Design Space for Visualizing Multimedia Documents in Spatio-Temporal-Thematic Urban Context

Akin Gulfidan[1], Pierre-Antoine Chiron[2], John Samuel[2,3], Valérie Gouet-Brunet[4], Gilles Gesquière[2]

[1] École Supérieure d'art et design, Saint-Etienne, France; akin.gulfidan@etudiant.esadse.fr
[2] Universite Claude Bernard Lyon 1, CNRS, INSA Lyon, Université Lumière Lyon 2, Lyon, France; (pierre-antoine.chiron, gilles.gesquiere)@univ-lyon2.fr
[3] CPE Lyon, Villeurbanne, France; john.samuel@cpe.fr
[4] Univ Gustave Eiffel, Géodata Paris, IGN, LASTIG, France; valerie.gouet@ign.fr

Keywords: Multimedia documents, 3D city model, Spatio-temporal-thematic structuring, Design-driven visualization, Iterative prototyping.

Abstract

Multimedia documents such as photographs, videos, texts, and web pages are essential for understanding how urban environments evolve, but exploring them in relation to a 3D city model remains difficult because spatial, temporal, and thematic relations must be handled simultaneously. This paper investigates how a design space for such visualizations can be constructed without reducing either the urban context or the informational richness of the documents. We adopt a design-driven methodology based on iterative prototyping: visualization propositions are developed from initial requirements, tested through simulation, confronted with user needs, and refined as new needs and limitations emerge. The paper first structures the corpus in a spatio-temporal-thematic model, then formalizes the main culling and display functions required for exploration. On that basis, we derive a family of complementary visualization propositions for tasks including spatial selection, temporal filtering, georeferenced media display, non-georeferencable document organization, link visualization, and document quantity analysis. Rather than proposing a single interface, the contribution is a structured design space and a set of concrete visualization propositions for multimedia exploration in 3D urban environments.

1. Introduction

Multimedia documents such as photographs, videos, text, and web pages are essential sources for understanding the transformation of urban space (Rose, 2013; Bruschke et al., 2017; Jaillot et al., 2021). Yet these documents are heterogeneous, unevenly georeferenced, and linked to city objects at multiple scales (Silva et al., 2022). Bringing them into a single interactive environment therefore raises a central visualization problem: how can one structure a design space for exploring a large multimedia corpus in relation to a 3D city model without losing either spatial context or the temporal and thematic relations between documents?

This paper addresses that problem by constructing a design space for visualizing multimedia documents in a 3D spatio-temporal-thematic context. Our objective is not to propose a single universal interface, but a coherent set of complementary visualizations that support exploration of documents, city geometry, and their relations. Working in 3D is especially relevant here because many interpretive questions depend on orientation, field of view, geometric hierarchy, and the position of documents relative to buildings and urban structures.

This research is developed within the ANR AGAPE project (ANR-24-CE38-7253)[1], an interdisciplinary collaboration bringing together seven partners specialized in computer science, artificial intelligence, multimedia, and human-computer interaction (LASTIG, LIRIS, Linkmedia, and ILDA), as well as in archives (Archives nationales de France), history (ACP), and media (France Télévisions). In such a context, the challenge is not only to visualize heterogeneous data, but also to articulate heterogeneous disciplinary questions, documentary practices, and expectations about how multimedia corpora should be explored in urban space.

The proposed visualizations are designed to address five main challenges:

- representing different document types while preserving the visual or semantic properties that make them informative;
- selecting relevant subsets of the corpus without severing the relations between visible and hidden elements;
- combining spatial, temporal, and thematic organization within views that remain readable;
- distributing information across multiple coordinated interfaces while maintaining cross-view links;
- designing exploratory views that reveal the structure and density of the corpus itself.

To answer these challenges, we distinguish two families of requirements. The first concerns *culling*: the user must be able to retrieve relevant subsets of documents according to city objects, time periods, or thematic criteria. Showing the entire corpus at once would quickly produce visual overload, especially in a 3D environment where screen space and stable viewpoints are limited. The second concerns *display*: once a subset has been selected, the interface must reveal the information necessary to interpret it, such as dates, view directions, links between documents, or the amount of documentation associated with a place.

Because no single representation can satisfy all of these requirements simultaneously, our approach is intentionally design-driven and modular. We define a set of visualization functions,

[1] AGAPE project website: https://agape-anr.github.io/

then develop and compare a series of propositions that combine these functions in different ways. Rather than optimizing a single interface, we seek to identify which visualization principles are most appropriate for which analytical tasks.

