STAG: System for ouTdoor Augmented reality using GeoWebXR

Corentin Gautier, Mathieu Brédif

Univ Gustave Eiffel, Géodata Paris, IGN, LASTIG, F-77454 Marne-la-Vallée, France - firstname.lastname@ign.fr

Keywords: Extended Reality, RTK GNSS, WebXR, Geospatial Visualization, Urban Planning

Abstract

Accurate and intuitive visualization of urban development projects is a persistent challenge in spatial planning and public participation. Recent advances in Extended Reality (XR) offer new opportunities to integrate geospatial data directly within the user’s real environment. This paper introduces GeoWebXR, an extension of the WebXR API designed to provide absolute georeferencing of the XR reference space via a standardized geopose. We present an outdoor proof-of-concept implementation that integrates a dual-antenna RTK GNSS receiver mounted on an XR headset. High-precision GNSS measurements are fused with the device’s local pose estimates to compute a consistent and accurate geopose, enabling decimeter-level alignment between virtual and physical environments. Leveraging GeoWebXR, WebGL applications can render georeferenced 3D content in situ through a web browser. We demonstrate this capability using the iTowns geospatial visualization framework to deliver an XR experience for urban planning. The system supports both 1:1-scale in-situ visualization and reduced-scale overview modes, enabling seamless multiscale exploration of planning scenarios. To mitigate cognitive overload in dense urban contexts, we implement several visualization and interaction strategies and conduct a preliminary usability evaluation.

Figure 1. Outdoor XR visualization approach, using Quest 3 headset and GNSS antenna as hardware to geolocalise the XR environment. Photo on the right show the result of urbanisation models integrated in the XR session.

1. Introduction

Urban project visualization represents a central challenge for spatial planning and citizen consultation. While traditional tools, such as maps or digital mock-ups, enable the representation of future developments, they often struggle to offer sufficient immersion for citizens and decision-makers to fully appropriate the space and understand its issues.

The rise of Extended Reality (XR) technologies opens new perspectives for overlaying digital data onto the real environment. These technologies introduce new ways of interacting with one’s physical surroundings, supporting not only spatial understanding but also building maintenance and the appropriation of urban transformations. Achieving an effective overlay of 3D urban data requires on-site integration at a 1:1 scale, along with coherent localization and orientation, to maintain credibility and usefulness for analysis or consultation. However, the relevance of such systems relies on a major technical challenge: the precise and automatic geolocation of 3D models in real space. A major issue also lies in the capacity to effectively integrate and represent complex 3D urban data in XR, without generating cognitive overload for the user. In fact, the density of information and the superposition of multiple virtual objects can quickly hinder scene readability and immersion, thus limiting understanding and spatial appropriation of urban planning projects. It is therefore essential to design visualization and interaction modes that facilitate access to information while preserving the immersive experience.

In this paper, we introduce a new approach for immersive, in situ visualization of 3D urban data and development projects using XR technologies called STAG (System for ouTdoor Augmented reality using GeoWebXR). Our method, describe in figure 1, leverages a dual-antenna RTK GNSS system to automatically and precisely align 3D models with the real world, enabling users to explore urban scenarios at a true 1:1 scale. We fuse global geospatial coordinates with the local XR reference frame through a dedicated software pipeline based on WebXR and iTowns, ensuring robust spatial consistency and seamless deployment on consumer headsets. To address the challenge of cognitive overload in dense urban environments, we design sev-

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-XI-4-2026-213-2026

eral visualization and interaction strategies. We provide users with interactive tools such as contextual metadata panels and scenario selection palettes that facilitate exploration and comparison of different urban planning options. These tools are available at various scales, from miniature palettes attached to the controllers to medium-scale mock-ups placed in the physical space, allowing users to adapt the experience to their needs and preferences.

Our approach is validated through an experimental setup involving real urban data and multiple project scenarios. We assess the usability and spatial appropriation enabled by our system, and discuss how it supports both expert analysis and citizen participation in urban planning processes.

2. Related works

The development of immersive XR systems for urban data visualization faces three main challenges: (1) achieving precise and robust geolocation (position and orientation) for outdoor use, (2) enabling effective and ergonomic visualization and interaction with dense 3D urban data, and (3) ensuring practical deployment on lightweight, user-friendly hardware.

2.1 Geopositioning and orientation in XR

A key technical barrier for outdoor XR is the reliable and automatic alignment of virtual content with the real world. While SLAM (Simultaneous Location And Mapping) algorithms provide accurate local tracking, they lack absolute georeferencing, which is essential for 1:1 overlays in urban environments. Hybrid approaches, such as the one by Ling et al. (2019), combine RTK GNSS and SLAM to anchor local tracking volumes globally, but require bulky external receivers and complex integration.

Recent works have explored multi-antenna GNSS configurations to directly estimate both position and heading. Zhao et al. (2017) and Hu et al. (2021) demonstrate that such setups improve real-time attitude determination, but highlight practical challenges: hardware complexity and sensitivity to signal obstruction. Ismael and Cornil (2025) proposes merging GNSS and inertial data for AR positioning, but their method requires the user to be facing north during initialisation, which limits natural use and flexibility. These studies show that there is still a need for a compact, user-friendly solution that provides both absolute position and orientation without calibration phases. An interesting approach is proposed by Orsholits et al. (2024) that combines GNSS and digital twin for outdoor sound experience. Their compact system composed of a RTK antenna connected to a headset allows for free movement in space and facilitates immersive exploration. But they didn’t propose an XR experience, an interesting approach should be visualize the digital twins they used in a AR environment.

