



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Politécnica Superior de Gandia

Reconstrucción 3D de espacios interiores mediante  
técnicas de Visión Artificial y el uso de una cámara RGB-D

Trabajo Fin de Grado

Grado en Tecnologías Interactivas

AUTOR/A: Cebolla Pardo, Álex

Tutor/a: Pérez López, David Clemente

Cotutor/a externo: ORTEGA PEREZ, MARIO

CURSO ACADÉMICO: 2021/2022

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCOLA POLITÈCNICA SUPERIOR DE GANDIA

Grado en Tecnologías Interactivas

---



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



ESCOLA POLITÈCNICA  
SUPERIOR DE GANDIA

# “Reconstrucción 3D de espacios interiores mediante técnicas de Visión Artificial y el uso de una cámara RGB-D”

*TRABAJO FINAL DE GRADO*

Autor/a:

**Alex Cebolla Pardo**

Tutor/a:

**Pérez López, David Clemente  
Ortega Pérez, Mario**

**GANDIA, 2022**

## RESUMEN

El principal objetivo del TFG es la recreación y reconstrucción 3D de múltiples instancias de espacios interiores, mediante el uso técnicas de Visión Artificial y usando como principal dispositivo de captura un sensor RGB-D. Para conseguir dicho objetivo se utilizará la implementación oficial del algoritmo BundleFusion, la cual permite obtener, de forma detallada y en tiempo real, una geometría con color del espacio físico de interés.

En este trabajo se explorarán las limitaciones de dicho método (BundleFusion), la configuración óptima de los parámetros de éste para diferentes casos de uso, la usabilidad de la herramienta (conjunto software/hardware) y el testeo de ésta frente a diferentes condiciones del entorno (p. ej. condiciones lumínicas del entorno), así como los posibles errores críticos del software. Finalmente, todas las conclusiones extraídas del estudio anterior se aplicarán en la creación de un novedoso dataset formado por múltiples instancias de espacios interiores (p. ej. una sala de reuniones o un cuarto de baño). De este modo, el usuario podrá fácilmente recrear en un sistema de Realidad Virtual el espacio físico, así como obtener otro tipo de información relevante (p. ej. plano 2D y medidas reales entre otros).

**Palabras clave:** Reconstrucción 3D; Visión Artificial; Realidad Virtual; Realidad Aumentada; RGB-D; Bundlefusion

## RESUM

El principal objectiu del present TFG és la recreació i reconstrucció 3D de diverses instàncies d'espais interiors, mitjançant l'ús de tècniques de visió artificial i utilitzant com a principal dispositiu de captura un sensor RGB-D. Per a aconseguir aquest objectiu s'utilitzarà la implementació oficial de l'algoritme BundleFusion, el qual ens permet obtenir de forma detallada i en temps real, una geometria amb un color de l'espai físic d'interés.

En aquest treball s'exploraran les distintes limitacions d'aquest mètode (BundleFusion), la configuració òptima dels diferents paràmetres d'aquest per a diferents situacions d'ús, la forma d'utilitzar la ferrament (conjunt software/hardware) i el testege d'aquest front a distintes condicions en l'entorn (p. ex. Condicions lumíniques de l'entorn), així com als possibles errors crítics del software. Finalment, totes les conclusions estretes de l'estudi realitzat s'inclouran a la creació d'un dataset compost per diverses instàncies d'espais interiors (p. ex. Un bany, una sala de reunions). D'aquesta manera, l'usuari podrà fàcilment recrear en un sistema de Realitat Virtual l'espai físic, així com obtenir altre tipus d'informació rellevant (p. ex. Una imatge en 2D i mesures reals entre altres.)

**Paraules clau:** Reconstrucció 3D; Visió Artificial; Realitat Augmentada; Realitat Virtual; RGB-D; Bundlefusion

## ABSTRACT

The main objective of this FDP is the recreation and 3D reconstruction of multiple instances of interior spaces, through the use of Computer Vision techniques and using as main capture device an RGB-D sensor. In order to achieve this goal, the official implementation of the BundleFusion algorithm will be used, which allows to obtain, in detail and in real time, a colored geometry of the physical space of interest.

This work will explore the limitations of this method (Bundlefusion), the optimal parameter configurations for different use cases, the usability of the tool (software/hardware set) and the testing of the tool against different environmental conditions (e.g., ambient lighting conditions), as well as the possible critical software errors. Finally, all the conclusions drawn from the previous study will be applied in the creation of a novel dataset consisting of multiple instances of indoor spaces (e.g., a meeting room or a bathroom). In this way, the user will be able to easily recreate in a Virtual Reality system the physical space, as well as to obtain other relevant information (e.g., 2D floor plan and real physical dimensions among others).

**Key words:** 3D reconstruction; Computer Vision; Augmented Reality; Virtual Reality; RGB-D; Bundlefusion

## TABLA DE CONTENIDOS

<b>1.- INTRODUCCIÓN.....</b>	<b>9</b>
<b>1.1.- OBJETIVOS.....</b>	<b>10</b>
<b>1.2.- METODOLOGÍA.....</b>	<b>10</b>
<b>1.3.- FASES .....</b>	<b>11</b>
<b>1.4.- ESTRUCTURA DEL DOCUMENTO.....</b>	<b>11</b>
<b>2.- MARCO TEÓRICO.....</b>	<b>12</b>
<b>2.1.- INTRODUCCIÓN A LA RECONSTRUCCIÓN 3D .....</b>	<b>12</b>
2.1.1.- DEFINICIÓN.....	12
2.1.2.- APLICACIONES/CASOS DE USO DE ESTA TECNOLOGÍA .....	12
2.1.2.1.- REALIDAD VIRTUAL .....	12
2.1.2.2.- REALIDAD AUMENTADA.....	14
2.1.2.3.- VISIÓN ARTIFICIAL .....	15
<b>2.2.- TÉCNICAS DE RECONSTRUCCIÓN 3D.....</b>	<b>15</b>
2.2.1.- FOTOGRAMETRÍA.....	16
2.2.2.- LUZ ESTRUCTURADA .....	17
2.2.3.- SLAM.....	18
<b>2.3.- HARDWARE UTILIZADO COMUNMENTE PARA LA RECONSTRUCCIÓN 3D.....</b>	<b>19</b>
2.3.1.- CÁMARA RGB .....	19
2.3.2.- CÁMARA RGB-D.....	20
2.3.2.1.- KINECT .....	21
2.3.2.2.- ZED .....	22
2.3.2.3.- OTROS.....	23
<b>2.4.- SOFTWARE.....</b>	<b>24</b>
2.4.1.- BUNDLEFUSION .....	24
<b>2.5.- DATASETS DE RECONSTRUCCIONES 3D DE ESPACIOS INTERIORES.....</b>	<b>26</b>
2.5.1.- SCAN-NET .....	26
Se observa cómo se han utilizado 21 clases distintas para la clasificación de los diferentes objetos que incluye el entorno. ....	27
2.5.2.- SUN RGB-D .....	27
2.5.3.- HABITAT .....	28
<b>3.- PROPUESTA DESARROLLO Y CREACIÓN.....</b>	<b>29</b>
<b>3.1.- HARDWARE UTILIZADO .....</b>	<b>29</b>

3.1.1.- CÁMARAS RGB-D.....	29
3.1.1.1.- CÁMARA INTEL REALSENSE D435 .....	30
3.1.1.2.- CÁMARA INTEL REALSENSE D415 .....	30
3.1.2.- MOCHILA HP BACKPACK G2 .....	31
3.1.3.- TABLET .....	32
3.1.4.- MONITOR USB-C .....	34
<b>3.2.- POR QUÉ SE HA UTILIZADO ESTE HARDWARE .....</b>	<b>35</b>
3.2.1.- CAPACIDAD DE PROCESAMIENTO.....	35
3.2.2.- AUTONOMÍA .....	35
3.2.3.- USABILIDAD .....	35
<b>3.3.- SOFTWARE UTILIZADO.....</b>	<b>36</b>
3.3.1.- DESCRIPCIÓN DEL ENTORNO DE DESARROLLO.....	37
3.3.1.1.- C++.....	37
3.3.1.2.- MICROSOFT VISUAL STUDIO.....	37
3.3.1.3.- WINDOWS .....	37
3.3.1.4.- SDK DE LA CÁMARA INTEL .....	37
3.3.1.5.- LIBRERÍAS UTILIZADAS.....	38
3.3.2.- MODIFICACIONES REALIZADAS SOBRE LA IMPLEMENTACIÓN OFICIAL DEL MÉTODO.....	38
<b>3.4.- DESCRIPCIÓN DEL FLUJO DE TRABAJO PARA LA CREACIÓN DEL DATASET DE ESPACIOS INTERIORES .....</b>	<b>38</b>
<b>4.- RESULTADOS.....</b>	<b>42</b>
4.1.- EJEMPLOS .....	42
4.2.- CONCLUSIONES DE USABILIDAD.....	45
4.3.- CONCLUSIONES DE LIMITACIONES TÉCNICAS.....	46
4.4.- PARÁMETROS ÓPTIMOS .....	47
4.5.- CONCLUSIÓN DEL TIEMPO EMPLEADO PARA UNA RECONSTRUCCIÓN 3D.....	48
<b>5.- CONCLUSIONES.....</b>	<b>48</b>
5.1.- CUMPLIMIENTO DE OBJETIVOS .....	48
5.1.- APORTACIONES .....	49
5.3.- TRABAJOS FUTUROS .....	50
<b>6.- BIBLIOGRAFÍA.....</b>	<b>50</b>

## ÍNDICE DE ILUSTACIONES

Ilustración 1. Preparación de operaciones quirúrgicas complejas.....	14
Ilustración 2. Ejemplo de uso de la Realidad Aumentada.....	15
Ilustración 3. Ejemplo de uso de la técnica de fotogrametría.....	17
Ilustración 4. Ejemplo de uso de la técnica de luz estructurada.....	18
Ilustración 5. Ejemplo de uso de la técnica SLAM.....	19
Ilustración 6. Imagen RGB-D correspondiente al mapa de profundidad.....	20
Ilustración 7. Mapa de profundidad de una escena.....	20
Ilustración 8. Modelos de cámaras Kinect.....	22
Ilustración 9. Zed 2 Cámara 3D Estéreo.....	23
Ilustración 10. Ejemplo de reconstrucción con el dispositivo iPhone 13 Pro.....	24
Ilustración 11. Clasificación de los diferentes objetos de un espacio en el dataset SCANNET.....	27
Ilustración 12. Clasificación de los diferentes objetos de un espacio en el dataset SUN- RGB-D.....	28
Ilustración 13. Ejemplo de reconstrucción 3D en el dataset Habitat-Matterport.....	29
Ilustración 14. Cámara Intel RealSense D435.....	30
Ilustración 15. Cámara Intel RealSense D415.....	31
Ilustración 16. HP Z VR Backpack G2 Workstation.....	32
Ilustración 17. Tablet Samsung Galaxy Tab A (2016).....	33
Ilustración 18. Monitor USB-C VSEVEN L156TCH-1G.....	34
Ilustración 19. Reconstrucción 3D de una sala de estar.....	42
Ilustración 20. Reconstrucción 3D de una sala de reuniones.....	42
Ilustración 21. Reconstrucción 3D de una habitación individual.....	42
Ilustración 22. Reconstrucción 3D de una sala de estar.....	42
Ilustración 23. Reconstrucción 3D de una tienda de deportes.....	43
Ilustración 24. Reconstrucción 3D de una pizzeria.....	43
Ilustración 25. Reconstrucción 3D de un almacén.....	43
Ilustración 26. Reconstrucción 3D de una sala de reuniones.....	43
Ilustración 27. Reconstrucción 3D de un taller mecánico.....	43
Ilustración 28. Reconstrucción 3D de un comedor de la universidad.....	43
Ilustración 29. Reconstrucción 3D de un baño individual.....	44
Ilustración 30. Reconstrucción 3D de un baño doble.....	44



Ilustración 31. Reconstrucción 3D de un baño doble.....	44
Ilustración 32. Reconstrucción 3D de una sala de reuniones .....	44
Ilustración 33. Reconstrucción 3D de una habitación de trabajo .....	44
Ilustración 34. Reconstrucción 3D de una sala de pruebas .....	44
Ilustración 35. Reconstrucción 3D de una sala de servidores .....	45
Ilustración 36. Reconstrucción 3D de un despacho.....	45

## ÍNDICE DE TABLAS

Tabla 1. Especificaciones de la Kinect v1 y la Kinect v2 .....	21
Tabla 2. Especificaciones de la cámara Zed 2 Cámara 3D Estéreo .....	23
Tabla 3. Especificaciones de la cámara RGB-D Intel RealSense D435.....	30
Tabla 4. Especificaciones de la cámara RBG-D Intel RealSense D415.....	31
Tabla 5. Especificaciones de la HP Z VR BackPack G2 Workstation.....	32
Tabla 6. Especificaciones de la tablet Samsung Galaxy Tab A .....	33
Tabla 7. Especificaciones del monitor VSEVEN L156TCH-1G .....	34

## 1.- INTRODUCCIÓN

En este trabajo se aborda la tarea de crear reconstrucciones 3D de estancias interiores, mediante sensores RGB-D y técnicas de Visión Artificial.