Methodologically, the work follows a research-through-design perspective, in which designing and reflecting are advanced together through iterative making and evaluation (Zimmerman et al., 2007). In that sense, the article adopts a “do and think” approach: we rapidly produced simulation prototypes, discussed and tested them across disciplinary perspectives, and refined or replaced them when new needs emerged. This process progressively led to a family of propositions, each addressing a specific subset of the overall problem while helping to delineate the overall design space.

Main Contributions. The main contributions of this paper are threefold. First, building on the spatio-temporal document model of Jaillot et al. (2021), we formalize the visualization problem through a distinction between culling and display functions; this functional decomposition and the resulting propositions are original to this work. Second, we derive from this structure a coherent set of visualization propositions covering complementary exploration tasks, from spatial selection and temporal filtering to georeferenced media placement and non-georeferencable document organization. Third, we show through iterative prototyping how these propositions progressively delineate a design space for multimedia exploration in 3D urban environments and reveal directions for future integration and evaluation.

Paper Organization. The remainder of the paper is organized as follows. Section 2 positions the work with respect to existing research on media visualization in 3D urban contexts and related visualization practices. Section 3 introduces the spatio-temporal-thematic structuring adopted for the corpus. Section 4 formalizes the culling and display functions that guide the design. Section 5 presents the visualization propositions. Section 6 discusses the main contributions, limitations, and extensions of the approach, and Section 7 concludes.

2. Related Works

Research most directly related to our work lies at the intersection of 3D urban visualization, historical media positioning, and multimedia analysis. Chagnaud et al. (2016) addressed the problem of documented 3D cities and proposed strategies for reducing overlap between images in urban scenes, notably through level-of-detail-based organization. Their work clearly establishes the importance of scale in media visualization. However, some of the proposed arrangements alter document placement in order to maintain readability. Our work follows the same concern for scale, but seeks to preserve georeferenced positions when possible and to treat level of detail as a filtering principle rather than only a layout strategy.

Before such 3D approaches, many spatial humanities and public-history projects already relied on map-based access to documents. In the broader context of the spatial turn, positioning photographs or testimonies on a 2D map made them easier to retrieve, compare, and interpret in relation to place. Platforms and services such as Historypin[2] or Mapillary[3] helped establish this logic of spatial access at large scale (Historypin, 2026; Mapillary, 2026). Our work can be seen as a continuation of these approaches, but in a setting where the third dimension, geometric hierarchy, and viewpoint become analytically important rather than optional.

The spatial integration of visual documents and information in 3D environments has also been explored by several works (Münster et al., 2017; Blettery et al., 2020; Lecat et al., 2021; Silva et al., 2022; Antoun et al., 2025). These works demonstrate the analytical value of placing historical photographs in the scene with their orientation, and in some cases their immersive alignment with reconstructed geometry. There is an important distinction between geolocating the depicted scene and geolocating the camera itself. The first helps identify the urban area represented by the document; the second gives access to the acquisition position and orientation, which can in turn reveal framing choices, lines of sight, and even recurrent photographer behaviour. Our propositions for georeferenced images and videos build directly on this line of work, while extending it toward a broader multimedia system that also includes selection, filtering, linking, and document quantity visualization, some of which are surveyed by Miranda et al. (2024).

Jaillot et al. (2021) proposed a framework for linking multimedia documents to time-evolving 3D city models. Their work is especially relevant because it formalizes relations between documents, city geometry, and temporal states. In this paper, the term *city model* refers to structured urban representations such as CityGML-based building models (Kutzner et al., 2022), time-evolving urban reconstructions, or documented heritage models used for navigation and analysis (Chagnaud et al., 2016; Jaillot et al., 2021; Biljecki et al., 2016). We adopt a compatible perspective, but shift the emphasis from data integration to visualization design. In other words, our contribution is not a new data model for urban cultural heritage, but a set of coordinated visualization propositions for exploring such a model once the links are available.

Our design decisions were also informed by work outside the specific domain of urban cultural heritage. In cultural analytics, Hochman and Manovich (2013) showed that plotting images as images rather than abstract symbols can reveal visual regularities that would otherwise remain hidden. This principle is valuable in our context because the appearance of documents is often analytically relevant: visual differences between archival and contemporary imagery, or between photographs taken from recurrent viewpoints, may themselves constitute meaningful evidence. For that reason, several of our propositions preserve documents as visible media objects rather than reducing them to points. More broadly, cartographic and information-design practices (Gautier et al., 2023) also remind us that the urban model must support, rather than compete with, the media layers; recent work on large-scale video-augmented geographic scenes confirms the continuing relevance of this problem for geo-visualization (Chen et al., 2024).

