



**Universidad  
Zaragoza**

Trabajo Fin de Grado

## **Caracterización de sensores RGB-d para aplicaciones de metrología industrial**

English title:

RGB-d sensor characterization for industrial metrology applications

Autor:

**Diego Lahuerta Bordonada**

Directores:

Ana Cristina Murillo Arnal  
Eduardo Montijano Muñoz

Escuela de Ingeniería y Arquitectura

Zaragoza, Enero de 2018



## DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

(Este documento debe acompañar al Trabajo Fin de Grado (TFG)/Trabajo Fin de Máster (TFM) cuando sea depositado para su evaluación).

D./D<sup>a</sup>. Diego Lahuerta Bordonada,

con nº de DNI 76973737M en aplicación de lo dispuesto en el art.

14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster)  
Grado en Ingeniería Electrónica y Automática, (Título del Trabajo)  
Caracterización de sensores RGB-d para aplicaciones de metrología industrial

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, 20 de enero de 2018

Fdo: DIEGO LAHUERTA BORDONADA

## Resumen

El objetivo de este proyecto fin de grado consiste en la realización de un análisis formal sobre cámaras con información de profundidad (RGB-d) enfocado a su uso en aplicaciones de metrología industrial. Muchas tareas de medición se realizan con sensores de altísima precisión, pero también muy complejos y costosos. Es interesante saber que tipo de medidas se pueden realizar con este tipo de sensores RGB-d de manera fiable, ya que son mucho más baratos y sencillos de utilizar.

El proyecto se ha dividido en varias fases, desde el estudio de la calibración inicial de los sensores, hasta la calidad de sus medidas en estático y de los valores de las reconstrucciones 3D obtenidas:

En primer lugar, el análisis se ha centrado en un estudio de las diferentes tecnologías que utilizan este tipo de cámaras y la caracterización de sus parámetros intrínsecos mediante técnicas de calibración. Se han seleccionado varias cámaras que abarquen el conjunto de tecnologías descritas para realizar una evaluación experimental más detallada de las mismas.

En una segunda etapa, se ha diseñado una batería de pruebas con el objetivo de evaluar distintas variables de interés en el ámbito de la metrología. Tras la obtención de los datos necesarios para realizar el estudio, se ha dividido el análisis en dos bloques.

Por un lado, en el primer bloque se ha evaluado la repetibilidad que ofrece cada cámara, definida como la capacidad del sensor de proporcionar medidas constantes en las capturas bajo las mismas condiciones, cualidad ineludible de cualquier elemento utilizado en metrología. El estudio sobre la repetibilidad se ha dividido a su vez en la evaluación de la vivacidad de los píxeles que componen las imágenes de profundidad y la variabilidad en las medidas registradas de los mismos.

Por otro lado, el segundo bloque de este análisis ha consistido en analizar la calidad de los datos tridimensionales que se obtienen de las imágenes de profundidad. La medida de la calidad de las nubes de puntos (datos tridimensionales obtenidos de las imágenes de profundidad), se ha llevado a cabo mediante el ajuste de un plano a un conjunto de nubes de puntos de un escenario con un elemento central plano.

Los resultados obtenidos en este trabajo ponen de manifiesto los puntos fuertes y débiles de los sensores RGB-d en un contexto de aplicación diferente del entorno para el que han sido diseñados (interacción con ordenadores). Como conclusiones interesantes cabe destacar que pese a los precios similares, hay gran diferencia en la robustez de las medidas de los distintos sensores estudiados. Se ha comprobado que se pueden conseguir medidas de repetibilidad y precisión de hasta 0.4 mm al fotografiar un escenario con un elemento central plano para una distancia de posicionamiento de la cámara de 0.5 metros con respecto al escenario.

# Índice general

|   |           |
|---|-----------|
| <b>1. Introducción</b>  | <b>1</b>  |
| 1.1. Motivación y Contexto . . . . .                              | 1         |
| 1.2. Objetivos y alcance . . . . .                                | 3         |
| 1.3. Estructura de la memoria . . . . .                           | 3         |
| <b>2. Cámaras RGB-d y su Calibración</b>                          | <b>5</b>  |
| 2.1. Técnicas de adquisición de imágenes 3D . . . . .             | 5         |
| 2.2. Calibración . . . . .  | 10        |
| <b>3. Repetibilidad de los sensores</b>                           | <b>14</b> |
| 3.1. Definición y evaluación de repetibilidad . . . . .           | 14        |
| 3.2. Resultados . . . . .   | 18        |
| 3.3. Discusión . . . . .  | 23        |
| <b>4. Ajuste 3D</b>   | <b>25</b> |
| 4.1. De la imagen de profundidad a la nube de puntos . . . . .    | 25        |
| 4.2. Ajuste de un plano a la nube de puntos . . . . .             | 25        |
| 4.3. Resultados . . . . .   | 27        |
| <b>5. Conclusiones</b>  | <b>31</b> |
| 5.1. Discusión de resultados. . . . .                             | 31        |
| 5.2. Trabajo futuro . . . . .                                     | 32        |
| <b>Anexos</b>   | <b>32</b> |
| <b>A. Error de reproducción en la calibración de cámaras</b>      | <b>33</b> |
| <b>B. Nubes de puntos y ajuste del plano</b>                      | <b>35</b> |
| B.1. Del modelo de la cámara a la nube de puntos . . . . .        | 35        |
| B.2. Función <i>pcfitplane</i> de Matlab . . . . .                | 36        |
| B.3. Resultados: ajuste de un plano a la nube de puntos . . . . . | 37        |
| <b>Bibliografía</b>   | <b>41</b> |



# Capítulo 1

## Introducción

Este capítulo describe en que consiste este proyecto, así como la motivación y el contexto que han llevado a su realización.

### 1.1. Motivación y Contexto

En los últimos años han entrado en el mercado un tipo de sensores de bajo coste (coste medio de unos 300 €), capaces de registrar tanto la profundidad como el color de las escenas que fotografían. El principal enfoque de estos sensores está dirigido hacia campos como la realidad virtual o el control gestual. Sin embargo, existen otros campos en los que estos sensores pueden introducir mejoras sustanciales y en los que su implantación no se ha estudiado adecuadamente. Significa, por tanto, una oportunidad el poder evaluar cual es el grado de confiabilidad que nos dan estos sensores *low-cost* en el mundo de la Metrología Industrial, donde es común que un sensor láser para el escaneado 3D de objetos tenga un coste de varios miles de euros y una precisión y repetibilidad de micrómetros.

**Metrología Industrial.** Como define el Centro Español de Metrología (CEM), máximo órgano técnico de España en el campo de la Metrología, la Metrología Industrial es la rama dentro de la Metrología encargada de asegurar “*el adecuado funcionamiento de los instrumentos de medida empleados en la industria y en los procesos de producción y verificación.*”

Por tanto, la Metrología Industrial interviene en los procesos industriales, cobrando en ellos un importante papel en cualquier sistema de calidad aplicado a fabricación.

**Sensores utilizados en metrología industrial.** Para satisfacer la exigente demanda de la industria, en la actualidad se han desarrollado máquinas multifuncionales capaces de combinar múltiples sensores (sensores ópticos con sensores de contacto, Figura 1.1).

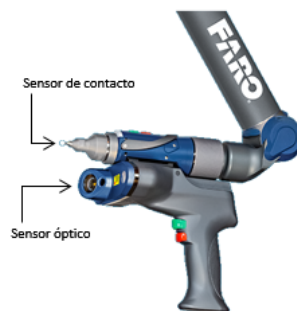


Figura 1.1: Ejemplo de sensor utilizado en aplicaciones industriales de metrología. Sensor FARO ScanArm. Imagen extraída de <http://factory-metrology.faro.com/es/analisis-dimensional/>

Los escáneres ópticos 3D utilizados actualmente en la industria son característicos por su alto coste y su precisión de medida. Hay un amplio rango de tecnologías que se utilizan para obtener medidas tridimensionales de una escena sin contacto. Por ejemplo, la Figura 1.2 muestra varios sistemas comerciales de medición 3D basados en distintas tecnologías.

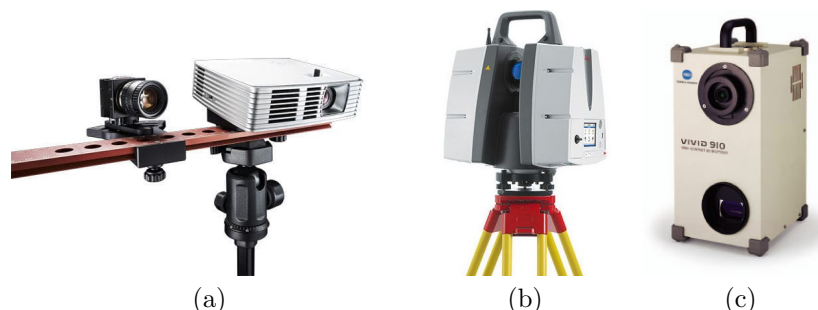


Figura 1.2: Ejemplos de tecnologías utilizadas para escaneo 3D, en entornos industriales, basadas en luz estructurada (a), tiempo de vuelo (b) o triangulación láser (c). Imágenes extraídas de <http://www.directindustry.es/>.

**Sensores 3D de bajo coste.** Los sensores de visión y profundidad (RGB-d, *Red Green Blue and Depth*) han sido desarrollados dentro del ámbito de los videojuegos, las interfaces gestuales y la realidad aumentada. Estos sensores presentan un enorme potencial en el campo de la Metrología Industrial de cara a la realización de modelados tridimensionales. Sin embargo, para poder utilizarlos en este contexto, hace falta una caracterización previa que permita determinar si estos sensores son capaces de ofrecer las precisiones requeridas actualmente en el ámbito de la metrología. En la Figura 1.3 podemos ver diferentes sensores RGB-d.



Figura 1.3: Ejemplos de sensores RGB-d de bajo coste: (a) Structure Sensor de Occipital (<https://structure.io/>), (b) ASUS Xtion Pro ([https://www.asus.com/es/3D-Sensor/Xtion\\_PRO/](https://www.asus.com/es/3D-Sensor/Xtion_PRO/)), (c) Kinect V2 (<https://developer.microsoft.com/en-us/windows/kinect/hardware>), (d) Orbbec Astra (<https://orbbec3d.com/product-astra/>).

Por lo general, estos dispositivos incluyen múltiples sensores, llegando a incorporar incluso micrófonos (como es el caso de la cámara Kinect V2). Para la realización de la caracterización de los sensores, solo serán ámbito de estudio el comportamiento que ofrecen los sensores involucrados en obtener la imagen de profundidad en las distintas cámaras.

Estos sensores han tenido una gran aceptación dentro de la comunidad científica dada su alta versatilidad para su desempeño en múltiples aplicaciones. Se pueden encontrar gran cantidad de estudios sobre la utilidad y evaluación de este tipo de sensores RGB-d de bajo coste para el modelado 3D de entornos interiores, [1, 2]. También se han realizado estudios acerca de los diferentes métodos de calibración utilizados para estas cámaras [3, 4]. En cambio, no se han encontrado estudios basados en la caracterización de las cámaras RGB-d con un enfoque hacia la Metrología Industrial.

