POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

Precision Outdoor XR Framework: Multi-Device GPS-Vision Localization with Persistent Beacon-Based Augmented Reality

Laurea Magistrale in Computer Science and Engineering - Ingegneria Informatica

Author: Mattia Zambetti

Advisor: Prof. Marco Gribaudo

Co-advisor: Marco Domenico Buttiglione

Academic year: 2024-2025

1. Introduction

Augmented Reality (AR) technologies have rapidly evolved over the past decade, yet large-scale, persistent, and globally accessible outdoor AR infrastructures remain largely unrealized. Current solutions are typically device-centric, platform-specific, and limited in spatial persistence or interoperability. This thesis investigates whether a scalable and globally distributed outdoor augmented reality ecosystem for immersive devices (e.g. headset or AR glasses) can be designed by extending established web-based architectural principles into a spatial computing paradigm.

The work proposes and validates a conceptual and technical framework referred to as World Wide Augmented Reality (WWAR). Rather than focusing on a single algorithm or hardware solution, the thesis adopts a system-engineering perspective, structuring the problem across communication layers, software architecture, spatial localization mechanisms, content provisioning, and governance models.

2. Problem Definition

The primary objective of this research is to determine the architectural and technological requirements necessary to enable a globally scalable outdoor AR infrastructure capable of supporting:

- Sub-meter spatial anchoring accuracy,
- Persistent georeferenced 3D content,
- Multi-device cooperative localization,
- Secure and distributed content provisioning.

A key conceptual contribution of the work is the formal analogy between the traditional World Wide Web and the proposed WWAR paradigm. While the Web organizes and indexes information through URLs and renders content independently of geographic context, WWAR introduces spatial anchoring as a primary dimension of resource identification. In this model, digital assets are indexed not only by network identifiers but also by geographic coordinates and contextual visibility constraints.

The research questions therefore address the feasibility of such a transition, examining communication protocols, client architectures, data formats, localization strategies, content indexing models, adoption barriers, and governance mechanisms.

3. Research Questions

The thesis is structured around eight research questions addressing the feasibility of a globally distributed outdoor augmented reality infrastruc-

ture.
The first research question investigates the architectural and technological components required to design a scalable World Wide Augmented Reality (WWAR) framework capable of achieving sub-meter spatial anchoring through multi-device sensor fusion and persistent spatial indexing.
The second research question focuses on communication architecture, examining which transport and application-layer features are necessary to support continuous, low-latency transmission of spatially dynamic 3D content in mobile outdoor environments.
The third and fourth research questions analyze the software layer and 3D content representation, investigating how a hybrid client architecture and standardized 3D formats can ensure interoperability, real-time rendering performance, and scalable content delivery.
The fifth research question addresses multi-device localization, exploring how GPS, VIO/VSLAM, smartphone-assisted telemetry, and network connectivity can be integrated into a cooperative sensor fusion model capable of stable outdoor anchoring.
The sixth research question examines content provisioning, discovery, and geospatial indexing mechanisms required for large-scale distributed AR ecosystems.
The seventh research question evaluates technological and ecosystem-level barriers preventing mainstream adoption.
Finally, the eighth research question investigates governance, security, and spatial ownership challenges in globally distributed AR infrastructures.
Together, these questions decompose the WWAR problem into architectural, technological, and systemic dimensions.

4. System Architecture and Technical Analysis

The core of the thesis develops a layered system architecture integrating communication, software, localization, and content management components into a unified infrastructure model.

4.1. Communication Layer

Outdoor AR systems require continuous exchange of dynamic 3D assets and localization data under mobility constraints. The analysis compares TCP-based HTTP/2 with QUIC-based HTTP/3 [1] and demonstrates the advantages of HTTP/3 in terms of reduced handshake latency, improved congestion control, connection migration support, and integrated encryption.
A comparative latency estimation was conducted between HTTP/2 (TCP-based) and HTTP/3 (QUIC-based) under mobile network assumptions (4G/5G conditions, 30–80 ms RTT). The elimination of head-of-line blocking and the reduced handshake overhead of QUIC result in an estimated 8–15% reduction in initial content delivery latency and improved connection stability under network switching scenarios. For dynamic 3D asset streaming (1–5 MB GLB assets), connection migration support further reduces reconnection overhead in high-mobility contexts, improving continuity of spatial rendering sessions.
The results indicate that modern QUIC-based transport layers provide a suitable foundation for low-latency spatial data streaming within a scalable infrastructure.

