iXR OSC & NDI: An Artist-Centric Framework for XR Interaction via Gaze and Hand Tracking

Gwangyu Lee*
MARTE Lab. Dept. of Multimedia
Dongguk University

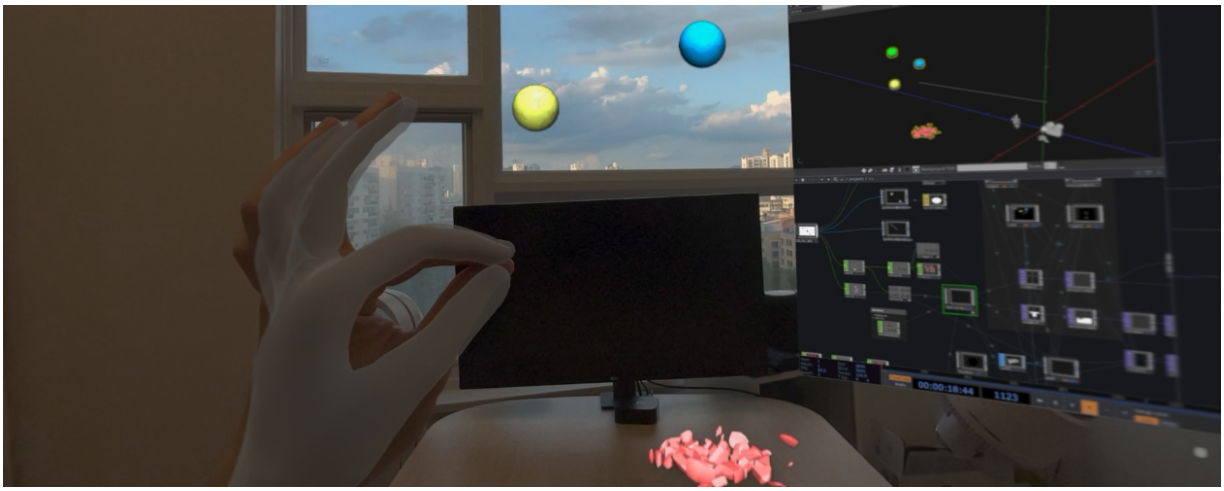


Figure 1: Overview of the MR Interaction System Featuring Gaze and Pinch-Based Control

ABSTRACT

With the release of Apple Vision Pro, interest in Mixed Reality (MR) has surged, along with the growing demand for Extended Reality (XR) content that spans Virtual Reality (VR), Augmented Reality (AR), and MR. This paper presents the iXR OSC & NDI system, designed for real-time XR interaction using OSC and NDI protocols, enabling seamless integration with tools like TouchDesigner. The system supports gaze-based interaction by combining eye tracking and pinch gestures, allowing intuitive control of virtual objects without complex programming. Tailored for artists, it facilitates easy creation and manipulation of XR content and proposes a scalable framework for use in interactive performances, exhibitions, and real-time media art environments.

Index Terms: Gaze Interaction, Apple Vision Pro, XR Content Creation, OSC, NDI, Artist-Centered Design.

1 INTRODUCTION

Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR) have long defined the technological and conceptual frameworks of spatial interfaces, explored by researchers in engineering, computer science, and human-computer interaction (HCI). In recent years, the term Extended Reality (XR) has emerged as a comprehensive umbrella term encompassing VR, AR, and MR. The rapid evolution of technology has led to the development of more affordable, lightweight devices and increasingly powerful software, enabling the widespread application of XR across domains such as education, healthcare, and the arts [8].

The release of Apple Vision Pro in 2024 marked a pivotal moment in the mainstream adoption of XR, catalyzing public interest and demand for immersive content. XR technology facilitates real-time interaction between physical and digital environments, seamlessly integrating virtual elements into the user's surroundings. Advancements in sensor and camera systems now allow for precise recognition of objects and gestures, enabling intuitive, controller-free interaction.