**Contexto de realización del proyecto.** Este proyecto se desarrolla en colaboración con el grupo de Robótica, Percepción y Tiempo Real y el grupo de investigación de Manufactura y Metrología Avanzada de la EINA, los cuales han facilitado diferentes sensores RGB-d, medidas de referencia y orientado sobre

los parámetros que resultan de interés a la hora de analizar estos sistemas de medida de bajo coste. Los detalles de las tareas y objetivos se describen a continuación.

## 1.2. Objetivos y alcance

**Objetivos.** Para realizar la caracterización de los sensores será necesario estudiar el comportamiento que ofrecen desde un punto de vista formal. Evaluando la repetibilidad de los sensores a diferentes distancias, la calidad de los píxeles que componen las imágenes de profundidad que proporcionan los sensores y, por último, la calidad de las nubes de puntos que se extraen de las imágenes de profundidad proporcionadas. Las tareas a realizar en este proyecto han sido:

**Estudio de las tecnologías disponibles.** En primer lugar se han estudiado las diferentes tecnologías disponibles para la adquisición de datos 3D. Seguido de un estudio pormenorizado de las cámaras disponibles para realizar el estudio agrupándolas según la tecnología que utilizan.

**Calibración de las cámaras.** Para realizar la calibración de las cámaras se ha estudiado el significado de calibración de la cámara y algunas de las técnicas más comunes que se emplean para el cálculo de sus parámetros intrínsecos [5, 6]. Finalmente, se han obtenido los parámetros intrínsecos de las cámaras a través de la App Single Camera Calibration, disponible en la Toolbox de Matlab: *Computer Vision System Toolbox*.

**Captura de datos.** Se ha diseñado una batería de experimentos de cara a la caracterización de los diferentes sensores (estudio de repetibilidad y estudio de la calidad de los datos 3D). Para capturar los datos necesarios se han utilizado las herramientas disponibles en el laboratorio del grupo de robótica.

**Estudio de repetibilidad.** Para este estudio, se ha implementado en Matlab el código necesario para procesar las imágenes capturadas con cada cámara. Se han definido las métricas y calculado las estadísticas más representativas, y se ha realizado un análisis comparativo sobre el comportamiento registrado por cada cámara ante el mismo escenario.

**Estudio de la calidad de los datos 3D.** Se ha estudiado la metodología a seguir para la obtención de los puntos tridimensionales (que forman nubes de puntos) a partir de: el modelo de cámara utilizado, los parámetros intrínsecos obtenidos en la calibración y las imágenes de profundidad capturadas.

Se han obtenido nubes de puntos de una escena con un elemento central plano, de manera que se ha evaluado la reconstrucción de un mismo escenario a diferentes distancias y con distintas cámaras. El análisis se ha llevado a cabo mediante el ajuste de un plano a la mismas, estudiando así los parámetros de los diferentes planos ajustados. El desarrollo del estudio se ha llevado a cabo mediante la implementación de software en Matlab aprovechando las funcionalidades ya disponibles en la *Computer Vision System Toolbox* de Matlab.

## 1.3. Estructura de la memoria

La memoria consta de cinco Capítulos: Introducción, Cámaras RGB-d y su Calibración, Repetibilidad de los sensores, Ajuste 3D y Conclusiones.

**Introducción.** El Capítulo 1 describe la motivación y contexto del trabajo, seguido de un análisis de los objetivos y el alcance del mismo. Finalmente se resume brevemente la estructura de la memoria.

**Cámaras RGB-d y su Calibración.** El Capítulo 2 abarca las tareas referentes al estudio de las tecnologías disponibles y la calibración de las cámaras.

**Repetibilidad de los sensores.** En el Capítulo 3 se realiza un estudio sobre la repetibilidad que ofrecen los sensores, para resolver el estudio, se ha dividido en la evaluación de la vivacidad de los píxeles que componen las imágenes de profundidad y la variabilidad en las medidas registradas de los mismos.

**Ajuste 3D.** En el Capítulo 4 se realiza un análisis de las nubes de puntos mediante el ajuste de un plano a un escenario con un elemento central plano, se evalúan tanto la profundidad como la inclinación del plano ajustado dando una visión robusta acerca de la calidad de las nubes obtenidas a partir de las imágenes de profundidad capturadas con las cámaras.

**Conclusiones.** En el Capítulo 5 contiene una visión global del estudio realizado así como un resumen de los diferentes resultados obtenidos, por último se enumeran posibles trabajos futuros.

## Capítulo 2

# Cámaras RGB-d y su Calibración

Este capítulo describe las distintas técnicas de adquisición de imágenes tridimensionales así como las cámaras empleadas en el estudio agrupándolas dentro de las tecnologías que usan cada una. En segundo lugar se estudian las técnicas de calibración disponibles con el objetivo de obtener los parámetros intrínsecos de las cámaras de estudio.

### 2.1. Técnicas de adquisición de imágenes 3D

En la última década se ha producido un intenso desarrollo en las técnicas de adquisición de imágenes 3D. Esto es debido a que, si se tiene en cuenta la variabilidad y complejidad de las diferentes situaciones, no todos los sensores y técnicas pueden ser apropiados para cada tipo de aplicación. Por el momento se puede decir que, para todo tipo de objetos y escenarios, no hay una sola técnica de modelado 3D capaz de satisfacer todos los requisitos de alta precisión geométrica, portabilidad, automatización, bajo costo, flexibilidad y eficiencia.

La gran mayoría de los sensores actuales utilizados para la adquisición de imágenes 3D funcionan principalmente mediante dos formas. En primer lugar encontramos la forma activa, que actúa proyectando luz controlada sobre los objetos y grabando las respectivas reflexiones que se dan en el escenario, conteniendo esta información sobre la forma. El segundo modo de operación se conoce como pasivo, donde el punto de acción principal es la adquisición de energía de un objeto que está transmitiendo. En ambas formas, el resultado dependerá siempre del entorno donde actúan.

Para la obtención de medidas, el uso de métodos tecnológicos activos es el más adecuado, donde obtenemos como resultados datos de rango, creando relaciones con las distancias entre la superficie del objeto medido y el sensor. También se pueden obtener medidas indirectas basándose en el uso de imágenes monoculares o mediante el uso de conocimientos previos sobre las propiedades de las superficies [7].

Como se puede ver en la Tabla 2.1, extraída de [7], se puede observar una clasificación de las actuales técnicas disponibles para la adquisición de datos 3D mediante métodos ópticos.

|                          | Triangulación | Retardo tiempo | Imagen monocular | Pasivo | Activo | Directo | Indirecto | Rango | Orientación superficie |
|--------------------------|---------------|----------------|------------------|--------|--------|---------|-----------|-------|------------------------|
| Triangulación laser      | X             |                |                  |        | X      | X       |           | X     |                        |
| Luz estructurada         | X             |                |                  |        | X      | X       |           | X     |                        |
| Visión estéreo           | X             |                |                  | X      |        | X       |           | X     |                        |
| Fotogrametría            | X             |                |                  | X      |        | X       |           | X     |                        |
| Tiempo de vuelo          |               | X              |                  |        | X      | X       |           | X     |                        |
| Interferometría          |               | X              |                  |        | X      | X       |           | X     |                        |
| Interferometría de Moiré |               |                | X                |        | X      |         | X         | X     |                        |
| Forma desde el enfoque   |               |                | X                |        | X      |         | X         | X     |                        |
| Forma de sombras         |               |                | X                |        | X      |         | X         | X     |                        |
| Gradientes de textura    |               |                | X                |        |        |         | X         |       | X                      |
| Forma del sombreado      |               |                | X                |        | X      |         | X         |       | X                      |
| Forma desde fotometría   |               |                | X                |        | X      |         | X         |       | X                      |

Tabla 2.1: Clasificación de técnicas de captación de imágenes 3D.

En las siguientes secciones se presentan las dos principales tecnologías empleadas para obtención de las imágenes de rango, necesarias para registrar la profundidad por los sensores RGB-d de uso común, así como los diferentes sensores utilizados en el estudio, agrupados según la tecnología que usan.

### 2.1.1. Luz Estructurada

Los sensores de profundidad basados en la tecnología de luz estructurada (*Structured Light*, SL en adelante), proyectan patrones de luz para registrar la profundidad de la escena fotografiada. Los patrones de luz pueden ser lineales o de puntos (motas). Puesto que la estructura del patrón de luz proyectado es

conocida, el mapa de profundidad del objeto (o escenario) sobre el que se proyecta puede ser reconstruido en base a la disparidad que se obtiene entre la iluminación del patrón y los puntos observados.

Un sensor basado en la tecnología SL requiere de un emisor y un receptor de luz infrarroja. El análisis de las variaciones de intensidad obtenidas en cada punto de la imagen capturada por la cámara infrarroja permite obtener el registro de la distancia. Por tanto, un sensor basado en la tecnología de luz estructurada debe recorrer tres pasos fundamentales para la obtención de dicha distancia. Primero obtiene el infrarrojo, a continuación, la disparidad entre lo emitido y lo recibido y, finalmente, en base a esa disparidad estima la profundidad o distancia [3].

Las cámaras utilizadas en el trabajo que utilizan esta tecnología son: ASUS Xtion Pro, Orbbec Astra y Structure Sensor<sup>1</sup>.

**ASUS Xtion Pro.** Esta cámara, desarrollada por el fabricante ASUS (Figura 2.1), está enfocada a la detección de gestos para el desarrollo de aplicaciones interactivas con el movimiento. Contiene una cámara RGB junto con una cámara de profundidad para captar la información 3D del escenario que fotografía. En la Tabla 2.2 podemos encontrar un resumen de las principales especificaciones técnicas de la cámara extraídas de [8].



Figura 2.1: Detalle de la cámara ASUS Xtion Pro. Imagen extraída de: [https://www.asus.com/es/3D-Sensor/Xtion\\_PRO/](https://www.asus.com/es/3D-Sensor/Xtion_PRO/).

|  |                     |
|--|---------------------|
| <b>Dimensiones (mm)</b>                        | 180 × 35 × 50 mm    |
| <b>Campo de visión (°)</b>                     | 58 H, 45 V, 70 D    |
| <b>Resolución imagen RGB (píxeles)</b>         | (640 × 480) : 30fps |
| <b>Resolución imagen Profundidad (píxeles)</b> | (640 × 480) : 30fps |
| <b>Distancia de uso (m)</b>                    | Entre 0.8 y 3.5     |
| <b>Precio aproximado (€)</b>                   | 147                 |

Tabla 2.2: Especificaciones Técnicas: ASUS Xtion Pro

<sup>1</sup>Este sensor en un principio entraba dentro del estudio pero, tras los diferentes problemas que ha dado durante su calibración, se decide desecharlo del mismo.

