



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



# UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Escultura

Telepresencia en las artes visuales y multimedia: cómo  
potenciar las interacciones a distancia.

Trabajo Fin de Máster

Máster Universitario en Artes Visuales y Multimedia

AUTOR/A: Palacio Restrepo, Luis Miguel

Tutor/a: Sanmartin Piquer, Francisco Javier

CURSO ACADÉMICO: 2023/2024

## Resum

La comunicació natural o física és inherentment tridimensional, i abasta no sols els intercanvis verbals, sinó també els senyals no verbals, com les expressions facials, el contacte visual i els gestos. No obstant això, l'arribada dels cascos de realitat virtual (HMD) ha plantejat desafiaments per a les converses realistes cara a cara. Els HMD obstruïxen una part important dels rostres dels participants, la qual cosa dificulta la creació d'imatges convinents de rostres sencers a causa dels angles de captura extrems i les distorsions de les lents.

En investigacions recents, s'ha posat l'enfocament en la reconstrucció de rostres ocults davall dels \*HMD. Si bé les solucions d'última generació existents demostren resultats de reconstrucció 3D fotorrealistes, sovint requereixen equips de laboratori d'alt cost i recursos computacionals substancials.

D'altra banda, este treball se centra en la implementació d'una aplicació que permet la visualització en temps real d'entorns 3D per a crear teleconferències immersives, on la virtualització dels elements es limita a la resolució dels sensors, sense la necessitat d'incorporar models 3D o avatars; així, millorant les interaccions remotes a través de les arts visuals i els mitjans interconnectats realistes. Algunes de les tecnologies involucren \*WebRTC, sensors de profunditat i motors en temps real.

**Paraules clau** Telepresència; Temps real, Sensors de profunditat; Entorns 3D

## **Abstract**

Natural or physical communication is inherently three-dimensional, encompassing not only verbal exchanges but also non-verbal cues such as facial expressions, eye contact, and gestures. However, the advent of virtual reality headsets (HMDs) has posed challenges for realistic face-to-face conversations. HMDs obstruct a significant portion of participants' faces, making it difficult to create convincing images of entire faces due to extreme capture angles and lens distortions.

Recent research has focused on reconstructing faces hidden beneath HMDs. While existing state-of-the-art solutions demonstrate photorealistic 3D reconstruction results, they often require high-cost laboratory equipment and substantial computational resources.

On the other hand, this work focuses on the implementation of an application that enables real-time visualization of 3D environments to create immersive teleconferences, where the virtualization of elements is limited to the resolution of the sensors, without the need to incorporate 3D models or avatars; thus, enhancing remote interactions through visual arts and realistic interconnected media. Some of the technologies involve WebRTC, depth sensors, and real-time engines.

**Keywords** Telepresence; Realtime, Depth Sensors; 3D Environments

## Resumen

La comunicación natural o física es inherentemente tridimensional, y abarca no solo los intercambios verbales, sino también las señales no verbales, como las expresiones faciales, el contacto visual y los gestos. Sin embargo, la llegada de los cascos de realidad virtual (HMD) ha planteado desafíos para las conversaciones realistas cara a cara. Los HMD obstruyen una parte importante de los rostros de los participantes, lo que dificulta la creación de imágenes convincentes de rostros enteros debido a los ángulos de captura extremos y las distorsiones de las lentes.

En investigaciones recientes, se ha puesto el enfoque en la reconstrucción de rostros ocultos debajo de los HMD. Si bien las soluciones de última generación existentes demuestran resultados de reconstrucción 3D fotorrealistas, a menudo requieren equipos de laboratorio de alto costo y recursos computacionales sustanciales.

Por otro lado, este trabajo se centra en la implementación de una aplicación que permite la visualización en tiempo real de entornos 3D para crear teleconferencias inmersivas, donde la virtualización de los elementos se limita a la resolución de los sensores, sin la necesidad de incorporar modelos 3D o avatares; así, mejorando las interacciones remotas a través de las artes visuales y los medios interconectados realistas. Algunas de las tecnologías involucran WebRTC, sensores de profundidad y motores en tiempo real.

**Palabras clave** Telepresencia; Tiempo real, Sensores de profundidad; Entornos 3D



# Contenidos

<b>1</b>	<b>Introducción</b>	<b>7</b>
1.1	Objetivos . . . . .	7
1.2	Metodología . . . . .	8
1.2.1	Análisis de requerimientos . . . . .	8
1.2.2	Diseño del sistema y desarrollo . . . . .	8
1.2.3	Prototipado . . . . .	9
1.2.4	Análisis y planteamiento de casos de uso . . . . .	9
<b>2</b>	<b>Contexto</b>	<b>11</b>
2.1	Estado del Arte . . . . .	12
2.1.1	Teleconferencias . . . . .	12
2.1.2	Telepresencia robótica . . . . .	13
2.1.3	Telepresencia en Realidad Virtual . . . . .	14
2.1.4	Telepresencia en Realidad Aumentada . . . . .	16
2.1.5	Sistemas hápticos de telepresencia . . . . .	17
<b>3</b>	<b>Memoria</b>	<b>19</b>
3.1	Prototipado . . . . .	19
3.1.1	Diseño del sistema . . . . .	20
3.1.2	Desarrollo . . . . .	22
<b>4</b>	<b>Conclusiones</b>	<b>27</b>
4.1	Análisis general . . . . .	27
4.2	Casos de uso (Objetivos de Desarrollo Sostenible) . . . . .	28
4.3	Pasos a seguir . . . . .	29



# 1 Introducción

## 1.1 Objetivos

- Explorar el potencial del concepto de telepresencia aplicado en Artes visuales y multimedia, abriendo la puerta para interacciones en tiempo real entre artistas y usuarios; democratizando el acceso al arte e incentivando expresiones innovadoras a través del desarrollo de herramientas que integran tecnologías emergentes.
- Desarrollar un sistema de telepresencia que permita la integración de múltiples dispositivos, incluyendo Realidad Virtual, Realidad Aumentada, y actuadores físicos para crear la ilusión de presencialidad física en ambientes a larga distancia; trascendiendo barreras geográficas y enriqueciendo la experiencia de los usuarios.
- Exponer aspectos positivos y negativos del uso práctico de la telepresencia, aquellos como la disminución de interacciones face-to-face, contemplando elementos de privacidad, y analizando las limitaciones técnicas actuales con respecto a calidad visual, latencia, y compatibilidad; haciendo énfasis en sus implicaciones éticas y en la importancia de tener un balance entre las interacciones físicas, digitales, y “figitales”.
- Proponer diferentes caminos de evolución tecnológica en términos de telepresencia que permitan el desarrollo de aplicaciones en diferentes sectores, incluyendo trabajo remoto, educación, cuidado de la salud, y entretenimiento.



## **1.2 Metodología**

En esta sección se propone un enfoque sistemático para desarrollar y evaluar un prototipo diseñado para transmitir imágenes de profundidad, poniendo a prueba las posibilidades de mejorar experiencias inmersivas basadas en el concepto de telepresencia.

### **1.2.1 Análisis de requerimientos**

En la fase inicial del proyecto, se plantea una revisión exhaustiva de la literatura para obtener información sobre los sistemas de telepresencia existentes, los métodos de captura de imágenes en profundidad y las técnicas de transmisión dentro del ámbito de las artes visuales y multimedia. Entre los objetivos se busca identificar los requisitos de los usuarios, las funcionalidades básicas, los escenarios de aplicación y las posibles limitaciones relacionadas con la resolución, fidelidad de la imagen, las técnicas de compresión, el ancho de banda de la red y las capacidades de procesamiento en tiempo real. Este examen exhaustivo permitiría discernir las distintas preferencias de los usuarios al tiempo que consideramos las limitaciones técnicas, lo que en última instancia da forma a los objetivos y plantear un alcance realista del proyecto.