4.2. Software Layer and Hybrid Client Architecture

The proposed WWAR client adopts a hybrid architecture combining web-based content distribution with native XR rendering frameworks. Purely browser-based approaches lack hardware-level access and performance guarantees, while fully native ecosystems risk fragmentation. The hybrid HTTP-based model enables interoperability, standardized content retrieval, and high-performance real-time rendering through established XR APIs.
At the rendering level, modern game engines constitute the technological backbone of the client application. A game engine provides scene management, real-time rendering pipelines, physics simulation, shader control, and hardware abstraction layers necessary for immersive extended reality experiences. The thesis evaluates both proprietary engines and open-source alternatives, with particular attention to the Godot engine. As an open-source framework, Godot offers modular architecture, active community support, cross-platform deployment capabilities, and no licensing costs.
The adoption of an open-source engine presents significant strategic advantages for a globally distributed WWAR infrastructure. First, it reduces development and deployment costs, lowering barriers for free widespread distribution. Second,

it facilitates transparency and extensibility, enabling the research community to audit, modify, and optimize spatial computing components. These characteristics align with the objective of defining a scalable and openly accessible spatial computing layer analogous to the World Wide Web.

To ensure interoperability across heterogeneous head-mounted devices, the software architecture integrates standardized XR APIs, most notably OpenXR. By abstracting hardware-specific implementations, OpenXR allows applications built on top of a game engine to operate across multiple XR platforms with minimal modification. This hardware abstraction layer mitigates ecosystem fragmentation and supports long-term scalability.

The resulting software stack combines web-compatible communication, open and extensible rendering engines, and standardized XR interfaces. This layered design enables mass distribution, cost-effective deployment, cross-device interoperability, and high-performance real-time spatial interaction.

4.3. 3D Data Formats and Content Representation

A distributed outdoor AR infrastructure requires standardized 3D representations that balance expressiveness, rendering performance, interoperability, and streaming efficiency. The thesis therefore includes a comparative evaluation of widely adopted formats such as glTF/GLB, USDZ, COLLADA, OBJ, and X3D, analyzing them in terms of binary compactness, support for physically based rendering (PBR), animation capabilities, metadata extensibility, and suitability for real-time transmission.

Legacy formats such as OBJ and COLLADA provide broad compatibility and simplicity but lack optimized binary encoding and efficient streaming mechanisms. Their text-based structure increases bandwidth requirements and parsing overhead, limiting scalability in large-scale distributed scenarios. X3D offers richer scene description capabilities but introduces additional structural complexity and limited adoption in real-time XR pipelines.

USDZ was examined due to its strong ecosystem integration and support for complex scene hierarchies and material definitions. While technically robust, its tight coupling with specific platform ecosystems reduces interoperability in heterogeneous multi-device environments.

In contrast, glTF (particularly in its binary GLB form) [2] emerged as the most suitable candidate for large-scale outdoor AR. Its compact binary encoding, native support for PBR materials, efficient GPU pipeline integration, and compatibility with modern rendering engines enable low-latency streaming and reduced bandwidth consumption. Furthermore, its extensibility model allows inclusion of custom metadata, such as geospatial coordinates and anchor references, which are essential for persistent outdoor alignment.

A comparative size analysis of equivalent 3D assets shows that binary GLB files typically achieve 20–35% smaller footprint compared to text-based OBJ or COLLADA representations, while preserving PBR materials and animation data. This reduction directly impacts bandwidth consumption and streaming latency in distributed outdoor scenarios, particularly under mobile connectivity constraints.

The thesis concludes that standardized, binary, and engine-compatible 3D formats represent a foundational requirement for scalable georeferenced content delivery, ensuring both rendering efficiency and cross-platform interoperability within the WWAR framework.

4.4. Multi-Device Sensor Fusion and Spatial Anchoring

Accurate outdoor spatial anchoring represents one of the most critical technical challenges addressed in this thesis. The work analyzes multiple localization paradigms drawn from both industry and academic research, including GPS-based positioning, Visual-Inertial Odometry (VIO), Visual SLAM (VSLAM), cloud-assisted relocalization, and cross-view geo-localization techniques.

GPS provides globally consistent absolute positioning but suffers from meter-level inaccuracies, multipath effects, and signal degradation. Vision-based approaches such as VIO provide high-frequency relative pose estimation by combining camera input and inertial measurements. VSLAM extends this capability by building persistent maps of the environment and enabling relocalization [3]. However, these methods accumulate drift over time and may fail under dynamic lighting or texture-poor conditions.

Recent research directions, including satellite-to-ground cross-view geo-localization and Google

Street View-assisted refinement, were examined as complementary techniques capable of aligning local visual maps with global coordinates. While promising, these approaches introduce computational overhead and dependence on external datasets.