2.2 Visualization, interaction, and cognitive load

Beyond geolocation, the effective visualization of urban data in XR requires careful management of information density and user interaction. Spur et al. (2020) and Rua Mota et al. (2022) explore multi-layered and spatio-temporal VR visualizations, revealing trade-offs between precision and interpretability. These approaches are interesting to explore complex datasets, but they do not address the specific challenges of in-situ 1:1 scale visualization in outdoor environments. Rzeszewski and Orylski (2021) and Höftberger et al. (2023) emphasize the importance of intuitive interfaces and participatory methods to maximize user engagement and understanding.

Several domain-specific solutions illustrate the strengths and limitations of current approaches. Marker-based AR for urban projects Boos et al. (2023) is effective for local engagement but is limited by its reliance on physical markers, which hinders scalability and flexibility. Temporal lenses for heritage visualization Fanini et al. (2021) enrich narrative potential but do not address robust outdoor georeferencing. Early work on annotated overlays and guided tours Reitmayr and Schmalstieg (2004) remains relevant for structured visits, but must be adapted for free, in-situ exploration at 1:1 scale. Seo and Yoo (2020) provides technical insights into 3D data integration and annotation pipelines in XR, which inspire our own approach to data fusion and interaction.

A recurring issue is cognitive overload due to the density of 3D objects. Dynamic transparency and occlusion management techniques, as discussed by Elmqvist et al. (2007) and Alfakhori et al. (2023), are essential to maintain scene readability and depth perception, especially in dense urban contexts. The work of Lee et al. (2023) defines several patterns for situated visualization in augmented reality. In our case, we use the ”ghost” pattern to highlight real buildings with geospatial data and provide calibration feedback to the user. For scenario visualization, the ”proxy” pattern is the most used, as it enables comparison between an in-situ object and an alternative version of it at a reduced scale, making detailed inspection easier.

2.3 Deployment platforms and geospatial integration

The emergence of WebXR, combined with geospatial libraries such as iTowns Picavet et al. (2016); Martí-Testón et al. (2023), offers a promising context for deploying XR experiences on consumer headsets without native installation. This approach facilitates the rendering of complex urban scenes and the integration of standard geospatial formats (e.g., CityGMLGröger and Plümer (2012), 3DTilesMao et al. (2020)). However, fusing global GNSS coordinates with the local XR frame remains a technical challenge, requiring robust and lightweight solutions for coordinate transformation and real-time updates. Our approach aligns with previous work, notably Ling et al. (2019), aiming at a compact and automatic integration of a dual-antenna RTK system capable of simultaneously providing position and geographic orientation. This eliminates the complexity linked to SLAM tracking for initial orientation and enables geospatial 3D overlay in XR, facilitating immersive exploration and comparison of urban development projects at 1:1 scale, in an environment accessible via WebXR. This hardware and software innovation addresses the specific needs of participatory urban planning. Höftberger et al. (2023) proposes participatory methodologies involving workshops and collaborative tools, emphasizing the necessity to combine technologies and social methods.

3. Proposed method

Our methodological approach builds on the necessity of precise georeferencing and reliable overlay of 3D urban models in an outdoor XR environment. We first present the overall GeoWebXR API that decouples the geolocation problem from XR visualization concerns, then we introduce a straightforward approach for the geolocation estimation and, finally, we detail the visualization techniques and interactions developed for the

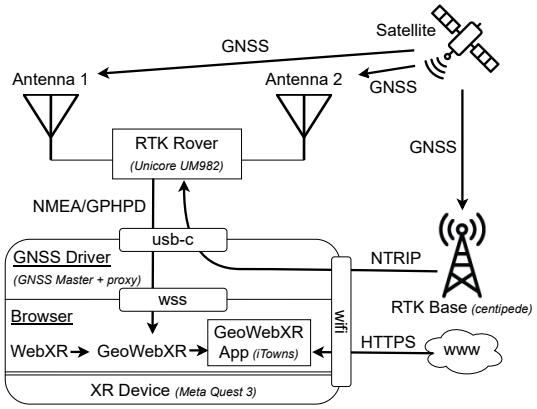

Figure 2. Overall schematic of the proposed setup illustrating the hardware and software pipeline for geolocation and geospatial XR visualization.

proposed use case. The diagram shown in Figure 2 summarizes the entire setup, from capturing position and orientation data to immersive and interactive visualization.

The entire approach has been designed to be reproducible, relying on open hardware and the development of our solution as open source. The hardware components used are detailed in the following section and consist of parts that are easily accessible to the community. The source code for our method is available on a public GitHub repository, and technical documentation has been written to facilitate its adoption and reuse.

3.1 GeoWebXR API

To enable 1:1 in-situ XR visualization, the absolute position, scale, and orientation of the webXR local frame must be estimated and taken into account. This enables the correct visualization of georeferenced digital content within a WebXR application. The WebXR API provides headset pose in a local frame but is unaware of its rigid transform to any physical frame such as ECEF or a local tangent plane. Worse, this local frame may drift overtime (e.g. when using the ’unbounded’ reference space). Recently, the Open Geospatial Consortium (OGC) has standardized this concept of geographically-anchored poses as a geopose, which encodes the pose of a real or digital object relative to a geographical frame of reference Smyth (2022). Here, we consider only the Chain standardization target that defines the local frame of a geopose by a Local Tangent Plane frame with an East North Up orientation (LTP-ENU) and a chain of rigid transforms (`Translate-Rotate`).

