoimage Generation'. In: *Geoinformatics FCE CTU* 6, pp. 291–299. ISSN: 1802-2669. DOI: 10.14311/gi.6.36 (cited on page 17).

Polidori, Laurent (2020). 'On Laussedat's Contribution to the Emergence of Photogrammetry'. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B2-2020, pp. 893–899. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLIII-B2-2020-893-2020 (cited on pages 13, 19).

Produit, Timothée, Nicolas Blanc, Sarah Composto, Jens Ingensand, and Loic Furhoff (2018). 'Crowdsourcing the Georeferencing of Historical Pictures'. In: *Proceedings of the Free and Open Source Software for Geospatial (FOSS4G) Conference*. DOI: 10.7275/awa4-4w48 (cited on pages 2, 27, 48, 54).

Pulli, Kari, Hugues Hoppe, Michael Cohen, Linda Shapiro, Tom Duchamp, and Werner Stuetzle (1997). 'View-Based Rendering: Visualizing Real Objects from Scanned Range and Color Data'. In: *Rendering Techniques '97*. Vienna: Springer, pp. 23–34. ISBN: 978-3-7091-6858-5. DOI: 10.1007/978-3-7091-6858-5_3 (cited on page 42).

Pulver, Yannick, Cédric Merz, Kathrin Koebel, Joe Scheidegger, and Arzu Çöltekin (2020). 'Telling Engaging Interactive Stories with Extended Reality (XR): Back to 1930s in Zurich's Main Train Station'. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* V-4-2020, pp. 171–178. ISSN: 2194-9050. DOI: 10.5194/isprs-annals-V-4-2020-171-2020 (cited on page 32).

Ray, Richard (1960). *Aerial Photographs in Geologic Interpretation and Mapping*. Report 373. U.S. Govt. Print. Off., pp. 1–4. DOI: 10.3133/pp373 (cited on page 18).

Read, Jane, Chad Chambers, and Marla Torrado (2020). 'Remote Sensing'. In: *International Encyclopedia of Human Geography*. Second Edition. Oxford: Elsevier, pp. 411–422. ISBN: 978-0-08-102296-2. DOI: 10.1016/B978-0-08-102295-5.10589-X (cited on page 18).

Reichmann, Michael (2006). *Making the Transition from Film to Digital*. San Jose, CA: Adobe, p. 12 (cited on page 12).

Remondino, Fabio, Isabella Toschi, Markus Gerke, Francesco Nex, David Holland, A. McGill, J. Talaya Lopez, and A. Magarinos (2016). 'Oblique Aerial Imagery for NMA-SOME Best Practices'. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B4, pp. 639–645. ISSN: 2194-9034. DOI: 10.5194/isprsarchives-XLI-B4-639-2016 (cited on page 19).

Rosenholtz, Ruth, Amal Dorai, and Rosalind Freeman (2011). 'Do Predictions of Visual Perception Aid Design?' In: *ACM Transactions on Applied Perception* 8.2, pp. 1–20. ISSN: 1544-3558, 1544-3965. DOI: 10.1145/1870076.1870080 (cited on page 43).

Roth, Robert (2012). 'Cartographic Interaction Primitives: Framework and Synthesis'. In: *The Cartographic Journal* 49.4, pp. 376–395. ISSN: 0008-7041, 1743-2774. DOI: 10.1179/1743277412Y.0000000019 (cited on pages 43, 81).

Roth, Robert (2017). 'Visual Variables'. In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Oxford: John Wiley & Sons, Ltd, pp. 1–11. ISBN: 978-1-118-78635-2. DOI: 10.1002/9781118786352.wbieg0761 (cited on pages xi, 47).

Rupnik, Ewelina, Mehdi Daakir, and Marc Pierrot Deseilligny (2017). 'MicMac: A Free, Open-Source Solution for Photogrammetry'. In: *Open Geospatial Data, Software and Standards* 2.1, p. 14. ISSN: 2363-7501. DOI: 10.1186/s40965-017-0027-2 (cited on pages 13, 71).

Schindler, Grant and Frank Dellaert (2010). 'Probabilistic Temporal Inference on Reconstructed 3D Scenes'. In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, CA, USA: IEEE, pp. 1410–1417. ISBN: 978-1-4244-6984-0. DOI: 10.1109/CVPR.2010.5539803 (cited on page 30).

