



UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Departamento de Electrónica e Computación

Tesis doctoral

**VISUALIZACIÓN AVANZADA E INTERACCIÓN APLICADA A
ESCENARIOS VIRTUALES**

Presentada por:

Roi Méndez Fernández

Dirigida por:

Julián Carlos Flores González

Enrique Castelló Mayo

Septiembre de 2016

Julián Carlos Flores González, Profesor Titular del Área de Lenguajes y Sistemas
Informáticos

Enrique Castelló Mayo, Profesor Titular del Área de Comunicación Audiovisual y
Publicidad

HACEN CONSTAR:

Que la memoria titulada **VISUALIZACIÓN AVANZADA E INTERACCIÓN APLICADA A ESCENARIOS VIRTUALES** ha sido realizada por **Roi Méndez Fernández** bajo nuestra dirección en el Centro Singular de Investigación en Tecnoloxías da Información de la Universidad de Santiago de Compostela, y constituye la Tesis que presenta para optar al título de Doctor.

Septiembre de 2016

Julián Carlos Flores González
Director tesis

Enrique Castelló Mayo
Director tesis

¿A quién va usted a creer, a mí o a sus propios ojos?

Groucho Marx

How do you define real? If you're talking about what you can feel, what you can smell, what you can taste and see, then real is simply electrical signals interpreted by your brain

Morpheus (The Matrix)

Agradecimientos

En primer lugar, quisiera agradecer a los profesores Julián Flores y Enrique Castelló la confianza que han depositado en mí para llevar a cabo este trabajo de investigación, el respeto y apoyo que mostraron siempre a mis decisiones y propuestas y su gran contribución al mismo.

También quisiera agradecer el apoyo y compañerismo de todas las personas que pasaron a lo largo de estos años, primero, por el grupo M.A.R. (Multimedia Animación y Realidad Virtual) y, posteriormente, por el grupo COGRADE (Gráficos por Computador e Ingeniería de Datos). Con ellos he aprendido las bases sobre las que se asienta un buen grupo de investigación.

Este trabajo no habría sido posible sin la ayuda, el esfuerzo y las horas de dedicación del equipo técnico de la Facultad de Ciencias de la Comunicación: Marisa, Reyes y muy especialmente Manuel Fidalgo.

Nunca daré las gracias lo suficiente a mi familia. A Castor y a Dulce por estar siempre ahí echándome una mano en todo y aconsejándome a la hora de tomar decisiones importantes. A Chelo por dar siempre el doscientos por cien en todo y a Tamara porque, a pesar de la distancia, se que siempre te tendré para ayudarme en lo que necesite.

A lo largo de estos años en el CITIUS he conocido a mucha gente que se han convertido en una parte imprescindible de mi vida. Gracias a Adrián, a Fer, a Vanesa, a Borja, a Diego y a todos los "pringadillos" por vuestra amistad. Gracias también a Gabriel, Moldes, Aida, Adrián, Castor, Luis, Pazos, Bea, Rebeca, Jose Luis, Borja y a todos mis amigos por ser esa isla en la que siempre puedo refugiarme y desconectar.

Finalmente quiero darle las gracias, de un modo especial, a Carla. Por acompañarme y aguantarme siempre, en los buenos y en los malos momentos. Por hacerme reír tanto. Por estar ahí cada día. Gracias.

Septiembre de 2016

Índice general

1	Introducción, objetivos y estructura de la tesis	1
1.1.	Introducción	1
1.2.	Trabajo de investigación: objetivos	6
1.2.1.	Objetivos principales	6
1.2.2.	Objetivos secundarios	9
1.3.	Estructura y organización de la tesis	10
2	Platós Virtuales de televisión	13
2.1.	Historia de los Platós Virtuales de Televisión	13
2.2.	Arquitectura Clásica de un Plató Virtual de Televisión	19
2.2.1.	Subsistema de <i>render</i>	22
2.2.1.1.	Ordenador de <i>render</i>	22
2.2.2.	Subsistema de sensorización y vídeo	23
2.2.2.1.	Cámaras	23
2.2.2.2.	Sensores	23
2.2.3.	Subsistema de composición	31
2.2.3.1.	Escenario y ciclorama	31
2.2.3.2.	<i>chroma keyer</i>	32
2.2.4.	Otros elementos a considerar	34
2.2.4.1.	Referencias para los actores	34
2.2.4.2.	Mezclador de Vídeo	36
2.2.4.3.	Mesa de sonido	37
2.3.	Análisis de casos de uso reales: noches electorales	37
2.3.1.	Elecciones en Reino Unido: 2014, 2015 y 2016	38

2.3.2.	Elecciones en España: autonómicas 2015, Generales 2015 y 2016 . . .	39
2.3.2.1.	Radio Televisión Canaria	40
2.3.2.2.	Antena 3	41
2.3.2.3.	La 1	42
2.3.2.4.	TV3	43
2.3.3.	Conclusiones	43
3	Retos de los platós virtuales de televisión y marco de trabajo	47
3.1.	Retos de los platós virtuales de televisión	47
3.2.	Marco de trabajo	49
3.2.1.	Introducción	49
3.2.2.	Plató virtual de trabajo	50
3.2.3.	Brainstorm Multimedia	53
3.2.3.1.	Software eStudio	55
4	Interacción en platós virtuales de televisión	57
4.1.	Diseño de una nueva arquitectura distribuida escalable para platós virtuales de televisión	58
4.2.	Implementación de una nueva arquitectura distribuida escalable para platós virtuales de televisión	61
4.2.1.	Implementación del <i>middleware</i>	61
4.2.2.	<i>Tracking</i> de Cámaras	71
4.2.2.1.	Implementación	71
4.2.2.2.	Pruebas	73
4.2.3.	<i>Tracking</i> de Objetos	75
4.2.3.1.	Implementación	75
4.2.3.2.	Pruebas	76
4.2.4.	<i>Tracking</i> de actores	77
4.2.4.1.	<i>Distance keying</i>	77
4.2.4.2.	Interacción directa	82
4.2.4.3.	Detección de gestos	86
4.2.5.	Actualización del sistema	91
4.2.5.1.	Implementación	92
4.2.5.2.	Pruebas	94

4.3.	Evaluación preliminar del sensor Microsoft Kinect V2 para su uso en platós virtuales de televisión con interacción natural	97
4.3.1.	Resultados	98
4.4.	Pruebas Integrales del sistema	101
4.4.1.	Rendimiento	101
4.4.2.	Caso de prueba I: “Copicat”	102
4.4.3.	Caso de prueba II: “La universidad responde”	103
4.5.	Conclusiones	105
5	Visualización avanzada en platós virtuales de televisión	107
5.1.	Iluminación en platós virtuales de televisión	108
5.1.1.	Análisis de iluminación de cicloramas	109
5.1.1.1.	Implementación	109
5.1.1.2.	Pruebas	114
5.1.2.	Iluminación de entornos virtuales	116
5.1.2.1.	Implementación	117
5.1.2.2.	Pruebas	123
5.1.3.	Sombras del mundo real en el mundo virtual	124
5.2.	Conclusiones	127
6	Visualización avanzada e interacción aplicada a entornos museísticos	129
6.1.	El Pórtico de la Gloria Virtual, aplicación interactiva en tiempo real para la exploración virtual del Pórtico de la Gloria	130
6.1.1.	Introducción	130
6.1.2.	Pórtico Virtual	131
6.1.3.	Desarrollo del proyecto	133
6.1.3.1.	Creación de contenido	133
6.1.3.2.	Planteamiento conceptual de la aplicación	134
6.1.4.	Arquitectura del sistema	136
6.1.4.1.	Sistema de visualización	136
6.1.4.2.	Sistema de control	137
6.1.5.	Conclusiones	138
6.2.	Los instrumentos virtuales del Pórtico de la Gloria	140
6.2.1.	Introducción	140

6.2.2.	Descripción del sistema	141
6.2.3.	Comportamiento del sistema	143
6.2.4.	Arquitectura software	144
6.2.5.	Exposición itinerante	145
6.2.6.	Conclusiones	146
6.3.	Conoce el Pórtico	148
6.3.1.	El Pórtico Virtual	148
6.3.2.	Restauración	150
6.3.3.	Galería de imágenes	150
6.3.4.	Exposición	151
6.3.5.	Conclusiones	151
6.4.	Módulo de visualización del interior del busto humano	152
6.4.1.	Introducción	152
6.4.2.	Objetivos	153
6.4.3.	Diseño del contenido	153
6.4.4.	Tecnologías empleadas y diseño hardware	154
6.4.4.1.	Sistema de Visualización	154
6.4.4.2.	Sistema de Interacción	156
6.4.5.	Implementación	157
6.4.5.1.	Prototipo 1	158
6.4.5.2.	Prototipo 2	159
6.4.5.3.	Prototipo 3	161
6.4.6.	Pruebas	162
6.4.7.	Conclusiones y trabajo futuro	164
7	Discusión de resultados, conclusiones, y trabajo futuro	167
7.1.	Introducción	167
7.2.	Discusión de resultados	167
7.2.1.	Diseño e implementación de una nueva arquitectura escalable y distribuida para platós virtuales de televisión	168
7.2.2.	Interacción en platós virtuales de televisión	168
7.2.3.	Visualización avanzada en platós virtuales de televisión	169
7.2.4.	Visualización avanzada e interacción en entornos museísticos	170
7.3.	Conclusiones	172

<i>Índice general</i>	XIII
7.4. Trabajo futuro	173
7.5. Publicaciones	175
7.5.1. Artículos publicados en Revista	175
7.5.2. Capítulos en libros	175
7.5.3. Artículos publicados en actas de congreso	176
Bibliografía	179
Índice de figuras	199
Apéndice	205

CAPÍTULO 1

INTRODUCCIÓN, OBJETIVOS Y ESTRUCTURA DE LA TESIS

1.1. Introducción

Desde la creación de los primeros ordenadores y, con mayor intensidad desde su popularización de cara al gran público, la forma de comunicación entre usuario y máquina ha sido un campo de investigación de capital importancia en el éxito o fracaso final de los productos informáticos. Dentro de esta comunicación el usuario transmite órdenes al ordenador a través de una serie de periféricos de entrada (ratón, teclado, micrófono, etc.) y este le devuelve información a través de los dispositivos de salida (pantallas, altavoces, actuadores, etc.). Parte del éxito de una aplicación viene dado por el modo en que se realiza esta entrada-salida de información, lo que se denomina “metáfora de interacción”. Una de las más populares es la utilización de un ratón (desarrollado en el Standford Research Laboratory en 1965 [45]) para interactuar con un entorno de ventanas (diseñado en 1968 en el NLS [44]). Otro tipo de metáforas, por ejemplo, son los gestos comunes en múltiples dispositivos con pantalla táctil, como pasar a la siguiente imagen en una galería fotográfica deslizando el dedo sobre la pantalla (un gesto denominado *swipe*) simulando arrastrar la fotografía hacia un lado.

El avance tecnológico y la aparición de nuevos dispositivos de entrada y salida, así como el aumento de la capacidad de cálculo de los ordenadores, ha permitido avanzar hacia una interacción más natural en la que la metáfora utilizada resulte intuitiva para usuarios no expertos. Tradicionalmente los usuarios se relacionan con los ordenadores a través de dispositivos

limitados y de una forma distinta a como lo hacen con los elementos del mundo real, lo que provoca que sea necesario un proceso de aprendizaje (a veces dilatado) para lograr un correcto manejo de los sistemas. En situaciones en las que el tiempo disponible para interiorizar el manejo de un dispositivo es limitado (un módulo en un museo, un punto de información, un cajero automático, etc.) cobra una especial importancia el aumento de la velocidad de aprendizaje y, por tanto, la sencillez y forma intuitiva de uso del mismo. Cuando los elementos con los que se interactúa simulan el mundo real, por ejemplo, haciendo uso de modelos 3D, la falta de la tercera dimensión tanto en los dispositivos de salida (las pantallas) como en los de entrada (ratones) hace más complejo su manejo de una forma natural. A esto se suma que la relación humana con el mundo es inherentemente multimodal. Las personas utilizan distintos sentidos de forma secuencial o en paralelo para explorar y manipular el entorno de una forma activa o pasiva [180], mientras que los dispositivos más comunes hoy en día únicamente hacen uso de la vista (y, en algunos casos, del oído). Actualmente, el máximo exponente de este tipo de interacción multimodal son los sistemas de realidad virtual y aumentada, ya que una de sus principales finalidades es integrarse con el mundo real (en el caso de la realidad aumentada) o simularlo (el caso de la realidad virtual). Además del tipo interacción, en este tipo de entornos tiene una importancia capital la calidad y coherencia de las imágenes sintéticas utilizadas. En la realidad virtual se ha de simular un mundo completamente generado por ordenador en el que el usuario se sienta inmerso¹ mientras que en la realidad aumentada el contenido CGI (*computer generated imagery*) debe de ser coherente con el entorno real para que su integración en el mismo sea correcta. Para la obtención de estas imágenes sintéticas de alta calidad y realismo a partir de modelos 3D se utilizan una serie de técnicas englobadas dentro de lo que se conoce como “visualización avanzada”.

En el año 1994 Milgram et al. [136] definían el espacio entre realidad y virtualidad como un continuo en el que se engloban las diferentes técnicas de la realidad mixta (figura 1.1). Así se establecen entre el mundo virtual y el mundo real dos clasificaciones en función de la importancia de ambos en la composición final: la realidad aumentada (cuando prevalece el mundo real sobre el virtual) y la virtualidad aumentada (cuando sobre el mundo virtual se superponen elementos reales). Tanto la realidad virtual como la aumentada están actualmente en fuerte expansión debido a la ubicuidad de los dispositivos móviles y el aumento de su potencia de cálculo así como al continuo desarrollo y mejora de los periféricos de entrada y

¹Se define inmersión en un medio como la sensación de alejamiento perceptivo del entorno real y de acercamiento a un entorno no real[8]

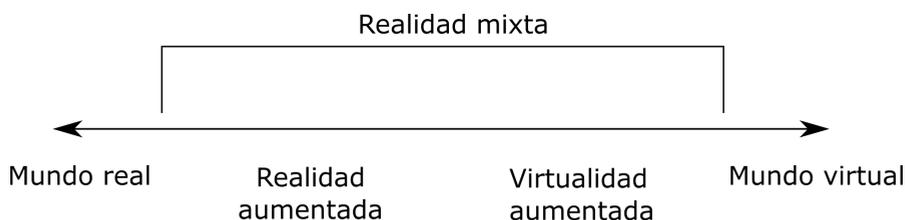


Figura 1.1: Continuo realidad-virtualidad de Milgram

salida. En este marco, ambas tecnologías están llegando a un número mayor de usuarios y conquistando nuevos mercados como el de la publicidad, la formación, etc.

En esta tesis se aborda la utilización de técnicas de visualización avanzada e interacción persona-ordenador en dos ámbitos de aplicación de las tecnologías asociadas a los mundos de la realidad virtual y aumentada: los platós virtuales de televisión y los museos interactivos.

En los últimos años la forma en que se presenta la información en televisión ha sufrido una gran transformación debido a la revolución tecnológica que ha supuesto la llegada de la era digital a los medios audiovisuales. El proceso de digitalización de la producción y la emisión televisiva ha consolidado un entorno audiovisual íntegramente digital [35]. La llegada de la televisión digital supuso una enorme evolución ya que no solo ha aportado imágenes de una mayor calidad, sino que también ha abierto las puertas a servicios como la recepción móvil de televisión, la interactividad, la televisión a la carta o los servicios multimedia [9]. La interactividad del contenido emitido a través de la televisión está empezando, también, a abrirse un hueco a través de plataformas como HBBTV [116]. Estos cambios han modificado por completo los procesos de producción y creación de productos audiovisuales [25]. La gestión informática de los procesos de producción se ha generalizado [24] y la introducción de elementos visuales generados por ordenador es común en la mayor parte de las parrillas televisivas del mundo.

En este contexto de modernización nacen, a finales de los años ochenta y principios de los noventa, los platós virtuales de televisión (PVT) como una aplicación de la realidad virtual y aumentada al campo televisivo. En ellos se combinan, en tiempo real, elementos reales grabados en plató con entornos 3D generados por ordenador. La integración entre ambos mundos se realiza utilizando la técnica del *chroma-key* [88] que utiliza variaciones de color para sustituir partes de la escena real por contenido virtual. La principal diferencia entre estos platós y un *chroma* al uso estriba en la utilización de modelos tridimensionales como entorno

sintético, frente a las tradicionales imágenes 2D. Esto permite obtener diferentes perspectivas del escenario realizando todo tipo de movimientos de cámara en tiempo real, así como situar objetos sintéticos delante o detrás de los reales. Para que esto sea posible se ha de conocer la posición y orientación de las cámaras, por lo que estas deben de estar sensorizadas. Asimismo, para que los actores tengan la capacidad de interactuar con objetos virtuales, sus acciones y movimientos deben de ser reconocidos por el sistema. El uso de esta tecnología introduce novedades también en las estructuras representativas del discurso televisivo. La puesta en escena se desmaterializa, abandonando su referente real en favor de la imagen de síntesis [34]. Desde un punto de vista técnico científico este tipo de entornos se encuentran a medio camino entre el mundo virtual y el mundo real en el continuo realidad-virtualidad, siendo un ejemplo claro de realidad mixta.

Las principales ventajas que presentan los platós virtuales de televisión son:

- **Creativas:** no ponen límites físicos a la creatividad a la hora de diseñar los entornos virtuales o los efectos utilizados durante una grabación. Las posibilidades creativas que ofrecen son prácticamente infinitas.
- **Técnicas:** la posibilidad de generar escenarios imposibles en el mundo real, rompiendo las leyes físicas así como la utilización de un solo estudio real para simular infinitos platós virtuales.
- **Económicas:** no es necesaria la construcción física de complejas escenografías, basta con modelarlas en 3D. Esto reduce enormemente los costes tanto de materiales como de personal. Al no existir físicamente los objetos no es necesario su almacenamiento, transporte o conservación, lo que redundará en un menor coste económico.

Estas tres ventajas tienen como consecuencia la incorporación de cambios en las diferentes fases de los procesos de preproducción, producción y postproducción de un programa de televisión [32]. Asimismo, al trabajar con escenografía virtual, la parte técnica adquiere un mayor peso, ya que mientras que en escenografía real se trabaja sólo con la señal de cámara, en escenografía virtual esa señal de cámara hay que incrustarla en un entorno virtual generado por ordenador utilizando información captada a través de una serie de sensores.

Como consecuencia de esta mayor complejidad técnica, en 1998 Andrew Wojdala describió los principales retos a los que se enfrentaban los estudios virtuales de televisión de aquella época [192]. Estos iban desde las limitaciones técnicas de los dispositivos o la iluminación

coherente del escenario virtual hasta la precisión de los sensores o la interacción de los actores con el entorno virtual. A día de hoy, la mayor parte de estas cuestiones no han sido todavía completamente resueltas debido, entre otros motivos, a que, a pesar del gran desarrollo tecnológico que han sufrido en su corta historia, la arquitectura hardware y software diseñada para los primeros prototipos en los noventa ha permanecido invariable, haciendo uso de una única sensorización para el seguimiento de las cámaras y dejando de lado la interacción del presentador con el entorno sintético. La aparición de nuevos dispositivos y técnicas que permiten dotar de mayores capacidades a los platós virtuales de televisión hacen necesaria una actualización de la arquitectura tradicional de cara a dotarla de una mayor flexibilidad, simplificando la inclusión de nuevos elementos en los flujos de trabajo de una grabación en directo que permitan incrementar el nivel de interactividad así como la automatización de ciertos procesos de cara a una integración más efectiva entre los mundos sintético y real. Asimismo debe de estudiarse el uso de técnicas de visualización avanzada que permitan aumentar la sensación de presencia² de los actores en el escenario virtual.

Además de en el televisivo, el desarrollo tecnológico ha facilitado cambios muy significativos en los modos de presentación de la información en distintos formatos. Dichos avances tienen su principal reflejo en el contexto digital cuyas características de multimedialidad, hipertextualidad e interactividad están posibilitando que los tradicionales géneros y formatos divulgativos se redefinan con el objetivo de aprovechar estas características [48]. Un ejemplo paradigmático es el género documental, que está adoptando un nuevo formato multimedia e interactivo conocido como “documental web”. Otros entornos, como los museísticos, no son ajenos a estos avances y se encuentran en una continua evolución de la forma de exponer su contenido hacia un paradigma de mayor implicación y participación por parte del visitante. Dentro de este marco, la utilización de algoritmos de visualización que permitan acceder al contenido de una forma más atractiva resulta clave. Asimismo, las metáforas de interacción utilizadas tendrán un papel crucial a la hora de evitar la frustración de los usuarios no expertos al entrar en contacto por primera vez con un módulo expositivo. La capacidad del visitante para interactuar de una forma natural e intuitiva con las piezas expuestas resultará de suma

²“Presencia” es un término muy utilizado en aplicaciones de realidad virtual. En este sentido se define como el grado en que los participantes sienten, de manera subjetiva, que se encuentran en una ubicación diferente a su ubicación física real debido a los efectos creados en sus sentidos por una simulación generada por computador [165]. De una manera más simple, podría definirse como “la sensación de encontrarse allí” [74]. En el entorno de los platós virtuales de televisión, esta definición podría llevarse a la tercera persona: “Presencia” es el grado en que los espectadores, sienten, de manera subjetiva, que ciertos elementos, como actores y atrezzo, se encuentran en una ubicación diferente a su ubicación física real (en este caso el plató donde se está grabando el contenido) debido a los efectos creados por una simulación generada por computador.

importancia para captar y mantener su atención de cara a una transmisión más efectiva del conocimiento. Es necesario, por tanto, analizar distintas técnicas de interacción y visualización avanzada en entornos museísticos que permitan hacer más atractiva para el gran público la transmisión de conocimiento y la puesta en valor del patrimonio pretendida por los diferentes espacios expositivos.

1.2. Trabajo de investigación: objetivos

A lo largo del presente texto se expondrá el trabajo de investigación desarrollado por el doctorando Roi Méndez, durante el período 2009-2016, e integrado dentro de la actividad investigadora del grupo COGRADE (Gráficos por Computador e Ingeniería de Datos) del CI-TIUS (Centro Singular de Investigación en Tecnologías de la Información) de la Universidad de Santiago de Compostela, bajo la dirección de los Doctores Julián Flores y Enrique Castelló. Esta labor de investigación se encuentra a caballo entre dos mundos: el tecnológico y el de la comunicación audiovisual. La parte tecnológica abarca desde la rama de la Computación Gráfica (CG), centrándose específicamente en las técnicas de Visualización Avanzada (VA), Realidad Virtual (RV) y Realidad Aumentada (RA) hasta la Interacción Persona-Ordenador (IPO). Por otro lado, dentro del mundo de la comunicación audiovisual, se ha trabajado en torno a la gestualidad natural de los individuos (comunicación no verbal), la optimización de la transmisión de la información y la ergonomía y usabilidad intuitivas de un plató de televisión durante los procesos de preproducción y grabación de programas en directo.

1.2.1. Objetivos principales

El objetivo principal del trabajo de investigación aquí presentado es la actualización de la arquitectura y flujos de trabajo de los actuales platós virtuales de televisión para emisiones en directo. Para ello se propone la incorporación del uso sinérgico de nuevos sensores, para una mayor interacción del presentador con el entorno sintético, y la utilización de técnicas de visualización avanzada que permitan mejorar la composición e integración finales entre los mundos virtual y real, mejorando así la sensación de presencia. Asimismo, se propone la utilización de diversas técnicas de visualización avanzada e interacción en otras áreas tales como los entornos museísticos.

El desarrollo de la investigación se ha realizado sobre un plató virtual de televisión emplazado en la Facultad de Ciencias de la Comunicación de la Universidad de Santiago de

Compostela y equipado en el año 2011 a través del programa Profesionales Digitales [160]. Brainstorm Multimedia [144], una de las multinacionales más reputadas del sector, es la empresa encargada del software y la sensorización del mismo. El plató está configurado en base a una arquitectura hardware y software tradicional que sigue siendo, a día de hoy, la más común en todo el mundo. Esta tiene su base en un único sistema de sensores conectado directamente a los ordenadores encargados de generar las imágenes del mundo sintético. En base a lo expuesto hasta este punto, los objetivos principales de esta tesis son:

1. **Actualizar la arquitectura hardware y software tradicional de un plató virtual de televisión para emisiones en directo, con la finalidad de agilizar la inclusión de nuevos dispositivos en el sistema y permitir su funcionamiento sinérgico.**

La arquitectura hardware y software tradicional de los platós virtuales de televisión hace que la configuración de los mismos sea poco flexible y que la actualización o ampliación de los dispositivos implicados sea costosa, tanto temporal como económicamente. Del mismo modo, generalmente, estos sistemas hacen uso de una única sensorización monolítica, lo que limita sus capacidades debido a la distinta naturaleza de los elementos objeto de seguimiento en el escenario. Por tanto, surge la necesidad de actualizar esta arquitectura, haciéndola más flexible y simplificando la incorporación de nuevos dispositivos. Asimismo, ante la creciente oferta de sensores específicos para tareas concretas (seguimiento del cuerpo de una persona, seguimiento de manos, seguimiento de marcadores, etc.), se presenta la posibilidad de su uso sinérgico de forma que, trabajando en paralelo, unos puedan contrarrestar las carencias de los otros, aumentando así las capacidades de los platós virtuales de televisión.

2. **Utilizar técnicas de visualización avanzada para mejorar la calidad de la integración entre los mundos virtual y real.**

Uno de los principales objetivos de los platós virtuales de televisión es la obtención de imágenes sintéticas coherentes con los elementos reales presentes en el escenario. Para este fin, además de las sensorizaciones, se utilizan técnicas de visualización avanzada que intentan buscar el equilibrio entre el rendimiento (ejecución en tiempo real) y el realismo de la imagen obtenida. Sin embargo, a día de hoy, la calidad de los escenarios virtuales que pueden verse en televisión presenta numerosas carencias, como la falta de una iluminación coherente, ausencia de sombras, etc. Se hace necesario, por tanto, profundizar en estas técnicas y plantear nuevas soluciones que permitan mejorar la calidad

de las imágenes generadas por ordenador y, consecuentemente, la composición final de los mundos virtual y real.

3. Definir, en base a la arquitectura y técnicas de visualización avanzada propuestas, nuevos flujos de trabajo que permitan acelerar y simplificar los procesos de preproducción y grabación de un programa en directo.

La preparación de una emisión en directo desde un plató virtual de televisión requiere semanas de ensayo previo por parte de los presentadores y del personal técnico implicado. Esto es debido a la ausencia prácticamente total de interacción real entre los presentadores y su entorno virtual, lo que hace necesario el aprendizaje y ensayo de coreografías para simularla visualmente. A partir de una nueva sensorización, implementada sobre la nueva arquitectura, se plantea el cambio de los flujos de trabajo previos automatizando procesos y reduciendo los tiempos de aprendizaje necesarios por parte del presentador.

Asimismo, es necesario definir un proceso que permita, a partir de un modelo 3D simple, obtener una visualización realista del mismo que sea coherente con las imágenes captadas por las cámaras en el escenario.

4. Aplicar técnicas de visualización avanzada e interacción a otros entornos de trabajo

El mundo de los platós virtuales de televisión está directamente relacionado con la visualización avanzada de escenarios sintéticos y la interacción “persona-ordenador”. Estas dos ramas de la informática tienen un amplio abanico de campos de aplicación. Por ello, y teniendo en cuenta la amplia experiencia de los integrantes del grupo COGRADDE en la creación de aplicaciones museísticas, se plantea la aplicación de las técnicas y tecnologías utilizadas en el desarrollo de esta investigación a la creación aplicaciones interactivas para museos.

1.2.2. Objetivos secundarios

Dentro del proceso necesario para afrontar los objetivos principales del trabajo de investigación aquí presentado será preciso dar respuesta a otras cuestiones contextuales, entre las que destacamos:

1. Creación y adaptación de contenidos y efectos visuales

Para poder realizar pruebas de los desarrollos realizados será necesario disponer de un escenario virtual, así como de una serie de efectos visuales que se habrán de crear o adaptar a las necesidades específicas de la investigación.

2. Adaptación de los actores y técnicos a la utilización de las tecnologías y formas de interacción propuestas

En la fase de pruebas los técnicos y actores se tendrán que adaptar a las nuevas tecnologías y formas de interacción propuestas. Se utilizarán nuevos sensores y se automatizarán procesos, por lo que variará el modo de relacionarse con el entorno virtual y la forma de llevar el control de la grabación en directo. Se buscará agilizar esta adaptación mediante la utilización de técnicas de interacción natural que permitan acortar este proceso de aprendizaje.

3. Utilización de sensores de bajo coste

Una de las características de los platós virtuales de televisión tradicionales es su elevado coste. Parte de este viene derivado de la utilización de grandes sensorizaciones que, a pesar de su gran precisión, limitan sus capacidades al estar especializadas en el seguimiento de un único tipo de elemento. En los últimos años han aparecido en el mercado numerosos sensores de bajo coste especializados en tareas específicas que, en general, no han sido probados dentro del marco de una producción profesional de televisión. Por este motivo, se plantea como un objetivo secundario de esta tesis la utilización de sensores de bajo coste con la finalidad de reducir el precio de esta tecnología y así acercarla a un mayor mercado.

4. **Definición un marco de trabajo colaborativo con la industria que permita tanto un acceso temprano a la tecnología como la transferencia efectiva de la investigación realizada.**

La tecnología de los platós virtuales de televisión y su evolución es directamente dependiente de las empresas que ofertan soluciones comerciales específicas, como es el caso de Brainstorm Multimedia en el plató de la Facultad de Ciencias de la Comunicación. Para la realización de cualquier investigación o desarrollo se hace imprescindible una colaboración directa que permita avanzar a ambas partes. Por ello se plantea la necesidad de definir un marco de colaboración que permita, por un lado, conocer de primera mano las necesidades de la industria y dirigir la investigación a resolverlas y, por otro, disponer de las últimas versiones de software, incluso antes de su lanzamiento comercial, de cara a conocer la orientación y el alcance de la evolución tecnológica.

1.3. Estructura y organización de la tesis

Se presenta en este punto la estructura y organización seguidas en la redacción de este trabajo de investigación. Se analizará cada uno de los capítulos expuestos identificando aquellos puntos de mayor relevancia para una mejor comprensión del texto en su conjunto.

De modo sinóptico, cabe distinguir la articulación temática de los contenidos que integran esta tesis doctoral en cuatro secciones diferenciadas: una de introducción y presentación de los objetivos y estructura de la tesis, una centrada en el trabajo desarrollado en el marco de los platós virtuales de televisión, otra enfocada a otros trabajos de investigación desarrollados para el entorno de los museos interactivos y, finalmente, una de discusión de resultados y conclusiones generales, así como de planteamiento de diversas líneas de trabajo futuro.

■ *Introducción, objetivos y estructura de la tesis*

Comprendidos en el capítulo 1. En él se realiza una introducción al mundo de la visualización avanzada e interacción general y también específico de los entornos en los que se ha desarrollado la presente tesis: los platós virtuales de televisión y los entornos museísticos. Se presentan también los objetivos principales y secundarios que han guiado el trabajo de investigación y la estructura y organización del documento.

- *Platós virtuales de televisión*

Consta de cuatro capítulos en los que se presenta la tecnología de los platós virtuales de televisión, así como sus retos y el proceso seguido para su superación centrados en la interacción “persona-ordenador” y la visualización avanzada. En el capítulo 2 se presenta un recorrido por la evolución histórica y el estado del arte de los PVT, así como por su funcionamiento y arquitectura hardware y software, describiendo, de una forma detallada, los diferentes elementos que la componen. Se analiza también la utilización que se hace de esta tecnología en la televisión actual a través de un recorrido por los programas especiales dedicados a diferentes noches electorales de los últimos años en España y Reino Unido. En el capítulo 3 se presentan los retos que se plantean en los platós virtuales de televisión para una mejor integración de los mundos virtual y real. Se analiza también el marco de trabajo sobre el que se ha llevado a cabo el proceso de investigación. En el capítulo 4 se presentan las soluciones propuestas para mejorar la capacidad interactiva de los PVT a partir de una nueva arquitectura hardware y software que permita agilizar la inclusión, actualización y eliminación de dispositivos, al mismo tiempo que superar las restricciones impuestas por un hardware y software propietarios. Se presenta la inclusión ágil en un plató virtual de televisión de tres sensores (cámaras de infrarrojos Optitrack, Microsoft Kinect y Leap Motion) y la actualización de uno de ellos sustituyéndolo por su nueva versión (Microsoft Kinect V2). Todos estos dispositivos se han utilizado para afrontar diferentes los retos de la interacción en PVT planteados en el capítulo 3. En el capítulo 5 se presentan una serie de técnicas de visualización avanzada que se han adaptado para su utilización en platós virtuales de televisión con la finalidad de mejorar la integración visual entre los mundos virtual y real. Estas técnicas van desde el análisis de la correcta iluminación del plató hasta la captura de luz de un entorno real y su aplicación a entornos virtuales. Se realizan también propuestas sobre posibles soluciones para realizar una proyección de sombras de actores reales sobre elementos sintéticos.

- *Visualización avanzada e interacción aplicada a entornos museísticos*

Esta sección se desarrolla en el capítulo 6 en el que se hace un repaso de una serie de técnicas de interacción y visualización avanzada que se han aplicado en la elaboración de módulos expositivos reales en museos interactivos. Se presentan los proyectos “El Pórtico de la Gloria virtual”, “Los instrumentos virtuales del Pórtico de la Gloria”, “Conoce el Pórtico” y “Módulo de visualización del interior del busto humano”. Para cada

uno de ellos se presentan el proceso seguido desde la idea original hasta su implementación final en un museo, así como los resultados obtenidos por la exposición en la que fueron incluidos.

- *Discusión de resultados, conclusiones y trabajo futuro*

La última sección está comprendida en el capítulo 7 e incluye una discusión pormenorizada de los resultados obtenidos en el proceso de investigación, un análisis de las conclusiones fruto de la elaboración de la presente tesis y una serie de propuestas del trabajo futuro hacia el que podrían orientarse futuras investigaciones.

CAPÍTULO 2

PLATÓS VIRTUALES DE TELEVISIÓN

A continuación se revisa el desarrollo de los platos virtuales de televisión a lo largo de la historia, repasando la arquitectura hardware y software que los define, así como otros elementos significativos que intervienen en una producción desarrollada en un estudio virtual para su emisión en directo. Finalmente se analizan una serie de casos reales de utilización de esta tecnología a modo de ilustración del estado actual de la misma.

2.1. Historia de los Platós Virtuales de Televisión

Heredera de la tradición teatral y más tarde cinematográfica, la televisión ha utilizado, desde sus inicios, la escenografía y la iluminación de los platós para situar a los actores en un lugar, tiempo y atmósfera adecuados para el programa que se estaba grabando. Esta escenografía estaba habitualmente fabricada en madera, metal, cartón piedra, etc. y sigue siendo, a día de hoy, la más común en la mayoría de los programas. Sin embargo, con el paso del tiempo y la consiguiente maduración de los espectadores y de su nivel de exigencia perceptiva, se hicieron imprescindibles nuevas ilusiones y efectos visuales que rompiesen los límites de un plató de televisión tradicional, naciendo, de esta forma, los platós virtuales de televisión [140].

El origen de los PVT se remonta a los años 30, cuando una solicitud de patente presentada por Goldsmith [62] describió la forma de insertar información de una cámara de televisión en la señal proveniente de otra cámara, obteniendo como resultado una composición de ambas. El término *inset* fue utilizado para describir el proceso en el que un área de la imagen procedente de una cámara (generalmente un actor) es utilizada para eliminar el área correspondiente de la

señal obtenida por una segunda cámara (generalmente un fondo real o sintético), de forma que la imagen obtenida por la primera coincida exactamente en el *agujero* generado en la señal de la segunda. Esta técnica permite generar la ilusión de presencia de ciertos elementos (por ejemplo un actor) en un lugar en el que realmente no se hallan. En esa época la televisión era en blanco y negro y se utilizaban diferencias de brillo para delimitar las áreas que debían de ser eliminadas de la señal. Con la llegada de la televisión en color esta técnica evolucionó hacia el actual *chroma key* [88], que utiliza diferencias de color para discernir si un área es parte del fondo de una grabación (figura 2.1).

En los primeros prototipos los elementos que se integraban con el mundo real eran imágenes 2D, pero la evolución tecnológica fue implantando paulatinamente entornos tridimensionales generados por ordenador. Esta tercera dimensión permitió la integración con imágenes reales desde distintos puntos de vista, e incluso realizar movimientos de cámara en tiempo real. Para que esto sea posible la posición y orientación de las cámaras debe de ser conocida, ya que estas han de ser simuladas en el entorno 3D de cara a obtener imágenes sintéticas coherentes. Para conseguir este efecto es necesaria la utilización de sensores capaces de realizar un seguimiento de las cámaras en tiempo real, puesto que numerosos programas de televisión son emitidos en directo. Si además los actores deben de interactuar con el entorno virtual, es



Figura 2.1: Imagen real y composición mediante la técnica de *chroma key*

necesario conocer también las posiciones y movimientos de estos.

Las dificultades técnicas que presentaba la sensorización de los diferentes elementos debido a la baja precisión de los sensores disponibles y su poca flexibilidad, así como la necesidad de que el proceso al completo se realizase en tiempo real, provocaron que los platós virtuales de televisión no fueran funcionales en sus inicios. El *boom* de la tecnología ocurrió a finales de los ochenta y principios de los noventa de la mano de una serie de empresas que comenzaron a superar las limitaciones tecnológicas existentes hasta ese momento. La compañía japonesa NHK comenzó a trabajar en la composición de imágenes utilizando cámaras virtuales [72], teniendo como resultado un método de *chroma keying* denominado *Synthevision* [166] en el cual, a partir de una imagen de alta definición almacenada y estática, el sistema era capaz de extraer una ventana¹ cuya posición, orientación y escala era modificable a partir de la señal procedente de una serie de sensores mecánicos. Esto permitía la simulación de movimientos de rotación de cámara, enfoque y zoom. A partir de esta experiencia, en 1991 NHK utilizó su primer prototipo de plató virtual para producir un documental científico llamado *Nanospace*. Paralelamente, el *Institut National de Audiovisual* (INA) francés desarrolló el *Synthetic TV Project* [46] centrado en platós virtuales de televisión para emisiones en diferido.

A mediados de los noventa, tanto el aumento de la potencia gráfica de los ordenadores (con la aparición de supercomputadores gráficos como el SGI Onyx) como los nuevos sensores diseñados para el *tracking*² de cámaras de vídeo, permitieron la integración entre imagen real y elementos virtuales 3D manteniendo una tasa de refresco estable entre los 50 y 60 fotogramas por segundo, permitiendo la grabación o emisión de la señal en tiempo real. Uno de los primeros ejemplos de esta tecnología es el *Mona Lisa Project* [19], que desarrolló e integró las diferentes tecnologías (algoritmos, software y hardware) implicadas en la construcción, manejo y visualización de modelos 3D en tiempo real en un plató virtual de televisión.

Los primeros PVTs comercializables fueron presentados a las grandes compañías de televisión en 1994, con motivo de la celebración del *International Broadcasting convention* (IBC) en Amsterdam. Estos primeros prototipos eran difíciles de configurar, poco fiables y además tenían un elevado coste, pero, sin embargo, esto no frenó su impulso y la tecnología siguió evolucionando rápidamente hasta finales de los años noventa. En 1995 la empresa ORAD presentó un sistema llamado *CyberSet* [96] que hacía uso de un sistema de seguimiento de cámaras basado en el reconocimiento de patrones sobre el fondo del plató real (figura 2.2).

¹En este caso se define ventana como un área rectangular que engloba una parte de la imagen y puede desplazarse a través de ella

²*Tracking* es la acción de seguir el movimiento de un determinado elemento en el espacio.



Figura 2.2: Sistema de *tracking* de cámaras Cyberset, creado por ORAD

En 1996 SGI presentó un nuevo motor gráfico que permitía, además de manejar escenas más complejas, crear efectos de desenfoque sobre los objetos virtuales.

En 1998, Andrew Wojdala describió los principales retos a los que se enfrentaban los estudios virtuales de televisión de aquella época [192]. Estos iban desde las limitaciones técnicas de los dispositivos o la iluminación coherente del escenario virtual hasta la precisión de los sensores o la interacción entre los actores y el entorno virtual. A día de hoy, muchas de estas cuestiones no han sido todavía completamente resueltas. Se creyó entonces, ante la rápida evolución que estaba experimentando, que la tecnología necesaria para el correcto funcionamiento de los PVTs iba a revolucionar la forma de crear los programas de televisión en un corto espacio de tiempo. Se vaticinaba que la potencia gráfica evolucionaría a una gran velocidad, pero el avance en los ordenadores de altas prestaciones no fue tan rápido como en los ordenadores personales y esto acabó frenando el impulso inicial.

Tras el parón producido en pleno umbral del siglo XXI, en los últimos años el desarrollo de los platós virtuales de televisión ha recuperado el impulso debido a la evolución técnica de todos los sistemas que forman parte de una producción. A día de hoy, una estación de trabajo o un PC con una tarjeta gráfica común tienen suficiente potencia como para realizar el *render*³ en tiempo real de modelos 3D complejos, y los nuevos sensores hacen posible la mejora continua de la composición entre los mundos virtual y real. Estas mejoras van desde la calidad y el realismo de las imágenes sintéticas generadas hasta la alineación de las cámaras o la posibilidad de interacción entre ambos mundos. Un ejemplo de los resultados que pueden

³*Render* es el proceso seguido por un computador para generar una imagen 2D a partir de un modelo 3D con unas características propias de iluminación, texturas, etc. desde un punto de vista determinado.



Figura 2.3: Imagen de la demostración pública de su tecnología realizada por Brainstorm Multimedia en el BIT Broadcast 2016

llegar a obtenerse en la actualidad puede observarse en la figura 2.3.

Actualmente el desarrollo de la tecnología está capitalizado por una serie de empresas multinacionales que introducen evoluciones continuas en sus productos para mantener su situación predominante en el mercado. Las principales multinacionales que ofrecen servicios integrales de platós virtuales de televisión son: Brainstorm Multimedia [144], Avid [10] tras la compra de ORAD [11] en 2015 y VIZrt [185]. Las tres trabajan con multitud de canales de televisión en todo el mundo implementando sus productos con distintos enfoques: mientras que ORAD y VIZrt ofrecen sistemas más cerrados a los que se pueden acoplar módulos para aumentar las prestaciones del plató, Brainstorm Multimedia ofrece una alternativa basada en código *Python* [55] que le permite al usuario final una mayor flexibilidad a la hora de manejar sus productos o integrarlos con una tecnología no soportada oficialmente por la compañía.

La mejor prueba de la recuperación de esta tecnología y su gran avance en los últimos años es su presencia en prácticamente todos los canales de televisión en la actualidad. Aunque sus campos de aplicación principales son los programas de deportes, la información meteorológica [21], las noches electorales y los informativos, el campo de aplicación y, por tanto, la capacidad de expansión de los platós virtuales de televisión es ilimitada, pudiendo ofrecer grandes alternativas también en el diseño de programas de entretenimiento. Prueba de ello es “La Ruleta de la Suerte” (figura 2.4), que se emite desde hace años en España en el canal Antena 3 y que hace uso de la tecnología de Brainstorm Multimedia.



Figura 2.4: Imagen del programa “La Ruleta de la Suerte”

Paralelamente a esta recuperación de la tecnología en los canales tradicionales, se están abriendo de forma continua nuevos mercados que hacen augurar grandes posibilidades de expansión. El uso extendido de platós de *chroma* en emisoras locales [107], así como el *boom* de los canales de televisión de bajo coste que emiten a través internet y no pueden permitirse un plató virtual de televisión tradicional por su elevado coste económico, convierte a estas dos áreas en potenciales clientes de una tecnología de platós virtuales más asequible. Por otro lado, se está generalizando el uso de estudios virtuales para docencia a distancia [202], ya que permiten generar materiales de gran calidad didáctica sin que ello implique una asunción de los elevados costes derivados la escenografía física tradicional [91]. Parte del futuro de esta tecnología, por tanto, está en la implementación de platós virtuales de bajo coste que la hagan asequible a este creciente mercado [195].

2.2. Arquitectura Clásica de un Plató Virtual de Televisión

A pesar de su corta historia, los platós virtuales de televisión han evolucionado rápidamente, si bien su desarrollo, aunque centrado principalmente en la parte técnica, no ha modificado sustancialmente los planteamientos iniciales, en lo que se refiere a arquitectura hardware y software. La configuración clásica se divide en tres subsistemas: *Render*, Sensorización y Vídeo y Composición (figura 2.5). Existe un cuarto bloque que estaría formado por el resto del equipo necesario para la emisión de un programa en directo, aunque no está directamente involucrado en la parte de virtualidad del estudio [60].

La función y componentes de cada uno de estos tres subsistemas es el siguiente:

- **Subsistema de render:** es el encargado de generar las imágenes sintéticas a partir de la información recibida de los sensores y del modelo 3D del entorno. Suele configurarse

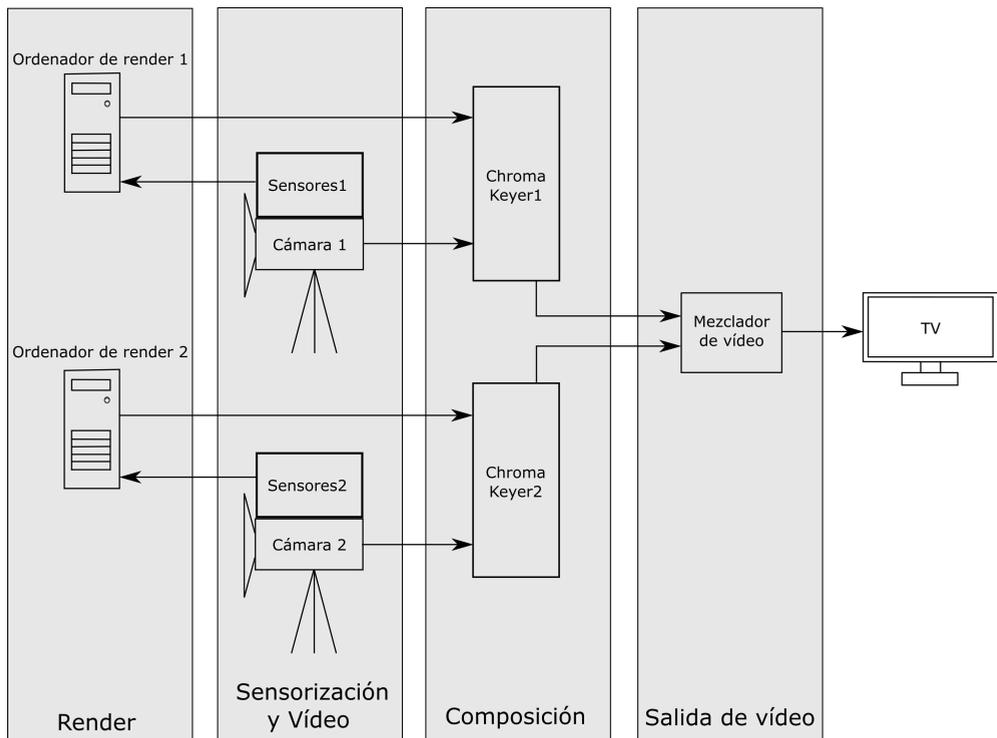


Figura 2.5: Esquema tradicional de un plató virtual de televisión con dos cámaras

en base a un ordenador y una instancia del software de render por cada cámara presente en el plató.

- **Subsistema de sensorización y vídeo:** es el encargado de grabar el mundo real y seguir el movimiento de cada cámara en el plató. Está compuesto por las cámaras y el sistema de *tracking*.
- **Subsistema de composición:** es el encargado de combinar en una sola imagen las señales recibidas del subsistema de *render* y de las cámaras. Está formado por un *chroma keyer* por cada par “ordenador de *render*-cámara”.

Todas las conexiones ente los diferentes subsistemas son directas. Los sensores mandan su información directamente a los ordenadores de render, y estos, al igual que las cámaras, envían las imágenes generadas a los *chroma keyers*.

El funcionamiento de un plató virtual se basa en dos líneas paralelas que acaban combinándose para formar la señal de emisión final. Por un lado, la cámara captura la imagen real de los actores y elementos reales que componen el escenario, junto con el fondo monocromo, y envía esta señal al *chroma keyer*. En paralelo, los sensores de la cámara envían información de su posición, orientación, zoom y foco al ordenador encargado de la generación del contenido 3D. El motor gráfico⁴ del software de *render* recibe esta información y sitúa una cámara virtual, previamente calibrada para que tenga las mismas distorsiones ópticas que su correspondiente real, en la posición y orientación indicada por los sensores. Se obtiene, de este modo, una cámara virtual que simula a su equivalente real.

Además de la imagen del escenario, el motor gráfico genera una máscara de valores alfa que servirá para conocer la posición relativa de los elementos virtuales y los reales. La imagen contiene dos valores: blanco, para las zonas de la imagen en las que se encuentran objetos virtuales que se sitúan en la composición final por delante del mundo real, y negro, para las zonas que se encuentran detrás. Esta máscara puede ser alterada en directo, permitiendo a los actores deambular con naturalidad alrededor de objetos virtuales (figura 2.6).

Existe un condicionante tecnológico decisivo, y es que todos estos cálculos, para un programa de emisión en directo, han de ser llevados a cabo en tiempo real, lo que significa generar 50 o 60 imágenes por segundo. Esto limita la complejidad geométrica de los entornos 3D, así como los efectos que se les pueden aplicar (iluminación, partículas, etc.), y supone uno de los

⁴Un motor gráfico es un software utilizado por aplicaciones y programas para la generación gráficos sintéticos



Figura 2.6: Escenario en el que una pantalla virtual se sitúa por delante del mundo real y máscara alfa codificando su posición relativa

equilibrios más complejos del mundo de los platós virtuales de televisión: el que se establece entre realismo y tiempo real.

Una vez se ha obtenido la imagen real de la cámara, y la virtual y la máscara alfa del motor gráfico, el *chroma keyer* lleva a cabo la composición de las tres dando como resultado la imagen final que será emitida en directo.

A continuación se presentan, de un modo pormenorizado, los diferentes componentes de cada uno de los subsistemas de un plató virtual de televisión, así como sus características, finalidad y funcionamiento.

2.2.1. Subsistema de *render*

Dentro del subsistema de *render* se generan las imágenes sintéticas que conforman el mundo virtual. Está formado por un elemento principal: el ordenador de render.

2.2.1.1. Ordenador de *render*

El ordenador de render es el equipo encargado de generar una imagen sintética coherente a partir de los datos recibidos de los sensores. En él se ejecuta un motor gráfico diseñado específicamente para platós virtuales de televisión, en el que, como se ha comentado previamente, se crean cámaras virtuales que mimetizan a las reales generando vistas del modelo 3D que funcionan como entorno virtual de la emisión. Además de esto, este tipo de programas le permiten al operador realizar todas las acciones necesarias para la producción de programas en directo: desde modificar la posición de objetos o lanzar efectos visuales, hasta insertar rótulos e incluso generarlos en tiempo real a partir de la información recibida de una base de datos.

Históricamente, para realizar todas estas acciones eran necesarios supercomputadores gráficos. Sin embargo, a día de hoy, una estación de trabajo común es capaz de realizar el *render* de grandes cantidades de polígonos manteniendo una velocidad de refresco de 60 fotogramas por segundo. Como resulta evidente, cuanto más potencia gráfica tenga el hardware, mayores escenarios virtuales y mejores efectos visuales podrá soportar. A pesar de esto, la característica más importante en un ordenador de *render* es su robustez, es decir, que tenga la capacidad de mantener una frecuencia de refresco estable, sin grandes picos o valles en su rendimiento, ya que un problema de este tipo repercutiría directamente en la sensación de presencia de los actores en el entorno sintético, provocando BIPs (*Break In Presence* o interrupciones de presencia)[161]. La consecuencia directa de una bajada en la tasa de refresco es un movimiento de la imagen real sin que la acompañe el entorno virtual y, a continuación, un desplazamiento brusco del entorno sintético a su posición correcta. Si se generan menos fotogramas por segundo de los que están siendo grabados por las cámaras, se produce como resultado un entorno virtual que se desliza a saltos mientras el mundo real se mueve de un modo suave y natural.

2.2.2. Subsistema de sensorización y vídeo

El subsistema de sensorización y vídeo es el encargado de capturar la imagen real del plató y realizar el *tracking* de los diferentes elementos que han de ser seguidos dentro del mismo. Sus principales elementos son las cámaras y los sensores.

2.2.2.1. Cámaras

Las cámaras en un estudio virtual de televisión no se diferencian en lo esencial de las de un plató al uso. Para su integración se emplean sensores externos que se acoplan a los trípodes o a la estructura de la propia cámara. Sin embargo, a la hora de producir una combinación correcta entre el mundo virtual y el real, no sólo es necesario conocer su posición y orientación, sino que, simultáneamente, deben reproducirse en el entorno sintético la configuración y características intrínsecas de la cámara. Es necesario crear una cámara virtual que imite con exactitud las características (distorsión de la lente, campo de visión, apertura, etc.) de la real, de modo que ambos mundos compartan la misma apariencia y sufran las mismas distorsiones.

Las cámaras actuales acostumbran a tener protocolos de comunicación que facilitan el acceso a su configuración, simplificando enormemente el cálculo de los parámetros de operación de las mismas. A través de un proceso de calibrado específico, la distorsión natural de la lente (diferente en cada cámara) puede ser modelada matemáticamente y posteriormente simulada en el entorno virtual. Es necesario combinar estos parámetros con la información recibida de los sensores, como el zoom y el foco, para que la integración final sea correcta. Toda esta información permite, por ejemplo, que ciertas zonas del plató virtual aparezcan desenfocadas y que, al variar el foco de la cámara, se pueda desenfocar a los actores mientras se enfoca el entorno virtual o viceversa. El resultado de la combinación de todos estos elementos busca que el manejo de la cámara en un estudio virtual de televisión se asemeje lo máximo posible al de un plató tradicional, simplificando de esta forma el trabajo al operador.