La reconstrucción 3D es el proceso por el cual se puede reproducir un objeto/escena real en un ordenador, manteniendo en éste sus dimensiones, forma y volumen. Se trata de un proceso de vital importancia en múltiples sectores/campos, como la Realidad Virtual, Realidad Aumentada o la robótica, por el que se obtienen recreaciones de espacios y objetos.

Si bien esta tarea puede ser abordada desde diferentes enfoques, en este trabajo se utiliza el método BundleFusion, el cual realiza el proceso íntegramente en GPU, permitiendo crear reconstrucciones 3D en tiempo real, con alto nivel de detalle y con información de color. Éste es un método que solventa una de las limitaciones más importantes de las técnicas clásicas de fusión volumétrica TSDF (truncated signed distance field), la falta de escalabilidad, la cual restringe la reconstrucción de espacios a tamaños muy reducidos. Además, este método también solventa el fenómeno conocido como drift, el cual degrada la calidad de la reconstrucción, debido a la acumulación de error, a lo largo del tiempo, en el sistema de seguimiento usado para estimar la posición y orientación relativa de la cámara respecto al entorno. Por lo tanto, este es el método óptimo para la creación del dataset objeto de este trabajo.

Existen diferentes bases de datos de reconstrucciones 3D de espacios interiores en la literatura, como pueden ser SCAN-NET, SUN RGB-D y HABITAT. Sin embargo, todas tienen en común unas limitaciones, consistentes en presentar muchas reconstrucciones incompletas, la falta de fidelidad de algunas de ellas y sobre todo, la no disponibilidad de: las imágenes RGB-D utilizadas por la cámara para la realización de la reconstrucción, las poses relativas de la cámara en cada frame, los parámetros intrínsecos y extrínsecos del sensor, así como los datos la aceleración y la velocidad angular en el caso de que el sensor RGB-D utilizado tuviese esta capacidad. Esta información es muy importante a la hora de realizar nuevas implementaciones sobre el método. Por dicho motivo, el presente trabajo propone la creación de un nuevo dataset que solvante estas limitaciones.

Por último, ninguno de los métodos estudiados ha propuesto un conjunto hardware óptimo para la creación de reconstrucciones 3D. En consecuencia, este trabajo ha propuesto un estudio de usabilidad con la finalidad de poder definirlo.

### 1.1.- OBJETIVOS

El objetivo principal de este trabajo es la recreación y reconstrucción 3D de múltiples instancias de espacios interiores, haciendo uso técnicas de Visión Artificial y usando un sensor RGB-D como dispositivo de captura. Dichas reconstrucciones se clasificarán y almacenarán como un dataset novedoso de espacios interiores. Para conseguir dicho objetivo se definen una serie de objetivos secundarios que se describen a continuación:

- Modificar la implementación oficial de BundleFusion para soportar otros modelos de cámaras RGB-D
- Explorar las limitaciones técnicas de BundleFusion
- Estudiar usabilidad de la herramienta
- Obtener una configuración óptima de parámetros para que el método se pueda usar en cualquier entorno

### 1.2.- METODOLOGÍA

Antes de comenzar el desarrollo del proyecto se definió el objetivo principal, así como los objetivos secundarios del trabajo, para después diseñar, a grandes rasgos, las diferentes etapas de desarrollo que habría que seguir en orden estricto para obtener un resultado satisfactorio. Dichas etapas se listan en el apartado 1.3.

Por lo que respecta a la parte de desarrollo, se empleó una cámara RGB-D, un ordenador, un dispositivo de visualización y control (monitor táctil) y una colección de herramientas software, donde el elemento principal es la implementación oficial del método BundleFusion, para la creación de reconstrucciones 3D. Además, dicha implementación se modificó utilizando C++ en Microsoft Visual Studio, así como la SDK de la cámara Intel RealSense, la librería MLib y OpenCV.

### 1.3.- FASES

Para la realización de este trabajo se siguieron un total de 6 etapas que se listan a continuación:

- Análisis del marco teórico
- Modificación implementación oficial BundleFusion
- Exploración de las limitaciones técnicas del método
- Definición de parámetros óptimos para cualquier entorno
- Estudio de usabilidad de la herramienta
- Creación del dataset de espacios interiores obtenido mediante el método de reconstrucción 3D.

### 1.4.- ESTRUCTURA DEL DOCUMENTO

El presente documento se ha dividido en 6 capítulos, Introducción, Marco teórico, Propuesta de desarrollo y creación, Resultados, Conclusiones y Bibliografía. El primero de ellos presenta la temática que aborda este trabajo, los objetivos propuestos, la metodología utilizada, las fases del desarrollo y por último la estructura del documento.

En el marco teórico, se introducirán diferentes conceptos para la comprensión total del trabajo.

En la propuesta, desarrollo y creación, se describe el material utilizado para los escaneos, tanto software como hardware. También se describe el algoritmo a utilizar BundleFusion, así como las modificaciones implementadas. Además, este apartado incluye la descripción del flujo de trabajo utilizado para obtener el máximo rendimiento del sistema en cada reconstrucción 3D.

En el capítulo Resultados se incluirán algunas capturas de las reconstrucciones 3D realizadas, junto con las conclusiones de diferentes aspectos referentes a la realización de los escaneos.

En el capítulo Conclusiones se describirá el cumplimiento de objetivos, junto con las diferentes aportaciones realizadas y posibles trabajos futuros.

Por último, el capítulo Bibliografía, presenta los recursos bibliográficos consultados para la elaboración de esta memoria.

## 2.- MARCO TEÓRICO

### 2.1.- INTRODUCCIÓN A LA RECONSTRUCCIÓN 3D

#### 2.1.1.- DEFINICIÓN

La reconstrucción 3D es el proceso por el cual se puede reproducir un objeto/escena real en un ordenador, manteniendo en éste sus dimensiones, forma y volumen. La información necesaria para realizar una reconstrucción se puede obtener de dispositivos, como, por ejemplo, láseres, cámaras, etc. Este proceso es de vital importancia en múltiples sectores/campos como son la Realidad Virtual o la Realidad Aumentada.

#### 2.1.2.- APLICACIONES/CASOS DE USO DE ESTA TECNOLOGÍA

A continuación, se presentan los casos de uso más relevantes de la reconstrucción 3D.

##### 2.1.2.1.- REALIDAD VIRTUAL

La realidad virtual (VR) es una tecnología que engloba un amplio conjunto de ideas, bajo el cual muchos investigadores y compañías han centrado sus trabajos. Se trata de un término que no tiene una definición única, sin embargo, tradicionalmente se ha definido como “un entorno tridimensional, interactivo y generado por ordenador en el cual se sumerge una persona” (Aukstakalnis, 1992) (Otegui, 2017).

Los puntos clave de esta definición son tres. El primero de ellos es que el entorno virtual es una escena tridimensional generada por ordenador, la cual requiere un elevado rendimiento gráfico para conseguir un adecuado nivel de realismo. El segundo es que el mundo virtual es interactivo. Un usuario requiere respuestas en tiempo real por parte del sistema para que éste tenga la sensación de interactividad. El último es que el usuario se encuentre inmerso en este mundo virtual. Para que esta inmersión parezca realista, el sistema de VR debe localizar exactamente todos los movimientos del usuario y determinar qué efectos se producirán sobre la escena que se muestra en el dispositivo de visualización empleado.

En este contexto, aparecen dos conceptos clave para caracterizar al sistema de VR, la presencia y la inmersión. Así la presencia es una característica subjetiva del sistema que identifica la sensación de estar en un lugar o entorno, incluso cuando uno se encuentra físicamente en otro (Heater 1992). Mientras que la inmersión se refiere a una característica cuantificable del sistema que mide su capacidad para mostrar un entorno generado por ordenador de forma que se aproxime a la experiencia real. La mayor o menor inmersión mide cómo el sistema de VR admite contingencias sensoriomotoras naturales para la percepción (Slater 2009).

Dicho lo cual, el nivel de inmersión en el entorno de VR dependerá de la tecnología empleada para su experimentación, esto es, dependerá de los dispositivos de entrada o sensores, los cuales capturan las acciones del participante, y de los dispositivos de salida o efectores, que generan los estímulos necesarios para los sentidos del participante. Actualmente, se está popularizando como “factor de forma” de la VR el Head Mounted Display (HMD) de 6 grados de libertad, debido principalmente al abaratamiento e integración de la tecnología (Soliman et. al. 2021). De hecho, el casco de VR ha relegado a un papel testimonial a otras tecnologías altamente inmersivas, como el CAVE (Cruz-Neira 1992), que era una habitación con unas dimensiones alrededor de 3x3x3 metros, habitualmente con tres paredes (frontal y laterales) y un suelo donde se proyectaban imágenes estereoscópicas, complementado por sonido 3d y seguimiento del usuario de 6 grados de libertad.

Como se ha comentado anteriormente, en este trabajo se van a crear mundos virtuales propios, que representarán la realidad y en los que se podrá interactuar. Concretamente, el TFG se centrará en la reconstrucción 3D, ya que, con esta técnica, cualquier espacio físico se va poder recrear de forma rápida y sencilla en un entorno de Realidad Virtual.

Algunos de los ejemplos de uso de Realidad Virtual pueden ser, videojuegos, educación, industria, marketing, patrimonio, salud (psicología, rehabilitación de distintas enfermedades, detección prematura de diferentes enfermedades mentales, preparación de operaciones quirúrgicas complejas), entre otros.



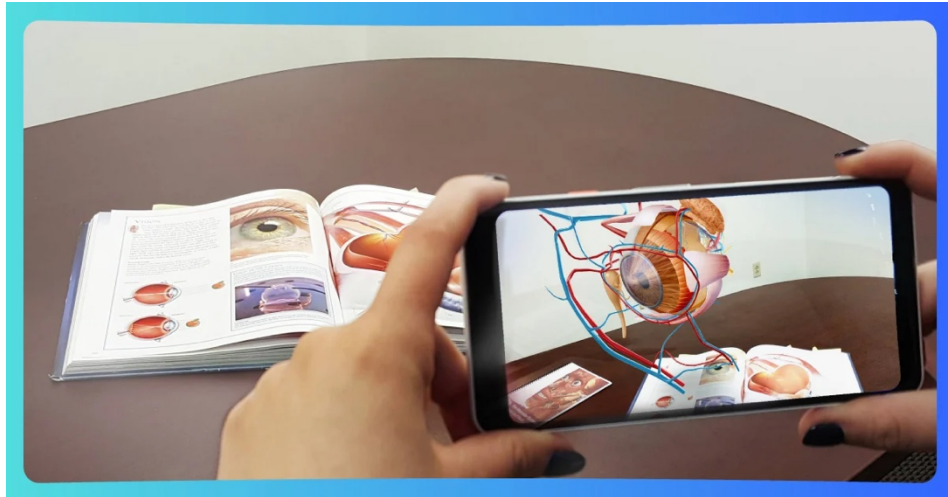
*Ilustración 1. Preparación de operaciones quirúrgicas complejas*

### *2.1.2.2.- REALIDAD AUMENTADA*

La definición más habitual de Realidad Aumentada (en inglés Augmented Reality o AR) es la de Ronald Azuma (1997), la cual se basa en 3 elementos. El primer elemento es la combinación de lo real y lo virtual, seguidamente establece que funciona en tiempo real, y, por último, que la escena está registrada en tres dimensiones. Ese registro en 3D es clave y se consigue cuando los objetos virtuales generados por ordenador están alineados correctamente con el mundo real en todas las dimensiones. Si existen errores en el ajuste, el usuario no tendrá la percepción de ver ambas imágenes, virtual y real, fusionadas. Además, el ajuste de las imágenes debe ser lo más exacto posible en todo momento, incluso cuando el usuario se esté moviendo, dado que éstos son los errores a los que el usuario es más sensible (Azuma 1997). Con esta definición, se pueden considerar sistemas AR aquéllos que utilicen como dispositivo de visualización HMDs u otro tipo de tecnologías tales como dispositivos basados en monitor, dispositivos móviles, etc., siempre y cuando éstos mantengan las características esenciales de la AR. Por ejemplo, eso no incluye películas o superposiciones 2-D en la imagen. Películas que poseen objetos virtuales fotorrealistas que han sido perfectamente fusionados con el entorno 3-D no pueden ser consideradas como AR porque no existe una interacción con ellos en tiempo real. De igual modo que no lo son, los objetos virtuales superpuestos a vídeos, ya que no están combinados con el mundo real en 3D.