The thesis proposes a sensor fusion strategy that integrates GPS for global anchoring, VIO/VSLAM for high-frequency local tracking, and smartphone-assisted telemetry to improve overall performances. The smartphone acts as a complementary sensing unit, enabling cooperative localization and enhancing robustness through multi-device communication.

Standalone GPS positioning in urban environments typically exhibits 2–5 m error under standard conditions. The integration of VIO/VSLAM reduces short-term relative pose error below 1% drift per traveled distance, while periodic GPS correction constrains long-term global deviation. The proposed cooperative multi-device fusion strategy is therefore expected to achieve sub-meter anchoring stability in controlled outdoor scenarios, assuming sufficient visual features and stable satellite visibility.

This cooperative multi-device architecture mitigates the limitations of individual tracking methods. Absolute positioning constrains long-term drift, while relative visual tracking ensures smooth real-time interaction. The analysis indicates that sub-meter outdoor anchoring is technically achievable through coordinated fusion and cross-device synchronization, shifting the focus from single-device optimization to system-level localization design.

4.5. Content Provisioning, Discovery and Indexing

A globally distributed AR infrastructure requires content management mechanisms that extend traditional web paradigms into spatially anchored environments. The thesis analyzes existing web indexing models based on URLs and centralized servers, comparing them with geospatial databases, distributed cloud architectures, and proximity-based discovery systems.

Conventional web systems rely on uniform resource identifiers and globally accessible endpoints without geographic constraints. In contrast, outdoor AR content must incorporate geospatial metadata, spatial boundaries, and contextual visibility rules. The concept of an Area of Use is introduced to regulate spatial ownership, visibility range, and contextual activation.

Cloud-based distributed indexing architectures are evaluated as scalable solutions for managing georeferenced 3D assets. Such systems enable persistent storage, global synchronization, and cross-session alignment. However, exclusive dependence on centralized indexing may introduce latency and scalability challenges in dense urban environments.

To address contextual discovery at local scale, the thesis examines proximity-based mechanisms such as Bluetooth Low Energy (BLE) beacons. BLE offers low-energy contextual signaling suitable for lightweight discovery, area activation, and initial content retrieval.

Research literature on edge computing, distributed spatial anchors, and decentralized content delivery networks provides further insight into architectural design. By combining cloud-based indexing with local proximity signaling, the proposed hybrid model preserves web-level scalability while enabling geographically constrained content visibility and persistent spatial alignment.

The resulting indexing architecture extends traditional Web principles by integrating URLs with geospatial anchoring and contextual filtering, defining a core component of the World Wide Augmented Reality paradigm.

4.6. Governance, Adoption Barriers and Systemic Constraints

The thesis recognizes that technical feasibility alone is insufficient for large-scale deployment. Several structural barriers are identified:

- Hardware limitations in ergonomics, battery life, and outdoor visibility;
- Network latency and bandwidth constraints;
- Ecosystem fragmentation and lack of unified standards;
- Security and spatial ownership concerns.

To address governance challenges, the work proposes adapting web security principles to spatial computing. Public-key infrastructures, encrypted communication channels, and spatial authorization boundaries are identified as essential components to prevent malicious content injection and preserve user trust in a distributed AR ecosystem.

The analysis emphasizes that large-scale adoption depends not only on improved hardware but on interoperability, standardization, and robust governance frameworks.

5. Main Contributions

The thesis provides the following primary contributions:

1. A formal architectural model for a globally distributed World Wide Augmented Reality infrastructure;
2. A system-level validation of communication, rendering, and indexing strategies suitable for large-scale outdoor AR;
3. A multi-device sensor fusion framework integrating GPS, VIO/VSLAM, and smartphone-assisted telemetry for improved outdoor localization;
4. A conceptual extension of Web principles into a spatial computing paradigm, redefining resource identification through geospatial anchoring.

Unlike device-centric AR approaches, this work shifts the focus toward infrastructure-centric design, demonstrating that coordinated integration of existing technologies can enable persistent and scalable outdoor AR systems.

6. Comparison with the State of the Art

The proposed WWAR framework has been evaluated against both current commercial solutions and existing academic research.

From a commercial perspective, platforms such as browser-based AR engines and mobile SDKs provide partial solutions focused on content visualization, marker-based tracking, or mobile location-based AR. However, these systems typically remain platform-specific, lack persistent geospatial indexing at scale, and do not provide an integrated infrastructure model combining communication protocols, multi-device localization, and governance mechanisms.