### **1.2.2 Diseño del sistema y desarrollo**

Basándonos en el análisis de requisitos y en una revisión exhaustiva de la literatura, iniciamos la conceptualización de la arquitectura y el diseño del sistema. Esto implica identificar los componentes de hardware y software necesarios para capturar, procesar y transmitir imágenes de profundidad. Simultáneamente, diseñamos la interfaz de usuario y los mecanismos de interacción. La arquitectura general del sistema abarca módulos para la captura, transmisión, renderización e interacción del usuario con las imágenes de profundidad.

Seleccionamos cuidadosamente una tecnología de cámara de profundidad adecuada, teniendo en cuenta factores como la resolución, la precisión y la disponibilidad. Además, se busca utilizar algoritmos diseñados para una compresión eficiente de imágenes de profundidad, logrando un equilibrio entre la calidad de la imagen y la utilización del ancho de banda. Finalmente, se implementaría un prototipo del sistema utilizando herramientas y entornos

de desarrollo de software, incorporando canales de procesamiento en tiempo real y protocolos de comunicación para la transmisión de datos.

### **1.2.3 Prototipado**

El prototipo del sistema se desarrolla de forma iterativa según las especificaciones de diseño establecidas. Esto implica la implementación de los algoritmos necesarios para la captura, procesamiento y transmisión de imágenes de profundidad, así como la integración de cualquier característica o funcionalidad adicional identificada durante la fase de diseño. Al mismo tiempo, se considera relevante la definición de un conjunto de métricas que permitan la evaluación del sistema, identificando su eficacia en la mejora de experiencias inmersivas basadas en telepresencia. Estas métricas abarcan calificaciones subjetivas de la experiencia del usuario, centrándose en la percepción de profundidad, el realismo y la satisfacción general, así como medidas objetivas de calidad de imagen, latencia y utilización de recursos del sistema.

### **1.2.4 Análisis y planteamiento de casos de uso**

Con base en los hallazgos de las primeras pruebas técnicas y posibles evaluaciones con usuarios, se identifican áreas de mejora en el prototipo. Implicando el refinamiento de los algoritmos, la optimización de la interfaz de usuario o la exploración de protocolos alternativos para la transmisión de imágenes de profundidad. Luego, el sistema se perfecciona de manera iterativa para abordar los problemas identificados, mejorar el rendimiento y la experiencia inmersiva.

Las mejoras y necesidades detectadas permiten a su vez identificar posibles contextos en los cuales la tecnología podría brindarse como una solución a problemáticas específicas; dando paso casos de uso en múltiples industrias, tales como el entretenimiento, la educación, la salud, entre otros.



## 2 Contexto

La noción de presencia ha suscitado un interés significativo tanto en los debates académicos como en los públicos. Sin embargo, los investigadores han propuesto definiciones diversas y a veces superpuestas de este concepto, lo que puede obstaculizar nuestro avance en la comprensión de los fenómenos relacionados con la presencialidad. Matthew Lombard y Matthew T. Jones presentan un marco para desentrañar estas conceptualizaciones y abogan por una terminología coherente a la hora de analizar y definir la presencia [14] (ver figura 2.1)

De este modo, se menciona que el concepto de telepresencia involucraría explícitamente el uso de tecnología, “una máquina, dispositivo u otra aplicación de las artes industriales humanas, incluyendo la televisión, la radio, el cine, el teléfono, las computadoras, la realidad virtual y los juegos de simulación; medios impresos tradicionales como periódicos, libros y revistas; y artes tradicionales como la pintura y la escultura” [1]

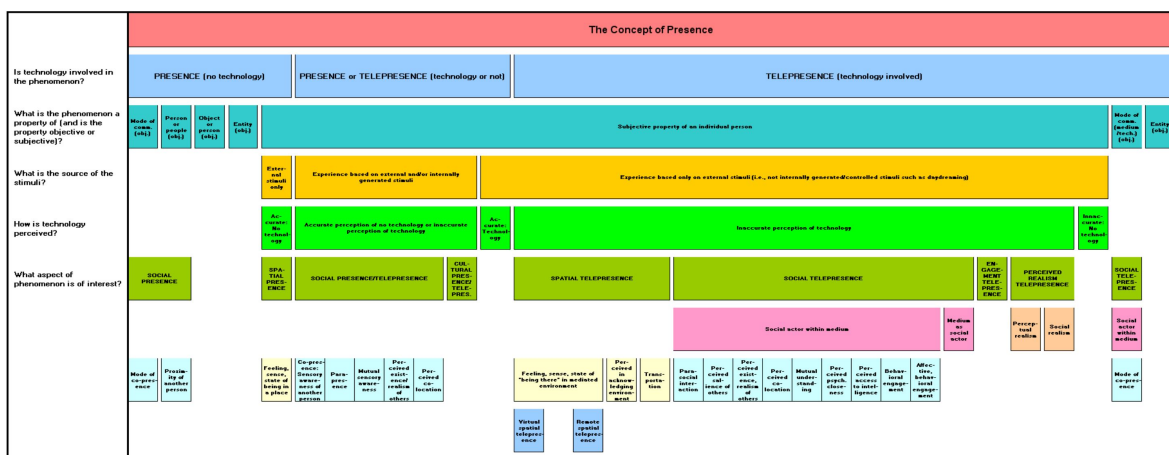


Figure 2.1: La terminología definida por Matthew Lombard & Matthew T. Jones se expone a detalle en el siguiente enlace: <https://matthewlombard.com/presence-definitions/>

## 2 Contexto

La telepresencia se ha descrito como la experiencia humana de estar presente en un lugar en tiempo real, a pesar de estar físicamente en otra locación. En otras palabras, es la sensación de estar inmerso en un entorno distante, aunque uno se encuentre físicamente en otro lugar. Esta ilusión perceptiva se evoca a través de tecnologías multimedia que incluyen el sonido, la visión y el tacto para crear la sensación de presencia [19].

Ya sea experimentar un mundo virtual, participar en una teleconferencia o controlar un robot de forma remota, la telepresencia difumina los límites entre los espacios físicos y virtuales, lo que nos permite interactuar con contextos distantes como si estuviéramos allí [18]. Adicionalmente, es relevante mencionar que a pesar de que experiencias visuales desempeñan un papel central, otros sentidos también pueden contribuir a la percepción general de la telepresencia.

### 2.1 Estado del Arte

La tecnología de telepresencia ha experimentado avances significativos que han transformado la forma en que nos conectamos y colaboramos a través de distancias físicas. Entre varias obras que mencionan el estado general de la industria y la tecnología, se puede resaltar “Immersed in Media” [13], una colección académica que explora la teoría de la presencia, la investigación y el diseño de tecnología relacionados con las experiencias inmersivas. Incluye contribuciones de académicos internacionales, cubre la teoría de la presencia, investiga los factores que contribuyen a la sensación de “estar allí” y explora el diseño de tecnología innovadora. Además, reconoce el impacto de las tecnologías emergentes como la realidad aumentada/virtual y las redes 5G en el panorama de los medios inmersivos. Adicionalmente, es válido mencionar las conferencias de tecnología como el SIGGRAPH-2023[2], donde se han llevado a varias ediciones del curso llamado “State of the Art in Telepresence” [11] [10]

#### 2.1.1 Teleconferencias

La telepresencia eleva la calidad del audio y el video en las reuniones virtuales, cerrando la brecha entre la presencia física y la colaboración remota. A medida que la tecnología continúa evolucionando, la elección entre la telepresencia y la videoconferencia tradicional se vuelve más matizada, pero ambas cumplen funciones esenciales en la comunicación moderna.