Schindler, Grant and Frank Dellaert (2012). '4D Cities: Analyzing, Visualizing, and Interacting with Historical Urban Photo Collections'. In: *Journal of Multimedia* 7.2, pp. 124–131. ISSN: 1796-2048. DOI: 10.4304/jmm.7.2.124-131 (cited on pages 30, 48, 141).

Schindler, Grant, Frank Dellaert, and Sing Bing Kang (2007). 'Inferring Temporal Order of Images From 3D Structure'. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, MN, USA: IEEE. ISBN: 978-1-4244-1179-5. DOI: 10.1109/CVPR.2007.383088 (cited on page 30).

Segal, Mark, Carl Korobkin, Rolf van Widenfelt, Jim Foran, and Paul Haeberli (1992). 'Fast Shadows and Lighting Effects Using Texture Mapping'. In: *ACM SIGGRAPH Computer*

Graphics 26.2, pp. 249–252. ISSN: 0097-8930. DOI: 10.1145/142920.134071 (cited on page 35).

Seitz, Steven and Charles Dyer (1996). 'View Morphing'. In: *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*. New York, NY, USA: ACM Press, pp. 21–30. ISBN: 978-0-89791-746-9. DOI: 10.1145/237170.237196 (cited on pages 37, 38).

Semmo, Amir, Tobias Isenberg, and Jürgen Döllner (2017). 'Neural Style Transfer: A Paradigm Shift for Image-Based Artistic Rendering?' In: *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering - NPAR '17*. Los Angeles, California: ACM Press, pp. 1–13. ISBN: 978-1-4503-5081-5. DOI: 10.1145/3092919.3092920 (cited on pages 43, 141).

Shade, Jonathan, Steven Gortler, Li-wei He, and Richard Szeliski (1998). 'Layered Depth Images'. In: *Proceedings of the 25th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '98*. ACM Press, pp. 231–242. ISBN: 978-0-89791-999-9. DOI: 10.1145/280814.280882 (cited on page 42).

Shih, YiChang, Wei-Sheng Lai, and Chia-Kai Liang (2019). 'Distortion-Free Wide-Angle Portraits on Camera Phones'. In: *ACM Transactions on Graphics* 38.4, pp. 1–12. ISSN: 07300301. DOI: 10.1145/3306346.3322948 (cited on page 42).

Shum, Heung-Yeung and Li-Wei He (1999). 'Rendering with Concentric Mosaics'. In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '99*. ACM Press, pp. 299–306. ISBN: 0201485605. DOI: 10.1145/311535.311573 (cited on page 34).

Shum, Heung-Yeung and Sing Bing Kang (2000). 'A Review of Image-Based Rendering Techniques'. In: *IEEE Computer Graphics and Applications*, p. 12 (cited on pages 34, 35).

Simon, Ian, Noah Snavely, and Steven M. Seitz (2007). 'Scene Summarization for Online Image Collections'. In: *2007 IEEE 11th International Conference on Computer Vision*. Rio de Janeiro, Brazil: IEEE, pp. 1–8. ISBN: 978-1-4244-1630-1. DOI: 10.1109/ICCV.2007.4408863 (cited on page 50).

Sloan, Peter-Pike, Michael Cohen, and Steven Gortler (1997). 'Time Critical Lumigraph Rendering'. In: *Proceedings of the 1997 Symposium on Interactive 3D Graphics - SI3D '97*. Providence, Rhode Island, United States: ACM Press, 17–ff. ISBN: 978-0-89791-884-8. DOI: 10.1145/253284.253296 (cited on page 34).

Snavely, Noah, Rahul Garg, Steven Seitz, and Richard Szeliski (2008). 'Finding Paths Through the World's Photos'. In: *ACM Transactions on Graphics* 27.3, pp. 1–11. ISSN: 0730-0301. DOI: 10.1145/1360612.1360614 (cited on pages 25, 42).

Snavely, Noah, Steven Seitz, and Richard Szeliski (2006). 'Photo Tourism: Exploring Photo Collections in 3D'. In: *ACM Transactions on Graphics* 25.3, pp. 835–846. ISSN: 0730-0301. DOI: 10.1145/1141911.1141964 (cited on pages 2, 25, 26, 48, 54).