We propose GeoWebXR, a new WebXR extension that defines a new reference space `"geo"` which extends the reference space `"unbounded"` by providing its geopose in the `XRSpace` object as `referenceSpace.geopose`. This architecture clearly decouples geolocation from geovisualization, such that a GeoWebXR application may be executed in the browser without any concern about the underlying hardware and software geolocation stack. The proposed GeoWebXR does not rely on any specific pose estimation approach, so that future implementation may leverage any localisation approach, such as vision, marker-based or lidar.

3.2 Geolocation

The geolocation step of our method aims to establish a robust correspondence between global geospatial positioning and the local spatial reference of the XR environment: the geopose of the GeoWebXR reference space. This geopose ensures that the virtual content precisely matches the real environment in translation and orientation, which is essential for achieving coherent geospatial augmented reality visualization.

We assume that the XR device already provides reliable estimates of the device position relative to the floating frame of the WebXR reference space and that the vertical direction of this local frame is correct as it is now commonly well estimated by processing IMU (inertial measurement unit) measurements. We further assume that the local frame is metric, so that scaling corrections are not necessary. Thus, the absolute positioning of the local frame relative to the physical world requires the estimation of four missing degrees of freedom : the absolute 3D position of the local frame origin and the unknown azimuth angle parametrizing the rotation of the local frame along the vertical direction of the LTP at its origin.

Basic GNSS sensors and magnetometers would be able to measure these missing parameters. However, the direct GNSS positioning and the estimated azimuth of the magnetic north are not accurate enough for our use case. Using preliminary georeferenced high accuracy acquisitions, techniques based on vision and/or lidar, may achieve the required accuracy. These more advanced geolocation techniques are however, out of the scope of this paper. We propose here a simple setup, based on direct georeferencing only.

Figure 3. Prototype: 2 antennas connected to a Unicore UM932, with a 30cm baseline, and attached to a Quest 3 XR headset.

3.2.1 Hardware: Our approach builds upon a dual-antenna RTK GNSS configuration that provides both the position in the WGS84 coordinate system and the vector between the two antennas (see Figure 3). This vector measures two of the three angles of the device rotation relative to the LTP frame located at the first antenna position, as it does not measure the rotation of the device along the line formed by the two antenna centers. Arranging the antennas vertically on the device would measure the vertical direction in the standard user position (standing position with a horizontal user gaze). This would be redundant with the IMU-based estimate of the vertical direction. We propose to place the antennas sideways, as users are more likely to look up or down (pitch) than to tilt their head sideways (roll). Using such a transversal baseline, the heading and roll angles are estimated but not the pitch (which would correspond to looking up or down).

The lever-arm offset and boresight angles are estimated using the CAD model of the system. More precise estimates of the rigid transforms between the frame attached to the GNSS antennas and each camera frame is out of the scope of this paper.

3.2.2 Geopose estimation: This fusion stage estimates the geopose of the WebXR reference space. Five coordinate systems come into play, considering a single camera. Figure 4 illustrates these frames and introduces the rotations R_z, R_x, R_c and R_g and the translations T_x and T_c ($T_z = T_g = 0$):

- LTP_{GNSS}: the local tangent plane frame centered at the master GNSS antenna.
- LTP_{XR}: the local tangent plane frame centered at the XR reference space origin.
- $GNSS$: the moving frame attached to the GNSS sensors (the baseline direction vector is $[1,0,0]^T$.
- XR: the frame of the XR reference space.
- $Camera$: the moving frame attached to the XR Camera.

The rotation $[0, R_z]$ has a vertical axis and accounts for the unknown azimuth of the XR reference space. $[T_x, R_x]$ is provided by the XR device through the WebXR API. $[T_c, R_c]$ is the calibrated rigid transform between the camera and GNSS frames. Finally, $[0, R_g]$ encodes the rotation of the GNSS frames relative to its LTP frame.

The dual antenna GNSS system estimates the baseline vector b in the LTP_{GNSS} frame, so we can write $R_g b = [1,0,0]^T$. Moreover, we assume that we can neglect the rotation between the LTP frames, such that we can write that :

$$R_g = R_c R_x R_z \quad (1)$$

$$R_z b = b' \quad \text{with} \quad b' = R_x^{-1} R_c^{-1} [1,0,0]^T \quad (2)$$

In this equation, a single parameter is unknown: the angle θ of the vertical rotation R_z. This yields a validity check that $b_z \simeq b'_z$ and that θ may be estimated from the 2D rotation from $\frac{(b_x, b_y)}{\|b_x, b_y\|}$ to $\frac{(b'_x, b'_y)}{\|b'_x, b'_y\|}$. Now that R_z is estimated, we can now compute R_g using Equation 1 and define the geopose of the XR reference space relative to the LTP_{GNSS} frame with the chain of transforms $([0, R_g], [T_c, R_c]^{-1}, [T_x, R_x]^{-1})$.