En el ámbito de las teleconferencias de realidad aumentada (RA), el objetivo es permitir que los usuarios remotos interactúen en 3D dentro de sus propios entornos físicos. Los métodos existentes a menudo priorizan la alta fidelidad (utilizando captura y reconstrucción volumétrica compleja) o la alta co-presencia (soluciones simples en cascos de realidad aumentada). Para lograr un equilibrio, este concepto propone utilizar avatares basados en video 2D de tamaño real (avatares de video) en las teleconferencias de RA. Un estudio piloto explora la ubicación óptima del avatar, considerando el impacto del campo de visión (FoV) en la proximidad percibida. Durante el estudio se ejecuta una prueba de concepto y las evaluaciones de los usuarios verifican su efectividad. Además, los estudios cuantitativos exploran los efectos del tamaño del campo de visión, con el objetivo de guiar los modelos de ubicación futuros para varios dispositivos de RA. [21]

En 2011, el sistema de telepresencia presentado por Andrew Maimone y Henry Fuchs en el Departamento de Ciencias Informáticas de la Universidad de Carolina del Norte aprovecha las cámaras de color y profundidad Microsoft Kinect™ para lograr la captura de escenas en 3D en tiempo real y una visualización en 3D estereoscópica con seguimiento de la cabeza. Cabe destacar que elimina la necesidad de que los usuarios usen equipos de seguimiento o visualización adicionales. El flujo de procesamiento acelerado por GPU del sistema maneja tareas como la fusión de datos, la generación de superficies y la corrección de color a velocidades impresionantes. [16]

### 2.1.2 Telepresencia robótica

La telepresencia robótica se refiere a una clase de dispositivos robóticos que permiten la interacción remota con entornos físicos. Estos robots suelen estar equipados con una cámara de vídeo, un sistema de audio y varios sensores. Los usuarios los controlan a través de una interfaz remota, como una computadora, una tableta o un teléfono inteligente. El propósito principal de estos robots es extender la presencia de un individuo a una ubicación diferente, lo que le permite participar virtualmente en reuniones, eventos u otras actividades sin la necesidad de viajar físicamente. Al proporcionar comunicación audiovisual en tiempo real, los robots de telepresencia mejoran la colaboración y facilitan la interacción efectiva a través de distancias geográficas. Encuentran aplicaciones en varios dominios, incluidos los negocios,

## 2 Contexto

la atención médica, la educación y las interacciones sociales.

La telepresencia robótica móvil (MRP) permite a los usuarios remotos acceder y navegar por entornos locales. Estos dispositivos MRP encuentran aplicaciones en dominios críticos como lugares de trabajo, museos, comercio, educación y atención médica. Cabe destacar que, durante la pandemia de Covid-19, facilitaron la accesibilidad a espacios cuando la asistencia física estaba restringida. Si bien los robots de telepresencia incorporan funciones autónomas como la prevención de colisiones, generalmente carecen de la capacidad de manipulación física directa del entorno. En consecuencia, su confiabilidad sigue siendo limitada, lo que dificulta su adopción generalizada. Algunas investigaciones han logrado exponer un enfoque innovador llamado Telepresencia robótica aumentada (ART, acrónimo en inglés). ART tiene como objetivo mejorar la inclusión, la accesibilidad y la independencia de los usuarios remotos de MRP al aumentar las posibilidades del entorno local. Técnicas como la realidad aumentada (RA), Internet de las cosas (IoT) y la actuación remota contribuyen a expandir el espacio para la interacción y la participación. [17]

En una colaboración entre la Universidad de Hong Kong y el University College de Londres, Ray Lc et al. exploraron un paradigma más interactivo para las actuaciones a distancia, permitiendo a los visitantes alterar el entorno del artista a través de cambios de música e iluminación. Mientras tanto, el artista interactuaba con un brazo robótico remoto. Los hallazgos clave incluyen la identidad distintiva del brazo robótico, el impacto de los cambios musicales en el comportamiento del artista y el público que se considera a sí mismo co-artista. Este trabajo informa futuras interacciones performativas virtualizadas más allá de los escenarios tradicionales. [12]

### 2.1.3 Telepresencia en Realidad Virtual

Este tipo de telepresencia implica el uso de dispositivos de realidad virtual, presentando entornos inmersivos que transportan a los usuarios dentro de un espacio virtual; usualmente compartido con otros usuarios, sin embargo, podríamos mencionar que el sentido de presencia puede variar dependiendo del nivel inmersivo que presenten los mundos virtuales, incluso en experiencias individuales.

Al usar un casco de realidad virtual, los participantes pueden experimentar la sensación

de estar presentes en el mismo lugar, aunque estén físicamente separados. El objetivo es crear una conexión perfecta entre los mundos físico y virtual, lo que permite a los usuarios interactuar como si realmente estuvieran "allí".

Adicionalmente, la comunicación en tiempo real a través de videos 360° promete permitir a los usuarios experimentar la telepresencia en entornos transmitidos de forma remota. A medida que más usuarios cambian a dispositivos móviles y se conectan a través de redes móviles, el potencial de las aplicaciones móviles se hace evidente. Algunos cascos de realidad virtual ahora incorporan capacidad móvil, lo que mejora aún más su versatilidad. Sin embargo, la transmisión de video en vivo de 360° de alta calidad a través de conexiones móviles presenta desafíos de ancho de banda. Para abordar esto, se han empleado técnicas como la transmisión adaptativa de la ventana gráfica y la renderización foveada para comprimir videos. Esta última es una técnica que aumenta la definición de las áreas en las que el ojo está mirando dentro del visor, es decir, actúa de enfoque sobre lo que el usuario está viendo en cada momento. Es importante recalcar que estos enfoques exigen una latencia ultrabaja para su efectividad, lo que históricamente ha limitado su uso en redes móviles tradicionales. En trabajos recientes, se presenta un sistema de telepresencia de realidad virtual de extremo a extremo que transmite aproximadamente videos 6K de 360° a través de radio de ondas milimétricas (mmW) 5G. El aprovechamiento de las tecnologías 5G junto con los nodos de computación de borde móviles reduce significativamente la latencia en comparación con las redes 4G existentes, lo que permite una compresión foveada eficiente y comparable a WiFi. Durante la evaluación técnica, basada en la relación señal-ruido máxima (PSNR) y FOVVideoVDP, se han demostrado transmisiones de video visualmente indistinguibles mientras se utilizaba hasta un 80% menos de datos que el video sin foveación. De este modo, el proyecto realizado por Huang et al, logra exponer un sistema de compresión de video en el contexto de una aplicación de videollamada inmersiva y telepresente [8].



### 2.1.4 Telepresencia en Realidad Aumentada

La realidad aumentada (RA), cuando se aplica a la telepresencia, combina elementos del mundo real con objetos virtuales. Permitiendo a los usuarios ver e interactuar con contenido digital dentro de su entorno real y creando una conexión sensorial realista entre lugares geográficamente distantes. En escenarios de telepresencia, la RA mejora la sensación de presencia e inmersión, haciendo que los usuarios se sientan realmente conectados a pesar de la separación física.