**Orbbec Astra.** La cámara Orbbec Astra (Figura 2.2), desarrollada por la empresa Orbbec, está diseñada para trabajar sobre un amplio rango de escenarios, desde el control de gestos hasta el desarrollo de nubes de puntos. Al igual que la ASUS, la cámara contiene sensores RGB e infrarrojo. Aparte, esta cámara también contiene un micrófono integrado, dándole mayor versatilidad. En la Tabla 2.3 podemos encontrar un resumen de las principales especificaciones técnicas de la cámara extraídas de [9].



Figura 2.2: Detalle de la cámara Orbbec Astra. Imagen extraída de: <https://orbbec3d.com/product-astra/>.

|  |                     |
|--|---------------------|
| <b>Dimensiones (mm)</b>                        | 165 × 30 × 40       |
| <b>Campo de visión (°)</b>                     | 60 H, 49.5 V, 73 D  |
| <b>Resolución imagen RGB (píxeles)</b>         | (640 × 480) : 30fps |
| <b>Resolución imagen Profundidad (píxeles)</b> | (640 × 480) : 30fps |
| <b>Distancia de uso (m)</b>                    | Entre 0.6 y 8.0     |
| <b>Precio aproximado (€)</b>                   | 170                 |

Tabla 2.3: Especificaciones Técnicas: Orbbec Astra.

**Structure Sensor.** La cámara Structure Sensor (Figura 2.3) ha sido desarrollada por la empresa Occipital. A diferencia de las anteriores cámaras, solo posee un sensor infrarrojo. Este hecho no influye en el desarrollo del estudio puesto que solo se hacen uso de los sensores de profundidad. Esta cámara está dimensionada para ser acoplada a un iPad y utilizar la cámara RGB de la *tablet*. En la Tabla 2.4 podemos encontrar un resumen de las principales especificaciones técnicas de la cámara extraídas de [10].



Figura 2.3: Detalle de la cámara Structure Sensor acoplada a un iPad. Imagen extraída de: <https://structure.io/>.

|  |                     |
|--|---------------------|
| <b>Dimensiones (mm)</b>                        | 119.2 × 28 × 29     |
| <b>Campo de visión (°)</b>                     | 58 H, 45 V          |
| <b>Resolución imagen RGB (píxeles)</b>         | -                   |
| <b>Resolución imagen Profundidad (píxeles)</b> | (640 × 480) : 30fps |
| <b>Distancia de uso (m)</b>                    | Entre 0.4 y 3.5     |
| <b>Precio aproximado (€)</b>                   | 325                 |

Tabla 2.4: Especificaciones Técnicas: Structure Sensor



### 2.1.2. Tiempo de Vuelo

La tecnología basada en Tiempo de Vuelo (*Time-of-Flight*, ToF en adelante), estima la distancia radial desde un punto de la imagen al sensor mediante el estudio del desplazamiento de las ondas electromagnéticas en el aire. La radiación electromagnética viaja en el aire a la velocidad de la luz  $c \approx 3 \times 10^8$  [m/s]. Entonces, la distancia  $\rho$  cubierta en el tiempo  $\tau$  por una radiación óptica es  $\rho = c\tau$ .

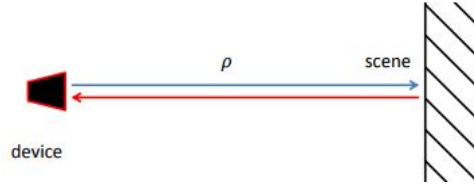


Figura 2.4: Principio de operación de las tecnologías que emplean el Tiempo de Vuelo para medir la profundidad. Imagen extraída de [11].

En la Figura 2.4 se muestra el esquema de medida típico para un sistema ToF. En tiempo 0 se transmite radiación desde el sensor, la cual viaja una distancia  $\rho$  hasta la escena. La radiación es reflejada en la escena y vuelve a recorrer la distancia  $\rho$  hasta el receptor que contiene el sensor. Si el tiempo empleado se denomina  $\tau$ , tenemos que para un tiempo  $\tau$  se ha recorrido una distancia igual a  $2\rho$ . Por tanto, la relación en este caso entre  $\rho$  y  $\tau$  es [11]

$$\rho = \frac{c\tau}{2}. \quad (2.1)$$

La cámara considerada en este trabajo que implementa esta tecnología es el sensor Kinect V2, de Microsoft.

**Kinect V2.** La cámara Kinect V2, desarrollada por Microsoft, representa la segunda generación de cámaras RGB-D desarrolladas por esta compañía tras el lanzamiento de la primera Kinect V1 en noviembre de 2010. El sensor se compone de una cámara RGB y una infrarroja, y la iluminación activa de la escena observada es proporcionada por tres proyectores infrarrojos. En la Figura 2.5 se puede ver una imagen de la cámara [4]. Aparte de las dos cámaras, el sensor dispone de cuatro micrófonos para captar el sonido, grabar audio, y poder localizar la fuente de sonido así como la dirección de la onda de audio. En la Tabla 2.5 podemos encontrar un resumen de las principales especificaciones técnicas de la cámara extraídas de [12].



Figura 2.5: Detalle de la cámara Kinect V2. Imagen extraída de <https://developer.microsoft.com/en-us/windows/kinect/hardware>.

|   |                       |
|---|-----------------------|
| Dimensiones (mm)                        | 249 × 66 × 67         |
| Campo de visión (°)                     | 60 H, 70 V            |
| Resolución imagen RGB (píxeles)         | (1920 × 1080) : 30fps |
| Resolución imagen Profundidad (píxeles) | (512 × 424) : 30fps   |
| Distancia de uso (m)                    | Entre 0.5 y 4.5       |
| Precio aproximado (€)                   | 90                    |

Tabla 2.5: Especificaciones Técnicas: Kinect V2

## 2.2. Calibración de las cámaras RGB-d

Una vez estudiadas las cámaras que se van a evaluar y las tecnologías que estas emplean para registrar las distancias de la escena que fotografían con el sensor de profundidad que contienen, el siguiente paso es la calibración de las mismas.

En cualquier proceso desarrollado dentro de la visión por computador tridimensional, como es el presente estudio, la calibración de las cámaras empleadas resulta un paso necesario de cara a la obtención de información métrica desde imágenes 2D [6]. En la presente caracterización de las cámaras para su uso en aplicaciones de metrología, la calibración de las mismas nos servirán para generar las nubes de puntos finales desarrolladas en el Capítulo 4.

La calibración de la cámara comprenderá el estudio del modelo de la cámara, la obtención de las capturas necesarias para calibrar el sensor de profundidad, la metodología a implementar y la obtención final de los parámetros intrínsecos de las cámaras.

### 2.2.1. Modelo de la cámara

El principal concepto de calibración está basado en el modelo de cámara estenopeica o *pin-hole*. Una cámara *pin-hole* es una cámara simple sin lente y con una sola abertura pequeña (Figura 2.6). Los rayos de luz pasan a través de la abertura y proyectan una imagen invertida en el lado opuesto de la cámara [6, 5].

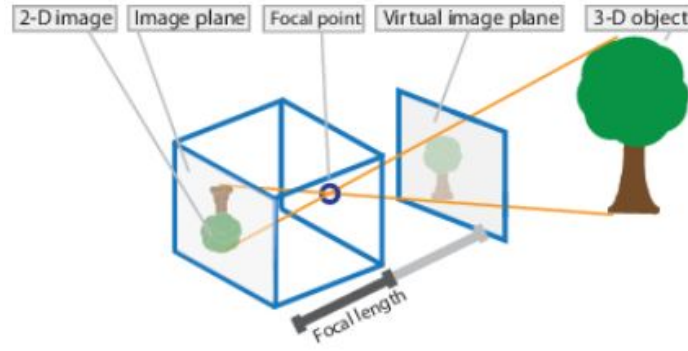


Figura 2.6: Esquema de un modelo de cámara *pin-hole*. Imagen extraída de [5].

Para la obtención de los parámetros intrínsecos de la cámara se ha utilizado la aplicación *Single Camera Calibration* de Matlab. La aplicación utiliza un algoritmo de calibración definido por Bouguet en [13]. Este algoritmo calcula los parámetros extrínsecos e intrínsecos de la matriz de la cámara. Siendo la matriz de la cámara

$$P = \begin{bmatrix} \mathbf{R} \\ \mathbf{t} \end{bmatrix} \mathbf{K}. \quad (2.2)$$

Los parámetros extrínsecos de la cámara consisten en una rotación,  $\mathbf{R}$ , y una traslación,  $\mathbf{t}$ , representando así la transformación rígida del sistema de coordenadas 3D del mundo al sistema de coordenadas 3D de la cámara. Los parámetros intrínsecos a su vez representan la transformación proyectiva que se realiza desde las coordenadas 3D de la cámara a las coordenadas 2D de la imagen.

Resulta entonces necesario conocer los parámetros recogidos en la matriz  $\mathbf{K}$ , siendo esta la matriz intrínseca de la cámara:

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & 0 \\ s & f_y & 0 \\ c_x & c_y & 1 \end{bmatrix}, \quad (2.3)$$

con  $f_x$  y  $f_y$  la distancia focal (entre el punto principal de la lente y el plano de la imagen) en píxeles, y  $c_x$  y  $c_y$  las coordenadas (en píxeles) del punto principal de la lente. Aparte, la inclinación del píxel:  $s$ , que resulta 0 si los ejes de la imagen son perpendiculares [5].

Una vez definidos los elementos constituyentes de la matriz de la cámara se inicia el calibrado de la misma. El proceso utilizado, resumido en la Figura 2.7, consta de cuatro pasos englobados dentro de dos fases:

1. Preparación (patrón y sensores) y captura de datos.
2. Calibración de la cámara (obtención de los parámetros intrínsecos).

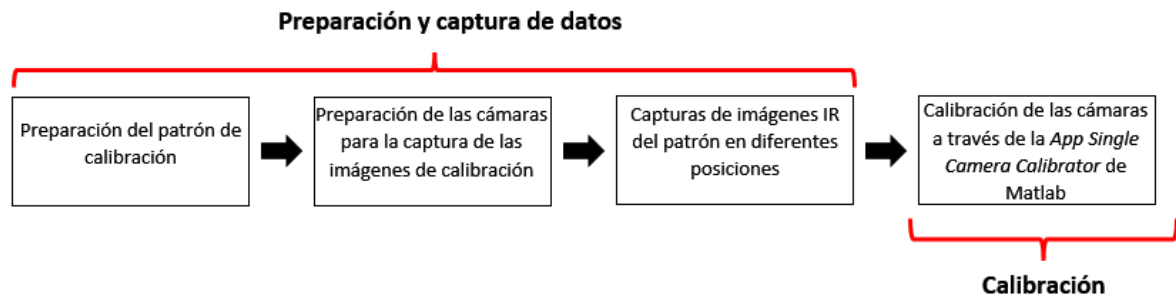


Figura 2.7: Proceso seguido para la calibración de las cámaras.

En un principio se procede a calibrar los cuatro sensores antes mencionados: ASUS Xtion Pro, Orbbec Astra, Kinect V2 y Structure Sensor. Debido a la gran cantidad de problemas encontrados utilizando el sensor Structure y, puesto que con los otros sensores ya se abarcan las dos tecnologías de captación de imágenes de profundidad, en el TFG se decide no realizar el análisis empírico de esta cámara. Por tanto, se realiza la calibración de los sensores: ASUS Xtion Pro, Orbbec Astra y Kinect V2.