2.2.2.2. Sensores

Como se ha comentado previamente, para que una cámara virtual se mueva de forma coherente dentro de un plató, su equivalente en el mundo real ha de estar sensorizada. Pero, si además de la alineación del punto de vista de la imagen, se desea que exista una interacción activa entre los mundos virtual y real, ha de realizarse un *tracking* del movimiento de otros elementos en la escena. En general, existen tres tipos de elementos que se deben sensorizar:

- **Cámaras:** es indispensable sensorizar las cámara para poder reproducir sus características en el mundo virtual. Las características más comunes que es necesario conocer son la posición, la orientación, el zoom y el foco.
- **Objetos:** a la hora de permitir la manipulación de objetos virtuales, una de las soluciones más comunes es asociarlos a objetos reales sensorizados que son ocluidos en la imagen final por los modelos 3D que controlan. Del mismo modo, si se desea que un objeto real tenga una interacción activa con el entorno virtual, ha de conocerse su posición y orientación en todo momento.
- **Usuarios:** si se desea que exista una interacción entre los actores, presentadores e invitados y el espacio virtual en el que se insertan, estos deberán estar sensorizados. Esta sensorización dependerá del tipo de interacción que se desee desarrollar. Existen, principalmente, tres tipos de interacción en los platós virtuales de televisión:
 - Oclusiones o cambio de posición relativa entre usuarios y escenografía virtual. Como se ha comentado anteriormente, un objeto 3D puede estar delante o detrás de un actor y, por tanto, su posición relativa puede variar a medida que uno u otro se desplazan en el espacio. Sensorizar la posición del actor permite automatizar este proceso y llevarlo a cabo en tiempo real.
 - Interacción directa entre una persona y objetos virtuales [120]. Un actor puede chocar con elementos virtuales y estos deberían de reaccionar de alguna forma para no afectar a la sensación de presencia. Asimismo, los objetos del mundo 3D deberían de ser directamente manipulables. Este tipo de interacción acostumbra evitarse por su complejidad intrínseca, utilizando referencias en el escenario real que eviten que los actores entren en contacto directo con elementos virtuales. Para que estas acciones sean posibles, ha de realizarse un seguimiento de la anatomía completa de los presentadores, ya que cualquier parte de su cuerpo puede entrar en contacto con el mundo sintético.
 - Activación de efectos en el mundo virtual. Este tipo de interacción consiste en la activación, por parte del presentador, de efectos predefinidos en el mundo virtual, como comenzar la reproducción de un vídeo, mostrar una infografía, iniciar una reproducción de sonido, etc. Este proceso puede ser realizado, principalmente, mediante el uso de dos técnicas: la utilización de un mando a distancia que controle los efectos mediante la pulsación de botones y la detección de gestos. Esta

última requiere de la sensorización del actor y el procesado de la señal obtenida para detectar y clasificar los movimientos que han de tener como consecuencia la activación de algún efecto.

Tradicionalmente los sensores utilizados en platós virtuales de televisión y aplicaciones de realidad virtual pueden clasificarse en tres tipos en función de la tecnología que utilizan para realizar el *tracking*:

- **Electromagnéticos:** estos sensores están basados en dos elementos principales: un emisor que crea tres campos electromagnéticos distintos y un receptor que analiza dichos campos para obtener su posición y orientación con respecto al emisor (figura 2.7). Estos sistemas fueron muy utilizados en aplicaciones de realidad virtual y tuvieron su campo de pruebas, también, en el mundo de los platós virtuales de televisión. Sin embargo, han caído en desuso debido a la gran cantidad de ruido que introducen por las distorsiones en los campos electromagnéticos producidas por dispositivos de uso cotidiano, tales como teléfonos móviles, y de modo muy especial, por la microfónica inalámbrica que caracteriza este tipo de platós, donde la ausencia de cables resulta esencial para la movilidad de los actores y la ausencia de interferencias visuales en los planos generales.
- **Mecánicos:** son los sensores más comunes para realizar el seguimiento del *pan*, *tilt*, *zoom* y foco de las cámaras cuando estas están montadas sobre una estructura fija (un



Figura 2.7: Sistema de *tracking* electromagnético en un entorno de realidad virtual CAVE



Figura 2.8: Sensor mecánico Vinten-Radamec

trípode, una grúa, etc.). Generalmente están formados por una serie de poleas y engranajes asociados a potenciómetros, situados en el soporte de la cámara, que son capaces de medir, de una forma muy precisa, las interacciones del operador con la misma (figura 2.8). Cuando el soporte es un trípode fijo es necesaria una configuración previa de su posición, que no podrá variar durante la grabación del programa. En caso de necesitar desplazar la cámara, estos sensores deberán acoplarse a un dispositivo de *travelling* que se mueva a través de unos raíles. Del mismo modo, si la cámara se sitúa en una grúa o *boom*, esta puede sensorizarse para obtener la posición y orientación en tiempo real.

Sin embargo, estos sensores no permiten el movimiento libre de la cámara en el espacio, ya que están limitados al rango que pueda alcanzar la maquinaria a la que están acoplados. No pueden ser utilizados, por ejemplo, para seguir el desplazamiento de una *steadycam*⁵.

- **Ópticos:** esta tecnología hace uso de técnicas de visión por computador para, utilizando una serie de cámaras, analizar las imágenes capturadas en busca elementos de referencia que le permitan calcular posiciones y orientaciones. Es, actualmente, la solución más extendida para la sensorización de diferentes tipos de entornos.

Dentro de los sensores ópticos se distinguen tres tipos:

- **Sensores de Infrarrojos:** se trata de cámaras con un rango de visión centrado en el espectro infrarrojo, invisible, por tanto, al ojo humano, lo que los hace ideales

⁵*Steadycam* es el nombre comercial del primer estabilizador de cámara y da nombre, en la actualidad, a todos los sistemas de este tipo. Permite el desplazamiento libre de la cámara a pie eliminando los movimientos bruscos producidos en la misma por la acción de caminar

para su utilización en platós virtuales de televisión (figura 2.9). Están compuestos por una serie de cámaras que detectan un conjunto de marcadores que pueden ser activos (cuando emiten luz infrarroja) o pasivos (cuando la reflejan). En ambos casos, estos marcadores son visibles, por lo que será necesario esconderlos en la imagen final de emisión.

Dentro de esta familia han surgido, en los últimos años, diversos sensores que han superado la necesidad de utilizar marcadores físicos. Destacan los dispositivos destinados a la captura de la anatomía humana, algunos de los cuales han llegado a un nivel de precisión que los hace ideales para ser utilizados en estudios biomecánicos [26]. Las dos técnicas más comunes utilizadas en estos dispositivos son la proyección de patrones de luz estructurada [52], como en los casos del Asus Xtion PRO [81] o del Microsoft Kinect [201], y la cámara de tiempo de vuelo [105] empleada por el Microsoft Kinect V2 [198].

- **Sensores de espectro visible:** estos sensores son cámaras estándar y en un plató virtual de televisión pueden ser las propias cámaras utilizadas en la grabación del programa. Al funcionar en el espectro visible detectan patrones de color a partir de los cuales, teniendo en cuenta su forma y tamaño, son capaces de calcular su posición y orientación respecto al sensor [103]. Aunque generalmente los patrones detectados están impresos sobre objetos físicos, existen sistemas que, mediante la



Figura 2.9: Sistema de sensores infrarrojos *outside-in*

sincronización de proyectores con cámaras de alta velocidad (120 fotogramas por segundo), permiten proyectar los patrones sobre elementos en el escenario manteniéndolos invisibles en la imagen final de emisión [67]. La principal ventaja de utilizar esta segunda opción reside en su capacidad de adaptación en el tiempo, permitiéndoles una visibilidad óptima desde las cámaras en todo momento [68].

Los sensores ópticos de espectro visible han sido ampliamente utilizados en los platós virtuales de televisión con dos finalidades básicas: localizar la posición de la cámara con respecto a un patrón fijo situado en el plató [196] o calcular la posición de un objeto dentro del escenario para asociarle un objeto virtual y así hacerlo interactivo [17].

Dentro de esta categoría también existen tecnologías, como la ofrecida por la empresa *Ncam* [172], que utilizan cámaras de profundidad para detectar puntos de referencia en la escena que permitan calcular los movimientos del sensor (figura 2.10). Esta tecnología puede resultar muy útil en escenas con gran cantidad de atrezzo y escenografía e incluso puede ser utilizada en exteriores, pero su funcionamiento se ve gravemente comprometido en un plató vacío, ya que el ciclorama carece de puntos singulares que tomar como referencia [181]. Asimismo, la empresa *Organic Motion* ofrece el sistema *Openstage* [141] capaz de realizar una captura de movimiento del cuerpo humano con calidad profesional, sin utilizar marcadores, a partir de las imágenes obtenidas por una serie de cámaras RGB.



Figura 2.10: Sensor ncam

Otros sensores ópticos: otro tipo de sensores ópticos, utilizados generalmente para el *tracking* 2D (localización en el plano del escenario, pero no altura) son los basados en láser [191]. Un emisor láser en combinación con un espejo en rotación crea un plano invisible. El sistema detecta intersecciones con ese plano y es capaz de calcular posiciones dentro del área del mismo. Utilizando este sistema se pueden realizar interacciones como la apertura de puertas virtuales mediante la aproximación del actor. Este tipo de sistemas también son muy eficaces para calcular la posición relativa del presentador y los objetos virtuales.

- **Sensores inerciales:** en los últimos años, con la expansión de los videojuegos y el *boom* de la realidad virtual, han aparecido en el mercado numerosos dispositivos que, haciendo uso de sensores inerciales, permiten diferentes tipos de interacción del usuario con los sistemas, pudiendo tener también viabilidad su aplicación en el campo de los platós virtuales de televisión. Al estar diseñados para interactuar con consolas u ordenadores, se trata de periféricos muy centrados en la interacción persona-ordenador, lo que los hace ideales para resolver los problemas de interactividad⁶ de los actuales PVTs. Ejemplo de este tipo de sensores son el Wiimote [147], el Ring Zero [82] o la muñequera MYO [100](figura 2.11).

Finalmente, existen soluciones híbridas, como la presentada por el sistema Matris, que combinaba sensores de inercia con el análisis de imagen [23].



Figura 2.11: Muñequera MYO

⁶Una de las características que dan valor a un plató virtual es el hecho de que el mundo virtual responda a las acciones de los actores del mundo real, esta cualidad se denomina interactividad.

Paralelamente a la clasificación antes propuesta, los sistemas de sensores ópticos con múltiples cámaras pueden clasificarse, también, en función de la posición relativa de los sensores y marcadores [174]. Según este criterio se diferencian dos configuraciones:

- **Configuración *outside-in*:** estos sistemas utilizan cámaras fijas montadas en torno al área de actuación y un conjunto de marcadores activos [137] o pasivos que se desplazan en el volumen de captura (figura 2.9). En este tipo de configuraciones resulta muy económico incluir nuevos marcadores, ya que, generalmente, estos constan simplemente de un conjunto de reflectores o emisores *led*. Sin embargo, aumentar el tamaño del volumen de captura resulta costoso, ya que implica aumentar el número de cámaras necesarias para cubrir el área al completo. Un ejemplo de este tipo de sistemas es Cam Trak ofrecido por la empresa Motion Analysis Corporation [27].
- **Configuración *inside-out*:** estos sistemas utilizan la imagen capturada por una cámara anclada al objeto del que se desea conocer su posición. Esta cámara detecta marcadores o patrones fácilmente identificables en el plató y cuya posición es conocida (generalmente se sitúan en el techo)(figura 2.12). A partir de esta, se calcula la posición de la propia cámara. Estos sistemas, fueron originalmente desarrollados para entornos de



Figura 2.12: Marcadores situados en el techo utilizados en una sensorización *inside-out* de la BBC

realidad aumentada con HMDs (*Head Mounted Displays*) [187], pero a lo largo de los años se han adaptado para su uso en los estudios virtuales de televisión. Ejemplos de este tipo de sistemas son el propuesto por Rahbar y Reza [159] o el free-d system [175] desarrollado por la BBC, que utiliza marcadores reflectantes circulares en el techo del plató y una pequeña cámara situada en posición vertical (apuntando hacia arriba) anclada a cada cámara del estudio.

2.2.3. Subsistema de composición

El subsistema de composición es el encargado de combinar la imagen real de la cámara, la virtual generada por el ordenador y la máscara alfa y obtener el fotograma final que será emitido. Los elementos principales para conseguir que esta composición se lleve a cabo de una forma correcta son el escenario, el ciclorama que lo rodea y el *chroma keyer*, encargado de llevar a cabo el proceso.

2.2.3.1. Escenario y ciclorama

El escenario es el lugar en el que se situarán los actores y el *atrezzo*, y lo que captarán las cámaras como fondo de la escena que se esté grabando. El escenario al completo se rodea por un ciclorama monocromo que permite la eliminación del fondo de la imagen mediante la técnica del *chroma key*. Esta técnica utiliza diferencias de color para eliminar partes no deseadas de la imagen, preservando únicamente aquellas zonas que no compartan tonalidad con el ciclorama de fondo.

Generalmente se utiliza esta técnica por ser la más eficiente y simple, ya que trabaja pixel a pixel eliminando aquellos de similar color al del *chroma key*. Frente a la ventaja que supone la sencillez y rapidez de este método, presenta la desventaja evidente de que la ropa del presentador, o los objetos que este manipule y lo rodeen, no podrán ser del mismo color que el fondo, ya que serían eliminados de la imagen, produciendo inconsistencias en el resultado final. Por otra parte, si se desea disponer de un objeto invisible al espectador, basta con pintarlo de ese mismo tono.

En el proceso de *chroma key* se puede utilizar cualquier color para el fondo, siempre y cuando este no esté presente en los objetos y actores que deben conservarse en la imagen final. Los colores más utilizados son el verde y el azul saturados. Tradicionalmente, con el uso de cámaras analógicas, se utilizaba el color azul, ya que los reflejos que provoca son menores y su presencia en la naturaleza y en la piel humana es menos prevalente. En la actualidad,

el color predominante es el verde, ya que los sensores de las cámaras de vídeo digitales son más sensibles a este tono y el ojo humano es más sensible a la luz verde [53]. Debido a esta mayor sensibilidad, la imagen capturada presenta menos ruido y, por tanto, se necesita menos luz para iluminar el fondo del ciclorama [154], evitando de esta forma reflejos no deseados sobre los elementos del plató. Sin embargo, si la ropa de los actores o parte del mobiliario fuese de este color, lo ideal sería un fondo azul. Del mismo modo, estos dos colores pueden presentar problemas con el color de los ojos de un actor, siendo necesario, en estos casos, el uso de lentillas de color. En este sentido se han propuesto técnicas alternativas como usar fondos con líneas de dos colores [197] o patrones de cuadrados del tipo tablero de ajedrez [5], pero al añadir complejidad al proceso, no han conseguido sustituir al uso de un solo tono. Incluso existen ejemplos de fondos reflectantes que no necesitan de una luz adicional y, en algunos casos, utilizan como iluminación únicamente un anillo de *leds* situados en torno a la lente de la cámara [106].

Existen otras técnicas para realizar esta diferenciación entre fondo a eliminar y contenido a mantener en la imagen final, como la estimación de profundidad, ya sea utilizando múltiples cámaras [176], o calculando el tiempo de vuelo de la luz infrarroja [79]. En algunos casos, cuando el fondo es estacionario, se pueden utilizar técnicas de extracción de fondos como las usadas en [177]. Pero el uso de estas técnicas no está extendido por no ser ninguna de ellas tan robusta ni eficiente como el *chroma key*.

En cuanto a la construcción del ciclorama se ha de tener en cuenta que, al tener el suelo y las paredes una orientación distinta, sus uniones deben de estar formadas por secciones redondeadas (figura 2.13), evitando las aristas y los cambios de color que estas implican [85]. En caso contrario las aristas aparecerían en la imagen final en forma de líneas.

2.2.3.2. *chroma keyer*

El *chroma keyer* es el encargado de realizar la integración entre los mundos virtual y real. Este proceso puede ser realizado en tiempo real por software en un PC convencional a través de un programa específico, o mediante un hardware diseñado para este cometido. La utilización de hardware dedicado suele ser la opción más utilizada ya que, debido a la estandarización del proceso, la utilización de un dispositivo específico asegura unas velocidades de refresco estables. Este dispositivo recibe tres señales de forma continua. Una es la señal que viene directamente de la cámara situada en el escenario, otra es la señal generada por el ordenador desde una cámara virtual, con las mismas características que la real, y la tercera es una



Figura 2.13: Detalle de la construcción de un ciclorama con la unión de paredes y suelo en curva

máscara de valores alfa que codifica las posiciones relativas de los objetos virtuales respecto al mundo real. El *chroma keyer* elimina de la señal de cámara todos los píxeles que compartan tono con la clave de color y, utilizando la máscara alfa, realiza la composición con los objetos virtuales del fondo y el frente. Se trata, por tanto de una composición de tres capas: la primera es el fondo 3D, la segunda es el entorno real y la tercera contiene los objetos virtuales situados delante del mundo real.

Además de realizar esta tarea, los *chroma keyer* permiten realizar dos ajustes de crucial importancia en un plató virtual de televisión:

- Permiten modificar el umbral utilizado para la aplicación del *chroma keying*. De esta forma, se puede adecuar este valor a las características de la iluminación, vestuario y configuración general del plató virtual para un programa específico, minimizando así los efectos no deseados, como halos, ruido de señal, etc.

- Permiten introducir un *lag*⁷ configurable entre la imagen capturada y la generada por ordenador. Este retraso es necesario para la sincronización de ambas señales ya que, mientras las imágenes captadas por la cámara llegan directamente al chroma keyer, la generación de imágenes sintéticas precisa, sin embargo, de un proceso necesariamente más complejo (captura de información por parte de los sensores, procesamiento de esta información y *render* del escenario) y cronológicamente más costoso, lo que produciría una asincronía en la composición final.

2.2.4. Otros elementos a considerar

2.2.4.1. Referencias para los actores

En un plató virtual de televisión los actores se mueven dentro de un espacio vacío, formado por un escenario rodeado por un ciclorama monocromo y, en el mejor de los casos, algún elemento de escenografía real que puede servir como referencia espacial. Sin embargo, desde el punto de vista del espectador, se observa cómo el presentador se desenvuelve en un entorno virtual, dando por hecho, de forma intuitiva, su presencia en el mismo y, por tanto, que los elementos virtuales que él ve son también visibles para el actor. Cualquier incoherencia entre la parte real y la virtual afecta directamente a la sensación de presencia transmitida. Por esto es de capital importancia dotar a los actores de referencias y herramientas que les permitan desenvolverse con soltura en el vacío del escenario.

De esta necesidad han surgido una serie de soluciones o herramientas, que permiten simplificar la navegabilidad del presentador en el escenario virtual [33]. Para este fin, tradicionalmente se han utilizado marcas físicas o volúmenes del mismo color que el ciclorama, de forma que son eliminados en el proceso de *chromakeying* (figura 2.14). Esta opción presenta varios inconvenientes como el hecho de que las referencias pueden variar al moverse los objetos virtuales (mientras que las marcas conservarían su ubicación original) o la influencia directa que tiene en el comportamiento del actor, afectando a su naturalidad.

Por este motivo, a lo largo de la historia de los platós virtuales de televisión se han desarrollado otras soluciones más avanzadas que permiten una retroalimentación entre el usuario y el entorno sintético, de forma que el actor reciba información sobre la localización de los

⁷*Lag* es un retraso producido en una telecomunicación que dificulta el desarrollo normal de la misma. En el caso de los platós virtuales de televisión el principal *lag* viene introducido por la diferencia de tiempo entre el instante de captura de una imagen por las cámaras y el tiempo necesario para generar el fotograma correspondiente a ese mismo instante de tiempo en el ordenador de *render*.



Figura 2.14: Uso de marcas físicas en el suelo del estudio

objetos virtuales y su volumen en tiempo real . A continuación se describen los tres tipos de soluciones más destacadas [75]:

1. **Visual:** dentro del campo de información visual se puede hacer una clasificación entre información mostrada en monitores e información proyectada. La primera muestra al actor la imagen que está siendo emitida a través de un monitor que se sitúa en su campo de visión (no afectando así a la naturalidad de su comportamiento). El usuario, al poder verse a sí mismo en el entorno sintético, recibe información directa que le permite orientarse en el mundo virtual. La segunda solución es utilizada en la actualidad por televisiones como la BBC [14] y consiste en la proyección del mundo virtual sobre el ciclorama [65], lo que da al presentador la información visual de una forma muy natural y directa sobre en qué punto específico del escenario se sitúan los diferentes objetos virtuales [200] . Para que estas proyecciones no sean visibles en la imagen final de emisión, han de ser monocromas, del mismo color que el ciclorama, haciendo además uso de luz polarizada no visible desde las cámaras [183] o sincronizadas con las cámaras para proyectar imágenes solo en los instantes en que la cámara no está

captando información (entre el final de la captura de un fotograma y el principio del siguiente).

Una solución más compleja es la presentada en [167], en la que la posición de un elemento virtual es transmitida a un sistema mecánico de poleas, que a su vez sitúa una esfera verde (del color del ciclorama) suspendida mediante cuerdas transparentes en el escenario real, de forma que el actor interactúa con un elemento real que es eliminado de la imagen en proceso de *chromakeying*.

2. **Vibrotáctil:** se basa en el uso de actuadores vibrotáctiles situados en diferentes partes de su cuerpo que se activan ante la proximidad de un objeto virtual [3]. Utilizando esta técnica puede enviarse información a diferentes puntos del cuerpo que estén cerca de colisionar con un objeto virtual (por ejemplo un brazo como en [95]) permitiéndole al actor evitar incoherencias visuales como una colisión con un objeto virtual en las que este no reaccione. Se han realizado estudios en los que, para ciertas tareas, se ha demostrado que este tipo de orientación es más efectivo, incluso, que las indicaciones por voz [188]. Por otro lado, Vierjahn et al. han estudiado y evaluado las diferentes formas de transmitir información de orientación vibrotáctil en [193, 184] llegando a la conclusión de que un cinturón con un pequeño número de actuadores separados 90 grados es suficiente realizar un guiado preciso del actor en diferentes direcciones.
3. **Acústica:** utiliza señales sonoras para orientar a los actores. Estas señales pueden ser audibles a todos los actores mediante la instalación de altavoces en el escenario o puede ser individualizado mediante el uso de audífonos o “pinganillos” invisibles para el espectador. Suele utilizarse para transmitir información simple como la entrada en una determinada área del escenario virtual o la cercanía a un objeto, pero también puede ser utilizado para dar información precisa sobre la dirección y distancia en la que se encuentran los objetos sintéticos [117].

2.2.4.2. Mezclador de Vídeo

El mezclador de vídeo es uno de los dispositivos que tienen en común los platos virtuales y los tradicionales. Permite al realizador elegir la señal que será emitida de entre todas las generadas en el plató. Diferentes tipos de señales son, por ejemplo, la grabada directamente por las cámaras, las generadas por los ordenadores de *render*, la proveniente de un reproductor de vídeo mostrando una grabación previa y, sobre todo, la composición de los mundos virtual

y real procedente de los *chroma keyers*. Cada una de estas señales de vídeo está asociada a un canal de la mesa, que a su vez se activa mediante un botón único, lo que le permite al operador seleccionar en cada momento que imagen verán los espectadores.

Asimismo, también suelen incluir una serie de efectos de transición que permiten pasar de una señal de vídeo a otra mediante fundidos a negro, movimientos de imagen, cortinillas, etc.

2.2.4.3. Mesa de sonido

La mesa de sonido es el dispositivo encargado de controlar todo el audio del plató. Esto implica, como mínimo, dos líneas en paralelo: una para la comunicación interna entre la sala de control y el escenario (que permita a los realizadores comunicarse con los cámaras, enviar instrucciones y a estos responderles) y otra para el audio de la emisión. Esta segunda línea recibe el sonido tanto de los micrófonos de los actores como de los ordenadores de *render* (ya que estos pueden reproducir clips de música) y la envían a emisión junto con la imagen procesada. Esta sería la configuración básica, pero podrían existir más fuentes de entrada (como varios reproductores) y más fuentes de salida (en caso de haber altavoces en el escenario).

Otro punto a tener en cuenta a la hora de configurar el sonido de un plató virtual de televisión es el *lag* que se ha de introducir al audio para que este coincida con la imagen de emisión. Al estar la señal de vídeo retardada por el *chroma keyer* por razones de sincronía, se debe de retrasar en la misma medida el sonido para que este se corresponda con el instante de tiempo correcto. La no coincidencia de las labiales (o movimiento sincrónico de los labios con la cadencia de las palabras pronunciadas) de los actores con la imagen que se está emitiendo produce un efecto de incomodidad en el espectador, afectando también a la sensación de presencia y realismo de la escena presentada.

2.3. Análisis de casos de uso reales: noches electorales

La mejor forma de analizar el estado actual de la tecnología, sus carencias, fortalezas y el uso que en la actualidad se hace de ella, consiste en observar los programas emitidos en los últimos tiempos que la han utilizado de un modo u otro. De entre los diferentes tipos de emisiones en las que es habitual el uso de plató virtuales de televisión destacan las noches electorales. Estas son idóneas para el uso de los estudios virtuales, ya que se trata de emisiones en directo en las que se ha de mostrar una gran cantidad de información que va llegando al

plató en tiempo real. Los grafismos generados por ordenador y actualizados en tiempo real han sido ampliamente utilizados en televisión y han ido ganando, en los últimos años, una mayor sensación de presencia gracias a la sensorización de las cámaras y a la interacción que los presentadores simulan o realizan con los mismos. A continuación se presenta el análisis de una serie de programas especiales emitidos en noches electorales en Reino Unido y España, dos de los países donde el uso de esta tecnología está más extendido.

2.3.1. Elecciones en Reino Unido: 2014, 2015 y 2016

Históricamente uno de los países que ha demostrado una apuesta más decidida por los estudios virtuales de televisión ha sido el Reino Unido y, más concretamente, la BBC. Este canal es una referencia a nivel mundial en el uso de estas tecnologías, y uno de los momentos en que esta cobra una mayor dimensión es en las noches electorales. A continuación se analiza su uso en los programas emitidos durante el recuento del voto de tres de los últimos procesos electorales de las islas: las elecciones locales y europeas de 2014, las generales de 2015 y el referéndum de independencia Escocés de 2016. Se comentan las tecnologías y efectos utilizados así como las mejoras que han ido apareciendo en el tiempo. Cabe destacar que, actualmente, la BBC confía en los productos de la empresa VIZrt [185], aunque anteriormente ha colaborado con otras compañías como Brainstorm Multimedia o incluso ha desarrollado tecnologías propias.

Las tecnologías empleadas en las diferentes noches electorales fueron siempre las mismas, lo que refleja tanto su fiabilidad como su buen comportamiento en emisiones en directo. En los tres casos el plató era completamente virtual, sin ningún elemento real más que el presentador, lo que permitió la utilización de efectos visuales que de otra forma hubiesen sido mucho más complejos debido a la necesidad de integrarlos con la escenografía real (figura 2.15). La sensorización de las cámaras se realizó mediante un sistema óptico por infrarrojos *inside-out* de la empresa Mo-Sys [113] en el que los patrones fijos se situaron en el techo. De esta forma las cámaras se podían mover libremente a lo largo de todo el plató sin perder su posición y orientación en ningún momento. Con la finalidad de orientar al presentador se utilizaron una serie de proyecciones sobre el croma que le permitían recibir información de en qué lugar se situaban los diferentes elementos generados por ordenador para facilitar una apariencia de interacción con el entorno. Sin embargo, a pesar de toda esta tecnología el control de las oclusiones entre el presentador y los objetos virtuales se realizó de forma manual, siendo un operador el encargado de indicarle al sistema cuando el actor se situaba delante y cuando



Figura 2.15: Imagen del programa especial de una noche electoral en la BBC

detrás de los mismos. Asimismo, las interacciones del presentador con el entorno se realizaron mediante un mando que este portaba en su mano restándole naturalidad a los movimientos de su brazo y mano derecha (figura 2.15). Este efecto destacó aún más frente a la comodidad y naturalidad con la que se movía el resto del cuerpo en un entorno con el que, de forma evidente, se observaba que estaba muy familiarizado.

Mención aparte merece el programa emitido en la noche electoral del referéndum por la independencia de Escocia en 2014. La utilización del sistema de tracking *inside-out* permitió, debido a su gran escalabilidad en cuanto a volumen de captura, la sensorización del interior de un edificio de varias plantas, permitiendo la combinación efectiva de elementos virtuales y reales en un enorme espacio. Esto tuvo como resultado la aparición y desplazamiento del presentador por distintas zonas y plantas del plató lo que, junto con una acertada iluminación, permitió una perfecta integración de los gráficos generados por ordenador en el entorno real 2.16.

2.3.2. Elecciones en España: autonómicas 2015, Generales 2015 y 2016

En los dos últimos años se han sucedido tres procesos electorales en España: las elecciones autonómicas y municipales de mayo de 2015 y las generales de diciembre de 2015 y junio de 2016. En todas ellas ha tenido una importancia capital en diferentes cadenas de televisión la utilización de platós virtuales de televisión, en concreto de las soluciones ofrecidas por la



Figura 2.16: Imagen del programa especial de la noche electoral sobre el referéndum escocés en la BBC

multinacional española Brainstorm Multimedia [144]. A lo largo de las tres noches electorales diferentes cadenas de ámbito estatal como Antena 3, 13 TV o La 1 así como televisiones autonómicas como Televisión Canaria han hecho uso de su tecnología. En el caso de TV3 se utilizaron tecnologías de la empresa ORAD [11]. A continuación se describen, de un modo pormenorizado, las técnicas utilizadas en cada caso.

2.3.2.1. Radio Televisión Canaria

En el caso de las elecciones autonómicas destaca la cobertura realizada por RTVC (Radio Televisión Canaria). En este caso se utilizó un plató completamente virtual, sin ningún tipo de elemento real a parte de un presentador (figura 2.17). Se hizo uso de diferentes cámaras pero sin movimiento real de las mismas (el movimiento es simulado mediante movimientos de las cámaras virtuales y el posicionamiento del presentador como un *billboard*⁸ en la escena). A pesar de la habilidad del presentador para moverse en el entorno virtual y simular una interacción con el mismo, esta no fue real ya que sus acciones no tienen ningún tipo de efecto sobre los elementos generados por ordenador. Del mismo modo la posición relativa del presentador y los diferentes elementos virtuales no tuvo ninguna variación durante toda la emisión. Asimismo, el presentador no manipuló ningún tipo de elemento virtual y se limitó a tener en las manos una serie de notas orientativas.

⁸Un *billboard* es un plano bidimensional, situado en un espacio tridimensional, que rota para mantenerse siempre mirando a la cámara.

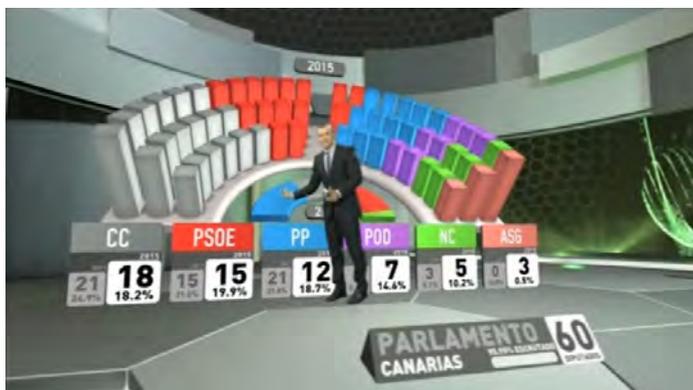


Figura 2.17: Imagen del programa especial de la noche electoral en RTVC

2.3.2.2. Antena 3

En este mismo proceso electoral, el programa especial emitido en Antena 3 combinó elementos reales del plató de informativos con grafismos digitales con los que se relacionaban los presentadores. A diferencia del caso anterior, la presencia de dos presentadores, así como su falta de experiencia con la tecnología provocó que la integración final no se limitase a aportar una mayor espectacularidad o atractivo visual, si no que supusiese un hándicap en el desarrollo del programa. En términos de las tecnologías implicadas se hizo uso de una cámara sensorizada mediante el sistema ncam [172], lo que permitió su movimiento real en directo. Sin embargo, no existió ningún tipo de interacción de los presentadores con el medio ni tampoco se controló la posición relativa de los mismos con respecto a los elementos virtuales que iban apareciendo en plató. Por este último motivo, en diferentes momentos del programa, se dio el caso de que, estando los presentadores físicamente más adelantados que los elementos sintéticos, estos aparecían por delante, afectando drásticamente a la sensación de presencia de los mismos en el entorno real (figura 2.18). En lo que corresponde a las referencias de las que disponían los presentadores para orientarse respecto a los objetos virtuales existía un círculo visible en el suelo que delimitaba la zona en la que aparecían los gráficos, no siendo suficiente esta información para un correcto comportamiento de los presentadores ya que este espacio era continuamente invadido por los mismos.



Figura 2.18: Imagen del programa especial de la noche electoral en Antena 3 en la que se aprecian errores en el posicionamiento relativo de los presentadores y los objetos virtuales

2.3.2.3. La 1

En el caso de las elecciones generales de 2016 en La 1 de Televisión Española se incorporaron nuevas tecnologías como la detección de gestos mediante la tecnología propietaria de Brainstorm Multimedia TrackFree integrada en su software Infinity Set [142], el seguimiento de cámaras en exteriores utilizando sensores ncam y el movimiento de cámaras virtuales en directo en el interior de modelos 3D de gran tamaño. La detección de gestos de la presentadora permite una interacción efectiva y visible entre el mundo virtual y el real, difuminando la línea que los separa. Por otro lado, la sensorización de cámaras en exteriores permite la utilización de elementos virtuales sobre el mundo real (en este caso el edificio del Congreso de los Diputados) y la utilización de movimientos de cámara para dar una mayor sensación de presencia de los mismos. Asimismo el paseo virtual por un modelo 3D del interior del edificio permite una integración más natural del escenario final (el hemiciclo en el que se sientan los diputados) donde se desarrolla el grueso del programa. Sin embargo no existió ningún control sobre la posición relativa de los objetos virtuales y la presentadora, permaneciendo esta invariable durante todo el programa. La acción de un actor rodeando un objeto virtual hace que este cobre importancia en escena y se integre de una forma más efectiva en su entorno y podría haber dado un mayor peso a la mesa situada durante toda la emisión en primer plano (figura 2.19). Del mismo modo, la iluminación de la presentadora no fue coherente con la del modelo 3D, afectando a la sensación de presencia de la misma en el hemiciclo.



Figura 2.19: Imagen del programa especial de la noche electoral en La 1

2.3.2.4. TV3

Un caso en el que se utilizaron tecnologías de una empresa distinta a Brainstorm Multimedia fue el seguimiento que se realizó en la televisión catalana TV3 de las elecciones municipales de 2015. En este caso, la solución empleada fue la ofrecida por otra de las grandes multinacionales del sector: Orad, comprada meses después por AVID. En el programa emitido en TV3 se hizo uso del movimiento en tiempo real de cámara a través de una sensorización mecánica basada en el sistema Stype Grip [169], un sistema de sensorización de grúas que permitió la integración efectiva de grafismos en exteriores (figura 2.20). Sin embargo, no se realizó ningún tipo de interacción real por parte de los presentadores con el entorno virtual ni se controló la posición de los mismos con respecto a los elementos sintéticos.

2.3.3. Conclusiones

A partir de los casos de uso expuestos puede observarse, en líneas generales, que las diferentes soluciones que ofrece la industria no presentan grandes diferencias en el resultado visual obtenido. Todas hacen uso de la arquitectura tradicional de los platós virtuales de televisión (presentada en el siguiente capítulo de esta tesis), con un único sistema de *tracking* encargado exclusivamente del posicionamiento de las cámaras. Cabe destacar el uso de distintos sensores de terceros y la compatibilidad que ofrecen los productos utilizados con los dispositivos de las principales compañías del sector.

Sin embargo, es un denominador común la falta de interacción real de los actores con su



Figura 2.20: Imagen del programa especial de la noche electoral en TV3

entorno virtual. La mayor parte de la relación que estos mantienen con el mundo sintético se basa en coreografías ensayadas, que cuando no se han practicado lo suficiente (como en el caso de Antena 3), producen incongruencias visuales o comportamientos artificiosos por parte de los presentadores. Otra opción común es la utilización de mandos a distancia que modifican el comportamiento natural del actor y distraen la atención del telespectador. Únicamente en el caso de las elecciones generales de 2016 en España, Brainstorm Multimedia introduce en la emisión de La 1 la detección de un gesto simple (*swipe* vertical). Un hecho tan simple como este permite a la presentadora tener las manos libres durante toda la emisión, pero pudiendo controlar el movimiento de un carrusel de resultados mientras ella expone los datos. La emisión gana de este modo en espectacularidad, capacidad de transmisión de información y en naturalidad en el comportamiento de la actriz.

Otro de los tipos de interacción de los que apenas se hace uso es el cambio de posición relativa de los objetos virtuales y el presentador. El hecho de que un actor rodee un objeto virtual hace que este gane peso y presencia en la escena; sin embargo, este efecto no suele utilizarse ya que no acostumbra a estar automatizado y requiere la dedicación casi exclusiva de un operario y su coordinación con el actor para realizar el cambio.

Finalmente, desde un punto de vista exclusivamente visual, destaca el paseo por el Congreso de los Diputados y el entorno en el que se desarrolla todo el programa (el hemiciclo) de las elecciones Generales de 2016 en La 1. Son modelos detallados y su único defecto es la iluminación, que no es coherente con la que recibe la presentadora, denotando así su no presencia en el escenario. En contraposición, cabe destacar el programa especial sobre el

referéndum escocés, en el que una iluminación inteligente del escenario real permitía una perfecta integración de los elementos virtuales. En general, se utilizan modelos simples con una iluminación plana, lo que facilita su *renderizado* en tiempo real, pero resta credibilidad a la composición final.

CAPÍTULO 3

RETOS DE LOS PLATÓS VIRTUALES DE TELEVISIÓN Y MARCO DE TRABAJO

3.1. Retos de los platós virtuales de televisión

En el 1998, A. Wojdala enumeraba los retos a los que deberían enfrentarse los platós virtuales de televisión en los siguientes años [192]. Dieciocho años después, y a pesar de la gran evolución que ha experimentado la tecnología, la mayor parte de ellos sigue sin tener una respuesta definitiva. Entre estos retos destacan:

1. **Realismo del *render* frente a tiempo real de emisión:** a pesar del gran avance en las técnicas de visualización avanzada, la obtención de imágenes fotorrealistas en tiempo real sigue siendo, a día de hoy, un reto. Equiparar los resultados obtenidos por producción audiovisual en directo a los alcanzados en la postproducción cinematográfica, continúa siendo, a día de hoy, el objetivo principal de los platós virtuales de televisión en este campo, pero sigue lejos de ser una realidad debido al alto coste computacional de los algoritmos que han de ejecutarse.
2. **La correcta alineación de las cámaras reales y virtuales:** al tratarse de un elemento imprescindible para la obtención de una correcta composición es, también, uno de los que más ha evolucionado. En los platós virtuales de televisión de hoy en día es posible seguir el movimiento libre de un número elevado de cámaras de una forma precisa y en tiempo real. Sin embargo, los sistemas de *tracking* capaces de realizar esta tarea tienen un precio elevado y sólo están al alcance de unas pocas cadenas de televisión. Por eso

es muy común que, en lugar de confiar en sistemas ópticos más flexibles, se siga confiando en sistemas mecánicos más limitados, pero al mismo tiempo más económicos, renunciando al movimiento libre de cámaras.

3. **Coherencia lumínica:** para obtener una sensación de presencia efectiva de los elementos reales en el mundo virtual, ambos entornos han de tener una iluminación coherente. Generalmente, el procedimiento seguido es el de imitar en el mundo real la configuración de iluminación del entorno sintético, buscando siempre iluminaciones planas y simples. Este modo de operar viene dado por la dificultad de generar efectos de luz realistas (como por ejemplo sombras) en tiempo real. El objetivo es la simulación realista en la escena sintética de la iluminación de los elementos reales presentes en el escenario.
4. **Oclusiones entre el mundo real y el virtual o *distance keying*:** consiste en realizar un seguimiento de la posición relativa de los elementos reales respecto a los virtuales. Al tratarse de un entorno tridimensional, los objetos sintéticos pueden estar situados delante o detrás de los objetos y actores presentes en el escenario. Realizando un *tracking* de los elementos de ambos mundos pueden controlarse las oclusiones generadas. Sin embargo, el funcionamiento por capas del *chroma keyer* hace que la posición relativa de un objeto sintético deba de ser la misma con respecto a todos los elementos reales. Así, dos actores (o dos partes del cuerpo de un mismo actor) no podrán estar uno delante y otro detrás de un mismo objeto virtual, ya que forman parte de la misma capa de imagen. A día de hoy, en la mayor parte de los platós virtuales de televisión, esta operación se sigue realizando de forma manual por parte de un operador.
5. **Interacción con los elementos virtuales:** como se ha demostrado previamente mediante el análisis de casos de uso reales, la interacción entre los actores y el mundo sintético sigue siendo uno de los mayores retos no resueltos en los platós virtuales de televisión. Generalmente se opta por coreografías largamente ensayadas que simulan interactividad, si bien restan naturalidad al comportamiento de los presentadores y que son susceptibles a fallos de sincronización [90]. También es común el uso de mandos a distancia desde los que el presentador controla los efectos visuales del plató, con resultados igualmente nocivos en lo que respecta a la naturalidad de su comportamiento.

El hecho de que estos retos, planteados hace casi dos décadas, no estén completamente resueltos a día de hoy patentiza su complejidad. Sin embargo existen, a día de hoy, tecnologías

que podrían permitir afrontar y avanzar en la solución de algunos de ellos. Los diferentes sensores que han aparecido en el mercado permiten desde una correcta alineación de cámaras hasta diferentes tipos de interacción por parte de los presentadores. Los algoritmos de visualización avanzada y el preprocesado de modelos 3D permiten aportar un mayor realismo a los entornos sintéticos, y la captura y utilización de iluminaciones reales hace posible mejorar la coherencia entre los mundos virtual y real. Es necesario, por tanto, dotar a los platós virtuales de televisión de la capacidad de incorporar y utilizar de forma efectiva estas técnicas.

3.2. Marco de trabajo

3.2.1. Introducción

Como se ha descrito previamente, los platós virtuales de televisión han evolucionado técnicamente de una forma notable desde sus primeros prototipos aparecidos en los noventa. Sin embargo, su arquitectura hardware y software se mantiene prácticamente intacta a día de hoy. Las grandes multinacionales que lideran los avances en la tecnología centran sus esfuerzos en mejorar puntos específicos de la misma (como la proyección de sombras, el *tracking* de cámaras en exteriores, etc.), dejando de lado cambios más profundos en la configuración de los platós.

Generalmente, estas compañías desarrollan el motor gráfico encargado de generar el *render* del mundo virtual y utilizan productos de terceros para el resto de tareas, como sensorización, *chroma keying*, etc. Esto las lleva a confiar en un único sistema de *tracking* centrado en el seguimiento de las cámaras dejando de lado la interactividad. En la propia arquitectura tradicional (figura 2.5) únicamente se contemplan los sensores asociados a las cámaras, y no se tiene en cuenta el *tracking* de otros elementos. Como resultado, se ha eliminado casi por completo la interactividad real en los platós virtuales de televisión o esta se ha reducido al manejo de mandos a distancia. Además, al hacer uso de hardware (los sensores y otros dispositivos) como de software (el motor gráfico) propietarios, existen numerosos problemas de compatibilidad a la hora de utilizar dispositivos novedosos. Si se desea combinar un motor gráfico con un sensor no soportado por el mismo han de realizarse numerosos ajustes, implementar *plugins* y añadir *drivers*, lo que resulta costoso tanto en términos económicos como temporales.

En esta tesis se plantea la modificación de la arquitectura tradicional de los platós virtuales de televisión, separando el subsistema de sensorización y vídeo en dos subsistemas para per-

mitir la utilización sinérgica de distintos de dispositivos. Esta nueva arquitectura actualiza a la tradicional aumentando su escalabilidad y flexibilidad, y facilitando así la incorporación, actualización y eliminación de dispositivos. Esto permitirá mejorar las capacidades de los platós virtuales de televisión mediante el movimiento libre de cámaras y una mayor interactividad en tres direcciones: manipulación directa de objetos virtuales, *distance keying* y reconocimiento de gestos. Se plantea también la utilización de técnicas de visualización avanzada de cara a una mejor integración entre los elementos sintéticos y reales, simulando la iluminación del escenario en entornos virtuales. Finalmente, es necesaria la creación de nuevos flujos de trabajo que permitan agilizar y simplificar los procesos necesarios para obtener una producción audiovisual en platós virtuales de televisión con emisiones en directo de una calidad profesional.

Este trabajo de investigación se ha realizado sobre un plató tradicional que se presenta a continuación.

3.2.2. Plató virtual de trabajo

El plató virtual de televisión en el que se ha desarrollado gran parte de esta tesis está situado en la Facultad de Ciencias de la Comunicación de la Universidad de Santiago de Compostela. Pese a ser diseñado y construido en 2011 su arquitectura es exactamente la misma que la presentada en 1998 por Gibbs et al. [60] (figura 3.1).

Los componentes de este plató son los siguientes:

- Un escenario de 3.3 metros de profundidad por 5.1 metros de ancho y 3 metros de altura de color verde construido íntegramente en madera.
- Dos cámaras Sony XDCAM EX [28] capaces de grabar contenido con resolución FHD 1920x1080 *pixels*. Disponen también de un objetivo gran angular Fujinon con zoom 14x que las hace ideales, por su amplio ángulo de visión, para platós con un espacio reducido. Disponen también de estabilizador óptico de imagen, autoenfoco y enfoque manual.
- Dos estaciones de trabajo HP Z800 [77] para realizar las tareas de *render*, cada una asociada a una cámara. Estos equipos cuentan con un procesador Intel Xeon x5660 a 2,8 GHz, 4 Gb de RAM y una tarjeta gráfica Nvidia Quadro 5800.

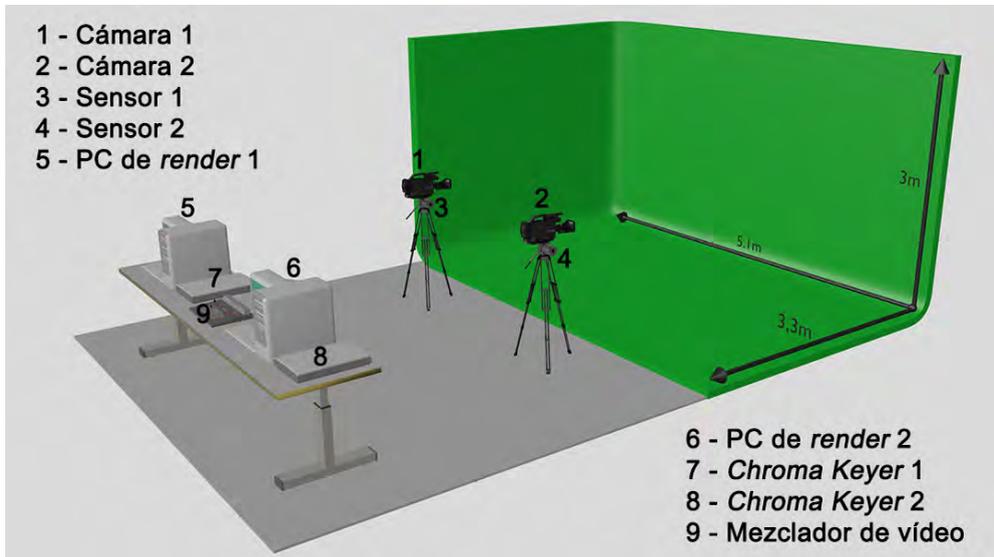


Figura 3.1: Diseño del plató virtual de televisión de la Facultad de Ciencias de la Comunicación de la Universidad de Santiago de Compostela

- Dos *chroma keyers* FOR-A MBP-100CK [108] cada uno asociado a una pareja Sony XDCAM EX-HP Z800. Se controlan desde una unidad de control MBP-RUA [109] que permite configurar los parámetros más importantes como el ajuste del valor del *chroma key*, el *lag* a introducir, etc.
- Sensores mecánicos de la empresa Vinten- Radamec [115] capaces de realizar un seguimiento del *pan*, *tilt*, *zoom* y foco de las cámaras.
- Un mezclador de vídeo FOR-A HVS-300HS [51] con hasta 4 entradas y 4 salidas de vídeo. Permite también el escalado de las señales, la superposición de vídeos y la división de la pantalla en diferentes áreas, una para cada señal de salida.
- Un sistema Nitris DX [12] para la grabación y edición de la salida de vídeo del plató.
- Una mesa de sonido Behringer Xenyx X2222USB [15] con 22 canales de entrada, interfaz USB y un procesador de efectos interno de 24 bits y 16 *presets*.

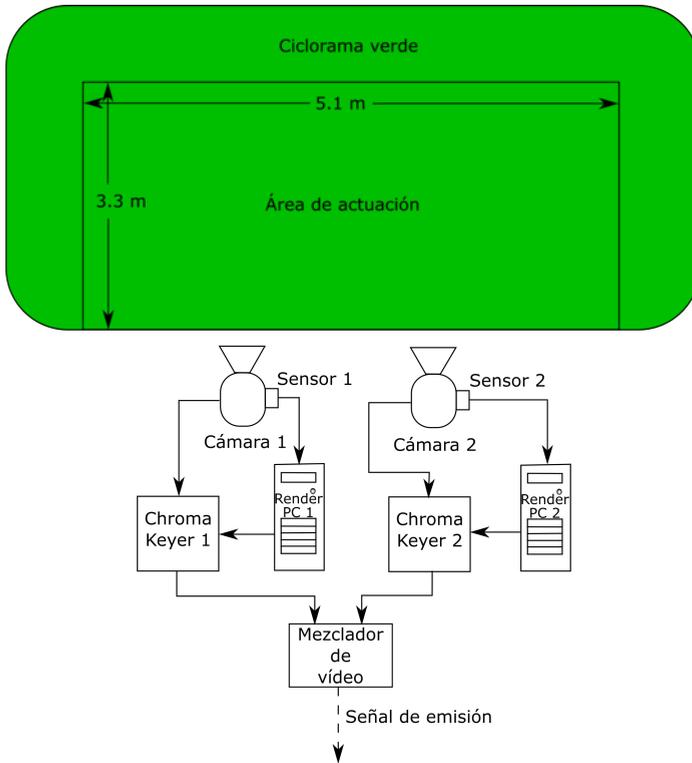


Figura 3.2: Esquema de conexiones del plató

Como se puede observar en la figura 3.2 la configuración del plató es exactamente la misma que la presentada en la figura 2.5 a pesar de estar diseñado 13 años más tarde. Este diseño plantea una serie de limitaciones:

1. Los únicos sensores presentes están asociados al movimiento de las cámaras, lo que elimina cualquier tipo de posibilidad de interacción real entre los actores y el mundo virtual.
2. La sensorización de las cámaras es mecánica y por tanto no permite el desplazamiento libre. Los sensores se sitúan sobre un trípode fijo que cada vez que es desplazado requiere de una nueva configuración en el motor gráfico.
3. Los sensores están directamente conectados a los equipos de *render*, por lo que una

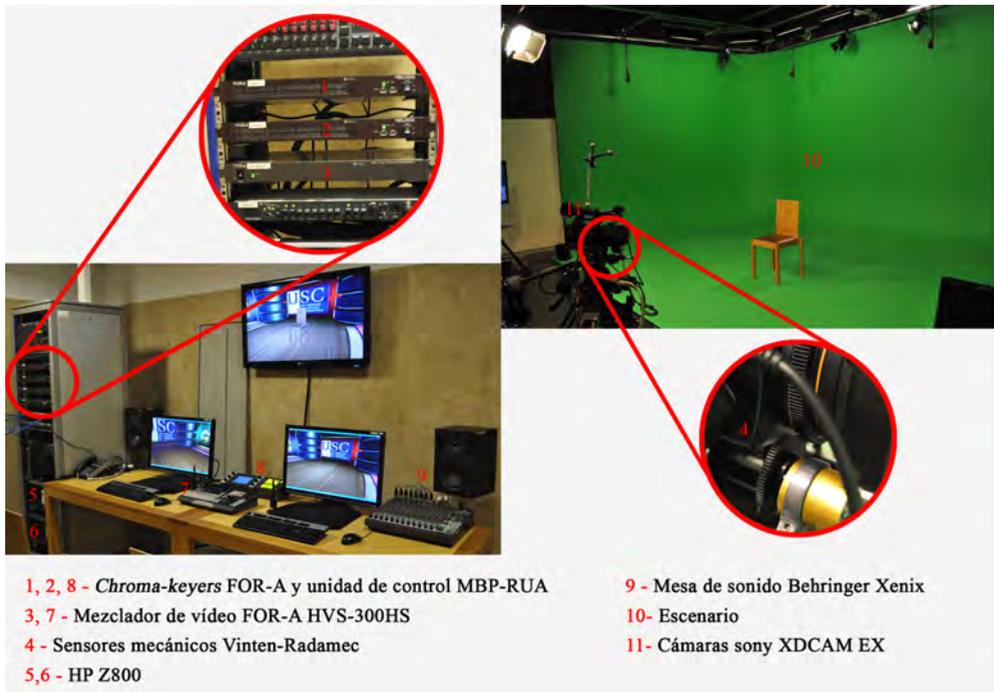


Figura 3.3: Componentes del plató de trabajo

modificación de los mismos requeriría hacer cambios profundos en la configuración del motor gráfico, especialmente en caso de existir la necesidad de incluir un sensor no soportado de forma nativa por el software.