Un avance fundamental que mejora la telepresencia es la evolución de los cascos de realidad aumentada (HMD, por sus siglas en inglés). Los investigadores se centran cada vez más en aprovechar la realidad aumentada (RA) y la realidad virtual (RV) para mejorar las operaciones remotas y la telepresencia. Para lograr una mayor sensación de telepresencia, es fundamental realizar un análisis exhaustivo de los datos de vídeo capturados en entornos remotos. La transmisión en tiempo real de los detalles del entorno y de los objetos al usuario se vuelve esencial. En nuestro estudio exhaustivo, profundizamos en las tecnologías que facilitan la transmisión de entornos de telepresencia basados en RV/RA a través de los HMD. Estas tecnologías se dividen en varias categorías, incluidas las nubes de puntos, los vóxeles, la síntesis de vistas y la estimación de la pose. Exploramos los desafíos críticos que enfrentan los investigadores en cada categoría y presentamos los últimos avances, brindando una perspectiva intuitiva para los lectores. [9]

Por otro lado, el concepto de Realidad Aumentada (RA) enfrenta múltiples desafíos; en el contexto de la comunicación remota, la teleconferencia de RA tiene como objetivo crear una experiencia donde los usuarios separados por la distancia física puedan interactuar como si estuvieran cara a cara. Entre los diversos enfoques, el enfoque basado en avatares de video se destaca por su equilibrio entre fidelidad y copresencia. Sin embargo, surge un problema crítico: las señales no verbales, en particular las relacionadas con la dirección de la mirada, a menudo se pierden o se comunican mal en las teleconferencias entre múltiples participantes. Tal como lo exponen Xuanyu Wang et al, las soluciones existentes han explorado diferentes estrategias, pero enfrentan limitaciones [20]:

- Visualizaciones basadas en pantalla: algunas soluciones utilizan superposiciones de pantallas para indicar la dirección de la mirada. Sin embargo, estas carecen de inmersión

y pueden distraer a los usuarios de la conversación.

- Hardware adicional: otros enfoques incorporan sensores o dispositivos adicionales para rastrear la mirada. Si bien son efectivos, esta complejidad los hace menos factibles para el uso diario.
- Representación de avatares virtuales: alterar la apariencia del usuario a través de avatares puede abordar la conciencia de la mirada, pero sacrifica la autenticidad de su apariencia real.

### 2.1.5 Sistemas hápticos de telepresencia

La telepresencia háptica se refiere a una tecnología que combina la retroalimentación táctil con la presencia remota, lo que permite a los usuarios sentir y manipular objetos en un entorno distante. A diferencia de la telepresencia tradicional, que se centra principalmente en los aspectos visuales y auditivos, la telepresencia háptica agrega una capa de interacción física. Al incorporar interfaces hápticas, los usuarios experimentan retroalimentación de fuerza, textura y otras sensaciones táctiles relacionadas con los objetos con los que interactúan de forma remota. Este enfoque multidimensional mejora la sensación de estar físicamente presente en un lugar distante, lo que reduce la brecha entre el mundo real y el virtual [18].

La integración de la retroalimentación visual y háptica en los sistemas de teleoperación es crucial para ampliar las capacidades humanas en diversos entornos. Si bien la teleoperación ha demostrado ser eficaz, persisten los desafíos debido a la insuficiencia de señales sensoriales desde ubicaciones remotas. Esta limitación afecta la guía del operador. En respuesta, Fanghao Huang et al han desarrollado un Sistema de Reconstrucción y Percepción Integrada Visual/Háptica (VHI-PRS). Este sistema proporciona a los operadores información visual en 3D y una guía háptica eficaz. En concreto, un método de aumento de telepresencia basado en la visión combina modelos de nubes de puntos reales con manipuladores virtuales, evitando la renderización de mallas que consume mucho tiempo. Además, un método basado en la háptica ofrece una retroalimentación de fuerza integral, que incluye fuerzas de guía y repulsivas, lo que reduce la carga de trabajo del operador. Un estudio de usuarios valida la eficacia de este innovador sistema en tareas prácticas [7].



## 3 Memoria

En esta sección, analizaremos en profundidad el planteamiento de un sistema que aprovecha WebRTC (comunicación web en tiempo real) para permitir la transmisión de video proveniente de cámaras RGB y de profundidad. Desde las consideraciones iniciales de diseño hasta la implementación práctica, se exploran los aspectos clave de la construcción de un sistema robusto y eficiente. De este modo, se hace camino para mejorar las experiencias de telepresencia, acortar las distancias físicas y fomentar conexiones significativas a través de las distancias.

### 3.1 Prototipado

Como parte del proceso de diseño se planteó como principal requerimiento que el sistema debería tener la capacidad de transmitir imágenes a color y mapas de profundidad en simultáneo, haciéndolos accesibles a través de internet; adicionalmente, la arquitectura debería permitir agregar más dispositivos y así habilitar la transmisión en tiempo real de entornos de mayor amplitud.

Entre múltiples opciones para establecer la conexión entre los usuarios, se ha destacado la utilización de dos tecnologías: Photon Network-Fusion 2 y WebRTC. Cada uno presentaba diferentes retos en términos de implementación, el primero estaría diseñado principalmente a entornos multijugador pero con la ventaja de ser aparentemente más amigable de implementar, pues la conexión a internet ya estaría integrada; y el segundo presenta soluciones estándar en la industria del software para la realización de videollamadas y transmisión de audio.

### 3.1.1 Diseño del sistema

Después de realizar la respectiva exploración técnica, el protocolo y paquete de desarrollo a integrar en la solución final sería WebRTC; aprovechando a su vez los paquetes de desarrollo implementados por Unity: "Unity Render Streaming". La arquitectura del sistema se presenta en la siguiente subsección. (ver figura 3.1)

WebRTC (Web Real-Time Communication) es un proyecto gratuito y de código abierto soportado por Apple, Google, Microsoft, Mozilla, y muchas otras organizaciones; el cual permite a los navegadores web y las aplicaciones móviles comunicarse en tiempo real a través de interfaces de programación de aplicaciones (API). De este modo, permite que la comunicación y la transmisión de audio y video funcionen directamente dentro de las páginas web, lo que elimina la necesidad de instalar complementos o descargar aplicaciones nativas. En esencia, WebRTC permite una comunicación fluida entre usuarios a través de la web, lo que facilita las llamadas de audio y video, el intercambio de datos y más.

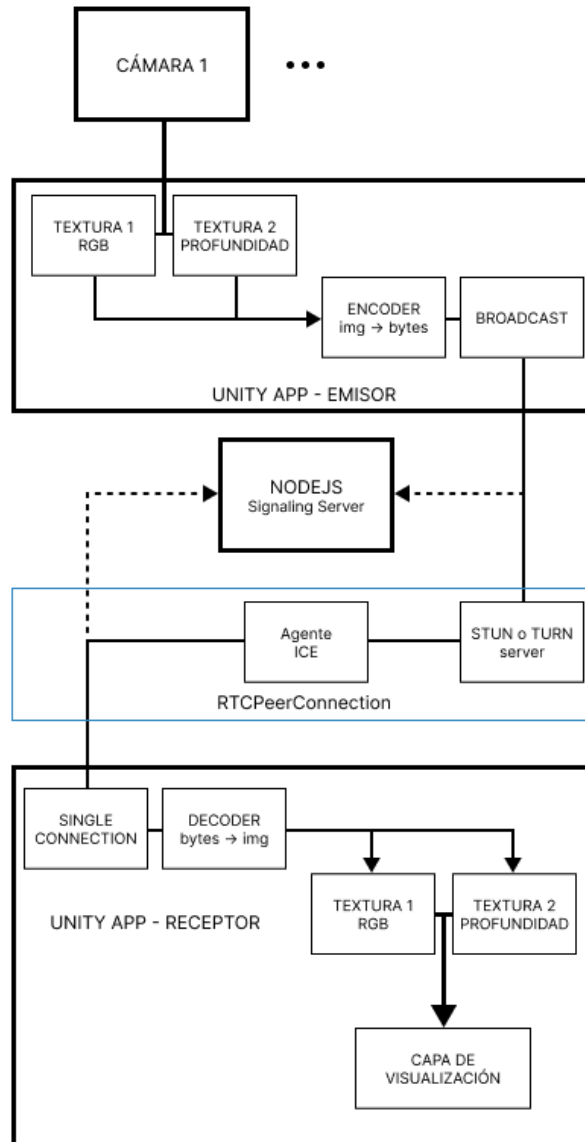


Figure 3.1: Diagrama que describe las conexiones entre usuarios. Similitudes pueden encontrarse en múltiples conferencias e investigaciones realizadas en los últimos años [15, 6, 5, 4]

### 3 Memoria

#### 3.1.2 Desarrollo

##### Configuración del sistema

Con la intención de identificar el correcto funcionamiento de las cámaras de profundidad fabricadas por Intel con referencia RealSense 435, es necesario descargar e instalar el software de visualización. En la figura 3.2 se puede observar dos dispositivos conectados, cada uno con su textura de color y profundidad respectivamente.