### 2.2.2. Preparación y captura de datos

Primero se ha preparado un patrón de calibración, de  $648 \times 504$  mm. Un patrón de calibración consiste en un tablero de ajedrez rectangular (con un ancho de cuadrado escogido de 72 mm). Una vez listo, se han puesto a punto las cámaras para la sesión de capturas del patrón.

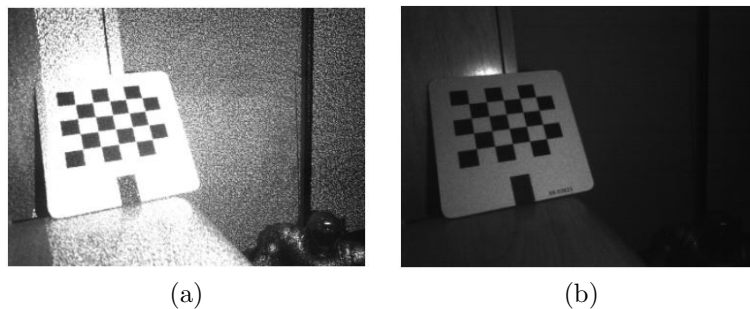


Figura 2.8: (a) Imagen capturada con la cámara infrarroja (IR) por defecto y (b) misma captura obstruyendo el emisor IR. (Imágenes extraídas de [http://wiki.ros.org/openni\\_launch/Tutorials/IntrinsicCalibration](http://wiki.ros.org/openni_launch/Tutorials/IntrinsicCalibration)).

Para poner a punto las cámaras antes de capturar las imágenes que se van a utilizar en la calibración se han tapado los emisores infrarrojos de las cámaras del patrón. Sobre todo en las cámaras que utilizan la tecnología *Structured Light* debido al patrón que proyectan (comentado en la sección 2.1.1). Obtener imágenes infrarrojas con una calidad óptima resulta de vital interés para el buen funcionamiento de la aplicación de calibración. Se puede ver en detalle como afecta la obstrucción del emisor infrarrojo en la calidad de la imagen infrarroja capturada por el sensor en la Figura 2.8.

Se han tomado entre 30 y 70 capturas del patrón en distintas posiciones (y orientaciones), con respecto el plano de la cámara. Durante la calibración, la cámara se ha encontrado en una posición estática en todo

momento. En todas las capturas se ha procurado que el patrón encontrara a una distancia aproximada de 1,5 metros con respecto de la cámara. En la Figura 2.9 se puede ver una reconstrucción de las distintas posiciones del patrón con respecto de la cámara.

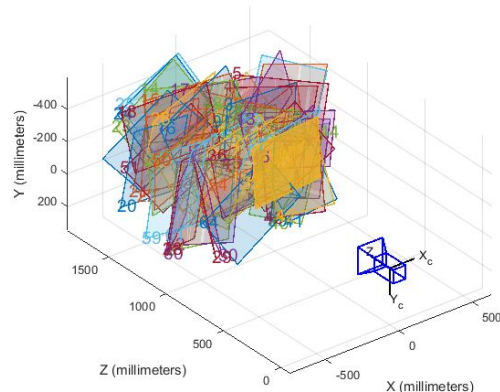


Figura 2.9: Diferentes posiciones del patrón con respecto de la cámara. Imagen extraída durante el proceso de calibración de la cámara ASUS Xtion Pro, *App Single Camera Calibration* de Matlab.

### 2.2.3. Tratamiento de los datos

Una vez se han recogido las diferentes imágenes, se han tratado con la aplicación *Single Camera Calibrator* de Matlab. El proceso de calibración realizado se puede resumir en el diagrama de flujo de la Figura 2.10.

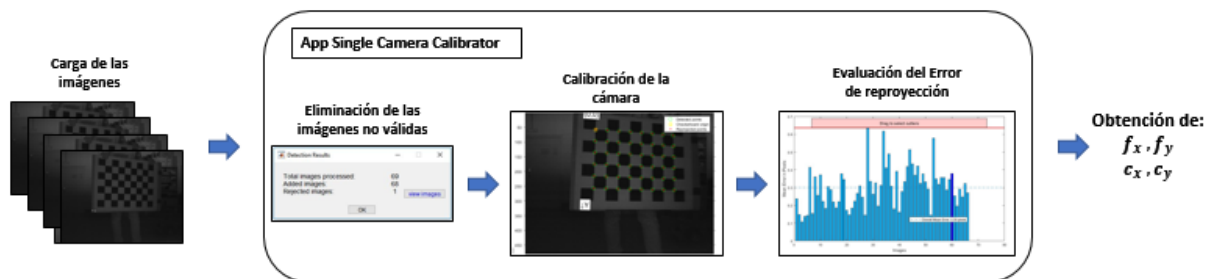


Figura 2.10: Detalle del proceso de calibración realizado con la *App Single Camera Calibration* de Matlab.

En el proceso de cargar las imágenes, la aplicación ha desechado aquellas en las cuales no ha conseguido reconocer el patrón. Esto es debido en su mayoría a una orientación relativa del patrón con respecto de la cámara de más de 45°. Con las imágenes válidas ya cargadas, se obtiene una primera calibración. La calidad de la calibración se evalúa mediante el error de reproyección. El error de reproyección medio obtenido para las tres cámaras resulta igual o menor de 0.30 píxeles, por lo que no se ha visto necesario realizar una segunda calibración de las cámaras.

Los diferentes errores de reproyección se pueden observar en el Anexo A. En dicho anexo, se pueden ver más detalles de estos resultados gráficos obtenidos para la calibración de cada una de las cámaras, junto con una explicación más detallada del significado de este error de reproyección en el proceso de calibración.

**Resultados de Calibración obtenidos.** Por último, los parámetros intrínsecos obtenidos de los tres sensores tras haber realizado la calibración se encuentran resumidos en la Tabla 2.6.

|                       | $f_x$    | $f_y$    | $c_x$    | $c_y$    |
|-----------------------|----------|----------|----------|----------|
| <b>Orbbec Astra</b>   | 568.3750 | 568.7463 | 317.0851 | 252.1317 |
| <b>Kinect V2</b>      | 361.6219 | 360.7421 | 251.1073 | 197.4387 |
| <b>ASUS Xtion Pro</b> | 564.9532 | 564.9737 | 319.1705 | 242.347  |

Tabla 2.6: **Parámetros intrínsecos** calculados durante la calibración de las cámaras (en píxeles):  $f_x$  y  $f_y$  representan las distancias focales,  $c_x$  y  $c_y$  son las coordenadas del punto principal.

Para las cámaras ASUS Xtion Pro y Orbbec Astra, las coordenadas esperadas para el punto principal teniendo en cuenta que la resolución de la imagen de profundidad que capturan es  $640 \times 480$  son (320, 240), mientras que para la Kinect V2, con resolución de la imagen de profundidad igual a  $512 \times 424$ , esperamos unas coordenadas para el punto principal entorno a (256, 212).

Teniendo en cuenta que los resultados obtenidos se aproximan a los valores ideales descritos dependiendo de la resolución de las cámaras y, atendiendo al significado geométrico de los parámetros obtenidos, que las distancias focales sean de similar longitud y los puntos principales se encuentren situados cerca del centro de la imagen, damos por válidas la calibraciones de las cámaras. Estos parámetros serán utilizados en las construcciones de las nubes de puntos del Capítulo 4.

## Capítulo 3

# Repetibilidad de los sensores

En este capítulo se realiza un estudio riguroso acerca de la repetibilidad que ofrecen los sensores que abarca el estudio. Se estudian los sensores considerados por separado, finalizando con una comparativa entre ellos dando una visión global acerca de la calidad de los mismos enfocados a la Metrología Industrial.

### 3.1. Definición y evaluación de repetibilidad

Dado un instrumento de medida, la repetibilidad que ofrece es uno de los parámetros más importantes a definir, ya que da cuenta acerca de la precisión que ofrece el instrumento. Por tanto, el estudio de la repetibilidad de los sensores es un estudio sobre el hardware de las cámaras.

En esta parte, se ha estudiado el tipo de respuesta que ofrecen los píxeles que conforman la imagen de profundidad proporcionada por cada cámara ante un mismo escenario que contiene como elemento central un plano, de manera que se pueda evaluar de forma rigurosa y en las mismas condiciones todos los píxeles de la imagen. El proceso llevado a cabo para realizar el estudio se encuentra resumido en el flujograma definido en la Figura 3.1.

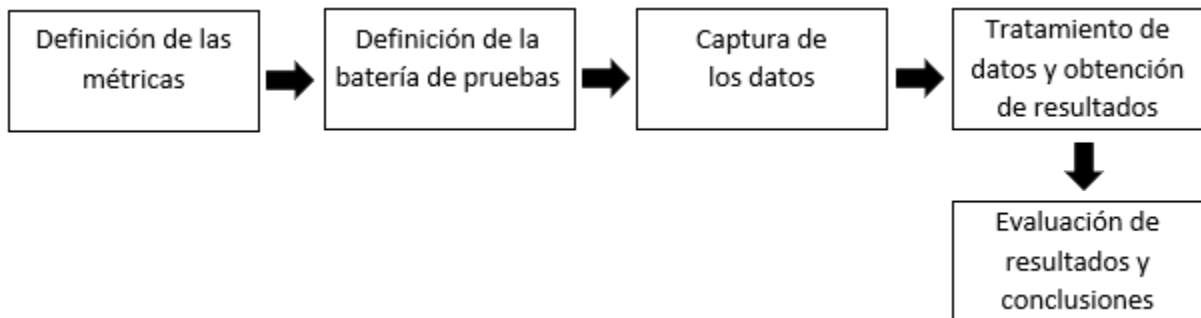


Figura 3.1: Flujograma del estudio de repetibilidad de las cámaras.

**Definición de las métricas.** Para realizar el estudio de la repetibilidad de las cámaras es necesario conocer la variabilidad de las medidas que han registrado los píxeles que conforman las imágenes de profundidad capturadas por las cámaras. Esta variabilidad viene definida mediante la desviación estándar que ofrece cada píxel al capturar un número determinado de medidas.

Durante el tratamiento de los datos se ha visto que los píxeles que conforman la imagen de profundidad no siempre registran un valor de medida válido, esto conlleva que, junto con la variabilidad de los píxeles sea necesario estudiar la vivacidad de los mismos. La vivacidad de los píxeles corresponde a determinar si los píxeles contenidos en las imágenes de profundidad tienen un comportamiento estable o no (entendiendo que un píxel mantiene un comportamiento estable cuando es capaz de medir distancia a lo largo de varias imágenes), independientemente de la medida que registren.