Por lo tanto, en AR lo que se pretende es complementar la realidad en lugar de reemplazarla, consiguiendo de este modo un mayor realismo en todas nuestras acciones.

Con la técnica de reconstrucción 3D se generarán objetos virtuales, los cuales van a poder ser insertados en un entorno de AR, superponiéndolos al mundo real (Innovae, 2022). El objetivo principal perseguido en AR es crear un sistema en el que no se distinga el mundo real de los objetos virtuales incluidos en él, logrando así una fusión total de ambos entornos.



*Ilustración 2. Ejemplo de uso de la Realidad Aumentada*

### *2.1.2.3.- VISIÓN ARTIFICIAL*

La visión artificial es una disciplina científica que permite dotar a un ordenador de la toma de decisiones. Es decir, permite a un ordenador comprender y percibir una o varias imágenes para así, actuar de una manera determinada dependiendo de la situación. Esta técnica se utiliza en diferentes campos, como, por ejemplo, en reconocimiento de objetos y reconstrucción de una escena, entre otros.

Actualmente la visión artificial está teniendo un peso muy importante en la industria 4.0, ya que el objetivo de ésta es automatizar y digitalizar todo tipo de procesos, ya sean, por ejemplo, controles de calidad, guiado de robots, control dimensional, etc.

## **2.2.- TÉCNICAS DE RECONSTRUCCIÓN 3D**

En este apartado se van a describir las principales técnicas que se pueden emplear en la reconstrucción de espacios 3D.



### 2.2.1.- FOTOGRAMETRÍA

La fotogrametría es una técnica que tiene como objetivo estudiar cualquier objeto, y con ello definir la forma, junto con la dimensión y posición del objeto. Con la realización de diferentes fotografías o vídeos desde diferentes perspectivas, y escalas este tipo de método es capaz de obtener los puntos característicos de estas imágenes mediante técnicas de visión artificial (p. ej. El método SIFT o el método SURF (Oyallon, E., Rabin, J., 2015)), para así, determinar la posición relativa (matriz de transformación homogénea  $4 \times 4$ ) de la cámara entre cada imagen capturada. Destacar que el método SIFT (Scale Invariant Feature Transform) permite detectar diferentes puntos característicos, así como los descriptores asociados a estos en las imágenes a partir del histograma orientado de gradientes (Lindeberg, T. 2012). De este modo, el método SIFT es robusto a cambios de escala, perspectiva e iluminación en la imagen.

A partir de la posición relativa de la cámara de cada una de las imágenes se obtiene, mediante técnicas de visión artificial, la nube de puntos 3D del objeto/escena. Finalmente se obtiene la malla 3D en color del objeto a partir de la nube de puntos de 3D.

Como se ha indicado, la fotogrametría es una técnica asequible para conseguir una reconstrucción, tan solo basta con tener un dispositivo para realizar las fotografías, como puede ser un móvil. Pero, por el contrario, esta técnica es computacionalmente costosa, se precisa de un ordenador que tiene que ser capaz de procesar cantidades elevadas de fotografías y poder generar un modelo digitalizado en 3D (Bitfab, 2022). Por otro lado, es importante destacar que las imágenes empleadas deben estar lo suficientemente texturizadas ya que en caso contrario no se podrán obtener suficientes puntos característicos mediante técnicas de visión artificial y, en consecuencia, no se podrá determinar la transformación relativa de esta imagen con el resto de imágenes.



*Ilustración 3. Ejemplo de uso de la técnica de fotogrametría*

### 2.2.2.- LUZ ESTRUCTURADA

La técnica de luz estructurada permite obtener la forma y características de un objeto físico, a partir de la proyección de patrones de luz sobre el mismo y mediante 2 cámaras que capturan la deformación que han sufrido los patrones al proyectarse sobre el objeto. Mediante técnicas de visión artificial se analiza dicha deformación para poder establecer la reconstrucción 3D (Manteca Fernández, F. 2018).

A diferencia de la fotogrametría, la luz estructurada es una técnica que no requiere de conocimientos previos para utilizarla, además, es una técnica muy rápida. Sin embargo, la luz estructurada sí requiere de una inversión inicial elevada ya que esta utiliza equipos caros, los cuales proporcionan mayor calidad en la reconstrucción que la técnica de fotogrametría. Además, cabe añadir que la técnica de luz estructurada está pensada para objetos que tienen un tamaño pequeño, y no para espacios físicos reales, como por ejemplo una habitación, los cuales son objeto de este trabajo.



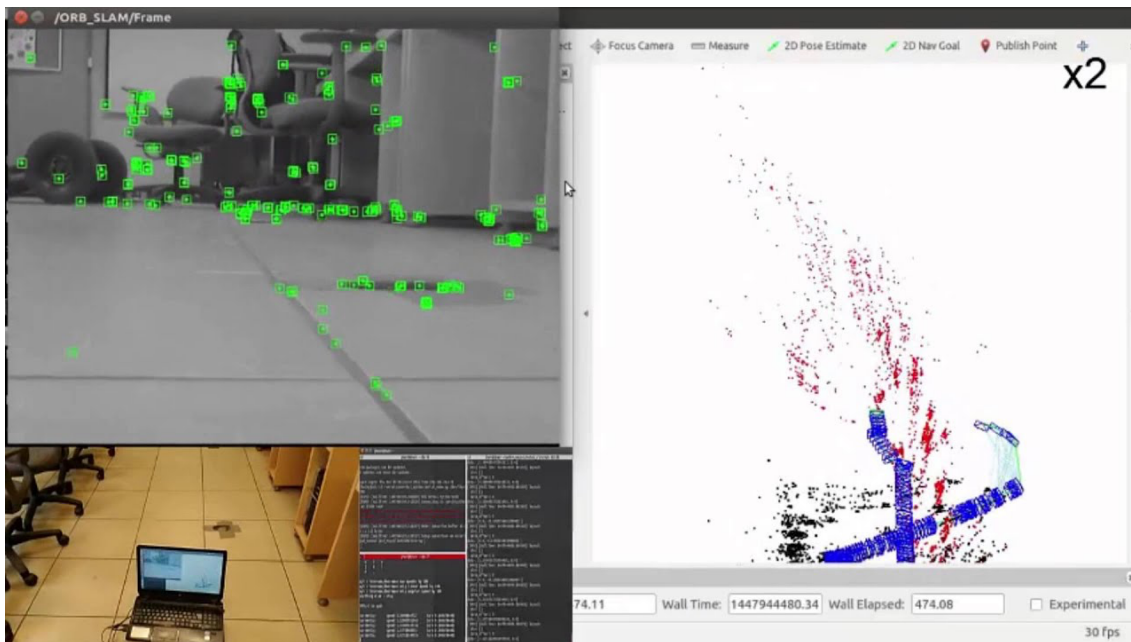
*Ilustración 4. Ejemplo de uso de la técnica de luz estructurada*

### 2.2.3.- SLAM

SLAM (Simultaneous Localization And Mapping) significa realizar un mapeo y obtener la posición tridimensional de manera simultánea. Esta técnica es comúnmente utilizada en el área de robótica. En otras palabras, SLAM tiene como principal objetivo poder obtener la información tridimensional en un área totalmente desconocida, y conseguir que ésta se vaya actualizando en el mapa del entorno al tiempo que se realiza un seguimiento en tiempo real. Esta técnica facilita la posible reconstrucción de un área extensa, bien sea interior o exterior. De hecho, se puede medir con el uso de drones, robots móviles, etc. Por lo tanto, la reconstrucción se convierte en un proceso sencillo y rápido, el cual se adaptará a nuestros recursos disponibles (Geoslam, 2020).

La técnica SLAM también permite obtener una reconstrucción 3D de un espacio físico. Sin embargo, actualmente esta técnica se está centrando en proyectos como vehículos autónomos y productos de uso doméstico como los robots de limpieza.

Este método también se utiliza en una de las asignaturas del grado, donde un robot TurtleBot reconocía un área desconocida y, además, localizaba y ubicaba los diferentes objetos añadidos a la escena, como se observa en la figura.



*Ilustración 5. Ejemplo de uso de la técnica SLAM*

## 2.3.- HARDWARE UTILIZADO COMUNMENTE PARA LA RECONSTRUCCIÓN 3D

En este apartado se presentan algunos de los sensores más comúnmente utilizados en la creación de reconstrucciones 3D.

### 2.3.1.- CÁMARA RGB

Las cámaras RGB (red, green, blue) capturan la capacidad de luz dentro del espectro visible, para ello disponen de un sensor con un filtro en forma de mosaico que incluye los colores primarios. Algunos dispositivos típicos de entrada RGB son las cámaras de video, cámaras de fotos, webcams, etc. Este tipo de cámara, a diferencia de cámaras RGB-D u otra cámara multiespectral, ofrecen menos información del entorno por lo que resultan mucho más económicas.

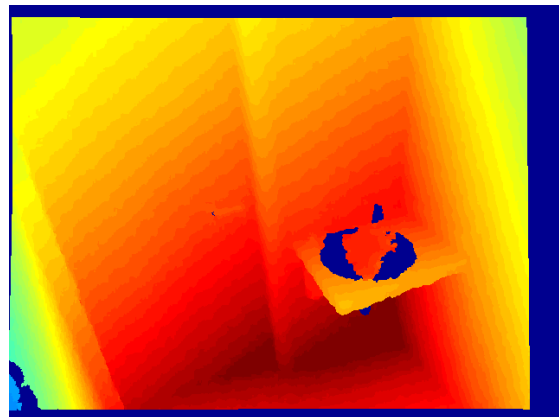
Estas cámaras comúnmente se están utilizando en la agricultura, mediante drones a los que se añade una cámara RGB, la cual es capaz de obtener imágenes de grandes extensiones. El inconveniente de estos dispositivos, es que al final captan lo que el ojo humano también es capaz de percibir, por lo que para un estudio donde se desee incidir más en el terreno y obtener más información, lo ideal sería utilizar una cámara multiespectral, la cual incorpora un sensor que es capaz de captar la radiación fuera del espectro visible (Insights, 2018).

### 2.3.2.- CÁMARA RGB-D

Los sensores o cámaras RGB-D (red, green, blue and depth) son dispositivos que proporcionan píxel a píxel tanto información de color como de profundidad de una escena. Por profundidad se entiende la distancia de un elemento con respecto a un plano horizontal de referencia situado en el origen de la cámara. La representación más habitual de la profundidad es una imagen 2D con una escala de colores, de forma que cada color representa una distancia. Además, conociendo los parámetros del sensor de profundidad, se puede obtener el valor real de la distancia para cada píxel de color. Actualmente, los sensores RGB-D combinan los sensores activos y pasivos.



*Ilustración 7. Imagen RGB-D*



*Ilustración 6. Mapa de profundidad de una escena*

Los sensores activos utilizan una fuente de energía, así, emitiendo sonido o luz pueden medir el reflejo emitido por los elementos impactados. Por otro lado, los sensores pasivos, a diferencia de los activos, utilizan la luz del medio ambiente que es reflejada en los objetos y luego recogida en la cámara. Una de las principales ventajas de utilizar este tipo de sensores pasivos es que ahorran recursos computacionales, a diferencia de las cámaras estereoscópicas. Estas últimas tienen que obtener la profundidad mediante la obtención de distintas imágenes, sin embargo, las cámaras RGB-D lo hacen obteniendo los valores de cada píxel.