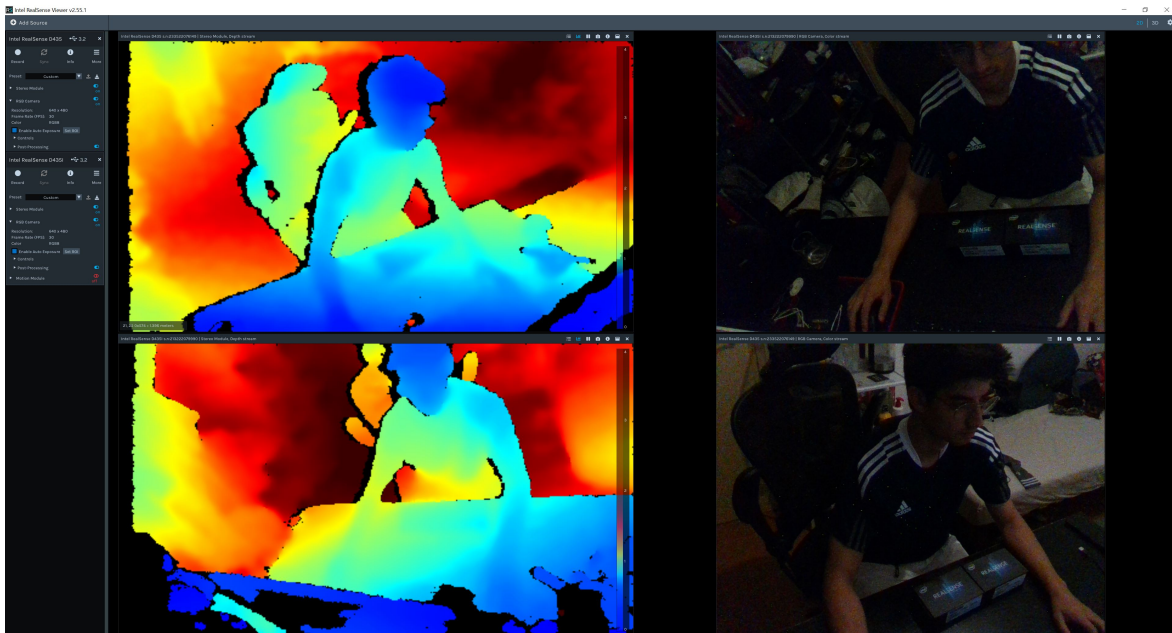
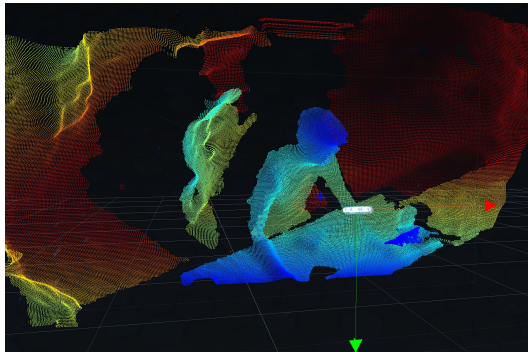
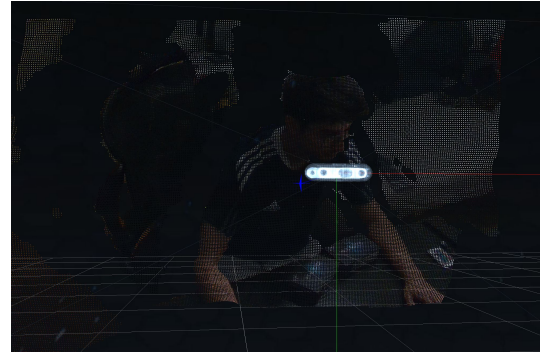


Figure 3.2: Intel RealSense View. En la parte izquierda se presentan las texturas de profundidad, utilizando un código de color conocido como "Jet"; donde el rojo representa lejanía y azul cercanía. Por otro lado, en su derecha se pueden encontrar la textura RGB correspondiente.

Del mismo modo, el visualizador de Intel permite apreciar la entrada de la cámara en tiempo real y en de modo nube de puntos; tanto el mapa de profundidad como a color, tal como se expone en la figura 3.3.



(a) Nube de puntos generada en el visualizador de Intel a partir de los datos en tiempo real que entregan las cámaras RealSense 435.



(b) Nube de puntos a color utilizada a partir de la textura RGB de la cámara.

Figure 3.3: Comparación de las nubes de puntos generadas por las cámaras Intel RealSense.

Posteriormente, se establece un entorno de desarrollo en Unity, incluyendo la comunicación a través del protocolo WebRTC y el respectivo soporte proporcionado por Intel para configurar y manipular las cámaras de profundidad. Con la intención de adquirir la señal de varias cámaras en simulatáneo, es necesario obtener el número serial correspondiente; proceso realizado a través del visualizador de Intel y expuesto en la figura 3.4.

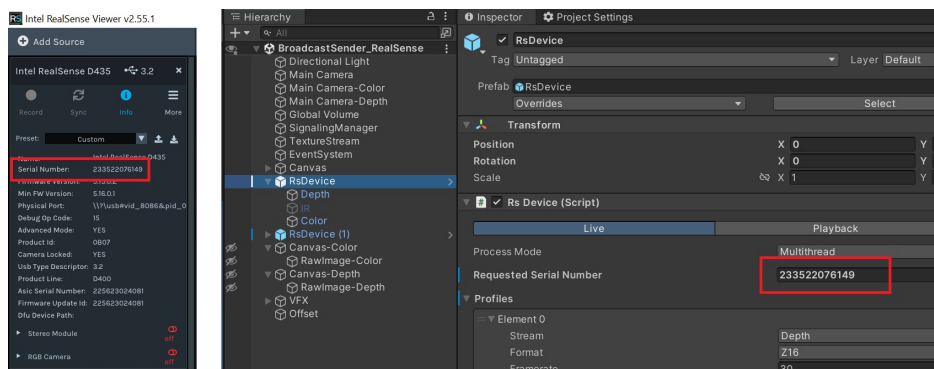


Figure 3.4: Intel RealSense View. En la parte izquierda se presentan las texturas de profundidad, utilizando un código de color conocido como "Jet"; donde el rojo representa lejanía y azul cercanía. Por otro lado, en su derecha se pueden encontrar la textura RGB correspondiente.



### 3 Memoria

#### Calibración y estimación de poses

Con la intención de reconstruir el ambiente y ubicar correctamente las cámaras en su respectiva posición dentro del ambiente virtual, se han explorado múltiples algoritmos; entre ellos destacan la utilización de AprilTags y la alineación de la nube de puntos con el punto más cercano iterativo (ICP, del inglés "Iterative Closest Point"). A continuación se exponen brevemente:

- Utilización de AprilTags

AprilTags es un sistema de etiquetas visuales desarrollado por investigadores de la Universidad de Michigan para proporcionar una localización de alta precisión y bajo consumo computacional. Durante esta exploración se logró implementar la calibración a través de la librería expuesta en la figura 3.5. Sin embargo, esta técnica requiere la utilización de un elemento externo a los dispositivos principales; dicho marcador debe ser visible por todas las cámaras involucradas en el sistema.