**Batería de pruebas.** Se ha definido un *setup* de pruebas donde el elemento central es un plano, para que todos los píxeles tengan el mismo valor ante una determinada distancia y poder así ejercer una comparación “justa” en la respuesta que ofrecen. Este plano lo formará una puerta negra a la que se ha pegado también el patrón de calibración utilizado en el Capítulo 2, la escena del plano contiene así diferentes texturas a partir de las cuales podremos deducir si el comportamiento que ofrecen las cámaras es dependiente de la textura de la escena que fotografían.

Tanto para la vivacidad como para la repetibilidad se han tomado 100 capturas iguales con las diferentes cámaras en estático (colocadas sobre un trípode) a diferentes distancias. El conjunto de las distancias (denominadas desde ahora en adelante como  $D$ ) utilizadas son: 0.5, 1.0, 1.5, 2.0, 2.5, y 3.0 metros. Obteniendo así un total de 600 imágenes de profundidad divididas en 6 tramos con las que poder analizar la calidad que ofrecen cada cámara en función de la distancia relativa al objeto. En la Figura 3.2, se puede ver en detalle el escenario utilizado para las pruebas.



Figura 3.2: Detalle de la estación de calibración de las cámaras. Además de la cámara en el trípode, se ve el plano dominante (puerta) con el patrón utilizado para la calibración de las cámaras, y las marcas en el suelo correspondientes a las diferentes distancias relativas con respecto al plano dominante en la escena.

**Captura de datos.** Con la batería de pruebas definida y montada (cámara sobre trípode y diferentes medidas plano-cámara en el suelo), se han capturado las imágenes a través de un software proporcionado por el grupo de investigación en el laboratorio del grupo de Robótica, Percepción y Tiempo Real. Entre las tres cámaras, se han capturado un total de 1800 imágenes de profundidad.

Una vez obtenidas todas las imágenes de profundidad necesarias para el análisis, éstas han sido tratadas a través de Matlab.

**Tratamiento de los datos.** Se ha desarrollado el software necesario para la evaluación tanto de la vivacidad de las cámaras como de la repetibilidad. Estos procesos se detallan a continuación.

### 3.1.1. Análisis de la vivacidad

El proceso de análisis de la vivacidad de los píxeles que conforman las imágenes de profundidad que capturan las cámaras se encuentra resumido en la Figura 3.3.

Este procedimiento se ha realizado tomando medidas con la cámara situada a distintas distancias del plano dominante de la escena: 1.0, 2.0, y 3.0 metros. Con esto se ha obtenido el comportamiento que ofrecen las diferentes cámaras en cuanto a la vivacidad de sus píxeles para distancias que se podrían considerar como corta, media y larga.

En primer lugar se han cargado las imágenes en el programa para poder trabajar con ellas. Una vez cargadas las 100 imágenes correspondientes a una distancia relativa de las consideradas en el estudio, se han recorrido todos los píxeles de todas las imágenes contabilizando si en el píxel se encontraba una medida válida (medida distinta de cero) o no. Con los valores obtenidos del conteo, que pueden ir de 0, el píxel está “muerto”, pues en ninguna de las 100 imágenes ha sido capaz de registrar una medida válida, a 100, el píxel en cuestión tiene una efectividad del 100 %, se ha construido la matriz de vivacidad: una

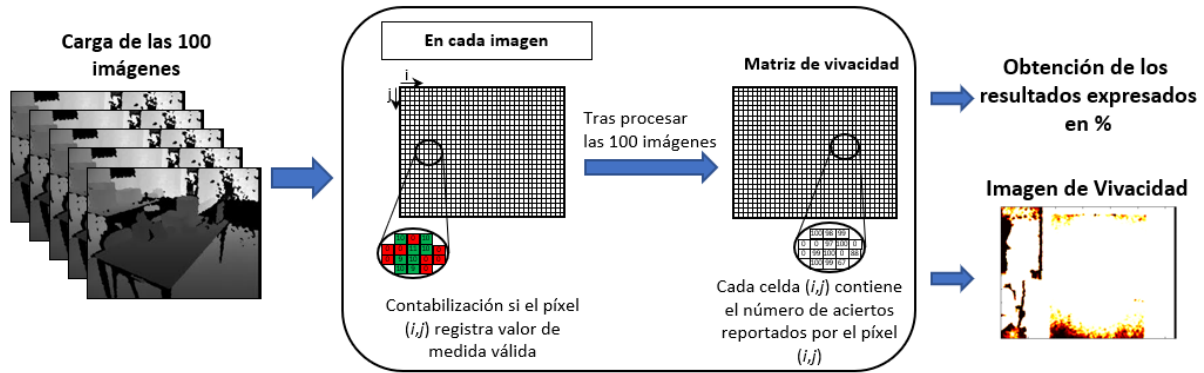


Figura 3.3: Metodología empleada para el análisis de la vivacidad de las cámaras.

matriz donde se ha alojado en cada posición  $(i, j)$  el conteo realizado sobre ese píxel  $(i, j)$ . Por último, de la matriz de vivacidad resultado del conteo, se han obtenido para cada cámara y cada distancia relativa:

- Una imagen de vivacidad, que consiste en un *heatmap* de la matriz de vivacidad con un *colorbar* de 0 a 100, valores correspondidos con el conteo previamente realizado.
- Una tabla de vivacidad donde, para las tres distancias consideradas se ha mostrado el porcentaje de píxeles sobre el total que conforman la imagen de profundidad que han tenido una tasa de aciertos (medida válida) del: 0 %, entre 0 y 25 %, entre 25 y 75 %, más del 75 %, y, por último, del 100 %.

### 3.1.2. Análisis de la repetibilidad

El análisis de la repetibilidad que ofrecen las cámaras se encuentra definido en la Figura 3.4. El procedimiento se ha utilizado para las tres cámaras, evaluando en este caso las seis distancias cámara-plano: 0.5, 1.0, 1.5, 2.0, 2.5, y 3.0 metros.

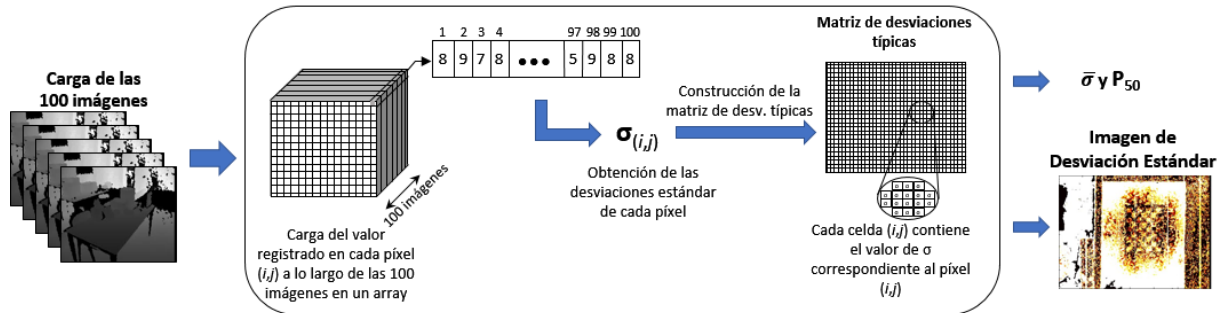


Figura 3.4: Metodología empleada para el análisis de la repetibilidad de las cámaras.

Al igual que para el análisis de la vivacidad el primer paso consisten en cargar las imágenes de profundidad que se van a tratar y establecer un recorrido por todos los píxeles a lo largo de ellas. Para estudiar la variabilidad de los píxeles que conforman las imágenes de profundidad en cuanto a las medidas que registran será necesario, primero, cribar los píxeles que no consigan registrar ninguna medida válida, esto es debido a que no nos interesa conocer la desviación que ofrece un píxel que está "muerto" (el píxel siempre ofrecerá la misma respuesta). Estos píxeles que siempre reporten medida cero se considerará que tienen una desviación estándar a lo largo de sus medidas de valor igual a 0. Una vez se han recorrido todos los píxeles a lo largo de todas las imágenes, se ha construido la matriz de desviaciones típicas, donde en cada posición  $(i, j)$  se aloja el valor de desviación estándar obtenido para el mismo píxel  $(i, j)$  (resultado obtenido al recorrer las cien imágenes). Por último, a partir de las diferentes matrices de desviación estándar calculadas para cada distancia relativa por cada cámara, se han obtenido:



- Una imagen de desviación estándar: para las distancias de 1.0, 2.0 y 3.0 metros. Consisten en 3 *heatmaps* que dan cuenta de una forma visual del comportamiento variable de los píxeles que conforman las imágenes, el rango de colores del *colorbar* va de 0 a 6 mm para las cámaras ASUS Xtion Pro y Orbbec Astra y de 0 a 9 mm para la Kinect V2.
- Una tabla de estadísticos de desviación estándar para las seis distancias relativas cámara-plano. Se han calculado la media y la mediana de los valores contenidos en la matriz de desviaciones típicas omitiendo las celdas  $(i, j)$  de la matriz que contuviera como valor un 0. De esta manera, se ha analizado la variabilidad que ofrecen los píxeles de cada cámara para cada distancia y su comportamiento al aumentarla.

## 3.2. Resultados

Siguiendo el proceso descrito en la Sección anterior (Figura 3.1), se han evaluado los resultados obtenidos para las distintas cámaras de manera individual y, por último, se han extraído conclusiones comparativas entre las tres cámaras.

### 3.2.1. ASUS XTION PRO

En la Tabla 3.1 se encuentran los resultados obtenidos en el análisis de la vivacidad de los píxeles que componen las imágenes de profundidad. Observando los porcentajes de píxeles tanto “muertos” como con tasa de acierto del 100 % podemos afirmar que esta cámara ofrece unos resultados estables a medida que se aleja de la escena del plano, siendo el porcentaje de píxeles con tasa de acierto del 100 % para la distancia relativa de 3m igual al 80 %.

| ASUS XTION PRO   |     |       |         |       |       |
|--|-----|-------|---------|-------|-------|
| Porcentaje de píxeles sobre el total de la imagen<br>con ratio de acierto/medida válida del: |     |       |         |       |       |
| $D$  | 0 % | <25 % | 25-75 % | >75 % | 100 % |
| 1m   | 3 % | 1 %   | 1 %     | 2 %   | 93 %  |
| 2m   | 4 % | 4 %   | 5 %     | 8 %   | 80 %  |
| 3m   | 6 % | 5 %   | 4 %     | 5 %   | 80 %  |

Tabla 3.1: Resultados de vivacidad expresados como tantos por ciento de los píxeles totales que conforman las imágenes de profundidad de la cámara ASUS Xtion Pro. Cada imagen en este experimento tiene 307200 ( $480 \times 640$ ) píxeles en total.

En la Tabla 3.2, se encuentran tanto la desviación típica media como la mediana de las matrices de desviaciones típicas obtenidas durante el análisis de la repetibilidad. Los resultados obtenidos dan cuenta de una variabilidad estable con el incremento de la distancia y la proximidad del valor de la mediana a la desviación típica media es un buen indicador de que no hay medidas puntuales con una alta desviación en la medida.

| ASUS XTION PRO |                     |               |
|----------------|---------------------|---------------|
| $D$ [m]        | $\bar{\sigma}$ [mm] | $P_{50}$ [mm] |
| 0.5            | 0.39                | 0.41          |
| 1.0            | 1.47                | 1.45          |
| 1.5            | 5.06                | 3.43          |
| 2.0            | 15.79               | 6.17          |
| 2.5            | 21.68               | 8.88          |
| 3.0            | 26.24               | 12.31         |

Tabla 3.2: Resultados de variabilidad obtenidos para la cámara ASUS Xtion Pro, para cada distancia de estudio: desviación típica media y percentil 50 extraídas de la matriz de desviaciones típicas calculada.