En cuanto al software utilizado para el *render*, se trata de eStudio [143], de la multinacional española Brainstorm Multimedia.

3.2.3. Brainstorm Multimedia

Brainstorm Multimedia es una empresa que nació en Valencia en 1993 como una compañía de soluciones gráficas 3D en tiempo real. Aunque ofrece servicios para publicidad y cine, su principal mercado son los platós virtuales de televisión para emisiones en directo. Su desarrollo en estos años la ha situado como una de las soluciones punteras a nivel mundial y

prueba de ello son algunos de sus clientes como Antena 3, RTVE, NHK, CNBC, REUTERS, BBC, Al Jazeera, Jordan TV, etc.

Ofrece múltiples soluciones en el mercado de los platós virtuales de televisión, adaptándose a las diferentes necesidades que han ido apareciendo en el sector. Inicialmente ofrecía únicamente su producto estrella, eStudio, un programa que unifica todas las herramientas necesarias para generar y manipular en tiempo real todos los elementos 3D de un plató. Sin embargo, con el paso del tiempo, ha creado nuevas soluciones específicas con menos capacidades pero, al mismo tiempo, más económicas. Un ejemplo es Aston, un sistema de creación de grafismos que permite al operador crear, manipular, animar y realizar cambios de última hora en cualquier tipo de texto, información o gráfico de un programa, incluso cuando este se está emitiendo en directo. Aston Elections es otra muestra de esta especialización, al igual que BrainNews, paquetes específicamente diseñados para noches electorales e informativos.

Su última creación es Infinity Set, un conjunto de herramientas que simulan un plató virtual de televisión al uso, pero en el que, generalmente, la cámara real no se desplaza. Esto se debe a que el sistema convierte al presentador en un elemento 3D más. El actor pasa a ser un plano en el mundo 3D que siempre está mirando a la cámara virtual (un (billboard)), y de esta forma, realizando movimientos de cámara, debidamente diseñados con ese fin, se puede simular el comportamiento habitual de un plató virtual de televisión. Este software ofrece proyecciones de sombras, integración con escenarios reales, etc. pero no permite el libre movimiento del presentador en el plató y la existencia de escenografía puede suponer un problema en algunos casos. Es una solución pensada para ahorrar costes simplificando alguna de las tareas más complejas a llevar a cabo en un estudio virtual.

Esta tesis se ha desarrollado sobre el paquete de software más completo, eStudio (figura 3.4), en una colaboración continua con Brainstorm Multimedia. Su aportación al presente trabajo no se ha limitado a una mera labor de soporte, si no que han acompañado el desarrollo apoyándolo con la asistencia gratuita del doctorando a dos cursos de formación, la cesión gratuita de tres licencias durante los años que ha durado la investigación y el forjado de una sólida colaboración con la Universidad de Santiago de Compostela que ha tenido como último resultado la cesión gratuita desde 2013 de 21 licencias del software eStudio para la capacitación de los alumnos de la Facultad de Ciencias de la Comunicación en escenografía virtual.

A continuación se presentan las principales características de eStudio que han sido utilizadas en el desarrollo de esta tesis.

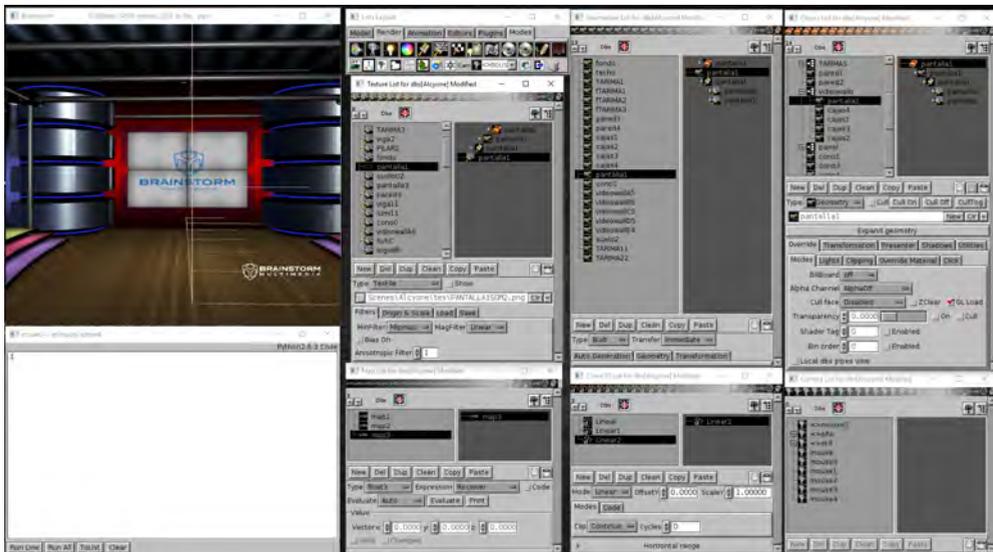


Figura 3.4: Interfaz de usuario de Brainstorm eStudio

3.2.3.1. Software eStudio

eStudio es un programa que contiene un conjunto de herramientas que ofrecen soluciones a diferentes necesidades que puedan surgir en un plató virtual de televisión. Esta versatilidad viene dada por su arquitectura abierta que permite desde el diseño, creación y manipulación en tiempo real de estudios virtuales y grafismos 3D hasta la creación de *plugins* y aplicaciones personalizadas.

Este software ha sido utilizado tanto para la emisión de programas en directo de noticias, deporte, entretenimiento y meteorología, como para publicidad, grafismos sobreimpresos, presentaciones interactivas y pre-visualización cinematográfica entre otros.

El desarrollo continuo de eStudio durante más de 20 años lo convierte en uno de los entornos más robustos del mercado, probando su fiabilidad desde hace años en complejos entornos de trabajo 24/7 en todo el mundo.

Algunas de sus características principales son:

- Ofrece la posibilidad de manipular objetos 2D y 3D, permitiendo modificar sus características principales como escala, posición, rotación, textura, material, etc.

- Es capaz de realizar *renders* y modelados avanzados. Permite la utilización de *shaders* y efectos visuales complejos como conjuntos de partículas, así como la creación y modificación de diferentes tipos de objetos.
- Posee una extensa librería de *plugins* que permiten desde la importación directa de archivos desde los programas de modelado más comunes, como Maya o 3D Max, hasta utilidades para la calibración de cámaras o importación directa desde Photoshop.
- Permite la conexión con bases de datos externas para la actualización automática de campos en tiempo real.
- Posee una API que permite crear programas compatibles y conectar eStudio con aplicaciones externas. Entre las funcionalidades que ofrece destaca la posibilidad de modificar tipos de datos directamente a través de la red. Estos se denominan *maps* en eStudio y para ser modificados a través de la red ha de ser del tipo *receiver*, lo que implica la creación de un *socket* a través del cual se recibe el valor del mismo. Estas variables pueden ser de varios tipos: *floats*, *enteros*, listas de *floats* o de enteros, *booleans*, etc., Utilizando estos valores es posible modificar a través de la red valores de posición, orientación, escala, etc. de elementos del mundo virtual. La API también ofrece la posibilidad de enviar instrucciones específicas en código *Python* a través de la red, pudiendo, de esta forma, modificar texturas, lanzar efectos, etc., de forma remota.

Sin embargo, a pesar de estas capacidades, el software presenta una serie de carencias comunes a las diferentes soluciones del mercado. La conexión directa de sensores hace necesaria la implementación de *plugins* y *drivers* específicos, lo que dificulta enormemente la inclusión de dispositivos no soportados. Asimismo, no ofrece soluciones de visualización avanzada que permitan la proyección de sombras realistas o la integración de una iluminación global que permita una perfecta composición entre los mundos virtual y real.

CAPÍTULO 4

INTERACCIÓN EN PLATÓS VIRTUALES DE TELEVISIÓN

Una de las claves para poder superar de forma definitiva los retos expuestos por Wojdala [192] anteriormente comentados es la evolución técnica de los dispositivos utilizados. Las grandes compañías (debido al imperativo de máxima fiabilidad de sus herramientas de producción audiovisual, concebidas para un uso en emisiones en directo que excluye la posibilidad de segundas tomas o correcciones, características de la postproducción audiovisual), suelen ser conservadoras y confiar en sistemas propietarios que ofrecen soluciones a una problemática concreta y están ampliamente probados. Por ejemplo, si nos referimos a sistemas de *tracking*, como se ha comentado previamente en el análisis de casos reales, la industria acostumbra a confiar en una única solución comercial para afrontar el seguimiento de las cámaras, dejando de lado otros elementos del plató. Como consecuencia, estos sistemas están dotados de una gran fiabilidad y precisión, pero resultan económicamente costosos y son poco flexibles al tener una finalidad específica y no poder abarcar las diferentes funcionalidades necesarias en un plató virtual de televisión interactivo.

Para superar estas limitaciones se propone el uso combinado de diferentes dispositivos que, trabajando en paralelo de forma sinérgica, permitan superar las limitaciones de los actuales platós virtuales de televisión. Este funcionamiento conjunto permitirá que un dispositivo cubra las carencias de los otros logrando, al combinar las capacidades de todos, solventar las diferentes necesidades técnicas de un plató interactivo. Un caso paradigmático que demuestra la validez de esta propuesta es el de los sensores. En un estudio virtual han de sensorizarse

diferentes elementos (cámaras, objetos y personas) y existen dispositivos especializados en cada una de estas tareas. El uso conjunto de sensores específicos permitirá la utilización de los sistemas de *tracking* más avanzados para cada finalidad concreta, aumentando, de esta forma, la precisión y capacidades del plató.

Asimismo, con la aparición en los últimos años de numerosos sensores de bajo coste, esta solución permitiría la creación de platós virtuales de televisión asequibles para los nuevos mercados (principalmente televisiones a través de internet y docencia), ampliando el alcance comercial de la tecnología.

Los dos principales inconvenientes que se han de afrontar para la implementación de esta solución son, por un lado, la arquitectura tradicional de los platós virtuales de televisión que no contempla el uso de diferentes sensorizaciones y por otro la necesidad de utilización de software propietario para la realización del *render*. La combinación de ambos implica que, para añadir un nuevo dispositivo al plató, ha de realizarse una conexión directa del mismo con el software de *render*. Para que esto sea posible han de utilizarse *plugins* específicos y realizar cambios en la configuración del programa para que este sea capaz de recibir y enviar información al dispositivo. El procesado de esta información, especialmente ante el funcionamiento sinérgico de diversos sensores, podría afectar directamente al rendimiento y llegar incluso a comprometer la obtención de imágenes sintéticas en tiempo real. Además, en caso de que el hardware no estuviese soportado por el sistema la posibilidad de conexión con el mismo dependería de la voluntad de la empresa de darle soporte.

Por tanto, se propone una actualización de la arquitectura tradicional de los platós virtuales de televisión hacia un diseño distribuido y escalable que permita la inclusión, actualización y eliminación ágil e independiente del software de *render* de los diferentes dispositivos involucrados en la grabación de programas en directo.

4.1. Diseño de una nueva arquitectura distribuida escalable para platós virtuales de televisión

El diseño de la nueva arquitectura se basa en la separación del subsistema de sensorización y vídeo en dos: el subsistema de sensorización y el subsistema de vídeo (figura 4.1). De esta forma se consigue, de un modo conceptual, separar los sensores de las cámaras ya que su finalidad será mucho más amplia que el seguimiento de las mismas. Dentro del subsistema de vídeo permanecen, por tanto, únicamente las cámaras que realizan la grabación del escenario.

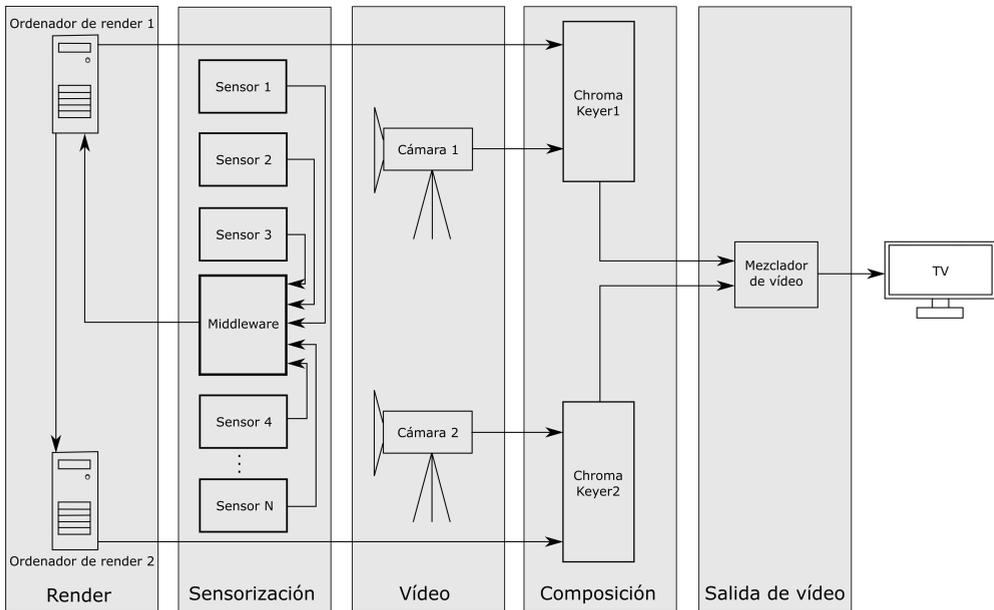


Figura 4.1: Nueva arquitectura distribuida y escalable para un plató virtual de televisión con dos cámaras

Por otro lado, el subsistema de sensorización está compuesto por un número N de sensores conectados a un *middleware* que hace al mismo tiempo de concentrador y unidad de procesamiento y que es el encargado de enviar la información obtenida al subsistema de *render*.

La arquitectura software es una arquitectura cliente-servidor de tres capas [42]. La capa de datos está formada por N servidores de datos conectados cada uno a un sensor. La capa de aplicación es un *middleware* que transforma la información recibida de la capa de datos y la envía al cliente. Finalmente, la capa cliente es la encargada de realizar el proceso de *render* a partir de los datos recibidos (figura 4.2).

El flujo de información desde los sensores hasta el programa de *render* es el siguiente: los sensores capturan información del escenario y la envían a los servidores de datos, que a su vez la remiten a la capa de aplicación. El *middleware* recibe, procesa y transforma la información para adecuarla al formato requerido y la envía a la capa cliente. Finalmente, el cliente modifica el entorno virtual en función de los datos recibidos y realiza el *render* del mismo. Cada servidor de datos está directamente conectado con el *middleware* que a su vez está conectado a un ordenador de *render* principal. En caso de existir más de una instancia del

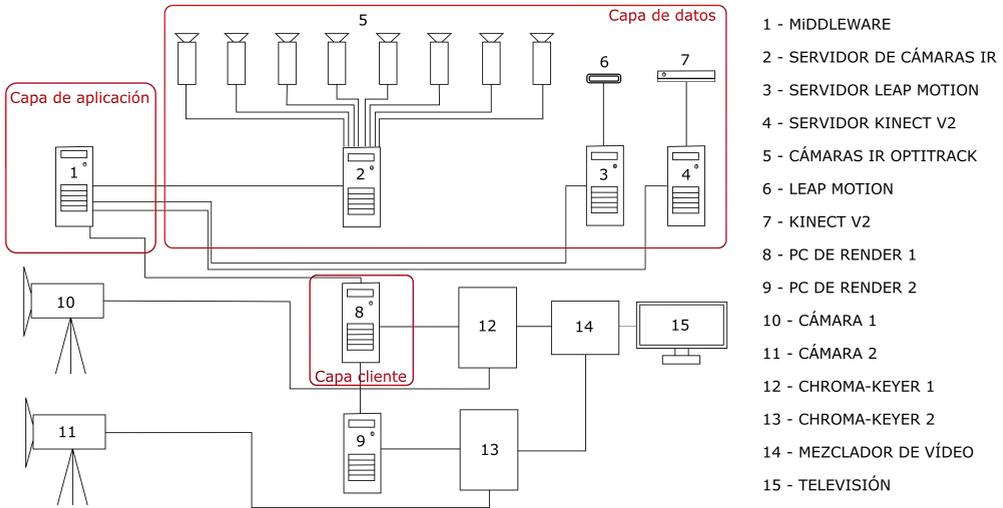


Figura 4.2: Esquema de la arquitectura cliente-servidor de tres capas implementada en el plató

software de *render* la información se distribuye directamente desde la principal a cada una de las secundarias. Tanto la capa cliente como la capa de aplicación se encuentran escuchando en red de forma continua y asíncrona, ya que cada sensor tiene un tiempo de refresco diferente y envía su información en cuanto esta está disponible.

En términos de hardware, la configuración ideal para esta arquitectura es la compuesta por un ordenador por cada servidor de datos (cada uno encargado de controlar su correspondiente sensor), otro para el *middleware* y otro más por cada instancia del software de *render* ejecutada en el plató (generalmente una instancia por cada cámara de vídeo)(figura 4.2). Sin embargo, esta arquitectura es también flexible en términos de hardware, ya que todo el software puede ser ejecutado en una misma máquina siempre y cuando el rendimiento se mantenga en un tiempo real estable.

Como consecuencia directa de su naturaleza distribuida, la arquitectura es independiente del sistema operativo utilizado por los diferentes dispositivos incluidos en la misma al ser estos independientes y estar conectados a través de la red. Esta es una característica de una importancia capital para la finalidad de este diseño, ya que permite la inclusión de todo tipo de hardware independientemente de la plataforma sobre la que ejecute su software. Por ejemplo, el sensor Microsoft Kinect V2 que requiere Windows 8 o superior podría estar enviando información a un software de *render* ejecutado en Windows XP al mismo tiempo que un sensor

Structure [83] conectado a un ipad.

Cabe resaltar que la arquitectura también hace posible la comunicación bidireccional, lo que permite incrementar las capacidades de los platós virtuales de televisión en dos direcciones: la obtención de información del escenario y la posibilidad de actuar en el mundo real en base a información generada por ordenador.

En conclusión, la principal ventaja de este diseño es su flexibilidad, permitiendo añadir un número ilimitado de dispositivos independientemente de la plataforma sobre la que se ejecuten y de forma independiente al software de *render* o hardware propietarios.

4.2. Implementación de una nueva arquitectura distribuida escalable para platós virtuales de televisión

La arquitectura diseñada se ha implementado sobre el plató de la Facultad de Ciencias de la Comunicación presentado previamente. Para demostrar su correcto funcionamiento y su validez para afrontar los retos planteados se han incluido nuevos sensores en el plató original que, funcionando de forma sinérgica, permiten realizar tanto el seguimiento de cámaras, objetos y presentadores como la detección de gestos de los actores. Esto incrementa las capacidades del plató y le permite automatizar procesos comunes en una producción profesional para televisión.

A la hora de realizar la implementación de la arquitectura, el primer paso ha sido la implementación del *middleware*, ya que es el punto central de la misma y donde se concentra toda la información recibida de los sensores.

4.2.1. Implementación del *middleware*

La capa de aplicación ha sido implementada utilizando .NET [133] (y por tanto el lenguaje c# [132]) y WPF (Windows Presentation Foundation) [135] para la parte de la interfaz gráfica. Tiene, principalmente, tres tareas: conectarse y recibir datos de los sensores, procesar la información recibida y enviar esta información al software de render, en este caso Brains-torm eStudio.

La conexión con los sensores ha sido implementada utilizando la librería VRPN (*Virtual Reality Peripheral Network*), un sistema de acceso a periféricos independiente del dispositivo y transparente para la red que se ha convertido en un estándar “de facto” en el mundo de la realidad virtual [80]. Esta librería permite la creación de servidores que codifican la infor-

mación recibida de los dispositivos y la envían a través de la red a clientes que facilitan el acceso directo y ordenado a los datos generados por los periféricos. VRPN soporta de forma nativa multitud de dispositivos como el sensor Advanced Realtime Tracking Flystick 2 [179], el InterSense IS-900 [173], el sensor Motion Node [194] o las librerías ARTToolkit [38]. La lista al completo puede consultarse en su página web [61]. Sin embargo, cualquier dispositivo puede ser soportado por VRPN gracias a una capa de abstracción que hace que todos los periféricos del mismo tipo sean accesibles de la misma forma. Por ejemplo, todos los dispositivos de *tracking* serán del tipo `vrpn_tracker` ya que todos capturan información semejante (posición, orientación, velocidad, etc.). Las abstracciones que ofrece la librería son:

- **Analog:** permite transmitir uno o más valores analógicos.
- **Button:** permite transmitir los eventos de pulsar y soltar uno o más botones del dispositivo. Además de botones físicos pueden ser utilizados para cualquier tipo de evento, como, por ejemplo, la detección de un gesto.
- **Dial:** permite transmitir rotaciones incrementales.
- **ForceDevice:** permite transmitir superficies y campos de fuerza.
- **Imager:** permite la transmisión de imágenes.
- **Sound:** permite la transmisión de sonido.
- **Text:** permite la transmisión de mensajes de texto.
- **Tracker:** permite la transmisión de la posición, orientación, velocidad y aceleración de un elemento en el espacio.

Cada dispositivo puede hacer uso de una o más de estas abstracciones dependiendo de los datos que capture. El uso de VRPN hace posible la conexión de sensores, a través de servidores de terceros, al *middleware* de forma directa. Esto implica que los servidores de datos solo han de ser implementados desde cero cuando no exista una solución en el mercado o cuando ésta no permita transmitir todos los datos necesarios para el uso del dispositivo en un plató virtual de televisión. Por ejemplo, un sensor puede realizar el seguimiento de las manos de un actor y detectar sus gestos y su servidor VRPN podría limitarse a transmitir únicamente la posición detectada, siendo necesaria la implementación de un nuevo servidor que incluya la transmisión la información gestual. Por otro lado, la extensión del uso de VRPN implica que

numerosos dispositivos disponen desde su lanzamiento de un servidor específico facilitando, de esta forma, su integración en la arquitectura presentada.

A pesar de esta capacidad para la incorporación a la arquitectura de servidores de terceros, el *middleware* ha de implementar uno o más clientes específicos para cada uno de los servidores de datos, lo que podría afectar a la agilidad del proceso de inclusión de nuevos dispositivos. Por ello se propone un esquema XML que permita la configuración de una serie de clientes VRPN por usuarios no expertos a través de una interfaz gráfica implementada ad hoc.

Este esquema describe las características de los sensores así como la información que capturan, permitiéndole al sistema automatizar la creación de clientes VRPN específicos. Para formalizar el desarrollo de esta herramienta y dada la amplia variedad de sensores existentes en el mercado, se realizó un estudio de las diferentes configuraciones presentes en los servidores VRPN de terceros, diferenciando, finalmente, dos tipos de sensores (figura 4.3):

- **MultiReceiverSensor:** en este caso cada elemento (*tracker*) que es capaz de detectar el sensor tiene un servidor específico encargado de transmitir, a través de la red, los datos del mismo. Por tanto, cada sensor tendrá asociados tantos servidores como elementos sea capaz de seguir. Así mismo, en el *middleware* serán necesarios el mismo número de clientes para recibir la información. Por ejemplo, un sensor que detecte la posición de cinco objetos en el escenario hará uso de cinco servidores y necesitará cinco clientes en la capa de aplicación.
- **MonoReceiverSensor:** esta configuración es la más común y hace uso de un solo servidor para transmitir la información de todos los *trackers* detectados por el sensor, identificando cada uno de ellos con un código numérico. Necesita, por tanto, un único cliente para recibir sus datos en la capa de aplicación.

Para cada uno de estos sensores el archivo de configuración XML almacenará la siguiente información (figura 4.3):

- **Name:** nombre del sensor que servirá para identificarlo en la interfaz gráfica.
- **Origin_Tracker:** nombre del *tracker* asociado a la posición del sensor. Este valor sólo se utiliza cuando el sensor es móvil y, por tanto, debe de estar sensorizado.
- **Position:** posición del sensor. Este valor es únicamente utilizado para sensores localizados en una posición fija.

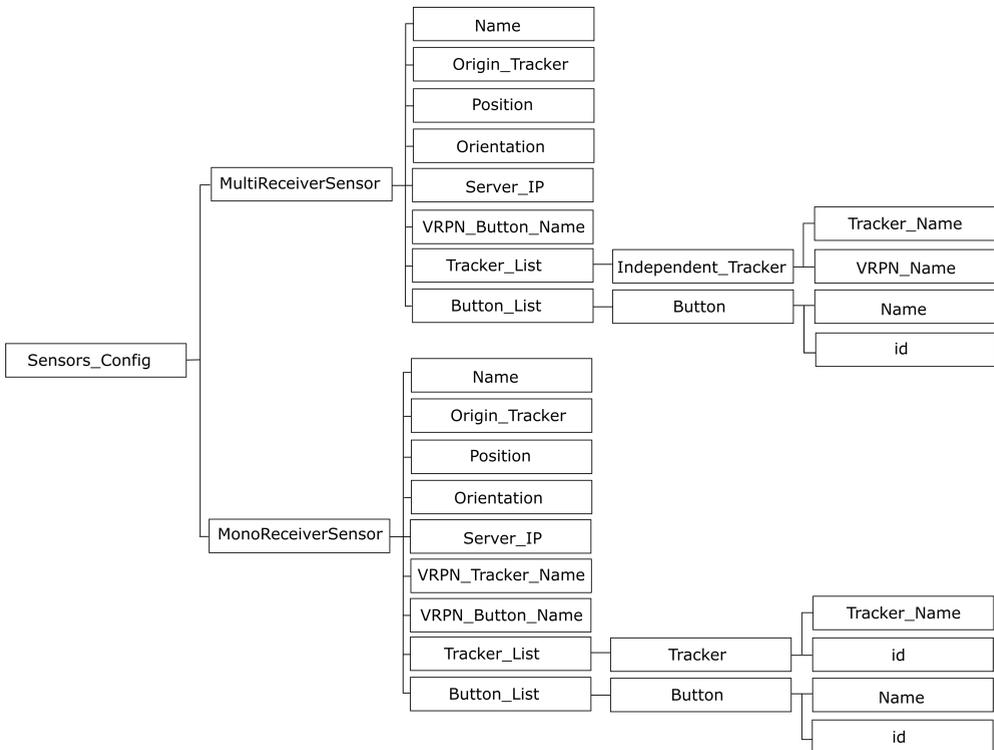


Figura 4.3: Esquema XML utilizado para la conexión de nuevos sensores con el *middleware*

- **Orientation:** orientación del sensor. Al igual que la posición, este valor sólo se utiliza cuando no varía a lo largo de la grabación del programa.
- **Server_IP:** dirección IP del servidor de datos asociado al sensor.
- **VRPN_Tracker_Name:** nombre del servidor de valores de tipo `vrpn.tracker`. Este dato se utiliza para la configuración de los sensores *MonoReceiverSensor* que hacen uso de un único servidor.
- **VRPN_Button_Name:** nombre del servidor de valores de tipo `vrpn.button`.
- **Tracker_List:** lista de *trackers* detectados por el sensor. En el caso de los *MultiReceiverSensors* para cada uno de ellos se guardará el nombre a mostrar en la interfaz gráfica (`Tracker_Name`) y el nombre del servidor de valores de tipo `vrpn.tracker`

(VRPN_Name). Para los *MonoReceiverSensors* se almacenará el nombre a mostrar en la interfaz gráfica (Tracker_Name) y el identificador numérico del *tracker* (id).

- **Button_List:** lista de botones del sensor. Para cada botón se guarda el nombre a mostrar en la interfaz gráfica (Name) y el identificador numérico del botón (id).

A partir de esta información la capa de aplicación es capaz de generar automáticamente clientes VRPN y conectarlos a sus correspondientes servidores de la capa de datos. Esto agiliza enormemente la inclusión de nuevos sensores en el sistema, ya que, en caso de existir un servidor VRPN de terceros para un determinado dispositivo, la conexión del mismo con el *middleware* se reduce a la creación de un archivo de configuración XML a través de una interfaz gráfica implementada ad hoc (figura 4.4). Este proceso, por tanto, se realiza en cuestión de minutos.

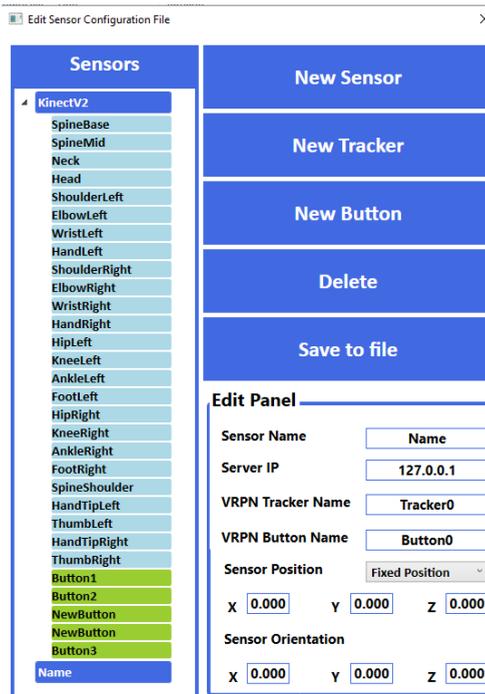


Figura 4.4: Interfaz gráfica implementada para la creación y edición de archivos de configuración XML para la conexión de sensores

Una vez recibida la información capturada por los sensores, la capa de aplicación ha de realizar un procesado de la misma para que esta pueda ser utilizada por el software de *render*. Se implementan tres funcionalidades directamente relacionadas con los retos de los platós virtuales de televisión presentados anteriormente:

1. **El envío de la posición y orientación de un *tracker*.** Permite al software de *render* conocer la posición de ciertos elementos reales y que, por tanto, estos puedan interactuar con el escenario virtual. Sin embargo, puesto que cada dispositivo funciona bajo su propio sistema de coordenadas independiente y que, en general, no coincide con el del plató virtual, es necesario establecer las posiciones relativas de los sensores respecto al origen de coordenadas del plató. Esto permitirá, mediante la composición de ambos sistemas, la obtención de la posición de los elementos detectados por los dispositivos respecto al sistema de referencia del escenario virtual.

Con el fin de realizar esta operación se parte de la posición y orientación del sistema de coordenadas del sensor para generar una matriz de transformación (M en la ecuación 4.1) que codifica su desplazamiento y rotación respecto al mundo virtual. Del mismo modo, se genera una matriz de transformación para cada *tracker* detectado por el dispositivo. Multiplicando cada una de ellas por la que codifica la posición del sensor en el mundo virtual se obtendrá una tercera matriz de la que se podrá obtener la posición y orientación global del *tracker* (ecuación 4.1). En caso de que el sensor únicamente sea capaz de detectar la posición de los elementos seguidos, bastará con multiplicar este valor directamente por la matriz de transformación para obtener la posición de los mismos respecto al sistema de coordenadas del plató.

2. **El control de la posición relativa de los objetos virtuales y los reales.** El *middleware* ha de conocer la posición de los objetos virtuales del plató, así como la de los actores para poder calcular cuando varía su posición relativa. Así, cuando esta cambie, la capa de aplicación enviará un mensaje al software de *render* para que, de forma automática, se actualice la máscara alfa.
3. **La activación de efectos en el mundo virtual a partir de eventos detectados en el real.** Cuando uno de los botones de los sensores es pulsado, o cuando se detecta un determinado gesto, esta información se transmite a través de la red en forma de pulsación de un *vrpn_button*. Al recibir esta notificación, el *middleware* comprueba qué

evento ha sido detectado y activa el efecto correspondiente, a través de la red, en el software de *render*.

$$\begin{aligned}
 & R(r_x, r_y, r_z) = R(\alpha, \beta, \lambda) \\
 & \quad T(t_x, t_y, t_z) \\
 \\
 & M = R(X, \alpha)R(Y, \beta)R(Z, \lambda)T(X, t_x)T(Y, t_y)T(Z, t_z) = \begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} \\ M_{10} & M_{11} & M_{12} & M_{13} \\ M_{20} & M_{21} & M_{22} & M_{23} \\ M_{30} & M_{31} & M_{32} & M_{33} \end{pmatrix} = \\
 \\
 & = \begin{pmatrix} \cos \beta \cos \lambda & -\cos \beta \sin \lambda & \sin \beta & t_x \\ \cos \lambda \sin \alpha \sin \beta + \cos \alpha \sin \lambda & \cos \alpha \cos \lambda - \sin \alpha \sin \beta \sin \lambda & -\cos \beta \sin \alpha & t_y \\ -\cos \alpha \cos \lambda \sin \beta + \sin \alpha \sin \lambda & \cos \lambda \sin \alpha + \cos \alpha \sin \beta \sin \lambda & \cos \alpha \cos \beta & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
 \\
 & \text{Si } (M_{02} = \sin \beta = 1), \text{ entonces } \begin{cases} \alpha = \text{atan2}(M_{10}, M_{11}) \\ \beta = \frac{\pi}{2} \\ \lambda = 0 \end{cases} \\
 \\
 & \text{Si } (M_{02} = \sin \beta = -1), \text{ entonces } \begin{cases} \alpha = -\text{atan2}(M_{10}, M_{11}) \\ \beta = -\frac{\pi}{2} \\ \lambda = 0 \end{cases} \\
 \\
 & \text{En cualquier otro caso } \begin{cases} \alpha = \text{atan2}(-M_{12}, M_{22}) \\ \beta = \text{asin}(M_{02}) \\ \lambda = \text{atan2}(-M_{01}, M_{00}) \end{cases}
 \end{aligned} \tag{4.1}$$

Como se ha comentado con anterioridad, el software de *render* utilizado en esta investigación ha sido eStudio. Esta aplicación permite la comunicación con aplicaciones de terceros a través de la red haciendo uso de su API. Para que la comunicación entre la capa de aplicación y la capa cliente presentada en los tres puntos anteriores sea posible se han utilizado dos herramientas principalmente. La primera es el tipo de dato *receiver map* que permite la modificación de variables del entorno virtual a través de la red y la segunda el envío de instrucciones remotas. eStudio permite el envío de fragmentos de código *Python* a través de la red, haciendo posible la activación remota de efectos, la variación de la posición relativa de los objetos virtuales, etc. Estas dos herramientas son también utilizadas para la transmisión de información entre el ordenador de *render* principal y los secundarios planteada en la arquitectura.

Con la finalidad de automatizar y agilizar la creación de estas conexiones con eStudio se ha diseñado un esquema XML de configuración llamado BST_Link en el que se guardan los siguientes datos (figura 4.5):

- **BST_IP:** dirección IP del ordenador de *render* principal.
- **Alpha_Reference_Tracker:** *tracker* que servirá como referencia de la posición del presentador para el cálculo de la posición relativa de este y los objetos virtuales que lo rodean.

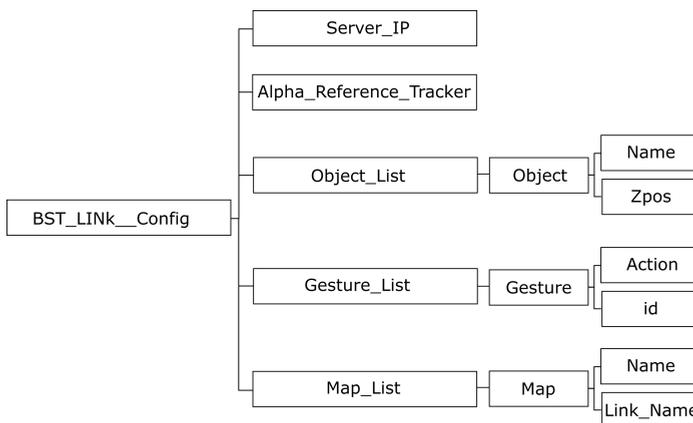


Figura 4.5: Esquema XML utilizado para la del *middleware* con Brainstorm eStudio

- **Object_List:** lista de objetos virtuales que pueden cambiar su posición relativa a lo largo de una emisión. Para cada objeto se almacena su nombre en eStudio (Name) y la profundidad a la que se encuentra (Zpos). De esta forma, cuando el presentador esté más alejado estará detrás y cuando esté más cerca se situará delante.
- **Gesture_List:** almacena la lista de gestos o eventos que se pueden detectar en el plató y que tienen un efecto en el mundo virtual. De cada uno de ellos se almacena el identificador del botón asociado (id) y la acción que debe lanzar en el entorno 3D (Action).
- **Map_List:** lista de *maps* que serán modificados remotamente en eStudio. Para cada uno de ellos se almacena un nombre (Name) que será el utilizado por la interfaz gráfica y otro nombre (Link_Name) que es el que adopta en el software de *render*.

Asimismo, se ha implementado una interfaz gráfica intuitiva que permite la edición y creación de archivos XML de configuración de la conexión con eStudio para usuarios no expertos (figura 4.6).

Utilizando estas características del *middleware* puede conectarse cualquier sensor del mercado a eStudio en cuestión de minutos utilizando un servidor VRPN de terceros y creando, a través de una interfaz gráfica específica, dos archivos de configuración XML. Asimismo, la interfaz gráfica del *middleware* resulta intuitiva y sencilla de usar, permitiendo visualizar todos los valores configurados en los archivos XML, así como los datos recibidos de los sensores y el estado de las distintas interacciones (figura 4.7).

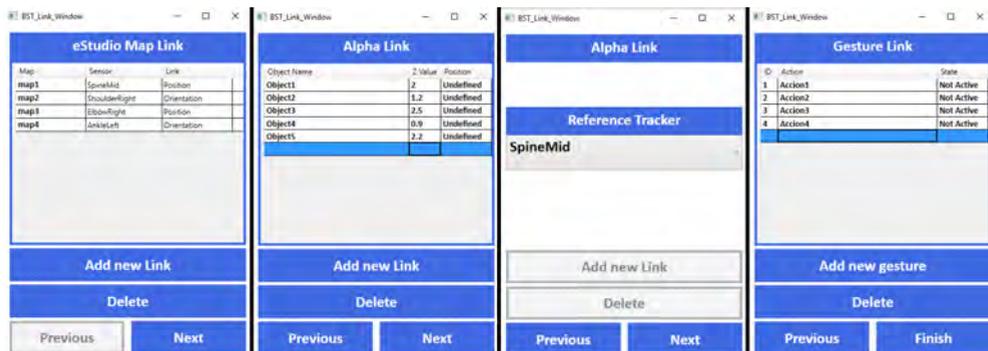


Figura 4.6: Diferentes pasos implementados en la interfaz gráfica implementada ad hoc, para la creación y edición de archivos XML de configuración de la conexión con eStudio



Figura 4.7: Interfaz gráfica del *middleware*

La implementación de la capa de aplicación abre un gran abanico de posibilidades al permitir la conexión ágil de cualquier dispositivo al motor gráfico (en nuestro caso el eStudio). A partir de este momento las capacidades del plató de trabajo pueden ser expandidas repitiendo el siguiente proceso de una forma iterativa:

1. Se analiza el plató virtual de televisión en busca de las necesidades que presenta. Se selecciona una de estas necesidades para ser solucionada y se plantean los requisitos funcionales necesarios para superarla.
2. Se realiza un estudio de los dispositivos que cumplen los requisitos funcionales planteados y se selecciona el más conveniente para el escenario concreto planteado.
3. El dispositivo se añade al sistema de forma ágil haciendo uso de la arquitectura distribuida y escalable presentada anteriormente.
4. Se comprueba el correcto funcionamiento del dispositivo para resolver el problema planteado.

La facilidad que ofrece la arquitectura presentada para conectar cualquier dispositivo hace posible el probar diferentes sensores hasta hallar una solución óptima así como la continua actualización de los ya incorporados. Del mismo modo permite el funcionamiento sinérgico de los dispositivos, combinando sus capacidades y aumentando así las del plató. Este funcionamiento conjunto hace que un dispositivo complementa a otros permitiéndole al sistema superar las limitaciones que estos presentan de forma individual y aislada. Por ejemplo, el sensor Microsoft Kinect es capaz de realizar el *tracking* del cuerpo de una persona pero no de los dedos de sus manos y el sensor Leap Motion está enfocado específicamente a esta tarea. Combinando ambos puede obtenerse un sistema capaz de realizar un seguimiento completo del cuerpo del usuario. A continuación se presenta el proceso seguido para dotar al plató de la Facultad de Ciencias de la Comunicación de movimiento libre de cámaras y diferentes posibilidades de interacción por parte del presentador mediante la inclusión de diferentes sensores específicos.

4.2.2. Tracking de Cámaras

La principal carencia que presentaba el plató virtual de televisión sobre el que se implementó la nueva arquitectura era la falta de un sistema de *tracking* para las cámaras que permitiese su desplazamiento libre en el escenario. Como se ha comentado previamente las cámaras se encontraban en una posición fija sobre sus trípodes sensorizados mediante sensores mecánicos que realizaban un seguimiento del *pan*, *tilt*, *zoom* y foco de las mismas. Cada vez que se necesitaba el desplazamiento de la cámara a una nueva posición se hacía imprescindible la calibración manual del sistema, midiendo la posición de la misma e introduciéndola manualmente en eStudio. Por tanto, no era posible la variación en tiempo real de su colocación.

El principal requisito funcional, asociado a la correcta alineación de cámaras reales y virtuales, que debía cumplir un dispositivo de cara a su inclusión en la nueva arquitectura del plató, era el de permitir realizar un seguimiento preciso de la posición y orientación de las cámaras, en tiempo real, en el escenario.

4.2.2.1. Implementación

De entre los sistemas comentados con anterioridad capaces de satisfacer el requisito funcional planteado, se optó por una solución óptica *outside-in* basada en tecnología infrarroja debido al tamaño fijo del plató, su relativo bajo coste, precisión y capacidad para realizar el

seguimiento de múltiples elementos. De entre las soluciones disponibles en el mercado se optó por la ofrecida por Optitrack [157] y su software Tracking Tools, ahora llamado Motive [156].

Esta solución permite realizar un seguimiento simultáneo de hasta 32 sólidos rígidos (grupos de marcadores reflectantes en una posición relativa fija y única) en tiempo real con una precisión submilimétrica. Su calibración es sencilla, a través de un proceso de barrido del espacio de captura con un *tracker* específico y el posicionamiento del origen de coordenadas y el plano del suelo mediante un marcador en forma de escuadra. También cuenta con un servidor VRPN del tipo *MultiReceiverSensor* que hace sencilla su inclusión en la arquitectura presentada. La configuración final instalada en el plató está formada por 8 cámaras de infrarrojos Flex 3 [155] repartidas de forma circular por la parrilla de focos del escenario y conectadas a un PC con procesador Intel Core 2 duo a 2,8 GHz, 4 Gb de RAM, tarjeta gráfica Nvidia Geforce 8400GS y sistema operativo Windows XP (figura 4.8). El coste total de esta configuración, incluida la licencia del software fue inferior a los 6500 euros.

Su incorporación al plató fue directa y se realizó en una mañana, el tiempo necesario para situar todas las cámaras en el escenario, realizar el cableado de las mismas y calibrar el sistema. Una vez realizada la instalación, su conexión a la capa de aplicación y por tanto a eStudio fue llevada a cabo en cuestión de minutos. Se utilizaron dos conjuntos de cuatro marcadores



Figura 4.8: Sistema de 8 cámaras Optitrack y Tracking Tools instalados en el plató

reflectantes (uno para cada cámara) a los que se les puso el nombre “Camera1” y “Camera2” en Tracking Tools (Figure 4.9). Se modificaron los archivos de configuración para incluir el nuevo sensor y estos dos *trackers* así como para asociar su posición y orientación a cuatro *receiver maps* formados por tres variables de punto flotante cada uno (“Camera1Position”, “Camera1Orientation”, “Camera2Position”, “Camera2Orientation”). Al iniciar tanto los servidores de Tracking Tools como el *middleware*, la posición y orientación de las cámaras fue accesible desde eStudio a través de la actualización en tiempo real de los *maps* asociados.

4.2.2.2. Pruebas

Las pruebas realizadas sobre el sistema de cámaras de infrarrojos de Optitrack se centraron en evaluar en las características técnicas del mismo, comprobar su correcta integración en el sistema y en verificar su funcionamiento para la implementación del movimiento libre de las cámaras en el plató.

Entre las características técnicas analizadas la primera fue la cobertura del sistema con la finalidad de comprobar si el volumen de captura generado tras la calibración era suficiente para que un operador pudiese desplazarse libremente en a lo largo del escenario con la cámara

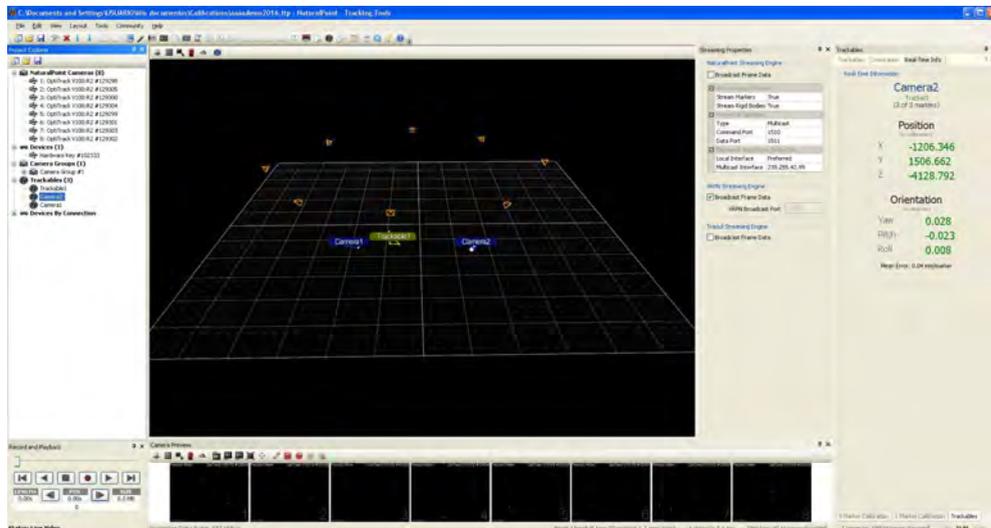


Figura 4.9: Interfaz gráfica de Tracking Tools enviando los datos de 3 trackers a través de la red haciendo uso de VRPN

al hombro. Para ello se desplazó una cámara sensorizada a lo largo del plató a tres alturas diferentes comprobando en que zonas el seguimiento de la misma era consistente y en cuales el sistema no podía calcular su localización:

- **1 metro:** desplazando la cámara a un metro del suelo se obtuvo una superficie útil de 5,66 metros de ancho por 5,20 metros de fondo.
- **2 metro:** con la cámara a dos metros del suelo el sistema ofrecía una cobertura de 4,60 metros de ancho por 4,10 metros de fondo.
- **Cámara al hombro (aproximadamente 1,5 metros):** la altura más común, ya sea desplazando la cámara en el hombro o manteniéndola sobre un trípode para grabar a actores que permanecen de pie ronda los 1,5 metros. A esta altura la cobertura del sistema de *tracking* es de 5 metros de ancho por 4,8 de fondo.

Todos estos resultados contienen la práctica totalidad de la zona de actuación del plató, que es de 5,10 metros de ancho por 3,3 metros de fondo, e incluyen aproximadamente un metro de espacio en la zona frontal de la misma, que es donde se suelen situar las cámaras por razones de diseño del ciclorama. Estas medidas prueban que la cobertura del sistema de 8 cámaras infrarrojas Optitrack es suficiente para desplazar una cámara sensorizada libremente a lo largo de del plató virtual.

El software Tracking Tools ofrece información sobre la calidad de la calibración realizada así como del rendimiento del sistema, lo que permite comprobar su correcto funcionamiento. Para la configuración instalada en el plató, el sistema presenta una latencia de 8,1 milisegundos desde el momento en que las cámaras capturan una imagen hasta el momento en que los datos de localización son enviados a través de la red al *middleware*. La velocidad de refresco de las cámaras es de 100 imágenes por segundo y el error medio del sistema es de en torno a 0,4 milímetros, imperceptible desde un punto de vista visual y por tanto suficiente para su uso en un plató virtual de televisión para emisiones en directo.

La correcta integración del sistema de *tracking* en el plató se analizó a través de una serie de pruebas de caja negra. Se situaron una serie de *trackers* en distintas posiciones conocidas dentro del escenario y se comprobó que los valores de entrada en el *middleware* (enviado por Tracking Tools a través de la red) eran correctos, que estos eran procesados de forma adecuada y posteriormente eran enviados al software de render en un formato compatible con su utilización en eStudio. Se realizaron diferentes pruebas variando el origen de coordenadas del sistema de *tracking*, obteniendo en todos los casos resultados satisfactorios.

En cuanto al funcionamiento del sistema para la correcta alineación de las cámaras reales y virtuales las pruebas realizadas fueron empíricas, centrándose en la coherencia visual del desplazamiento de ambas cámaras. En primer lugar se realizó una comparativa entre las rotaciones capturadas por los sistemas mecánicos del plató original y las obtenidas por el sistema Optitrack, no encontrando diferencias substanciales en la orientación de las imágenes finales obtenidas. En segundo lugar se desplazó la cámara al hombro a lo largo del volumen de captura, obteniendo en todo momento un desplazamiento coherente de los elementos reales situados en el escenario y el entorno virtual, consiguiendo, de este modo, una integración correcta entre ambos mundos.

4.2.3. Tracking de Objetos

Teniendo disponible el movimiento libre de cámaras, la siguiente limitación detectada en el plató sobre el que se desarrolló la implementación fue la ausencia total de interacción entre el presentador y el entorno sintético. Dentro de los diferentes tipos de interacción posibles, se decidió afrontar en primer lugar la manipulación directa de objetos virtuales a través de interfaces tangibles (objetos reales manipulables)(figura 4.10).

El requisito funcional, asociado a la interacción directa basada en el *tracking* de objetos, que debía cumplir un dispositivo de cara a su inclusión en la nueva arquitectura del plató, era, por tanto, el de permitir, de una forma discreta, el seguimiento en tiempo real del movimiento de un objeto en el escenario.

4.2.3.1. Implementación

Teniendo en cuenta que una cámara es un objeto, el reducido tamaño de los marcadores reflectantes y su probada cobertura y precisión así como su capacidad para cumplir el requisito funcional planteado, se decidió utilizar el sistema de cámaras infrarrojas Optitrack instalado para el desplazamiento libre de cámaras.

Al estar ya conectado con eStudio no se realizó ningún cambio en la configuración del plató más que la inclusión de un nuevo sólido rígido, formado por tres reflectantes unidos a una escuadra metálica negra, para su utilización como interfaz tangible en la fase de pruebas. Este nuevo marcador se denominó de forma genérica “Tracker1” y fue añadido en los archivos de configuración asociándolo a los *receiver maps* de eStudio “Tracker1Position” y “Tracker1Orientation”.



Figura 4.10: Manipulación de una pantalla a través de una interfaz tangible basada de sensores Optitrack

4.2.3.2. Pruebas

Habiendo probado el correcto funcionamiento del sistema de cámaras infrarrojas Optitrack para el desplazamiento libre de las cámaras del plató, las pruebas del seguimiento de objetos se centraron en la coherencia visual obtenida en la manipulación de los mismos por parte del presentador. Como se ha comentado previamente, se utilizó como interfaz tangible una escuadra metálica negra, a la que se asoció una pantalla reproduciendo un vídeo. El tamaño del objeto virtual se configuró para ser lo suficientemente amplio como para ocultar la escuadra a la vista del espectador.

Con esta configuración, el presentador puede manipular la interfaz tangible, teniendo co-

mo resultado una composición coherente en la que la pantalla responde a sus movimientos en tiempo real, sin introducir ningún *lag* apreciable y dando sensación de presencia. El objeto virtual, además de desplazarse siguiendo los movimientos del actor, responde a los cambios en la iluminación del plató virtual al tiempo que reproduce un vídeo, lo que hace que la percepción de una manipulación real y directa sea más efectiva.

Sin embargo, al tratarse de una combinación de tres imágenes 2D (la imagen del mundo real capturada por las cámaras, la imagen del fondo generada por ordenador y la imagen de los objetos situados delante del mundo real generada a partir de la máscara alfa) existen situaciones en las que una correcta alineación de ambos mundos resulta imposible. En el caso anterior, por ejemplo, cuando la pantalla se sitúa en posición horizontal, el espectador puede ver la interfaz tangible ya que no es posible que esta se encuentre dentro del objeto virtual. La solución en estos casos es la utilización de objetos del color del ciclorama para que sean borrados por el *chroma keyer* durante el proceso de integración. El problema persistiría con los reflectantes, pero su discreto tamaño y la posibilidad de situarlos en distintas zonas del objeto real permiten minimizar las posibilidades de que sean visibles en la emisión final.

4.2.4. Tracking de actores

Una vez probado el correcto funcionamiento del sistema para el seguimiento de cámaras y objetos, el siguiente paso para aumentar su interactividad era el *tracking* del movimiento de los actores con tres finalidades principales: realizar de forma automática el *Distance Keying* (control de la posición relativa de objetos virtuales y elementos reales), manipulación directa de objetos sin ningún tipo de interfaz tangible y detección de gestos.

4.2.4.1. Distance keying

En primer lugar se abordó la problemática del cálculo de posiciones relativas de actores y objetos virtuales.

Al no disponer el plató virtual de ningún dispositivo de seguimiento de actores, las posibilidades de interacción eran inexistentes. El control de las posiciones de los objetos virtuales respecto al mundo real se realizaba de forma manual, por lo que un operador tenía que estar centrado en estos eventos en lugar de poder realizar tareas de mayor valor añadido. Para evitar este derroche de capital humano se planteó la automatización del proceso.