From a research perspective, existing frameworks explore individual components of outdoor AR systems, such as location-based advertising platforms, mobile AR toolkits, or SLAM-based localization pipelines. While these works contribute valuable technical advancements, they often address isolated subsystems rather than defining a unified, web-scalable infrastructure model.

The main distinction of the proposed WWAR framework lies in its system-level integration. Instead of optimizing a single localization method or content distribution strategy, this thesis formalizes an infrastructure-centric architecture that combines standardized communication layers, hybrid application models, multi-sensor fusion, distributed geospatial indexing, and spatial governance principles.

Compared to isolated AR frameworks, the proposed WWAR paradigm enables persistent georeferenced content with infrastructure-level interoperability, shifting AR from application-centric deployment to network-level spatial computing.

7. Limitations

Despite its architectural coherence, the proposed WWAR framework presents several limitations.

First, cross-view geo-localization techniques depend on large-scale external datasets (e.g., satellite imagery or street-level imagery), introducing computational overhead and data availability constraints.

Second, BLE-based proximity mechanisms require physical infrastructure deployment, increasing operational costs in dense environments.

Third, sub-meter accuracy assumptions rely on favorable satellite visibility and sufficient environmental visual features; highly dynamic or texture-poor scenes may degrade performance.

Finally, large-scale indexing infrastructures raise privacy and regulatory challenges related to spatial ownership and persistent georeferenced content.

These limitations highlight that while technically feasible, large-scale WWAR deployment requires economic, infrastructural, and regulatory alignment.

8. Future Directions

Future research should extend this architectural model through quantitative validation and large-scale simulations. Performance benchmarking under varying user densities, bandwidth constraints, and concurrency scenarios would provide empirical confirmation of scalability assumptions.

Additional research directions include:

- Edge-computing-assisted localization and 5G-based synchronization,
- AI-driven cross-view geo-localization techniques,
- Privacy-preserving localization and spatial indexing,

- Standardized spatial authentication frameworks.

The progressive convergence of communication, localization, and governance technologies suggests that the emergence of a globally distributed outdoor AR layer is a matter of consolidation rather than invention.

9. Architectural Synthesis and System-Level Validation

The systematic investigation of the eight research questions confirms that a globally distributed outdoor augmented reality infrastructure is technically feasible within current technological constraints.

The architectural analysis demonstrates that existing web communication protocols, when optimized for mobility and low latency, can sustain dynamic spatial data transmission. The software and data representation evaluation confirms that hybrid client architectures and standardized 3D formats enable interoperable and scalable content delivery.

The localization study establishes that sub-meter outdoor anchoring can be achieved through coordinated multi-device sensor fusion combining absolute and relative positioning methods. The content indexing analysis shows that geospatial metadata and proximity-based discovery mechanisms can extend traditional web paradigms into spatial computing environments.

The assessment of adoption barriers and governance challenges further indicates that large-scale deployment depends on interoperability, standardization, and trust mechanisms rather than on fundamental technological breakthroughs.

Overall, the research questions collectively validate the central hypothesis of this thesis: the emergence of a World Wide Augmented Reality layer is not limited by missing technologies, but by the need for architectural consolidation, system-level integration, and coordinated standardization.

10. Conclusions

This research demonstrates that the realization of a persistent, scalable, and globally accessible outdoor augmented reality infrastructure is technically plausible using existing technologies. The principal challenge lies in system integration, interoperability, and governance rather than in the absence of enabling components.

By proposing a unified architectural framework and validating its technological foundations, this thesis contributes to the consolidation of spatial computing infrastructures and provides a structured baseline for future research and large-scale deployment.

The transition from web-based information browsing to geospatially anchored digital interaction represents a structural evolution in human–computer interaction. The technical groundwork for this evolution is already available; its large-scale realization depends on coordinated architectural development and standardization efforts.

11. Acknowledgements

I would like to express my sincere gratitude to my supervisor, Professor Marco Gribaudo, for his guidance, support, and feedback throughout this research. I am also deeply thankful to his PhD student, Marco Domenico Buttiglione, for his advices and assistance in navigating the challenges of this project. Their expertise and mentorship have been essential to the completion of this work.

References

[1] Mike Bishop. Rfc 9114: HTTP/3. Internet Engineering Task Force (IETF), 6 2022. Status: Internet Standard.

[2] Khronos Group. glTF 2.0 specification. Khronos Group official specification, 2017.

[3] Carlos Campos, Richard Elvira, Juan J. Gómez Rodríguez, José M. M. Montiel, and Juan D. Tardós. ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM. *IEEE Transactions on Robotics*, 37(6):1874–1890, 2021.