En la Figura 3.5, se puede observar como el 3 % de píxeles con tasa de acierto/medida válida del 0 % para la distancia cámara-plano de 1m, se encuentran en su mayoría en los laterales izquierdo y derecho de la imagen, siendo más pronunciado el lado izquierdo (Fig. 3.5 (a)), aparte, se puede observar siguiendo las imágenes de vivacidad como la parte derecha de la imagen contiene una franja con píxeles “muertos” que se mantiene constante a pesar de capturar diferentes escenarios (aumento de la distancia), mientras que la franja de la parte izquierda (muy pronunciada a la distancia de 1 metro), desaparece para 2 y 3 metros. Este hecho, junto con la aparición de píxeles que no son capaces de registrar medidas en las

partes inferiores de la puerta (por debajo de la franja blanca) observables en las imágenes de vivacidad (b) y (c) de la Figura 3.5 indican los problemas del sensor para capturar medidas cuando la escena contiene cambios bruscos en las texturas de los materiales (zonas amplias de la imagen que pasan de blanco a negro). También se puede observar el decremento del porcentaje de píxeles que tienen tasa de acierto del 100 % a medida que se aumenta la distancia relativa cámara-plano. Esto es debido en gran medida a la escena que capta la cámara, fotografiándose objetos con texturas que dificultan que los píxeles involucrados en esa zona puedan medir distancias.

También es necesario resaltar cómo afecta la textura de los objetos en la imagen a la calidad de las medidas registradas por estos. En la Figura 3.5 (b) y (c), se visualiza un mejor comportamiento en cuanto a variabilidad de medida de distancia de los píxeles que miden la pared blanca con respecto los píxeles que miden la puerta negra, ambos elementos de diferentes texturas. Predominando el color negro (variabilidad baja) en la pared y el color blanco (variabilidad igual o mayor que el límite) en la puerta.

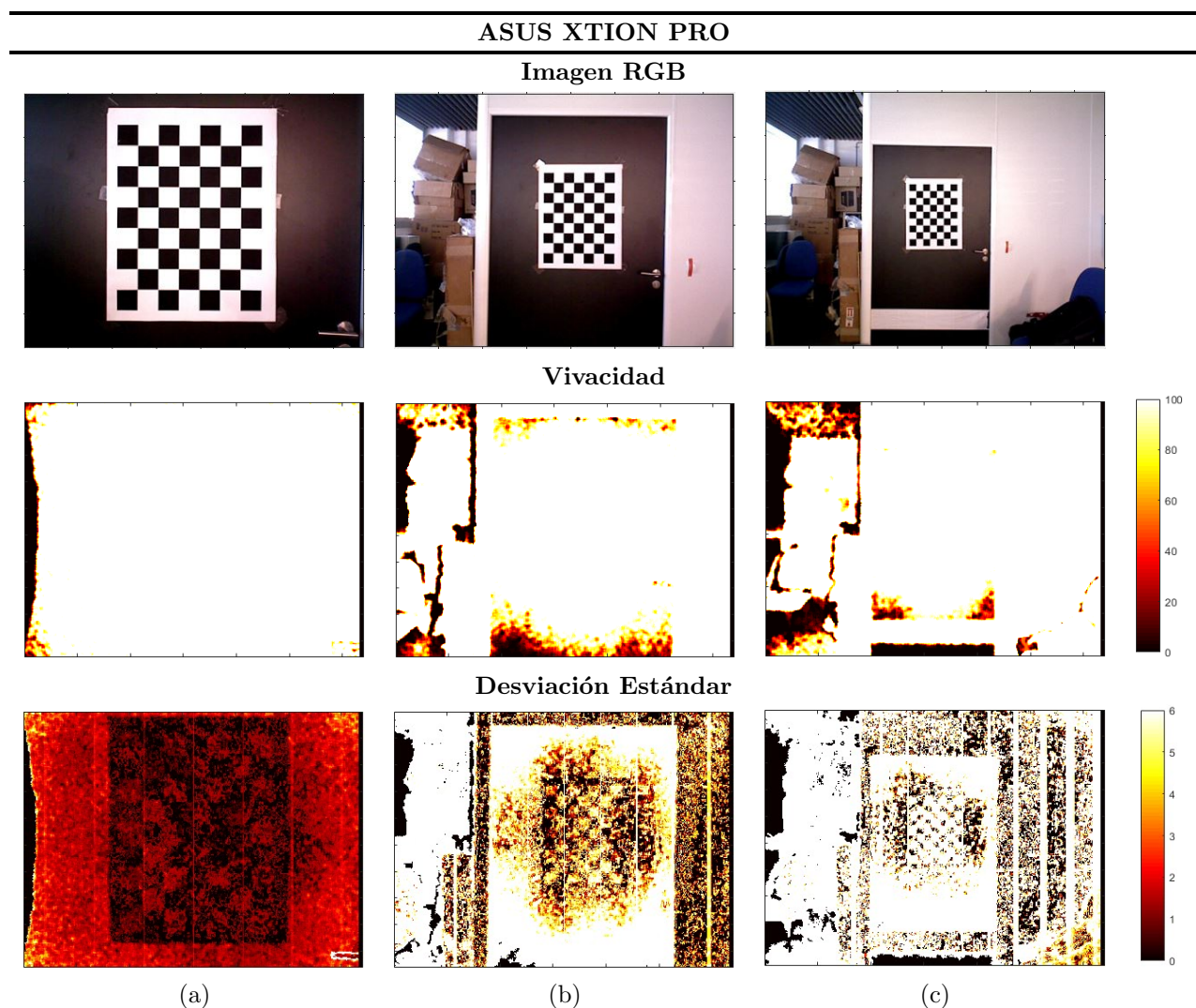


Figura 3.5: Resultados de repetibilidad por píxel en la cámara ASUS Xtion Pro, tomando una imagen a distintas distancias de la puerta (elemento central de la escena): 1m (a), 2m (b), y 3m (c). De arriba a abajo, imagen RGB, *heatmap* de vivacidad (frecuencia de medida correcta en cada píxel) y *heatmap* de la desviación estándar (en mm) de las medidas en cada píxel en estático.

### 3.2.2. ORBBEC ASTRA

Los resultados del análisis de vivacidad de los píxeles de esta cámara se muestran en la Tabla 3.3. Esta cámara ofrece unos porcentajes de píxeles que dan una tasa de acierto del 100 % bastante más moderados en comparación con la anterior, siendo del 81 % para distancias cortas (1m) y del 66 % para distancias largas (3m).

| ORBBEC ASTRA   |      |       |         |       |       |
|--|------|-------|---------|-------|-------|
| Porcentaje de píxeles sobre el total de la imagen<br>con ratio de acierto/medida válida del: |      |       |         |       |       |
| $D$  | 0 %  | <25 % | 25-75 % | >75 % | 100 % |
| 1m   | 5 %  | 1 %   | 1 %     | 12 %  | 81 %  |
| 2m   | 14 % | 10 %  | 6 %     | 9 %   | 62 %  |
| 3m   | 17 % | 6 %   | 4 %     | 6 %   | 66 %  |

Tabla 3.3: Resultados de vivacidad expresados como tantos por ciento de los píxeles totales que conforman las imágenes de profundidad de la cámara Orbbec Astra. Cada imagen en este experimento tiene 307200 ( $480 \times 640$ ) píxeles en total.

Visualizando las imágenes de vivacidad de la Figura 3.6, se observa una dependencia fuerte del comportamiento que ofrecen los píxeles a la textura del objeto que fotografían. Teniendo la cámara problemas a la hora de registrar medidas de profundidad en las zonas donde se produce un cambio brusco de texturas (Figura 3.6 (b), zonas limítrofes con el marco de la puerta).

Esta cámara también contiene un porcentaje de píxeles “muertos” que no son dependientes de la escena, habiendo una franja negra de carácter pronunciado en la zona derecha en todas las imágenes de vivacidad de la Figura 3.6.

Los resultados obtenidos en cuanto a la variabilidad de los píxeles, mostrados en la Tabla 3.4, son los mejores, comparados con las otras dos cámaras, por lo que podemos decir que los píxeles que si consiguen registrar medidas en la imagen contienen una variabilidad media muy baja, siendo de 20 mm con mediana de 12 mm para la distancia cámara-plano máxima de estudio de 3 metros.

| ORBBEC ASTRA |                     |               |
|--------------|---------------------|---------------|
| $D$ [m]      | $\bar{\sigma}$ [mm] | $P_{50}$ [mm] |
| 0.5          | 0.40                | 0.41          |
| 1.0          | 4.42                | 1.43          |
| 1.5          | 5.57                | 3.77          |
| 2.0          | 11.15               | 6.01          |
| 2.5          | 15.23               | 8.44          |
| 3.0          | 19.46               | 11.71         |

Tabla 3.4: Resultados de variabilidad obtenidos para la cámara Orbbec Astra, para cada distancia de estudio: desviación típica media y percentil 50 extraídas de la matriz de desviaciones típicas calculada. Todos los valores están expresados en mm.

Al igual que sucede con la cámara ASUS Xtion Pro, que utiliza la misma tecnología para registrar las distancias de la escena, esta cámara no es capaz de registrar la distancia a la que se encuentran las ventanas de cristal, se puede observar el comportamiento que ofrece la cámara ante este material en las imágenes (b) y (c) de la Figura 3.6.