Por tanto, el principal requisito funcional, asociado a la automatización del proceso de *distance keying*, que debía cumplir un dispositivo de cara a su inclusión en la nueva arquitectura

del plató, era el de permitir, de una forma invisible al espectador, el seguimiento en tiempo real de los desplazamientos del presentador en el escenario.

Implementación

Aunque se planteó la utilización del sistema de cámaras de infrarrojos para seguir el movimiento del presentador en el escenario este se desechó debido a la visibilidad de los marcadores reflectantes y la imposibilidad de ocultarlos con objetos virtuales superpuestos. La opción de situar los marcadores en la espalda del actor no es tampoco una solución válida ya que afecta directamente a la naturalidad de su comportamiento en el escenario, al tener que estar pendiente, en todo momento, de que estos no sean visibles desde las cámaras de vídeo. Por tanto, era necesario, de cara a cumplir el requisito funcional planteado, un sensor capaz de realizar el seguimiento de la anatomía del actor sin la utilización de ningún tipo de marcador [37]. Los principales sensores con estas capacidades en este momento del desarrollo eran el Microsoft Kinect [201] y el Asus Xtion PRO LIVE [81]. Ambos utilizan la tecnología de Primesense basada en la proyección de patrones de luz estructurada en el espectro infrarrojo y, por tanto, tienen un rendimiento y precisión muy similares [63]. Su principal diferencia estriba en que el Microsoft Kinect cuenta con un motor que le permite modificar su inclinación vertical y el Asus Xtion PRO LIVE puede trabajar a 60 fotogramas por segundo reduciendo su resolución y por tanto su precisión. La velocidad estándar de funcionamiento de ambos dispositivos es de 30 imágenes por segundo, siendo esta una cifra que se sitúa dentro del tiempo real. Se decidió incorporar el sensor Microsoft Kinect (figura 4.11) al sistema debido a su coste más económico (en torno a los 150 euros), su amplia red de distribuidores y su extensa documentación y soporte.

El volumen de captura de este dispositivo es el adecuado para las dimensiones del plató de trabajo ya que es capaz de realizar un seguimiento de usuarios desplazándose entre 50 centímetros y 5 metros de distancia del sensor. En nuestro caso, el escenario tiene una profundidad de 3.3 metros, por lo que el presentador se mantiene siempre dentro del rango per-



Figura 4.11: Sensor Microsoft Kinect

mitido. La distancia óptima de funcionamiento del sensor se encuentra entre 1 metro y tres metros [89] y coincide con el principal rango de trabajo en nuestro estudio. Más allá de los 4 metros, el error del sensor aumenta de forma cuadrática, alcanzando los 4 centímetros a 5 metros debido a la poca resolución de la cámara de profundidad. Esta precisión variable así como su elevada latencia, de en torno a los 250 milisegundos [20], son sus principales limitaciones. La variabilidad de la precisión puede evitarse manteniendo a los actores siempre dentro de la distancia de captura óptima y la latencia puede ser contrarrestada aumentando el *lag* introducido en la señal real por el *chroma keyer*.

Por otro lado, el Microsoft Kinect es capaz de capturar la posición de hasta dos actores, reconstruyendo un esqueleto de compuesto por 20 huesos que permite conocer la posición de diferentes partes de su anatomía [110]. Esta característica puede ser utilizada para un control de la posición relativa entre actores y objetos virtuales con un mayor nivel de detalle, ya que las interacciones suelen producirse a través de las extremidades y utilizar un punto en el centro del cuerpo de actor como referencia podría no ser lo óptimo en algunos casos [76].

Para incluir el sensor en la arquitectura se utilizó el software *Flexible and Articulated Skeleton Toolkit* (FAAST) [171] (figura 4.12), una herramienta para la integración de los sensores PrimeSense en aplicaciones de realidad virtual a través de un servidor VRPN. FAAST codifica la información del dispositivo en forma de un *MonoReceiverSensor* con 24 puntos de *tracking*. La inclusión del sensor, por tanto, fue directa, requiriendo únicamente la modificación de los archivos de configuración. En sensor se conectó a un PC con procesador Intel Core 2 duo a 2,8 GHz, 4 Gb de RAM, tarjeta gráfica Nvidia Geforce 8400GS y sistema operativo Windows 7.

Pruebas

La correcta integración del sensor Microsoft Kinect en el plató se analizó a través de una serie de pruebas de caja negra. Se situaron distintas partes de la anatomía de un usuario en posiciones conocidas dentro del escenario y se comprobó que los valores de entrada en el *middleware* (enviados por FAAST a través de la red) eran correctos, que estos eran procesados de forma adecuada y posteriormente eran enviados al software de render en un formato compatible con su utilización en eStudio. Se realizaron diferentes pruebas variando el origen de coordenadas del sensor, obteniendo en todos los casos resultados satisfactorios.

Las pruebas de funcionamiento del sensor Microsoft Kinect para el control de la posición relativa de los actores y los objetos virtuales se realizaron utilizando dos configuraciones dis-

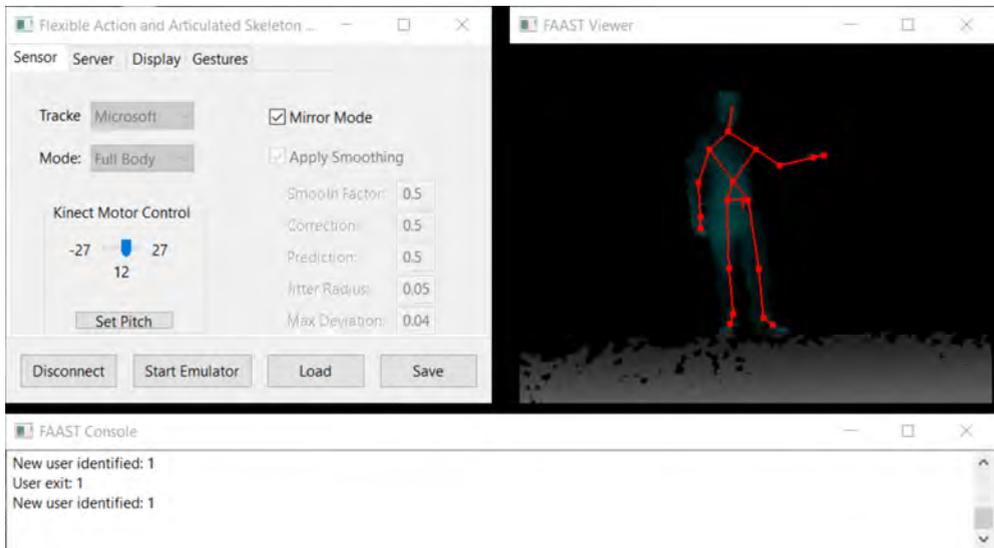


Figura 4.12: Interfaz del software FAAST mostrando los puntos anatómicos detectados por sensor Microsoft Kinect

tintas. En la primera el dispositivo se situó en una posición fija y conocida y en la segunda se asoció a un marcador del sistema Optitrack para permitir su libre movimiento en el plató. Esta segunda opción, permite asociar el sensor a una cámara asegurando que el presentador esté siempre dentro del espacio de captura del mismo cuando esté siendo enfocado. Esto resulta especialmente útil en platós virtuales de gran tamaño en los que, para tener una cobertura de toda su superficie, serían necesarios numerosos sensores que podrían provocar interferencias mutuas. Como se ha expuesto con anterioridad la composición de posiciones, entre la del propio sensor y los datos que este captura, para obtener la localización global de los actores se realiza de forma automática en el *middleware*.

La prueba realizada consistió en situar una marca a 2 metros del origen de coordenadas global y, en el mundo virtual, situar un objeto en esa misma posición. Se empleó el sistema de control del valor alfa incluido en el *middleware* para que el presentador caminase alrededor del objeto virtual entrando en escena por detrás de este y saliendo por delante. También se comprobó que el cambio de la posición relativa del objeto virtual se realizaba en la posición correcta atravesándolo y confirmando que el cambio se realizaba en la posición de la marca física del escenario. Se utilizó el torso como referencia de la posición del actor por ser uno

de los puntos anatómicos captados por el Microsoft Kinect con mayor precisión. El funcionamiento del sensor fue correcto en ambos casos, permitiéndole al presentador rodear objetos virtuales y atravesarlos (figura 4.13). Estas dos pruebas se realizaron en las dos configuraciones posibles (posicionamiento del sensor fijo y móvil) sin que se apreciaran diferencias en el comportamiento del sistema.

También se comprobó el uso de otras partes del cuerpo como puntos de referencia de la posición del actor, en busca de un *distance keying* avanzado, no basado en una única referencia. Se utilizó la mano para, sin variar la posición global del presentador, cambiar su posición relativa respecto a un objeto sintético. Esta capacidad resulta interesante debido a que la mayor parte de las interacciones humanas son llevadas a cabo con las manos. Se demostró el correcto funcionamiento para desplazamientos amplios pero, debido al ruido introducido por el sensor, se constató la existencia de una zona intermedia en la que la posición del objeto variaba de forma continua, provocando un parpadeo del objeto sobre la mano del presentador.

Por tanto, tras la fase de pruebas se concluyó que el sensor tiene la suficiente precisión para realizar un control de la posición relativa de los objetos virtuales y el mundo real simple, basado en un punto de referencia situado en el torso. Sin embargo, para un *distance keying*



Figura 4.13: Actor pasa de estar detrás de un objeto virtual a estar delante de forma automática

más avanzado se hace necesario el uso de un sensor con mayor precisión en la captura de las extremidades de los actores.

4.2.4.2. Interacción directa

El segundo tipo de interacción que se abordó fue la manipulación directa de objetos virtuales por parte de los actores. Esto implica tanto la detección de colisiones como el movimiento acompasado y coherente de los elementos sintéticos con respecto a las acciones de los presentadores.

El principal requisito funcional, asociado a una interacción directa de los presentadores con los objetos virtuales, que debía cumplir un dispositivo de cara a su inclusión en la nueva arquitectura del plató, era el de permitir el seguimiento, en tiempo real, de diferentes partes de la anatomía del presentador, teniendo una especial relevancia el *tracking* de las manos del actor dentro del escenario.

Implementación

El principal sensor que cumplía el requisito funcional planteado para la interacción directa de los presentadores con los objetos virtuales era el Microsoft Kinect que, al estar ya incluido en el sistema, no necesitó más configuración que la asociación de varios de sus *trackers* a *receiver maps* de eStudio, para así poder asociar sus movimientos a objetos del mundo virtual.

Este sensor tiene como uno de sus grandes inconvenientes el ruido de los datos que genera, especialmente en las extremidades. Esto no supone un problema mayor para la detección de



Figura 4.14: Sensor Leap Motion

colisiones, pero sí que resulta un inconveniente insalvable a la hora de manipular objetos directamente, ya que estas interacciones se realizan, generalmente, con las manos. El resultado visual, como se comprobó en la fase de pruebas, es que los objetos tiemblan en la mano del presentador, afectando de una manera definitiva a la sensación de presencia de los mismos.

Por tanto se hizo necesaria la inclusión de un sensor que, cumpliendo el requisito funcional planteado, fuese capaz de realizar un seguimiento del movimiento de las manos del presentador con la mayor precisión y el menor ruido posible. De entre las diferentes alternativas del mercado se eligió el sensor Leap Motion ya que no necesita que el presentador utilice ningún tipo de marcador (figura 4.14). Su precisión va desde menos de 0.2 milímetros en entornos estáticos a 1.2 milímetros en dinámicos [70], siendo mucho más preciso que el Microsoft Kinect y haciendo prácticamente imperceptible el ruido generado por el dispositivo. Realiza un seguimiento de las manos con una velocidad de refresco de la información de alrededor de 115 muestras por segundo con un *lag* en torno a los 5 milisegundos (figura 4.15). Su principal limitación es el tamaño limitado del volumen de captura que ofrece. Para un seguimiento óptimo de las manos estas deben de situarse en el entorno de los 250 milímetros sobre el sensor [189], lo que lo convierte en un dispositivo idóneo para programas en los que el presentador se sitúe tras una mesa, como los informativos. El precio del dispositivo ronda los 80 euros.

Se conectó el sensor a un nuevo PC con procesador Intel Core 2 duo a 2,8 GHz, 4 Gb de RAM, tarjeta gráfica Nvidia Geforce 8400GS y sistema operativo Windows 7. Para incluirlo

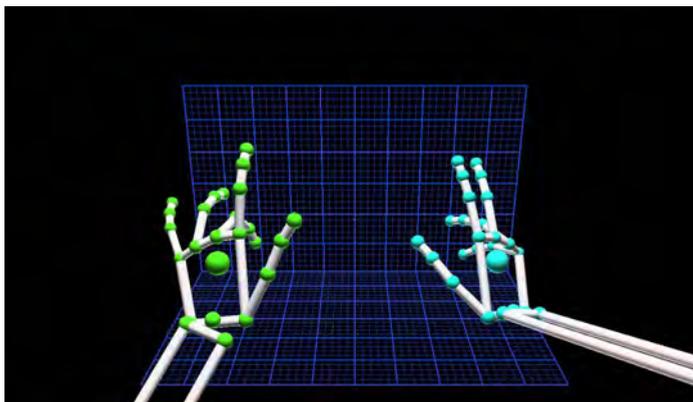


Figura 4.15: Representación gráfica de los datos capturados por Leap Motion

en el sistema se creó un servidor VRPN capaz de enviar la posición y orientación de dos manos y todos sus dedos encapsulados en forma de un solo *MonoReceiverSensor* (Figure 4.16). Para la implementación se utilizó el lenguaje c# y WPF para la interfaz gráfica. Una vez implementado, la conexión del sensor con el *middleware* se realizó añadiendo en el archivo de configuración una nueva conexión en la que se incluían *RightHandTracker* y *LeftHandTracker* y sus respectivos dedos. Para su prueba en plató se enlazó el valor de las dos manos con dos *maps* de eStudio.

Pruebas

Se realizaron pruebas basadas en la asociación de objetos virtuales a la posición de las manos de un presentador. De esta forma se elimina la necesidad de utilizar interfaces tangibles y por tanto, se evitan los problemas que esta solución conlleva (figura 4.17). Las pruebas se realizaron utilizando el Microsoft Kinect y el Leap Motion, con la finalidad de poder realizar una comparativa visual del funcionamiento de ambos. Para poder realizar una comparación entre los resultados obtenidos en la manipulación de objetos directa y a través de una interfaz tangible su utilizó el mismo modelo 3D de pantalla con reproducción de vídeo que en la fase de pruebas previa.



Figura 4.16: Interfaz gráfica implementada para el servidor VRPN de Leap Motion en el momento de la detección del gesto *swipe*

El sensor Microsoft Kinect demostró no ser lo suficientemente preciso para esta manipulación ya que, como se ha comentado con anterioridad, la información del seguimiento de las manos que captura contiene demasiado ruido y hace que el objeto virtual manipulado tiemble de forma continua rompiendo de esta forma la sensación de presencia del mismo. Se redujo el ruido mediante la incorporación de un suavizado de la señal, pero el nivel de suavizado necesario para evitar el temblor de los objetos virtuales hacía que los movimientos del actor no fuesen directamente transmitidos y los modelos 3D manipulados dieran la impresión de flotar tras las manos en lugar de responder a sus acciones.



Figura 4.17: Actor manipulando un objeto virtual a través de la sensorización proporcionada por el Microsoft Kinect

Por otro lado, se analizó la correcta integración del sensor Leap Motion en el plató a través de una serie de pruebas de caja negra. Se situaron las manos de un usuario en diferentes

posiciones conocidas dentro del escenario (dentro del limitado rango de trabajo del sensor) y se comprobó que los valores de entrada en el *middleware* (enviados por el servidor VRPN implementado ad hoc a través de la red) eran correctos, que estos eran procesados de forma adecuada y posteriormente eran enviados al software de render en un formato compatible con su utilización en eStudio. Se realizaron diferentes pruebas variando la posición del sensor en el escenario, obteniendo en todos los casos resultados satisfactorios.

Una vez comprobada la correcta integración del dispositivo en el sistema, se realizaron pruebas basadas en la asociación de objetos virtuales a la posición de las manos de un presentador. El sensor Leap Motion, gracias a su mayor precisión, resultó más efectivo que el Microsoft Kinect, permitiendo una manipulación directa y eficaz de los modelos 3D en tiempo real y con una ausencia prácticamente total de ruido. El objeto asociado a la posición de la mano siguió sus movimientos de una forma continua y suave, respondiendo sin retardo a las acciones de la misma.

A pesar de las ventajas de este tipo de interacción, desde el punto de vista del usuario, resulta más intuitiva y sencilla la manipulación de una interfaz tangible, ya que en la interacción directa con elementos virtuales no se dispone de ninguna referencia en el mundo real de las modificaciones que se están realizando en la posición y orientación de los mismos.

Por tanto, es posible concluir la preferencia, por parte de los presentadores, del uso de un elemento real para manipular el entorno virtual lo que, unido a su precisión, hace recomendable la utilización de este método en la mayoría de las situaciones. Para escenarios y efectos en los que la presencia de una interfaz tangible no sea posible se hace necesario el uso de un sensor con mayor precisión que el Microsoft Kinect, como el Leap Motion.

4.2.4.3. Detección de gestos

El último tipo de interacción propuesto para su implementación en el plató de trabajo de cara a mejorar la integración del presentador en el entorno sintético fue la detección de gestos que permite, a través del reconocimiento de una serie de movimientos predefinidos, liberar a los actores del uso de dispositivos visibles para controlar eventos en el mundo virtual. Como se ha expuesto previamente en el análisis de casos reales, en la mayor parte de los estudios virtuales de televisión la interacción entre el presentador y el entorno 3D (cuando esta existe) se realiza, generalmente, a través de un mando a distancia. La necesidad de tener este objeto en la mano resta naturalidad a los movimientos del presentador y distrae la vista del espectador hacia un dispositivo sin ningún valor desde el punto de vista comunicativo. Por ello se propone

la utilización de sensores sin marcadores que permitan la detección de gestos del presentador con la finalidad de utilizarlos para lanzar efectos en el entorno virtual.

Dentro de los gestos que se pueden realizar en un plató virtual de televisión se diferencian, principalmente, dos tipos: Gestos amplios, y gestos discretos.

Gestos amplios

Son aquellos que el actor realiza frente a la cámara y que se basan en grandes desplazamientos que implican diferentes partes del cuerpo. Un gesto amplio puede ser desde un *swype* realizado desplazando la mano delante del cuerpo, a un salto o a levantar la mano por encima de la cabeza. Se caracterizan por no necesitar un grado de precisión muy elevado en la señal para su reconocimiento y por ser tolerantes al ruido de la misma ya que, al estar caracterizados por grandes desplazamientos, estos pueden ser diferenciados de una forma sencilla de pequeñas variaciones en los datos capturados.

El principal requisito funcional, asociado a la detección de gestos amplios, que debía cumplir un dispositivo de cara a su inclusión en la nueva arquitectura del plató, era el de permitir el *tracking* en tiempo real e invisible al espectador, de diferentes partes de la anatomía del presentador que pudiesen ser utilizadas en la realización de un gesto.

Implementación

Un sensor ideal para la detección de gestos amplios es el Microsoft Kinect ya que el ruido de su señal se ve contrarrestado por la naturaleza de estos, permitiendo reconocer movimientos que impliquen cualquier parte del cuerpo. Al estar el dispositivo ya integrado en el plató virtual se optó por su utilización con esta finalidad.

FAAST, el software utilizado para la inclusión del dispositivo en el sistema, incluye una interfaz gráfica que permite, de una forma intuitiva y visual, crear gestos basados en reglas. Para esto dispone de una sintaxis sencilla que representa los componentes espaciales y temporales básicos de una acción [170]. Estas reglas individuales forman primitivas que, a pesar de su sencillez, pueden ser combinadas simultáneamente y en secuencias temporales para crear interacciones complejas. Mediante esta técnica se pueden definir gestos del tipo:

1. Inclinar el cuerpo a la derecha al menos 15 grados
2. Saltar más de 20 centímetros

3. Situar la mano derecha por encima del hombro derecho al menos 15 centímetros
4. Situar el pie izquierdo por encima de la rodilla derecha al menos 10 cm

Todas estas acciones pueden definir un gesto por sí mismas o crear movimientos más complejos mediante su combinación. Las acciones 2 y 3 podrían combinarse para que la característica a detectar fuese un salto con la mano derecha levantada. Incluso se podría poner una restricción de tipo temporal, por ejemplo: levantar la mano derecha y, en los tres segundos posteriores, realizar un salto.

De esta forma, pueden definirse de un modo natural e intuitivo diferentes tipos de gestos. La principal ventaja de esta metodología es su sencillez y eficacia para acciones simples. Sin embargo para movimientos más complejos, como el dibujo de formas en el aire, la dificultad de definirlos aumenta hasta hacer, en algunos casos, imposible su detección de una forma robusta.

Una vez detectado un gesto, FAAST puede producir diferentes tipos de salida como pulsaciones de teclado y ratón, pausa o detención del propio software o la pulsación de botones VRPN a través de la red. Esta última alternativa es la utilizada para su conexión con el *middleware* y, por tanto, para la integración de la detección de gestos amplios en el plató.

Pruebas

Para la fase de pruebas de gestos amplios se planteó la creación y prueba de cinco gestos:

1. *Swipe* con la mano derecha hacia el lado izquierdo.
2. Levantar la mano derecha por encima de la cabeza.
3. Dibujar un círculo en el aire.
4. Realizar un movimiento de abajo a arriba con las dos manos
5. Sentarse

En la definición de estas acciones a través de las reglas proporcionadas por FAAST se obtuvieron resultados dispares. Para las acciones más sencillas que únicamente requerían un cambio de posición desde un punto a otro como la 1, 2 y 4 su definición fue rápida y su detección posterior robusta y eficaz. Se definió una posición inicial aproximada de la mano,

una posición final y el tiempo máximo de duración del cambio entre ambas. Para estos tres gestos el porcentaje de acierto del sistema es cercano al 100%. Sin embargo, para gestos más complejos como el de sentarse (en el que entra en juego todo el cuerpo) o dibujar un círculo en el aire la definición a través de esta metodología resulta más compleja. El principal problema para detectar el gesto de sentarse reside en el ruido que aparece en la señal del sensor al no tener esta una visión directa de las caderas del actor, lo que hace que utilizar la relación de posición entre estas y las rodillas no sea posible. Se utilizó, por tanto, la distancia entre hombros y rodillas para caracterizar la acción, lo que implica la necesidad de adaptar la definición del gesto a cada presentador en caso de existir diferencias de altura considerables. No fue posible, sin embargo, hallar una definición robusta para el gesto de dibujar un círculo en el aire. La complejidad intrínseca del gesto que implica el paso de la mano por diferentes posiciones relativas respecto al cuerpo, así como la variabilidad del tamaño del mismo hizo que la definición mediante reglas no fuese una solución viable.

Al realizarse la detección de uno de estos gestos, se registra la pulsación de un botón VRPN con un identificador numérico único que está asociado en el *middleware* a una acción en eStudio que a su vez lanza un evento en el mundo virtual.

De esta forma, se probó la correcta integración de la detección de gestos del presentador en las nuevas capacidades del plató. Así mismo, se llegó a la conclusión de que, para gestos sencillos y amplios, la definición a través de reglas planteada por FFAST es una solución viable, rápida y robusta que puede ser llevada a cabo por usuarios no expertos al utilizar una sintaxis basada en el lenguaje natural. Para la detección de gestos más complejos, sin embargo, ha de optarse por otro tipo de soluciones más complejas, como las basadas en inteligencia artificial.

Gestos discretos

Son aquellos que no se basan en grandes movimientos si no que buscan discreción y naturalidad. Generalmente se realizan con los dedos de la mano y permiten interacciones más sobrias y menos espectaculares. En un plató virtual de televisión estos gestos son ideales para un entorno informativo, en el que la naturalidad del presentador combinada con cierto grado de sobriedad es un punto clave para la credibilidad del discurso.

El principal requisito funcional, asociado a la detección de gestos discretos, que debía cumplir un dispositivo de cara a su inclusión en la nueva arquitectura del plató, era el de permitir un seguimiento preciso, invisible para el espectador y en tiempo real, de parte de la

anatomía de un presentador destinada a la realización de gestos (principalmente las manos).

Implementación

La detección de este tipo de gestos hace imprescindible un dispositivo con una alta precisión y que no introduzca grandes cantidades ruido en su señal ya que, cualquier variación en la misma podría ser confundida por el sistema con una acción predefinida, dando lugar a falsos positivos.

De entre los sensores disponibles en el que más se ajusta al requisito funcional planteado es el Leap Motion, ya que su gran precisión combinada con su capacidad para realizar el seguimiento de los dedos de la mano le permite detectar gestos sutiles y naturales. El propio SDK del dispositivo incorpora la detección de cuatro tipos de acciones que fueron integradas en el servidor VRPN implementado para su evaluación.

Aunque existen soluciones más avanzadas para la creación de gestos a través de técnicas como las redes neuronales convolucionales [122] o la combinación del sensor con otros dispositivos [87, 119] la finalidad de la presente tesis se centra en la integración ágil del sensor y la prueba preliminar de las posibilidades que este podría aportar a un plató virtual de televisión. Por este motivo no se ha desarrollado un detector de gestos propio y se ha optado por el análisis del comportamiento del sensor para la detección de sus gestos predefinidos.

Pruebas

El proceso de pruebas se centró en la detección de los cuatro gestos predefinidos por el sensor y en la transmisión a eStudio de las instrucciones consecuencia de estos. Para ello se realizó un estudio de la robustez de la detección a través de su eficacia en 100 repeticiones de cada acción. Los resultados mostraron un porcentaje de acierto superior al 80 % en todos los casos, llegando al 92 % en los gestos de deslizamiento (*swipe*) y del 96 % en el gesto circular. Los gestos de pulsación en pantalla y toque con el dedo se quedaron en el 87 % y 81 % respectivamente (figura 4.18). Se detectó un problema, principalmente, con la robustez del seguimiento de los dedos (no así de las manos al completo), lo que implicó la aparición de numerosos falsos positivos y falsos negativos, que provocaron que las acciones que implican un movimiento más sutil tuviesen un menor porcentaje de acierto. Sin embargo, y a pesar de su insuficiente robustez para emisiones en directo, el sistema permite comprobar la correcta integración del mismo en el plató. A modo de trabajo futuro será necesario profundizar en

técnicas de detección de gestos que permitan mejorar estos resultados y acercarse en la medida de lo posible al 100 % de éxito deseable en un estudio virtual de televisión.

La integración de la detección de gestos del Leap Motion en el sistema fue directa, lanzando los efectos correspondientes a cada acción en eStudio a través del *middleware* de forma automática y sin retardo.

4.2.5. Actualización del sistema

Una vez incluidos en la arquitectura los sistemas de *tracking* basados en las cámaras de infrarrojos Optitrack, el Microsoft Kinect y el Leap Motion (figura 4.19), se habían afrontado los principales retos interactivos de un plató virtual de televisión aumentando las capacidades del mismo de una forma ágil y sin modificar la configuración de los elementos que previamente estaban presentes en el estudio. Como se ha comentado con anterioridad, una de las principales ventajas de esta nueva arquitectura es la facilidad que ofrece para actualizar el hardware y esta capacidad se probó mediante la sustitución de uno de los sensores por una versión actualizada del mismo.

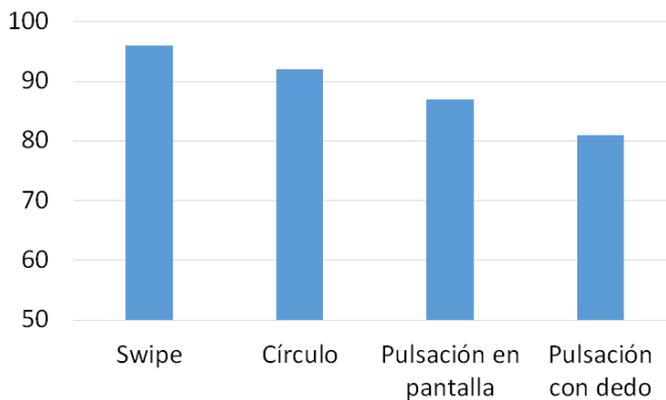


Figura 4.18: Porcentaje de acierto de los diferentes gestos detectados por el sensor Leap Motion

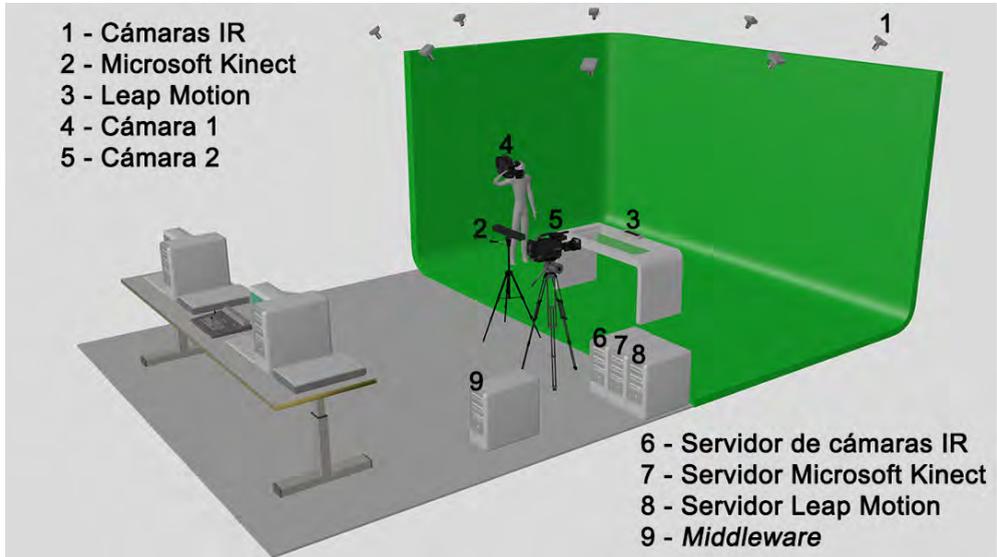


Figura 4.19: Configuración final del plató

4.2.5.1. Implementación

El dispositivo sustituido fue Microsoft Kinect cuya nueva versión, el Microsoft Kinect V2 (figura 4.20), mejora sustancialmente las capacidades de su predecesor manteniendo su precio en torno a los 150 euros. Incrementa su resolución obteniendo como resultado una mayor precisión que se ve reflejada en un error que varía entre menos de 2 milímetros y 4 milímetros dependiendo de la distancia al usuario manteniendo la tasa de refresco de 30



Figura 4.20: Sensor Microsoft Kinect V2

imágenes por segundo [198]. También aumenta su campo de visión así como el número de puntos anatómicos que detecta, permitiendo realizar el seguimiento de 25 partes del cuerpo humano [6] en un mayor volumen de captura [43]. Este incremento de los puntos detectados permite, por ejemplo, saber si el usuario tiene la mano abierta o cerrada en función de la posición de su dedo gordo. Una prueba de la precisión de este nuevo dispositivo es el hecho de que el error que presentan los ángulos obtenidos para los huesos que unen estos puntos se mantiene generalmente por debajo de 10 grados [190].

Además, el conjunto de herramientas que ofrece el SDK del dispositivo incluye el software Kinect Studio V2.0, que permite grabar la información capturada por el sensor, y la aplicación Visual Gesture Builder que, a partir de una captura grabada con Kinect Studio, permite crear nuevos gestos de una forma ágil y sencilla. Para ello hace uso de técnicas de inteligencia artificial que, a partir de una serie de muestras de un determinado gesto, permiten al sistema aprender las características del mismo y ser, a posteriori, capaz de detectarlo en tiempo real. El proceso a seguir para la creación de un gesto es el siguiente:

1. Se captura y graba al usuario interactuando con el sensor y realizando repetidas veces la acción que se desea aprender. Se realizan, como mínimo, dos capturas distintas que servirán una para entrenar al sistema y la otra para probar la eficacia de la detección.
2. Se realiza una caracterización de la interacción grabada utilizando una línea de tiempo en la que se le indica de forma manual al sistema en qué momentos se realiza el gesto (cuando empieza y cuando termina) y en cuáles no (figura 4.21).
3. Se entrena al sistema utilizando el algoritmo *adaboost* [162] que hace uso de una serie de características obtenidas por el sensor para caracterizar el gesto que se está creando.
4. Se comprueba la eficacia de la detección utilizando los datos obtenidos en el paso 3 sobre una grabación de prueba.
5. Se genera un archivo que permitirá importar la detección del gesto programáticamente utilizando las herramientas específicas incluidas en el SDK de forma directa para el programador.

Siguiendo estos pasos se pueden crear infinitos gestos que, a posteriori, pueden ser utilizados en aplicaciones interactivas para conseguir una interacción más natural por parte del usuario. Aunque se han desarrollado proyectos centrados en la detección de gestos realizados

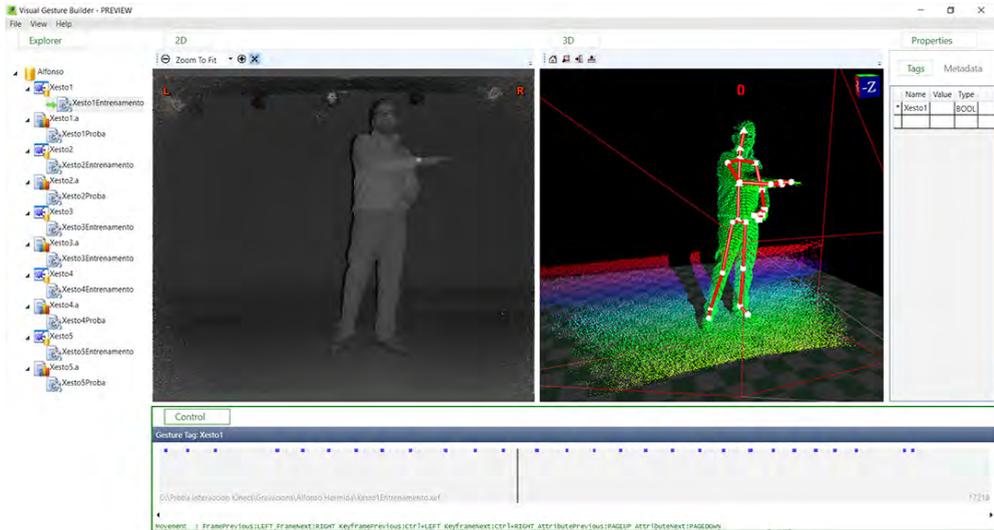


Figura 4.21: Proceso de caracterización de gestos

únicamente con las manos [104], en el entorno de los platós virtuales de televisión se hace necesaria la utilización de todo el cuerpo ya que este entra en relación, en toda su extensión, con el entorno virtual.

El sensor se conectó a un PC con procesador Intel Core i7-4710HQ a 2.5 GHz, 8 Gb de RAM, tarjeta gráfica Nvidia Geforce GTX850M y sistema operativo Windows 10. Para la inclusión en el sistema se implementó un servidor VRPN que incluye tanto el envío de la posición y orientación de los 25 puntos que detecta el sensor en forma de *MonoReceiverSensor* como la transmisión de los gestos reconocidos como pulsaciones de botones VRPN (figura 4.22).

4.2.5.2. Pruebas

Las pruebas realizadas se centraron en comprobar la correcta integración del sensor en el sistema en analizar su eficacia ante los tres tipos de interacción para los que había sido probada su anterior versión. *Distance keying*, manipulación directa y detección de gestos.

En primer lugar, al igual que en los anteriores casos, se llevaron a cabo una serie de pruebas de caja negra de cara a determinar la correcta integración del dispositivo en el sistema. Se situaron diferentes partes de la anatomía de un usuario en distintas posiciones conocidas

dentro del escenario y se comprobó que los valores de entrada en el *middleware* (enviados por el servidor VRPN implementado ad hoc a través de la red) eran correctos, que estos eran procesados de forma adecuada y posteriormente eran enviados al software de render en un formato compatible con su utilización en eStudio. Se realizaron diferentes pruebas variando la posición del sensor en el escenario, obteniendo en todos los casos resultados satisfactorios.

Para el control de la posición relativa de los elementos reales y virtuales se comprobó una mejora en su eficacia utilizando como punto de referencia las extremidades del actor. El aumento de la precisión del sensor hace posible una detección robusta del momento en el que una extremidad cambia su posición relativa respecto a un objeto virtual, reduciendo notablemente la zona de ambigüedad en la que esta varía continuamente. No se observaron diferencias notables en el caso de utilización del torso como referencia, ya que la primera versión del dispositivo realizaba esta tarea correctamente.

Se pudo observar, también, una mejora notable en la manipulación directa de objetos virtuales debido a la reducción del ruido de la señal. Sin embargo, la señal sigue presentando pequeñas variaciones que introducen temblores en los objetos manipulados. Con un ligero suavizado de la misma se consigue reducir de forma notable esta problemática sin afectar a la reactividad de los elementos sintéticos a los movimientos del actor. Para esta tarea es un dispositivo que se encuentra a medio camino entre el ruido introducido por el Microsoft



Figura 4.22: Interfaz gráfica implementada para el servidor VRPN asociado al sensor Microsoft Kinect V2

Kinect original y la precisión del Leap Motion, haciéndolo usable pero no siendo la solución ideal para estos fines.

En cuanto a la detección de gestos amplios se siguió el mismo proceso que el llevado a cabo para el análisis de la eficacia del Microsoft Kinect original utilizando la detección por reglas. Se crearon los mismos gestos utilizando las herramientas de aprendizaje que ofrece el SDK:

1. *Swipe* con la mano derecha hacia el lado izquierdo.
2. Levantar la mano derecha por encima de la cabeza.
3. Dibujar un círculo en el aire.
4. Realizar un movimiento de abajo a arriba con las dos manos.
5. Sentarse.

La creación de estos gestos utilizando técnicas de inteligencia artificial no presentó diferencias en la dificultad de su caracterización por parte del sistema. Mediante el aprendizaje a partir de la grabación de 50 repeticiones de cada gesto se obtuvo un porcentaje de acierto en la detección superior al 95% en todos los casos, lo que convierte a esta aproximación en una solución viable para la creación de gestos naturales personalizados para cada programa.

Esta capacidad de aprendizaje permite proponer un cambio en los flujos de trabajo previos a la grabación de un programa, substituyendo las semanas de ensayo de coreografías por interacciones reales del presentador que hagan uso de su gestualidad natural, aprendida por el sistema. Esta nueva forma de trabajo permite que sea el presentador quien cree los gestos que lanzarán efectos en directo dentro del plató virtual. Se plantea que, al hacer uso la gestualidad natural propuesta por el actor, la curva de aprendizaje del mismo será mucho más pronunciada y por tanto se acortarán los tiempos de preparación y se reducirá el estrés del presentador durante el rodaje. Para ello se ha realizado un estudio preliminar que trata de sentar las bases para un posterior estudio en profundidad que confirme o desmienta la capacidad del sensor Kinect V2 para capturar gestos naturales de interacción utilizando las herramientas de inteligencia artificial que se ofrecen con el mismo y posteriormente aplicarlos al entorno de los platós virtuales de televisión de uso comercial.

4.3. Evaluación preliminar del sensor Microsoft Kinect V2 para su uso en platós virtuales de televisión con interacción natural

Para la realización del estudio se ha dispuesto de tres muestras de usuarios diferenciadas: cuatro sujetos sin ninguna experiencia en platós virtuales de televisión, tres usuarios con poca experiencia en este tipo de entornos pero que conocen el medio (estudiantes de Comunicación Audiovisual) y un presentador de televisión profesional experimentado en el uso de estudios virtuales. El tamaño reducido de los grupos viene determinado por la dificultad de conseguir voluntarios con las características requeridas (especialmente el grupo de expertos) y por tratarse de un estudio preliminar que pretende vislumbrar posibles potencialidades del sensor de cara a un estudio de mayor profundidad. Se han definido estos tres grupos diferentes de usuarios para observar posibles diferencias en el tipo de gestualidad y su reacción frente a la tecnología utilizada. A todos los sujetos se les ha pedido que definan cinco gestos que ellos propondrían para lanzar cinco efectos distintos en tiempo real sin poner ningún tipo de condición previa como limitaciones del sensor, partes del cuerpo a utilizar en el gesto, etc. Para ello se les ha dado un tiempo límite de un minuto para la creación del gesto (de forma que este sean lo más intuitivo posible). Se han grabado dos series del sujeto hablando a cámara y repitiendo la acción propuesta 20 veces, intercalando el gesto con su comportamiento natural al expresarse. Posteriormente se ha utilizado una muestra para entrenar al sistema y la otra para evaluar la calidad de detección.

Los efectos a lanzar definidos en el experimento son los siguientes:

1. **Aparición de una gráfica desde el suelo.** Hacer que aparezca de modo creciente un gráfico que aparente estar saliendo del suelo.
2. **Dar paso a un nuevo vídeo dentro de una de las pantallas virtuales presentes en el plató.** Ejemplo de esto es el cambio de mapa durante la presentación de la predicción meteorológica.
3. **Mostrar un vídeo a pantalla completa.** El plató desaparece de la pantalla que es ocupada por un vídeo. Ejemplo de esto es la entrada de una noticia en un telediario.
4. **Hacer aparecer un objeto virtual en la mano del presentador.** El presentador realiza un gesto que provoca la aparición de un objeto virtual (una pantalla, una tablet, etc.) manipulable en su mano.

5. Lanzar una animación de borrasca en el mapa del tiempo durante la presentación de la predicción meteorológica.

Los efectos planteados combinan acciones cotidianas que se pueden ver a menudo en televisión (acciones 2, 3 y 5) con efectos menos comunes (1 y 4) con el fin comprobar si existe alguna diferencia por parte del usuario a la hora de crear gestos específicos. Por otro lado se combina una acción muy específica (hacer aparecer una borrasca en un mapa) con acciones más generales (como dar paso a un vídeo que puede tener cualquier tipo de contenido).

En paralelo a la captura de información por parte del sensor se realizó una observación del comportamiento y reacciones de los usuarios durante el proceso de creación y grabación de los gestos. Finalmente se realizaron entrevistas personales para conocer las impresiones de los usuarios respecto a la idea principal del estudio (el aprendizaje por parte del sistema de los gestos realizados por el usuario) y sus sensaciones durante el proceso de grabación de los mismos.

4.3.1. Resultados

Para el análisis de los resultados se han contabilizado los falsos positivos y falsos negativos aparecidos durante la evaluación de los diferentes gestos al utilizar el resultado del entrenamiento sobre la grabación realizada para el control. Estos resultados se han agrupado de dos formas distintas: por grupo de pertenencia (no expertos, usuarios con conocimientos y experto) y por gesto.

Por grupo de pertenencia los usuarios no expertos han sido los que han presentado un menor número de falsos positivos y falsos negativos global (figura 4.23). Esto viene dado por la simpleza y claridad de los gestos escogidos por este grupo. Intuitivamente han realizado movimientos bien diferenciados unos de otros y alejados de lo que es la expresión habitual de una persona mientras habla, alejándolos de la zona frontal entre el pecho y el abdomen, donde la mayor parte de la gesticulación natural tiene lugar. Esto le permite al sistema diferenciar cuando el usuario está simplemente exponiendo un tema y cuando está realizando un gesto. Los usuarios con conocimientos del medio han realizado gestos más complejos e indefinidos, por lo que el sistema ha tenido más problemas a la hora de identificarlos ya que en ciertas ocasiones se confundían con gestualidades naturales de expresión al hablar a la cámara. Esto se debe a que, pensando en el televidente, buscaban movimientos discretos que no resultaran llamativos o confusos para el espectador. Por otro lado, el experto es el sujeto que logra un

100% de acierto en la detección de más gestos (tres de los cinco), sin embargo, su mayor grado de error en el caso 1 hace que, globalmente, acumule más errores que los usuarios no expertos. No obstante, observando el comportamiento del usuario durante el estudio así como los gestos elegidos se puede afirmar que se encuentra en el punto medio entre la claridad de los usuarios no expertos y el pensar continuamente en el televidente y la capacidad comunicativa del movimiento, logrando altas tasas de acierto por parte del sistema con gestos bien diferenciados, claros y naturales.

En cuanto a la caracterización global de los diferentes gestos, todos son reconocidos con más de un 94% de acierto excepto en el caso 4: “hacer aparecer un objeto virtual en la mano del presentador” (figura 4.24). Esto se explica por la naturaleza de los gestos realizados y las limitaciones del sensor (visibilidad del gesto desde el sensor, claridad del gesto realizado, etc.). Los usuarios tienden a situar ambas manos delante del cuerpo, entre la cintura y el pecho, una zona en la que, al expresarnos naturalmente, acostumbramos a moverlas. De esta forma se dificulta al sistema la diferenciación entre cuando se está realizando un gesto y cuando no, disparando los falsos positivos hasta el 16% en global, pero llegando al 47% en algún caso individual. En cuanto a los tipos de gestos realizados varían desde los que son muy similares en todos los usuarios (caso 1, caso 2 y caso 5) y los que presentan una mayor variabilidad (caso 3 y caso 4). Esto se debe a que los primeros son casos reales que se ven habitualmente en televisión y que tienen una clara similitud con una interfaz táctil con la que hoy en día la

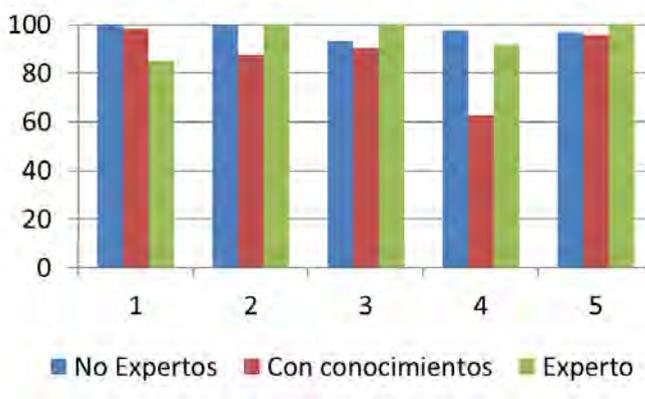


Figura 4.23: Porcentaje de acierto del sistema, por grupo de usuarios, en la detección de los cinco gestos propuestos.

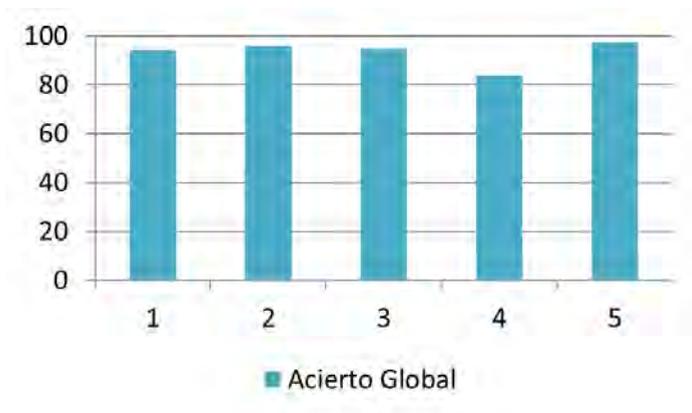


Figura 4.24: Porcentaje de acierto global del sistema en la detección de cada uno de los cinco gestos propuestos.

sociedad está familiarizada (manipular pantallas, arrastrar objetos, etc.), lo que hace que todos los sujetos acaben optando por propuestas similares. Los casos 3 y 4, sin embargo, no sólo presentan gestualidades diferentes en cada usuario, si no que durante la realización del estudio fueron los casos en los que más dudas aparecieron en los sujetos y más tiempo necesitaron para proponer un gesto que les resultase intuitivo. Estos dos casos son casuísticas que ningún usuario había visto (que el presentador de paso a un vídeo en pantalla completa con un gesto y hacer aparecer un objeto virtual en su mano) y por ello presentan una mayor variabilidad, no afectando esto, sin embargo, a la capacidad del sensor para detectar las acciones finalmente propuestas.

Por otro lado cabe destacar, también, que, en general, el sistema produce más falsos positivos que falsos negativos, tendiendo a confundir movimientos naturales de las manos durante la expresión oral con gestos predefinidos. Sin embargo, puede afirmarse que el porcentaje de acierto del sistema es muy alto (mayor del 94% en casi todos los casos) utilizando solamente 20 repeticiones como muestra de entrenamiento (por motivos de tiempo disponible por parte de los sujetos), lo que permite aventurar que con un entrenamiento más intenso (con más muestras de entrenamiento y grabaciones de mayor duración) en un entorno de producción los resultados aquí presentados podrían ser mejorados.

Durante la realización del estudio se mantuvo una entrevista informal con los diferentes sujetos con el fin de conocer su opinión sobre el paradigma de control en plató propuesto.

De entre las opiniones recabadas cabe destacar la del experto, que afirmó que de ser posible la implementación del nuevo flujo de trabajo propuesto en entornos de producción reales supondría un gran avance para el estado actual de la tecnología. Durante la entrevista informó de que para la preparación de un programa de presentación de los resultados electorales el tiempo de ensayos previos necesarios en el plató virtual había sido de cuatro semanas a tiempo completo, lo que le había supuesto un gran nivel de esfuerzo y estrés. Expresó por tanto, su opinión de que el tener una opción que permita reducir estos tiempos facilitando la interacción del presentador era algo realmente necesario en la industria. Asimismo, destacó que la interacción gestual propuesta (que el presentador sea capaz de modificar el entorno en lugar de usar mandos o coreografías previamente entrenadas) supondría un gran avance comunicativo en diferentes tipos de programas, especialmente los informativos (campo en el que él trabaja en la actualidad).

Por tanto, de este estudio preliminar, se deduce que el sensor empleado puede ser apto para su uso en la detección de gestos naturales del presentador en platós virtuales de televisión. Esta primera impresión debe de ser corroborada por un análisis en profundidad con una mayor muestra poblacional (especialmente de expertos) y mayores muestras para el aprendizaje y validación de los gestos en un entorno de producción real.

4.4. Pruebas Integrales del sistema

Una vez finalizada la integración de los diferentes sensores y la actualización del Microsoft Kinect, así como las pruebas de funcionamiento de los mismos, se realizaron tres pruebas integrales en el plató con la finalidad de demostrar el rendimiento de la solución propuesta así como la viabilidad de su uso para el rodaje en directo de programas de televisión.

4.4.1. Rendimiento

La primera prueba se centró en el rendimiento del sistema. Por un lado se evaluó el funcionamiento de la red local y el *middleware* en busca de posibles *lags* introducidos entre la imagen capturada por las cámaras y la imagen generada por ordenador.

Por un lado se analizó el rendimiento de la red formada por siete ordenadores: cuatro servidores VRPN enviando información del sistema de cámaras Optitrack, el Microsoft Kinect, el Leap Motion y el Microsoft Kinect V2, un ordenador hospedando el *middleware* que recibe los datos de los sensores y envía información a eStudio y dos ordenadores de render

(uno por cada cámara del plató) de los cuales el principal recibe información de la capa de aplicación la utiliza y la distribuye al secundario. Con esta configuración de red el ping medio de la conexión entre los diferentes ordenadores se sitúa por debajo de un milisegundo, siendo despreciable, por tanto, el retardo introducido para el rendimiento final del sistema.

Por otro lado se utilizó esta misma configuración para analizar el tiempo necesario para procesar, por parte del *middleware*, los datos de los sensores y enviarlos al software de *render*. La capa de aplicación recibe los datos obtenidos por los sensores, realiza la transformación de los mismos a coordenadas globales del plató, calcula la posición relativa de los diferentes objetos virtuales, actualiza los valores de los *maps* y lanza efectos en el mundo virtual provocados por acciones en el mundo real. Este procesado de la información de cuatro sensores en paralelo se realiza en menos de un milisegundo siendo, por tanto, un retardo despreciable en la integración final de las imágenes generadas.

Finalmente, eStudio realiza la actualización de los valores de los *maps* 100 veces por segundo, siendo esta una velocidad de refresco mayor que la de la mayoría de los sensores presentes en el sistema y superior al tiempo real exigido por los platós virtuales de televisión.

Por todo esto, puede afirmarse que el rendimiento global del sistema es suficiente para la realización de grabaciones en tiempo real. Para demostrar esta afirmación se utilizó el plató en la grabación en tiempo real de dos episodios para dos programas distintos en los que se utilizaron todos los sensores integrados en la arquitectura.

4.4.2. Caso de prueba I: “Copycat”

El primer ejemplo de uso del plató virtual de trabajo fue la grabación de parte del episodio piloto para la serie web “Copycat” (figura 4.25). Los dos fragmentos del mismo han sido grabados utilizando la tecnología virtual fueron un programa informativo que daba inicio al episodio y una pequeña píldora informativa en la que una presentadora informa de una nueva tecnología de audio utilizada en la grabación del mismo que requería, por parte del espectador, el uso de cascos.

Los sensores utilizados en estas dos secuencias fueron los siguientes:

1. **Sistema de cámaras Optitrack:** fue utilizado para realizar el seguimiento del movimiento de las cámaras.
2. **Microsoft Kinect:** utilizado para controlar la posición relativa de la presentadora y un objeto virtual.

3. **Leap Motion:** utilizado en el programa informativo para que la presentadora, a través de un gesto *swipe*, diese paso a la siguiente noticia.

Las dos partes del capítulo fueron grabadas en directo y su incorporación al montaje final no requirió de ningún tipo de postproducción. Esto demuestra la utilidad del uso de estas tecnologías para ahorrar tiempo de producción y por tanto coste final del producto también en emisiones en diferido.