This work presents the development and application of iXR OSC & NDI, an XR content creation system optimized for Apple Vision Pro. Traditional XR development on this platform typically requires proficiency in programming environments such as Unity (a cross-platform game engine developed by Unity Technologies) or Swift (Apple's general-purpose language), along with a multi-stage build process, which poses significant barriers for media artists.

iXR OSC & NDI addresses these challenges by supporting Open Sound Control (OSC), a protocol widely used for real-time control in media art, and Network Device Interface (NDI), a standard for high-quality video streaming over networks. OSC conveys numerical tracking data (hand positions, gaze vectors) while NDI carries the rendered video frames; details are provided in Section Sec. 2.1. This enables seamless integration with artist-friendly platforms such as TouchDesigner (a node-based visual programming language for real-time multimedia) and Max/MSP/Jitter (a visual programming language for music and multimedia), allowing artists to focus on creative processes rather than complex coding workflows.

This paper analyzes the system architecture and implementation of iXR OSC & NDI, exploring its application in fields such as interactive performance, media art, and real-time exhibitions. The paper also discusses its influence on creative workflows and its scalability across multiple Apple Vision Pro devices, offering future directions for XR content creation.

2 DESIGN OF IXR OSC & NDI

iXR OSC & NDI was developed in the Unity environment, utilizing the OscJack package for OSC [7] and the KlakNDI package

*e-mail: gwangyulee@dongguk.edu

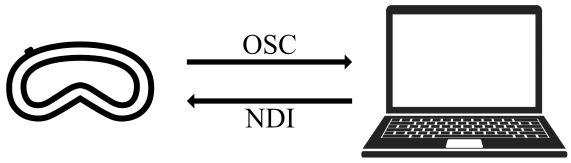


Figure 2: XR Data Integration Flow

for NDI [6]. While the official NDI Software Development Kit (SDK) is currently available only for iOS and macOS, it was successfully adapted for visionOS due to its architectural compatibility with iOS. A demonstration project showcasing KlakNDI on visionOS has been made publicly available on GitHub [3], and iXR OSC & NDI was officially released on the Apple App Store on August 23, 2024 [5].

2.1 Overall Structure

Hand and gaze tracking data captured by Apple Vision Pro are transmitted via OSC to a host computer, which generates the corresponding graphics. OSC carries numerical control data (for example, 3D hand coordinates and gaze vectors), while NDI transmits the rendered frames back to the headset. Offloading rendering to an external machine distributes computational load and enables higher-fidelity visuals and smoother frame rates.

This architecture also supports collaborative environments, where multiple artists on the same network can share and receive visual output via NDI without building or deploying code. Such a setup facilitates real-time interaction, monitoring, and rapid iteration of XR content. The overall data integration process—capturing hand and gaze input, generating visuals, and streaming the output—is illustrated in Fig. 2, which outlines the XR Data Integration Flow that underpins this system.

2.2 UX/UI Design

Upon launching iXR OSC & NDI, a dialogue window prompts users to input the IP address and port number to configure network settings and select an NDI source for display on Apple Vision Pro. Users may also choose whether their hand data will be visualized in the XR environment—an option particularly useful for VR-centric content development. Clicking the ”Start” button transitions the system into XR mode, initiating hand-tracking data and gaze vectors transmission to the designated IP and port. Fig. 3 shows the configuration dialogue window, where users can set these parameters before entering XR mode.

2.3 Multimodal Interaction System

To achieve truly immersive and intuitive XR experiences, iXR OSC & NDI integrates multiple input modalities—specifically hand gestures and gaze direction—into a unified interaction framework. Hand-tracking data and event-driven gaze vectors are both transmitted via OSC, allowing synchronized, real-time control over visual and physical transformations. Such multimodal designs resonate with the “superpower interaction” framework, in which sensory and motor extensions are grounded in naturalistic user expectations to maintain immersion and ease of control [4].

2.3.1 Hand Tracking

Apple Vision Pro’s hand-tracking feature allows for precise recognition of user hand movement, facilitating natural interaction. In this system, hand-tracking data is transmitted in real time via OSC, enabling gesture-based manipulation of multimedia content. This