In our experience, the pose estimation of the XR system is very reliable and the drift of the XR reference space is minimal. Since the geopose of the XR reference space is relatively stable, in this first prototype, we use a simple approach. It is estimated once at start-up after the GNSS system warm-up and GNSS fix. Then the user may enable new estimates on demand by pressing a controller button. A limitation of this approach is that the dynamics of the system are not taken into account. This means that quantities provided by the GNSS system and the WebXR API may correspond to different timestamps, which is problematic if the user is moving. Thus, we ask the user to remain static at (re)initialization time. This issue could be addressed by keeping track of the measurement timestamps during the fusion. This time-aware fusion could be implemented either within XR host. Otherwise, the time alignment could be estimated by aligning redundant measurements such as the GNSS and XR trajectory, or the b_z and b'_z values, which are respectively measured by the GNSS and XR systems.

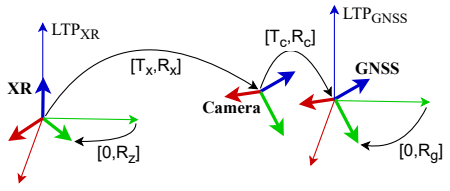

Figure 4. Frames and transformations involved in the estimation of the geopose of the WebXR reference space by fusing WebXR local estimates and GNSS global estimates.

To go further, the fusion of the GNSS and XR measurements should be performed through continuous updates (e.g. using Kalman filtersWelch et al. (1995)), to provide a better and more robust estimate of the slowly evolving geopose of the XR reference space in the presence of possibly noisy or erroneous measurements. This framework provides a consistent transformation system for global–local fusion, ensuring stable and metrically accurate augmented visualization of 3D urban scenes within XR environment.

3.3 Visualization

The geolocation component described in Section 3.2 provides a necessary foundation for immersive XR visualization of 3D urban data. Integrating 3D urban data into XR scene requires a transformation method to keep virtual objects aligned with the physical environment as the user moves. The visualization workflow integrates georeferenced 3D urban data into the XR scene through the GeoWebXR API described in Section 3.1. This requires addressing challenges related to cognitive overload, spatial comprehension, and user interaction with complex urban models (see Figure 5).

3.3.1 Georeferenced 3D data using GeoWebXR: Through the GeoWebXR API, at render time, the geopose of the reference space may be accessed. Knowing the geopose of the XR reference space and the geoposes of each georeferenced WebGL object, the transform between the WebXR's reference space and the local space of each georeferenced WebGL drawable object may be updated. If the WebGL application uses a scene graph, entire subtrees of drawable objects may be georeferenced by defining the geopose of their root node. Then, the update only needs to recompute the transform between the local spaces defined by the GeoWebXR geopose and by the geopose of each of the subtree roots, instead of having to recompute separately the transforms of every drawable of these subtrees. There is a rigid transform between the ECEF frame and the LTP frame centered at the geodetic coordinates (λ, φ, h). The semi-major axis a and the eccentricity e, derived from the WGS84 parameters, may be used to express the origin $P(\lambda, \varphi, h)$ of the LTP frame in ECEF coordinates. The rotation may likewise be expressed as $R(\lambda, \varphi)$. For illustration, $R_{ENU}(\lambda, \varphi)$ is given here for the East-North-Up (ENU) convention.

$$P(\lambda,\varphi,h) = \begin{bmatrix} (\frac{a}{\sqrt{1-e^2\sin^2\varphi}} + h)\cos\varphi\,\cos\lambda \\ (\frac{a}{\sqrt{1-e^2\sin^2\varphi}} + h)\cos\varphi\,\sin\lambda \\ \left(\frac{a(1-e^2)}{\sqrt{1-e^2\sin^2\varphi}} + h\right)\sin\varphi \end{bmatrix} \quad (3)$$

$$R_{ENU}(\lambda,\varphi) = \begin{bmatrix} -\sin\lambda & \cos\lambda & 0 \\ -\cos\lambda\sin\varphi & -\sin\lambda\sin\varphi & \cos\varphi \\ \cos\lambda\cos\varphi & \sin\lambda\cos\varphi & \sin\varphi \end{bmatrix} \quad (4)$$

Denoting X_i and X_{ECEF} the coordinates of a point in the LTP frames of two geoposes ($i = 1$ or 2), one can write, using the ECEF to LTP rotations R_i and the LTP origins in ECEF coordinates P_i, we have $X_i = R_i(X_{ECEF} - P_i)$ and thus:

$$X_2 = R_2\left(R_1^T X_1 + (P_1 - P_2)\right) \quad (5)$$

Moreover, the chain of transforms that define the local frame of a geopose relative to its base frame also results in a rigid transform, denoted H_i. Using homogeneous 4x4 matrices, the resulting overall transform $H_{1\to 2}$ between the local frames of 2