### 2.3.2.1.- KINECT

La cámara Kinect fue creada por Microsoft, para la consola XBOX 360. Su lanzamiento oficial tuvo lugar en 2010, como un controlador de juego, el cual permitía interactuar con la consola mediante el movimiento corporal, ya que era capaz de mapear el esqueleto del jugador en 3D. Por lo que respecta al hardware, era un dispositivo que incorporaba una cámara infrarroja, una cámara RGB convencional, un proyector láser infrarrojo y un array de micrófonos.

Cabe destacar que en los primeros años aparecieron desarrollos de software no oficiales para PC, los cuales convertían al dispositivo en un escáner 3D. De hecho, en la actualización a Windows 10 se incorporó una aplicación, esta vez de manera oficial, llamada 3D SCAN, la cual permitía al usuario utilizar la Kinect como escáner 3D. Las Kinect más utilizadas para realizar escaneados son las Kinect que han sido llamadas como v1 y v2. La Kinect v1 utiliza la triangulación combinada con la luz estructurada, a diferencia de esta, la Kinect v2 utiliza el tiempo de vuelo. La ventaja del cambio de versión fue que el tiempo de vuelo era mucho más efectivo y, además, ya podía detectar objetos situados más alejados. También, esta nueva versión hacía que no afectasen las interferencias con otras señales.

Además, como ya se ha comentado anteriormente, este dispositivo se puede utilizar como escáner 3D, es decir, para reconstruir objetos. Para ello se precisa de una Kinect ya sea la v1 o la v2, un adaptador para poder conectarla a un ordenador, y por último un ordenador, ya sea de sobremesa o portátil, pero es necesario que éste cuente con un puerto USB.

Estas son las principales características de las Kinect, tanto la versión 1 como la versión 2 (Etapé, 2014).

	<b>Kinect 1 (Xbox 360, PC)</b>	<b>Kinect 2 (Xbox One, PC)</b>
Resolución	640x480 30 fps 4:3	1920x1080 30 fps 16:9
Distancia mínima de uso	1.82 m	1.37 m
Ángulos de visión	57° horizontal, 43° vertical	70° horizontal, 60° vertical
Latencia	102ms	20ms
Detección de personas	4	6

Tabla 1. Especificaciones de la Kinect v1 y la Kinect v2



Kinect V1



Kinect V2

Ilustración 8. Modelos de cámaras Kinect

### 2.3.2.2.- ZED

Existen diferentes modelos de cámaras Zed, a continuación, se describen sus principales características. En primer lugar, la cámara tiene una resolución de 2K, cuenta con la capacidad de detectar profundidad y también rastrear el movimiento. La Zed utiliza la tecnología de visión estéreo.

Para definir qué es la visión estéreo podríamos definir qué es la visión estereoscópica, ya que estas poseen las mismas capacidades. La visión estereoscópica es la capacidad que tienen nuestros ojos de ver dos imágenes desde dos puntos de vista ligeramente distintos, y conseguir una imagen tridimensional (Álvarez-Gutiérrez, E. L., Jiménez-López, F. R., 2019).

Por tanto, la cámara Zed, gracias a sus dos cámaras y el proceso de triangulación proporciona una comprensión tridimensional de la escena que esté escaneando.

En la siguiente tabla se pueden observar las principales características de la cámara Zed 2 Cámara 3D Estéreo (Polvillo Díaz, M, 2016).

<b>Zed 2 Cámara 3D Estéreo</b>	
Resolución de salida	2x(2208x1242) 15fps (2.2K) 2x(1920x1808) @30fps (1080p) 2x(1280x720) @60fps (720p) 2x(640x480) @100fps (VGA)
Dimensiones	175x30x33 mm

Ángulo de visión	Máximo 110°
Latencia	102ms
Peso	150g

*Tabla 2. Especificaciones de la cámara Zed 2 Cámara 3D Estéreo*



*Ilustración 9. Zed 2 Cámara 3D Estéreo*

#### *2.3.2.3.- OTROS*

Otro sensor útil en la creación de reconstrucciones 3D es el sensor LIDAR (laser detection and ranging), el cual es un láser pulsado que mide el tiempo que tarda la señal en volver a la fuente por el cual se ha emitido. Con esta tecnología se consigue una mayor precisión.

Un dispositivo móvil que incorpora un sensor LIDAR es el iPhone13. Su funcionamiento en estos dispositivos móviles ayuda a que el enfoque sea mucho más rápido, y que, en entornos con poca iluminación se consiga obtener la máxima calidad. Actualmente este es el sensor de captura más utilizado en las aplicaciones de Realidad Aumentada en dispositivos iPhone de última generación (Sastre, 2021).

Por otra parte, desde los modelos S7 de Samsung, la marca incorpora la tecnología Dual Pixel. Ésta es otra tecnología útil para la creación de reconstrucciones 3D. Se trata de un sistema de autoenfoco por detección de fase, el cual agrega un receptor de señal de enfoque a cada píxel. En esta tecnología, cada píxel está compuesto por dos fotodiodos separados generalmente menos de un milímetro, de forma que se captan dos imágenes, que pueden ser consideradas como un par estéreo. Generalmente se aplica tanto en videos como en fotografías consiguiendo un mayor autoenfoco en objetos que están en movimiento (Euronics, 2020) (Fabara, 2021).





*Ilustración 10. Ejemplo de reconstrucción con el dispositivo iPhone 13 Pro*

## 2.4.- SOFTWARE

En el siguiente apartado se presenta el software utilizado en este trabajo para la creación de reconstrucciones 3D.

### 2.4.1.- BUNDLEFUSION

La reconstrucción 3D de espacios interiores es una técnica de vital importancia en múltiples áreas de interés como son Realidad Virtual o robótica. Si bien existen múltiples enfoques para abordar esta tarea, nos centraremos en aquellos que mediante el uso de técnicas de visión artificial y el uso de una cámara RGB-D o de profundidad permiten obtener de forma precisa y en tiempo real una geometría 3d del espacio físico de interés. Generalmente estos métodos usan como técnica fundamental para realizar la fusión volumétrica el uso de TSDF (truncated signed distance field) de Curless y Levoy (Levoy, M., Pulli, K., Curless, B. et al. 2000). KinectFusion (Roth, H., & Vona, M., 2012) fue el primer trabajo en demostrar el potencial de esta técnica para realizar la reconstrucción 3D en tiempo real usando como sensor de profundidad una Kinect. Concretamente el núcleo de este sistema se basa en la representación de una escena volumétrica 3D de forma implícita mediante TSDF en la que cada nuevo frame de entrada se registra y fusiona con el volumen TSDF en tiempo real. Destacar que este proceso se realiza íntegramente en la GPU.

Una de las limitaciones más importantes de este trabajo es su falta de escalabilidad, ya que la cuadrícula uniforme de vóxeles utilizada requiere mucha memoria, limitando considerablemente el volumen de reconstrucción que se puede obtener y en consecuencia su uso se limita a realizar la reconstrucción 3D de espacios interiores de un tamaño muy reducido. Con el fin de solventar esta limitación diversos trabajos (p. ej. VoxelHashing (Nießner, M., Zollhöfer, M., Izadi, S. et al. 2013) y BundleFusion) proponen explotar la dispersión de las cuadrículas del TSDF diseñando estrategias de subdivisión espacial jerárquica con estructuras de datos eficientes.

Otra de las principales limitaciones de este tipo de métodos de reconstrucción 3D es que el sistema de tracking usado que se encarga de estimar la posición y orientación relativa de la cámara respecto al entorno acumula error a lo largo del tiempo. Este fenómeno es conocido drift y repercutirá directamente de forma negativa en la calidad de la reconstrucción 3D obtenida.

Con el fin de solventar estas limitaciones el método de BundleFusion incorpora la información de color de la cámara RGB-D para realizar la reconstrucción de 3D de forma más precisa minimizando el drift de la cámara. Concretamente para resolver este problema en cada frame calculan los puntos característicos de la imagen RGB capturada por la cámara mediante el algoritmo SIFT (Lowe, D. G. 1999) en GPU para realizar la re-localización y refinamiento de las poses de la cámara (loop closer). Además, también incorporan al método de reconstrucción el algoritmo BundleAdjustment que permite refinar de forma global las poses de la cámara obtenidas por el método de tracking, y en consecuencia obtener resultados más precisos a la hora de realizar la reconstrucción. Destacar que estos procesos se realizan de forma simultánea al método de fusión. Finalmente se obtiene la malla en color de la escena aplicando el algoritmo Marching Cubes (Newman, T. S., & Yi, H. 2006) sobre el volumen TSDF previamente creado.

Algunos de los requisitos que debe de cumplir son los siguientes:

- Modelado de superficies de alta calidad: Se requiere que la reconstrucción 3D sea de alta calidad, no tenga ruido en la escena y, además, sea consumible para cualquier aplicación gráfica estándar.

- Escalabilidad: El sistema debe soportar el escaneo tanto de superficies pequeñas al igual que con espacios más extensos. Debe preservar la estructura global del entorno y mantener una precisión alta.
- Consistencia del modelo global: Dependiendo del espacio que vayamos a escanear, es necesario corregir las distintas desviaciones de la posición y también, errores de estimación. Esta corrección se vuelve complicada debido a que va a ser a tiempo real, pero nos debe permitir la revisión de áreas que previamente hayan sido escaneadas o la corrección de distintos bucles que nos hayan podido ocurrir durante el escaneo
- Seguimiento robusto de la cámara: En muchas ocasiones durante la reconstrucción 3D, la cámara puede perder la continuidad, esto es debido a que falla en superficies 3D que no estén suficientemente texturizadas o tengan una geometría demasiado simple, por ejemplo, paredes de color blanco entre otros.
- Actualizaciones de modelos sobre la marcha: Como bien se ha citado anteriormente, el escaneo va a ser a tiempo real, por lo que el sistema, tiene que ser capaz de ir integrando los datos en una representación 3D y poder llegar a visualizar estos de una forma interactiva.

## 2.5.- DATASETS DE ESPACIOS FISICOS INTERIORES GENERADOS MEDIANTE TECNICAS DE RECONSTRUCCION 3D

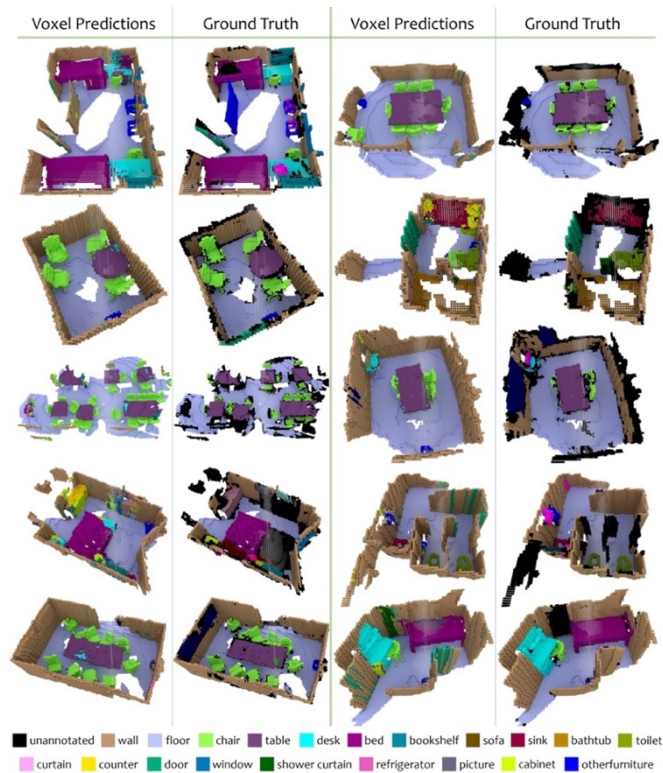
A continuación, se describen algunas de los datasets más importantes espacios físicos interiores generados mediante técnicas de reconstrucción 3D.

### 2.5.1.- SCANNET

El dataset Scannet (Dai, A., Chang, A., X. et al. 2017) es uno de los más importantes y utilizados por la comunidad científica. Este contiene un total de 1513 escaneos, los cuales se han realizado en un total de 707 escenas reales. Esto suma un total de 2,5 millones de imágenes obtenidas con un sensor RGB-D.

Respecto al hardware, han utilizado como sensor RGB-D, Microsoft Kinect v1 y para visualizar el escaneo en tiempo real han utilizado el iPad Air 2, el cual se sincronizaba con el sensor de profundidad. Con este hardware se obtienen diferentes imágenes RGB-D sincronizadas.