Snowden, Robert, Peter Thompson, and Tom Troscianko (2011). *Basic Vision: An Introduction to Visual Perception*. Revised Edition. OCLC: ocn656767547. Oxford: Oxford University Press. ISBN: 978-0-19-957202-1 (cited on pages ix, 44, 45).

Spindler, Martin, Marco Bubke, Tobias Germer, and Thomas Strothotte (2006). 'Camera Textures'. In: *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques - GRAPHITE '06*. Kuala Lumpur, Malaysia: ACM Press, p. 295. ISBN: 978-1-59593-564-9. DOI: 10.1145/1174429.1174480 (cited on pages 41, 68).

Steadman, Philip (2002). 'The Camera Obscura'. In: *Vermeer's Camera: Uncovering the Truth Behind the Masterpieces*. Oxford University Press, pp. 4–24. ISBN: 978-0-19-280302-3 (cited on page 10).

Stefani, Chiara, Chawee Busayarat, Julie Lombardo, Livio De Luca, and Philippe Véron (2013). 'A Web Platform for the Consultation of Spatialized and Semantically Enriched Iconographic sources on Cultural Heritage Buildings'. In: *Journal on Computing and Cultural Heritage* 6.3, pp. 1–17. ISSN: 1556-4673, 1556-4711. DOI: 10.1145/2499931.2499934 (cited on page 30).

Stich, Timo, Christian Linz, Georgia Albuquerque, and Marcus Magnor (Oct. 2008). 'View and Time Interpolation in Image Space'. In: *Computer Graphics Forum* 27.7, pp. 1781–1787. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/j.1467-8659.2008.01323.x (cited on page 38).

Szeliski, Richard and Heung-Yeung Shum (1997). 'Creating Full View Panoramic Image Mosaics and Environment Maps'. In: *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques - SIGGRAPH '97*. ACM Press, pp. 251–258. ISBN: 978-0-89791-896-1. DOI: 10.1145/258734.258861 (cited on page 34).

Szirmay-Kalos, László, László Szécsi, and Mateu Sbert (2008). 'Local Illumination Rendering Pipeline of GPUs'. In: *GPU-Based Techniques for Global Illumination Effects*. Synthesis lectures in computer graphics and animation. Morgan & Claypool, pp. 23–41. ISBN: 978-1-59829-559-7 (cited on page 68).

Thurley, Simon (2005). 'Into the Future: Our Stategy for 2005-2010'. In: *Conservation Bulletin*. Vol. 49. English Heritage. London, pp. 26–27 (cited on page 8).

Treue, Stefan (2003). 'Visual Attention: The Where, What, How and Why of Saliency'. In: *Current Opinion in Neurobiology* 13.4, pp. 428–432. ISSN: 09594388. DOI: 10.1016/S0959-4388(03)00105-3 (cited on page 46).

Triggs, Bill, Philip McLauchlan, Richard Hartley, and Andrew Fitzgibbon (2000). 'Bundle Adjustment — A Modern Synthesis'. In: *Vision Algorithms: Theory and Practice*. Vol. 1883. Berlin, Heidelberg: Springer, pp. 298–372. ISBN: 978-3-540-44480-0. DOI: 10.1007/3-540-44480-7_21 (cited on page 16).

Vass, Gergely and Tamás Perlaki (2003). *Applying and Removing Lens Distortion in Post Production*. Tech. rep. Budapest, Hungary: Colorfront Ltd. (cited on pages 40, 79, 141).

Verhoeven, Geert (2016). 'Basics of Photography for Cultural Heritage Imaging'. In: *3D Recording, Documentation and Management of Cultural Heritage*. Whittles Publishing, pp. 127–251. ISBN: 978-1-84995-168-5 (cited on pages 9–11).

Vries, Joey (2020). 'Shaders'. In: *Learn OpenGL: Graphics Programming*. Kendall & Welling, pp. 25–53. ISBN: 978-90-90-33256-7 (cited on page 58).

Wehrwein, Scott, Kavita Bala, and Noah Snavely (2015). 'Shadow Detection and Sun Direction in Photo Collections'. In: *2015 International Conference on 3D Vision*. Lyon, France: IEEE, pp. 460–468. ISBN: 978-1-4673-8332-5. DOI: 10.1109/3DV.2015.58 (cited on page 25).