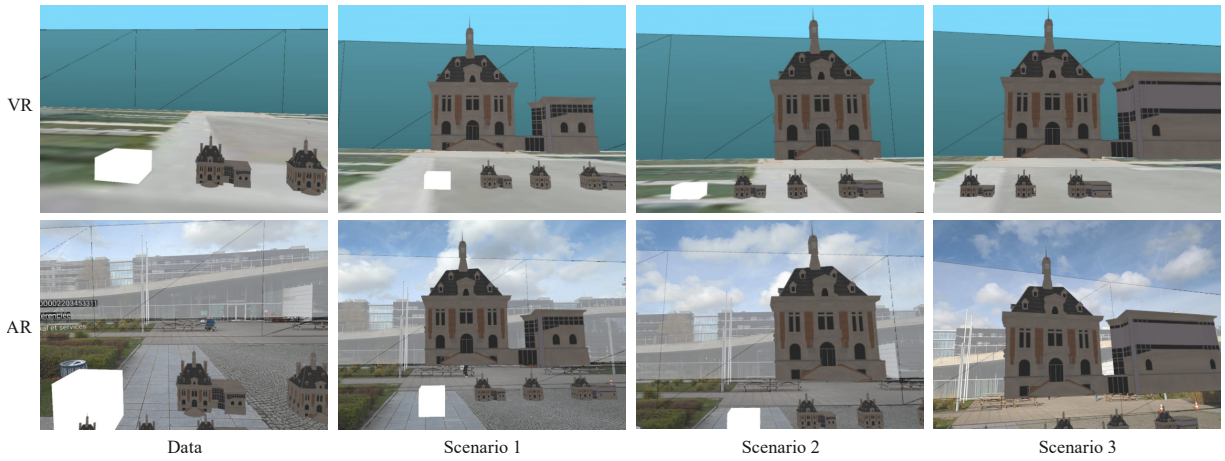

Figure 5. Visualization modes comparison: (top row) VR interior mode showing existing building data and three renovation scenarios; (bottom row) same scenarios registered at 1:1 scale in outdoor AR, precisely aligned with real environment using GeoWebXR.

geoposes is then finally:

$$H_{1\to 2} = H_2 \begin{bmatrix} R_2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & P_1 - P_2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R_1^T & 0 \\ 0 & 1 \end{bmatrix} H_1^{-1} \quad (6)$$

Please note that, for numerical stability, the translations between the LTP origins P_i and the ECEF origin have been coalesced into a single translation. As the resulting transformation $H_{1\to 2}$ between two geopose frames is a rigid transform (hence a homography), it may be used seamlessly in a computer graphics pipeline such as WebGL to express the transform between the root node of the XR reference space and each node featuring a geopose that georeferences its local space.

3.3.2 Digital twin data context: To establish a clear spatial reference for users, we incorporate the urban digital twin Ferré-Bigorra et al. (2022) based on open-source 3D geospatial data representing existing buildings and urban features at the project site. The digital twin data is setup using the iTowns library, which efficiently handles large-scale 3D geospatial datasets and supports various standard formats such as CityGML and 3DTiles. The GeoWebXR API initializes the georeferenced scene ensuring that these datasets are accurately aligned with the real-world environment. These datasets are initially rendered with full opacity to provide immediate visual feedback, immersing users in a virtual replica of their surroundings. To maintain depth perception and allow simultaneous visualization of physical and digital content, we apply a dynamic transparency mode called ”ghost” pattern Lee et al. (2023) to the 3D building models that the user can activate with the controller. To further enhance building envelope visibility, we add a lightweight wireframe overlay around the transparent structures. This wireframe delineates building geometry more clearly, improving perceptual accuracy and depth discrimination.

The transparent mode can be enabled or disabled by the user using the XR controller, allowing them to switch between a fully immersive digital twin view and a blended augmented reality experience (Figure 5). The user can also interact with the 3D geospatial data to access contextual information about specific buildings or urban features. When a user points and selects a building through the XR controllers, a descriptive panel displays its characteristics, such as construction date, building type, and height. This contextual interaction strengthens spatial understanding and supports analysis or decision-making during public consultation.

3.3.3 In-situ 3D model: Initializing the XR scene with existing data establishes a spatial reference for overlaying 3D models of planned urban developments. These models, often created from architectural software, include detailed geometries and realistic textures to communicate design intentions. Their integration requires precise georeferencing to align perfectly with the real-world environment. In the case where the 3D models are not georeferenced, we perform a manual alignment using control points shared between the terrain and the model to ensure proper positioning. The model is integrated at a 1:1 scale and georeferenced with GeoWebXR API and overlaid onto the transparent 3D data, allowing users to visualize proposed changes directly within their physical context (Figure 5). Architectural or urban competitions often involve several design scenarios representing different spatial configurations or budget constraints. Each scenario appears as an independent 3D model that can be explored individually to facilitate comparison. Figure 8 shows an example where a building extension is varied to simulate three development strategies. Interesting approaches to highlight differences between scenarios and the digital twin could be to set color themes depending on the evolution (demolition, new construction, modification). It would facilitate the perception of changes between scenarios exploration conditions. To further enhance scenario comparison, we recommend assigning color themes based on change type (e.g., red for demolition, green for new construction, yellow for modification). Such visual encoding would strengthen the perception of spatial differences between scenarios and improve decision-making during exploration.

The 1:1 scale integration allows users to walk around the proposed development, observing it from multiple viewpoints and distances, as if physically present at the project site. However, to visualize different forms of urbanization users need to choose the one to display in the real world. We proposed two interactive modes to facilitate scenario exploration and selection.

3.3.4 Medium-Scale Mock-up Mode: We used the design pattern ”proxy” described by Lee et al. (2023) to letting the user handle hug object such a building. This first exploration mode involves medium-scale mock-ups positioned in the nearby physical space. The different models scenario are georeferenced thanks to the GeoWebXR API, they are oriented depending on the users needs. They are placed on pedestrals around the user, allowing side-by-side comparison before selecting one for full-scale visualization (see Figure 6). The scale of these models

is reduced to fit within the user's immediate reach, typically around 1:20 to 1:50 scale depending on the initial model scale. This setup enables users to examine design details and spatial relationships in a more manageable format. Users can walk around these mock-ups, gaining a comprehensive understanding of each scenario's spatial configuration and design intent. We also provide a default mode where only the existing conditions are displayed using a neutral ”white cube” model to temporarily mask future development models. A scenario is selected when the user points at a medium-scale model and presses the selection button on the XR controller. Once a scenario is selected, the chosen model appears at its precise geospatial location in 1:1 scale. The user can then walk around it freely within the XR space, observing the proposed intervention from multiple viewpoints and distances, as if physically present at the project site.