A continuación, se muestran diferentes ejemplos tanto de la reconstrucción 3D obtenida, así como la segmentación semántica de los principales elementos de la escena de alguna de las instancias reales reconstruidas.



*Ilustración 11. Clasificación de los diferentes objetos de un espacio en el dataset SCAN-NET*

Se observa cómo se han utilizado 21 clases distintas para la clasificación de los diferentes objetos que incluye el entorno.

### 2.5.2.- SUN RGB-D

SUN RGB-D contiene 10335 imágenes RGB-D con anotaciones de 2D y 3D así como la reconstrucción 3D del entorno obtenida a partir de estas imágenes RGB-D.

Los creadores de este dataset, al igual que los de Scannet, recalcan que actualmente el “cuello de botella” se está produciendo en el tamaño de los ficheros, ya que como se ha comentado anteriormente, si éstos son muy grandes la compresión se vuelve muy difícil de realizar.

Se han utilizado diferentes sensores de profundidad, más concretamente han utilizado cuatro sensores, estos son, Intel RealSense 3D camera, con la cual dejaron los parámetros predeterminados que venían de fábrica, Asus Xtion LIVE PRO, en la cual utilizaron los parámetros predeterminados de la biblioteca OpenNI, y, por último, tanto la Microsoft Kinect v1 como la Microsoft Kinect v2 donde estas cámaras sí que las calibraron con el uso de la caja de herramientas de calibración estándar (Song, S., Lichtenberg, S. P., & Xiao, J. 2015)

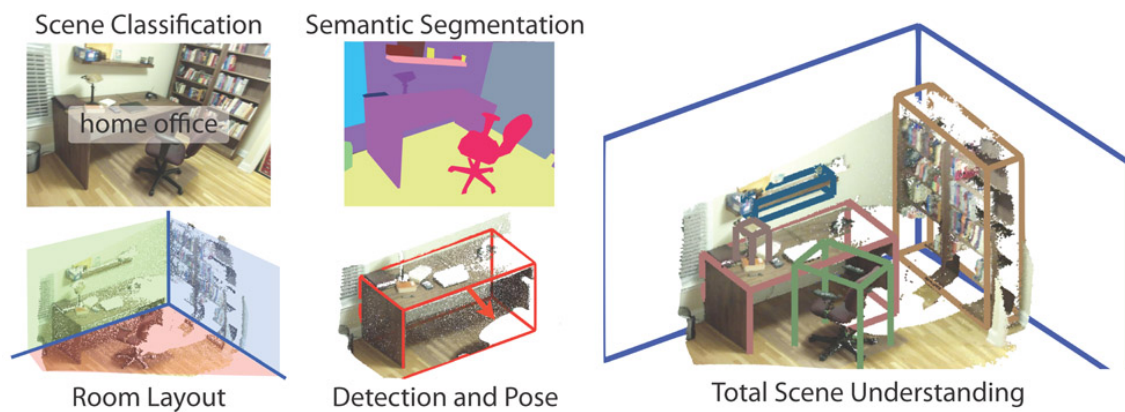


Ilustración 12. Clasificación de los diferentes objetos de un espacio en el dataset SUN-RGB-D

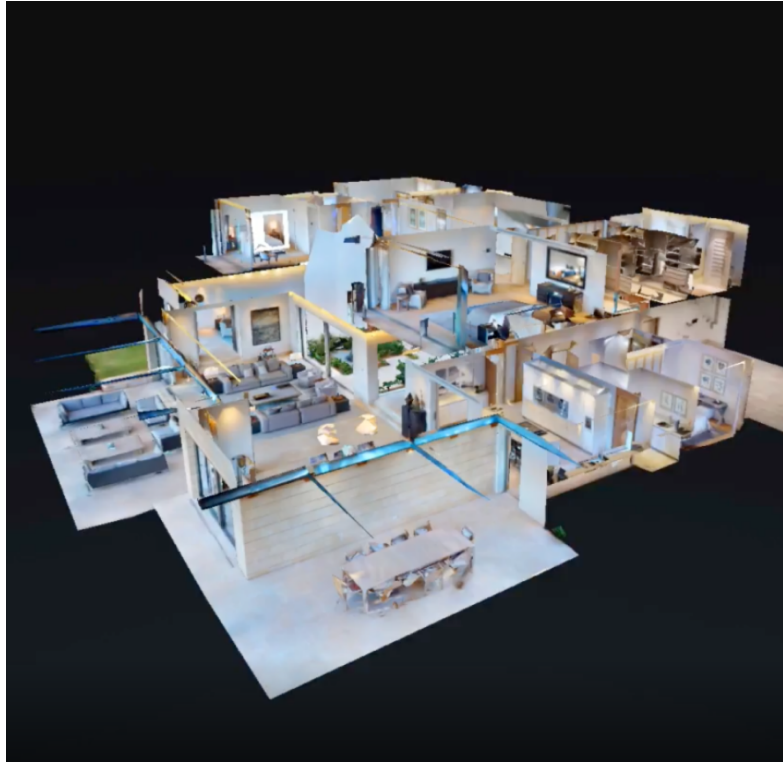
### 2.5.3.- HABITAT

Habitat-Matterport 3D es un dataset que contiene la reconstrucción 3D de 1000 espacios físicos diferentes. En este caso, la reconstrucción se ha centrado en espacios de gran tamaño, como pueden ser pisos, tiendas, etc. Habitat-Matterport contiene un total de  $112.500m^2$  navegables, los cuales contiene 10600 habitaciones, alrededor de unos 1920 pisos.

Este dataset se caracteriza por la escala, la integridad y la fidelidad. Se ha conseguido en este dataset la reducción de las reconstrucciones incompletas, ya que en los datasets mencionados anteriormente había ciertos puntos del entorno que no se había podido completar la reconstrucción. A diferencia de los datasets anteriores, éste proporciona reconstrucciones más precisas y con una mayor fidelidad visual.

Para la creación de este dataset se ha utilizado un sensor de profundidad basado en Matterport 2, junto con el software Matterport, Inc (Ramakrishnan, Santhosh K., et al. 2021).

A continuación, se puede observar el modelo de la reconstrucción 3D de una vivienda privada.



*Ilustración 13. Ejemplo de reconstrucción 3D en el dataset Habitat-Matterport*

## 3.- PROPUESTA DESARROLLO Y CREACIÓN

### 3.1.- HARDWARE UTILIZADO

Para la realización de este trabajo se empleó un montaje hardware bastante específico. Básicamente, se utilizó como elemento de captura una cámara RGB-D, un ordenador con altas capacidades gráficas, un dispositivo de visualización y un dispositivo de entrada de datos en el ordenador, tipo ratón/teclado. En este apartado se describen todos estos elementos hardware.

#### 3.1.1.- CÁMARAS RGB-D

Durante la realización de este trabajo se han utilizado dos cámaras RGB-D diferentes que son descritas a continuación.

### 3.1.1.1.- CÁMARA INTEL REALSENSE D435

La cámara Intel Realsense D435 es una cámara RGB-D USB que ofrece una solución estereoscópica, donde se obtiene información precisa de la profundidad. Cabe destacar que su FOV es ideal para el desarrollo de aplicaciones de Realidad Aumentada y Realidad Virtual junto con la robótica.

	<b>Intel Realsense D435</b>
Dimensiones	90x25x25 mm
Peso	72g
Cámara RGB	1920x1080
Cámara de profundidad	1920x1080
Distancia máxima	14m
Distancia mínima	0.2m
Depth FOV	86° (H) x 57° (V) x 94°(D)
Velocidad de salida de los datos	400Hz
Precio	350€

Tabla 3. Especificaciones de la cámara RGB-D Intel RealSense D435



Ilustración 14. Cámara Intel RealSense D435

### 3.1.1.2.- CÁMARA INTEL REALSENSE D415

La cámara Intel Realsense D415 utiliza visión estéreo para calcular la profundidad. Se trata de una cámara RGB-D que cuenta con alimentación USB, sensor RGB, sensores de profundidad y un proyector de infrarrojos.

	<b>Intel Realsense D415</b>
Dimensiones	99x20x23 mm
Cámara RGB	1920x1080
Cámara de profundidad	1280x720
Distancia máxima	10m
Distancia mínima	0.2m
Depth FOV	64° (H) x 41° (V) x 77°(D)
Precio	395€

*Tabla 4. Especificaciones de la cámara RGB-D Intel RealSense D415*



*Ilustración 15. Cámara Intel RealSense D415*

### 3.1.2.- MOCHILA HP BACKPACK G2

Como dispositivo de procesamiento para realizar la reconstrucción 3D se utilizó la mochila HP Z VR Backpack G2 Workstation la cual se caracteriza por su portabilidad y su capacidad de procesamiento.

Estas son algunas de las características principales de la mochila HP Z VR.



<b>HP Z VR Backpack G2 Workstation</b>	
Procesador	Intel Core i7-8850H (frecuencia base de 2,6 GHz, hasta 4,3 GHz con tecnología Intel Turbo Boost, 9 MB de caché L3, 6 núcleos)
RAM	32 GB DDR4-2666 SDRAM ECC (2 x 16 GB)
Almacenamiento	512 GB SSD
GPU	Nvidia GeForce RTX 2080 con 8 GB GDDR6
Sistema operativo	Windows 10 Pro 64
Dimensiones	33,3 x 23,6 x 6,09 cm
Peso	4,66 kg (incluido baterías y arnés)
Puertos y conectores	1 combo de auriculares y micrófono; 1 HDMI; 1 Thunderbolt (USB Type-C); 2 USB 3.0; 2 mini-DisplayPort 1.4
Baterías	2 (una para la CPU y otra para la GPU)
Precio	5030 €

Tabla 5. Especificaciones de la HP Z VR Backpack G2 Workstation



Ilustración 16. HP Z VR Backpack G2 Workstation

### 3.1.3.- TABLET

Para realizar la captura del entorno era necesario visualizar en tiempo real el proceso de escaneado, en primera instancia se empleó la Samsung Galaxy Tab A, la cual es una tableta de 10,1 pulgadas con sistema operativo Android. En dicha Tablet se utilizaba un sistema de control remoto de escritorio para visualizar y controlar la sesión que ejecuta

la mochila HP. Concretamente se utilizó Parsec, que como característica principal ofrece el control remoto con escritorio acelerado por hardware.

<b>Samsung Galaxy Tab A (2016)</b>	
Procesador	Samsung Exynos 7870 (4 x 1,6 GHz Cortex-A53, 4 x 1 GHz Cortex A-53)
RAM	2 GB LPDDR3
Almacenamiento	16 GB
Resolución	1.920 x 1.200
Sistema operativo	Android 6.0 Marshmallow
Dimensiones	255,4 x 155,3 x 8,2 mm
Peso	525 gramos
Cámaras	Posterior de 8 megapíxeles. Frontal de 2 megapíxeles
Baterías	7300 mAh
Precio	279 €

*Tabla 6. Especificaciones de la tablet Samsung Galaxy Tab A*



*Ilustración 17. Tablet Samsung Galaxy Tab A (2016)*

### 3.1.4.- MONITOR USB-C

Posteriormente al uso de la tablet, se decidió utilizar un monitor ligero con capacidad táctil, concretamente el modelo VSEVEN L156TCH-1G. Este monitor USB-C tiene 15,6”, y supera notablemente el rendimiento de la tablet que se estaba usando anteriormente, tanto por su tamaño como por la respuesta táctil.

<b>VSEVEN L156TCH-1G</b>	
Puertos de entrada	USB-C y HDMI
Resolución	1.920 x 1.200
Dimensiones	35.5 × 1.1 × 22.3 cm
Peso	0,77 kg
Resolución máxima	1920 x 1080
Tiempo de respuesta	15ns
Precio	350 €

*Tabla 7. Especificaciones del monitor VSEVEN L156TCH-1G*



*Ilustración 18. Monitor USB-C VSEVEN L156TCH-1G*

## 3.2.- POR QUÉ SE HA UTILIZADO ESTE HARDWARE

En el siguiente apartado se presentan las principales características del hardware que se ha utilizado para la realización de este estudio.