Finally, some of our propositions connect with photogrammetric reconstruction methods. Structure from Motion (SfM) techniques (Ozyesil et al., 2017), as implemented for example in Meshroom, make it possible to derive point clouds and 3D structure from overlapping photographs. This is relevant to our video propositions because georeferenced image sequences can generate additional spatial information that is itself visualizable alongside the city model. The point-cloud extension considered

[2] Historypin website: https://www.historypin.org/
[3] Mapillary website: https://www.mapillary.com/

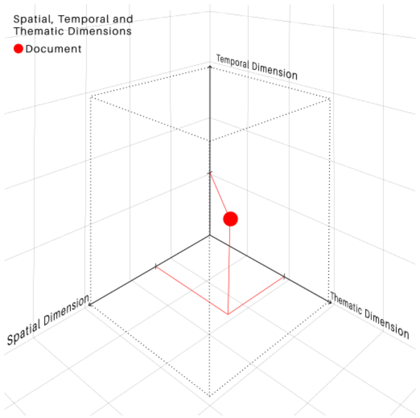


Figure 1. Documents are organized in a spatio-temporal-thematic model following Jaillot et al. (2021).

later in the paper therefore relies on a well-established reconstruction pipeline rather than on an ad hoc representation.

Taken together, these related works point to an open gap. Existing research has studied georeferenced image placement, the linking of multimedia documents to city models, or large-scale media visualization, but fewer works propose a unified design space spanning georeferenced and non-georeferencable documents, culling and display functions, and coordinated views for 3D urban exploration. The present paper aims to fill that gap through a structured set of visualization propositions.

3. Structuring Multimedia Documents

Our corpus contains heterogeneous documents such as images, videos, text, and web pages. Following Jaillot et al. (2021), we describe them through three dimensions: spatial, temporal, and thematic. This overall organization is summarized in Figure 1.

3.1 Spatial Dimension

The spatial dimension links each document to the smallest relevant city object, called a *building component*. In our current corpus, this lowest stable level is typically a wall, facade segment, doorway, or other persistent part of a building envelope. We privilege such built elements because they remain identifiable across time and across documentary sources. In principle, other objects such as trees or street furniture could also be documented, but they are less stable, less systematically modeled, and therefore less suitable as a common anchor for the whole corpus. Documents attached to components can then be aggregated upward to buildings, blocks, or larger urban entities (Jaillot et al., 2021). For example, a photograph of a painted shopfront can be linked to one facade segment and later retrieved either at the facade level or at the scale of the whole building.

3.2 Temporal Dimension

The temporal dimension places documents and city objects on a timeline. A document may be associated with a single date or with a time interval. This is important because urban analysis often compares states rather than isolated events. For example, a planning report may describe a renovation campaign from 1972 to 1975, while a photograph may document only one moment within that interval.

3.3 Thematic Dimension

The thematic dimension groups documents based on semantic similarity, defined here through manually assigned or automatically extracted descriptors such as topics, documented urban processes, object types, events, or keywords. Documents with related content lie close to each other even when they differ in date or spatial attachment. For example, photographs, reports, and video excerpts dealing with demolition may belong to the same thematic neighborhood even if they concern different streets or different years. Together, the three dimensions describe where a document is linked, when it belongs, and what it is about (Chen et al., 2025).

4. Functionalities of the Visualization

We distinguish two groups of functions: **culling functions**, which filter the corpus, and **display functions**, which reveal selected information. Following the space–time partitioning principles discussed by Bach et al. (2014), these functions operate on the spatial, temporal, and thematic dimensions introduced above. For readability, each function is identified by the prefix **F** (for example, F1, F2, . . .), while the visualization propositions introduced in the next section are identified by the prefix **P** (P1, P2, . . .). Figure 2 illustrates the main culling operations in the spatio-temporal-thematic space, Figure 3 shows the distribution of culling and display functions by dimension.

4.1 Culling Functions

Not every culling function has an equivalent on every dimension because the three dimensions do not describe the same kind of object. The spatial dimension is organized around identifiable entities and hierarchies, which makes object-based selection possible. The temporal dimension is organized around dates and intervals, which is why it is more naturally filtered than individually "picked" in the same way as a building component. The thematic dimension expresses semantic proximity rather than explicit geometric or chronological boundaries, which makes thematic filtering fundamentally different again.

F1: Select an Individual Building Component Retrieves all documents linked to one building component. This function acts only on the spatial dimension because it requires a discrete urban object that can serve as an anchor.

F2: Select a Building Component Hierarchy Extends F1 from a single component to all components belonging to one building hierarchy. For example, selecting one building retrieves documents linked to its walls, openings, or other associated components.

F3: Select Multiple Component Hierarchies Selects several buildings at once and therefore several component hierarchies in the spatial dimension. This function also covers selection within a spatial zone or perimeter when that area contains multiple buildings or components to retrieve together.