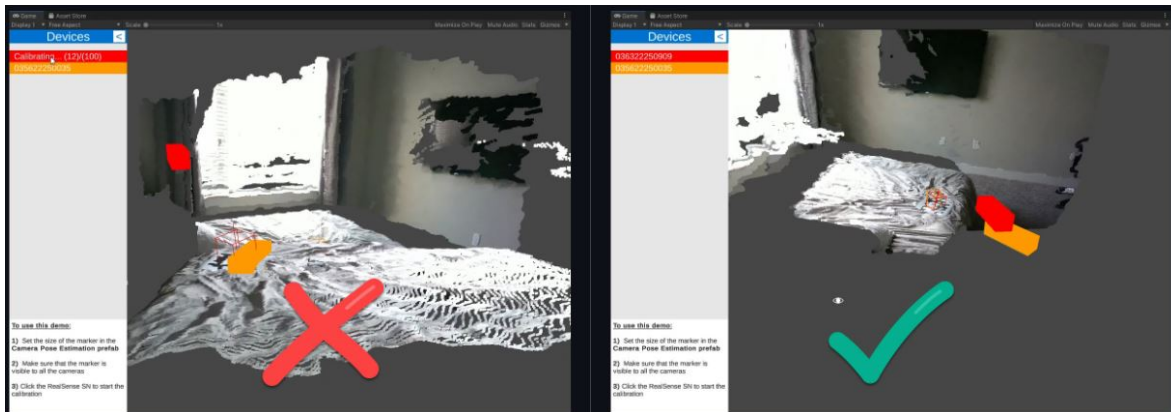


Figure 3.5: Ejemplo de calibración a partir del uso de April Tags. El paquete de desarrollo puede encontrarse en el siguiente repositorio: Multi-camera Calibration for RealSense in Unity

- Punto más cercano iterativo (ICP)

Finalmente, después de investigar en profundidad el funcionamiento de los algoritmos empleados en la reconstrucción de espacios, fue posible realizar una implementación del algoritmo ICP; el cual se plantea inicialmente con el objetivo de identificar puntos que coincidan entre una nube de puntos y un modelo 3D, como se expone en la figura 3.6. Así, también demostró ser eficaz para encontrar similitudes entre las nubes de puntos generadas por dos cámaras que apuntan al mismo ambiente desde perspectivas diferentes; permitiendo complementar la reconstrucción sin la necesidad de marcadores externos.

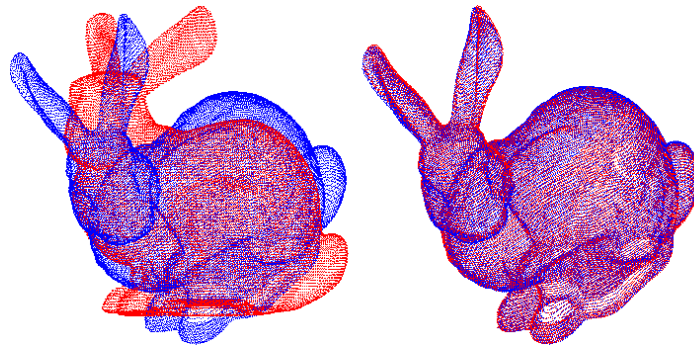


Figure 3.6: Calibración de una nube de puntos en un modelo 3D de un conejo. Más información se puede encontrar en la documentación del algoritmo y su implementación en Matlab [3] o la versión en OpenCV.

### 3 Memoria

#### Prototipo

Como resultado del análisis previo y la definición técnica, se presenta una solución compuesta por tres aplicaciones; dos de ellas hacen parte del desarrollo realizado en el entorno de renderizado en tiempo real: Unity3D. Por otro lado, la tercera aplicación es un "signaling-server" que se encuentra disponible como recurso entre los ejemplos del paquete Unity Render Streaming y disponible en el Github oficial.

Así, se ha logrado desplegar el servidor de señalización utilizando los servicios de hosting para WebApp de DigitalOcean. La conexión se evidencia en la figura 3.7

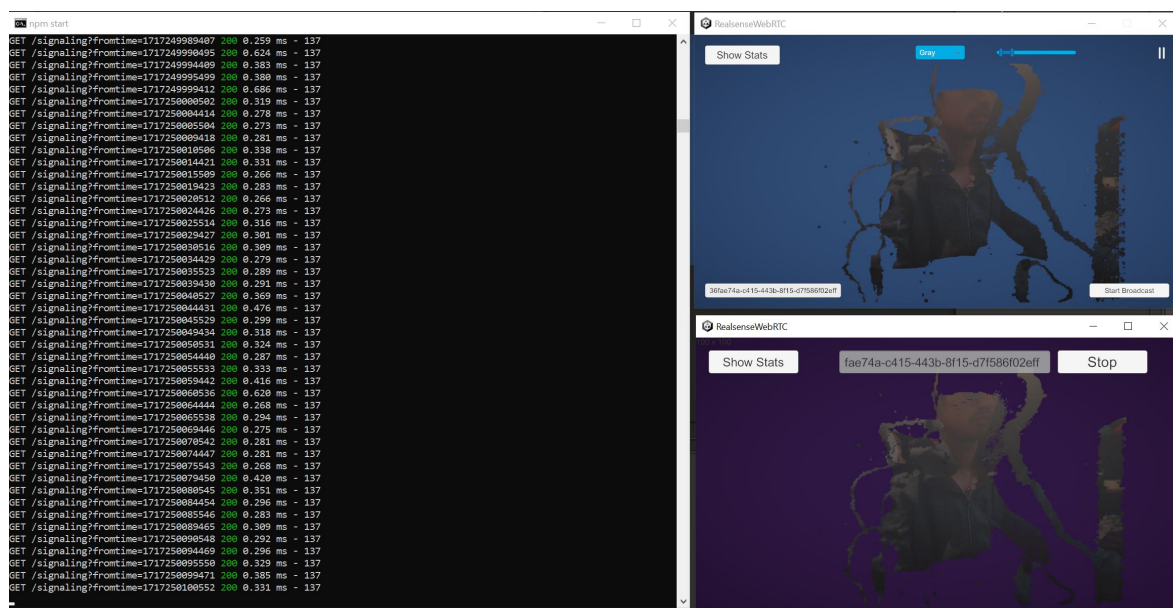


Figure 3.7: En la parte izquierda se presenta el servidor de señalización siendo ejecutado de forma local. En la parte superior derecha se muestra la aplicación que detecta la cámara de profundidad, separa las respectivas texturas y las envía a través de la conexión WebRTC. Así, la aplicación que está expuesta en la parte inferior derecha puede recibir las imágenes y transformarlas en partículas que permiten visualizar la teleconferencia en tres dimensiones.

Adicionalmente, los recursos respectivos pueden encontrarse en los siguientes repositorios:  
WebApp UnityApp

## 4 Conclusiones

### 4.1 Análisis general

Con la llegada del internet y la democratización de la tecnología, las videollamadas se han convertido en algo común e indispensable en la vida diaria de muchas personas. De hecho, gran mayoría de trabajos dependen estrictamente de videoconferencias y conexiones remotas.