### 3.2.1.- CAPACIDAD DE PROCESAMIENTO

El sistema presentado en este trabajo se caracteriza por el uso de la tecnología CUDA (Compute Unified Device Architecture). Gracias al uso de esta tecnología se consigue extraer el máximo rendimiento del sistema. CUDA es una extensión del lenguaje de programación C que permite aprovechar toda la capacidad de procesamiento paralelo de las GPUs, pudiendo éstas resolver problemas complejos computacionales y consiguiendo el máximo rendimiento, en el caso de este trabajo, una reconstrucción 3D en tiempo real (Oriol Bermúdez, E, 2010).

El objetivo principal de esta tecnología es llegar a aprovechar la arquitectura multiprocesador de las tarjetas gráficas, las cuales van a poder realizar cálculos de forma paralela.

### 3.2.2.- AUTONOMÍA

Con el fin de poder desplazarse por una estancia a reconstruir sin ataduras de ningún tipo, se escogió la mochila HP Z VR Backpack G2 Workstation, descrita en el apartado 3.1.2. En la mochila se incluían dos baterías portátiles junto con el estuche de carga.

Además, se compraron dos baterías extras, las cuales permitían doblar la autonomía de la mochila a la hora del escaneo. Con este conjunto de baterías cargadas completamente se podían realizar escaneos durante 3 horas aproximadamente.

Con esta autonomía se podían realizar varios escaneos en entornos distintos, sin la necesidad cargar las baterías cada vez que se realizaba la reconstrucción de un espacio.

### 3.2.3.- USABILIDAD

Respecto a la usabilidad habría que comentar diferentes aspectos ya que en principio resultó un poco difícil realizar los escaneos de una manera correcta y óptima. Inicialmente, para el proceso de captura, se utilizaba como elemento de visualización y

entrada de datos la tablet Samsung y el software Parsec. Dicha configuración era muy sensible a los cortes de red, ya que, ante esta circunstancia, se cortaba la visualización y el control del proceso de reconstrucción, por lo que antes de empezar se tenía que asegurar la disponibilidad de una conexión Wifi y que ésta ofreciera cobertura para toda la zona a escanear.

Como ya se ha comentado, este inconveniente se debe al uso de Parsec, ya que dicho Software requiere de una elevada y continua tasa de transmisión. Se intentó solventar esta dependencia haciendo uso de un smartphone actuando de hotspot, para la mochila y la tablet, pero tampoco se obtuvieron buenos resultados. Por este motivo se decidió utilizar un monitor táctil USB-C que llevaría a cabo las mismas funciones que la tablet sin necesidad de Parsec, y, por tanto, tampoco de conexión de red.

Asimismo, con el cambio de la tablet al monitor, se consiguió eliminar el LAG del sistema, ya que éste, cuando se utilizaba Parsec, en numerosas ocasiones daba problemas a la hora de poder estar visualizando el escaneo de manera fluida, sin embargo, el monitor Vseven, al estar conectado directamente al ordenador con el cable USB-C, proporcionaba el máximo rendimiento.

Otro aspecto a comentar es el del tamaño de la pantalla. En los primeros escaneos se utilizó la Samsung Galaxy Tab A (2016), que como ya se describió, tiene un tamaño de 10 pulgadas, circunstancia que dificultaba la tarea de escaneado, ya que en la misma pantalla se visualizaba en tiempo real el resultado del escaneado, y el teclado del sistema, para controlar la aplicación de captura. Al cambiar de dispositivo de visualización y control por el monitor USB-C VSEVEN, se pasó a disponer de una pantalla de 15,6", donde todos los problemas de visualización y control anteriormente experimentados desaparecieron.

### 3.3.- SOFTWARE UTILIZADO

En este apartado se presenta el software utilizado para realizar las modificaciones en la implementación oficial de BundleFusion, las cuales se describen posteriormente.

### 3.3.1.- DESCRIPCIÓN DEL ENTORNO DE DESARROLLO

#### 3.3.1.1.- C++

C++ es un lenguaje de programación diseñado por Bjarne Stroustrup el cual fue una extensión del previamente conocido como C. C++ a diferencia de C, incluye la programación orientada a objetos (OOP).

La principal ventaja de este lenguaje de programación es que se trata de un lenguaje de programación multiparadigma. Un lenguaje multiparadigma permite aplicar diferentes conjuntos de metodologías y técnicas a un software, gracias a ello se consigue resolver una mayor cantidad de problemas. Asimismo, proporciona soluciones rápidas y eficientes.

#### 3.3.1.2.- *MICROSOFT VISUAL STUDIO*

Microsoft Visual Studio es un entorno de desarrollo integrado (IDE en inglés) compatible con Windows y MacOS. Permite la programación en múltiples lenguajes, tales como C++, C#, Visual Basic .NET, F#, Java, Python, Ruby y PHP. En este trabajo se empleó la versión de 2019.

#### 3.3.1.3.- *WINDOWS*

Microsoft Windows 10 ha sido el sistema operativo empleado para la realización de este trabajo de fin de grado. Windows es un sistema operativo multitarea que ha sido creado por Microsoft. Windows contiene diferentes programas, los cuales permiten aprovechar los recursos del ordenador.

#### 3.3.1.4.- *SDK DE LA CÁMARA INTEL*

Intel RealSense SDK 2.0 es una biblioteca multiplataforma que ha sido diseñada para las cámaras RGB-D Intel RealSense, ésta se incluye en las series D400 y L500. Este SDK permite acceder vía programación al buffer de la cámara y así recuperar la información de color y profundidad de la escena. Además, permite acceder a la calibración intrínseca y extrínseca (Herrero, 2019) (IntelRealSense, 2021).

### 3.3.1.5.- LIBRERÍAS UTILIZADAS

Además, se han utilizado un conjunto de librerías en C++, de las cuales destacan la librería MLib (Spark, A. 2015) y OpenCV (Bradski, G., Kaehler, A. 2000). Estas dos librerías facilitan tanto el procesamiento y análisis de las imágenes obtenidas por el sensor RGB-D así como de los modelos 3D obtenidos durante el proceso de reconstrucción.

### 3.3.2.- MODIFICACIONES REALIZADAS SOBRE LA IMPLEMENTACIÓN OFICIAL DEL MÉTODO

La principal modificación realizada sobre la implementación oficial consistió en añadir el soporte completo de las cámaras Intel RealSense, ya que dicha implementación no las soportaba. Asimismo, se implementó un método que activaba la exposición automática de la cámara utilizada en el proceso de captura. Es decir, de forma automática, la cámara variaba la apertura y la velocidad de obturación para obtener una correcta iluminación de las imágenes captadas, y así independizar el proceso de captura de las condiciones lumínicas de la estancia.

Además, se ha modificado el método de fusión de volumétrica en caso de que opcionalmente se requiere un mayor nivel de detalle en la reconstrucción. Para realizar esta tarea se ha utilizado el método de fusión volumétrica de la librería Open3D.

Por último, se han post-procesado los modelos 3D obtenidos durante el proceso de reconstrucción eliminando el ruido y artefactos de estos (p. ej. Eliminación de las componentes conexas de tamaño muy pequeño).

## 3.4.- DESCRIPCIÓN DEL FLUJO DE TRABAJO PARA LA CREACIÓN DEL DATASET DE ESPACIOS INTERIORES

Con el fin de demostrar la eficacia del método de reconstrucción implementado en este proyecto se realizaron múltiples pruebas de este método sobre diferentes espacios interiores reales. Como resultado de estas pruebas se creó un dataset (ExperienceRooms) con la información y resultados más relevantes obtenidos durante el proceso de reconstrucción de cada una de las escenas. Estos espacios interiores fueron

principalmente extraídos de instancias reales del Instituto de Investigación e Innovación en Bioingeniería (I3B) en la UPV, aunque también se capturaron algunos espacios públicos y privados fuera de Valencia. De forma más precisa se realizó la reconstrucción 3D de 24 diferentes espacios interiores:

- 4 despachos
- 1 sala de servidores
- 4 talleres de trabajo
- 2 almacenes
- 2 comedores
- 2 baños
- 1 tienda de ropa
- 1 pizzería
- 1 cocina
- 2 habitaciones
- 2 salas de reuniones
- 2 salas de proyección

El dataset creado, ExperienceRooms, permite complementar la información del dataset de Scannet, y puede ser de interés para la comunidad científica en el campo de scene understanding.

Por otro lado, destacar que además se realizaron múltiples instancias de algunos de los espacios físicos reconstruidos en las que se variaron algunos de los parámetros de reconstrucción (p. ej. el voxel size para realizar la fusión volumétrica o la distancia máxima de la cámara de profundidad), la trayectoria realizada, así como las condiciones del entorno (p.ej. diferentes condiciones lumínicas o diferente disposición espacial de los principales elementos de ésta).

De cada espacio físico reconstruido se almacenó en el dataset la siguiente información:

- Archivo .sens. Este archivo contiene tanto las imágenes RGB-D utilizadas por la cámara para realizar la reconstrucción, así como las poses relativas de la cámara en cada frame (matriz de transformación homogénea 4x4) y los parámetros



intrínsecos y extrínsecos del sensor. Además, este archivo contiene los datos del IMU (Inertial Measurement Unit, incluyendo la aceleración y la velocidad angular en el caso de que el sensor RGB-D utilizado tuviese esta capacidad (p. ej. Intel Realsense D435i).

- Archivo .ply. Malla en color del espacio físico reconstruido. Destacar que esta malla ya está limpia de ruido y artefactos y está situada en el sistema de coordenadas de Scannet.
- Archivo .xml. Este archivo describe textualmente el tipo de escena y los principales elementos que conforman ésta (mesas, sillas, etc.).
- Archivo .yaml. Este archivo contiene las matrices de transformación que permiten situar en el sistema de coordenadas de Scannet la malla inicialmente obtenida por BundleFusion.

En particular, para realizar la reconstrucción 3D se utilizó como sensor RGB-D la cámara Intel RealSense D415. Las imágenes RGB del sensor se capturaron a una resolución de 640x480 y una tasa de 60fps, mientras que las imágenes de profundidad se capturaron a una resolución de 1280x720 y una tasa de 60 fps, siendo éstas automáticamente registradas con las imágenes RGB del sensor. Como resultado, estas imágenes de profundidad tienen una resolución de 640x480 píxeles.

Cabe destacar, que al igual que el dataset de Scannet, se descartaron algunos de los resultados obtenidos en el proceso de reconstrucción cuando:

- La secuencia de imágenes RGB-D utilizada para realizar la reconstrucción 3D era demasiado corta.
- Se obtenía un alto residual en la reconstrucción.
- El porcentaje de frames alienados respecto el total de frames de la secuencia de imágenes RGB-D obtenido al realizar la reconstrucción 3D era bajo.

Por último, se describen las diferentes tareas que se realizaban para preparar el equipo de captura antes de empezar a realizar una reconstrucción 3D. Se puede dividir el flujo de trabajo en 5 procesos.

- **Hardware:** Primera tarea y casi la más importante, es tener las baterías de la mochila cargadas con antelación. La mochila funciona con 2 baterías, las cuales proporcionan una autonomía de unas 1.5 horas escaneando entornos. Hay que tener en consideración que se disponía de otra pareja de baterías por lo que se doblaba la autonomía. Gracias a la sencillez del hardware tan solo queda conectar el monitor portátil con un cable USB-C a la mochila y colocar sobre el monitor la cámara RGB-D.
- **Software:** Punto clave en este estudio. Uno de los objetivos de este trabajo es la obtención de una configuración óptima para que éste se pueda usar en cualquier entorno. La preparación del flujo de trabajo incluye la edición de diferentes parámetros de la configuración para ir probando éstos durante el proceso de reconstrucción. Esta configuración se encontraba en el fichero del proyecto nombrado como “ZParametersDefault.txt”.
- **Prueba:** A continuación, el objetivo era probar el rendimiento del sistema antes de proceder a realizar la reconstrucción de un entorno. Con este rápido test se observaba el rendimiento del sistema y la adaptación a los diferentes entornos lumínicos.
- **Entorno:** En primer lugar, había que desplazarse al espacio a capturar y comprobar allí el estado de éste. La acción más común era comprobar las condiciones lumínicas del espacio. Además, también era muy importante intentar predecir algunos puntos en el entorno que iban a ser conflictivos, ya que, para una reconstrucción óptima, se necesitan distintas imágenes de referencia capturadas por la cámara RGB-D. Éstas deberían ser lo suficientemente descriptivas, tanto en apariencia como en geometría.
- **Comprobación:** Por último, y no menos importante, al finalizar una reconstrucción se comprueba el resultado, visualizando el archivo con extensión ply mediante el software MeshLab (Cigonini, P., Ranzuglia, G., Callieri, M. et al. 2011). El monitor USB-C aparte de mejorar el rendimiento, también ha sido clave por su gran tamaño de pantalla, el cual, permitía visualizar de una forma totalmente adecuada la reconstrucción del entorno en él. Fue útil ya que a simple vista se podía comprobar si la reconstrucción era correcta.