F4: Select a Single Document Selects one document as a point in the spatio-temporal-thematic space.

F5: Select Multiple Documents Extends F4 to a set of documents distributed across the three dimensions.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XII-4-W1-2026-179-2026 | © Author(s) 2026. CC BY 4.0 License.

(a) F1 (b) F2 (c) F3

(d) F4 (e) F5 (f) F6 (g) F2 + F6

Figure 2. Culling functions illustrated in the spatio-temporal-thematic space.

F6: Filter Documents by Time Period Filters documents within a temporal interval and therefore acts only on the temporal dimension. Unlike F2, which expands a spatial hierarchy, F6 defines a continuous interval with start and end bounds.

F7: Filter in the Thematic Dimension Filters documents according to semantic proximity or theme. Culling functions can also be combined, for example to retrieve documents attached to one building during a selected time interval.

4.2 Display Functions

F8: Display Document Dates Displays document dates on a timeline and therefore acts on the temporal dimension.

F9: Display the Document in Geographical Context Shows georeferenced documents in the 3D city model, linking media content to its broader spatial context such as the surrounding street, facade, or block.

F10: Display Non-Georeferencable Documents Displays documents without direct spatial placement through a media-based organization rather than geometric position, for example as thumbnails, grouped panels, or ordered clusters based on type, date, or theme.

F11: Display Links between Documents Shows metadata-based relationships between documents in the thematic dimension.

F12: Display Links between a Document and Associated City Objects Shows the explicit relation between a document and the city objects to which it is attached. Compared with F9, which emphasizes surrounding spatial context, F12 focuses on the attachment link itself, for example between one photograph and the facade, wall, or building component it documents.

F13: Display City Geometry Displays the city model itself as the spatial support for exploration.

F14: Display Georeferenced Image View Direction Displays the orientation of georeferenced images, typically as a position-and-direction marker or viewing vector, to reveal how the scene was captured.

F15: Display Georeferenced Image Field of View Extends F14 by showing the visible area of an image, making overlaps between documents explicit. We distinguish F14 and F15 only for images and videos because camera pose, viewing direction, and field of view are specific to capture devices; the same kind of specialization is not meaningful for a text or a generic web page.

F16: Visualize Document Quantity Measures how strongly city objects are documented by counting attached documents along the geometric hierarchy.

5. Visualization Propositions

All propositions were prototyped in *Visual Language (VL)* within the *VVVV* IDE,[4] using a lightweight metadata table that stores document identifiers, type, date or interval, thematic tags, camera pose when available, and links to CityGML building or

[4] VVVV documentation: https://thegraybook.vvvv.org/

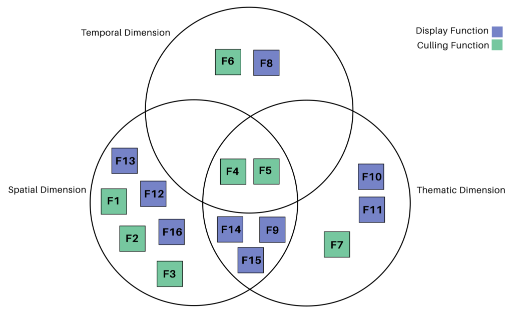
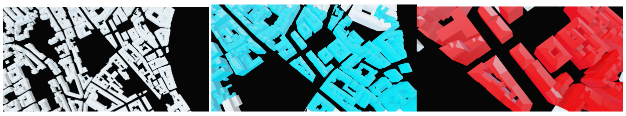


Figure 3. Distribution of culling and display functions by dimension.

component identifiers. The prototype queries this table in memory to apply the culling functions and updates the coordinated 3D, timeline, and document views; the 3D buildings are derived from the CityGML dataset for Lyon. Lyon is used as a test case because an existing CityGML urban model and reusable open multimedia sources are available, but the propositions depend only on generic document metadata and city-object identifiers and are therefore transferable to other CityGML-based or similarly structured 3D city models. Table 1 synthesizes the functions they address.

5.1 P1: LOD Filtering

P1 uses camera distance as a proxy for the active level of detail (LOD), allowing documentary scale to be filtered without moving documents from their georeferenced positions (Figure 4). It addresses F7 in the specific case where theme is understood as documentary scale; its main limit is that links to hidden LODs remain implicit.



(a) City level (b) District level (c) Building level

Figure 4. LOD filtering. Camera distance drives the visible thematic level.