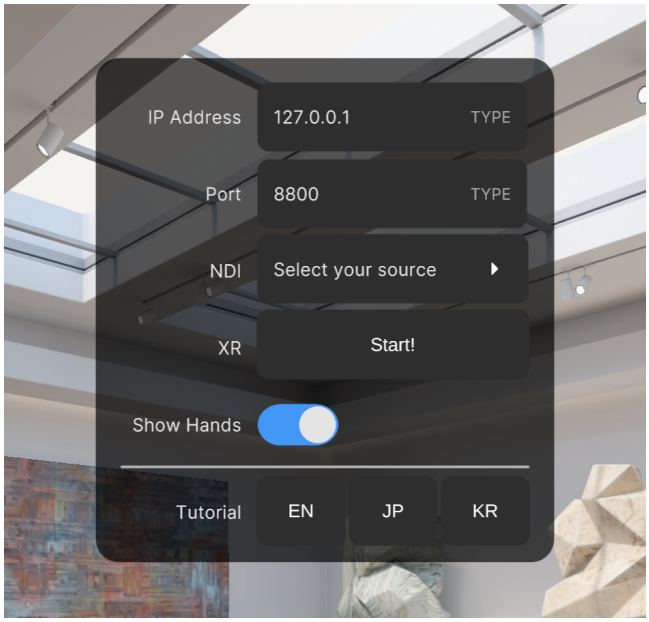


Figure 3: Dialogue Window

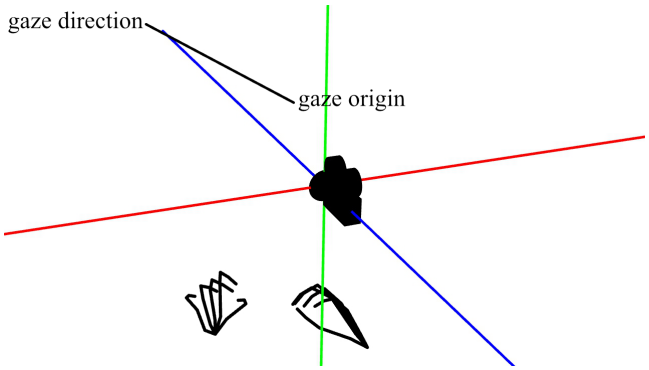


Figure 4: Visualization of Hand Positions and Gaze Vector

expands creative possibilities in domains such as live performance and immersive installation.

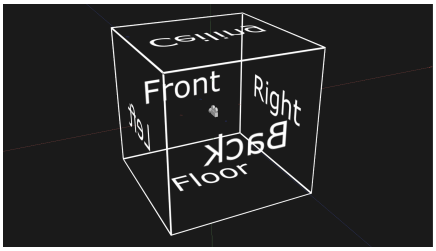
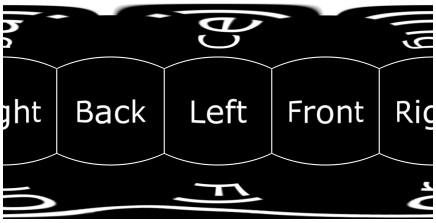
The transmitted data includes positional and rotational values for multiple points on each hand. Gesture recognition enables specific actions to be mapped to different content parameters.

2.3.2 Gaze Tracking

Although Apple Vision Pro supports eye tracking, continuous real-time gaze data is restricted due to privacy considerations. Developers can access gaze data only upon detecting a pinch event. In this system, these gaze vectors—comprising origin and direction in XYZ coordinates—are transmitted via OSC only when a pinch gesture is detected.

This approach facilitates intentional gaze-based interaction while respecting privacy constraints. By combining gaze direction with hand gestures, users can intuitively select and control elements in the XR environment. This multimodal interaction paradigm enhances the expressive capabilities of immersive media applications. Fig. 4 illustrates the combined visualization of hand positions and gaze vectors during interaction.

Table 1: Comparison of XR Content Rendering Perspectives

3D Geometry Viewer Scene	
Equirectangular Rendered Image	
View from Apple Vision Pro	

2.4 XR Rendering & Networking Design

This application simplifies the XR content creation process for artists with limited programming experience, enabling seamless integration into existing workflows. It utilizes the equirectangular rendering format, commonly used for 360-degree panoramic media, which maps spherical coordinates onto a rectangular image at a 2:1 aspect ratio [1]. This format enhances compatibility across platforms and allows easy adaptation from traditional 2D environments.

A visual comparison of different XR content perspectives—including the 3D scene, equirectangular output, and Apple Vision Pro view—is presented in Tab. 1 to clarify the rendering pipeline.

Dynamic alpha transparency toggles enable smooth transitions between VR and MR modes. When the background alpha transparency is set to 0, the system operates in MR mode, revealing the surrounding physical environment; when set to 1, it runs in a fully immersive VR environment. The system is compatible with major multimedia platforms such as TouchDesigner, Max/MSP/Jitter, Unity, and Unreal Engine, thereby bridging traditional media production and immersive XR environments.