White, Sean and Steven Feiner (2009). 'SiteLens: Situated Visualization Techniques for Urban Site Visits'. In: *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*. Boston, MA, USA: ACM Press, p. 1117. ISBN: 978-1-60558-246-7. DOI: 10.1145/1518701.1518871 (cited on page 32).

Willis, Kenneth (2014). 'The Use of Stated Preference Methods to Value Cultural Heritage'. In: *Handbook of the Economics of Art and Culture*. Vol. 2. Elsevier, pp. 145–181. DOI: 10.1016/B978-0-444-53776-8.00007-6 (cited on page 8).

Wolf, Paul, Bon Dewitt, and Benjamin Wilkinson (2014a). 'Introduction'. In: *Elements of Photogrammetry with Applications in GIS*. Fourth Edition. New York: McGraw-Hill Education, pp. 1–16. ISBN: 978-0-07-176112-3 (cited on pages 17–19).

Wolf, Paul, Bon Dewitt, and Benjamin Wilkinson (2014b). 'Terrestrial and Close-Range Photogrammetry'. In: *Elements of Photogrammetry with Applications in GIS*. Fourth Edition. New York: McGraw-Hill Education, pp. 485–509. ISBN: 978-0-07-176112-3 (cited on page 20).

Wu, Bo, Linfu Xie, Han Hu, Qing Zhu, and Eric Yau (2018). 'Integration of Aerial Oblique Imagery and Terrestrial Imagery for Optimized 3D Modeling in Urban Areas'. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 139, pp. 119–132. ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2018.03.004 (cited on page 20).

Xu, Rui and Donald C Wunsch (2015). *Clustering*. OCLC: 1164479728. John Wiley & Sons. ISBN: 978-0-470-38277-6 (cited on pages ix, 49).

Zitnick, Larry, Sing Bing Kang, Matt Uyttendaele, Simon Winder, and Rick Szeliski (2004). 'High-Quality Video View Interpolation Using a Layered Representation'. In: *ACM SIGGRAPH*. Vol. 23. Association for Computing Machinery, Inc., pp. 600–608. DOI: 10.1145/1186562.1015766 (cited on page 38).

List of Publications

Full Paper Reviewed

Evelyn Paiz, Mathieu Brédif, and Sidonie Christophe (2020). 'Geometric Distortion of Historical Images for 3D Visualization'. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* V-2-2020, pp. 649–655. DOI: https://www.doi.org/10.5194/isprs-annals-V-2-2020-649-2020.

Extended Abstract Reviewed

Evelyn Paiz, Mathieu Brédif, and Sidonie Christophe (2021). 'Cluttering Reduction for Interactive Navigation and Visualization of Historical Images'. In: *Proceedings of the ICA* 4, pp. 81–87. DOI: https://www.doi.org/10.5194/ica-proc-4-81-2021.

Extended Abstract Reviewed with no Proceedings

Evelyn Paiz (May 2019).Image Based Rendering of Large Historical Image Collections. *Eurographics Doctoral Consortium*. Genova, Italy. URL: https://hal.archives-ouvertes.fr/hal-02301561

Communications

Evelyn Paiz (Feb. 2020). Image Based Rendering of Large Historical Image Collections. *Webinar 3D GDR MAGIS*. Online. URL: https://github.com/VCityTeam/MAGIS-AP3D/blob/master/Media/README.md

Evelyn Paiz (Oct. 2020). Geometric Distortion of Historical Images for 3D Visualization. *Journées de la Recherche IGN*. Online. URL: http://recherche.ign.fr/jr/jr20.php