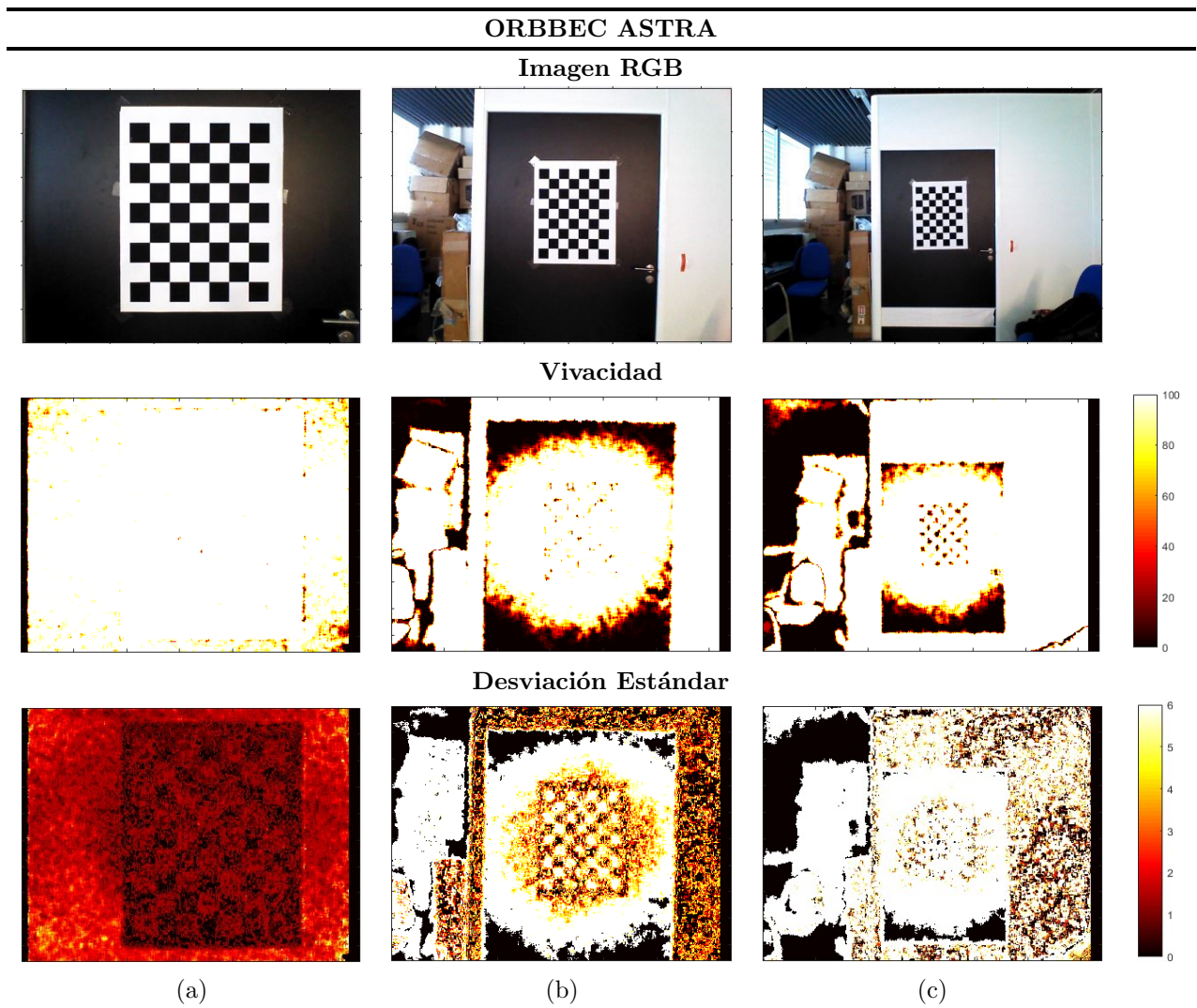


Figura 3.6: Resultados de repetibilidad por píxel en la cámara Orbbec Astra, tomando una imagen a distintas distancias de la puerta (elemento central de la escena): 1m (a), 2m (b), y 3m (c). De arriba a abajo, imagen RGB, *heatmap* de vivacidad (frecuencia de medida correcta en cada píxel) y *heatmap* de la desviación estándar (en mm) de las medidas en cada píxel en estático.

### 3.2.3. KINECT V2

La cámara Kinect V2, a diferencia de las otras dos cámaras del estudio, utiliza la tecnología *Time-of-Flight* para registrar las medidas de profundidad (Capítulo 2). Esta cámara es la más susceptible a ofrecer mayor rendimiento en términos de vivacidad de los píxeles con colores claros que con colores oscuros como se ve a lo largo de las imágenes de vivacidad de la Figura 3.7.

Puesto que el elemento central de la escena fotografiada es una puerta negra, los resultados de vivacidad mostrados en la Tabla 3.5 denotan que la cámara tiene dificultades para registrar medidas en este escenario. Analizando los resultados contenidos en la tabla, vemos que el porcentaje de píxeles con una efectividad del 100 % van de un 62 % (para la distancia de 1 metro) a un 50 % (para la distancia de 3 metros), pero, estos píxeles no se transforman en píxeles “muertos” en su mayoría como sucede con las otras dos cámaras, sino que se encuentran contenidos en los diferentes ratios de acierto.

| KINECT V2  |     |       |         |       |       |
|--|-----|-------|---------|-------|-------|
| Porcentaje de píxeles sobre el total de la imagen<br>con ratio de acierto/medida válida del: |     |       |         |       |       |
| $D$  | 0 % | <25 % | 25-75 % | >75 % | 100 % |
| 1m   | 4 % | 11 %  | 9 %     | 14 %  | 62 %  |
| 2m   | 1 % | 20 %  | 8 %     | 9 %   | 62 %  |
| 3m   | 5 % | 29 %  | 10 %    | 6 %   | 50 %  |

Tabla 3.5: Resultados de vivacidad expresados como tantos por ciento de los píxeles totales que conforman las imágenes de profundidad de la cámara Kinect V2. Cada imagen en este experimento tiene 217088 ( $424 \times 512$ ) píxeles en total.

Los resultados obtenidos en el análisis de la variabilidad de los píxeles dan cuenta de lo mencionado anteriormente, obteniendo unos valores medios de desviación típica en la Tabla 3.6 muy altos (distinto orden de magnitud) en comparación con los valores contenidos en la celda correspondiente al percentil 50 para cada distancia.

| KINECT V2 |                     |               |
|-----------|---------------------|---------------|
| $D$ [m]   | $\bar{\sigma}$ [mm] | $P_{50}$ [mm] |
| 0.5       | 3.97                | 1.55          |
| 1.0       | 132.53              | 7.47          |
| 1.5       | 442.91              | 6.87          |
| 2.0       | 589.49              | 9.94          |
| 2.5       | 692.37              | 66.93         |
| 3.0       | 729.94              | 9.29          |

Tabla 3.6: Resultados de variabilidad obtenidos para la cámara Orbbec Astra, para cada distancia de estudio: desviación típica media y percentil 50 extraídas de la matriz de desviaciones típicas calculada.

En la Figura 3.7, se puede ver como esta cámara es fuertemente dependiente de la textura del objeto que fotografía, siendo prácticamente incapaz de medir distancias a objetos donde predomina el color negro. Es necesario resaltar que la resolución de la imagen RGB en esta cámara es distinta de la resolución de la imagen de profundidad captada siendo la imagen RGB de  $1920 \times 1080$  mientras que la imagen de profundidad es de  $512 \times 424$ . A diferencia de las otras dos cámaras del estudio, esta cámara no posee franjas de píxeles “muertos” a los lados de la imagen.



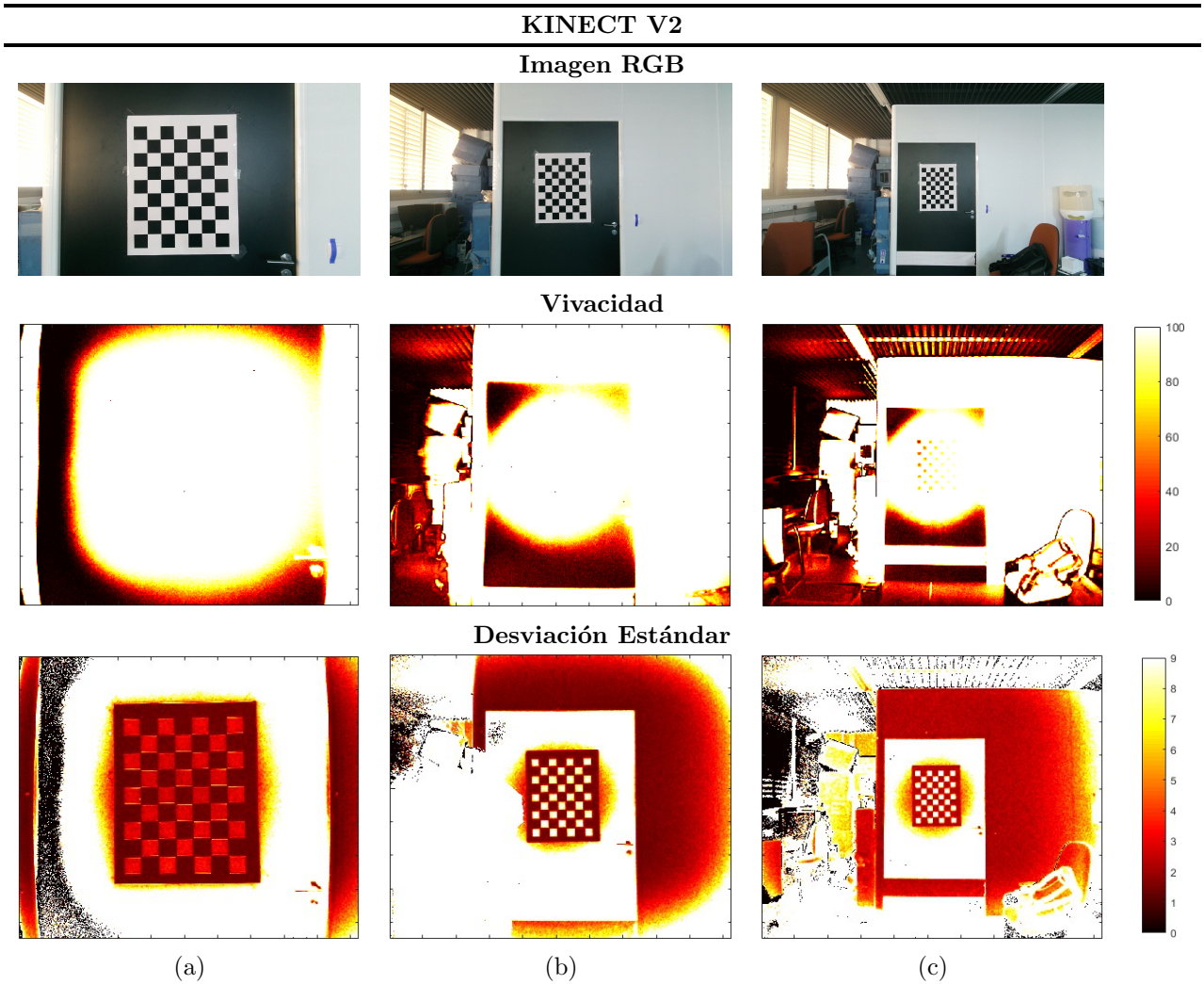


Figura 3.7: Resultados de repetibilidad por píxel en la cámara Kinect V2, tomando una imagen a distintas distancias de la puerta (elemento central de la escena): 1m (a), 2m (b), y 3m (c). De arriba a abajo, imagen RGB, *heatmap* de vivacidad (frecuencia de medida correcta en cada píxel) y *heatmap* de la desviación estándar (en mm) de las medidas en cada píxel en estático.

### 3.3. Discusión

Finalmente, se realiza una discusión comparativa de las tres cámaras a partir del análisis individualizado.

En las tres cámaras se ha visto como influyen sobre el registro de la profundidad los bordes pronunciados y los cambios de texturas bruscos. El contraste de la puerta negra con los marcos blancos ha sido tolerado por las cámaras en diferente medida, siendo la cámara ASUS Xtion Pro la que mejor resultado ha dado.