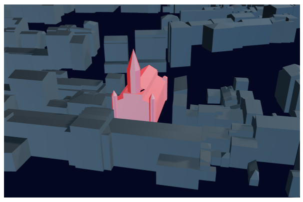
5.2 P2: Building Selection

Selection in the 3D viewer supports F1, F2, and F3 through three complementary interactions: whole-building selection, component picking by ray casting, and perimeter-based area selection. A first test used direct color change on the geometry, but this conflicted with textures and later color-based encodings, so we retained a distance-adaptive bounding-box highlight (Figure 5).

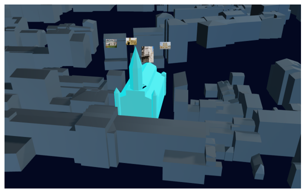
P2 responds to F1, F2, and F3. Its strength is to unify building, component, and area selection in one visual language, as illustrated by Figure 5; its main limit is the difficulty of selecting occluded components directly in 3D, so it should be complemented by non-spatial selection methods such as search or thematic filters.

5.3 P3: Building Component Tree

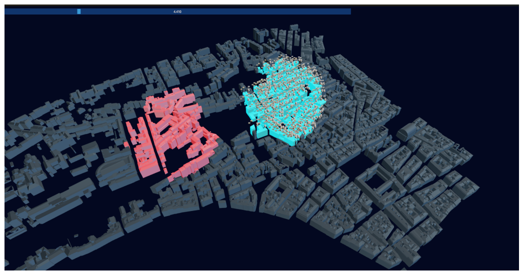
The component tree complements 3D selection by exposing the building hierarchy in a dedicated adjacent view. Tree-map, floor-plan, and 3D-integrated variants were tested (Figure 6). P3 responds to F1 and F13 by making hidden hierarchies explicit, but deep descendants remain hidden until their parents are opened.



(a) Hover by color

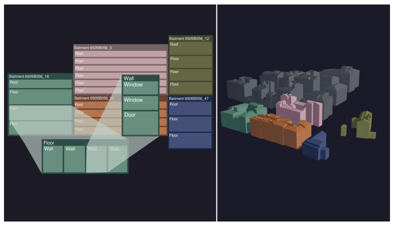


(b) Selection by color

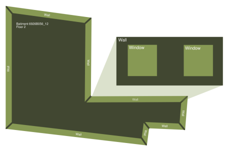


(c) Area cursor

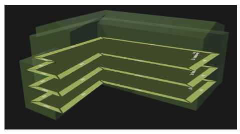
Figure 5. Building selection alternatives: from material-dependent color changes to distance-adaptive bounding-box highlighting and area selection.



(a) Tree-map



(b) Floor-plan variant



(c) Inserted in 3D

Figure 6. Hierarchical navigation alternatives for building components.

5.4 P4: Dome Timeline Visualization

P4 explores a first way to combine time with georeferenced images by projecting them onto a dome or sphere whose radius corresponds to time (Figure 7). It responds to F4, F8, F9, F13, and F14 by revealing repeated viewpoints and temporal clusters, but it distorts geographic position and becomes ambiguous when several buildings are involved. This exploratory proposition motivated the later separation between true spatial placement in 3D and a dedicated timeline.

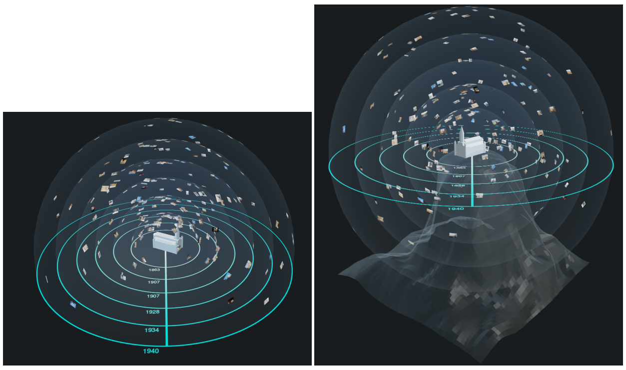
The dome keeps only the upper hemisphere, which reinforces the idea of a grounded observation space and avoids sending images below the urban surface. The sphere offers a more symmetric distribution around the center, but it is less intuitive in relation to gravity and to the built environment. Recent images were placed farther from the center so that older documents remain close to the observed place and can be read as a historical core, while newer documents expand outward as later additions.

5.5 P5: Georeferenced Image Display

P5 keeps images at their true locations in the 3D city model and delegates time to the timeline. Two readings are possible: an

Proposition	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16
P1							✓									
P2	✓	✓	✓													
P3	✓												✓			
P4				✓				✓	✓				✓	✓		
P5				✓					✓				✓	✓	✓	
P6					✓	✓		✓								
P7				✓					✓				✓	✓	✓	
P8										✓						
P9				✓			✓	✓		✓	✓	✓				
P10																✓

Table 1. Summary of propositions and the functions they address.