Figura 4.25: Fotograma del episodio piloto de la serie web Copycat en el que la presentadora manipula una fotografía a través de una interfaz tangible

4.4.3. Caso de prueba II: “La universidad responde”

Se realizó también la grabación de un episodio para “La universidad responde”, un programa coordinado por los diferentes servicios de medios audiovisuales de diferentes universidades españolas en el que los investigadores muestran los resultados de su trabajo de un modo divulgativo para acercar el mundo científico al público en general (figura 4.26). El formato se basa en una pregunta inicial que el investigador responde en aproximadamente dos minutos de tiempo. En este caso la pregunta fue: ¿Qué elementos se pueden sensorizar en un plató virtual de TV y para qué sirven estas sensorizaciones? Para responderla se utilizaron los siguientes sensores:

1. **Sistema de cámaras Optitrack:** fue utilizado para realizar el seguimiento del movimiento de las cámaras y para realizar el seguimiento de una interfaz tangible manipula-

da por el investigador, asociada a una pantalla virtual.

2. **Microsoft Kinect V2:** fue utilizado para tres tareas: controlar la posición relativa entre el presentador y un objeto virtual, realizar una manipulación directa de un objeto virtual (la pantalla antes manipulada a través de una interfaz tangible) a través de la asociación de este a la mano izquierda del actor y la detección de gestos previamente aprendidos por el sistema empleando la metodología anteriormente presentada. Estos gestos lanzaban acciones como el cambio de la imagen mostrada en una de las pantallas o la asociación de un objeto virtual a la mano del presentador y su posterior vuelta a su posición inicial.

Estos dos ejemplos exploran una mínima parte de las potencialidades que presenta la combinación sinérgica de sensores en un plató virtual de televisión. La capacidad de incluir nuevos dispositivos y tecnologías en la arquitectura propuesta de una forma ágil y sencilla abre un mundo de posibilidades creativas relacionadas con la interacción entre los mundos virtual y real. Además, se ha demostrado la viabilidad de la utilización de sensores de bajo



Figura 4.26: Fotograma del episodio de “La universidad responde” en el que el actor realiza un gesto que es detectado por el sensor Microsoft Kinect V2

coste para tareas específicas, lo que permite acercar los platós virtuales de televisión a canales de bajo presupuesto que hasta ahora no podían plantearse el uso de esta tecnología.

4.5. Conclusiones

Se ha diseñado, implementado y probado una nueva arquitectura distribuida para platós virtuales de televisión que permite agilizar la inclusión de nuevos dispositivos. La utilización de conexiones de red hace que la arquitectura sea independiente tanto del sistema operativo de los diferentes componentes como del hardware necesario para ejecutarlo, permitiendo evitar los problemas tradicionales de compatibilidad de los diferentes dispositivos con los softwares de *render* propietarios. La utilización de esquemas de configuración XML junto con las librerías VRPN permite la incorporación de servidores de terceros de una forma directa sin la necesidad, por parte del usuario, de implementar ningún servicio, reduciendo de forma notable el tiempo necesario para la incorporación de un dispositivo al sistema. En este sentido se ha implementado una interfaz gráfica intuitiva que hace accesible la configuración de los dispositivos involucrados en un plató virtual de televisión para usuarios no expertos.

Se han añadido tres sensores (Cámaras de infrarrojos Optitrack, Microsoft Kinect y Leap Motion) y se ha actualizado uno (Microsoft Kinect V2) en un plató de trabajo situado en la Facultad de Ciencias de la Comunicación de la Universidad de Santiago de Compostela. Esto aumentado las capacidades del mismo incluyendo el movimiento libre de cámaras, seguimiento de objetos y tres tipos distintos de interacción para los actores (*distance keying*, interacción directa y detección de gestos). Esta nueva configuración permite, también, la automatización de una serie de procesos, lo que facilita la grabación de programas en directo.

Finalmente, se ha propuesto un nuevo flujo de trabajo para la preparación de emisiones en platós virtuales basado en el aprendizaje (a través de técnicas de inteligencia artificial) por parte del sistema de los gestos naturales del presentador, reduciendo de esta forma los tiempos de ensayo necesarios y reduciendo el estrés del presentador en la emisión en directo.

Como resultado del diseño implementación y prueba de esta nueva arquitectura y flujos de trabajo para platós virtuales de televisión se presentaron artículos en el XXII Congreso Español de Informática Gráfica [131], en el VI Congreso de TV Digital Interactiva - IV Jornadas Iberoamericanas de Aplicaciones y Usabilidad de la TVDi 2015 [128] y en el XVII Congreso Internacional de Interacción Persona-Ordenador [127].

CAPÍTULO 5

VISUALIZACIÓN AVANZADA EN PLATÓS VIRTUALES DE TELEVISIÓN

Como se ha comentado previamente, la calidad de la integración en un plató virtual de televisión depende tanto de la interacción entre los mundos real y sintético como de la calidad visual del *inset*. Esta segunda característica depende principalmente de dos factores: la correcta alineación y calibración de las cámaras virtuales y el realismo del *render* obtenido. El primer factor depende, principalmente, de la calidad de la sensorización con la que esté equipado el plató. El segundo está directamente relacionado con el realismo o verosimilitud, así como de la coherencia visual de la imagen generada por ordenador. Este siempre es necesario ya que, aunque el entorno 3D no busque parecerse a la realidad, a la hora de obtener una buena integración será necesario que al menos la iluminación del mundo real y virtual sean coherentes.

Es en este punto en el que entran en juego tanto la calidad del motor gráfico como las técnicas de visualización avanzada que nos permite emplear. Dentro de estas técnicas destacan por su importancia aquellas relacionadas con la iluminación de escenas 3D y, más concretamente, las enfocadas a simular la luz real del escenario.

5.1. Iluminación en platós virtuales de televisión

A la hora de iluminar del escenario real se han de tener en cuenta dos grupos de elementos: uno formado por los actores, *atrezzo* y escenografía y otro formado únicamente por el ciclorama. Los primeros han de ser iluminados conforme al tono que se le quiera dar al producto final, ya que la luz es una parte de capital importancia en la ambientación de los programas de televisión. La correcta colocación y orientación de los focos resulta clave a la hora de reducir el ruido producido en la imagen final, por reflejos y sombras, en el proceso de *chroma keying*. Por otro lado, la iluminación es una de las características clave a tener en cuenta a la hora de diseñar un ciclorama que envuelva el escenario. Para que el sistema sea capaz de eliminar de forma óptima el fondo de la escena este debe de recibir una luz uniforme y difusa, evitando sombras y brillos que causen variaciones de color que dificulten la aplicación de la clave de color. También debe de tenerse en cuenta la forma redondeada del ciclorama a la hora de situar y calibrar la intensidad de los focos en el plató. Asimismo, deben reducirse en la medida de lo posible las sombras proyectadas por los actores y la escenografía, ya que estas suponen un oscurecimiento y, por tanto, un cambio de color en el ciclorama y podrían provocar la aparición de manchas o ruido en la composición final de la imagen. Como regla general, para obtener un buen resultado ha de crearse una diferencia de iluminación que ayude a incrementar la diferencia de color, separando a los sujetos del fondo [86]. Para conseguir este efecto se utilizan dos configuraciones distintas: una para la escena basada en el esquema por defecto de un plató tradicional con luz frontal, lateral y contra [163] y otra enfocada exclusivamente al ciclorama. Para la escena debe buscarse una luz ni muy fuerte ni muy suave, obtenida, generalmente, mediante la utilización de focos *Fresnel* [118]. Si se desean efectos más complejos existe también la alternativa de utilizar proyectores como fuentes lumínicas [18]. Para el fondo se utilizan habitualmente luminarias grandes y blancas, generalmente tubos fluorescentes. Lo ideal es hacerlo con tubos específicos que saturan el verde o el azul, y eliminan las dominantes rojas [7]. Para asegurar que la escena se separa correctamente del fondo, el ciclorama ha de recibir una intensidad de, aproximadamente, 2/3 de la utilizada para los actores, *atrezzo* y escenografía [49].

Por otro lado, uno de los puntos de mayor importancia para que la integración entre el mundo virtual y el real sea coherente es la iluminación de escenario 3D. Las luces que dan color a los actores y escenografías deben de tener su correspondencia con las que iluminan los modelos sintéticos para que la sensación de presencia en el entorno sea efectiva. Asimismo, los actores deben de proyectar sombras realistas sobre los objetos virtuales ya que de forma

intuitiva, el cerebro humano las utiliza para determinar exactamente en qué punto un cuerpo entra en contacto con otro. La ausencia de estas produce un efecto de no contacto entre los elementos virtuales y reales, dando la impresión de que los actores “vuelan” sobre el entorno tridimensional. Sin embargo, obtener una iluminación realista es un proceso muy costoso computacionalmente y puede comprometer la tasa de refresco en tiempo real. Por ello es necesario la utilización de técnicas de visualización avanzada para simular efectos complejos de una forma eficiente y efectiva.

5.1.1. Análisis de iluminación de cicloramas

Como se ha subrayado en el apartado anterior la calidad de la iluminación del ciclorama tiene una importancia capital en el resultado obtenido de aplicar la técnica del *chroma key*. Para que en la imagen real grabada por las cámaras los bordes entre el fondo y la escena sean claros y definidos y las variaciones de color en el ciclorama mínimas han de disponerse una serie de focos específicamente situados para lograr una luz difusa y uniforme. Aun disponiendo de un equipamiento especializado su colocación e intensidad puede variar a lo largo del tiempo, produciendo variaciones lumínicas con consecuencias directas en la calidad de la composición final.

Para evitar estas variaciones es necesario realizar un análisis de la luminosidad¹ del ciclorama periódicamente. En la realización estos estudios se suelen emplear sensores específicos, pero resultan procesos tediosos (al tener que ir tomando numerosas muestras a lo largo de todo el escenario) que generalmente han de ser llevados a cabo por expertos. Por ello se ha implementado una solución visual, rápida y efectiva que permite dar una orientación a usuarios no expertos sobre la correcta colocación y calibración de la iluminación del fondo de un plató virtual de televisión.

5.1.1.1. Implementación

Existen numerosas técnicas en el campo de la visualización avanzada en entornos 3D que permiten extraer la iluminación de una escena real y mapearla sobre una virtual para obtener *renders* realistas [123]. Todas ellas realizan un análisis de la distribución lumínica de la escena real y, por tanto, pueden ser utilizadas para comprobar la correcta iluminación del ciclorama de un plató virtual. Muchas de estas técnicas subdividen una imagen del entorno en zonas re-

¹Luminosidad se define como la cantidad de energía lumínica que llega a un área por segundo [139]

presentativas que posteriormente son utilizadas para dar luz a escenas sintéticas [98, 4, 149]. De entre estas técnicas destaca la de muestreo de corte medio, propuesta por Paul Debevec en [40], que realiza una clasificación de la imagen en zonas rectangulares para posteriormente llevar a cabo un muestreo que permite generar un número finito de luces puntuales capaces de simular la iluminación real de la escena. Esta solución se encuentra ampliamente referenciada en la literatura y destaca frente a otras propuestas por su bajo coste computacional y la calidad del muestreo obtenido. Para que el modelo de iluminación generado sea correcto, las luces puntuales obtenidas deben de tener una distribución en el espacio 3D coherente con la imagen original del mismo. En el muestreo de corte medio, estas posiciones se obtienen de la subdivisión de la imagen en regiones de igual energía, dando como resultado un algoritmo simple y eficiente que puede representar escenarios complejos con tan solo 64 puntos de luz. Esta capacidad de dividir imágenes en regiones con la misma energía es la utilizada en la propuesta aquí presentada para realizar un análisis visual de la distribución uniforme de la luz en el ciclorama de un plató virtual de televisión.

El algoritmo utiliza imágenes e alto rango dinámico (HDR), una técnica que permite la codificación de un rango de luminancia más amplio entre las zonas iluminadas y las zonas oscuras de una imagen que en las técnicas digitales tradicionales. Esto hace posible representar, de una forma más precisa, la amplia variedad de niveles de luz existentes en el mundo real, que pueden ir desde la luz directa del sol o un foco a un punto totalmente en la sombra [138]. Se empleó esta técnica porque permite capturar de una forma fidedigna las diferencias de luminancia entre diferentes zonas de la imagen, objetivo final de la herramienta que se estaba desarrollando. Para obtener estas imágenes es necesario realizar una serie de fotografías con diferentes niveles de apertura para capturar las diferentes intensidades de iluminación de la escena [39]. Estas imágenes se combinan en una sola que incluye el rango de luminancia capturado al completo. Para llevar a cabo este proceso se ha utilizado el software Adobe Photoshop [112], Luminance HDR [1] y hdrshop [30].

El cálculo de la energía total se realiza sobre una versión monocroma de la imagen en la que se almacena la energía lumínica de la misma, en lugar de sobre los colores originales RGB de la fotografía HDR (figura 5.1). Esta imagen en escala de grises está formada por píxeles que almacenan la información de luminancia de cada punto y se calcula mediante la media ponderada de los canales de color siguiendo los pesos recomendados por el autor, basados en la ITU-R BT.709 (*International Telecommunication Union Recommendation Broadcasting Series*

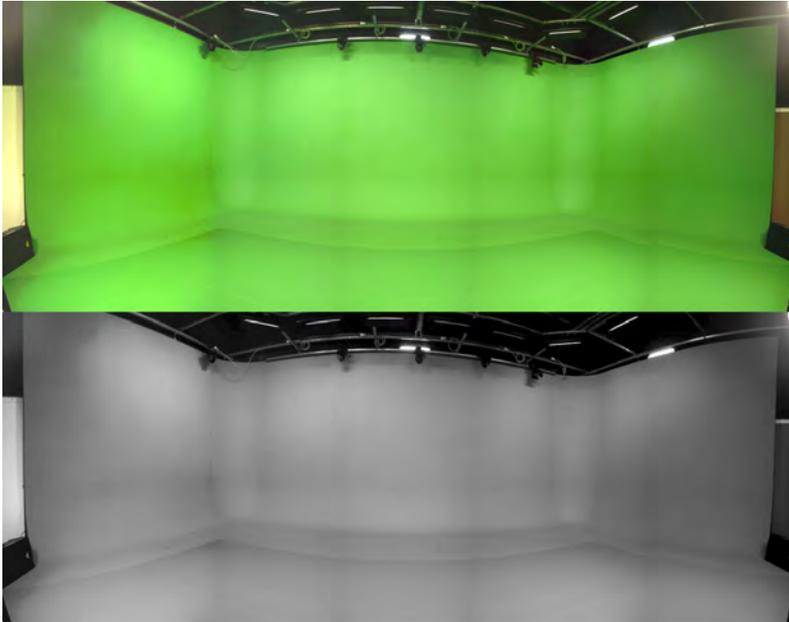


Figura 5.1: Imagen HDR del escenario (arriba) e imagen de luminancia correspondiente (abajo)

Television) (ecuación 5.1).

$$Y = 0,2125 * R + 0,7154 * G + 0,0721 * B \quad (5.1)$$

Para la subdivisión de la imagen, el muestreo de corte medio se inspira en el algoritmo de corte medio de la cuantificación de color de Paul Heckbert [73] y por ello realiza el particionado de una imagen en formato rectangular latitud-longitud en 2^n regiones de similar energía lumínica de la siguiente forma:

1. Se introduce la imagen al completo en una lista de regiones como una única región.
2. Por cada región de la lista, se subdivide a lo largo de la dimensión más larga (latitud o longitud) de forma que la energía de las dos nuevas regiones sea la misma.
3. Si el número de iteraciones es menor que n (para 2^n regiones), se repite el paso 2.

4. Se sitúa un punto de luz en el centro o baricentro de cada región y se le asigna el color medio de los píxeles de la misma.

Computacionalmente, el punto más costoso del proceso es el cálculo de los límites de las diferentes regiones. Para acelerarlo se utiliza la técnica de la tabla de área sumada o *summed area table* (SAT) [31]. Una SAT es una estructura que permite calcular, de forma eficiente, la suma de valores de un subgrupo rectangular en una matriz bidimensional, aumentando así el rendimiento global del algoritmo. Se trata de una tabla en la que el valor de cada punto (x,y) corresponde a la suma de todos los puntos situados más arriba y a la izquierda sumados con el propio punto (x,y) (ecuación 5.2). El cálculo de estos valores es poco costoso desde un punto de vista computacional ya que el valor de cada punto depende únicamente de otros cuatro: el situado a su izquierda, el situado en diagonal, el situado encima y el propio punto objeto de cálculo (ecuación 5.3). Una vez se ha completado esta tabla resulta trivial computar el peso de cualquier región rectangular de vértices A, B, C y D (figura 5.2) ya que, como se observa en la ecuación 5.4, únicamente es necesario utilizar cuatro valores de la SAT.

$$sat(x,y) = \sum_{x' \leq x, y' \leq y} i(x',y') \quad (5.2)$$

$$sat(x,y) = i(x,y) + sat(x-1,y) + sat(x,y-1) - sat(x-1,y-1) \quad (5.3)$$

$$\sum_{A(x) < x' \leq C(x), A(y) < y' \leq C(y)} i(x',y') = sat(A) + sat(D) - sat(B) - sat(C) \quad (5.4)$$

Por tanto se calcula la tabla de área sumada de la imagen de luminancia para simplificar el punto número 2 del proceso del algoritmo de corte medio. Posteriormente se divide la imagen en regiones de igual energía, repitiendo el proceso hasta que el número de zonas rectangulares

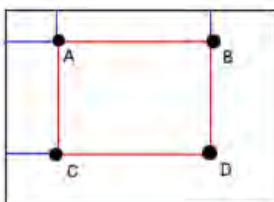


Figura 5.2: Área formada por los vértices A, B, C y D

es el deseado. La cantidad de áreas fruto de estas subdivisiones será siempre múltiplo de dos ya que en cada paso del algoritmo se dividen a la mitad todas las obtenidas en la anterior iteración, multiplicando por dos su número.

Una vez calculadas las 2^n regiones es necesario definir la posición exacta de la luz puntal que las representará. Esta se sitúa en el baricentro de cada área rectangular, es decir, en el píxel que, dentro de la región, tiene a su izquierda la misma luminancia que a su derecha y en su parte superior la misma que en su parte inferior. El cálculo de este píxel se realiza mediante una iteración, primero en la dimensión horizontal (componente x) hasta encontrar el punto medio, y, en segundo lugar, en la dimensión vertical (componente y). El punto (x, y) resultante se sitúa en el baricentro de la región. Finalmente, el color del punto de luz que representa a cada área se calcula como la media del color de todos los píxeles que la forman. Por tanto, la subdivisión en regiones de igual energía y el cálculo de la posición de las luces se realiza sobre la imagen de luminancia y el del color de los puntos de luz sobre la imagen original en color.

De cara a facilitar la utilización del algoritmo por parte de usuarios no expertos se implementó una aplicación en C++ [29] haciendo uso de Windows Forms [134] para la interfaz de usuario y OpenCV [84] para el procesamiento de las imágenes (figura 5.3). La aplicación resultante permite la utilización de imágenes HDR en formato “OpenEXR” y la exportación, tanto del resultado visual final como de la imagen de luminancia, en los formatos “jpg” y “png”. Así

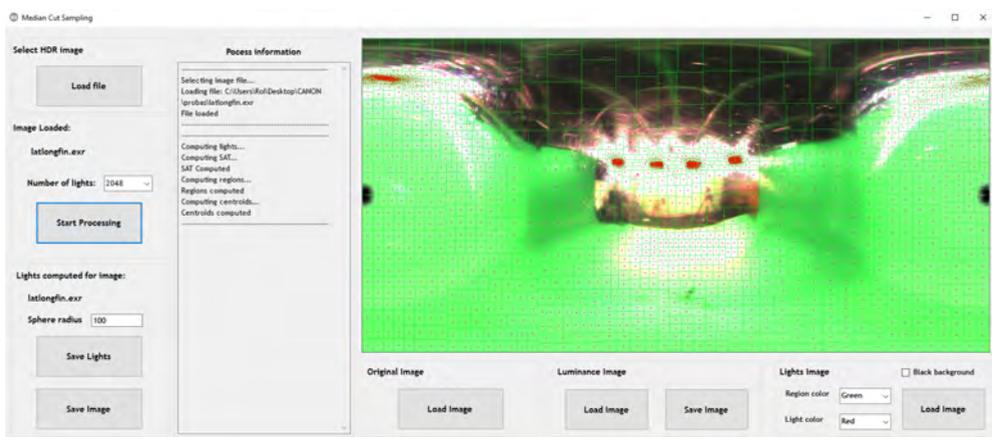


Figura 5.3: Interfaz gráfica de la aplicación implementada para la ejecución del algoritmo de muestreo de corte medio

mismo, permite de seleccionar el número de regiones en que se subdividirá la imagen (entre 2 y 2048), así como los colores que se utilizarán para mostrar las regiones y sus baricentros de cara a una mejor visualización del resultado.

5.1.1.2. Pruebas

El software fue implementado durante el proceso de diseño y configuración del plató virtual de televisión de la Facultad de Ciencias de la Comunicación, por lo que fue probado tras su montaje, en la fase de calibrado. Como se ha comentado en la introducción, la iluminación del ciclorama debe de ser uniforme para evitar cambios de color provocados por zonas con distinta luz que puedan afectar al proceso de *chroma keying*. Desde el punto de vista de un análisis visual, la imagen de un escenario de *chroma* procesado utilizando el algoritmo de muestreo de corte medio debería de tener como resultado una cuadrícula perfecta de regiones en las cuales la posición de sus baricentros coincidiera con el centro geométrico de las mismas. La aparición de diferencias substanciales en el tamaño de las regiones implicaría la existencia de zonas del ciclorama con mayor energía lumínica que otras y, por tanto, diferen-

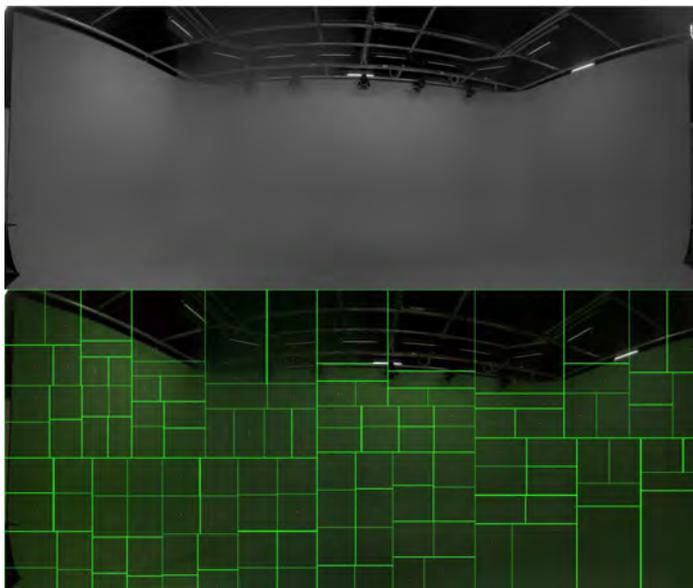


Figura 5.4: Primer análisis de iluminación del plató. Imagen de luminancia (arriba) y áreas de igual energía lumínica (abajo)

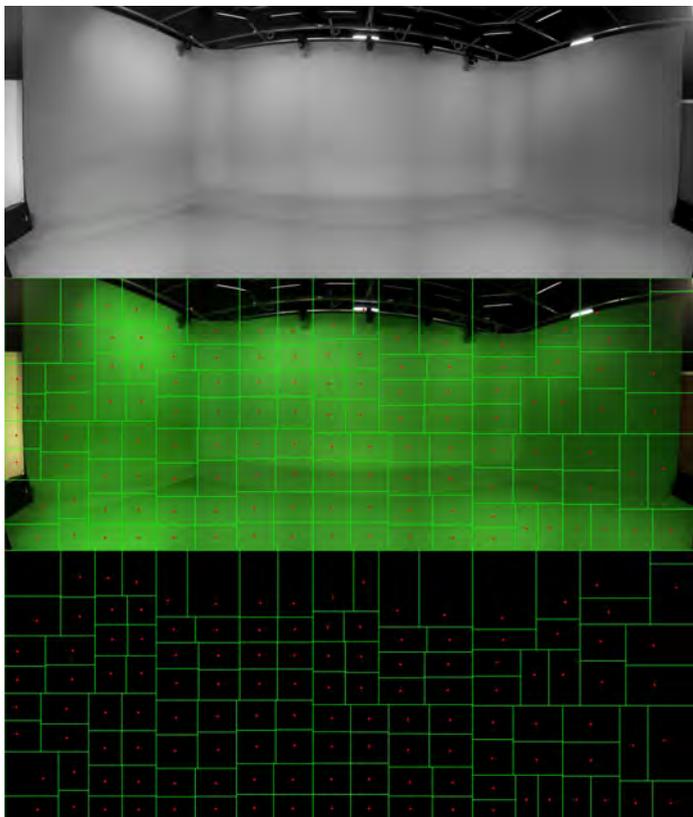


Figura 5.5: Segundo análisis de iluminación del plató. Imagen de luminancia (arriba) y áreas de igual energía lumínica (abajo)

cias de iluminación que podrían comprometer un correcto discernimiento entre el fondo y la escena que se desea conservar en la imagen final.

Al finalizar el proceso de instalación del plató se obtuvo una imagen panorámica HDR del escenario y se aplicó el algoritmo de muestreo de corte medio sobre la misma. El resultado obtenido, como se puede observar en la figura 5.4, mostró una cuadrícula irregular, denotando diferencias importantes en la iluminación en varias zonas del ciclorama. Esta información visual generada por el software, fue utilizada para la recolocación y calibración de los focos, buscando obtener la luz uniforme deseada. Tras el cambio en la configuración se repitió el proceso. Se obtuvo una imagen panorámica HDR y se procesó obteniendo como resultado

una cuadrícula mucho más regular que la del primer análisis, en la que los baricentros de las diferentes regiones se sitúan en el centro de las mismas 5.5. Finalmente, de cara a probar la validez de la solución propuesta, un experto realizó un análisis de la iluminación utilizando sensores específicos, confirmando la correcta configuración lumínica del ciclorama. La sencillez del proceso, que puede ser llevado a cabo por usuarios no expertos, así como su rapidez lo hacen ideal para la realización de análisis periódicos que permitan detectar variaciones en la luz del escenario que pudiesen comprometer el correcto funcionamiento del proceso de *chroma keying*.

5.1.2. Iluminación de entornos virtuales

Además de la iluminación del escenario, en un plató virtual de televisión, es de crucial importancia la calidad y coherencia del *render* generado por los ordenadores. La calidad viene dada por el modelado, los materiales, las texturas y los algoritmos utilizados mientras que la coherencia viene dada, principalmente, por la alineación de las cámaras virtual y real y similitud entre la luz en el mundo real y el sintético.

Desde sus inicios la obtención de imágenes sintéticas con una iluminación realista, en tiempo real, ha sido uno de los principales retos de los platós virtuales de televisión. A lo largo de la historia, en el campo de la visualización avanzada, se han utilizado diversas técnicas como el trazado de rayos [158] o la radiosidad instantánea [102] para poner una solución a este problema. Sin embargo, estos algoritmos no están libres de artefactos y presentan un difícil equilibrio entre la velocidad de *render* y la calidad del mismo, lo que podría comprometer su ejecución en tiempo real en ciertos escenarios. Por estos motivos y pese a ser utilizados ampliamente en la industria del videojuego en los últimos años, su uso no se ha extendido a los platós virtuales de televisión. Ninguna de las soluciones comerciales actuales, como Brainstorm eStudio, ofrece algoritmos de iluminación global que permitan obtener en directo imágenes sintéticas realistas. En la actualidad una de las técnicas más extendidas para la simulación de este efecto en tiempo real es el *many-light rendering* o “*render* con muchas luces” [36]. Esta solución ofrece un marco matemático para reducir el problema de la simulación de iluminaciones globales realistas al cálculo de la iluminación directa desde numerosos puntos de luz virtuales. Frente a los algoritmos tradicionales basados en el transporte de luz, el *many-light rendering* ofrece un equilibrio entre la calidad y la velocidad que hace posible obtener imágenes sin artefactos en una fracción de segundo [99].

La similitud entre la luces presentes en el mundo real y el virtual se puede conseguir

mediante dos aproximaciones: la simulación en el escenario de la configuración lumínica del mundo virtual [59] o la captación de la iluminación real y su posterior mapeado al mundo sintético. Esta segunda metodología permite la no modificación de la colocación y calibración de los focos en el plató, haciendo posible mantener una configuración estable y evitando variaciones que pudiesen comprometer la calidad del *chroma keying*.

Se propone, por tanto, un nuevo flujo de trabajo que permita capturar la iluminación real del escenario y aplicarla al entorno virtual, igualando la luz en ambas imágenes y aumentando, de este modo, la coherencia visual de la integración final.

5.1.2.1. Implementación

El flujo de trabajo propuesto se basa en tres pasos: la obtención de una imagen HDR de 360 grados del escenario, la utilización del algoritmo de muestreo de corte medio sobre esta para obtener la posición de las luces puntuales representativas, y la aplicación de las mismas sobre la escena sintética.

1. **Obtención de una imagen HDR de 360 grados del escenario** El método más común para la obtención de una imagen HDR de 360 grados de un entorno, cuando no se dispone de un sensor específico, es la utilización de una esfera reflectante pulida. El primer paso del proceso consiste en la obtención de una serie de fotografías de la misma con diferentes grados de exposición (figura 5.6). Posteriormente, estas fotografías

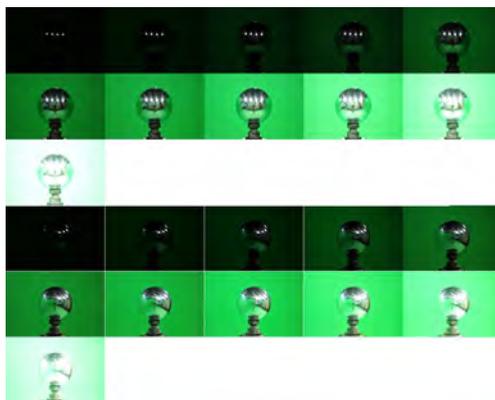


Figura 5.6: Series de diez imágenes tomadas con diferentes niveles de exposición para la generación de las imágenes HDR en ángulo de 90 grados

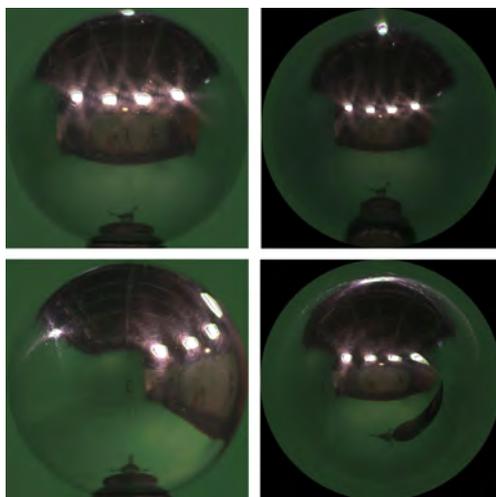


Figura 5.7: Imágenes HDR originales (izquierda) y mapeadas con rotación (derecha)

se combinan para obtener una imagen HDR. Este proceso puede ser realizado mediante la utilización de diferentes paquetes de software. En el planteamiento del presente flujo de trabajo se ha utilizado del programa Luminance HDR ya que incluye la técnica de generación de imágenes de alto rango dinámico desarrollada por Paul Debevec, autor del algoritmo de muestreo de corte medio que se aplicará posteriormente sobre las mismas. El resultado de combinar las fotografías de la figura 5.6, utilizando esta técnica, puede observarse la figura 5.7. En el centro de las imágenes originales (figura 5.7) puede observarse la presencia de la cámara y el trípode que sostiene la esfera. Este objeto ocluye parte del entorno y puede provocar la pérdida, en la imagen, de parte de las condiciones lumínicas del entorno. Así mismo, las partes reflejadas en el borde de la esfera se deforman y distorsionan perdiendo parte de la información y dando como resultado una imagen de baja calidad cuando se realiza mapeado para convertir la fotografía original en un mapa de entorno. De cara a disminuir esta problemática es necesario tomar dos fotografías de la esfera desde diferentes ángulos para, al combinarlas, eliminar la presencia de la cámara y reducir la pérdida de información en los bordes (figura 5.6). El ángulo ideal entre estas dos imágenes es de noventa grados, ya que en caso de situarse la cámara en la parte posterior de la esfera (que intuitivamente podría parecer la solución más acertada) esta se encontraría en la región de pérdida de información de la

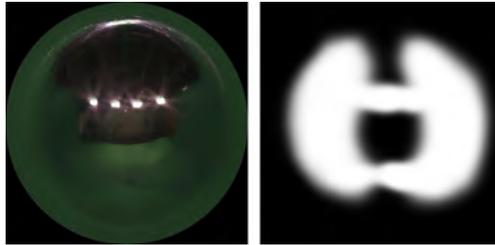


Figura 5.8: Resultado final combinando dos imágenes a 90 grados y máscara utilizada en la unión

fotografía original y viceversa. Un ejemplo de estas dos imágenes puede observarse en la figura 5.7.

Para la combinación de las dos imágenes, la eliminación de la cámara de las mismas, la reducción de la pérdida de información en los bordes y la creación del mapa de entorno se ha hecho uso del software HDRShop. El proceso a seguir es el siguiente:

1. Eliminación del fondo de la imagen recortando la esfera en ambas fotografías (figura 5.7).
2. Alineación de las dos fotografías y cálculo de la rotación exacta entre ambas. Para facilitar el proceso se le indica al software la posición dos puntos comunes en ambas imágenes. Una vez que el ángulo es conocido se aplica una rotación 3D para alinearlas (figura 5.7).
3. Combinación de ambas imágenes. Para ello se utiliza una máscara a la que se le asigna el valor 1 en las zonas en las que se conservará la imagen original sin rotación, el valor 0 en las que se utilizará la imagen rotada y valores intermedios en la transición entre ambas para conseguir una unión continua (figura 5.8). Esta máscara puede ser creada mediante cualquier programa de edición de imagen (en proceso de diseño de este flujo de trabajo se utilizó el software Adobe Photoshop). Finalmente, HDRShop permite la combinación de las tres imágenes utilizando la técnica de *alpha blending*, en el que, siendo las imágenes A y B las dos vistas de las esferas y C la máscara, se combinan utilizando la operación 5.5. El resultado es una imagen en la que se ha eliminado la cámara y se ha reducido la pérdida de información en los bordes de la esfera (figura 5.8).

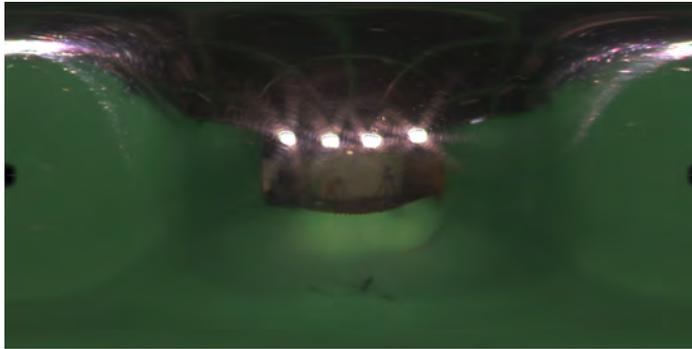


Figura 5.9: Imagen “Latitud-Longitud” obtenida en el proceso de captura de la iluminación del entorno

$$I = A * C + B * (1 - C) \quad (5.5)$$

4. Mapeado de la imagen de la esfera a un mapa de entorno aplicando sobre ella la transformación panorámica “Latitud-Longitud” de HDRShop, dando como resultado una imagen rectangular que, al ser proyectada sobre un objeto esférico en un entorno 3D, produce una cúpula que permite observar el entorno real (figura 5.9).

2. Obtención de la posición de las luces virtuales

Para la obtención de la posición de las luces en el escenario virtual se utiliza el algoritmo de muestreo de corte medio sobre el mapa de entorno generado en el paso uno. A través de este procesado se calcula la posición y el color de cada luz correspondientes al baricentro y la media de colores de cada región respectivamente (figura 5.10). Los puntos obtenidos están situados sobre la imagen y, por tanto, se encuentran en el espacio 2D, por lo que es necesario mapearlos a un entorno 3D que permita integrarlos en la escena virtual. Como el mapa de entorno se obtiene a partir de una esfera, la geometría sobre la que se deben de mapear las luces es esférica (figura 5.11).

$$(x, x_0)^2 + (y, y_0)^2 + (z, z_0)^2 = r^2 \quad (5.6)$$

$$\begin{aligned}
 x &= x_0 + r * \sin \theta * \cos \varphi \\
 y &= y_0 + r * \sin \theta * \sin \varphi \quad (0 \leq \varphi \leq 2\pi, 0 \leq \theta \leq \pi) \\
 z &= z_0 + r * \cos \theta
 \end{aligned}
 \tag{5.7}$$

La ecuación de la esfera (ecuación 5.6) permite realizar este cálculo utilizando el mapeado mostrado en las ecuaciones 5.7 para una esfera de radio r y centro (x_0, y_0, z_0) . Para conocer el tamaño de cada pixel en radianes se dividió el número de pixeles horizontales y verticales por π y por $\pi/2$ respectivamente. Los ángulos φ y θ pueden ser calculados multiplicando la posición (x, y) del baricentro por el número de radianes por pixel. De este modo se obtienen posiciones 3D para las luces extraídas de la imagen en 2D.

Se incorporó este cálculo a la aplicación que implementa el algoritmo de muestreo de corte medio. El software permite seleccionar el radio de la esfera de mapeo de forma manual así como la exportación de las posiciones 3D y el color de las luces a un archivo de texto para su posterior utilización.

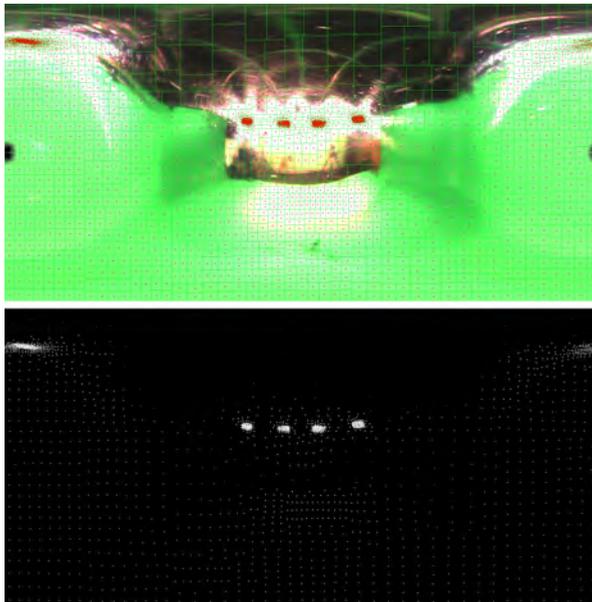


Figura 5.10: 2048 luces obtenidas del procesado de la imagen

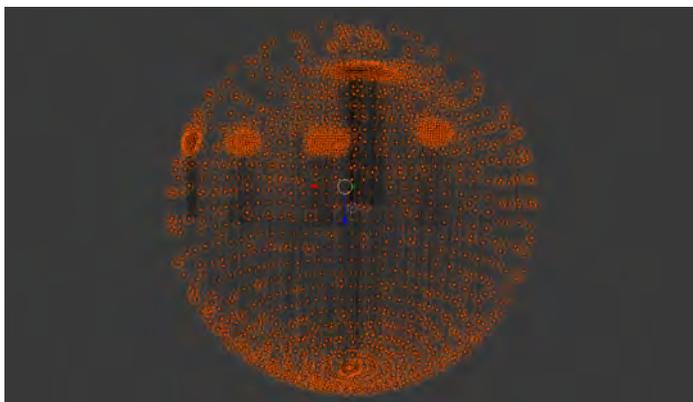


Figura 5.11: Mapeado de las 2048 luces obtenidas sobre una esfera

3. Aplicación de las luces sobre el entorno virtual

Las luces obtenidas en los puntos 1 y 2 pueden ser utilizadas en un entorno 3D, principalmente, a través de dos técnicas: la utilización de algoritmos de iluminación global a partir de muchas luces en eStudio o el *baking* de texturas [66] durante el proceso de creación del escenario. La primera opción requeriría la implementación por parte de Brainstorm Multimedia de los algoritmos de visualización avanzada necesarios soportar el “*render* con muchas luces” en tiempo real, siendo esta una solución que no puede ser implementada por el usuario final pero que permitiría simular, incluso, variaciones de luz en tiempo real. La segunda opción consiste en preprocesar las texturas y materiales de los objetos virtuales para que almacenen información sobre la iluminación global de entorno, como se puede observar en la figura 5.12. Este proceso se denomina *baking* de texturas y es ampliamente utilizado en el *render* en tiempo real de escenarios complejos, ya que permite almacenar en los materiales y texturas de los objetos información sobre su geometría y la iluminación global de la escena. De esta forma, si la luz no varía en el tiempo, se pueden obtener visualizaciones complejas sin afectar al rendimiento.

La combinación del “*render* con muchas luces”, junto con otras técnicas como el *ambient occlusion* [164], y el *baking* de texturas permite obtener visualizaciones realistas de escenarios sintéticos en tiempo real. Si la iluminación es capturada en el escenario del plató, se obtendrá una unificación de la luz de los mundos real y virtual, mejorando de esta forma la calidad de la integración de la imagen final.

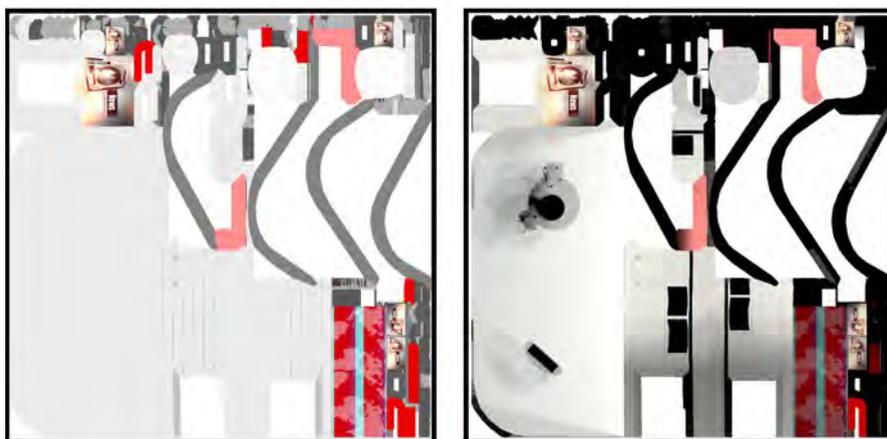


Figura 5.12: Textura sin información adicional (izquierda) y textura con información de iluminación global tras un proceso de *baking* (derecha)

5.1.2.2. Pruebas

Este flujo de trabajo se probó mediante su puesta en práctica en el plató virtual de la Facultad de Ciencias de la Comunicación. En primer lugar se realizó la captación la luz del escenario mediante una imagen HDR de 360 grados. A través el procedimiento anteriormente descrito se obtuvo un mapa de entorno sobre el que se aplicó el algoritmo de muestreo de corte medio haciendo uso de la aplicación desarrollada en el marco de la presente tesis. Se obtuvieron mapeos de 64, 256, 512, 1024 y 2048 luces representativas del entorno sobre una esfera de cinco metros de radio, que fueron exportados a archivos de texto para su importación en programas de modelado 3D.

El procesado de los modelos 3D se realizó en Blender [54], un entorno de software libre para modelado y animación 3D que permite, entre otras funcionalidades, la utilización de los algoritmos de *ambient occlusion* y *baking* de texturas. La importación de la posición y el color de las luces se realizó mediante un *script Python* desarrollado ad hoc. En la figura 5.13 se pueden observar los resultados del *render* de un escenario virtual simple utilizando diferente número de luces.

Se procesaron de diferentes modelos 3D variando el número de luces de las escenas y, tras la aplicación de los algoritmos de visualización avanzada previamente descritos, se realizó la exportación de los mismos en formato “obj” por motivos de compatibilidad con eStudio.

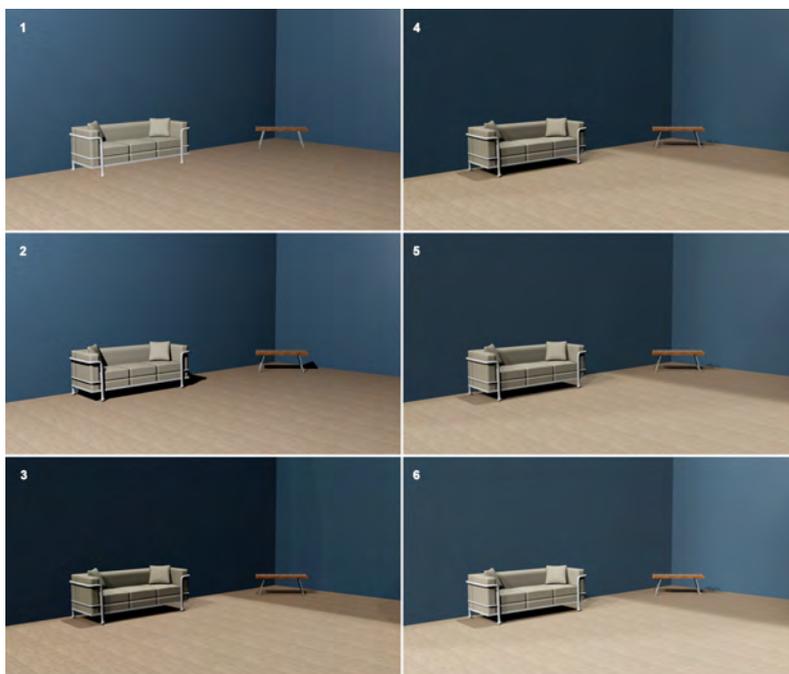


Figura 5.13: Escenario iluminado sin sombras (1), con sombras duras simples(2), con 64 luces (3), con 512 luces (4), con 1024 luces (5) y con 2048 luces (6)

Finalmente, estos entornos se importaron en el software de *render* y se probó su uso como escenarios virtuales de televisión. En la imagen 5.14 puede observarse un ejemplo en el que se compara un estudio virtual sin preprocesado y el mismo entorno tras la aplicación del flujo de trabajo presentado con 1024 luces.

5.1.3. Sombras del mundo real en el mundo virtual

Para obtener una integración verosímil de los actores y el mundo virtual estos han de proyectar sombras sobre los objetos sintéticos. En cerebro humano, de forma intuitiva, utiliza las sombras como una de las colas de información más importantes a la hora de localizar el punto de unión entre dos elementos. En numerosos casos, los presentadores en los platós virtuales de televisión no proyectan ninguna sombra sobre el suelo sintético, dando la impresión de estar flotando sobre el mismo en lugar de estar pisándolo. Esta es una problemática que ha afectado a esta tecnología desde sus comienzos y no presenta una solución sencilla.

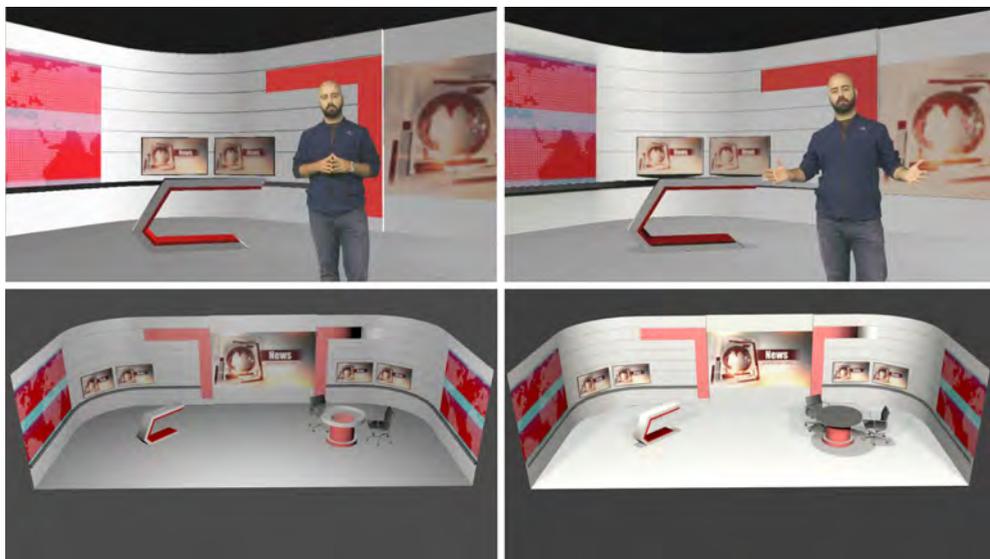


Figura 5.14: Plató virtual antes y después de aplicar la técnica *baking* de texturas con iluminación global. Vista cercana con el presentador en escena (arriba) y vista global del modelo 3D (abajo)

Las sombras en el mundo real estas están formadas por dos zonas (la umbra y la penumbra) cuya zona de transición es gradual y, por tanto, carece de bordes claramente definidos. Sin embargo, este efecto gradual es costoso desde un punto de vista computacional y por ello existen diferentes soluciones que permiten aproximarlos de una forma más eficiente [186]. Originalmente, los algoritmos de visualización avanzada únicamente soportaban sombras duras (figura 5.13-2), pero el avance de la capacidad de cómputo de las tarjetas gráficas ha hecho posible el cálculo de sombras blandas similares a las del mundo real [71]. Entre estas aproximaciones destacan las que hacen uso de mapas [69] y volúmenes [97] de sombra generados desde el punto de vista de las luces. Estas soluciones permiten su adaptación para la ejecución en tarjeta gráfica, lo que hace posible su ejecución en tiempo real para escenas en las que no haya ni un gran número de fuentes de luz ni de objetos ya que, para cada origen lumínico se genera un mapa o un volumen de sombra.

En el caso de un plató virtual de televisión, la dificultad para que un actor proyecte sombras sobre objetos virtuales es todavía mayor, ya que este no existe en el mundo sintético. Una solución simple, pero también efectiva consiste en, conociendo la posición del presen-

tador (que ha de estar sensorizado), situar bajo sus pies una textura simulando una sombra difusa. Aunque el efecto producido estará lejos de la sombra real del presentador en el escenario, proporcionará al cerebro del espectador un punto de referencia del contacto del actor con el suelo virtual y por tanto mejorará la sensación de presencia del mismo, minimizando la apariencia de “flotación” producida por la ausencia de sombreado. Esta solución ha sido implementada y probada realizando un seguimiento del actor con el sensor Microsoft Kinect V2 y situando una textura semitransparente bajo sus pies. También puede aplicarse este efecto sobre objetos reales siempre que estos estén siendo seguidos por el sistema de *tracking* o no vayan a ser desplazados. El resultado de la aplicación de esta técnica puede observarse en la figura 5.15.

Por otro lado, para conseguir efectos más complejos, como la proyección de sombras sobre objetos y no únicamente sobre el suelo, una aproximación es la utilización de un maniquí virtual simple que siga los movimientos del presentador (su cuerpo debe de estar sensorizado). Este modelo 3D se utilizaría, en un primer paso, para la generación de los mapas de sombras que posteriormente serían difuminados, utilizando un desenfoque gaussiano [41]. Finalmente el maniquí sería eliminado de la escena y esta sería *renderizada* con las sombras difusas



Figura 5.15: Imagen de una silla y un presentador sin sombra (izquierda) y de la misma silla y presentador con una sombra simple (derecha)

proyectadas sobre los objetos. De este modo el actor podría proyectar sombras blandas sobre cualquier objeto del entorno virtual. Las principales dificultades de implementación que plantea esta solución son, por un lado, la precisión de la sensorización, que en caso contener ruido podría provocar temblores en la sombra, y por otro, la necesidad de realizar varios pasos de *render* para generar el mapa de sombras, difuminarlo y realizar el *render* final de la escena, lo que podría comprometer el rendimiento en tiempo real del sistema. La implementación de esta tecnología requiere una modificación en el proceso de *render* de eStudio, por lo que no pudo ser probada en el transcurso de esta tesis y se plantea como un trabajo futuro.

5.2. Conclusiones

Se ha implementado un sistema de análisis de la iluminación de cicloramas para platós virtuales de televisión que permite, a usuarios no expertos, realizar una aproximación visual y rápida de la homogeneidad lumínica del escenario de cara a una mejora de la calidad del proceso de *chroma keying*. Aunque es deseable la confirmación de resultados por parte de un experto, este sistema permite intuir errores en la iluminación debidos a una mala configuración inicial (como en el caso de prueba presentado anteriormente) o la degradación de las lámparas y la pérdida de calibración de los focos fruto del paso del tiempo. La sencillez del proceso, así como su rapidez lo hacen ideal para la realización de análisis periódicos.

Por otro lado se ha diseñado y probado un nuevo flujo de trabajo que permite la captación de entornos lumínicos reales y su aplicación sobre escenarios virtuales de televisión de cara a una mayor coherencia lumínica y una mejor integración de los mundos sintético y real.

Finalmente se han planteado dos soluciones para abordar el reto de la proyección de sombras de elementos reales sobre el entorno virtual: la utilización de texturas simples y el uso de un maniquí virtual del presentador que permita efectos más complejos.