3 Application in XR Media Art

A key advantage of iXR OSC & NDI is its ability to transmit computer-generated graphics to Apple Vision Pro in real time via NDI. This functionality is highly applicable to XR gaming, data visualization, and virtual production.

NDI also supports multicast transmission, enabling a single XR source to be streamed to multiple Vision Pro devices simultaneously. This capability facilitates collaborative workflows and audience participation in settings such as live concerts, performances, and exhibitions. The following sections detail XR content creation using TouchDesigner and outline potential applications.

TouchDesigner was selected as the primary platform for XR content development due to its robust visual programming features and native OSC and NDI support. The project files produced for this work are available on GitHub [2].

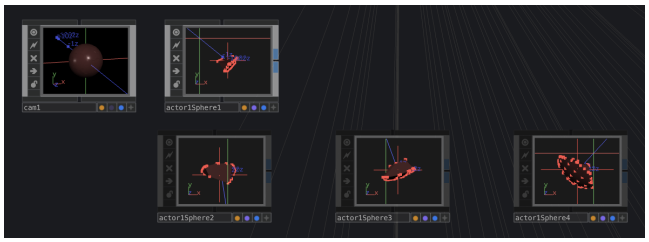


Figure 5: TouchDesigner patch for sphere-based interactive visual system

3.1 Interaction Design Concept

The XR content responds in real time to hand gestures, gaze direction, and audio input, enabling rich multimodal interaction. Users can manipulate visual elements through pinching and gaze, supporting intuitive engagement with virtual environments. Audio input captured via a computer-connected microphone is analyzed in real time; when the measured level exceeds a predefined decibel threshold, the system toggles between MR and VR modes (see Sec. 3.5 for details).

The system allows users to trigger physical events—such as object drops and fragmentation—through the combination of gaze tracking and gestures.

3.2 Visual Programming with Fragmented Sphere Objects

The primary interactive visual is a 3D fragmented sphere. Four such spheres were implemented as assets, with their positions modulated by noise using a Noise TOP. The speed of this modulation is influenced by grab gestures, while gaze input triggers object drops and fragmentation on collision with the floor. To support visual transitions, the system switches between Physically Based Rendering Material (PBR MAT) for VR and Phong Material (Phong MAT) for MR, controlled by audio input. Fig. 5 illustrates a portion of the TouchDesigner patch used to assemble multiple fragmented sphere assets into a unified interactive object.

3.3 Gesture-Based Interaction with Hand Tracking

Grab and release gestures are interpreted based on the distance between the middle fingertip and the wrist. Instead of relying on a fixed threshold to distinguish between gestures, the system dynamically maps this distance to control the noise object's translation speed along the Z-axis. In this mapping, shorter distances correspond to slower movement (interpreted as a grab), while longer distances correspond to faster movement (interpreted as a release).

For instance, a fully closed hand (grab) yields a middle fingertip–wrist distance of about 0.02 (normalized units), producing a Z-axis noise translation speed of 0.02. A fully open hand (release) measures around 2 units (speed 2), while a partially open hand (1 unit) produces a moderate speed of 1. In practice, the speed is clamped to a configurable range (e.g., [0.1, 0.4]) to account for tracking calibration differences and prevent extreme motions. This design follows prior work highlighting the benefits of preserving sensorimotor regularities in MR interaction design, which improves predictability and reduces cognitive load for users [4].

Tab. 2 presents a visual comparison among grab, partial grab, and release gestures, highlighting differences in hand posture and the corresponding movement speed. A motion trail effect (motion blur) was intentionally applied to the images for the purpose of this paper, to illustrate relative speed differences—it was not present in the actual system.

Table 2: Comparison of Hand Gestures and Associated Movement Speeds

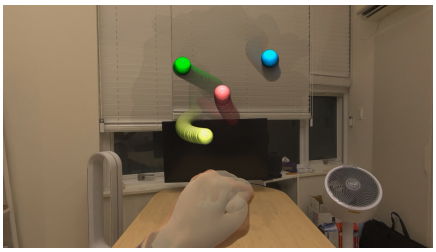
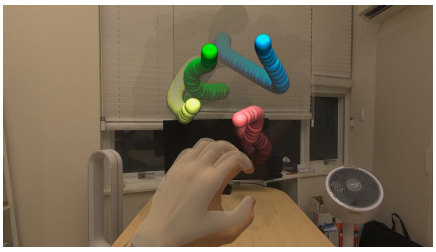
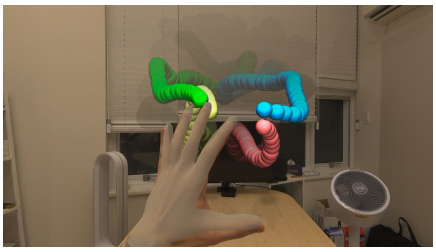
Grab (Slow Movement)	
Partial Grab (Medium Movement)	
Release (Fast Movement)	