(a) Dome view (b) Sphere view

Figure 7. Dome-timeline concept for georeferenced images.

external view for spatial context and an immersive overlay that compares the photograph directly with the reconstructed scene (Figure 8).

(a) External view (b) Immersive view

Figure 8. Two complementary readings of a georeferenced image.

In the immersive overlay, the observer is placed at the optical center of the camera so that the image can be compared with the reconstructed environment from the acquisition viewpoint. P5 responds to F4, F9, F13, F14, and F15. Its main strength is accurate spatial placement with explicit direction and frustum cues; its main limitation is that temporal reading still depends on a separate timeline.

5.6 P6: Timeline

The timeline addresses F6 and F8. It is displayed as a dedicated view adjacent to the 3D scene, showing dates, interval selection, and scale-dependent aggregation when the visible temporal range becomes too large (Figure 9). In this view, point color encodes document type, while zooming or changing the temporal granularity progressively merges nearby documents into clusters and separates them again when more detail is requested.

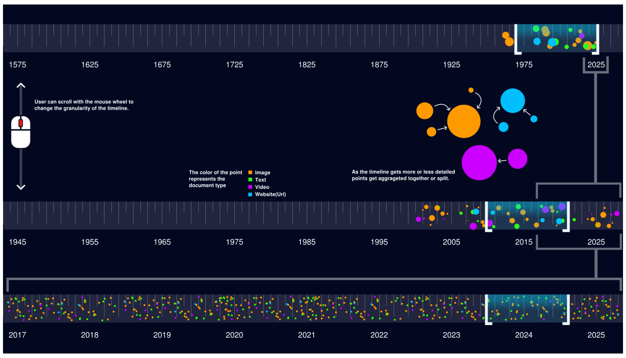


Figure 9. Timeline aggregation and declustering. Point color encodes document type, and changing the temporal granularity merges nearby documents into clusters or splits them back into individual items.

P6 responds to F6 and F8. It supports date display and period filtering, but dense periods still require aggregation to remain readable. The unused vertical axis could encode type or theme, and the selection model could also be extended to compare multiple periods.

5.7 P7: Georeferenced Video Display

Videos extend P5 by turning image placement into a time-varying object. P7 can rely on sampled key frames, trajectories, view directions, and optional point clouds, as illustrated in Figures 10, 11, and 12. It responds to F4, F9, F13, F14, and F15, but dense trajectories and noisy reconstructions can reduce readability.

(a) Frame in 3D (b) Camera and look trajectories (c) Target line

Figure 10. Core spatial cues for georeferenced video.

The target line denotes the segment between the camera position and the current look-at point, making the viewed direction explicit in space. Figure 11 shows how multiple sampled frames can also be displayed together to reveal overlap and movement along the sequence.

5.8 P8: Scroll Viewer

As a baseline for F10, non-georeferencable documents can be shown in a simple scrollable viewer. P8 offers a readable entry point for such material, but it does not structure documents by time, theme, or links, which motivates P9.

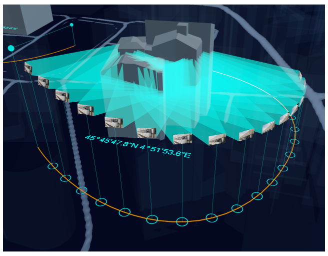
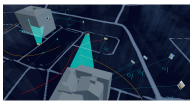
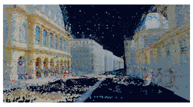


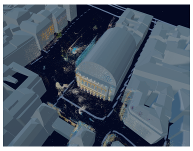
Figure 11. Video-specific interaction through multiple sampled frames.



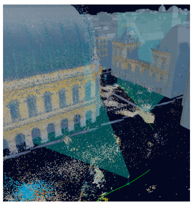
(a) Videos and photos



(b) Immersive point cloud



(c) Zoomed-out point cloud



(d) Frames with point cloud

Figure 12. Extensions of georeferenced video display.

5.9 P9: 3D Non-Georeferencable Document Graph

P9 organizes non-georeferencable documents around the buildings to which they are related. Although these documents cannot be placed at an exact camera position or precise 3D location, they still retain a coarse spatial attachment, typically at the building or component level. Date is mapped to angle around a radial timeline, while the remaining axes can encode theme, type, or relevance (Figure 13).

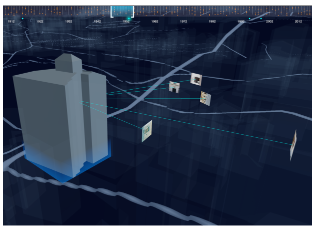


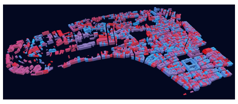
Figure 13. Organization of non-georeferencable documents around related buildings.