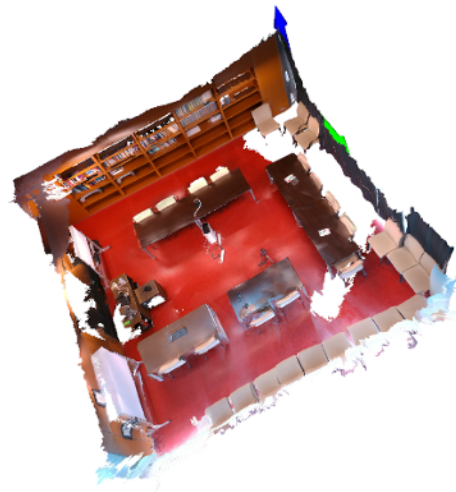
## 4.- RESULTADOS

### 4.1.- EJEMPLOS

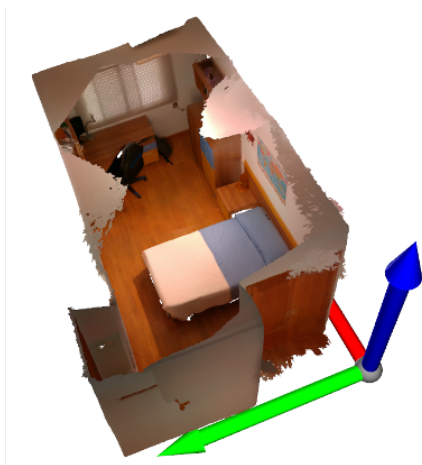
A continuación, con el fin de demostrar el desempeño del método de reconstrucción 3D de espacios interiores presentado en este trabajo, se mostrarán una serie de imágenes que representan algunas de las reconstrucciones 3D obtenidas. Este dataset está compuesto por un total de 73 reconstrucciones 3D realizadas en 24 espacios físicos diferentes.



*Ilustración 19. Reconstrucción 3D de una sala de estar*



*Ilustración 20. Reconstrucción 3D de una sala de reuniones*



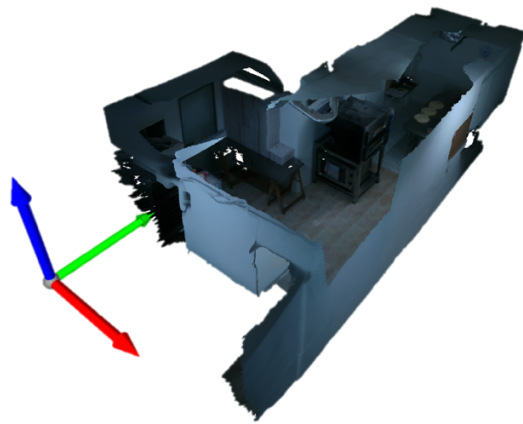
*Ilustración 21. Reconstrucción 3D de una habitación individual*



*Ilustración 22. Reconstrucción 3D de una sala de estar*



*Ilustración 23. Reconstrucción 3D de una tienda de deportes*



*Ilustración 24. Reconstrucción 3D de una pizzeria*



*Ilustración 25. Reconstrucción 3D de un almacén*



*Ilustración 26. Reconstrucción 3D de una sala de reuniones*



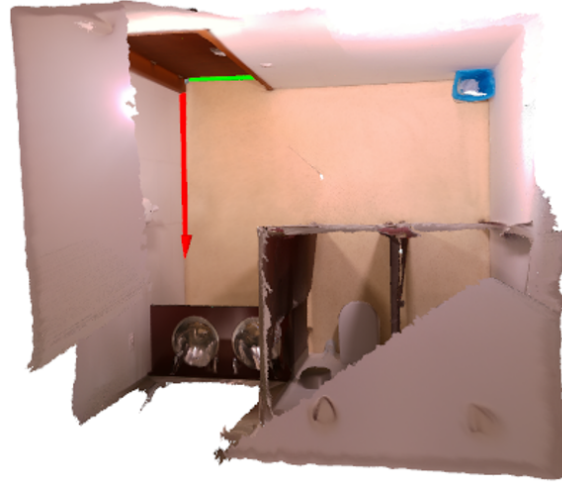
*Ilustración 27. Reconstrucción 3D de un taller mecánico*



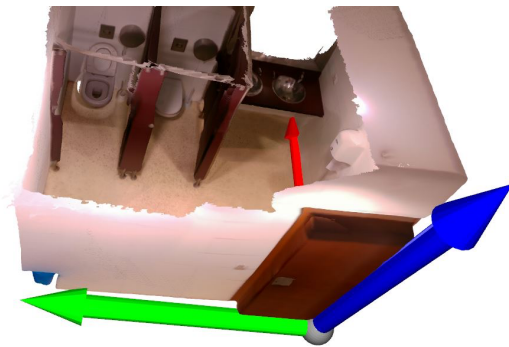
*Ilustración 28. Reconstrucción 3D de un comedor de la universidad*



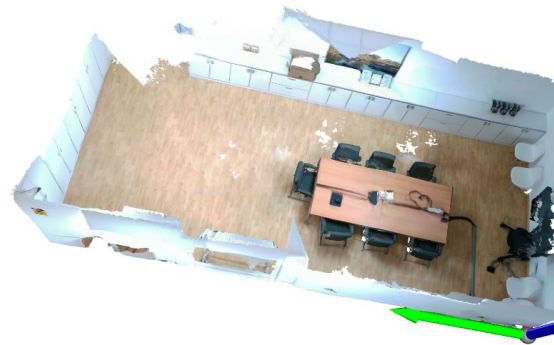
*Ilustración 29. Reconstrucción 3D de un baño individual*



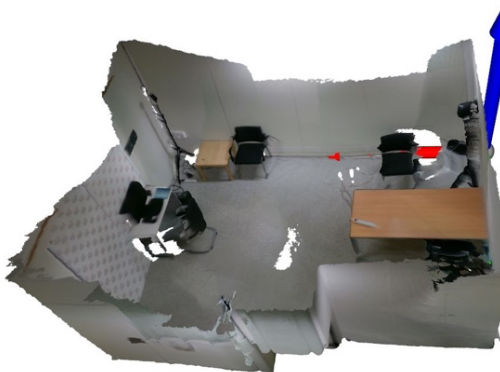
*Ilustración 30. Reconstrucción 3D de un baño doble*



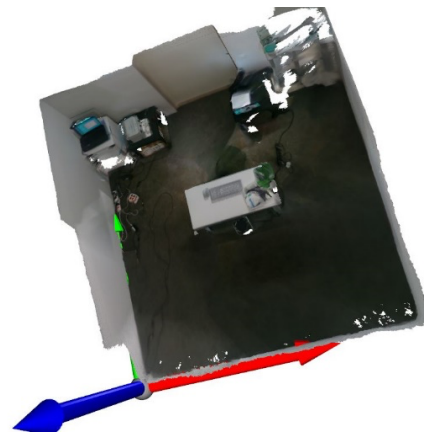
*Ilustración 31. Reconstrucción 3D de un baño doble*



*Ilustración 32. Reconstrucción 3D de una sala de reuniones*



*Ilustración 33. Reconstrucción 3D de una habitación de trabajo*



*Ilustración 34. Reconstrucción 3D de una sala de pruebas*



*Ilustración 35. Reconstrucción 3D de una sala de servidores*



*Ilustración 36. Reconstrucción 3D de un despacho*

Se puede observar como la ilustración 19 y 22 representan un comedor de una casa, por otro lado, las reconstrucciones 3D de las ilustraciones 23 y 25 representan entornos con unas dimensiones mucho mayores a las anteriores. En este último caso se obtuvo una reconstrucción 3D de una tienda de deportes y de un almacén.

#### 4.2.- CONCLUSIONES DE USABILIDAD

En el siguiente apartado se describen las conclusiones de usabilidad más destacables obtenidas en el presente trabajo.

Inicialmente, como bien se ha comentado anteriormente, se utilizó la tablet Samsung junto con Parsec, donde el rendimiento no fue el esperado debido a los cortes de red. Se intentó solventar este inconveniente usando un teléfono móvil como hotspot, dicha configuración mejoraba la experiencia, pero no lo suficiente para realizar una reconstrucción 3D precisa. Así que, con esta experiencia, se decidió comprar un monitor USB-C, el cual no requiere que la mochila este conectada a la red, por lo que se elimina una variable crítica del sistema de captura.

También, con el monitor USB-C se consiguió una mayor facilidad para el escaneado debido al tamaño de la pantalla. En la tablet Samsung el tamaño es de 10", mientras que en el monitor VSeven se incrementa el tamaño hasta 15,6". Con este cambio, se mejoró el agarre del monitor y se evitaba producir toques en la pantalla indeseados. Además, se mejoraba notablemente el manejo del sistema, ya que se podía visualizar la aplicación y un teclado virtual para introducir los comandos, al mismo tiempo y sin problemas.

Por otro lado, la mochila facilitó en gran parte los escaneos. Primero, por la ergonomía de la misma, ya que esta se adapta perfectamente a la espalda. Además, con el conjunto de baterías se podían realizar escaneados durante 1,5h. En la realización del presente trabajo, se contaba con un conjunto extra de baterías, el cual proporcionaba el doble de tiempo para realizar escaneos.

Sin embargo, el proceso de poner la mochila era complicado, ya que, tanto el monitor como el sensor RGB-D se conectan a la mochila, por lo que en la puesta a punto se debía sujetar el monitor y ponerse la mochila.

Como conclusión general de este apartado, se ha conseguido definir un conjunto de hardware óptimo para la realización de reconstrucción 3D de espacios interiores.

#### 4.3.- CONCLUSIONES DE LIMITACIONES TÉCNICAS

En el siguiente apartado se van a describir las limitaciones técnicas que más afectan al escaneado de un entorno en el contexto de este trabajo.

La mayor limitación que se ha encontrado, por lo que respecta a las características del entorno a escanear, es la dificultad para capturar paredes o superficies planas con colores uniformes. Para obtener un buen resultado del escaneado del entorno, se debe considerar que cada una de las paredes que componen dicho entorno, debe contener algún tipo de objeto que actúe como punto de anclaje, ya que, si no es así, el sistema es incapaz de encontrar los puntos solapados de las diferentes imágenes capturadas, y por tanto falla el algoritmo de reconstrucción.

Por otro lado, también se debe tener en cuenta que todo tipo de cristales, ventanas, o espejos son problemáticos. Dicho esto, las televisiones o cualquier otro tipo de pantalla también presentan problemas, tanto encendidas como apagadas. Esto es debido a que el sensor de profundidad no puede obtener información precisa en este tipo de superficies.

El tamaño del entorno también es una limitación técnica, ya que, si se pretende obtener una reconstrucción 3D, con un nivel de detalle alto de una habitación con un gran tamaño, el sistema puede llegar a presentar problemas de falta de memoria. Sin embargo, en las habitaciones más pequeñas, el detalle de la reconstrucción 3D es mejor, debido a que un entorno con menor tamaño permite escanear la misma zona más veces.

Por último, no se puede utilizar un tiempo mayor a 10 minutos para capturar una estancia, ya que se consume toda la memoria del sistema.

#### 4.4.- PARÁMETROS ÓPTIMOS

A continuación, se describen los parámetros más influyentes a la hora de la adquisición de datos en una reconstrucción 3D de una escena. Los siguientes parámetros se encuentran en el fichero “zParametersDefault.txt”.