List of Figures

1.1 Virtual View of a 3D Model and a Historical Photo 4

2.1 Cultural Heritage 8
2.2 Heritage Cycle 8
2.3 Pinhole Camera 9
2.5 Primary Components Inside a Camera 10
2.4 *View from the Window at Le Gras* 10
2.6 Exposure Triangle 11
2.7 Comparison Between Analog and Digital Photography 12
2.9 Digitized Image Sample 13
2.8 Process of Digitization 13
2.10 Process of Photogrammetry 14
2.11 Pinhole Camera Model (Geometric Representation) 14
2.12 Comparison Between Sparse and Dense Point Clouds 16
2.13 Aerial Photography 17
2.15 Camera Orientation for Vertical Aerial Photographs 18
2.14 Captured Ground Area of Aerial Imagery 18
2.16 Camera Orientation for Oblique Aerial Photographs 19
2.17 Camera Orientation for Terrestrial Photographs 20
2.18 Image Gallery 21
2.19 Functions for Geovisualization. 22
2.20 Image Gallery Along with a 2D Map Model 23
2.21 Co-Visualization of Terrestrial Historical Image and *Google Street View* 24
2.22 3D Model and Spatially Placed Image 25
2.23 Co-Visualization of a Terrestrial Image and a Reconstructed Sparse 3D Point Cloud . . . 26
2.24 Co-Visualization of a Street-View Photo and a Reconstructed 3D Model 26
2.25 Co-Visualization of an Aerial Picture and a Contemporary 3D Geographical Environment 27
2.26 Spatio-Temporal Triangle. 27
2.27 Image Gallery Along with a Synchronized Spatio-Temporal View 28
2.28 Spatio-Temporal Interface 29
2.29 Co-Visualization of a Historical Photograph and a Time-Evolving 3D City Model. . . . 29
2.30 Time-Varying 3D Model 30
2.31 Virtual Reality Environment for Photo Exploration 31
2.32 Virtual Reality System for a Museum 31
2.33 Mixed Reality Mobile Application 32
2.34 Augmented Reality Annotations Platform 33
2.35 Different Rendering Equations 34
2.36 Image-Based Rendering Techniques 34
2.37 Projective Texturing 35
2.38 Real-Time Projective Multi-Texturing 36

2.39 Interpolation Between Images . . . 36
2.40 View Interpolation . . . 37
2.41 View Morphing . . . 37
2.42 Video View Interpolation . . . 38
2.43 Uncertainty in the Projection with Approximate Scene Geometry . . . 39
2.44 Silhouette-Aware Filtered Blending . . . 39
2.45 Image Transition Using Ambient Point Cloud . . . 40
2.46 Image Mapping . . . 40
2.47 Deformed Grid . . . 41
2.48 Panorama View Generated Using Dynamic Mesh Refinement . . . 41
2.49 Lighting Changes . . . 42
2.50 Spatio-Temporal Interpolation . . . 43
2.51 Human Eye . . . 44
2.52 Visual Perception Process . . . 45
2.53 Bottom-Up and Top-Down Processing . . . 46
2.55 Visual Salience Comparison . . . 46
2.54 Scan Path of a Viewer . . . 46
2.56 Primary Visual Variables . . . 47
2.57 3D Bookmarks . . . 48
2.58 Thumbnail Slider Menu . . . 49
2.59 Clustering of Photos in a 2D Map . . . 49
2.60 Scalable Insets . . . 50
2.61 Orientations from Spatialized Images . . . 50

3.1 Evolution of the Main Photo Visualization Techniques . . . 53
3.2 Illustration of the Virtual View of a 3D Environment Textured by an Unprojected Old Photograph . . . 55
3.3 Many Historical Images Located in a 3D Scene. . . . 56
3.4 Objectives of Research . . . 57
3.5 Shading Process . . . 58
3.6 Rendering Spectrum . . . 59
3.7 Geometrical Components in an Image Projection . . . 59
3.8 Historical and View Camera . . . 60

4.1 Overview of Different Camera Models . . . 64
4.3 Depiction of an Ideal and Distorted Image Point . . . 65
4.2 Aerial Photograph with a Strong Distortion . . . 65
4.4 3D City Model and Projected 2D Image . . . 66
4.5 Rendering with Distortion . . . 67
4.6 One-Pass Rendering . . . 68
4.7 Two-Pass Rendering . . . 69
4.8 Extrapolating Distortion Models for 3D Visualization . . . 70
4.9 Sample Behavior of the Distortion Model . . . 72
4.10 Use of Different Camera Models for Projective Texturing . . . 75
4.11 Extrapolation Radius . . . 76

4.12 Extrapolation Radius for View Interpolation . . . 77
4.13 One-Pass vs. Two-Pass Rendering . . . 78
4.14 Visualization of a Photograph in Two Different 3D Scene Models . . . 78