3.3.5 Miniature Palette Mode: The second exploration mode utilizes a miniature palette of design options attached to the XR controllers, allowing users to quickly switch between different scenarios. The scale of these miniatures is around 1:100 to 1:200 (according to the size of the model), making them easy to handle and not overload the user (see Figure 7). In this mode, users can move around their environment while having access to all the scenarios. As the medium-scale mock-up mode, a neutral ”white cube” model is also available to temporarily mask future development models, displaying only existing conditions. When a user selects a scenario from the miniature palette, the corresponding full-scale model is activated onsite, replacing any previously displayed model. This approach streamlines the exploration process, making it more efficient and flexible.

4. Experimentation

This section presents the experimental setup used to evaluate our semi-automatic geolocation and XR visualization method, and the two complementary scenario exploration modes. We first describe the technical implementation and the in situ model used for testing, then detail the expert user study conducted with ten researchers.

4.1 Technical Implementation

The whole design is described in the Figure 2. The hardware assembly mounts a Unicore UM982 GNSS receiver with two EM-901A antennas on a Meta Quest 3 headset. We have built 2 prototypes with different baselines (20cm and 30cm) to explore the trade-off between system compactness and orientation accuracy. The mechanical mount is composed of a rigid and lightweight aluminium profile attached to the headset using a 3D printed clip.

Our geolocation pipeline leverages the GNSS Master Android application to retrieve NTRIP (Network and Transport of RTCM via Internet Protocol) messages from a nearby RTK base of the centipede[1] network and transfer them to the GNSS receiver in RTK Rover mode. The UM982 chip performs the RTK correction and estimates the position and baseline direction, which it back to GNSS Master as GPHPD sentences(NMEA standard[2]). This middleware, in turn forwards these GPHPD sentences over TCP/IP. We implemented a TCP/IP to Websocket Secure (WSS) proxy that streams these messages so that they may be accessed from the WebXR application. In the future, a single middleware application could be implemented that would combine the functionalities of the GNSS Master application and the WSS proxy. The GNSS receiver is connected to the XR headset using USB-C and the XR headset connects to the RTK base and to the GeoWebXR-enabled website using wifi to any wifi hotspot (e.g. fixed infrastructure or wifi-sharing smartphone). The XR application runs within Meta Quest 3's browser using the WebXR API[3]. The fusion of the WebXR local positioning and the GNSS global positioning is performed in javascript. We access the GPHPD sentences via WSS to decode the absolute position of the GNSS frame and its baseline direction b. At the following WebXR animation frame, the cached GNSS position and baseline direction are used to estimate the geopose of the WebXR reference space. If the browser implemented the proposed GeoWebXR API, the geopose of the XR reference space would already be available in the `XRSpace` object.

The AR visualization has been implemented using the iTowns Picavet et al. (2016) library for efficient management and rendering of geospatial data directly on-device, which is based on the THREE.js library. Now that the geopose of the XR reference space is available, the application may update the transforms from the world space (the XR reference space) to the `THREE.Groups` which are roots of sub-trees of georeferenced assets. This technical implementation serves as the basis for our outdoor XR visualization. It is a suitable approach that meets our need for a compact and easy-to-use system.

4.2 Use-case: city hall modification

Our experimentation is based on a 3D digital model of a city hall (see Figure 8), inspired of a city in northern France. We chose to use this building as our first experiment because of

[1] `https://docs.centipede.fr/`
[2] `https://www.nmea.org`
[3] `https://immersiveweb.dev`

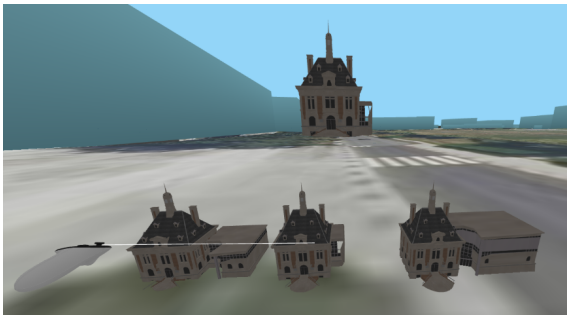

Figure 6. Medium scale selection mode. Scenario 2 is selected by the user and displayed at the in-situ localisation.

Figure 7. Miniature palette mode. Scenarios are displayed along the XR controller and can be selected.

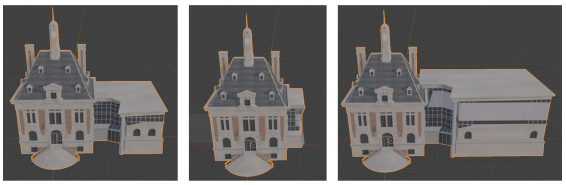

Figure 8. From left to right, design scenarios 1, 2 and 3.