Teniendo en cuenta lo anterior, sería válido mencionar que las soluciones tecnológicas actuales cumplen con las tareas requeridas para una comunicación efectiva; sin embargo, contiene limitaciones inherentes al mundo virtual bidimensional. De este modo, la implementación de sistemas para teleconferencias en tres dimensiones podría abrir un abanico de posibilidades en términos de visualización, análisis de la información, e interacción; como mínimo podría estar un paso más cerca a la realidad, sin la necesidad de agregar una capa de abstracción puramente generada o simulada en el ambiente virtual (como es el caso de los avatares en 3D). Recientes investigaciones han logrado avances considerables con relación al sentido de presencia en experiencias remotas, principalmente aplicando técnicas de reconstrucción facial, logrando la generación de modelos 3D fotorrealistas. A pesar de su eficacia, estas soluciones suelen requerir equipos y recursos computacionales costosos, además de exponer el contenido como una simulación de la realidad. Por otro lado, el enfoque sugerido a lo largo de este trabajo, presenta teleconferencias inmersivas y realistas sin la necesidad de modelos 3D o avatares; evitando dicha capa de abstracción.

## 4 Conclusiones

Así, la solución planteada permitiría teleconferencias con un grado de fidelidad superior, lo cual contemplaría la siguiente evolución: surgiendo desde las llamadas telefónicas donde solamente se transmitía señales de audio, continuando con las videollamadas que incluyen tanto audio como imágenes bidimensionales, y finalmente el grado que se sugiere es agregar tanto audio como vídeo; pero este último incluiría imágenes de profundidad, permitiendo la transmisión en tiempo real de escenarios completos y tridimensionales.

La solución planteada se limita a la capacidad de los sensores en términos de resolución y profundidad, sin embargo, estas limitaciones pueden abordarse al aplicar capas de procesamiento de imagen con algoritmos de inteligencia artificial que ya han sido investigados en múltiples ocasiones y actualmente están presentes en muchos dispositivos móviles. De ese modo, se ha logrado plantear un sistema apto para teleconferencias inmersivas en campos de las artes visuales, aprovechando tecnologías como WebRTC, sensores de profundidad y motores de renderizado en tiempo real.

### 4.2 Casos de uso (Objetivos de Desarrollo Sostenible)

Al incrementar la fidelidad con la cual se reproducen los entornos, elevando el sentido de presencia, y permitiendo el acceso de forma remota, se plantean los siguientes casos de uso aplicables a los Objetivos de Desarrollo Sostenible (ODS):

- **Entretenimiento:** Una gran cantidad de eventos son transmitidos en tiempo real a través de las redes sociales y otras plataformas de contenido, incluyendo aquellas plataformas emergentes como los servicios de streaming para cascos de Realidad Virtual y Realidad Aumentada. Por eso, la solución podría brindarse para transmitir ambientes y situaciones con niveles de interacción superior debido al contexto tridimensional que plantea. Conciertos, eventos deportivos, shows de televisión, y series inmersivas son algunos de los casos de uso que podrían utilizar el sistema. De cierto modo, la democratización de los eventos y el entretenimiento tiene un impacto clave en la reducción de emisiones a causa del transporte y los vuelos a larga distancia.

- Educación: No es una novedad que la virtualidad se ha hecho camino en los últimos años, haciendo que muchos estudiantes prefieran permanecer en casa; sin embargo, muchos ejercicios requieren contemplar escenas realistas y tridimensionales para poder compartir el conocimiento. La transmisión de aulas o laboratorios podría añadir una capa de visualización requerida para lograr niveles de inmersión superiores.
- Salud y bienestar: Durante y después de la contingencia del COVID-19, servicios médicos comenzaron a realizar diagnósticos virtuales a través de videollamadas. Así, es normal pensar que teleconferencias en tres dimensiones podría brindar puntos de vista más realistas, llegando a diagnósticos más precisos; igualmente aplicado a seguimiento de pacientes y rutinas de fisioterapia.

### 4.3 Pasos a seguir

Se propone la ejecución de pruebas que integren una mayor cantidad de cámaras, llevando al límite las capacidades de transmisión en tiempo real; lo que permitiría identificar las posibilidades de las conexiones actuales y dar pie para las potenciales redes 5G.

Adicionalmente, se sugiere mejorar la implementación de la capa de visualización, adelantándose en las capacidades de renderizado que ofrece el entorno de desarrollo o incluso haciendo transmisiones multiplataforma. Referencias como los efectos visuales desarrollados por Keijiro, podrían ser un buen inicio (ver figura 4.1).

## 4 Conclusiones

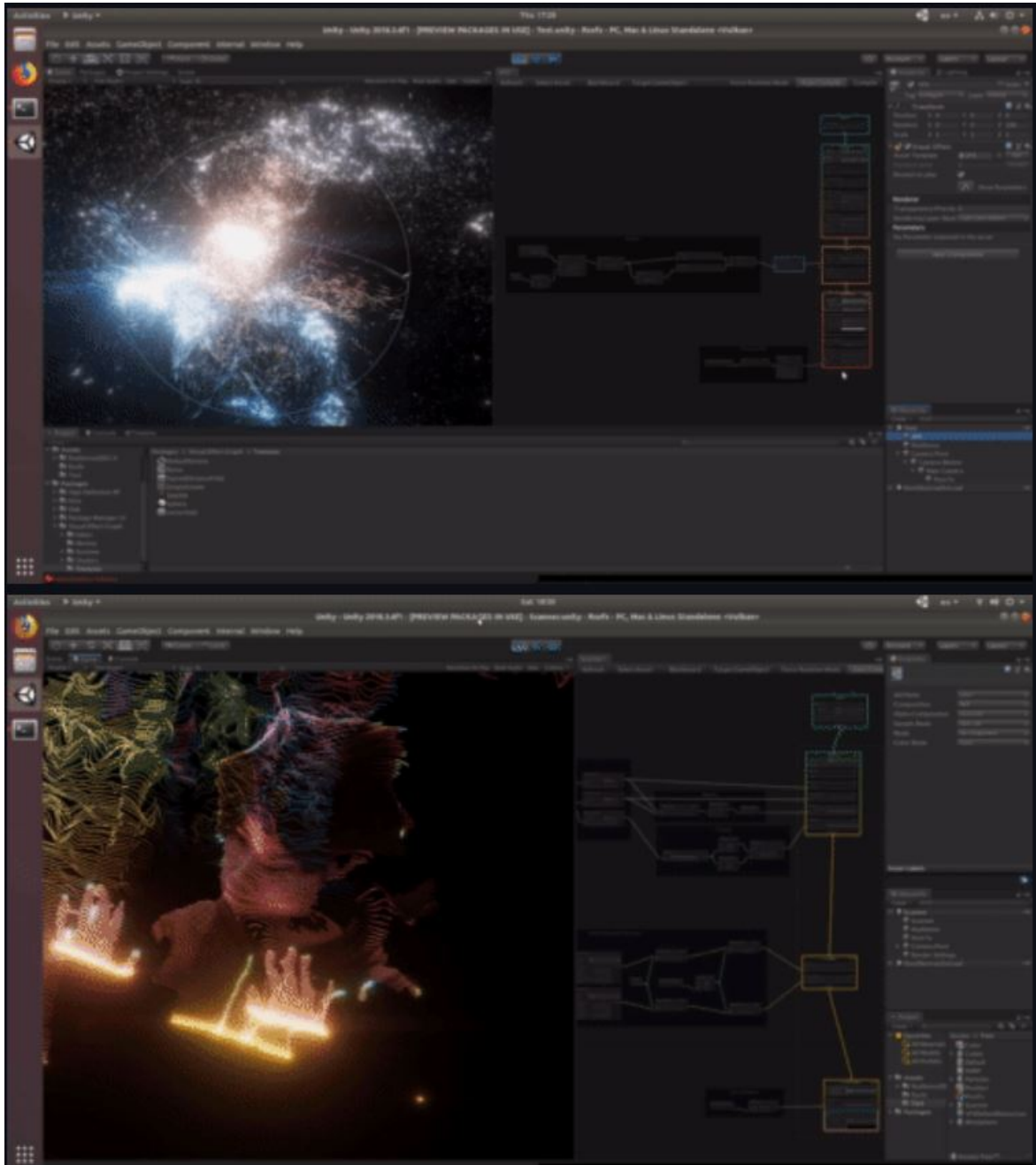


Figure 4.1: Algunos ejemplos expuestos en el repositorio público del artista y desarrollador Keijiro. Algunos más pueden verse en el canal oficial de Unity-Japan