Metadata is shown on hover, and links are displayed on demand only for the selected document. P9 responds to F4, F7, F8, F10, F11, and F12; its main limitations are overlapping radial timelines and possible duplication for multi-building documents.

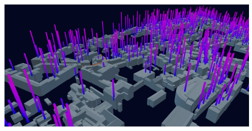
5.10 P10: Document Quantity

P10 is an exploratory view for F16 in which several encodings were tested to show how strongly buildings or components are documented. Figures 14a and 14b summarize two of these encodings. Heatmaps are immediate but weak for close values, while bars make comparison more explicit.

Panels work best in top view, and bars make comparison more explicit; overall, billboard bars offer the best readability, especially with clustering. This quantity view could also be coupled



(a) Heatmap



(b) Bar chart

Figure 14. Quantity encoding alternatives.

with P6 so that bars represent only the counts within the selected period.

6. Discussions and Future Work

This paper contributes a structured design space for exploring multimedia documents in relation to 3D urban geometry rather than a single finalized interface. The propositions suggest that coordinated views are more suitable than a monolithic visualization, especially for photographs and videos, which are both abundant and spatially informative in urban documentation. Table 1 summarizes how the propositions cover the identified functions.

Several limitations nevertheless remain: spatial selection is more developed than multi-document interaction, link display is still partial, temporal comparison is limited to a single interval, and the framework assumes a fixed geometric state of the city. The propositions were discussed with project partners from archives, history, media, HCI, and 3D visualization, but no controlled user study has yet been conducted; future work should therefore address richer multi-document interaction, on-demand links, comparison between periods, time-dependent geometry, stronger coordination between views, and task-based user evaluation.

7. Conclusion

This paper proposed a design space for visualizing multimedia documents in a spatio-temporal-thematic context connected to a 3D city model. By distinguishing culling and display functions, it structured a family of propositions for spatial selection, temporal filtering, georeferenced media placement, organization of non-georeferencable documents, link visualization, and document quantity analysis.

The results show that multimedia exploration in urban space is best supported by several coordinated views rather than by a single representation. The framework is intended as a basis for future systems that combine multimedia corpora and urban digital twins while remaining analytically useful and visually manageable.

Acknowledgements

The authors gratefully acknowledge the French National Research Agency (ANR) and the AGAPE (ANR-24-CE38-7253) project for providing funding for this research. We also thank Vanessa Peña-Araya and Anastasia Bezerianos for their valuable advice. We are grateful to the VCity Project[5] members for their ongoing support, in particular Corentin Gautier for guidance on space-time cube operations, and Jey Puget Gil, Diego Vinasco-Alvarez, and Marwan Ait-Addi for feedback on our propositions. Finally, we acknowledge Wikimedia Commons, from which all images constituting our simulated dataset were sourced from the free-to-use Lyon media collection.

[5] VCity project website: https://projet.liris.cnrs.fr/vcity/

Code and Data Availability

The interactive simulations of the visualization propositions, the prototype patches written in Visual Language within the VVVV IDE, and a sample of the document metadata structure described in Section 5 are openly available in the GitHub repository https://github.com/VCityTeam/UD-ANR-AGAPE-Research (18 June 2026). The repository also documents how CityGML building identifiers are linked to the document metadata, so that the propositions can be reproduced on other CityGML-based 3D city models.

References

Antoun, M., Elhajj, I. H., Sayour, M., Papadopoulos, P. K., Zarpalas, D., Rizk, A., Islam, A. T., Huerta, I., Toledo, L., Arvanitis, G., Moustakas, K., Jevtić, A., Agustí, I., Vona, F., Voigt-Antons, J.-N., Warsinke, M., Bailer, W., Thallinger, G., Howkins, B., Poder, I., Ruess, M., Asmar, D., 2025. Interactive digital twins enabling responsible extended reality applications. *Scientific Reports*, 15(1), 34539. https://doi.org/10.1038/s41598-025-17855-9.

Bach, B., Dragicevic, P., Archambault, D., Hurter, C., Carpendale, S., 2014. A Review of Temporal Data Visualizations Based on Space-Time Cube Operations.

Biljecki, F., Ledoux, H., Stoter, J., 2016. An improved LOD specification for 3D building models. *Computers, Environment and Urban Systems*, 59, 25–37. https://doi.org/10.1016/j.compenvurbsys.2016.04.005.