5.1 Cartographic Interactions . . . 81
5.2 Immersiveness Spectrum . . . 82
5.3 3D City Model and Various Projected 2D Images . . . 82
5.4 Task 1: Visualizing One or Several Images . . . 84
5.5 Task 2: Exploring Images in Space and Time . . . 84
5.6 Our Bookmark . . . 86
5.7 Three Designs for a Bookmark . . . 86
5.8 Bookmark Icons . . . 87
5.9 Designs for the Placing of Thumbnails . . . 88
5.10 Clustered Thumbnail Representation . . . 88
5.11 Our Visual Variables and Their Usage . . . 89
5.12 Our Heat Map to Identify Areas of Interest . . . 90
5.13 Our Style Representation for 3D Models . . . 91
5.14 Our Time Axis with Histogram Representation . . . 92
5.15 Comparison of Bookmarks . . . 94
5.16 Thumbnail Placements Around the 3D Scene . . . 95
5.17 Colored Frame . . . 96
5.18 Use of Style Representation for the Radiometry of the 3D Models. . . . 97
5.19 Heatmaps. . . . 97
5.20 Histogram Representations for Timeline . . . 98

6.1 Heat Maps of Geographic Locations of Historical Dataset . . . 101
6.2 MRU Photo Collection Sample . . . 102
6.3 Lapie Photo Collection Sample . . . 103
6.4 Combier Photo Collection Sample . . . 103
6.5 Henrard Photo Collection Sample . . . 104
6.6 Charles Gros Photo Collection Sample . . . 104
6.7 BD ORTHO® . . . 105
6.8 Cartographic Representation of the BD TOPO® . . . 106
6.9 Cartographic Representation of the BD TOPO® . . . 106
6.10 Image States and Triggering User Interactions . . . 107
6.11 Class Diagram Modeling a Historical Image . . . 108
6.12 General Architecture . . . 110
6.13 UML Class Diagram . . . 112
6.14 UML Package Diagram . . . 113
6.15 GUI of HISTOVIS when System Just Launched . . . 114
6.16 UML Object Diagram (Init State) . . . 115
6.17 UML Object Diagram (Loaded Dataset) . . . 115
6.18 GUI of HISTOVIS . . . 116
6.19 Viewpoint Markers inside HISTOVIS . . . 118
6.20 Thumbnails Around the 3D View of HISTOVIS . . . 119

6.21 Distortion Model Applied on the View and Historical Camera Inside HISTOVIS 120
6.22 Frames Around the Projected Photographs in HISTOVIS 121
6.23 Projected Heat Map in HISTOVIS 122
6.24 Style Representation inside HISTOVIS 123
6.25 Timeline Next to 3D View in HISTOVIS 124
6.26 Use Case Scenario 1 125
6.27 Use Case Scenario 2 127
6.28 Use Case Scenario 3 129
6.29 Graphical Results on the Temporal Selection Task 131
6.30 Graphical Results on the Bookmark Interaction Task 132
6.31 Graphical Results on the Color Differentiation Task 133
6.32 Graphical Results on the Clustering Spatialized Thumbnails Task 135

A.1 Mock-Up 1: Selecting a Time Frame 143
A.2 Mock-Up 2: Interacting with Bookmarks 144
A.3 Mock-Up 3: Exploring through Space 145
A.4 Mock-Up 4: Co-Visualizing 2D Images and 3D Objects 146

B.1 Test Protocol: Loading a Photo Collection in HISTOVIS 147
B.2 Test Protocol: The Timeline Segment in HISTOVIS 148
B.3 Test Protocol: Interaction with Bookmarks in HISTOVIS 150
B.4 Test Protocol: Differentiation of Photo Collections in HISTOVIS. 151
B.5 Test Protocol: Thumbnails in HISTOVIS. 152

List of Tables

4.1 Datasets Used for Testing Our Approaches 75

6.1 ALEGORIA project datasets 101
6.2 Inputs of the HISTOVIS Prototype 109
6.3 Experimental Datasets for Prototype 124

C.1 Results 1: Timing of Test Users 154
C.2 Results 2: Additional Comments from Test Users 154
C.3 Results 3: Temporal Selection Task 155
C.4 Results 4: Interaction with Bookmarks Task 156
C.5 Results 5: Use of Frame and Color for Differentiation Task 157
C.6 Results 6: Thumbnails and Clustering task 158