## Bibliography

- [1] Various Authors, ed. *PRESENCE 2000: 3rd Annual International Workshop on Presence*. Delft, The Netherlands: ITC, BT, and SONY, 2000. ISBN: 90-386-1571-X.
- [2] Various Authors, ed. *SIGGRAPH 2023: 50th Annual Conference on Computer Graphics and Interactive Techniques*. Los Angeles, CA, USA: ACM, 2023. ISBN: 978-1-4503-XXXX-X.
- [3] Tolga Birdal. *ICP Registration using Efficient Variants and Multi-Resolution Scheme*. <https://www.mathworks.com/matlabcentral/fileexchange/47152-icp-registration-using-efficient-variants-and-multi-resolution-scheme>. MATLAB Central File Exchange. Retrieved September 2, 2024. 2024.
- [4] Agnieszka Chodorek, Robert R. Chodorek, and Krzysztof Wajda. “Improving the Efficiency of WebRTC Layered Simulcast Using Software Defined Networking”. In: *Broadband Communications, Networks, and Systems*. Ed. by Wei Wang and Jun Wu. Cham: Springer Nature Switzerland, 2023, pp. 13–28. ISBN: 978-3-031-40467-2.
- [5] Rushali Deshmukh et al. “Video Conferencing using WebRTC”. In: *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*. 2023, pp. 857–864. DOI: 10.1109/ICSCDS56580.2023.10104791.
- [6] Ian Gonsler et al. “Integrating Interfaces into Furniture: New Paradigms for Ubiquitous Computing, Mixed Reality, and Telepresence within the Built Environment”. In: *Proceedings of the 6th Media Architecture Biennale Conference*. MAB ’23. <conf-loc>, <city>Toronto</city>, <state>ON</state>, <country>Canada</country>, </conf-loc>: Association for Computing Machinery, 2024, pp. 167–172. ISBN: 9798400716355. DOI: 10.1145/3627611.3627629. URL: <https://doi.org/10.1145/3627611.3627629>.



## Bibliography

- [7] Fanghao Huang et al. “Telepresence augmentation for visual and haptic guided immersive teleoperation of industrial manipulator”. In: *ISA Transactions* (2024). ISSN: 0019-0578. DOI: <https://doi.org/10.1016/j.isatra.2024.05.003>. URL: <https://www.sciencedirect.com/science/article/pii/S0019057824002064>.
- [8] Xincheng Huang, James Riddell, and Robert Xiao. “Virtual Reality Telepresence: 360-Degree Video Streaming with Edge-Compute Assisted Static Foveated Compression”. In: *IEEE Transactions on Visualization and Computer Graphics* 29.11 (2023), pp. 4525–4534. DOI: 10.1109/TVCG.2023.3320255.
- [9] Myungjin Kim et al. “Visual Information Transmission Methods for Virtual/Augmented-Reality-Based Telepresence”. In: *Korea Digital Contents Society* 24.10 (2023), p. 2509. DOI: 10.9728/dcs.2023.24.10.2509.
- [10] Jason Lawrence et al. “State of the art in telepresence”. In: *ACM SIGGRAPH 2022 Courses*. SIGGRAPH ’22. Vancouver, British Columbia, Canada: Association for Computing Machinery, 2022. ISBN: 9781450393621. DOI: 10.1145/3532720.3535673. URL: <https://doi.org/10.1145/3532720.3535673>.
- [11] Jason Lawrence et al. “State of the Art in Telepresence (Part 1)”. In: *ACM SIGGRAPH 2023 Courses*. SIGGRAPH ’23. , Los Angeles, California, Association for Computing Machinery, 2023. ISBN: 9798400701450. DOI: 10.1145/3587423.3595484. URL: <https://doi.org/10.1145/3587423.3595484>.
- [12] RAY LC, Sijia Liu, and Qiaosheng Lyu. “IN/ACTive: A Distance-Technology-Mediated Stage for Performer-Audience Telepresence and Environmental Control”. In: *Proceedings of the 31st ACM International Conference on Multimedia*. MM ’23. <conf-loc>, <city>Ottawa ON</city>, <country>Canada</country>, </conf-loc>: Association for Computing Machinery, 2023, pp. 6989–6997. ISBN: 9798400701085. DOI: 10.1145/3581783.3613791. URL: <https://doi.org/10.1145/3581783.3613791>.
- [13] M. Lombard et al. *Immersed in Media: Telepresence Theory, Measurement Technology*. Jan. 2015, pp. 1–332. ISBN: 978-3-319-10189-7. DOI: 10.1007/978-3-319-10190-3.
- [14] Matthew Lombard and Matthew T. Jones. “Defining Presence”. In: *Immersed in Media: Telepresence Theory, Measurement & Technology*. Ed. by Matthew Lombard et al.

- Cham: Springer International Publishing, 2015, pp. 13–34. ISBN: 978-3-319-10190-3. DOI: 10.1007/978-3-319-10190-3\_2. URL: [https://doi.org/10.1007/978-3-319-10190-3\\_2](https://doi.org/10.1007/978-3-319-10190-3_2).
- [15] *MAB '23: Proceedings of the 6th Media Architecture Biennale Conference*. New York, NY, USA: Association for Computing Machinery, 2023. ISBN: 9798400716355.
- [16] Andrew Maimone and Henry Fuchs. “Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras”. In: *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. 2011, pp. 137–146. DOI: 10.1109/ISMAR.2011.6092379.
- [17] Gisela Reyes-Cruz et al. “Augmented Robotic Telepresence (ART): A Prototype for Enhancing Remote Interaction and Participation”. In: *Proceedings of the First International Symposium on Trustworthy Autonomous Systems*. TAS '23. New York, NY, USA: Association for Computing Machinery, 2023. ISBN: 9798400707346. DOI: 10.1145/3597512.3597532. URL: <https://doi.org/10.1145/3597512.3597532>.
- [18] Xiaojun Shen and Shervin Shirmohammadi. “Telepresence”. In: *Encyclopedia of Multimedia*. Ed. by Borko Furht. Boston, MA: Springer US, 2006, pp. 843–848. ISBN: 978-0-387-30038-2. DOI: 10.1007/0-387-30038-4\_233. URL: [https://doi.org/10.1007/0-387-30038-4\\_233](https://doi.org/10.1007/0-387-30038-4_233).
- [19] Ingvar Tjostheim, Wolfgang Leister, and J. A. Waterworth. “Telepresence and the Role of the Senses”. In: *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016*. Ed. by Don Berkich and Matteo Vincenzo d’Alfonso. Cham: Springer International Publishing, 2019, pp. 169–187. ISBN: 978-3-030-01800-9. DOI: 10.1007/978-3-030-01800-9\_9. URL: [https://doi.org/10.1007/978-3-030-01800-9\\_9](https://doi.org/10.1007/978-3-030-01800-9_9).
- [20] Xuanyu Wang, Weizhan Zhang, and Hongbo Fu. “A3RT: Attention-Aware AR Teleconferencing with Life-Size 2.5D Video Avatars”. In: *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 2024, pp. 211–221. DOI: 10.1109/VR58804.2024.00044.

## *Bibliography*

- [21] Xuanyu Wang et al. “Real-and-Present: Investigating the Use of Life-Size 2D Video Avatars in HMD-Based AR Teleconferencing”. In: *arXiv preprint arXiv:2401.02171* (2024). arXiv: 2401.02171 [cs.HC].