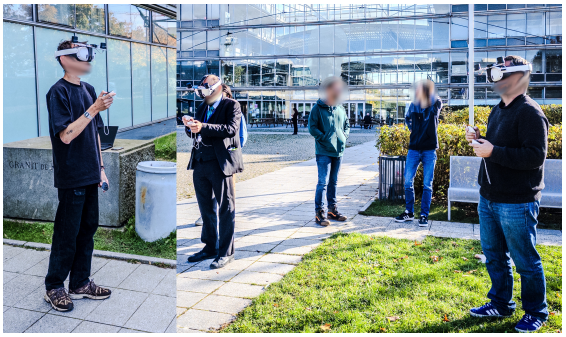

Figure 9. Experimentation session with the group of researchers

its simplicity in terms of spatial interpretation and appropriation. We can also geolocate it fairly easily wherever we want thanks to its size. We conducted an expert evaluation with 10 researchers who were comfortable working with 3D scenes but not necessarily with XR environments (see Figure 9). Qualitative data were collected through interviews conducted with the researchers during and after the experiment, following a 10 minutes introductory presentation. Each session lasted approximately 10 minutes and took place next to the laboratory, in an open area where participants could move freely. The in-situ model was geolocated in front of this open space and positioned on a real building. Different tasks were assigned to participants to evaluate the usability of each interaction method and the overall fluidity of the system. The first task consisted of selecting a scenario through the palette mode and observing the resulting in-situ model display. The second task followed the same procedure, but using the medium-scale selection method instead. Finally, participants were asked to move freely through the available space in order to observe the in-situ model from multiple viewpoints and assess its integration into the environment.

4.2.1 Model used and geospatial data: This model was developed by a professional graphic designer with precise metric scaling to match real-world dimensions, enabling accurate representation within the geospatial reference system. We overlay this architectural model onto open-source 3D building data from the French national mapping agency[4], consisting of cadastral footprint extrusions accessible via a WFS (Web Feature Service). This data provides a comprehensive urban context covering the experimental site, facilitating meaningful immersive visualization.

Three alternative scenarios simulate different budget constraints on the city hall renovation project: the base design, a reduced budget version with simplified geometry, and an expanded, more ambitious scheme. The three versions of the project have their geometry expressed in the same LTP-ENU frame and may thus be georeferenced and visualized in-situ in XR to enable straightforward spatial comparison.

[4] `https://www.ign.fr`

4.2.2 Miniature Palette Mode: In this mode, the three scenario models along with a neutral white cube are positioned horizontally slightly above the XR controller grip. For the demonstration, we add a textual label that indicates the scenario number below the miniature models, guiding the user and reinforcing selection clarity. Feedback collected from our researcher participants highlighted that this interaction was appreciated for its intuitiveness and ease of use. Users quickly understood how to select a scenario to display the corresponding full-scale model (1:1). We observed that this mechanism also facilitated smooth transitions when switching between scenarios. However, some users suggested that aligning the miniature models horizontally in line with the user’s arm extension could improve pointing precision and comfort. This adjustment might further enhance interaction fluidity by making it more ergonomic to direct the controller towards the models. Overall, the miniature palette mode offers a compact, always-accessible selection interface that effectively supports scenario exploration within the immersive environment.

4.2.3 Medium-Scale Mock-up Mode: The medium-scale models are positioned horizontally, side by side, between the user’s initial location and the full-scale model (1:1), facing the user to provide a clear, comprehensive overview of all scenarios. During user sessions with researchers, this mode encouraged closer exploration and free movement around the models, allowing observation from multiple angles. Participants appreciated this larger representation compared to the miniature palette, which facilitated the inspection of finer details, from a bird-eye view on these reduced-size models. However, this mode required more frequent transitions between interacting with mock-ups and the full-scale scene. Some participants initially had difficulty locating the medium-scale models due to their placement and orientation, suggesting that clearer cues or alignment could improve discoverability.

4.2.4 In-situ full-scale visualization: Once a scenario was selected using either of the two modes, the corresponding full-scale model was displayed replacing an actual building in the real world. The city-hall model was accurately overlaid onto the existing building footprint, allowing users to visualize the proposed renovation directly within its physical context. After the selection, some users took time to observe the full-scale model from different angles and distances. The in-situ model was geolocalised around 10 meters from the user initial position, which allowed them to walk around it and explore the model in detail. Participants appreciated the ability to see how the proposed design integrated with the surrounding environment, providing a tangible sense of scale and spatial relationships. We observe that users walk in front of the model to see the facade details, then move to the sides to understand how it fits in the world. Some users suggested that adding informational overlays, like which part of the real building would be modified or destructed to create the proposal model, could enhance understanding of the project’s impact.

5. Conclusion

This paper presents our proposed method STAG a System for ouTdoor Augmented reality using GeoWebXR. The GeoWebXR API decouples geolocation from visualization, fusing global GNSS coordinates with the local XR reference frame to ensure precise 1:1 scale alignment. However, in our current implementation, this fusion is performed only at startup; continuous updates would improve robustness to GNSS noise and real-time

device movement. Our approach is fully open source and relies on commercially accessible hardware components, lowering barriers to adoption and enabling reproducible research. To mitigate cognitive overload in dense urban environments, we implemented dynamic transparency, contextual metadata panels, and dual interaction modes for scenario comparison. Preliminary evaluation with researchers demonstrated the system's usability interaction design, with users effectively exploring and comparing urban design alternatives in situ. It should be noted that our hardware provides heading and roll estimation via dual-antenna baseline, while pitch estimation remains IMU-dependent, and GNSS accuracy may degrade significantly in dense urban canyons. Additionally, our user study was limited to 10 researchers; formal evaluation with diverse user groups (citizens, urban planners, architects) is essential to validate effectiveness.