CAPÍTULO 6

VISUALIZACIÓN AVANZADA E INTERACCIÓN APLICADA A ENTORNOS MUSEÍSTICOS

En paralelo a la investigación principal entorno a los platós virtuales de televisión se desarrollaron una serie proyectos que hicieron uso de parte de las tecnologías utilizadas y desarrolladas. Uno de los principales campos de trabajo del grupo COGRADE (Gráficos por Computador e Ingeniería de Datos) y previamente del grupo MAR (Multimedia, Animación y Realidad Virtual) ha sido la implementación de módulos para museo sacando partido de técnicas visualización avanzada e interacción persona ordenador no tradicional. Fruto de estos trabajos son proyectos como *EL ACUARIO VIRTUAL* [121], *EL PLANETARIO VIRTUAL* [151], *EL BOTAFUMEIRO VR* [50] y *AS CARRILANAS* [150]. Continuando esta línea de trabajo del grupo se han llevado a cabo dos proyectos museísticos: “El Pórtico de la Gloria Virtual” y el “Módulo de visualización del interior torso humano”. El primero está formado por tres módulos: “El Pórtico de la Gloria virtual” un recorrido virtual a través del monumento, “Los instrumentos virtuales del Pórtico de la Gloria” que permite la exploración y manipulación de las reconstrucciones virtuales de los instrumentos del Pórtico y “Conoce el Pórtico” una aplicación para dispositivos móviles que permite trasladar parte de la experiencia museística in situ de la exposición a los hogares de los usuarios. El segundo consiste en un módulo para la Domus, Casa del Hombre de A Coruña que permite a los visitantes observar el interior de un torso humano. A continuación se presentan estos desarrollos de un modo más

detallado:

6.1. El Pórtico de la Gloria Virtual, aplicación interactiva en tiempo real para la exploración virtual del Pórtico de la Gloria



Figura 6.1: Detalle de la 1ª fase de ejecución del proyecto Pórtico Virtual

6.1.1. Introducción

El Pórtico de la Gloria es, sin ningún género de dudas, una de las cumbres artísticas y arquitectónicas del Románico español y una de las obras maestras del arte Europeo de todos los tiempos. Finalizado en el año 1188 (según la inscripción grabada en el propio dintel de la obra) su autoría se debe al genio del Maestro Mateo y su taller escultórico. Genio, al que cada vez más expertos [145] a la luz de los nuevos datos obtenidos, asignan la autoría y la planificación de la remodelación de la fachada occidental de la Catedral de Santiago de Compostela a comisión del Rey Don Fernando II de León, marco en el cual se englobaría la construcción de tan magnífico conjunto monumental. Esto nos situaría ante un maestro constructor-escultor de dimensiones renacentistas.

El Pórtico es una obra que encierra un complejo simbolismo tanto conceptual como espiritual, y sobre la que se han propuesto numerosas teorías interpretativas que abarcan desde la visión clasicista, que plantea el significado de la obra como una alegoría de las tres iglesias: la Católica, la de Israel y la Gentil [47], hasta las hipótesis de influencia francesa, que nos presentan el Pórtico como una representación del Apocalipsis, el Fin de los Tiempos y el Juicio Final [22]. Las últimas investigaciones obligan, con cierta lógica, a analizar el pórtico en el contexto general de la obra acometida por el Maestro Mateo, incluyendo, por lo tanto, la Cripta de la Catedral, situada bajo el Pórtico, con sus metafóricas claves de bóveda significando el sol y la luna, y la Tribuna, en cuya clave aparece una encarnación pétreo del cordero de dios; lo que nos presentaría el conjunto monumental como la Jerusalén descrita por San Juan en el Apocalipsis, descendiendo de los cielos para dirigirse a la humanidad [199].

Desde hace más de ocho años la Fundación Pedro Barrié de la Maza desarrolla una intervención, diseñada y acometida en conjunto con el Arzobispado de Santiago y el Cabildo de la Catedral, y en colaboración con el Ministerio de Cultura y la Xunta de Galicia, denominada Programa Catedral de Santiago de Compostela [13]. El Programa Catedral, con una inversión de 3.000.000 euros por parte de la Fundación Pedro Barrié de la Maza, se marca como objetivo poner a disposición de la sociedad un proyecto cultural integral que contribuya de forma decisiva a la difusión de la importancia de la conservación de bienes culturales tan destacados como el Pórtico de la Gloria. Con este fin, la Fundación Barrié gestiona, desde el museo situando en la Cripta del Pórtico de la Gloria, un innovador programa de visitas guiadas gratuitas, en el que las nuevas tecnologías de información y comunicación abren nuevas perspectivas de educación, concienciación y puesta en valor del Patrimonio Histórico-Artístico. A continuación se presenta una de las principales actuaciones realizadas bajo el marco de este programa, denominada “Pórtico Virtual”.

6.1.2. Pórtico Virtual

Pórtico de la Gloria Virtual se concibe como un proyecto de largo recorrido, desarrollado en diferentes fases y centrándose en un ambicioso conjunto de objetivos: en una primera fase, a corto plazo y debido principalmente a restricciones temporales, se marca como meta la realización de una reconstrucción virtual del Pórtico de la Gloria 6.1, de carácter educativo y didáctico, sobre un modelo geométrico 3D simplificado así como una aplicación que permita la exploración del monumento en tiempo real (siendo éste el trabajo aquí expuesto). En una segunda fase, a medio plazo, se realizará, utilizando técnicas de visualización y optimización

de modelos extensos (construyendo una base de datos de LOD ¹ discretos multi-resolución), una aplicación que permita la visualización y exploración en tiempo real del modelo original utilizado en ambos casos: un escaneado laser de precisión milimétrica del Pórtico y su elementos circundantes, realizado en 2009 por la Universidad de Cottbus. Estos datos originales no cuentan con información de color, siendo uno de los principales retos de este proyecto el realizar una reconstrucción artística de la capa de textura.

Teniendo en cuenta el alcance marcado por la primera fase del proyecto, la aplicación se concibe como una herramienta pedagógica cuyos objetivos principales se pueden focalizar en dos aspectos básicos: por un lado, servir como complemento informativo sobre el conjunto monumental para su utilización, principalmente, por el personal del museo como sistema de apoyo a las visitas guiadas; mientras que, por otro lado, ha de ser un vehículo de transmisión y puesta en valor de conceptos relativos al “Pórtico de la Gloria”, haciendo hincapié en su significado y en su extrema importancia y valor artístico, así como informar a los visitantes del museo de los trabajos de restauración y conservación que se están llevando a cabo sobre el mismo, concienciando de la urgente necesidad de proteger y preservar el patrimonio histórico del deterioro que el paso del tiempo, y los agentes agresores, provocan de manera inexorable.

En conjunción con los comisarios a cargo de la exposición, y de una serie de expertos en todas las áreas implicadas en el desarrollo de un trabajo de esas características, se consensuan una serie de requisitos que marcaran, junto con los objetivos mencionados en el párrafo anterior, la aproximación seguida de cara a la implementación de la aplicación. A continuación destacamos aquellos más relevantes, o que condicionan de forma más significativa el proyecto:

- Experiencia Interactiva lo que implica una alta capacidad de respuesta
- Posibilidad de visita “en-raíles”
- Representación visual realista del Pórtico de la Gloria
- Representación de información adicional como deterioro, restauración, etc.

¹La técnica LOD (*Level Of Detail*) en computación gráfica implica la disminución de la complejidad de un modelo 3D a medida que se aleja del espectador o de acuerdo con otros parámetros como la importancia del objeto o su velocidad de desplazamiento. Esto permite aumentar la eficiencia de la representación 3D disminuyendo la carga de computacional del *render* de la misma. La menor calidad visual del modelo pasa inadvertida para el usuario debido a la distancia del objeto o su rápido desplazamiento.

6.1.3. Desarrollo del proyecto

En el siguiente apartado se expondrá, de forma resumida, la metodología seguida a lo largo del desarrollo del proyecto, desde la fase obtención de la geometría y texturas, al diseño e implementación de la aplicación, haciendo hincapié en las características particulares de este trabajo.

6.1.3.1. Creación de contenido

El primer punto clave de este proyecto es convertir los datos obtenidos “en bruto”, a partir del escaneado laser, en un formato manejable, tanto por tamaño de la geometría como por tipo de archivo, por los programas de DCC habituales (en este caso Autodesk MAYA) para posibilitar a los artistas gráficos la labor de texturizado. A su vez, es en este momento cuando se identifica cual va a ser el nivel de detalle ofrecido por la aplicación en tiempo real, de forma que se trabaje con el mismo nivel de detalle que en la fase de texturizado. Para ello se intenta llegar a un compromiso basándose en una serie de condiciones, intentando balancear los siguientes puntos: minimizar la pérdida de información visual (recurriendo a trasladar esa información a las texturas, cuando sea necesario, rugosidad superficie, imperfecciones, etc.); geometría total manejable sin técnicas de LOD en tiempo real (evitando posibles artefactos característicos del LOD: *popping*², *flickering*, etc.); garantizar una rápida retroalimentación y movimiento de cámara.

Partiendo de los datos originales, y dividiendo el Pórtico de la Gloria en 11 zonas de interés, se realiza una simplificación que varía entre el 70% y el 90%, en función del tamaño, complejidad geométrica e importancia de cada zona. Esta primera simplificación es llevada a cabo con la herramienta MESHDOCTOR (Geomagic Studio) [58] por la universidad de Cottbus. En una segunda fase de pre-procesado se realiza una optimización de las mallas correspondientes a cada zona de interés utilizando las herramientas proporcionadas, a tal efecto, por el proyecto Open Source, OpenSceneGraph [148], a través de la aplicación sobre la geometría algoritmos de colapso de aristas y de *teselación* eficiente.

En la figura 6.2 se puede comprobar que el nivel de simplificación permite conservar prácticamente toda la información visual necesaria para una aplicación de estas características, mientras que se minimiza la carga computacional requerida para el procesado y gestión en tiempo real de la geometría. El resultado final produce una malla de unos 5 millones de

²Se denomina *popping* al cambio brusco de calidad en el objeto 3D que se está visualizando al pasar de un nivel de detalle a otro.



Figura 6.2: Izquierda: datos sin simplificar de la cabeza del Apóstol Santiago (+300K Vértices). Derecha: simplificación al 70%.

triángulos para todo el Pórtico de la Gloria. Sobre esta malla simplificada y optimizada se trabaja en la fase de texturizado. Para ello se realiza un estudio fotográfico del conjunto artístico para obtener las fuentes de textura, estudio que, lamentablemente, tuvo un alcance limitado debido a que los trabajos de conservación que se están llevando a cabo sobre el pórtico restringen la toma de datos. En base a lo anterior se decidió abordar esta fase en dos etapas: en una primera se realiza el estudio y texturizado del tímpano y el parteluz (presentada en este texto) y en una fase posterior se abordará el estudio y texturizado de los pilares y arcos laterales. El trabajo de texturizado se enfoca desde una perspectiva pedagógico-artística, no-fotorrealista, pero conservando fielmente todos los elementos característicos presentes en el monumento: restos de policromía, textos labrados y pintados, marcas de deterioro, etc., garantizando el rigor científico de la información representada 6.1. Los tamaños de textura finales varían entre los 2048x2048 píxeles y 512x512 píxeles, siendo la carga total de la escena, en esta primera fase, de, aproximadamente, 1GB de información de textura. Así mismo se recurre a implementar técnicas de LOD de materiales/texturas, de cara a garantizar la máxima calidad de visualización posible 6.3.

6.1.3.2. Planteamiento conceptual de la aplicación

Teniendo en cuenta los objetivos y requisitos enunciados en el apartado 2, se concibe la aplicación, de forma conceptual, como un sistema de exploración virtual interactiva del Pórtico de la Gloria. Para ello, como ya se ha comentado, se divide el monumento, teniendo en cuenta la alta carga simbólica del mismo y diversos factores de relevancia histórica o cultural, en un conjunto de 11 zonas de interés, que los usuarios de la aplicación podrán visualizar



Figura 6.3: Resultados de la fase de texturizado. Izquierda Modelo 3D texturizado del Apóstol San Juan. Derecha: fotografía de la misma escultura en su emplazamiento en el Pórtico.

individualmente y explorar a demanda, a la vez que escuchan una serie de locuciones relativas a cada una de las zonas, en las que se explican el significado de cada zona, junto con diversas consideraciones históricas, y referentes a la conservación y restauración del monumento.

Según avanzan estos trabajos de restauración y conservación, el conocimiento que se obtiene sobre el Pórtico de la Gloria va aumentando de manera notable. Esta nueva información se irá incorporando a la aplicación gradualmente, según se vaya consolidando ante los expertos en patrimonio, así los usuarios serán capaces de contemplar el monumento tal cual fue concebido en el siglo XII. En esta primera fase se ha re-construido la policromía del eje central del Típano: el Pantocrátor, como se puede ver en la 6.4.

Desde el punto de vista funcional se contemplan dos modos de “visita”, un primer modo en el que el usuario puede ir explorando a su voluntad los diferentes puntos de interés, y un segundo modo en el que de forma automática se recorre el monumento al completo mientras que se escucha una locución continua, ofreciendo este último modo las funcionalidades de una visita “en-raíles”.



Figura 6.4: Reconstrucción de la capa original de policromía del Pantocrátor basada en los trabajos de restauración.

6.1.4. Arquitectura del sistema

La aplicación se implementa en base al paradigma de computación distribuida teniendo dos componentes fundamentales: un sistema de control y un sistema de visualización, siguiendo el modelo clásico cliente-servidor. El sistema de control tendrá carácter de servidor, mientras que el sistema de visualización hará las funciones de cliente. Para facilitar esta conexión se han utilizado las librerías VRPN, un estándar de facto para la conexión de periféricos en aplicaciones de realidad virtual. Este planteamiento resulta necesario, entre otros factores, debido al balance de carga computacional que se ha de establecer, dado que el sistema de visualización soporta una carga gráfica elevada.

6.1.4.1. Sistema de visualización

Como se ha comentado previamente existen dos modos de utilización de la aplicación que marcan la implementación del software del sistema de visualización: el modo interactivo y el

modo “en-raíles”.

El modo de visualización interactivo utiliza una cámara dinámica que permite contemplar los diferentes puntos de vista del pórtico desde diferentes perspectivas, en función de las peticiones realizadas por los usuarios desde el sistema de control. Se utilizan animaciones generadas en tiempo real para las transiciones entre puntos de interés, siempre pasando a través del denominado “punto de reposo”, que es la posición de cámara que permite contemplar todo el Pórtico de la Gloria en su conjunto (y en la cual la aplicación comienza su funcionamiento), de forma que el usuario nunca pierda la visión global del monumento, y tenga una referencia de en qué zona del mismo se encuentra el punto de interés que está contemplando. Para cada zona de interés existe una locución explicativa en paralelo a la cual, de forma sincronizada, se disparan eventos especiales como: la aparición de infografías, resaltar algún elemento del monumento o la transición entre estado actual y estado original policromado 6.4. El usuario puede acercar o alejar la cámara o moverla, de forma orbital, en dos ejes, siendo posible contemplar interesantes perspectivas que de otro modo resultarían inaccesibles. En cualquier momento puede solicitar la vista de otro punto de interés, o una perspectiva general de todo el Pórtico. El modo de visualización automático utiliza una animación de cámara generada y exportada de un software de DCC, para garantizar una planificación lo más cinematográfica posible. Esta animación recorre de forma secuencial los 11 puntos de interés, mientras se escucha una locución continua sobre todo el monumento, disparándose en momentos clave los eventos especiales, al igual que el modo interactivo. La duración total de este recorrido es de 20 minutos

El componente de la aplicación encargado de la generación de los gráficos ha sido desarrollado utilizando el motor de render de código abierto OGRE3D [114], ajustando la carga gráfica total a las capacidades del sistema hardware, en este caso equipado con una GPU de última generación AMD 5970 2GB. A la hora de instalar la aplicación en su emplazamiento definitivo se completa el sistema de visualización con un proyector de alta definición (HD) NEC con 5000 *lumens*, y una pantalla panorámica de 5x3 m.

6.1.4.2. Sistema de control

Se concibe el sistema de control con una interfaz de tipo táctil en la que, sobre un esquema del Pórtico de la Gloria, se destacan las diferentes zonas de interés numeradas (se incluye una leyenda para saber a qué zonas del pórtico corresponde cada número) 6.5. El usuario pulsará sobre el número de la zona que desee visualizar para obtener una vista en detalle, y podrá ir

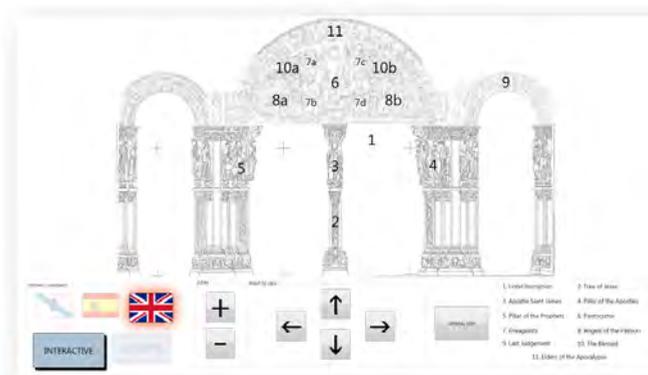


Figura 6.5: Detalle de la interfaz de usuario del sistema de control

variando la perspectiva de la cámara con un conjunto de botones que permiten hacer zoom o moverse en diferentes ejes. Además se incorpora un botón para solicitar una visión general del pórtico, botones para la selección de idioma, tanto de los textos como de las locuciones, entre tres idiomas (gallego, castellano e inglés), y un botón para realizar la transición entre el modo interactivo y el modo guiado o automático. En este último modo la interfaz de control se simplifica mostrando únicamente los botones de selección de idioma, de accionamiento de la visita “guiada”, y para regresar al modo interactivo. Véase detalle de la interfaz en la 6.5. La aplicación se desarrolla utilizando Windows Presentation Foundation [135], lo que permite utilizar todas las funcionalidades del novedoso GUI Aero del sistema operativo Windows, la interfaz gráfica se diseña en lenguaje XAML, y todo el software se apoya en el .NET Framework 4.0 permitiendo programar los diferentes eventos táctiles necesarios para el control del sistema. A su vez el sonido de la aplicación se gestiona desde este componente de la aplicación, para minimizar la carga computacional del sistema de visualización.

El software se instala en un ordenador tipo todo en uno, Asus EEPc Top ET2203, con una pantalla táctil capacitiva de 21.6 pulgadas, a una resolución 1080p. Se incorpora una tarjeta de sonido USB externa 5.1.

6.1.5. Conclusiones

Se ha presentado la primera fase de un proyecto de mayor alcance cuyo principal objetivo es desarrollar una herramienta de exploración y ayuda al estudio y conservación de un

monumento tan característico y complejo como es el Pórtico de la Gloria. En esta primera aproximación, partiendo de datos científicos rigurosos, se intenta construir una aplicación destinada a la exploración básica del monumento, orientada hacia el público general, dando un enfoque didáctico de forma que se haga más accesible la presentación de la información. Se presenta una interfaz de usuario transparente, directa, y accesible a todo tipo de público, que utilizando técnicas de selección de destino, permite la navegación libre por aquellos puntos de mayor interés del monumento, mientras se complementa la información visual con locuciones explicativas e información adicional a través de infografías y transiciones de textura. Además se incluye un modo “automático” que permite recorrer todos los puntos de interés de forma continua, sin necesidad de interacción por parte del usuario, muy útil en el caso de experiencias multiusuario. A través de la retroalimentación recibida de los responsables del museo, se constata que la aplicación es muy bien recibida por los usuarios, sin presentar problemas de adaptación, como cabe esperar, a las metáfora de interacción planteada.

Como resultado de este proyecto se presentaron artículos en el III Congreso Internacional de Arqueología e Informática Gráfica, Patrimonio e Innovación [152] y la revista VAR (*Virtual Archaeology Review*) [153].

6.2. Los instrumentos virtuales del Pórtico de la Gloria



Figura 6.6: (1) Reconstrucción del pórtico de la gloria. (2) Detalle de uno de los músicos.

6.2.1. Introducción

Como se ha comentado en el apartado anterior en el interior de la Catedral de Santiago de Compostela se encuentra una de las piezas maestras de románico, el Pórtico de la Gloria. El monumento consta de tres arcos. En el izquierdo se representan las escrituras del viejo testamento, con los justos esperando la llegada del salvador. En el arco derecho se muestra el infierno y en el central aparece representada la gloria, con Cristo en el centro. En torno al tímpano del arco central se sitúan 24 ancianos afinando instrumentos musicales, una imagen descrita en el Libro del Apocalipsis. El conjunto era, originalmente, polícromo, pero con el paso de los siglos ha sufrido los efectos del tiempo la humedad y diversas restauraciones que han modificado su color original y su aspecto.

Además de promover la conservación, estudio y exposición del Pórtico a través de la visita virtual propuesta en el módulo “El Pórtico de la Gloria Virtual” la Fundación Pedro Barrié de la Maza acompaña la experiencia de un estudio y reconstrucción virtual detallada de los instrumentos que portan los músicos del arco central. A continuación se presenta el sistema interactivo desarrollado para la exhibición de dichas reconstrucciones.

Al situarse el tímpano del Pórtico de la Gloria a seis metros de altura (alrededor de 9.5 en

su punto más alto), es imposible distinguir, observándolo desde el suelo, los pequeños detalles que caracterizan cada una de las figuras del arco y reconocer el instrumento que cada anciano está afinando 6.6-1. Estos detalles si se podían observar en las reproducciones de los instrumentos realizadas por [111] y expuestas en el museo de la Cripta del Pórtico de la Gloria. No obstante, al igual que muchas piezas de museo (como cerámicas, esculturas, etc.) los instrumentos musicales requieren unos criterios específicos para su protección y mantenimiento por los que los visitantes se veían obligados a observarlos en el interior de una vitrina, eliminando, de este modo, todo tipo de interacción con la pieza [78].

Para hacer los instrumentos del Pórtico de la Gloria más accesibles, se planteó la necesidad de desarrollar un sistema virtual interactivo, tangible y fotorrealista, introduciendo modelos 3D de los instrumentos [64], de forma que los visitantes pudiesen disponer de una forma de observar, interactuar y escuchar el sonido de los instrumentos reconstruidos. Este sistema debía estar centrado en la accesibilidad para usuarios no expertos, ya que estaba destinado a su exposición pública. La reconstrucción de los instrumentos fue realizada a partir de los planos originales y de fotografías tomadas de la reconstrucción realizada por López-Calo [111].

6.2.2. Descripción del sistema

Para el desarrollo del sistema se ha creado un conjunto de hardware y software que simplifique la interacción entre el usuario y el ordenador, reduciendo, de esta forma, la curva de aprendizaje del usuario, siendo este capaz de interactuar con el sistema sin problemas al poco tiempo de entrar en contacto con el sistema.

Con este objetivo se planteó la hipótesis de que la forma más natural de manejar un instrumento es coger un objeto real con las manos e interactuar con él como si realmente se estuviese manipulando la pieza real. Se optó entonces por que el sistema tuviese una interfaz tangible que el usuario podría manejar libremente y que esta sería la única forma de interacción entre el usuario y el sistema.

Tras estudiar diversas opciones (la utilización del Wiimote, un cubo con marcadores en sus caras), etc.) se decidió construir esta interfaz tangible utilizando técnicas de visión por computador tradicionalmente utilizadas en la implementación de aplicaciones de realidad aumentada. Una cámara capturaría imágenes de este componente manipulable, que dispondría de un solo marcador situado en su parte inferior para simplificar el tracking. Se debería construir la aplicación de forma que el usuario fuese capaz de operar con el sistema utilizando únicamente la posición y orientación de ese marcador en cada momento.



Figura 6.7: Montaje final del hardware del sistema.

Se decidió utilizar un disco como interfaz tangible por la forma semicircular del arco en el que se sitúan los ancianos del apocalipsis. La similitud de formas nos permite situar el objeto en el interior del arco, simplificando de esta forma la selección del anciano que se desea ver. El disco dispone de una flecha que, apuntando a una cierta posición del arco, permite al usuario seleccionar uno de los ancianos 6.7-5.

Se diseñó un puesto de información compuesto por un mueble que contiene todo el hardware necesario para un funcionamiento adecuado del sistema 6.7-1: un PC con una tarjeta gráfica Nvidia GTS 250, una pantalla full HD de 21 pulgadas, una webcam, el disco con marcador y unos cascos para escuchar las explicaciones sobre los instrumentos y su sonido 6.7.

En la parte superior del mueble hay una imagen del arco central del Pórtico de la Gloria con la representación de los músicos en torno a un agujero biselado 6.7-2 sobre el que descansa el disco. El disco cuenta con un patrón en la parte inferior y una flecha y logos del museo en la parte superior 6.7-4. Este disco va unido al mueble mediante un cable para evitar su hurto, ya que está pensado para ser expuesto en un lugar público. El usuario sostendrá en sus manos el interfaz tangible, por lo que el cable debe de tener la longitud suficiente para permitirle libertad de movimientos pero debe de ser lo suficientemente corto como para evitar interferir en la captura de la cámara o molestar al usuario.

La pantalla se sitúa sobre la imagen del arco de los ancianos con un ángulo de 30 grados para permitir una correcta visualización de los colores de la aplicación y evitar reflejos no deseados al ser para un usuario estándar un ángulo óptimo desde el que mirar a la pantalla 6.7-3. Sobre esta se sitúan las instrucciones de uso y a su derecha los auriculares que usará el usuario para escuchar tanto las explicaciones del sistema como el sonido de los instrumentos.

Bajo la imagen del Pórtico de la Gloria y el disco, en el interior del armario, se sitúa una cámara en un ángulo de 40 grados; un ángulo de visión óptimo para seguir el patrón adherido al disco al disponer así de un mayor ángulo de visión del exterior del armario. Bajo la cámara se sitúan tres tubos fluorescentes tapados por una superficie de metacrilato blanca, combinación que genera una iluminación blanca y difusa óptima para las aplicaciones de visión por computador, ya que evita reflejos sobre el disco o variaciones de color en el patrón en movimiento.

6.2.3. Comportamiento del sistema

El comportamiento del sistema se ha simplificado al máximo. Cuando el disco está sobre su soporte, en el centro del arco, podemos navegar por él simplemente apuntando al músico que se desea visualizar (el instrumento que queremos seleccionar) con la flecha sobre el disco. Entonces la aplicación moverá la cámara hasta un primer plano del músico seleccionado 6.8-1.

Tras seleccionar el músico, se puede acceder al instrumento que está afinando al levantar el disco de su soporte. Tas cogerlo, la aplicación desplaza la cámara a una posición de visión global del arco, el instrumento reconstruido aparece en primer plano y la explicación del instrumento correspondiente, así como su sonido, comienzan a sonar. Una vez en este estado, podemos manipular la reconstrucción virtual moviendo el disco libremente. Cuando el usuario rota el disco y varía su posición el instrumento se comportará de la misma manera, rotando y haciendo zoom consistentemente 6.8-2.

Para seleccionar un instrumento diferente el usuario debe colocar de nuevo el disco en su posición de reposo con la flecha apuntando en al músico deseado. Al bajar el disco la reconstrucción del instrumento y el sonido desaparecen y la cámara vuelve al primer plano de la figura pétreo apuntada.

La aplicación dispone también de un modo de inicio en el que el usuario escucha una explicación del Pórtico de la Gloria como conjunto. Para entrar en este modo el usuario debe de situar el disco sobre el soporte apuntando al logo de "Inicio/Start", que mueve la cámara a una vista general del arco de los ancianos mientras comienza la explicación anteriormente

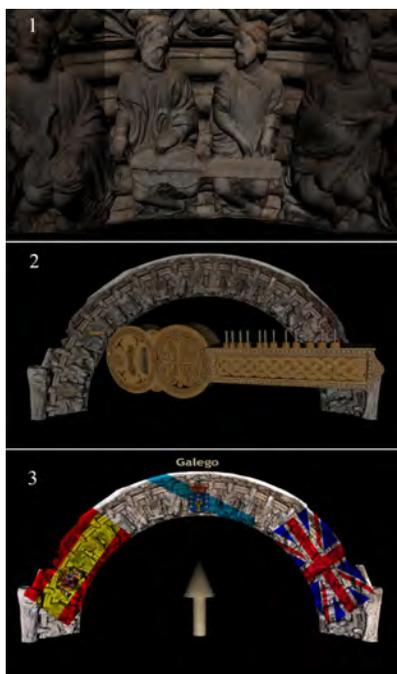


Figura 6.8: Capturas de la aplicación en funcionamiento. Arriba: selección de instrumento. En medio: interacción con el instrumento. Abajo: selección de idioma

nombrada. También se accede a este modo cuando la aplicación pasa más de un minuto sin ser manipulada, con la finalidad de que el usuario se encuentre con esta explicación global cuando entre en contacto con la aplicación.

Finalmente se permite la selección de idioma utilizando la misma metáfora de interacción. Se rota el disco hasta la posición de selección de idioma, al entrar en esta se levanta y aparece una flecha, junto con tres banderas representativas, que le permite al usuario seleccionar entre los tres lenguajes disponibles 6.8-3. Para realizar la selección de un nuevo idioma se debe depositar el disco en su posición apuntando a la bandera seleccionada.

6.2.4. Arquitectura software

La aplicación final está formada por dos módulos que se comunican mediante una conexión VRPN. El primero utiliza técnicas de visión por computador para obtener la rotación y traslaciones de la interfaz tangible a través del espacio y la segunda permite mostrar por

pantalla la información sobre el Pórtico y los instrumentos. Para lo primero se ha utilizado ARToolkit [38], una librería que permite detectar patrones predefinidos dentro de una imagen. A partir de la orientación de la figura situada en el interior de patrón y sus cuatro esquinas 6.7-4, ARToolkit permite calcular las rotaciones y translaciones que relacionan el sistema de coordenadas de la cámara con el del patrón almacenando, en tiempo real, esta información en una matriz de homografía. De este modo, haciendo uso de una cámara web y analizando las imágenes que esta captura, es posible realizar el seguimiento de un patrón que se desplaza a través del campo de visión de la cámara. En el sistema presentado, el patrón del que se realiza este seguimiento está adherido al disco que manipula el usuario 6.7-4.

Una vez que el patrón ha sido detectado y la matriz de homografía calculada, la aplicación empaqueta la información de rotaciones y traslaciones del objeto utilizando la librería VRPN y la envía a través de la conexión previamente establecida.

El segundo módulo muestra la reconstrucción del arco en el que se encuentran los ancianos y los instrumentos, encargándose también de la gestión de la lógica del programa y el sonido. Para llevar a cabo estas tareas se ha utilizado Ogre Graphics Engine [114] para el renderizado y OpenAL [178] para la gestión el sonido. Esta parte de la aplicación gestiona la interacción del usuario con la máquina a partir de los datos recibidos del módulo de visión por computador.

La principal ventaja de tener dos aplicaciones es que si el patrón es perdido por la cámara en algún momento no afectará a la imagen mostrada ya que el módulo de visión por computador no envía información, y por tanto las posiciones y rotaciones en el módulo de *render* no son actualizadas, pero se mantendrá la velocidad de refresco de la imagen. Se consigue por lo tanto una independencia en el funcionamiento de ambas aplicaciones, lo que hace que aunque la cámara capture a 25 fotogramas por segundo (y por tanto envíe información 25 veces por segundo) la aplicación de *render* mantendrá la velocidad de refresco máxima que le permita la tarjeta gráfica, teniendo en cuenta el número de polígonos que maneja y los cálculos que han de realizarse para obtener el *render*. De este modo se mantiene, en todo momento, el tiempo real en la visualización. En el caso de la implementación presentada se obtienen 130 imágenes por segundo.

6.2.5. Exposición itinerante

Tras el gran éxito de la exposición situada en la cripta del Pórtico de la Gloria se decidió la creación de una versión itinerante de la misma que permitiera acercar el monumento

a distintas partes del mundo. La principal dificultad que presentaba esta configuración móvil era la necesidad de acortar en la medida de lo posible los tiempos de instalación y montaje. En el caso de los instrumentos del pórtico de la gloria, la utilización de técnicas de visión por computador hacía necesario un proceso de recalibrado para cada nueva instalación que dificultaba su transporte. Por ello se creó una versión itinerante del módulo en la que se sustituía la utilización de una interfaz tangible por un *trackball* con dos botones, teniendo como resultado una interacción más tradicional pero un sistema más robusto y resistente a los desplazamientos.

Gracias a la utilización de las librerías VRPN, la sustitución del periférico de entrada de datos fue trivial, necesitando únicamente la implementación de un servidor de datos asociado al *trackball* y la adaptación de la información recogida por este al rango de funcionamiento de la interfaz tangible.

6.2.6. Conclusiones

En este proyecto se ha implementado un sistema para la visualización y manipulación de los instrumentos del Pórtico de la Gloria diseñando un hardware y una arquitectura software específicos. El sistema permite la visualización de la reconstrucción texturizada del arco en el que están representados los ancianos del Apocalipsis así como la visualización y manipulación de la reconstrucción virtual de los instrumentos que están afinando.

Se han utilizado técnicas de visión por computador para dotar al sistema de la máxima accesibilidad para usuarios no expertos, de forma que un usuario sin conocimientos de informática pueda acercarse a un puesto de información y manipular el sistema fácilmente al cabo de unos minutos. Al utilizar una cámara para detectar el patrón el usuario puede manejar el disco de la misma forma que lo haría si tuviese el instrumento real en sus manos, estableciendo, de esta forma, una interfaz natural e intuitiva entre el usuario y la información mostrada.

Los 18 instrumentos que aparecen representados en el Pórtico de la Gloria han sido reconstruidos siguiendo el trabajo de López-Calo [111] 6.9. Los 24 ancianos del Apocalipsis han sido texturizados en base al actual estado de conservación de la policromía del pórtico a partir de fotografías del mismo.

Desde su instalación en julio de 2010, el sistema ha tenido más de 100.000 visitas resultando en un alto grado de satisfacción de los usuarios con la exposición.

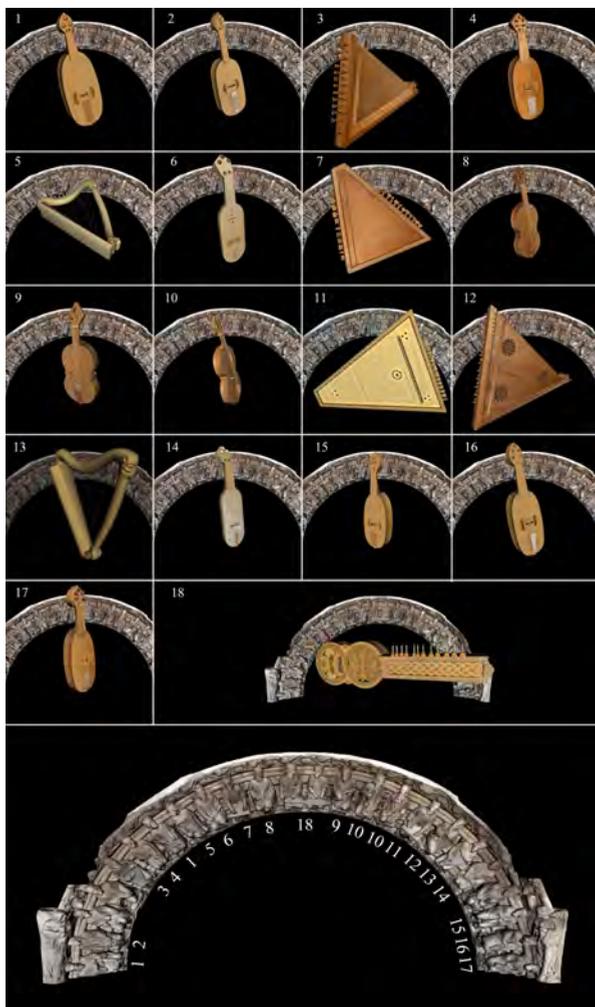


Figura 6.9: Reconstrucción de los 18 instrumentos del pórtico y su localización en el monumento

Finalmente, gracias a la utilización de una arquitectura distribuida mediante el uso de las librerías VRPN, fue posible una rápida adaptación del sistema para facilitar traslado dentro de la versión itinerante de la exposición que recorrió distintas ciudades europeas como Berlín, Londres o Lisboa.

Como resultado de este proyecto se presentaron artículos en el III Congreso Internacional

de Arqueología e Informática Gráfica, Patrimonio e Innovación [129], en el congreso CAA (*Computer applications and quantitative methods in Archaeology*) [124] y en la revista VAR (*Virtual Archaeology Review*) [130].

6.3. Conoce el Pórtico

Tras la implementación de los dos sistemas previamente comentados (“El Pórtico de la Gloria virtual” y “Los instrumentos virtuales del Pórtico de la Gloria”) y su exhibición en museos de diferentes partes de Europa como Berlín, Londres o Lisboa, se propuso la creación de una aplicación para dispositivos móviles que permitiese acercar esta experiencia, y por tanto el conocimiento del Pórtico de la Gloria, a usuarios de todo el mundo, independientemente de donde se encontrasen. La propuesta inicial fue la creación de una aplicación para iPod, iPad y iPhone llamada *Conoce el Pórtico*.

Esta aplicación está dividida en cuatro partes principales: El Pórtico Virtual, Restauración, Galería de imágenes y Exposición.

6.3.1. El Pórtico Virtual

Tanto el Pórtico de la Gloria virtual como los Instrumentos virtuales del Pórtico de la gloria son sistemas basados en el *render* en tiempo real de modelos 3D de gran tamaño, haciendo uso de un poder de cómputo gráfico que un dispositivo móvil no puede ofrecer. Por tanto, la versión para móviles de la visita virtual al monumento necesita de nuevos paradigmas de visualización e interacción.

El paradigma de interacción de la versión museística del Pórtico de la Gloria virtual se basaba en el movimiento libre del usuario en torno a ciertos puntos de interés del monumento, utilizando una pantalla táctil independiente como dispositivo de control. Asimismo, el paradigma de visualización hacía uso de grandes cantidades de polígonos y diferentes niveles de detalle en las texturas para lograr *renders* realistas. Al no ser posible el traslado directo tanto de la forma de interacción como de la visualización a un iPod, iPhone o iPad, se simplificaron computacionalmente manteniendo la filosofía de fondo. Para ello el modelo 3D original fue substituido por un *render* de alta resolución del mismo y la exploración continuó realizándose del mismo modo, a través de puntos de interés, pero en dos dimensiones. Al tratarse de una imagen obtenida de un modelo 3D y, a pesar de la ausencia de rotaciones, se conserva, en par-

te, la sensación de profundidad y por tanto se obtiene como resultado una experiencia similar a la que se puede vivir en el museo.

El control se realiza directamente sobre la pantalla táctil del dispositivo. El usuario puede realizar diferentes gestos habituales en las pantallas táctiles de hoy en día (pellizcar para acercar o alejar la imagen, arrastrar el dedo para desplazarla, etc.) permitiendo, también el movimiento libre a través del monumento. Se conservan 11 puntos de interés que el usuario puede visitar con detalle pulsando sobre el número identificativo 6.10. Cuando esto ocurre se reproduce una explicación del significado e historia de las esculturas de esa área, incluyendo el uso de fotografías reales del monumento. Mientras el usuario se encuentra dentro de un punto de interés puede desplazarse libremente por la imagen de la zona permitiendo, haciendo *zoom* sobre la misma o centrando la imagen en la escultura concreta que está siendo presentada en el audio.

También está disponible una visita “sobre raíles” que consiste en un vídeo que recorre todos los puntos de interés del modelo 3D, explicándolos y recorriendo el monumento al completo, permitiendo hacer uso de todas las posibilidades comunicativas que ofrece esta



Figura 6.10: Vista global y detallada del monumento en la aplicación

tecnología. Este modo le permite al usuario vivir la experiencia de visita “sobre raíles” de la aplicación museística desde sus dispositivos móviles.

6.3.2. Restauración

Esta parte de la aplicación pretende transmitir la complejidad del proceso de restauración del Pórtico de la Gloria que está siendo llevado a cabo con la ayuda de la fundación Barrié desde el año 2008 y cuya finalización se prevé para el año 2020. Debido a la duración de la obra es preciso utilizar herramientas divulgativas que le permitan al ciudadano de a pie comprender la necesidad de llevar a cabo estas actuaciones y su dilación en el tiempo.

El objetivo principal de los procesos de conservación del patrimonio cultural es mantener su estado original lo largo del tiempo. Para ello es necesario estudiar en profundidad las causas originales del problema de conservación (humedades, erosión, etc.) para poder atacarlas directamente y evitar que la degradación continúe.

Este módulo de la aplicación explica a los usuarios los problemas encontrados por los encargados de patrimonio de la Fundación Barrié en el Pórtico de la Gloria desde el inicio de la restauración, así como los pasos que se han seguido y los que se van a seguir para recuperar el monumento y detener su deterioro. Para ello se ha diseñado un mapa de estudios 6.11 en el que se combinan imágenes y texto para presentar, de una forma atractiva, los trabajos que se están llevando a cabo. Esta imagen ha sido pensada específicamente para resultar visualmente atractiva y facilitar al usuario la comprensión de los diferentes pasos que sigue un proceso de restauración. Para ello se presentan los problemas encontrados en diferentes partes del monumento acompañados de fotografías a modo de ejemplo. También se presenta un vídeo en el que algunos expertos explican verbalmente, de un modo meramente informativo, el proceso que se está llevando a cabo.

6.3.3. Galería de imágenes

Este módulo presenta una serie de fotografías en el formato tradicional de galería de imágenes presente en todos los dispositivos móviles de la actualidad. Permite realizar *zoom* a través de una doble pulsación o mediante un pellizco y navegar por las diferentes fotografías a través de gestos o utilizando dos flechas situadas a la derecha e izquierda de la pantalla. Una única pulsación sobre la imagen muestra u oculta diferentes opciones de navegación como ir

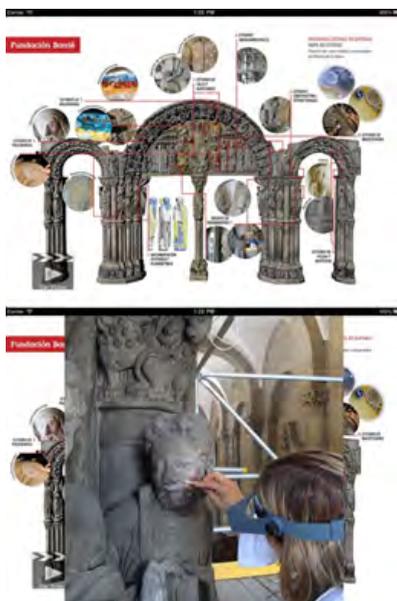


Figura 6.11: Mapa de estudios del monumento e imagen detallada del proceso

a la siguiente imagen, volver a atrás, ir a la primera fotografía o acceder a una con un número de identificación conocido utilizando una barra de desplazamiento.

6.3.4. Exposición

Este módulo funciona a modo de publicidad de la exposición itinerante que ha viajado a lo largo de Europa. Le permite al usuario visualizar vídeos explicando el proceso de desarrollo tanto del Pórtico de la Gloria Virtual como de los Instrumentos virtuales del Pórtico, como funcionan y el porqué de su motivación. Se pretende, a través de esta parte de la aplicación, promover las visitas de los usuarios a la exposición itinerante cuando esta se encuentre cerca de su domicilio y animar a las instituciones a llevar la experiencia virtual del Pórtico de la Gloria a diferentes partes del mundo.

6.3.5. Conclusiones

En este proyecto se ha implementado una versión móvil para iPod, iPad y iPhone de la experiencia presencial en museo de la exposición “El Pórtico de la Gloria Virtual”, permi-

tiéndoles a los usuarios acceder a información sobre el monumento, su actual estado y el proceso de restauración que se está llevando a cabo desde cualquier parte del mundo. Se ha simulado la experiencia de visita a la reconstrucción virtual mediante una imagen de alta resolución y la utilización de recursos audiovisuales como videos, locuciones y fotografías.

Además tener una utilidad didáctica, la aplicación tiene también la finalidad publicitaria de dar a conocer el monumento, su restauración y la exposición itinerante.

Como resultado este proyecto fue incluido en el capítulo *Development and Creativity in Ubiquitous Technologies in University: CIDUS Contribution* del libro *Emerging Perspectives on the Mobile Content Evolution* indexado en SCOPUS.

6.4. Módulo de visualización del interior del busto humano

6.4.1. Introducción

En los últimos años los museos, en especial los científicos, han sufrido una gran evolución pasando de meras salas de exposición a espacios donde el usuario puede interactuar con los elementos que se muestran. Incluso museos clásicos como el Louvre han incorporado elementos como la Nintendo 3DS o el sensor Kinect para crear una experiencia más interactiva con el visitante. En este tipo de entornos tiene una importancia capital el hecho de mantener la atención del visitante y transmitirle conocimientos de una forma clara, amena y entretenida, promoviendo un aprendizaje por experimentación que haga más atractiva la información que se desea transmitir. Si, por otra parte, los potenciales visitantes son niños debemos desarrollar paradigmas educativos que permitan difuminar, en la medida de lo posible, la frontera entre el aprendizaje y el juego.

El caso de estudio se centra en el diseño e implementación de un contenido interactivo para "La Domus, Casa del Hombre de A Coruña-[2], museo interactivo, encuadrado en el conjunto de museos científicos de la ciudad, que centra su actividad en diferentes aspectos del ser humano como especie, proporcionando, entre otras cosas, un recorrido completo a través de su anatomía. Se trata de un museo orientado principalmente a un público infantil y familiar, por lo que ha de tenerse en cuenta el aspecto lúdico a la hora de desarrollar la instalación. Esta instalación se enmarca dentro de una nueva exposición denominada "Juego de Neuronas" compuesta por 13 módulos interactivos, que recorren cada una de las 8 inteligencias descritas por Gardner [57] y algunos aspectos de la anatomía y fisiología del cerebro. Se plantea crear un contenido interactivo que permita al visitante observar las características

anatómicas del cerebro, explorarlo y visualizar las diferentes partes del mismo, así como los elementos orgánicos del busto humano. Por lo tanto, este módulo cubre la parte de la exposición centrada en la anatomía del cerebro.

6.4.2. Objetivos

El objetivo principal de este proyecto es desarrollar un sistema que permita la exploración anatómica del cerebro humano mediante una interacción natural por parte del usuario.

Para lograr este objetivo se plantearon dos requisitos funcionales principales:

- Implementar una herramienta de visualización de imagen médica (TAC, RMI o similar) que permita explorar el modelo mediante un plano de corte en los tres ejes cartesianos del espacio, pudiendo seleccionar la posición del plano y el eje de corte del mismo en cada momento.
- Implementar un sistema de interacción intuitivo, natural y atractivo del usuario con el sistema, evitando el uso de periféricos clásicos como teclados, ratones y, cada vez más, pantallas táctiles. Este sistema debe permitir que el usuario defina en todo momento el plano de corte deseado y su posición en el espacio.

6.4.3. Diseño del contenido

El paradigma utilizado para el diseño del sistema se basa en la arquitectura clásica de un expositor de museo. Estos expositores son urnas en cuyo interior se expone el elemento de interés. En nuestro caso se sustituye la urna por un monitor y se introduce un componente de interacción del usuario con el contenido que permita incrementar el valor expositivo y docente del mismo. De esta forma, el diseño debe tener una arquitectura similar, tal como se muestra en la 6.12. Está formado por una pantalla en la que realiza la visualización de las imágenes del cráneo y uno o varios sensores para permitir la interacción natural del usuario con el sistema.

Puesto que se trata de un contenido interactivo se define una zona de control en la que el usuario puede moverse manteniendo el control del sistema, fuera de la cual el usuario perderá el control del mismo. Esto es necesario debido a que, al estar situado el módulo en una zona con tránsito de visitantes se debe discernir cuál de los posibles usuarios en la zona cercana al quiosco tiene el control. Esta diferenciación se realizará mediante la restricción del funcionamiento del sistema a la zona de control especificada.

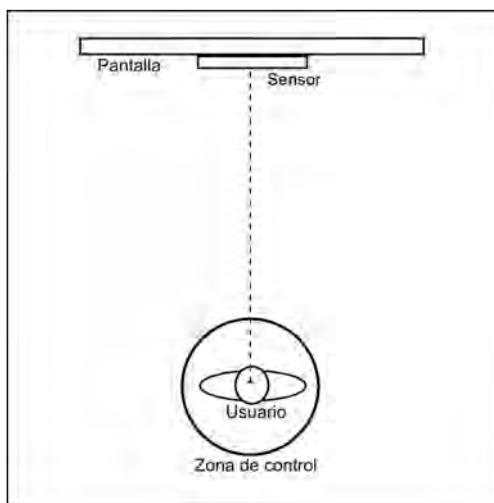


Figura 6.12: Esquema de diseño del módulo

6.4.4. Tecnologías empleadas y diseño hardware

En base a los objetivos planteados se realizó un estudio de las diferentes herramientas de software libre tanto para la visualización como para la interacción del usuario con el sistema. Como fue comentado con anterioridad el sistema consta de dos partes: un entorno gráfico que permite la visualización del modelo y un sistema de interacción para controlar dicha visualización.

6.4.4.1. Sistema de Visualización

Se ha de construir una visualización tridimensional a partir de la información contenida en 275 imágenes con una resolución de 512x512 pixels pertenecientes al Visible Human Project [12]. Para realizar este proceso se planteó el uso de un motor gráfico especializado en esta tarea: VTK (*The Visualization Toolkit*) [94]. No obstante, los resultados de rendimiento desaconsejaron su uso, ya que las velocidades de refresco obtenidas no eran estables y caían en ocasiones por debajo del tiempo real necesario para toda aplicación interactiva. Teniendo esto en cuenta se optó por un motor gráfico orientado al render en tiempo real, decantándonos por OSG (*Open Scene Graph*) [148] por tratarse de uno de los más extendidos, así como por la posibilidad de integración que ofrecía tanto con ITK (*Insight Segmentation and Registra-*

tion Toolkit) [93] como con DCMTK (*DICOM Toolkit*) [182], herramientas que permiten leer e interpretar diferentes formatos de imagen médica y en especial DICOM, formato que se ha convertido en un estándar en este campo y que presenta grandes dificultades de lectura, interpretación y generación debido a su complejidad interna [16]. Con esta herramienta se obtuvo una visualización en tiempo real y con una tasa de refresco estable, condiciones indispensables para el sistema planteado.

En este punto se planteó la sustitución de las imágenes iniciales en formato DICOM (en escala de grises) por unas equivalentes en formato png (en color), ya que la información cromática ofrece una mejor percepción de los diferentes tejidos y estructuras presentes en el modelo aumentando también la espectacularidad de la visualización. Al disponer del conjunto de datos completo del *Visible Human Project* [168] y utilizar OSG como motor de *render* el cambio fue directo, ya que este motor gráfico permite la importación de volúmenes como conjuntos de imágenes en diferentes formatos. Este cambio permitió, además, aumentar la resolución de la visualización al tener las imágenes en color un mayor detalle que su equivalente DICOM.

Este nuevo conjunto de imágenes precisó de un preprocesado ya que la zona exterior al cráneo estaba completamente cubierta por la mezcla de gelatina y agua empleada para la congelación del sujeto y, al realizar la visualización se visualizaba como un bloque azul. Para su procesado se definió una zona de interés (en la que estaba contenido el busto) descartando todos los elementos exteriores a la misma. La eliminación de la región exterior azul se realizó mediante un proceso de aplicación de umbrales. Al ser la parte anatómica de la imagen una región cerrada y al tener los tejidos de la misma una pigmentación bien diferenciada de la región a eliminar bastó con seleccionar el color azul y aplicarle una serie de umbrales. En cuanto a las zonas que contienen mucosas en el interior del cráneo (aparecen también azules debido a la penetración del líquido de congelación en su interior) se decidió no procesarlas ya que la pigmentación resultado de la congelación permite a un usuario no experto detectarlas con mayor facilidad. El conjunto de datos final está formado por 275 imágenes con una resolución de 1024x608 pixels (figura 6.13).

El proceso seguido a la hora de construir un modelo 3D a partir de una serie de cortes 2D es el siguiente: primero se cargan las imágenes en memoria. Una vez cargadas se construye una textura 3D que se visualizará a través de un proceso de trazado de rayos. Para realizar los cortes requeridos en los tres ejes cartesianos se hace uso de tres planos de corte (uno por cada eje) que, combinados con la textura 3D y el trazado de rayos permiten *renderizar* todos los

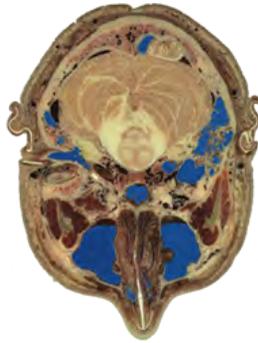


Figura 6.13: Imagen PNG procesada

detalles tanto exteriores como interiores del volumen visualizado.

6.4.4.2. Sistema de Interacción

Al tratarse de un museo principalmente dirigido a público infantil y familiar el diseño de una interacción intuitiva y al mismo tiempo atractiva resulta indispensable.

En los últimos años han aparecido una serie de sensores avanzados que permiten enriquecer, a través de los nuevos dispositivos de captura de movimiento de bajo coste, la interacción del usuario con el ordenador, creando nuevas interfaces naturales y metáforas de acción que añaden una tercera dimensión además de aumentar la implicación física del usuario en la interacción con la máquina [56]. Persiguiendo un control más natural del sistema se plantea la utilización de estos sensores con el fin de evitar el uso de periféricos tradicionales como el ratón y el teclado buscando simplificar la realización de una serie de tareas por parte del usuario. Los tipos de sensorización planteados fueron los siguientes:

- Sensorizaciones mediante cámaras de infrarrojos y marcadores reflectantes: el usuario tendría una herramienta que manejaría en el aire y a través de la cual podría interactuar con el sistema. El principal inconvenientes de este tipo de sensorización es la necesidad de que el usuario maneje una herramienta, con las posibilidades de hurto y rotura que esto implica.

- Sensorizaciones mediante cámaras RGB y marcadores: las mayores dificultades que presenta esta tecnología son tanto el evitar oclusiones en los patrones manejados por el usuario, como el obtener la configuración óptima de iluminación de la sala o la presencia de elementos físicos manipulables que pueden ser hurtados o que se pueden deteriorar con el uso.
- Sensorización mediante el uso de sensores PrimeSense (Asus Xtion o Microsoft Kinect): la principal ventaja de esta tecnología es que trabaja sin la necesidad de usar patrones ni marcadores de ningún tipo. Emplean un proyector y cámara de infrarrojos para detectar usuarios en su zona de acción y reconstruye automáticamente un esqueleto del mismo, siguiendo sus movimientos en el espacio.