- `s_sensorIdx`. Tiene un valor por defecto de 6 para el uso con las cámaras Intel Realsense D415/435. En el caso de leer un fichero con el formato “.sens” el valor de la variable debería de ser 8.
- `s_windowWidth` y `s_windowHeight`. Estos dos parámetros permiten cambiar el tamaño de la ventana donde se visualiza la reconstrucción. Por defecto está establecido a 640x480. El usuario podría cambiar el valor de dichas variables a 1280x720 si en ese tamaño se encuentra más comodidad.
- `s_sensorDepthMax`. Indica la máxima distancia que recoge el sensor de profundidad, viene dada en metros. Por defecto tiene un valor de 4m y en el dataset de Scannet tiene un valor de 5.5m. Cuando se realice una reconstrucción de una habitación o espacios no muy grandes, con un valor de 4m del `s_sensorDepthMax` será suficiente, pero, por el contrario, si el espacio tiene unas dimensiones considerables se debería cambiar el valor a 5 o 6 metros.
- `s_SDFVoxelSize`. Uno de los parámetros más importantes del sistema, por defecto tiene un valor de 0.007 metros. Cuanto más pequeño sea este valor, mayor precisión se va a conseguir en la reconstrucción, sin embargo, empleará muchos recursos y podría producir un error en el sistema. Por tanto, el valor mínimo para la reconstrucción debe ser 0.007m. En Scannet el valor utilizado es de 0.01 metros. A diferencia del valor anterior, éste proporciona una reconstrucción menos detallada pero el rendimiento del sistema es mejor.
- `s_binaryDumpSensorUseTrajectory = false`. Con este valor se fuerza el recálculo de la trayectoria del “.sens” y se obtienen valores con más precisión.
- `s_binaryDumpSensorFile`. Este parámetro contiene el path donde se encuentra el fichero .sens (P. ej. “dump/dump.sens”).



- `s_marchingCubesMaxNumTrinagles`. Indica el número máximo de triángulos para realizar la reconstrucción 3D de la malla mediante el algoritmo Marching Cubes. Por defecto tiene un valor de 6000000. En principio este parámetro no se va a modificar, solo se vería afectado en el caso de que se produjese un error a la hora de realizar la reconstrucción.
- `s_recordData = true`. Permite guardar archivos `.sens` o `.sensor` al pulsar la tecla “8”.

#### 4.5.- CONCLUSIÓN DEL TIEMPO EMPLEADO PARA UNA RECONSTRUCCIÓN 3D

El tiempo empleado en realizar la reconstrucción es bastante variable ya que éste depende de muchos factores, pero se podría hacer una estimación de alrededor de 6 a 7 minutos para obtener una reconstrucción con la geometría perfecta y con un alto nivel de detalle.

Los factores más influyentes son el tamaño de la habitación, y el nivel de detalle de la reconstrucción que se pretende obtener. El primero, es bastante lógico, ya que, si se realiza un escaneado de, por ejemplo, un comedor de una empresa o de una universidad, éste tendrá un tamaño bastante considerable. En este tipo de entornos se prioriza que la geometría aparezca completa, sin agujeros, sacrificando el nivel de detalle que era un poco menor. Para conseguir un resultado con un nivel de detalle elevado es necesario que, durante el proceso de captura, se realicen unas 2 o 3 pasadas por el mismo lugar. En conclusión, para obtener una reconstrucción 3D con un buen nivel de detalle se suele emplear unos 7 minutos.

### 5.- CONCLUSIONES

#### 5.1.- CUMPLIMIENTO DE OBJETIVOS

En el siguiente apartado se describe el grado de cumplimiento obtenido en los objetivos planteados para el TFG.

El objetivo principal era crear un novedoso dataset (`ExperienceRooms`) formado por la reconstrucción 3D de múltiples espacios interiores. En particular, `ExperienceRooms`, cuenta con un total de 24 escenas y 73 reconstrucciones que representan múltiples

espacios físicos de diferentes características (p. ej. habitaciones de viviendas, diferentes salas de reuniones de la universidad, una tienda de deportes, etc.).

Por lo que respecta a los objetivos secundarios, también se han conseguido. En primer lugar, se ha modificado la implementación oficial de BundleFusion para soportar la cámara Intel RealSense. En segundo lugar, se han encontrado las limitaciones técnicas del algoritmo, por lo que respecta a las características físicas de las estancias a capturar (cristales, espejos, paredes sin detalles, etc.). En tercer lugar, se ha estudiado la usabilidad de la herramienta a nivel de software y hardware, obteniendo una configuración óptima que permite capturar estancias de manera sencilla y rápida.

Por último, se ha definido un conjunto de parámetros óptimo para realizar la reconstrucción 3D. A partir de la experimentación se obtuvieron dichos valores, de hecho, se pudieron establecer diferentes conjuntos de valores dependiendo de las características físicas del entorno a reconstruir.

En resumen, se ha conseguido un grado de cumplimiento de los objetivos planteados para este trabajo del 100%.

## 5.1.- APORTACIONES

La principal contribución de este trabajo es la creación de un novedoso, heterogéneo y amplio dataset, ExperienceRooms, formado por los resultados obtenidos tras realizar la reconstrucción 3D de diferentes espacios interiores. Se trata de un dataset que puede ser de interés para la comunidad científica en el campo de scene understanding.

Otra contribución de este trabajo ha sido la modificación de la implementación oficial de BundleFusion, de forma que este método pueda utilizar como sensor de captura las cámaras Intel RealSense. Por otro lado, se modificó el método de captura de forma que en éste se pueda activar o desactivar la exposición automática de las cámaras, de este modo, el método de reconstrucción se adecuaba mejor a las condiciones lumínicas del entorno.

Por otra parte, se ha definido un conjunto hardware óptimo para la creación de reconstrucciones 3D, en el que destaca un monitor táctil USB-C que facilita la tarea del escaneado.

### 5.3.- TRABAJOS FUTUROS

Como principal trabajo futuro se plantea la posibilidad de reemplazar la mochila HP utilizada para este trabajo, por un dispositivo portable tipo smartphone, de este modo, realizar reconstrucciones 3D, se convertiría en un proceso sencillo y asequible para un gran número de personas. Para conseguirlo, se debería reducir el coste computacional y así, poder ejecutarlo en dispositivos con menos potencia.

También se podría ampliar el dataset ExperienceRooms con la inclusión de espacios exteriores, ya que se ha probado la funcionalidad del sistema en dichos entornos.

Además, se podrían aprovechar la reconstrucción 3D de estos espacios para realizar mediciones reales de las estancias y los objetos contenidos en ellas (paredes, cuadros, pilares, etc.). Adicionalmente, con la obtención de medidas del entorno, se podría crear un plano 2D que formara parte del propio dataset.

Por otra parte, se podrían crear espacios 3D virtualizados a partir de las reconstrucciones obtenidas en este trabajo y alguna base de datos de modelos 3D, de forma que se podrían crear diferentes aplicaciones/juegos de Realidad Virtual en los que el espacio de juego sería el modelo previamente obtenido.

Para concluir, los modelos obtenidos también podrían ser imprimidos en 3D, obteniendo una maqueta a escala del entorno escaneado.

## 6.- BIBLIOGRAFÍA

Álvarez-Gutiérrez, E. L., & Jiménez-López, F. R. (2019). Generación de Mapa Global 2D y SLAM usando LiDAR y una Estéreo Cámara para el seguimiento de movimiento de un robot móvil. *Iteckne*, 16(2), 144-156.

Aukstakalnis, S., Blatner D. Silicon Mirage (1992). The Art and Science of Virtual Reality. Peachpit Press, Berkeley, CA, USA, 1992.

- Bitfab. (2022). *bitfab*. Recuperado el Marzo de 2022, de bitfab.io:  
<https://bitfab.io/es/blog/fotogrametria/>
- Bradski, G., & Kaehler, A. (2000). OpenCV. *Dr. Dobb's journal of software tools*, 3, 120.
- Cignoni, P., Ranzuglia, G., Callieri, M., Corsini, M., Ganovelli, F., Pietroni, N., & Tarini, M. (2011). MeshLab.
- Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6), 64-73.
- Dai, A., Chang, A. X., Savva, M., Halber, M., Funkhouser, T., & Nießner, M. (2017). Scannet: Richly-annotated 3d reconstructions of indoor scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5828-5839).
- Dronica. (7 de 11 de 2018). *dronica*. Recuperado el Marzo de 2022, de dronica.es:  
<https://www.dronica.es/blog-de-escaneo-3d/nube-de-puntos-que-es-y-para-que-sirve/>
- Estapé, J. A. (28 de Marzo de 2014). *computerhoy*. Obtenido de  
<https://computerhoy.com/noticias/hardware/asi-es-kinect-20-windows-pc-10937>
- Euronics. (2020). Obtenido de euronics: <https://www.euronics.es/blog/que-es-la-tecnologia-dual-pixel-como-funciona-y-donde-se-encuentra/>
- Fabara, S. (24 de Febrero de 2021). *xatakafoto*. Obtenido de  
<https://www.xatakafoto.com/fotografia-con-movil/samsung-presenta-nuevo-sensor-para-moviles-50mp-nueva-tecnologia-autoenfoco-deteccion-fase>
- Geoslam. (2020). Obtenido de geoslam: <https://geoslam.com/what-is-slam/>
- Insights, A. (2018). Obtenido de Aerial Insights: <https://www.aerial-insights.co/blog/camara-multiespectral/>
- González, I. (2022). *kandio*. Obtenido de kandio: <https://kandiolatam.com/blog/cual-es->

la-importancia-de-manejar-los-lenguajes-de-programacion-multiparadigma

- Heater, C. (1992). Being there: The subjective experience of presence. *Presence Teleoperators Virtual Environ.*, 1(2), 262-271
- Lindeberg, T. (2012). Scale invariant feature transform.
- Levoy, M., Pulli, K., Curless, B., Rusinkiewicz, S., Koller, D., Pereira, L., ... & Fulk, D. (2000, July). The digital Michelangelo project: 3D scanning of large statues. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques* (pp. 131-144).
- Lowe, D. G. (1999, September). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision* (Vol. 2, pp. 1150-1157).
- Manteca Fernández, F. (2018). Generación de modelos 3D mediante luz estructurada.
- Newman, T. S., & Yi, H. (2006). A survey of the marching cubes algorithm. *Computers & Graphics*, 30(5), 854-879.
- Nießner, M., Zollhöfer, M., Izadi, S., & Stamminger, M. (2013). Real-time 3D reconstruction at scale using voxel hashing. *ACM Transactions on Graphics (ToG)*, 32(6), 1-11.
- Oyallon, E., & Rabin, J. (2015). An analysis of the SURF method. *Image Processing On Line*, 5, 176-218.
- Oriol Bermúdez, E., Algoritmos de reconstrucción de escenas 3D paralelizados en GPU para aplicaciones en tiempo real. 2010.
- Otegui Castillo, J., La realidad virtual y la realidad aumentada en el proceso de marketing. 2017.

- Ramakrishnan, S. K., Aaron Gokaslan, A., Wijmans, E., Oleksandr Maksymets, O., Clegg, A., Turner, J., Undersander, E., Galuba, W., Westbury, A., Chang, A. X., Savva, M., Zhao, Y. & Batra, D. (2021). Habitat-matterport 3D dataset (HM3D): 1000 large-scale 3D environments for embodied AI. *arXiv preprint arXiv:2109.08238*.
- Rodríguez, H. (28 de Abril de 2021). *crehana*. Obtenido de crehana: <https://www.crehana.com/blog/desarrollo-web/que-es-opencv/>
- Roth, H., & Vona, M. (2012, September). Moving Volume KinectFusion. In *BMVC* (Vol. 20, No. 2, pp. 1-11).
- Sastre, A. d. (3 de Junio de 2021). *cincodias.elpais*. Obtenido de [https://cincodias.elpais.com/cincodias/2021/06/03/smartphones/1622725656\\_659041.html](https://cincodias.elpais.com/cincodias/2021/06/03/smartphones/1622725656_659041.html)
- Slater, M. (2009). Place Illusion and Plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society of London*, 364, 3549–3557. <https://doi.org/10.3389/frobt.2016.00074>
- Soliman, M., Pesyridis, A., Dalaymani-Zad, D., Gronfula, M., & Kourmpetis, M. (2021). The application of virtual reality in engineering education. *Applied Sciences*, 11(6), 2879.
- Song, S., Lichtenberg, S. P., & Xiao, J. (2015). Sun rgb-d: A rgb-d scene understanding benchmark suite. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 567-576).
- Spark, A. (2015). Machine learning library (MLlib) guide. *Apache Spark*.
- Polvillo Díaz, M., Generación de datasets multi-sensor para mapeado y localización sin GPS de aeronaves no tripuladas. 2016.
- University, S. S. (s.f.). *rgb-d.cs.princeton.edu*. Obtenido de SUN RGB-D: A RGB-D Scene Understanding Benchmark Suite: <https://rgb-d.cs.princeton.edu/paper.pdf>