Table 3: Comparison of Gaze and Gaze & Pinch Interactions

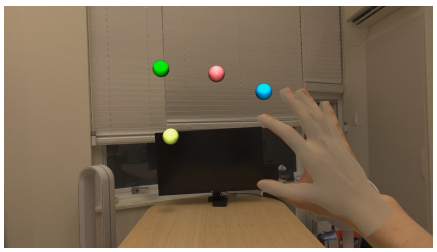
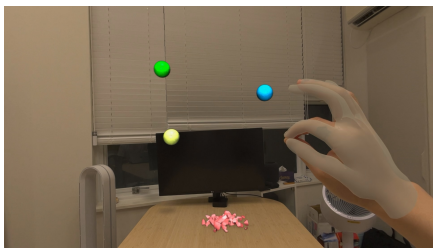
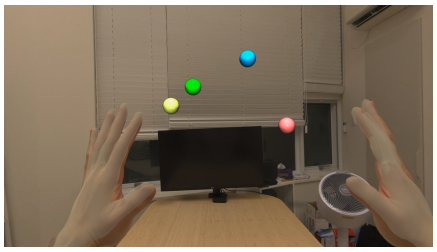
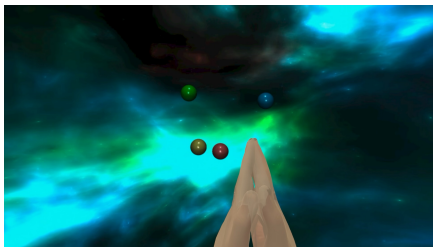
Gaze	
Gaze & Pinch	

Table 4: Comparison of MR to VR Transition by Audio

Before Clapping (Phong MAT)	
After Clapping (PBR MAT)	

3.4 Gaze-Based Interaction and Control

When the user gazes at a sphere and performs a pinch gesture, the system calculates the distance between the gaze vector and the sphere's position. If the distance falls within a defined threshold, the sphere drops under simulated gravity. A Bullet Solver COMP, which enables real-time rigid body dynamics simulation, is used for physics simulation. Upon ground contact, the sphere shatters into fragments that scatter realistically.

Tab. 3 compares the two interaction modes, gaze alone versus gaze combined with a pinch gesture, illustrating how the latter triggers a dynamic physical response.

3.5 Audio-Based Interaction

User sound input is analyzed in real time via a computer-connected microphone (hand and gaze tracking use Apple Vision Pro's built-in sensors). When sound exceeds a specified decibel threshold, alpha transparency toggles between 0 (MR) and 1 (VR). An alpha transparency of 0 activates MR mode with Phong MAT, while an alpha transparency of 1 activates VR mode with PBR MAT. This controls immersive mode transitions and associated material properties based on audio cues.

This interaction design allows users to switch between MR and VR modes without a controller, using only their sound as an input. When the user claps or makes a sufficiently loud sound, the system detects the audio threshold and switches the environment accordingly. Tab. 4 shows the transition from MR (Phong MAT) to VR (PBR MAT), triggered by a single clap sound.

4 Conclusion and Future Work

XR is no longer merely a technological trend—it is fundamentally reshaping how humans perceive, experience, and interact with digital media. By merging physical and virtual realms and enabling active user participation, XR provides immersive experiences beyond traditional audiovisual media. In the arts, XR is becoming a core medium for creativity and user engagement. Its spatial interactivity enables users not just to view content, but to step inside and interact with it. The increasing demand for XR content is driven by this evolving paradigm. As devices like Apple Vision Pro become more accessible, opportunities for intuitive, sensor-rich communication expand. This marks a shift toward more immersive and participatory media environments.