Teniendo en cuenta que las condiciones lumínicas del museo no podían ser controladas sin afectar a toda la exposición se descartó el uso de cámaras RGB. También se evitó la manipulación de un objeto como interfaz de control (un mando de consola o elemento manipulable con captura de movimiento) por los problemas de mantenimiento que este podría presentar así como por la posibilidad de hurto del mismo, descartando, por tanto, el uso de cámaras de infrarrojos o mandos de control. Se decidió emplear sensores PrimeSense para crear una interacción natural entre el usuario y el sistema a través de los movimientos de su cuerpo.

El esquema de interacción es el siguiente: el visitante del museo llega al módulo, se sitúa de pie, en una posición indicada mediante un gráfico en el suelo, de forma frontal a la instalación y el sensor lo detecta y obtiene de sus gestos la información necesaria para manipular el sistema. Finalmente, la imagen que se muestra en pantalla varía siguiendo los gestos realizados por el usuario. Para la detección de gestos se decidió optar por un desarrollo ad hoc en lugar de emplear gestos estándar ya que se entendió que estos se adecuarían mucho mejor a una interacción natural y efectiva [146].

6.4.5. Implementación

La implementación del sistema se planteó como un desarrollo por prototipos, se crearon diferentes prototipos y se probaron en sujetos experimentales, hasta obtener una implementación intuitiva del sistema, de modo que este a penas necesitase de instrucciones de uso y la interacción con el mismo surgiese de forma natural. También se tuvieron en cuenta diversos factores para evitar la fatiga del usuario [92] como una posición natural y relajada del mismo

durante la interacción, la no necesidad de mantener una tensión postural innecesaria en manos o cabeza para tener una interacción efectiva, etc.

6.4.5.1. Prototipo 1

La exploración tradicional de modelos 3D se basa en el zoom, la rotación y la traslación [101] convirtiendo la interacción del usuario con el sistema (generalmente a través de interfaces 2D) en algo complejo y que requiere de un aprendizaje. Sin embargo, el paradigma planteado reduce la interacción al desplazamiento de planos a través de los ejes cartesianos, lo que simplifica el uso del sistema al ser estas transformaciones independientes. Por ser tres los planos de corte planteados, se definieron tres zonas de acción distintas que permiten al usuario operar en cada una sobre un plano de corte distinto.

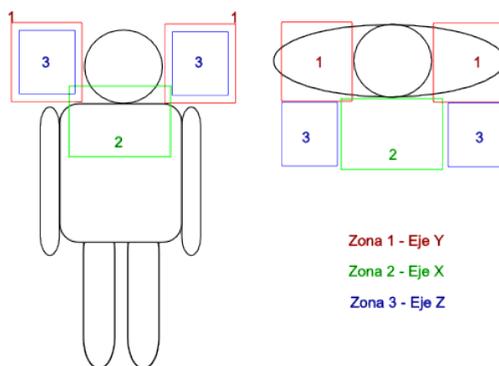


Figura 6.14: Zonas de acción del prototipo 1.

Se definieron las zonas de interacción que pueden verse en la 6.14 y cada una de ellas se asoció con un eje. Este primer prototipo restringe las zonas de acción al entorno de la cabeza del usuario asumiendo una metáfora exploratoria basada en que si se explora el busto, la forma más simple de interacción para el usuario era asumir que lo que está visualizando es su propio cráneo y que, al desplazar las manos en torno a él, puede explorar su interior. La zona 1 se situó frente al pecho y permite realizar cortes en el modelo en el sentido horizontal (eje X). La zona 2, a la derecha o izquierda de la cabeza, permite realizar cortes en sentido vertical (eje Y) y la zona 3, situada delante de la zona 2 permite explorar el interior del busto en profundidad (eje Z).

El prototipo se probó de forma empírica con 6 usuarios experimentales (entre los que se encontraban los organizadores de la exposición) que utilizaron el sistema en base a objetivos (detectar el nervio óptico, visualizar el cerebelo, etc.). A pesar de las ventajas que presentaba esta propuesta, como el ser válida tanto para zurdos como para diestros o lo acertado de la metáfora de interacción que facilitaba la comprensión del funcionamiento al usuario, empíricamente se pudieron detectar una serie de problemas operativos:

- Problemas derivados del poco recorrido de la mano entorno a la cabeza, ya que el modelo 3D se construye a partir de 275 cortes y el desplazamiento de la mano para mover el plano a través del mismo se reducía a 25 centímetros. Esto dificultaba la visualización de pequeños detalles al avanzar el plano de corte una gran distancia con un movimiento sutil de la mano (11 imágenes por centímetro de recorrido de la mano).
- El ruido en la señal producida por el sensor derivaba en un movimiento continuo de los planos de corte, impidiendo una visualización clara de los mismos. Este ruido se veía incrementado por la cercanía de las manos a la cabeza.
- La indefinición entre zonas de acción. Cuando la mano pasa de una zona de acción a otra la perspectiva que visualiza el usuario cambia, confundiéndolo al no ser capaz de identificar los límites de las diferentes zonas de acción, situación a la que contribuye el tamaño reducido de las mismas.

6.4.5.2. Prototipo 2

Dados los problemas que planteaba el prototipo 1 se decidió implementar dos cambios básicos en la concepción del sistema.

Por un lado se decidió aumentar el recorrido de la mano ampliando las zonas de acción que pasaron del entorno de la cabeza a ocupar todo el entorno del tronco (de 25 cm a 50 cm de recorrido) para permitir al usuario tener un mayor control del detalle que deseaba ver en cada momento (5,5 imágenes por cada centímetro de recorrido de la mano). Al ser más grande el rango de movimiento de la mano el plano se mueve más lentamente respecto al movimiento de la misma permitiendo observar detalles que, de otra forma, serían imposibles de visualizar y reduciendo, de esta forma, la intensidad de los temblores en la imagen debidos al ruido de la señal del sensor.

Por otro lado, se decidió también hacer una mayor diferenciación de zonas de trabajo, para que en cada zona espacial sucediese sólo una cosa, buscando simplificar de esta forma,

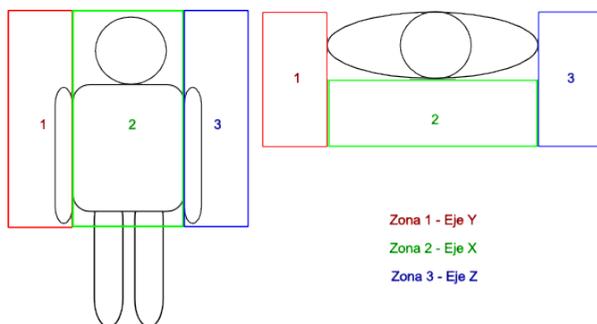


Figura 6.15: Zonas de acción del prototipo 2.

la comprensión de qué se manipulaba en cada zona. Se definieron las tres zonas como se muestra en la 6.15. A la derecha del hombro derecho se permite al usuario interactuar con el eje vertical de corte, en la zona delantera del torso y cabeza se interactúa con el eje horizontal de corte y a la izquierda del hombro izquierdo, se realiza la interacción del usuario con el eje de corte de profundidad.

Para probar el funcionamiento de este prototipo se realizaron pruebas empíricas a 7 usuarios no expertos. Estas pruebas estaban compuestas por 9 tareas cronometradas (localizar el nervio óptico, localizar el bulbo raquídeo, etc.) y un breve cuestionario para evaluar el nivel de frustración del usuario así como la usabilidad del sistema. Las tareas fueron realizadas con éxito por la totalidad de los usuarios, que invirtieron un tiempo medio de 13 segundos en realizar cada una de ellas. De las respuestas al cuestionario se dedujo que los usuarios consideraban el prototipo atractivo, así como intuitivo una vez habían interiorizado la mecánica de funcionamiento del mismo. Sin embargo, en los momentos iniciales, se sentían perdidos al no saber qué dirección exploraban con cada mano, lo que les producía frustración al pasar de una tarea a otra. También salió a relucir la dificultad de algún usuario diestro para manejar el plano de profundidad (Z) con la mano izquierda.

Al ser la primera impresión y el manejo intuitivo del sistema de capital importancia en un museo interactivo, así como la accesibilidad y que la aplicación sea usable para el usuario independientemente de su pericia con su mano menos hábil, se decidió cambiar radicalmente el enfoque que se estaba dando a la interacción, eliminando por completo las zonas de actuación diferenciadas.

6.4.5.3. Prototipo 3

En el tercer prototipo se cambió el enfoque de la interacción planteada, centrando la misma en la dirección principal de movimiento de la mano. Esta solución consiste en que la visualización es modificada y el objeto explorado en la dirección del eje en que la mano del usuario presente un mayor desplazamiento. Esta solución plantea tres ventajas principales respecto a las dos anteriores:

- Permite una diferenciación más intuitiva del eje de corte que se está utilizando en cada momento al no importar en que parte del espacio se encuentre la mano, sino la dirección en la que se mueve.
- Permite la utilización de ambas manos indiferentemente.
- Ofrece la posibilidad de tener un mayor rango de movimiento en los tres ejes, al no estar este limitado por su correspondiente zona acción.

Se repitieron las pruebas que se habían realizado en el prototipo anterior sobre los mismos 7 usuarios. De nuevo las 9 tareas fueron completadas con éxito por todos los participantes no observándose diferencias notables en los tiempos empleados en uno y otro caso (en torno a los 13 segundos de media) para completarlas. Sin embargo si se apreciaron diferencias en las respuestas al cuestionario. Se subrayaba la facilidad para el manejo del mismo, lo intuitivo que resultaba y lo preciso que era en el movimiento. Se señalaba también alguna dificultad a la hora de cambiar de tipo de vista, pero la valoración general, tanto de usabilidad como de atractivo, fue muy positiva eliminando la frustración inicial que si habían sentido en la utilización del anterior prototipo.

Se determinó, por tanto que este sería el prototipo final y se pasó a su implementación para ser probado con una muestra poblacional más representativa de los futuros usuarios que el sistema tendría en el museo.

En esta implementación final se determinó que el sistema reaccionaría a movimientos de las manos de los usuarios cuando éstas se encontrasen por encima de la cintura y que, en caso de no ser así, el sistema permanecería en reposo. Asimismo se decidió que en caso de interacción de las dos manos al mismo tiempo se le daría preferencia a la derecha por ser la de mayor frecuencia poblacional.

La configuración hardware final del sistema siguió el diseño planteado inicialmente, adaptándose a las características del espacio de que disponía en el museo (figura 6.16).

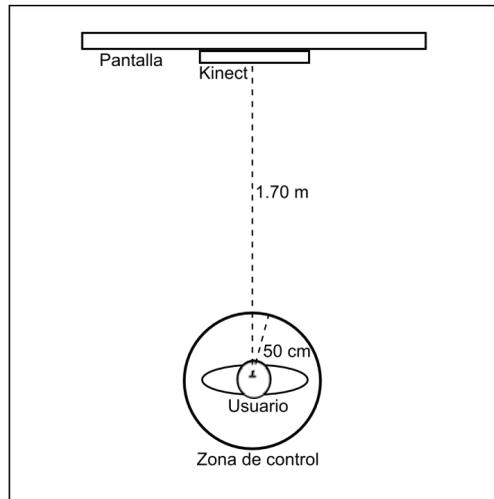


Figura 6.16: Configuración final del sistema

6.4.6. Pruebas

En la fase de prototipado también se observó la necesidad de dar un punto visual de referencia para que el usuario supiese en todo momento que plano de corte estaba visualizando y en qué posición respecto al busto se encontraba. Para esto, en una primera aproximación se optó por replicar el modelo en perspectiva y situarlo en el interior de un cubo transparente a través del cual se desplazaban los planos de corte de forma visual y solidaria con el plano de corte real. Esta visualización, a pesar de lo vistoso no resultó efectiva para conocer la posición exacta del plano respecto al modelo (por motivos de perspectiva) por lo que se decidió optar por una solución más esquemática y que transmitiera la información de una forma más simple y directa. Se emplearon dos siluetas (una de frente y otra de perfil) que se visualizan en paralelo a la imagen del modelo 3D. Según el plano de corte que se esté explorando en cada momento varía la silueta que se está visualizando y, dependiendo de la posición del plano que actúa sobre el modelo, se desplaza una línea a través de la silueta indicándole al usuario qué zona del busto está visualizando. Dos flechas indican la dirección en la que se está desplazando el plano en cada momento simplificando el control por parte del usuario así como la comprensión del funcionamiento del sistema 6.18.

Aprovechando una serie de visitas de escolares al centro de investigación y dado el enfoque familiar del museo se realizaron una serie de pruebas de uso del sistema en las que

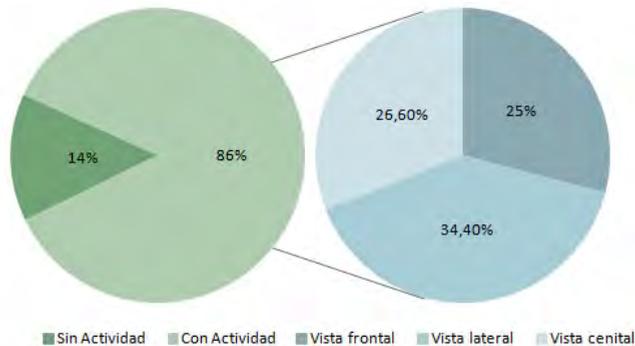


Figura 6.17: Resultados del proceso de pruebas en porcentajes de tiempo

participaron 41 estudiantes de entre 8 y 9 años. La prueba consistió en dejarlos manipular libremente el sistema sin ponerles límite de actuación o tiempo y una puesta en común posterior de la experiencia. Se debe de subrayar que el tiempo de utilización no fue totalmente libre al encontrarse la prueba en medio de una visita que debía de cumplir unos tiempos aunque sí que es un escenario que se puede asemejar a una visita de escolares en grupo al museo en la que hay un tiempo limitado y numerosas instalaciones para visitar.

El tiempo medio que estuvieron interactuando con el sistema fue de 61 segundos de los cuales un 75

Dentro de la fase de actividad (ya sea con la mano derecha o la izquierda) los usuarios interactuaron con la vista frontal (eje Z) un 29% del tiempo, con la vista superior (eje Y) un 40% del tiempo y con la vista lateral (eje X) un 31% (figura 6.17). Destaca ligeramente la vista superior al ser la que posee una mayor resolución, al tratarse de los cortes originales a partir de los cuales se realizó la reconstrucción del modelo 3D. Esta mayor nitidez se refleja en un mayor interés de los usuarios por los pequeños detalles visibles en ese eje.

Más allá de los datos obtenidos, los escolares afirmaron sentir una mayor curiosidad por el sistema que si hubiese sido manejado a través de un periférico clásico y que la interacción les había resultado más amena. Insistiendo repetidamente en que les habría gustado probar un sistema similar a través del que se pudiese observar el cuerpo humano al completo.

Desde el punto de vista de una supervisión visual exterior se puede afirmar que, en general, este tipo de interacción atraía la atención de los usuarios pero, en ciertos momentos se convertía más en una finalidad que en un medio, ocupando el juego de “retar” a la máqui-

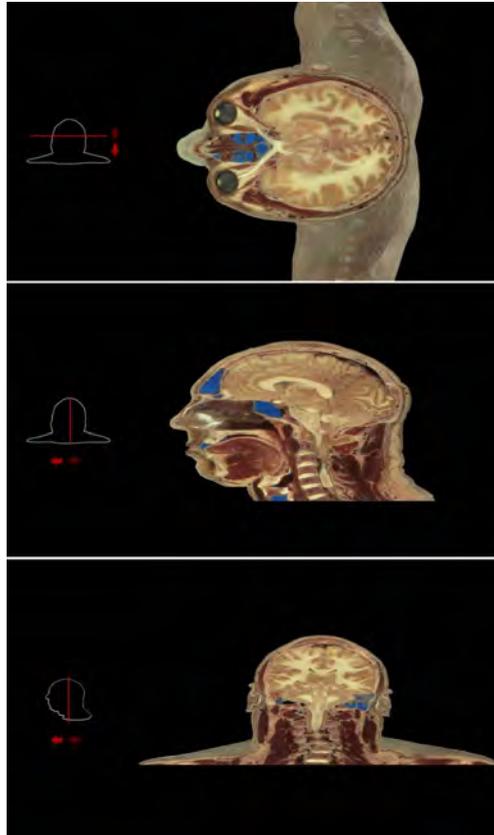


Figura 6.18: Diferentes visualizaciones presentes en la aplicación final.

na con movimientos rápidos una mayor parte de la atención del niño que la visualización propiamente dicha.

6.4.7. Conclusiones y trabajo futuro

Se ha desarrollado un sistema de visualización y exploración del cerebro humano a partir de una visualización basada en texturas 3D, controlado de una manera natural e intuitiva a través de una interfaz natural para su exhibición en el museo de divulgación científica Casa Domus. El funcionamiento del sistema se ha ido probando a lo largo de su desarrollo por prototipos y se ha validado en una prueba final con usuarios potenciales del mismo.

El sistema ha sido instalado en el museo, con gran aceptación por parte de los organizadores y visitantes de la exposición.

Como trabajo futuro se plantea la mejora del cambio de plano de corte utilizado en cada momento, ya que, debido al ruido de la señal del sensor se descartan los pequeños cambios en la dirección del movimiento para esta finalidad, dificultando, en ocasiones, la interacción del usuario con el sistema.

También se plantea la necesidad de introducir información visual y auditiva sobre los diferentes órganos y tejidos que se encuentran en el busto, indicando su nombre, una breve explicación de su funcionalidad etc. Esto incrementaría, tanto el valor divulgativo de la instalación, como el interés del usuario.

Como resultado de este proyecto se presentaron artículos en el XVI Congreso Internacional de Interacción Persona-Ordenador [125] y la revista NOVATICA [126].

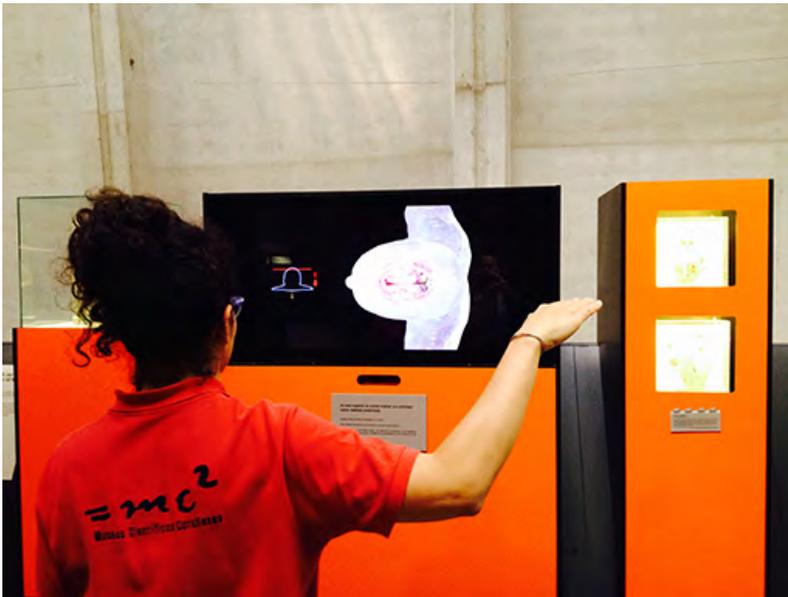


Figura 6.19: Instalación final en el museo.

CAPÍTULO 7

DISCUSIÓN DE RESULTADOS, CONCLUSIONES, Y TRABAJO FUTURO

7.1. Introducción

Una vez expuesto el trabajo de investigación llevado a cabo, es momento de extraer aquellas conclusiones más relevantes, así como las aportaciones originales realizadas e identificar los campos de trabajo futuro que se abren a partir de los desarrollos presentados a lo largo de los capítulos anteriores.

Debido al amplio espectro de tópicos abordado en esta tesis resulta útil realizar un acercamiento compartimentado a las conclusiones, resultados y trabajo futuro en cada uno de ellos. De este modo, además de unos resultados generales directamente relacionados con los objetivos de la tesis, podemos agrupar el trabajo realizado en cuatro grandes categorías: diseño e implementación de una nueva arquitectura escalable y distribuida para platós virtuales de televisión, interacción en platós virtuales de televisión, visualización avanzada en platós virtuales de televisión y visualización avanzada e interacción en entornos museísticos.

7.2. Discusión de resultados

A continuación se presenta un recorrido pormenorizado a través de los diferentes tópicos que se han abordado en la presente tesis, presentando los resultados obtenidos en cada uno de ellos.

7.2.1. Diseño e implementación de una nueva arquitectura escalable y distribuida para platós virtuales de televisión

La interacción real del presentador con el entorno virtual en los platós virtuales de televisión, como se pudo comprobar mediante el análisis de casos reales de uso actuales, es prácticamente inexistente. Esto viene dado por la arquitectura clásica que contemplaba un único sistema de sensores para realizar el seguimiento de las cámaras y dejaba de lado el seguimiento de otros elementos. Para solucionar esta problemática se ha diseñado, implementado y probado una evolución de la arquitectura tradicional. Se trata de una arquitectura distribuida y escalable que permite incluir, eliminar y actualizar dispositivos del plató de una forma ágil y que, haciendo uso de servidores VRPN, permite que estas acciones sean realizadas por usuarios no expertos. Asimismo, permite que los diferentes dispositivos añadidos funcionen de forma sinérgica, trabajando en paralelo para dotar, en conjunto, de mayores capacidades al plató virtual de televisión. Utilizando esta nueva arquitectura pueden realizarse diferentes sensorizaciones combinando dispositivos de forma que unos cubran las carencias de otros para obtener finalmente las funcionalidades deseadas para un programa específico.

La nueva arquitectura ha sido implementada sobre un plató tradicional y se ha probado su funcionamiento mediante la inclusión y prueba de tres sensores y la actualización de uno de ellos. Los dispositivos incluidos han sido el sistema de cámaras de infrarrojos de Optitrack, Leap Motion y Microsoft Kinect, siendo este último substituido por su segunda versión, Microsoft Kinect V2, cuando esta estuvo disponible en el mercado. Todos estos sensores son de bajo coste, lo que permite reducir el precio de esta tecnología y acercarla a nuevos mercados como pueden ser las cadenas de televisión locales de bajo presupuesto, la emisión por internet o la docencia a distancia.

7.2.2. Interacción en platós virtuales de televisión

La combinación de sensores incluida en la nueva arquitectura del plató de trabajo ha permitido aumentar sus capacidades interactivas en diferentes direcciones:

- Ha permitido el movimiento libre de cámaras en el escenario gracias al sistema de cámaras Optitrack.
- Se ha podido realizar el seguimiento de objetos reales y, por tanto, su utilización como interfaz tangible para la manipulación de elementos virtuales mediante las cámaras

Optitrack.

- Ha hecho posible la automatización del proceso de *distance keying* gracias a los sensores Microsoft Kinect primero y Microsoft Kinect V2 en una segunda fase.
- Se ha conseguido una manipulación directa de objetos por parte del presentador haciendo uso del Microsoft Kinect V2 y del Leap Motion.
- Se ha implementado la detección de gestos para el control de efectos en el plató. Se han hecho pruebas con la caracterización de acciones mediante reglas y mediante la utilización de inteligencia artificial. Este tipo de interacción puede ser llevado a cabo utilizando los sensores Microsoft Kinect original y V2 o el sensor Leap Motion.

La utilización de la inteligencia artificial para el aprendizaje de los gestos naturales de los actores ha permitido proponer un cambio de flujo del flujo de trabajo previo a la grabación de un programa en directo en plató virtuales de televisión, substituyendo las semanas de ensayo previo de coreografías por la proposición por la utilización de una interacción real propuesta por el presentador. El hecho de que el sistema sea capaz de aprender los gestos naturales de los actores permite que estos diseñen sus propias acciones y que al ser partícipes del proceso necesiten un menor tiempo de ensayo para dominar el manejo del sistema reduciendo también los niveles estrés durante la grabación. Se ha realizado un estudio preliminar para determinar la validez del sensor Microsoft Kinect V2 y las herramientas incluidas en su SDK para la detección de gestos naturales, obteniendo resultados prometedores.

7.2.3. Visualización avanzada en plató virtuales de televisión

En el campo de la visualización avanzada aplicada a plató virtuales de televisión se ha avanzado en dos aspectos fundamentales: por un lado, se ha propuesto un método rápido y visual para verificar el estado de la iluminación del ciclorama, permitiendo comprobar, sin la necesidad de un experto, el correcto estado de la instalación de luz. El proceso hace uso de imágenes HDR que pueden ser obtenidas de una manera sencilla a través de software específico y permite descubrir variaciones en la intensidad de ciertas lámparas, desplazamientos o movimiento de algún foco, etc., asegurando una iluminación difusa y uniforme del fondo y por tanto una configuración óptima para aplicar la técnica del *chroma key*.

Por otro lado, se ha utilizado la información obtenida en el análisis de la iluminación del ciclorama para iluminar el entorno sintético y así unificar las luces de ambos mundos

mejorando la composición final. Para ello se han obtenido luces puntuales a partir de la imagen HDR que son mapeadas a una esfera de radio variable y exportadas a un archivo de texto para su posterior utilización. Se ha propuesto el uso de estas luces obtenidas directamente del escenario, a través de las técnicas de “*render* con muchas luces”, en eStudio y se han hecho pruebas preprocesando un entorno sintético y almacenando la información de la iluminación en las texturas de los objetos a través del *baking* de texturas, mostrando una mejora evidente en la calidad de la imagen y la composición. Fruto de este proceso se ha propuesto un flujo de trabajo para la captura de la iluminación real de un escenario y su aplicación a platós virtuales de televisión que permite obtener una iluminación coherente en ambos mundos afrontando este proceso de una manera ordenada y clara.

También se ha estudiado la utilización de la proyección de sombras del presentador sobre el escenario virtual. Se han propuesto diferentes algoritmos para su implementación en eStudio y se han hecho pruebas con una sombra constante en el suelo del entorno virtual y seguimiento de la posición del presentador mediante el sensor Microsoft Kinect V2. También se ha probado la implementación de los mapas de sombras de varianza con un maniquí simple formado por esferas proyectando sombras difusas sobre un escenario, no pudiendo probarlo en eStudio.

7.2.4. Visualización avanzada e interacción en entornos museísticos

Parte de las técnicas y tecnologías utilizadas en el desarrollo de la nueva arquitectura para platós virtuales de televisión han sido utilizadas durante el desarrollo de esta tesis en el diseño e implementación de varias aplicaciones con fines educativos y de puesta en valor del patrimonio.

Dentro del proyecto “El Pórtico de la Gloria Virtual” se han implementado tres sistemas:

- **El Pórtico de la Gloria Virtual:** un recorrido virtual a través del monumento que le permite al usuario desplazarse a voluntad y en tiempo real a través de una reconstrucción 3D del Pórtico de la Gloria obtenida a partir de un escaneo láser. El control del sistema se realiza de forma remota a través de un ordenador “todo en uno” con pantalla táctil y la conexión entre este y el ordenador encargado de realizar el *render* del modelo 3D se realiza a través de VRPN.
- **Los instrumentos virtuales del Pórtico de la Gloria:** se trata de una instalación museística que permite, mediante la manipulación de una interfaz tangible y la utilización

de una cámara de vídeo y marcadores de realidad aumentada, la observación y manipulación de los instrumentos musicales representados en el Pórtico de la Gloria. La conexión entre ARtoolkit (la librería de realidad aumentada utilizada) y la aplicación de *render*, pese a ejecutarse en el mismo ordenador, se realiza a través de VRPN.

- **Conoce el Pórtico:** aplicación para dispositivos móviles Apple que permite simular la experiencia de “El Pórtico de la Gloria Virtual” sin necesidad de ir físicamente al museo.

Dentro de esta línea de trabajo se ha implementado un módulo para la visualización del interior del busto humano, que en la actualidad forma parte de la exposición permanente “Juego de neuronas” de La Domus, casa del hombre de A Coruña. El control de la aplicación encargada del *render* se realiza a través de un sensor Microsoft Kinect conectado a la misma haciendo uso de las librerías VRPN.

Todos estos desarrollos han estado (o aún están) expuestos en diversos espacios (Cripta del Pórtico de la Gloria, sede de la Fundación Pedro Barrié de la Maza, Victoria and Albert Museum, Casa Domus, etc.) con gran éxito y aceptación entre el público. En el caso de los relacionados con el Pórtico de la Gloria incluso han recorrido, en una exposición itinerante, diversas ciudades de Europa.

La separación del dispositivo de control de la parte de *render* que implica la utilización de VRPN permite la actualización o sustitución del periférico sin afectar a la parte de visualización o lógica de la aplicación. Esta capacidad fue utilizada para sustituir, en la exposición itinerante de “Los instrumentos virtuales del Pórtico de la Gloria”, la cámara de vídeo y la interfaz tangible, susceptibles a errores de calibración, por un *trackball* más robusto en los desplazamientos. Del mismo modo, se implementó un prototipo para el control de la visualización del interior del busto humano haciendo uso del sensor Leap Motion. El utilizar una arquitectura de aplicación distribuida permitió que estas modificaciones se realizaran de una forma ágil y sencilla, no afectando al comportamiento del *render* ni a la lógica de la aplicación. Por tanto, en aplicaciones de tipo museístico que hacen uso de dispositivos de entrada complejos, es recomendable, siempre que sea posible, separar la parte de control de los periféricos del resto de la aplicación ya que esto hará posible variar el modo de interacción del usuario final sin modificar el comportamiento del sistema.

7.3. Conclusiones

Se puede afirmar que se han abordado con éxito los diferentes objetivos de la investigación, obteniendo las siguientes conclusiones de la misma:

1. Se ha diseñado, implementado y probado una nueva arquitectura distribuida y escalable que permite la inclusión, eliminación y actualización de sensores de forma ágil a través del uso de servidores VRPN de terceros y archivos de configuración XML permitiendo el funcionamiento sinérgico de múltiples dispositivos de bajo coste.
2. Se han implementado y probado tres formas diferentes de interacción: la automatización del posicionamiento relativo de los elementos virtuales y reales, la manipulación directa de objetos sintéticos y la activación de efectos visuales mediante la detección de gestos.
3. Se ha propuesto y probado el uso del algoritmo de muestreo de corte medio para la captura, el análisis y la utilización de la iluminación real del escenario en entornos sintéticos. Asimismo se han utilizado técnicas como el *baking* de texturas o el *ambient occlusion* para mejorar el realismo y la integración del mundo virtual con el real. También se han propuesto y probado soluciones para la problemática de la proyección de sombras del presentador sobre los elementos virtuales.
4. Se ha desarrollado un nuevo flujo de trabajo que pretende formalizar una metodología de captura de la iluminación real del escenario y su aplicación sobre la escena sintética para la obtención de una mejor composición entre los mundos real y virtual.
5. Se ha desarrollado un nuevo flujo de trabajo para el proceso de preparación previo a una emisión en directo. En él se hace uso de técnicas de inteligencia artificial para el aprendizaje, por parte del sistema, de la gestualidad natural del presentador y su aplicación posterior a la interacción con el entorno virtual, reduciendo los tiempos de ensayo necesarios para la preparación de una emisión en directo.
6. Se han implementado cuatro sistemas para museos interactivos que hacen uso de técnicas de visualización avanzada y metáforas de interacción novedosas para transmitir el conocimiento a los usuarios de una manera atractiva y diferente. Tres de ellos siguen, a día de hoy, expuestos en museos con un gran éxito de público.

Asimismo, se han abordado varios de los retos planteados por Wojdala en el 98 y que, a día de hoy, siguen sin una solución definitiva:

1. La mejora del realismo de la imagen manteniendo el tiempo real de emisión
2. La correcta alineación de las cámaras reales y virtuales
3. La iluminación coherente de los mundos virtual y real
4. La automatización del *distance keying*
5. La interacción de los actores con los elementos virtuales.

Finalmente, cabe destacar el establecimiento de un marco de colaboración entre la Universidad de Santiago de Compostela y la multinacional Brainstorm Multimedia que ha tenido como resultado el acceso gratuito del doctorando a dos cursos de formación así como la obtención por parte de la Universidad de Santiago de Compostela de 24 licencias de software de forma gratuita. Asimismo, se ha colaborado con la empresa en la realización de pruebas de productos en fase beta para su validación. La transferencia de los avances realizados en la presente tesis se abordará tras la defensa de la misma.

7.4. Trabajo futuro

En el campo de la interacción persona-ordenador en platós virtuales de televisión, se plantea como trabajo futuro la utilización de la nueva arquitectura y sensores de bajo coste, integrados en el plató de la Facultad de Ciencias de la Comunicación de la Universidad de Santiago de Compostela, en la grabación de un programa profesional que haga uso de las capacidades ampliadas del plató virtual de televisión como son los diferentes tipos de interacción entre el presentador y su entorno y el movimiento libre de cámaras. Por otro lado, se propone una continua renovación y evolución de la configuración del plató mediante la inclusión de nuevos sensores que puedan aportar nuevas capacidades. Asimismo, es necesario probar la arquitectura para una comunicación bidireccional permitiéndole al sistema actuar sobre el mundo real, a través de dispositivos de salida, a partir de información generada de forma sintética. De este modo se podrá dotar a los actores de diferentes tipos de retroalimentación que les permita orientarse con mayor facilidad en el escenario vacío. Se plantea también la integración de la detección de instrucciones por voz que permitan a los presentadores activar efectos mediante una serie de frases predefinidas. Del mismo modo, es necesario profundizar en las técnicas y tecnologías de detección de gestos, buscando acercarse lo más posible el ratio ideal de un 100% de efectividad en el reconocimiento. Finalmente, se deben de buscar estrategias que

permitan un *distance keying* complejo, incluyendo el control de varios actores e implicando todas las partes de su anatomía en el cálculo de la posición relativa entre ellos y los diferentes objetos virtuales del entorno 3D.

En lo relativo a la aplicación de técnicas de visualización avanzada a escenarios virtuales de televisión, se hace necesaria la inclusión en eStudio tanto de algoritmos que simulen una iluminación global en tiempo real como de la posibilidad de utilizar maniqués virtuales invisibles que, ligados a una sensorización del presentador, permitan que este proyecte sombras blandas sobre los objetos virtuales en directo. Estos dos elementos, combinados con la captura de iluminación real del escenario implementada, permitirían un gran avance en la calidad de la integración entre los mundos real y sintético mejorando la sensación de presencia de los actores.

7.5. Publicaciones

7.5.1. Artículos publicados en Revista

R. Méndez, J. Flores y R. Arenas

Interacción natural en museos, un caso real de estudio: desarrollo de un módulo para la exposición “Juego de Neuronas”

Revista: Novática

Número: 235 / Páginas: 23-28 / Fecha: 31/05/2016

ISSN: 2444-6629

R. Méndez, A. Otero, S. Jarque y J. Flores

Exploración en tiempo real de la reconstrucción virtual de los instrumentos del Pórtico de la Gloria

Revista: Virtual Archaeology Review

Volumen: 3 / Número: 6 / Páginas: 49-53 / Fecha: 11/2012

ISSN: 1989-9947

A. Otero, R. Méndez y J. Flores

Pórtico de la Gloria Virtual, aplicación interactiva en tiempo real para la exploración virtual del Pórtico de la Gloria

Revista: VAR (Virtual Archaeology Review)

Volumen: 3 / Número: 5 / Páginas: 18-23 / Fecha: 05/2012

ISSN: 1989-9947

7.5.2. Capítulos en libros

E. Castelló, R. Méndez, A. M. L. Gómez, J. Flores y E. Sánchez-Vila

Development and Creativity in Ubiquitous Technologies in University: CIDUS Contribution

Libro: Emerging Perspectives on the Mobile Content Evolution

Páginas: 284-312 / Fecha: 2015

Editorial: IGI Global

ISBN: 978-1-4666-8838-4

R. Méndez

Visualizing 3D models with fine-grain surface depth. Visualizing 3D models with fine-grain surface depth

Fecha: 2011

Editorial: LAP-Publishing

ISBN: 978-3-8443-0251-6

7.5.3. Artículos publicados en actas de congreso**R. Méndez, J. Flores, E. Castelló y R. Arenas**

Evaluación preliminar del sensor Kinect V2 para su uso en platós virtuales de televisión con interacción natural

Congreso: XVII Congreso Internacional de Interacción Persona-Ordenador

Páginas: 143 - 149 / Fecha: 14/06/2016 - 16/06/2016

Lugar de celebración: Salamanca

ISBN: 978-84-9012-629-5

R. Méndez, J. Flores, E. Castelló, R. Arenas y S. Villaroya

Sensorización avanzada para estudios virtuales de televisión

Congreso: VI Congreso de TV Digital Interactiva - IV Jornadas Iberoamericanas de Aplicaciones y Usabilidad de la TVDi 2015

Páginas: 275-281 / Fecha: 14/10/2015 - 16/10/2015

Lugar de celebración: Mallorca

ISBN: 978-84-608-5567-5

D. Luaces, S. Villarroya, R. Méndez, J. R. R. Viqueira y J. M. Cotos

Sistema para la extracción de entidades geográficas asociadas a telediarios: construcción de mapas para lengua castellana y gallega a través de HBBTV

Congreso: VI Congreso de TV Digital Interactiva - IV Jornadas Iberoamericanas de Aplicaciones y Usabilidad de la TVDi 2015

Páginas: 263-273 / Fecha: 14/10/2015 - 16/10/2015

Lugar de celebración: Mallorca

ISBN: 978-84-608-5567-5

R. Méndez, J. Flores y R. Arenas

Exploración del busto humano en tiempo real mediante interacción natural con fines educativos

Congreso: XVI Congreso Internacional de Interacción Persona-Ordenador

Páginas: 63-69 / Fecha: 07/09/2015 - 09/09/2015

Lugar de celebración: Vilanova i la Geltrú

ISBN: 978-1-4503-3463-1 pp.266-270

G. Sanmartín, J. Flores; P. Arias, J. Cudeiro y R. Méndez

Motion Capture for Clinical Purposes, an Approach Using PrimeSense sensors

Congreso: AMDO (Articulated Motion and Deformable Objects)

Páginas: 133 - 136 / Año: 11/09/2012 - 14/09/2012

Lugar de celebración: Mallorca

G. Sanmartín, J. Flores; P. Arias, J. Cudeiro y R. Méndez

Captura de movimiento para propósitos clínicos: una aproximación usando sensores PrimeSense

Congreso: XXII Congreso Español de Informática Gráfica

Páginas: 133-136 / Fecha: 11/09/2012 - 14/09/2012

Lugar de celebración: Jaén

R. Méndez, G. Sanmartín, D. Mera, J. Flores y E. Castelló

Diseño de un sistema escalable para la sensorización de platós virtuales de televisión con emisiones en directo

Congreso: XXII Congreso Español de Informática Gráfica

Páginas: 137-140 / Fecha: 11/09/2012 - 14/09/2012

Lugar de celebración: Jaén

ISBN: 978-3-905673-92-0

R. Méndez, A. Otero, S. Jarque y J. Flores

Exploración en tiempo real de la reconstrucción virtual de los instrumentos del Pórtico de la Gloria

Congreso: III Congreso Internacional de Arqueología e Informática Gráfica, Patrimonio e Innovación

Páginas: 267–271 / Fecha: 21/06/2011 - 24/06/2011

Lugar de celebración: Sevilla

ISBN: 978-84-695-6946-7

A. Otero, R. Méndez y J. Flores

Pórtico de la Gloria Virtual, aplicación interactiva en tiempo real para la exploración virtual del Pórtico de la Gloria

Congreso: III Congreso Internacional de Arqueología e Informática Gráfica, Patrimonio e Innovación

Páginas: 65-69 / Fecha: 21/06/2011 - 24/06/2011

Lugar de celebración: Sevilla

ISBN: 978-84-695-6946-7

G. Sanmartín, R. Méndez y J. Flores

Simulación de elementos elásticos mediante FFD con volúmenes de Bézier

Congreso: XX Congreso Español de Informática Gráfica

Año: 2010

Lugar de celebración: Valencia

Bibliografía

- [1] *Luminance HDR*. <http://qtpfsgui.sourceforge.net/>, 2015. [En línea; último acceso 17-agosto-2016].
- [2] A Coruña), Museos Científicos Coruñeses (Ayuntamiento de: *Domus - Museos Científicos Coruñeses (=mc2)*. <http://mc2coruna.org/es/>, 2016. [En línea; último acceso 15-abril-2015].
- [3] A. Zerroug, A. Cassinelli, M. Ishikawa: *Virtual Haptic Radar*. En *ACM SIGGRAPH ASIA 2009 Sketches*, 2009.
- [4] Agarwal, S., R. Ramamoorthi, S. Belongie y H. W. Jensen: *Structured Importance Sampling of Environment Maps*. *ACM Trans. Graph.*, 22(3):605–612, Julio 2003, ISSN 0730-0301.
- [5] Agata, H., A. Yamashita y T. Kaneko: *Chroma Key Using a Checker Pattern Background*. *IEICE - Trans. Inf. Syst.*, E90-D(1):242–249, Enero 2007, ISSN 0916-8532.
- [6] Alabbasi, H., A. Gradinaru, F. Moldoveanu y A. Moldoveanu: *Human motion tracking evaluation using Kinect V2 sensor*. En *E-Health and Bioengineering Conference (EHB), 2015*, páginas 1–4, Nov 2015.
- [7] Armenteros, M.: *Posproducción digital*. Bubok, 2011.
- [8] Armenteros, M. y M. Fernández: *Inmersión, presencia y flow*. *Contratexto*, (19):165–177, 2011.
- [9] Armenteros, M. y F. Utray: *La Señal de Vídeo*. Madrid: E-Archivos Universidad Carlos III de Madrid, 2013.

- [10] Avid: *www.avid.com*. <http://www.avid.com/>. [En línea; último acceso 19-julio-2016].
- [11] Avid: - *Orad Forum - Avid - Mastering VideoGraphics*. <http://www.orad.tv/>, 2014. [En línea; último acceso 19-julio-2016].
- [12] Avid: *MC_NitrisDX_ds_US_sec*. [https://www.avid.com/static/resources/documents/datasheets/MC_NitrisDX_ds_US_sec](https://www.avid.com/static/resources/documents/datasheets/MC_NitrisDX_ds_US_sec.pdf). 2016. [En línea; último acceso 9-abril- 2016].
- [13] Barrié, Fundación: *Fundación Barrié*. http://www.fundacionbarrie.org/programa_catedral, 2016. [En línea; último acceso 23-agosto-2016].
- [14] BBC: *BBC - Homepage*. <http://www.bbc.com/>, 2016. [En línea; último acceso 15-febrero-2016].
- [15] Behringer: *X2222USB - Analog Mixers - Mixers - Behringer - Categories - MUSIC Group*. <https://www.music-group.com/Categories/Behringer/Mixers/Analog-Mixers/X2222USB/p/P0A0L>, 2016. [En línea; último acceso 9-abril- 2016].
- [16] Bidgood, W. D. y S. C. Horii: *Introduction to the ACR_NEMA DICOM standard*. *RadioGraphics*, 12(2):345–355, March 1992.
- [17] Billinghamurst, M., H. Kato y I. Poupyrev: *The MagicBook: a transitional {AR} interface*. *Computers & Graphics*, 25(5):745 – 753, 2001, ISSN 0097-8493. Mixed realities - beyond conventions.
- [18] Bimber, O., A. Grundhöfer, S. Zollmann y D. Kolster: *Digital illumination for augmented studios*. *Journal of Virtual Reality and Broadcasting*, 2006.
- [19] Blonde, L., M. Buck, R. Galli, W. Niem, Y. Paker, W. Schmidt y G. Thomas: *A virtual studio for live broadcasting: the Mona Lisa project*. *IEEE MultiMedia*, 3(2):18–29, Summer 1996, ISSN 1070-986X.
- [20] Brooks, A. y A. Czarowicz: *Markerless motion tracking: Ms kinect & organic motion openstage*. En *9th International Conference on Disability, Virtual Reality and Associated Technologies*, volumen 9, página 435–437, 2012.
- [21] Cantos, F. J. L.: *La divulgación científica en clave de espectáculo. Simulación gráfica en programas de televisión*. En *IV Congreso Internacional Latina de Comunicación Social: Comunicación, control y resistencias*, 2012.

- [22] Carbó, F.: *El Pórtico de la Gloria: Misterio y Sentido*. Ediciones Encuentro, 2009, ISBN 987-8499200057.
- [23] Chandaria, J., G. A. Thomas y D. Stricker: *The MATRIS project: real-time markerless camera tracking for Augmented Reality and broadcast applications*. *Journal of Real-Time Image Processing*, 2(2):69–79, 2007, ISSN 1861-8219.
- [24] Clemente, J.: *Introducción al software de gestión en la producción audiovisual*. Editorial Fragua, 2004, ISBN 84-7074-167-5.
- [25] Clemente, J.: *La incorporación de las nuevas tecnologías al audiovisual: una aproximación a la informatización de la gestión de la producción audiovisual*. *Area abierta*, (9), 2004.
- [26] Corazza, S., L. Mündermann, A. M. Chaudhari, T. Demattio, C. Cobelli y T. P. Andriacchi: *A Markerless Motion Capture System to Study Musculoskeletal Biomechanics: Visual Hull and Simulated Annealing Approach*. *Annals of Biomedical Engineering*, 34(6):1019–1029, 2006, ISSN 1573-9686.
- [27] corporation, Motion Analysis: *Motion Analysis corporation, the Motion Capture Leader*. <http://www.motionanalysis.com/index.html>, 2016. [En línea; último acceso 20-abril-2016].
- [28] Corporation, Sony: *PMW-EX! (PMWEX1) : Descripción General : España : Sony Professional*. <http://www.sony.es/pro/product/broadcast-products-camcorders-xdcam/pmw-ex1/overview/>, 2016. [En línea; último acceso 9-abril- 2016].
- [29] cplusplus.com: *cplusplus.com - The C++ Resources Network*. <http://www.cplusplus.com/>, 2016. [En línea; último acceso 17-agosto-2016].
- [30] Creative Technologies, USC Institute for: *www.hdrshop.com*. <http://www.hdrshop.com/>, 2013. [En línea; último acceso 17-agosto-2016].
- [31] Crow, F. C.: *Summed-area Tables for Texture Mapping*. *SIGGRAPH Comput. Graph.*, 18(3):207–212, Enero 1984, ISSN 0097-8930.
- [32] Cubillo, E. Galán: *Escenografía virtual en TV. Análisis del uso de escenografía virtual en la realización de un programa de televisión*. *Revista Latina de Comunicación Social*, (63):31–42, 2008.

- [33] Cubillo, E. Galán: *El trabajo del presentador de televisión en un escenario virtual*. Revista Latina de Comunicación Social, (64):143–150, 2009.
- [34] Cubillo, E. Galán: *La realidad virtual en televisión: el paso del átomo al bit*. Ámbitos, (19):9–24, 2010.
- [35] Cubillo, E. Galán: *Use of virtual sets in the broadcasting of major events*. Revista Latina de Comunicación Social, (66):63–78, 2011.
- [36] Dachsbacher, C., J. Křivánek, M. Hašan, A. Arbree, B. Walter y J. Novák: *Scalable Realistic Rendering with Many-Light Methods*. Computer Graphics Forum, 33(1):88–104, 2014, ISSN 1467-8659.
- [37] Daemen, J., P. Haufs-Brusberg y J. Herder: *Markerless actor tracking for virtual (TV) studio applications*. En *Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA), 2013 International Joint Conference on*, páginas 790–796, Nov 2013.
- [38] DAQRI: *Open Source agostomented Reality SDK* artoolkit.org. <https://artoolkit.org/>. [En línea; último acceso 20-febrero-2016].
- [39] Debevec, Paul E., Malik y Jitendra: *Recovering High Dynamic Range Radiance Maps from Photographs*. En *ACM SIGGRAPH 2008 Classes, SIGGRAPH '08*, páginas 31:1–31:10, New York, NY, USA, 2008. ACM.
- [40] Debevec, P.: *A Median Cut Algorithm for Light Probe Sampling*. En *ACM SIGGRAPH 2008 Classes, SIGGRAPH '08*, páginas 33:1–33:3, New York, NY, USA, 2008. ACM.
- [41] Donnelly, W. y A. Lauritzen: *Variance Shadow Maps*. En *Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games, I3D '06*, páginas 161–165, New York, NY, USA, 2006. ACM, ISBN 1-59593-295-X.
- [42] Eckerson, W.: *Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications*. En *Open Information Systems 10*, Enero 1995.
- [43] Egorova, L. y A. Lavrov: *Determination of workspace for motion capture using Kinect*. En *Power and Electrical Engineering of Riga Technical University (RTUCON), 2015 56th International Scientific Conference on*, páginas 1–4, Oct 2015.

- [44] Engelbart, D. C. y W. K. English: *A Research Center for Augmenting Human Intellect*. En *Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I*, AFIPS '68 (Fall, part I), páginas 395–410, New York, NY, USA, 1968. ACM.
- [45] English, W. K., D. C. Engelbart y M. L. Berman: *Display-Selection Techniques for Text Manipulation*. En *IEEE Transactions on Human Factors in Electronics*, volumen 8, páginas 5–15, 1967.
- [46] Fellous, A.: *Virtual Worlds and Multimedia*. páginas 127–133. John Wiley & Sons, Inc., New York, NY, USA, 1993, ISBN 0-471-9397-22.
- [47] Ferreiro, A. López: *El Pórtico de la Gloria*. Follas Novas, 1892, ISBN 978-1167511370.
- [48] Fidalgo, M. I. R. y S. M. Arnáiz: *Los nuevos documentales multimedia interactivos: construcción discursiva de la realidad orientada al receptor activo*. *Historia y Comunicación Social*, 18(0), 2014, ISSN 1988-3056.
- [49] Fitt, B. y J. Thornley: *Lighting Technology: a Guide for the Entertainment Industry*. Elsevier, 2001.
- [50] Flores, J., J. M. Ferro, J. Taboada y J. E. Arias: *The "Botafumeiro" VR: Virtual Reality in the Liturgy of the Middle Ages*. *Presence: Teleoper. Virtual Environ.*, 12(2):222–228, Abril 2003, ISSN 1054-7460.
- [51] FOR-A, HD/SD 1M/E Video Switcher HVS 300HS/300RPS Products: *Render engine for 3D real-time graphics, eStudio*. <http://www.for-a.com/products/hvs300hs/hvs300hs.html>, 2016. [En línea; último acceso 9-abril-2016].
- [52] Forster, F.: *A High-Resolution and High Accuracy Real-Time 3D Sensor Based on Structured Light*. En *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, páginas 208–215, June 2006.
- [53] Foster, J.: *The Green Screen Handbook*. Wiley Publishing Inc, 2010.
- [54] Foundation, Blender: *blender.org - Home of the Blender project - Free and Open 3D Creation Software*. <https://www.blender.org/>, 2016. [En línea; último acceso 16-agosto-2016].

- [55] Foundation, Python Software: *Welcome to Python.org*. <https://www.python.org/>, 2016. [En línea; último acceso 10-agosto-2016].
- [56] Francese, R., I. Passero y G. Tortora: *Wiimote and Kinect: Gestural User Interfaces Add a Natural Third Dimension to HCI*. En *Proceedings of the International Working Conference on Advanced Visual Interfaces, AVI '12*, páginas 116–123, New York, NY, USA, 2012. ACM, ISBN 978-1-4503-1287-5.
- [57] Gardner, H.: *Frames of Mind: The Theory of Multiple Intelligences*. Basic Books, 1993, ISBN 9780465025107.
- [58] Geomagic: *Geomagic Community*. <http://www.geomagic.com/en/community>, 2016. [En línea; último acceso 23-agosto-2016].
- [59] Ghosh, A., M. Trentacoste, H. Seetzen y W. Heidrich: *Real Illumination from Virtual Environments*. En *ACM SIGGRAPH 2005 Sketches*, SIGGRAPH '05, New York, NY, USA, 2005. ACM.
- [60] Gibbs, S., C. Arapis, C. Breiteneder, V. Lalioti, S. Mostafawy y J. Speier: *Virtual studios: an overview*. *IEEE MultiMedia*, 5(1):18–35, Jan 1998, ISSN 1070-986X.
- [61] GitHub, Inc.: *Home . vrpn/vrpn Wiki . GitHub*. <https://github.com/vrpn/vrpn/wiki>, 2016. [En línea; último acceso 20-febrero-2016].
- [62] Goldsmith, A. N.: *Color micro-facsimile system*. (US 2073370), Marzo 1937.
- [63] Gonzalez-Jorge, H., B. Riveiro, E. Vazquez-Fernandez, J. Martínez-Sánchez y P. Arias: *Metrological evaluation of Microsoft Kinect and Asus Xtion sensors*. *Measurement*, 46(6):1800 – 1806, 2013, ISSN 0263-2241.
- [64] Goodall, S., P.H. Lewis, K. Martinez, P. A. S. Sinclair, F. Giorgini, M. J. Addis, M. J. Boniface, C. Lahanier y J. Stevenson: *SCULPTEUR: Multimedia Retrieval for Museums*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, ISBN 978-3-540-27814-6.
- [65] Grau, O., T. Pullen y G. A. Thomas: *A combined studio production system for 3-D capturing of live action and immersive actor feedback*. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3):370–380, March 2004, ISSN 1051-8215.