Blettery, E., Lecat, P., Devaux, A., Gouet-Brunet, V., Saly-Giocanti, F., Brédif, M., Delavoipière, L., Conord, S., Moret, F., 2020. A spatio-temporal web-application for the understanding of the formation of the Parisian metropolis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, VI-4/W1-2020, 45–52. https://doi.org/10.5194/isprs-annals-VI-4-W1-2020-45-2020.

Bruschke, J., Niebling, F., Maiwald, F., Friedrichs, K., Wacker, M., Latoschik, M. E., 2017. Towards browsing repositories of spatially oriented historic photographic images in 3d web environments. *Proceedings of the 22nd International Conference on 3D Web Technology*, Web3D '17, Association for Computing Machinery, New York, NY, USA.

Chagnaud, C., Samuel, J., Servigne, S., Gesquière, G., 2016. Visualization of Documented 3D Cities. *Proceedings*, Proceedings, Liège, Belgium, 87–93.

Chen, C., Han, Y., Galinski, A., Calle, C., Carney, J., Ye, X., van Westen, C., 2025. Integrating Urban Digital Twin with Cloud-Based Geospatial Dashboard for Coastal Resilience Planning: A Case Study in Florida. *Journal of Planning Education and Research*, 45(4), 743-759. https://doi.org/10.1177/0739456X251316185.

Chen, Q., Chen, J., Sun, K., Huang, M., Chen, G., Liu, H., 2024. A Parallel-Optimized Visualization Method for Large-Scale Multiple Video-Augmented Geographic Scenes on Cesium. *ISPRS International Journal of Geo-Information*, 13(12). https://doi.org/10.3390/ijgi13120463.

Gautier, C., Delanoy, J., Gesquiere, G., 2023. Representation of urban geometry evolution through space-time cube. *2023 27th International Conference Information Visualisation (IV)*, 414–419.

Historypin, 2026. A community archive of geolocated historical photographs and stories. https://www.historypin.org/. (28 August 2026).

Hochman, N., Manovich, L., 2013. Zooming into an Instagram City: Reading the local through social media. *First Monday*. https://doi.org/10.5210/fm.v18i7.4711.

Jaillot, V., Rigolle, V., Servigne, S., Samuel, J., Gesquière, G., 2021. Integrating multimedia documents and time-evolving 3D city models for web visualization and navigation. *Transactions in GIS*, 25(3), 1419–1438. https://doi.org/10.1111/tgis.12734. Publisher: Wiley.

Kutzner, T., Chaturvedi, K., Kolbe, T. H., 2022. CityGML 3.0: New Functions Open Up New Applications. *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, 88(1), 43–61.

Lecat, P., Blettery, E., Delavoipiere, L., Saly-Giocanti, F., Conord, S., Gouet-Brunet, V., Devaux, A., Bredif, M., Moret, F., 2021. Restituer les bidonvilles de Nanterre : l'apport d'un outil de visualisation 3D a un projet de sciences sociales. *Humanites numeriques*, 3. https://doi.org/10.4000/revuehn.1946.

Mapillary, 2026. A platform for crowdsourced, georeferenced street-level imagery. https://www.mapillary.com/. (28 August 2026).

Miranda, F., Ortner, T., Moreira, G., Hosseini, M., Vuckovic, M., Biljecki, F., Silva, C. T., Lage, M., Ferreira, N., 2024. The State of the Art in Visual Analytics for 3D Urban Data. *Computer Graphics Forum*, 43(3), e15112. https://doi.org/10.1111/cgf.15112.

Münster, S., Friedrichs, K., Kröber, C., Bruschke, J., Henze, F., Maiwald, F., Niebling, F., 2017. URBAN HISTORY IN 4 DIMENSIONS – SUPPORTING RESEARCH AND EDUCATION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-2/W5, 525–530. https://doi.org/10.5194/isprs-archives-XLII-2-W5-525-2017.

Ozyesil, O., Voroninski, V., Basri, R., Singer, A., 2017. A Survey of Structure from Motion. https://doi.org/10.48550/arXiv.1701.08493. arXiv:1701.08493 [cs].

Rose, G., 2013. Visual Culture, Photography and the Urban: An Interpretive Framework. *Przegląd Kulturoznawczy*, 17(3), 193–204. https://www.ceeol.com/search/article-detail?id=183576. (28 August 2026).

Silva, J. P., Pereira, J. P., Santos, J., 2022. Digital Twin 3D Platform for Collaborative Representation and Editing of IFC Documents Applied to the BIM Methodology. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-4/W1-2022, 411–418. https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-411-2022.

Zimmerman, J., Forlizzi, J., Evenson, S., 2007. Research through design as a method for interaction design research in hci. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, Association for Computing Machinery, New York, NY, USA, 493–502.