This paper introduced iXR OSC & NDI, an XR content creation tool designed for the Apple Vision Pro. Unlike conventional XR development, which typically requires expertise in Unity or Swift, this system leverages OSC and NDI protocols to enable artists to develop content using tools they already know. This lowers the barrier to entry and fosters more inclusive content creation.

As shown in Fig. 6, the system supports receiving OSC data from multiple Apple Vision Pro users simultaneously while sending processed visual output via NDI. This architecture enables collaborative XR experiences where multiple users interact in a shared virtual environment with real-time data synchronization and high-quality rendering offloaded to external computers.

Beyond individual content creation, iXR OSC & NDI can also

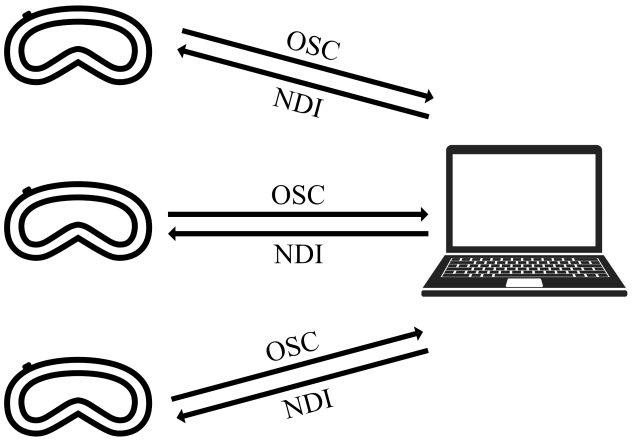


Figure 6: Receiving OSC from Multiple Users & Sending NDI

function as a network-based XR streaming solution. Its multicast capability supports deployment scenarios such as XR concerts, where content from a single source is shared across multiple headsets. For example, an artist using TouchDesigner can stream live visuals to multiple users, supporting remote collaboration and shared experiences. This system is also well-suited for educational and theatrical contexts, including virtual galleries, online lectures, and distributed performance art.

As discussed in Sec. 2.1, offloading rendering to an external computer distributes the processing load and facilitates collaborative workflows. While standalone headsets may experience limitations in real-time graphic rendering due to processing overload, offloading visual computation to a connected computer helps distribute the processing load more efficiently—resulting in higher-quality visuals and smoother performance.

Additionally, future works could examine how maintaining sensorimotor regularities in multimodal XR systems influences learnability, immersion, and perceived agency, as suggested in prior MR interaction research [4]. We also aim to explore real-world applications in education, performance, and interactive media art through cross-disciplinary collaboration. This research contributes to the accessibility and scalability of XR content development and offers practical value across artistic, educational, and academic domains.

REFERENCES

[1] A. Araújo. Guidelines for drawing immersive panoramas in equirectangular perspective. In *Proceedings of the 8th International Conference on Digital Arts*, ARTECH '17, p. 93–99. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3106548.3106606 3

[2] G. Lee. iXR-OSC-NDI. https://github.com/gwangyu-lee/iXR-OSC-NDI, 2025. Accessed: 2025-07-18. 3

[3] G. Lee. KlakNDI-visionOS-Demo. https://github.com/gwangyu-lee/KlakNDI-visionOS-Demo, 2025. Accessed: 2025-07-18. 2

[4] J. Li and P. O. Kristensson. On the benefits of sensorimotor regularities as design constraints for superpower interactions in mixed reality. *IEEE Transactions on Visualization and Computer Graphics*, 31(5):2568–2578, 2025. doi: 10.1109/TVCG.2025.3549876 2, 3, 5

[5] A. A. Store. iXR OSC & NDI. https://apps.apple.com/us/app/ixr-osc-ndi/id6642664920, 2025. Accessed: 2025-07-18. 2

[6] K. Takahashi. KlakNDI. https://github.com/keijiro/KlakNDI, 2024. Accessed: 2025-07-18. 2

[7] K. Takahashi. OscJack. https://github.com/keijiro/OscJack, 2024. Accessed: 2025-07-18. 1

[8] A. Çöltekin, I. Lochhead, M. Madden, S. Christophe, A. Devaux, C. Pettit, O. Lock, S. Shukla, L. Herman, Z. Stachoň, P. Kubíček, D. Snopková, S. Bernardes, and N. Hedley. Extended reality in spatial sciences: A review of research challenges and future directions. *ISPRS International Journal of Geo-Information*, 9(7), 2020. doi: 10.3390/ijgi9070439 1