- [66] Gritz, L.: *A Recipe for Texture Baking*. RenderMan in Production, SIGGRAPH, páginas 45–54, 2002.
- [67] Grundhöfer, A., M. Seeger, F. Häntsch y O. Bimber: *Coded Projection and Illumination for Television Studios*. En *EUROGRAPHICS 2007*, volumen 26, 2007.
- [68] Grundhofer, A., M. Seeger, F. Hantsch y O. Bimber: *Dynamic Adaptation of Projected Imperceptible Codes*. En *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, páginas 181–190, Nov 2007.
- [69] Guennebaud, G., L. Barthe y M. Paulin: *Real-time Soft Shadow Mapping by Back-projection*. En *Proceedings of the 17th Eurographics Conference on Rendering Techniques*, EGSR '06, páginas 227–234, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association, ISBN 3-905673-35-5.
- [70] Guna, J., G. Jakus, M. Pogacnik, S. Tomazic y J. Sodnik: *An Analysis of the Precision and Reliability of the Leap Motion Sensor and Its Suitability for Static and Dynamic Tracking*. *Sensors*, 14(2):3702, 2014, ISSN 1424-8220.
- [71] Hasenfratz, J. M., M. Lapierre, N. Holzschuch y F. Sillion: *A survey of Real-Time Soft Shadows Algorithms*. En *Eurographics State-of-The-Art Reports*, páginas 1–20. Eurographics, Eurographics, 2003.
- [72] Hayashi, M.: *Image compositing based on virtual cameras*. *IEEE MultiMedia*, 5(1):36–48, Jan 1998, ISSN 1070-986X.
- [73] Heckbert, P.: *Color Image Quantization for Frame Buffer Display*. *SIGGRAPH Comput. Graph.*, 16(3):297–307, Julio 1982, ISSN 0097-8930.
- [74] Heeter, C.: *Being There: The Subjective Experience of Presence*. *Presence: Teleoper. Virtual Environ.*, 1(2):262–271, Mayo 1992, ISSN 1054-7460.
- [75] Herder, J., J. Daemen, P. Haufs-Brusberg y I. Abdel Aziz: *Four Metamorphosis States in a Distributed Virtual (TV) Studio: Human, Cyborg, Avatar, and Bot – Markerless Tracking and Feedback for Realtime Animation Control*, páginas 16–32. Springer International Publishing, Cham, 2015, ISBN 978-3-319-17043-5.

- [76] Hough, G., C. Athwal, I. Williams y Ian: *Advanced Occlusion Handling for Virtual Studios*, páginas 287–294. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, ISBN 978-3-642-32645-5.
- [77] HP Development Company, L.P.: *HPZ800 - HP España*. <http://www8.hp.com/es/es/campaign/ebon-security/hp-z800.html>, 2016. [En línea; último acceso 9-abril-2016].
- [78] Huang, C. R., C. S. Chen y P. C. Chung: *Tangible photorealistic virtual museum*. IEEE Computer Graphics and Applications, 25(1):15–17, Jan 2005, ISSN 0272-1716.
- [79] Iddan, G.J. y G. Yahav: *Three-dimensional imaging in the studio and elsewhere*. 4298(1):48–55, 2001, ISSN 0277786X.
- [80] II, R. M. Taylor, T. C. Hudson, A. Seeger, H. Weber, J. Juliano y A. T. Helser: *VRPN: A Device-independent, Network-transparent VR Peripheral System*. En *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST '01, páginas 55–61, New York, NY, USA, 2001. ACM, ISBN 1-58113-427-4.
- [81] Inc., ASUSTeK Computer: *Multimedia - XtionPRO LIVE - ASUS*. https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/. [En línea; último acceso 2-agosto-2016].
- [82] Inc., Logbar: *Ring Zero*. <http://ringzero.logbar.jp/>, 2015. [En línea; último acceso 20-abril-2016].
- [83] Inc., Occipital: *Structure Sensor - 3D scanning, agostomented reality, and more for mobile devices*. <http://structure.io/>, 2016. [En línea; último acceso 22-febrero-2016].
- [84] Itseez: *OpenCV OpenCV*. <http://opencv.org/>, 2016. [En línea; último acceso 17-agosto-2016].
- [85] Jaksic, B., B. Gara, M. Petrovic, P. Spalevic y L. Lazic: *Analysis of the Impact of Front and Back light on Image Compression with SPIHT Method during Realization of the Chroma Key Effect in Virtual TV Studio*. Acta Polytechnica Hungarica, 12(2):71–88, 2015.
- [86] Jaksic, B. S., B. Gara, M. B. Petrovic, P. Lj. Spalevic y I. M. Milosevic: *Analysis of the Effects of Front and Back Lights in Chroma Key Effects During Implementation in*

- Virtual TV Studio*. International Journal of Computers and Communications, 7(3):90–97, 2013.
- [87] Kainz, O. y F. Jakob: *Approach to Hand Tracking and Gesture Recognition Based on Depth-Sensing Cameras and EMG Monitoring*. Acta Informatica Pragensia, 2014(1):104–112, 2014.
- [88] Kennedy, R. C. y F. J. Gaskins: *Electronic Composites in Modern Television*. Proceedings of the IRE, 46(11):1798–1807, Nov 1958, ISSN 0096-8390.
- [89] Khoshelham, K. y S. O. Elberink: *Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications*. Sensors, 12(2):1437, 2012, ISSN 1424-8220.
- [90] Kim, N., W. Woo, G. J. Kim y C. M. Park: *3-D Virtual Studio for Natural Inter-Acting*. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 36(4):758–773, July 2006, ISSN 1083-4427.
- [91] Kim, S.: *Virtual studio system for augmented reality amp; Chroma key processing*. En *Advanced Communication Technology (ICACT), 2011 13th International Conference on*, páginas 762–765, Feb 2011.
- [92] Kim, Y., G. A. Lee, D. Jo, U. Yang, G. Kim y J. Park: *Analysis on virtual interaction-induced fatigue and difficulty in manipulation for interactive 3D gaming console*. En *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, páginas 269–270, Jan 2011.
- [93] Kitware: *Kitware. Itk - segmentation and registration toolkit*. <http://www.itk.org/>. [En línea; último acceso 15-enero-2013].
- [94] Kitware: *Vtk - the visualization toolkit*. <http://www.vtk.org/>. [En línea; último acceso 15-enero-2013].
- [95] Klapdohr, M., B. Wöldecke, D. Marinos, J. Herder, C. Geiger y W. Vonolfen: *Vibrotactile Pitfalls: Arm Guidance for Moderators in Virtual TV Studios*. En *Proceedings of the 13th International Conference on Humans and Computers, HC '10*, páginas 72–80, Fukushima-ken, Japan, Japan, 2010. University of Aizu Press.
- [96] Kogler, R. A.: *Virtual set design*. IEEE MultiMedia, 5(1):92–96, Jan 1998, ISSN 1070-986X.

- [97] Kolivand, H. y M. S. Sunar: *Survey of Shadow Volume Algorithms in Computer Graphics*. IETE Technical Review, 30(1):38–46, 2013.
- [98] Kollig, T. y A. Keller: *Efficient Illumination by High Dynamic Range Images*. En *Proceedings of the 14th Eurographics Workshop on Rendering*, EGRW '03, páginas 45–50, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association, ISBN 3-905673-03-7.
- [99] Krivanek, J., M. Hasan, A. Arbre, C. Dachsbacher, A. Keller y B. Walter: *Optimizing Realistic Rendering with Many-light Methods*. En *ACM SIGGRAPH 2012 Courses*, SIGGRAPH '12, páginas 7:1–7:217, New York, NY, USA, 2012. ACM, ISBN 978-1-4503-1678-1.
- [100] Labs, Thalmic: *Myo Gesture Control Armband - Wearable Technology by Thalmic Labs*. <https://www.myo.com/>, 2016. [En línea; último acceso 20-abril-2016].
- [101] Lacolina, S. A., A. Soro y R. Scateni: *Natural Exploration of 3D Models*. En *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity*, CHIItaly, páginas 118–121, New York, NY, USA, 2011. ACM, ISBN 978-1-4503-0876-2.
- [102] Laine, S., H. Saransaari, J. Kontkanen, J. Lehtinen y T. Aila: *Incremental Instant Radiosity for Real-time Indirect Illumination*. En *Proceedings of the 18th Eurographics Conference on Rendering Techniques*, EGSR'07, páginas 277–286, Aire-la-Ville, Switzerland, Switzerland, 2007. Eurographics Association, ISBN 978-3-905673-52-4.
- [103] Lalioti, V. y A. Woolard: *Mixed Reality productions of the future*. En *IBC 2003 Conference, International Broadcasting Convention*, 2003.
- [104] Lan, Yuhai, Jing Li y Zhaojie Ju: *Data fusion-based real-time hand gesture recognition with Kinect V2*. En *2016 9th International Conference on Human System Interactions (HSI)*, páginas 307–310, July 2016.
- [105] Lange, R. y P. Seitz: *Solid-state time-of-flight range camera*. IEEE Journal of Quantum Electronics, 37(3):390–397, Mar 2001, ISSN 0018-9197.
- [106] LEDchromaKEY.com: *LEDchromaKEY.com - Light-ring and Retro-reflective Chroma-key Systems*. <http://ledchromakey.com/>, 2014. [En línea; último acceso 10-agosto-2016].

- [107] Lee, J. S., S. Park y Ch.Y. Kim: *A Study of Changes in Production by Domestic Broadcasters Using Virtual Studio*. J. Inform. and Commun. Convergence Engineering, 9:117–123, 2011.
- [108] Limited, For A Company: *HD/SD Chroma Keyer - MBP-100CK - Products - FOR-A*. <http://www.for-a.com/products/mbp100ck/mbp100ck.html>, 2016. [En línea; último acceso 9-abril- 2016].
- [109] Limited, For A Company: *Virtual System Platform - MBP-1244, MBP-12CK/12RU/RUA - Products - FOR-A*. http://www.for-a.com/products/virtual_system_platform/virtual_system_platform.html#Tab1, 2016. [En línea; último acceso 9-abril-2016].
- [110] Livingston, M. A., J. Sebastian, Z. Ai y J. W. Decker: *Performance measurements for the Microsoft Kinect skeleton*. En *2012 IEEE Virtual Reality Workshops (VRW)*, páginas 119–120, March 2012.
- [111] López-Calo, J.: *Los instrumentos del Pórtico de la Gloria. Su reconstrucción y la música de su tiempo*. Fundación Pedro Barrié de la Maza Conde de Fenosa, 1994, ISBN 84-87819-70-2.
- [112] Ltd., Adobe Systems Software Ireland: *Descargar Adobe Photosop CC Programa de edición de fotos*. <http://www.adobe.com/es/products/photoshop.html>, 2016. [En línea; último acceso 17-agosto-2016].
- [113] Ltd., Mo Sys Engineering: *mo-sys camera motion systems*. <http://www.mo-sys.com/>, 2015. [En línea; último acceso 20-abril-2016].
- [114] Ltd, Torus Knot Software: *OGRE - Open Source 3D Graphics Engine Home of a marvelous rendering engine*. <http://www.ogre3d.org>, 2016. [En línea; último acceso 23-agosto-2016].
- [115] Ltd, Vitec Videocom: *Vinten Radamec - www.vintenradamec.com*. <http://www.vintenradamec.com/>, 2016. [En línea; último acceso 9-abril- 2016].
- [116] Luaces, D., S. Villarroja, R. Méndez, J. R. R. Viqueira y J. M. Cotos: *Sistema para la extracción de entidades geográficas asociadas a telediarios: construcción de mapas para lengua castellana y gallega a través de HBBTV*. En *VI Congreso de TV Digital*

- Interactiva – IV Jornadas Iberoamericanas de Aplicaciones y Usabilidad de la TVDi*, páginas 263–273, 2015.
- [117] Ludwig, P., J. Buchel, J. Herder y W. Vonolfen: *InEarGuide - a navigation and interaction feedback system using in ear headphones for virtual tv studio productions*. En *9th Workshop Virtuelle und Erweiterte Realitat der GI-Fachgruppe VR/AR, Dusseldorf*, 2012.
- [118] Lyver, D. y G. Swainson: *Basics of Video Lighting*. Focal Press, Oxford, Inglaterra, 1999.
- [119] Marin, G., F. Dominio y P. Zanuttigh: *Hand gesture recognition with leap motion and kinect devices*. En *2014 IEEE International Conference on Image Processing (ICIP)*, páginas 1565–1569, Oct 2014.
- [120] Marinos, D., B. Wöldecke y C. Geiger: *Poster: Prototyping natural interactions in virtual studio environments by demonstrationby - Combining spatial mapping with gesture following*. En *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, páginas 153–154, March 2013.
- [121] Martín, R., J. Flores, A. Otero, E. Gutiérrez y J. Taboada: *Sistemas de visualización e interacción 3D: Un acuario virtual*. En *VI Congreso de Interacción persona-ordenador, AIPO 2005*, Sept 2005.
- [122] McCartney, R., J. Yuan y H. P. Bischof: *Gesture Recognition with the Leap Motion Controller*. En *Proceedings of the 19th International Conference on Image Processing, Computer Vision, & Pattern Recognition, IPCV'15*, 2015.
- [123] Mei, X., M. Jaeger y B. Hu: *An Effective Stratified Sampling Scheme for Environment Maps with Median Cut Method*. En *International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06)*, páginas 384–389, July 2006.
- [124] Méndez, R., Otero, A., Flores y J.: *A real-time exploration of the virtual reconstruction of the “Pórtico de la Gloria” instruments*. En *Computer applications and quantitative methods in Archaeology 2012*, 2012.
- [125] Méndez, R., J. Flores y R. Arenas: *Exploración del busto humano en tiempo real mediante interacción natural con fines educativos*. En *XVI Congreso Internacional de Interacción Persona-Ordenador*, páginas 63–69. AIPO, 2015.

- [126] Méndez, R., J. Flores y R. Arenas: *Interacción natural en museos, un caso real de estudio: Desarrollo de un módulo para la exposición “Juego de Neuronas”*. Novatica, (235):23–28, 2016, ISSN 0211-2124.
- [127] Méndez, R., J. Flores, E. Castelló y R. Arenas: *Evaluación preliminar del sensor Kinect V2 para su uso en platós virtuales de televisión con interacción natural*. En *XVII Congreso Internacional de Interacción Persona-Ordenador*, 2016.
- [128] Méndez, R., J. Flores, E. Castelló, R. Arenas y S. Villaroya: *Sensorización avanzada para estudios virtuales de televisión*. En *VI Congreso de TV Digital Interactiva - IV Jornadas Iberoamericanas de Aplicaciones y Usabilidad de la TVDi 2015*, páginas 275–281, 2015.
- [129] Méndez, R., A. Otero, S. Jarque y J. Flores: *Exploración en tiempo real de la reconstrucción virtual de los instrumentos del Pórtico de la Gloria*. En *III Congreso Internacional de Arqueología e Informática Gráfica, Patrimonio e Innovación*, página 267–271, 2011, ISBN 978-84-695-6946-7.
- [130] Méndez, R., A. Otero, S. Jarque y J. Flores: *Exploración en tiempo real de la reconstrucción virtual de los instrumentos del Pórtico de la Gloria*. *Virtual Archaeology Review*, 3(6):49–53, 2012, ISSN 1989-9947.
- [131] Méndez, R., G. Sanmartín, D. Mera, J. Flores y E. Castelló: *Diseño de un sistema escalable para la sensorización de platós virtuales de televisión con emisiones en directo*. En *XXII Congreso Español de Informática Gráfica*, 2012.
- [132] Microsoft: *C#*. <https://msdn.microsoft.com/es-es/library/kx37x362.aspx>, 2016. [En línea; último acceso 17-noviembre-2015].
- [133] Microsoft: *.NET - Powerfull Open Source Cross Platform Development*. <https://www.microsoft.com/net>, 2016. [En línea; último acceso 17-noviembre-2015].
- [134] Microsoft: *Windows Forms*. [https://msdn.microsoft.com/es-es/library/dd30h2yb\(v=vs.110\).aspx](https://msdn.microsoft.com/es-es/library/dd30h2yb(v=vs.110).aspx), 2016. [En línea; último acceso 17-agosto-2016].
- [135] Microsoft: *Windows Presentation Foundation*. [https://msdn.microsoft.com/es-es/library/ms754130\(v=vs.110\).aspx](https://msdn.microsoft.com/es-es/library/ms754130(v=vs.110).aspx), 2016. [En línea; último acceso 17-noviembre-2015].

- [136] Milgram, P., H. Takemura, A. Utsumi y F. Kishino: *Augmented reality: a class of displays on the reality-virtuality continuum*. Volumen 2351, páginas 282–292, 1995.
- [137] Mitsumine, H., Y. Yamanouchi, T. Fukaya, H. Okubo y S. Inoue: *Camera Parameter Estimation Method Using Infrared Markers for Live TV Production*. En *2008 IEEE Virtual Reality Conference*, páginas 275–276, March 2008.
- [138] Méndez, R. En *Visualizing 3D models with fine-grain surface depth*. LAP-Publishing, 2011, ISBN 978-3-8443-0251-6.
- [139] Mobsby, N.: *Lighting Systems for TV Studios*. Entertainment Technology Press, 2001.
- [140] Moshkovitz, M.: *The Virtual Studio. Technology & Techniques*. Focal Press, Oxford, Inglaterra, 2000.
- [141] Motion, Organic: *Openstage 2.0 technical overview*. <http://www.organicmotion.com/openstage-2-0-technical-overview/>. [En línea; último acceso 10-agosto-2016].
- [142] Multimedia, Brainstorm: *Infiniti Set, the best virtual studio solution*. <http://www.brainstorm3d.com/products/infinity-set/>, 2016. [En línea; último acceso 30-julio-2016].
- [143] Multimedia, Brainstorm: *Render engine for 3D real-time graphics, eStudio*. <http://www.brainstorm.es/products/estudio/>, 2016. [En línea; último acceso 30-julio-2016].
- [144] Multimedia, Brainstorm: *Virtual studios solution provider, Brainstorm Multimedia*. <http://www.brainstorm3d.com/>, 2016. [En línea; último acceso 30-julio-2016].
- [145] Munchmeyer, A. y S. Kruse: *Master Mateo – Skilled Artist or Medieval Engineer?* En *Proceedings of the Third International Congress on Construction History*, 2009.
- [146] Nielsen, M., M. Störing, T. B. Moeslund y E. Granum: *A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, ISBN 978-3-540-24598-8.
- [147] Nintendo: *Wii Remote - Nintendo - Wikia*. <http://nintendo.wikia.com/wiki/Wiimote>. [En línea; último acceso 20-abril-2016].

- [148] Openscenegraph: *Openscenegraph*. <http://www.openscenegraph.org>. [En línea; último acceso 15-enero-2013].
- [149] Ostromoukhov, V., C. Donohue y P. M. Jodoin: *Fast Hierarchical Importance Sampling with Blue Noise Properties*. *ACM Trans. Graph.*, 23(3):488–495, Agosto 2004, ISSN 0730-0301.
- [150] Otero, A., J. Flores y R. Arenas: *As carrilanas: Simulación de juegos tradicionales mediante técnicas de realidad virtual*. En *Arqueológica 2.0. I Congreso Internacional de Arqueología e Informática Gráfica*, June 2009.
- [151] Otero, A., J. Flores, R. Martín y E. Gutiérrez: *Técnicas de Visualización Avanzada y Museística: Un Planetario Virtual*. En *VI Congreso de Interacción persona-ordenador, AIPO 2005*, Sept 2005.
- [152] Otero, A., R. Méndez y J. Flores: *Pórtico de la Gloria Virtual, aplicación interactiva en tiempo real para la exploración virtual del Pórtico de la Gloria*. En *III Congreso Internacional de Arqueología e Informática Gráfica, Patrimonio e Innovación*, páginas 65–69, 2011, ISBN 978-84-695-6946-7.
- [153] Otero, A., R. Méndez y J. Flores: *Pórtico de la Gloria Virtual, aplicación interactiva en tiempo real para la exploración virtual del Pórtico de la Gloria*. *Virtual Archaeology Review*, 3(6):49–53, 2012, ISSN 1989-9947.
- [154] Petrovic, M., B. Jaksic, P. Spalevic, I. Petrovic y V. Dakovic: *The Analysis Background on the Effect of Chroma-Key in Virtual TV Studio*. *INFOTECH 12*, páginas 937–941, 2012.
- [155] Point, Natural: *Flex 3 - An affordable motion capture camera - Optitrack*. <https://www.optitrack.com/products/flex-3/>, 2016. [En línea; último acceso 23-mayo-2016].
- [156] Point, Natural: *Motive:Tracker - Motion capture and 6DOF object tracking - Optitrack*. <https://www.optitrack.com/products/motive/tracker/>, 2016. [En línea; último acceso 23-mayo-2016].
- [157] Point, Natural: *Optitrack - Motion Capture Systems - Optitrack*. <http://optitrack.com/>, 2016. [En línea; último acceso 23-mayo-2016].

- [158] Pomi, A. y P. Slusallek: *Interactive Ray Tracing for Virtual TV Studio Applications*. Journal of Virtual Reality and Broadcasting, 2(1), 2005, ISSN 1860-2037.
- [159] Rahbar, K. y H. Reza Pourreza: *Inside looking out camera pose estimation for virtual studio*. Graphical Models, 70(4):57 – 75, 2008, ISSN 1524-0703.
- [160] Red.es: *Profesionales Digitales*. <http://www.profesionalesdigitales.es/profdig/>. [En línea; último acceso 1-agosto-2016].
- [161] Rey, B., V. Parkhutik y M. Alcañiz: *Breaks in Presence in Virtual Environments: An Analysis of Blood Flow Velocity Responses*. Presence, 20(3):273–286, June 2011, ISSN 1054-7460.
- [162] Schapire, R. E.: *A Brief Introduction to Boosting*. En *Proceedings of the 16th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'99*, páginas 1401–1406, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.
- [163] Schiller, I. y R. Koch: *Data Structures for Capturing Dynamic Scenes with a Time-of-Flight Camera*. Dynamic 3D Imaging Workshop at DAGM 2009, Dyn3D, páginas 42–57, 2009.
- [164] Shanmugam, P. y O. Arikan: *Hardware Accelerated Ambient Occlusion Techniques on GPUs*. En *Proceedings of the 2007 Symposium on Interactive 3D Graphics and Games, I3D '07*, páginas 73–80, New York, NY, USA, 2007. ACM, ISBN 978-1-59593-628-8.
- [165] Shim, W. y G. J. Kim: *Designing for Presence and Performance: The Case of the Virtual Fish Tank*. Presence, 12(4):374–386, Aug 2003, ISSN 1054-7460.
- [166] Shimoda, S., M. Hayashi y Y. Kanatsugu: *New chroma-key imaging technique with Hi-Vision background*. IEEE Transactions on Broadcasting, 35(4):357–361, Dec 1989, ISSN 0018-9316.
- [167] Simsch, J. y J. Herder: *SpiderFeedback: Visual Feedback for Orientation in Virtual TV Studios*. En *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology, ACE '14*, páginas 12:1–12:8, New York, NY, USA, 2014. ACM, ISBN 978-1-4503-2945-3.

- [168] Spitzer, V., M. J. Ackerman, A. L. Scherzinger y D. Whitlock: *The visible human male: A technical report*. Journal of the American Medical Informatics Association, 3(2):118–130, April 1996.
- [169] Stype: *Stype GRIP agostomented reality and Virtual studio systems Stype Grip*. <http://www.stypegrip.com/>, 2016. [En línea; último acceso 20-abril-2016].
- [170] Suma, E. A., D. M. Krum, B. Lange, S.n Koenig, A. Rizzo y M. Bolas: *Adapting user interfaces for gestural interaction with the flexible action and articulated skeleton toolkit*. Computers & Graphics, 37(3):193 – 201, 2013, ISSN 0097-8493.
- [171] Suma, E. A., B. Lange, A. Rizzo, D. Krum y M. Bolas: *FAAST: The Flexible Action and Articulated Skeleton Toolkit*. En *IEEE Virtual Reality*, páginas 247–248, 2011.
- [172] Technologies, Ncam: *ncam - Ncam homepage*. <http://www.ncam-tech.com/>, 2016. [En línea; último acceso 10-junio-2016].
- [173] Thales Visionix, Inc.: *InterSense Precision Motion Tracking Solutions IS-900 System*. <http://www.intersense.com/pages/20/14>, 2016. [En línea; último acceso 20-febrero-2016].
- [174] Thomas, G.: *Mixed reality techniques for TV and their application for on-set and pre-visualization in film production*. En *International Workshop on Mixed Reality Technology for Filmmaking*, páginas 31–36, 2006.
- [175] Thomas, G. A., J. Jin, T. Niblett y C. Urquhart: *A versatile camera position measurement system for virtual reality TV production*. En *Broadcasting Convention, 1997. International*, páginas 284–289, Sep 1997.
- [176] Thomas, G.A., M. Koppetz y O. Grau: *New methods of image capture to support advanced post-production*. IEE Seminar Digests, 2003(10393):1–1, 2003.
- [177] Thongkamwitoon, T., S. Aramvith y T.H.Chalidabhongse: *An adaptive real-time background subtraction and moving shadows detection*. En *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volumen 2, páginas 1459 –1462 Vol.2, june 2004.
- [178] TimeDoctor: *OpenAL: Coss Platform 3D Audio*. <https://www.openal.org/>. [En línea; último acceso 23-agosto-2016].

- [179] Tracking, ART Advanced Realtime: *Flystick2 - Interaction - Products - ART Advanced Realtime Tracking*. <http://www.ar-tracking.com/products/interaction/flystick2/>. [En línea; último acceso 20-febrero-2016].
- [180] Turk, M.: *Multimodal interaction: A review*. *Pattern Recognition Letters*, 36:189 – 195, 2014, ISSN 0167-8655.
- [181] Tykkälä, A. Tommi, Andrew I. Comport, J. K. Kämäräinen y H. Hartikainen: *Live RGB-D camera tracking for television production studios*. *Journal of Visual Communication and Image Representation*, 25(1):207 – 217, 2014, ISSN 1047-3203. *Visual Understanding and Applications with RGB-D Cameras*.
- [182] V., Offis e.: *Offis. Dicom software made by offis - dcmtk - dicom toolkit*. <http://dicom.offis.de/index.php.en>, 2016. [En línea; último acceso 21-enero-2013].
- [183] Vidal, B.: *Chroma Key Visual Feedback Based on Non-Retroreflective Polarized Reflection in Retroreflective Screens*. *IEEE Transactions on Broadcasting*, 58(1):144–150, March 2012, ISSN 0018-9316.
- [184] Vierjahn, T., B. Woldecke, C. Geiger y J. Herder: *Improved direction signalization technique employing vibrotactile feedback*. En *11th Virtual Reality International Conference, VRIC'2009*, 2009.
- [185] VIZrt: *Vizrt.com*. <http://www.vizrt.com/>. [En línea; último acceso 28-julio-2016].
- [186] Wang, L., J. Zhen y Z. Shen: *Real-Time approximate soft shadow rendering with bidirectional penumbra map*. En *Educational and Information Technology (ICEIT), 2010 International Conference on*, volumen 1, páginas V1–248–V1–254, Sept 2010.
- [187] Ward, M., R. Azuma, R. Bennett, S. Gottschalk y H. Fuchs: *A Demonstrated Optical Tracker with Scalable Work Area for Head-mounted Display Systems*. En *Proceedings of the 1992 Symposium on Interactive 3D Graphics*, I3D '92, páginas 43–52, New York, NY, USA, 1992. ACM, ISBN 0-89791-467-8.
- [188] Weber, B., S. Schätzle, T. Hulin, C. Preusche y B. Deml: *Evaluation of a vibrotactile feedback device for spatial guidance*. En *World Haptics Conference (WHC), 2011 IEEE*, páginas 349–354, June 2011.

- [189] Weichert, F., D. Bachmann, B. Rudak y D. Fisseler: *Analysis of the Accuracy and Robustness of the Leap Motion Controller*. *Sensors*, 13(5):6380, 2013, ISSN 1424-8220.
- [190] Wiedemann, L. G., R. Planinc, I. Nemeč y M. Kampel: *Performance evaluation of joint angles obtained by the kinect V2*. En *Technologies for Active and Assisted Living (TechAAL), IET International Conference on*, páginas 1–6, Nov 2015.
- [191] Wöldecke, B., D. Marinos, P. Pogscheba, C. Geiger, J. Herder y T. Schwirten: *RadarTHEREMIN - Creating musical expressions in a virtual studio environment*. En *VR Innovation (ISVRI), 2011 IEEE International Symposium on*, páginas 345–346, March 2011.
- [192] Wojdala, A.: *Challenges of virtual set technology*. *IEEE MultiMedia*, 5(1):50–57, Jan 1998, ISSN 1070-986X.
- [193] Wöldecke, B., T. Vierjahn, M. Flasko, J. Herder y C. Geiger: *Steering Actors Through a Virtual Set Employing Vibro-tactile Feedback*. En *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, TEI '09*, páginas 169–174, New York, NY, USA, 2009. ACM, ISBN 978-1-60558-493-5.
- [194] Workshop, Motion: *MotionNode: Miniature Inertial Measurement Unit*. <https://msdn.microsoft.com/es-es/library/kx37x362.aspx>, 2016. [En línea; último acceso 20-febrero-2016].
- [195] Xiao, Q., G. Zhao y R. Li: *Three Dimensional Low Cost Virtual Studio System*. En *Knowledge Acquisition and Modeling, 2008. KAM '08. International Symposium on*, páginas 741–744, Dec 2008.
- [196] Xirouhakis, Y. S., A. I. Drosopoulos y A. N. Delopoulos: *Efficient optical camera tracking in virtual sets*. *IEEE Transactions on Image Processing*, 10(4):609–622, Apr 2001, ISSN 1057-7149.
- [197] Yamashita, A., T. Kaneko, S. Matsushita, y K. T. Miura: *Region Extraction with Chromakey Using Stripe Backgrounds*. *IEICE - Trans. Inf. Syst.*, E87-D(1), Enero 2004.
- [198] Yang, L., L. Zhang, H. Dong, A. Alelaiwi y A. E. Saddik: *Evaluating and Improving the Depth Accuracy of Kinect for Windows v2*. *IEEE Sensors Journal*, 15(8):4275–4285, Aug 2015, ISSN 1530-437X.

- [199] Yzquierdo, R.: *El maestro Mateo y la terminación de la catedral románica de Santiago*. Los caminos de Santiago. Arte, Historia y Literatura, páginas 253–284, 2005.
- [200] Zerroug, A., A. Cassinelli y M. Ishikawa: *MagicLensVS - Towards a flexible framework for quick setup of visual feedback in a virtual studio*. En *3D User Interfaces (3DUI), 2013 IEEE Symposium on, Orlando, FL, 2013*, páginas 183–184, 2013.
- [201] Zhang, Z.: *Microsoft Kinect Sensor and Its Effect*. IEEE MultiMedia, 19(2):4–10, Feb 2012, ISSN 1070-986X.
- [202] Zhao, G., T. Ming, C. Di, L. Quan y G. Ming: *Research on low-cost virtual studio system for education resource authoring*. En *IT in Medicine and Education, 2008. ITME 2008. IEEE International Symposium on*, páginas 198–201, Dec 2008.

Índice de figuras

1.1.	Continuo realidad-virtualidad de Milgram	3
2.1.	Imagen real y composición mediante la técnica de <i>chroma key</i>	14
2.2.	Sistema de <i>tracking</i> de cámaras Cyberset, creado por ORAD	16
2.3.	Imagen de la demostración pública de su tecnología realizada por Brainstorm Multimedia en el BIT Broadcast 2016	17
2.4.	Imagen del programa “La Ruleta de la Suerte”	18
2.5.	Esquema tradicional de un plató virtual de televisión con dos cámaras	19
2.6.	Escenario en el que una pantalla virtual se sitúa por delante del mundo real y máscara alfa codificando su posición relativa	21
2.7.	Sistema de <i>tracking</i> electromagnético en un entorno de realidad virtual CAVE	25
2.8.	Sensor mecánico Vinten-Radamec	26
2.9.	Sistema de sensores infrarrojos <i>outside-in</i>	27
2.10.	Sensor ncam	28
2.11.	Muñequera MYO	29
2.12.	Marcadores situados en el techo utilizados en una sensorización <i>inside-out</i> de la BBC	30

2.13.	Detalle de la construcción de un ciclorama con la unión de paredes y suelo en curva	33
2.14.	Uso de marcas físicas en el suelo del estudio	35
2.15.	Imagen del programa especial de una noche electoral en la BBC	39
2.16.	Imagen del programa especial de la noche electoral sobre el referéndum escocés en la BBC	40
2.17.	Imagen del programa especial de la noche electoral en RTVC	41
2.18.	Imagen del programa especial de la noche electoral en Antena 3 en la que se aprecian errores en el posicionamiento relativo de los presentadores y los objetos virtuales	42
2.19.	Imagen del programa especial de la noche electoral en La 1	43
2.20.	Imagen del programa especial de la noche electoral en TV3	44
3.1.	Diseño del plató virtual de televisión de la Facultad de Ciencias de la Comunicación de la Universidad de Santiago de Compostela	51
3.2.	Esquema de conexiones del plató	52
3.3.	Componentes del plató de trabajo	53
3.4.	Interfaz de usuario de Brainstorm eStudio	55
4.1.	Nueva arquitectura distribuida y escalable para un plató virtual de televisión con dos cámaras	59
4.2.	Esquema de la arquitectura cliente-servidor de tres capas implementada en el plató	60
4.3.	Esquema XML utilizado para la conexión de nuevos sensores con el <i>middleware</i>	64
4.4.	Interfaz gráfica implementada para la creación y edición de archivos de configuración XML para la conexión de sensores	65
4.5.	Esquema XML utilizado para la del <i>middleware</i> con Brainstorm eStudio	68
4.6.	Diferentes pasos implementados en la interfaz gráfica implementada ad hoc, para la creación y edición de archivos XML de configuración de la conexión con eStudio	69
4.7.	Interfaz gráfica del <i>middleware</i>	70
4.8.	Sistema de 8 cámaras Optitrack y Tracking Tools instalados en el plató	72

4.9.	Interfaz gráfica de Tracking Tools enviando los datos de 3 trackers a través de la red haciendo uso de VRPN	73
4.10.	Manipulación de una pantalla a través de una interfaz tangible basada de sensores Optitrack	76
4.11.	Sensor Microsoft Kinect	78
4.12.	Interfaz del software FFAST mostrando los puntos anatómicos detectados por sensor Microsoft Kinect	80
4.13.	Actor pasa de estar detrás de un objeto virtual a estar delante de forma automática	81
4.14.	Sensor Leap Motion	82
4.15.	Representación gráfica de los datos capturados por Leap Motion	83
4.16.	Interfaz gráfica implementada para el servidor VRPN de Leap Motion en el momento de la detección del gesto <i>swipe</i>	84
4.17.	Actor manipulando un objeto virtual a través de la sensorización proporcionada por el Microsoft Kinect	85
4.18.	Porcentaje de acierto de los diferentes gestos detectados por el sensor Leap Motion	91
4.19.	Configuración final del plató	92
4.20.	Sensor Microsoft Kinect V2	92
4.21.	Proceso de caracterización de gestos	94
4.22.	Interfaz gráfica implementada para el servidor VRPN asociado al sensor Microsoft Kinect V2	95
4.23.	Porcentaje de acierto del sistema, por grupo de usuarios, en la detección de los cinco gestos propuestos.	99
4.24.	Porcentaje de acierto global del sistema en la detección de cada uno de los cinco gestos propuestos.	100
4.25.	Fotograma del episodio piloto de la serie web Copycat en el que la presentadora manipula una fotografía a través de una interfaz tangible	103
4.26.	Fotograma del episodio de “La universidad responde” en el que el actor realiza un gesto que es detectado por el sensor Microsoft Kinect V2	104
5.1.	Imagen HDR del escenario (arriba) e imagen de luminancia correspondiente (abajo)	111
5.2.	Área formada por los vértices A, B, C y D	112

5.3.	Interfaz gráfica de la aplicación implementada para la ejecución del algoritmo de muestreo de corte medio	113
5.4.	Primer análisis de iluminación del plató. Imagen de luminancia (arriba) y áreas de igual energía lumínica (abajo)	114
5.5.	Segundo análisis de iluminación del plató. Imagen de luminancia (arriba) y áreas de igual energía lumínica (abajo)	115
5.6.	Series de diez imágenes tomadas con diferentes niveles de exposición para la generación de las imágenes HDR en ángulo de 90 grados	117
5.7.	Imágenes HDR originales (izquierda) y mapeadas con rotación (derecha) . .	118
5.8.	Resultado final combinando dos imágenes a 90 grados y máscara utilizada en la unión	119
5.9.	Imagen “Latitud-Longitud” obtenida en el proceso de captura de la iluminación del entorno	120
5.10.	2048 luces obtenidas del procesado de la imagen	121
5.11.	Mapeado de las 2048 luces obtenidas sobre una esfera	122
5.12.	Textura sin información adicional (izquierda) y textura con información de iluminación global (derecha) tras un proceso de <i>baking</i>	123
5.13.	Escena iluminada sin sombras (1), con sombras duras simples(2), con 64 luces (3), con 512 luces (4), con 1024 luces (5) y con 2048 luces (6)	124
5.14.	Plató virtual antes y después de aplicar la técnica <i>baking</i> de texturas con iluminación global. Vista cercana con el presentador en escena (arriba) y vista global del modelo 3D (abajo)	125
5.15.	Imagen de una silla y un presentador sin sombra (izquierda) y de la misma silla y presentador con una sombra simple (derecha)	126
6.1.	Detalle de la 1ª fase de ejecución del proyecto Pórtico Virtual	130
6.2.	Izquierda: datos sin simplificar de la cabeza del Apóstol Santiago (+300K Vértices). Derecha: simplificación al 70%.	134
6.3.	Resultados de la fase de texturizado. Izquierda Modelo 3D texturizado del Apóstol San Juan. Derecha: fotografía de la misma escultura en su emplazamiento en el Pórtico.	135
6.4.	Reconstrucción de la capa original de policromía del Pantocrátor basada en los trabajos de restauración.	136
6.5.	Detalle de la interfaz de usuario del sistema de control	138

6.6.	(1) Reconstrucción del pórtico de la gloria. (2) Detalle de uno de los músicos.	140
6.7.	Montaje final del hardware del sistema.	142
6.8.	Capturas de la aplicación en funcionamiento. Arriba: selección de instrumento. En medio: interacción con el instrumento. Abajo: selección de idioma	144
6.9.	Reconstrucción de los 18 instrumentos del pórtico y su localización en el monumento	147
6.10.	Vista global y detallada del monumento en la aplicación	149
6.11.	Mapa de estudios del monumento e imagen detallada del proceso	151
6.12.	Esquema de diseño del módulo	154
6.13.	Imagen PNG tras procesada	156
6.14.	Zonas de acción del prototipo 1.	158
6.15.	Zonas de acción del prototipo 2.	160
6.16.	Configuración final del sistema	162
6.17.	Resultados del proceso de pruebas en porcentajes de tiempo	163
6.18.	Diferentes visualizaciones presentes en la aplicación final.	164
6.19.	Instalación final en el museo	165

Apéndice

Ejemplo de archivo XML de configuración para la conexión de la capa de aplicación con la capa de datos

```

<?xml version="1.0" encoding="utf-8"?>
<Sensors_Config>
  <MultiReceiverSensor>
    <Name>TrackingTools</Name>
    <Position>
      <x>0.0</x>
      <y>0.0</y>
      <z>0.0</z>
    </Position>
    <Orientation>
      <x>0.0</x>
      <y>0.0</y>
      <z>0.0</z>
      <w>1</w>
    </Orientation>
    <Origin_Tracker></Origin_Tracker>
    <Server_IP>192.168.0.13</Server_IP>
    <Server_Port>3883</Server_Port>
    <Tracker_List>
      <Tracker>
        <Tracker_Name>Camera1</Tracker_Name>
        <VRPN_Name>Camera1</VRPN_Name>
        <id>0</id>
      </Tracker>
      <Tracker>
        <Tracker_Name>Camera2</Tracker_Name>
        <VRPN_Name>Camera2</VRPN_Name>
      </Tracker>
      <Tracker>
        <Tracker_Name>Tracker1</Tracker_Name>
        <VRPN_Name>Tracker1</VRPN_Name>
      </Tracker>
    </Tracker_List>
    <Button_List>
      </Button_List>
  </MultiReceiverSensor>
  <MonoReceiverSensor>
    <Name>LeapMotion</Name>
    <Position>
      <x>1.6</x>
      <y>2.4</y>
      <z>1.2</z>
    </Position>
    <Orientation>
      <x>0.0</x>
      <y>0.0</y>
      <z>0.0</z>
      <w>1</w>
    </Orientation>
  </MonoReceiverSensor>
</Sensors_Config>

```

```
<Server_IP>192.168.0.15</Server_IP>
<Server_Port>3883</Server_Port>
<VRPN_Tracker_Name>LeapMotionTracker</VRPN_Tracker_Name>
<VRPN_Button_Name>LeapMotionButton</VRPN_Button_Name>
<Tracker_List>
  <Tracker>
    <Tracker_Name>Hand 1</Tracker_Name>
    <id>1</id>
  </Tracker>
  <Tracker>
    <Tracker_Name>Hand 2</Tracker_Name>
    <id>2</id>
  </Tracker>
</Tracker_List>
<Button_List>
  <Button>
    <Name>Swipe</Name>
    <id>0</id>
  </Button>
  <Button>
    <Name>Circle</Name>
    <id>1</id>
  </Button>
  <Button>
    <Name>Finger Tap</Name>
    <id>2</id>
  </Button>
  <Button>
    <Name>Finger Click</Name>
    <id>3</id>
  </Button>
</Button_List>
</MonoReceiverSensor>
</Sensors_Config>
```

Ejemplo de archivo XML de configuración para la conexión de la capa de aplicación con la capa cliente

```

<?xml version="1.0" encoding="utf-8"?>
<BST_Link_Config>
  <Server_IP>192.168.0.12</Server_IP>
  <Alpha_Reference_Tracker>SpineMid</Alpha_Reference_Tracker>
  <Object_List>
    <Object>
      <Name>Screen1</Name>
      <Zpos>0.2</Zpos>
    </Object>
    <Object>
      <Name>Table1</Name>
      <Zpos>2.1</Zpos>
    </Object>
    <Object>
      <Name>Screen2</Name>
      <Zpos>1.2</Zpos>
    </Object>
  </Object_List>
  <Gesture_list>
    <Gesture>
      <Action>itemset("startVideoScreen1","ACTION.RUN")</Action>
      <ID> 1 </ID>
    </Gesture>
    <Gesture>
      <Action>itemset("nextVideoScreen1","ACTION.RUN")</Action>
      <ID> 2 </ID>
    </Gesture>
    <Gesture>
      <Action>itemset("startEffect1","ACTION.RUN")</Action>
      <ID> 3 </ID>
    </Gesture>
    <Gesture>
      <Action>itemset("startEffect2","ACTION.RUN")</Action>
      <ID> 4 </ID>
    </Gesture>
  </Gesture_list>
  <Map_List>
    <Map>
      <Name>Camera1</Name>
      <LinkPosition>Camera1Position</LinkPosition>
    </Map>
    <Map>
      <Name>Camera1</Name>
      <LinkOrientation>Camera1Orientation</LinkOrientation>
    </Map>
    <Map>
      <Name>Camera2</Name>
      <LinkPosition>Camera2Position</LinkPosition>
    </Map>
  </Map_List>
</BST_Link_Config>

```

```
<Map>
  <Name>Camera2</Name>
  <LinkOrientation>Camera2Orientation</LinkOrientation>
</Map>
<Map>
  <Name>Tracker1</Name>
  <LinkPosition>Tracker1Position</LinkPosition>
</Map>
<Map>
  <Name>Tracker1</Name>
  <LinkOrientation>Tracker1Orientation</LinkOrientation>
</Map>
<Map>
  <Name>Hand 1</Name>
  <LinkPosition>RightHandTracker</LinkPosition>
</Map>
<Map>
  <Name>Hand 2</Name>
  <LinkPosition>LeftHandTracker</LinkPosition>
</Map>
<Map>
  <Name>HandLeft</Name>
  <LinkPosition>HandLeftPosition</LinkPosition>
</Map>

</Map-List>
</BST-Link-Config>
```

Interfaz gráfica del *middleware*

- 1- Cargar archivos XML de configuración
- 2- Editar y crear archivos XML de configuración
- 3- Panel de visualización de sensores
- 4- Sensores
- 5- Trackers detectados por el sensor
- 6- Dirección IP de eStudio
- 7- Panel de visualización de información
- 8- Dirección IP del servidor VRPN
- 9- Nombre del servidor de trackers
- 10- Nombre del servidor de botones
- 11- Posición fija del sensor
- 12- Tipo de posición del sensor
- 13- Orientación del sensor
- 14- Botón de conexión con los servidores VRPN
- 15- Botón de conexión con eStudio
- 16- Panel de conexión con maps de eStudio
- 17- Nombre del receiver map en eStudio
- 18- Nombre del tracker del que tomar la información
- 19- Información del tracker a linkar (posición u orientación)
- 20- Panel de activación de efectos
- 21- ID del gesto detectado
- 22- Acción a activar en eStudio
- 23- Estado del efecto activado
- 24- Panel de control de distance keying
- 25- Tracker utilizado como referencia y valor de profundidad
- 26- Nombre del objeto virtual
- 27- Valor de profundidad del objeto en el escenario virtual
- 28- Posición relativa del presentador y el objeto

Interfaz gráfica implementada para la creación de archivos XML de configuración para la conexión de la capa de aplicación con la capa de datos



1- Panel de visualización de sensores

2- Sensores

3- Trackers detectados por el sensor

4- Botones detectados por el sensor

5- Botón para crear un nuevo sensor

6- Botón para añadir un nuevo tracker

7- Botón para añadir un nuevo botón al sensor

8- Botón de eliminación del elemento seleccionado

9- Botón para guardar en un archivo XML

10- Panel de edición de datos

11- Nombre del sensor

12- Dirección IP del servidor VRPN

13- Nombre del servidor de trackers

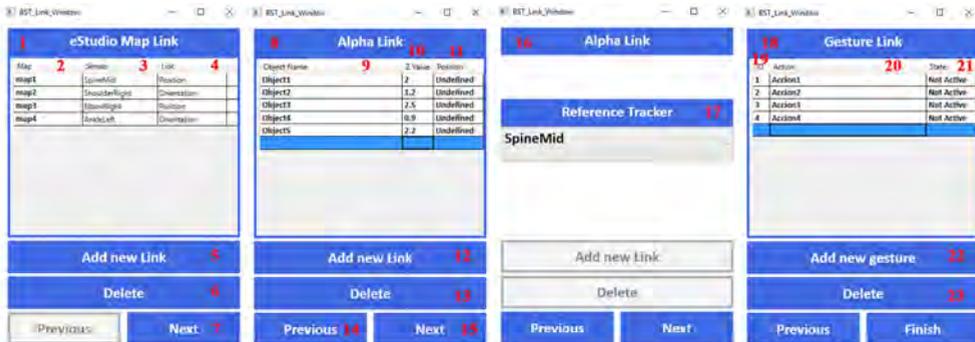
14- Nombre del servidor de buttons

15- Posición del sensor

16- Tipo de posición del sensor

17- Orientación del sensor

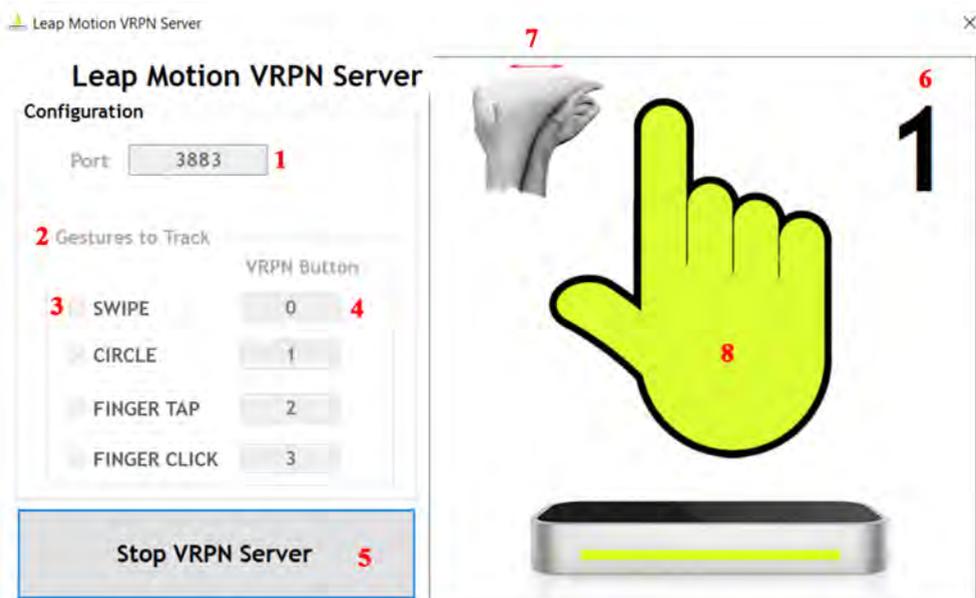
Interfaz gráfica implementada para la creación de archivos XML de configuración para la conexión de la capa de aplicación con la capa cliente



- 1- Ventana de creación de *links* a maps
- 2- Nombre del *map* en eStudio
- 3- Nombre del *tracker* asociado al *map*
- 4- Información del *tracker* a asociar
- 5- Añadir un nuevo *link*
- 6- Eliminar *link* a *map*
- 7- Siguiete ventana (Alpha link)
- 8- Ventana de posicionamiento de objetos virtuales
- 9- Nombre del objeto virtual
- 10- Posición de profundidad del objeto
- 11- Posición relativa respecto al presentador
- 12- Añadir *link* a objeto virtual

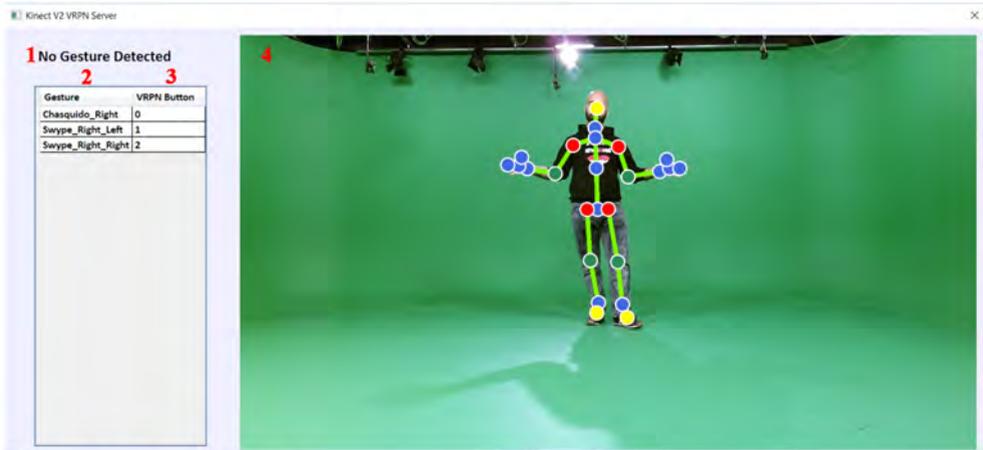
- 13- Eliminar *link* a objeto virtual
- 14- Ventana previa (eStudio Map Link)
- 15- Siguiete ventana (Alpha Link)
- 16- Ventana de selección de referencia alfa
- 17- *Tracker* referencia
- 18- Ventana de creación de *links* a efectos
- 19- ID del gesto asociado
- 20- Efecto a activar en eStudio
- 21- Estado del efecto
- 22- Añadir nuevo *link* a efecto
- 23- Eliminar *link* a efecto

Interfaz gráfica del servidor VRPN implementado para el sensor Leap Motion



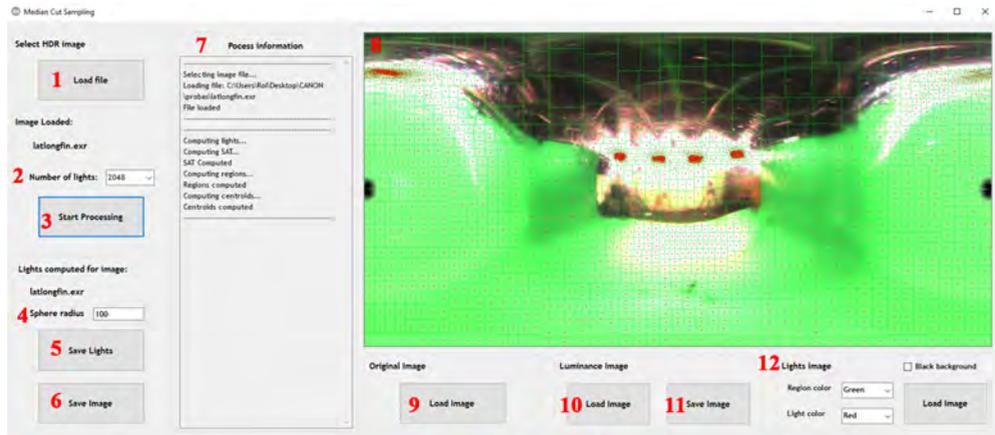
- 1- Puerto de conexión con el servidor VRPN**
- 2- Panel de gestión de gestos**
- 3- Gestos disponibles y *checkbox* para su activación**
- 4- ID del VRPN_Button asociado a cada gesto**
- 5- Botón de inicio y parada del servidor VRPN**
- 6- Indicador del número de manos detectadas**
- 7- Gráfico de indicación de gesto detectado**
- 8- Gráfico de indicación de mano detectada**

Interfaz gráfica del servidor VRPN implementado para el sensor Microsoft Kinect V2



- 1- Indicador de gesto detectado**
- 2- Lista de gestos aprendidos por el sistema**
- 3- Lista de valores de los VRPN_Buttons asociados a cada gesto**
- 4- Imagen de color capturada por el sensor Microsoft Kinect V2 con los esqueletos detectados superpuestos**

Interfaz gráfica implementada para el algoritmo de muestreo de corte medio



- 1- Cargar imagen HDR**
- 2- Número de regiones**
- 3- Iniciar procesado**
- 4- Radio de la esfera de mapeo**
- 5- Exportar luces en formato txt**
- 6- Guardar imagen de resultado**

- 7- Panel de información**
- 8- Área de visualización de resultados**
- 9- Cargar imagen original**
- 10- Cargar imagen de luminancia**
- 11- Guardar imagen de luminancia**
- 12- Configurar vista de las regiones y luces puntuales**