Our method could be enhanced with the implementation of Kalman-filter-based global-local fusion to handle real-time GNSS updates, latency, and noise. Integrate vision-based localization as a fallback in GNSS-denied environments, improving system resilience in dense urban settings. And finally, to promote interoperability, we could formalize the GeoWebXR API proposal within the OGC.

6. Acknowledgments

This work was supported by government funding administered by the French National Research Agency under the France 2030 Plan, reference ANR-21 EXES 007. Special thanks go to the members of the LASTIG lab for their participation and for the fruitful discussions during the user study.

References

Alfakhori, M., Sardi Barzallo, J. S., Coors, V., 2023. Occlusion handling for mobile AR applications in indoor and outdoor scenarios. *Sensors*, 23(9), 4245. https://www.mdpi.com/1424-8220/23/9/4245.

Boos, U. C., Reichenbacher, T., Kiefer, P., Sailer, C., 2023. An augmented reality study for public participation in urban planning. *Journal of Location Based Services*, 17(1), 48–77.

Elmqvist, N., Assarsson, U., Tsigas, P., 2007. Employing dynamic transparency for 3d occlusion management: Design issues and evaluation. *IFIP Conference on Human-Computer Interaction*, Springer, 532–545.

Fanini, B., Ferdani, D., Demetrescu, E., 2021. Temporal lensing: an interactive and scalable technique for Web3D/WebXR applications in cultural heritage. *Heritage*, 4(2), 710–724.

Ferré-Bigorra, J., Casals, M., Gangolells, M., 2022. The adoption of urban digital twins. *Cities*, 131, 103905.

Gröger, G., Plümer, L., 2012. CityGML–Interoperable semantic 3D city models. *ISPRS Journal of Photogrammetry and Remote Sensing*, 71, 12–33.

Höftberger, K., Konrath, A., Berger, A., Allerstorfer, D., Krebs, R., 2023. Xr-supported communication in green urban projects. participating in urban change through virtual and augmented reality. CORP–Competence Center of Urban and Regional Planning.

Hu, X., Thevenon, P., Macabiau, C., 2021. Attitude determination and rtk performances amelioration using multiple low-cost receivers with known geometry. *Proceedings of the 2021 International Technical Meeting of The Institute of Navigation*, 439–453.

Ismael, M., Cornil, M., 2025. Real-Time Kinematic Positioning and Optical See-Through Head-Mounted Display for Outdoor Tracking: Hybrid System and Preliminary Assessment. *arXiv preprint arXiv:2509.09412*.

Lee, B., Sedlmair, M., Schmalstieg, D., 2023. Design patterns for situated visualization in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 1324–1335.

Ling, S., Elvezio, L., Pradhan, P., Geiger, A., 2019. A hybrid rtk gnss and slam outdoor augmented reality system. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*.

Mao, B., Ban, Y., Laumert, B., 2020. Dynamic online 3D visualization framework for real-time energy simulation based on 3D tiles. *ISPRS International Journal of Geo-Information*, 9(3), 166.

Martí-Testón, A., Muñoz, A., Gracia, L., Solanes, J. E., 2023. Using WebXR metaverse platforms to create touristic services and cultural promotion. *Applied Sciences*, 13(14), 8544.

Orsholits, A., Qian, Y., Nardini, E., Obuchi, Y., Tsukada, M., 2024. Platone: An immersive geospatial audio spatialization platform. *2024 IEEE International Conference on Metaverse Computing, Networking, and Applications (MetaCom)*, IEEE.

Picavet, V., Brédif, M., Konini, M., Devaux, A., 2016. iTowns, framework web pour la donnée géographique 3D. *Revue XYZ*, 147, 49–52.

Reitmayr, G., Schmalstieg, D., 2004. *Collaborative augmented reality for outdoor navigation and information browsing*.

Rua Mota, R. et al., 2022. A comparison of spatiotemporal visualizations for 3D urban analytics. *IEEE Transactions on Visualization and Computer Graphics*, 28(1), 233–243. https://ieeexplore.ieee.org/document/9579953.

Rzeszewski, M., Orylski, M., 2021. Usability of WebXR visualizations in urban planning. *ISPRS International Journal of Geo-Information*, 10(11), 721.

Seo, D., Yoo, B., 2020. Interoperable information model for geovisualization and interaction in XR environments. *International Journal of Geographical Information Science*, 34(7), 1323–1352.

Smyth, C. S., 2022. OGC GeoPose 1.0 Data Exchange Standard. `http://www.opengis.net/doc/IS/geopose/1.0`, Open Geospatial Consortium (OGC).

Spur, M., Tourre, V., David, E., Moreau, G., Le Callet, P., 2020. Mapstack: Exploring multilayered geospatial data in virtual reality. *11th international conference on information visualization theory and applications*.

Welch, G., Bishop, G. et al., 1995. An introduction to the Kalman filter.

Zhao, L., Li, N., Li, L., Zhang, Y., Cheng, C., 2017. Real-time GNSS-based attitude determination in the measurement domain. *Sensors*, 17(2), 296. https://www.mdpi.com/1424-8220/17/2/296.