Ninguna de las tres cámaras ha sido capaz de captar la distancia a la que se encontraban las ventanas, reportando valores de profundidad nulos, por lo que, en este experimento, tanto la tecnología SL como ToF han sido incapaces de lidiar con el reflejo que produce el cristal.

De entre las tres cámaras, solo la cámara Kinect V2 es capaz de registrar medidas de profundidad a lo largo y ancho de toda la matriz de píxeles que conforman la imagen de profundidad, puesto que tanto ASUS como Orbbec contienen franjas verticales de píxeles “muertos” en los extremos laterales, siendo más destacada la franja derecha que se visualiza en las imágenes de vivacidad de la cámara Orbbec

(Figura 3.6), constante en las tres distancias estudiadas.

En la Tabla 3.7 se muestra de una manera visual qué cámara ha reportado mejores resultados en el experimento dependiendo de la condición que se busque. El *tick* verde significa que la cámara ha sacado buenos resultados en ese apartado, la cruz roja significa que la cámara no ofrece buenos resultados (en comparación con las demás) y la raya ondulada significa que podría darse como buena la cámara dependiendo de los requerimientos de la aplicación.

|                                     | ASUS Xtion Pro | Orbbec Astra | Kinect V2 |
|-------------------------------------|----------------|--------------|-----------|
| <b>Vivacidad<br/>Dist. Corta</b>    | ✓              | ~            | ✗         |
| <b>Vivacidad<br/>Dist. Media</b>    | ✓              | ✗            | ✗         |
| <b>Vivacidad<br/>Dist. Larga</b>    | ✓              | ✗            | ✗         |
| <b>Tolerancia<br/>Color Negro</b>   | ✓              | ✗            | ✗         |
| <b>Variabilidad<br/>Dist. Corta</b> | ✓              | ✓            | ~         |
| <b>Variabilidad<br/>Dist. Media</b> | ✓              | ✓            | ✗         |
| <b>Variabilidad<br/>Dist. Larga</b> | ~              | ✓            | ✗         |

Tabla 3.7: Comparativa entre las tres cámaras a partir del análisis de repetibilidad realizado. Teniendo en cuenta que: Dist. Corta equivale a un rango de 0 a 1 metro, Dist. Media equivale a un rango de 1 a 2 metros, y Dist. Larga va de 2 a 3 metros.



## Capítulo 4

# Ajuste 3D

En este capítulo se ha analizado la calidad de las nubes de puntos que se obtienen a partir de las imágenes de profundidad. Por lo tanto, el análisis se ha realizado a nivel de escena, en lugar de a nivel de píxel.

### 4.1. De la imagen de profundidad a la nube de puntos

Como se ha visto en el Capítulo 3, las imágenes de profundidad capturadas con las cámaras constan de un número determinado de píxeles, donde cada píxel contiene registrado un valor de medida. Las nubes de puntos representan la reconstrucción tridimensional de la escena fotografiada con el sensor de profundidad de las cámaras. Para la obtención de las coordenadas tridimensionales  $[X, Y, Z]^T$  de los puntos que conforman una nube de puntos se aplican las ecuaciones:

$$X = \frac{(i - c_x) \times z}{f_x}, \quad (4.1)$$

$$Y = \frac{(j - c_y) \times z}{f_y}, \quad (4.2)$$

$$Z = z, \quad (4.3)$$

donde  $i$  y  $j$  son las componentes horizontal y vertical de las coordenadas del plano de la imagen de profundidad,  $c_x$  y  $c_y$  son las coordenadas del punto principal de la cámara,  $z$  representa la medida de profundidad registrada en el píxel y, por último,  $f_x$  y  $f_y$  representan la distancia focal descompuesta en los ejes  $x$  e  $y$ . Las coordenadas del punto principal y la distancia focal utilizada para cada cámara han sido los calculados en el Capítulo 2. El proceso para deducir las ecuaciones (4.1), (4.2) y (4.3) a partir del modelo de la cámara se encuentra en el Anexo B (Sección B.1).

### 4.2. Ajuste de un plano a la nube de puntos

Para evaluar la calidad de la reconstrucción proporcionada por la nube de puntos se ha realizado un experimento donde se ha evaluado el ajuste de dicha nube a un plano dominante en la escena. Para la realización de este experimento se han utilizado las imágenes de profundidad del Capítulo 3 y los parámetros intrínsecos obtenidos durante la calibración de las cámaras descrita en el Capítulo 2.

El análisis de la calidad de la nube de puntos donde el elemento central es una superficie plana se ha fundamentado en la ecuación del plano. Teniendo en cuenta que la ecuación general del plano viene definida por

$$Ax + By + Cz + D = 0, \quad (4.4)$$

los planos ajustados a la nube de puntos de la puerta, al estar la cámara más o menos perpendicular a la misma, vendrán definidos por:

$$0x + 0y + 1z + D = 0, \quad (4.5)$$

donde  $D$  será igual a las diferentes distancias de las estaciones cámara-puerta.

El proceso realizado, descrito en la Figura 4.1, analiza si los diferentes planos ajustados para cada distancia de estudio de cada cámara se corresponden al resultado ideal descrito en la Ecuación (4.5).

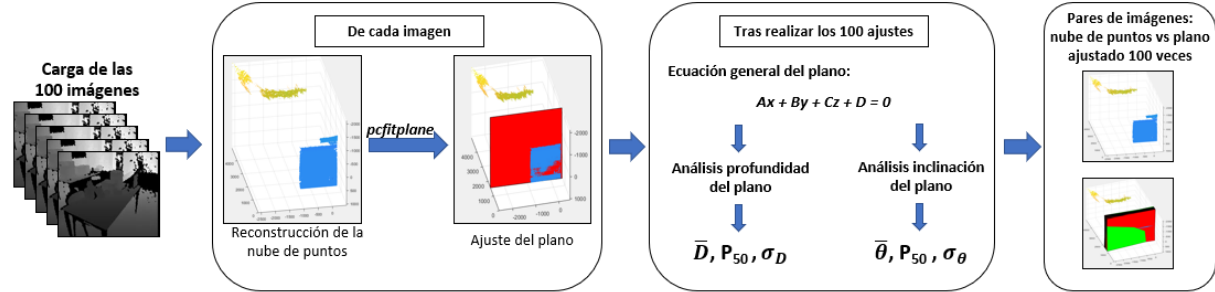


Figura 4.1: Metodología empleada para cada distancia cámara-escena en el análisis de la calidad de la nube de puntos.

En primer lugar, se han obtenido para cada distancia las cien nubes de puntos correspondientes con las cien imágenes de profundidad capturadas por cada cámara.

Se ha ajustado un plano a cada nube de puntos obtenida mediante la función *pcfitplane* de Matlab. Esta función utiliza el algoritmo *M-estimator Sample Consensus* (MSAC), una variante del algoritmo *RANdom SAMple Consensus* (RANSAC). Tanto la función *pcfitplane* como el algoritmo que usa se encuentran explicados en mayor detalle en el Anexo B, Sección B.2.

Para fijar los umbrales en el ajuste que se introducen como entrada de la función *pcfitplane* se han tenido en cuenta los resultados obtenidos en el Capítulo 3, donde quedaba patente la variabilidad que tenían los píxeles de la imagen de profundidad a la hora de registrar las distancias, además de los parámetros intrínsecos de las cámaras. El umbral de distancia fijado para cada distancia ha sido: 50, 100, 150, 200, 250 y 300 mm.

Siguiendo con la metodología descrita en la Figura 4.1, tras ajustar los planos se ha evaluado el porcentaje de *inliers* y *outliers* obtenidos de media, dándonos una primera aproximación acerca de cómo de bueno ha sido el ajuste y, por ende, de cómo de buena es la nube de puntos. Los *inliers* y *outliers* que resultan tras el ajuste del plano están fuertemente ligados, tanto a las distancias máximas definidas, como a la escena donde se encuentra la superficie plana a la que se realiza el ajuste. A medida que se aleja la cámara es esperable que el número de *inliers* disminuya, puesto que el plano deja de ser el elemento dominante de la escena.

Puesto que el modelo geométrico ajustado a la nube de puntos es un plano, el siguiente paso ha consistido en evaluar la distancia a la que ha sido ajustado con respecto de la cámara así como la inclinación de los diferentes planos ajustados con respecto la situación ideal descrita en la Ecuación (4.5).

Si descomponemos la ecuación del plano, se han evaluado por separado:

- La componente  $D$  para el análisis sobre la profundidad de los planos.
- Las componentes  $(A, B, C)$  para el análisis sobre la inclinación de los planos.

Tanto en el análisis de profundidad como en el de la inclinación de los planos ajustados se han obtenido la media, la mediana y la desviación estándar del conjunto de los 100 planos ajustados a las 100 nubes de puntos, pudiendo evaluar de manera rigurosa la calidad de las nubes de puntos extraídas de las imágenes de profundidad.

Por último, se muestran para las distancias de 0.5 y 2.5 metros pares de imágenes de manera que en la parte superior del par se visualiza una nube de puntos de las 100 que se obtienen en la distancia específica y, en la parte inferior del par, la misma nube de puntos más el montante del plano ajustado 100 veces.

### 4.3. Resultados

Los resultados numéricos obtenidos se muestran a lo largo de tres tablas. La Tabla 4.1 recoge los *inliers* y *outliers* obtenidos de media al ajustar el plano a cada nube de puntos, los resultados se muestran en modo porcentual de manera que expresan el porcentaje de píxeles sobre el total. La Tabla 4.2 muestra los resultados obtenidos en el análisis de la profundidad del plano ajustado a las nubes de puntos obtenidas a lo largo de las diferentes distancias de estudio. Por último, el análisis sobre la inclinación del mismo plano ajustado se encuentra en la Tabla 4.3.

| ASUS XTION PRO |             |              | ORBBEC ASTRA |              | KINECT V2   |              |
|----------------|-------------|--------------|--------------|--------------|-------------|--------------|
| $D$ [mm]       | Inliers [%] | Outliers [%] | Inliers [%]  | Outliers [%] | Inliers [%] | Outliers [%] |
| 500            | 93          | 7            | 65           | 35           | 64          | 36           |
| 1000           | 96          | 4            | 93           | 7            | 79          | 21           |
| 1500           | 90          | 10           | 73           | 27           | 69          | 31           |
| 2000           | 76          | 24           | 63           | 37           | 69          | 31           |
| 2500           | 73          | 27           | 63           | 37           | 59          | 41           |
| 3000           | 70          | 30           | 62           | 38           | -           | -            |

Tabla 4.1: *Inliers* y *Outliers* expresados de manera porcentual (sobre el total de píxeles que conforman la imagen en cada cámara), obtenidos de media al ajustar el plano a las 100 nubes de puntos analizadas en cada distancia por cámara.

Los resultados de la Tabla 4.1 son indicadores sobre el ajuste realizado. Teniendo en cuenta que cada píxel con medida válida resulta en un punto tridimensional los valores porcentuales contenidos en las columnas encabezadas por *inliers* representan los puntos de la nube de puntos que han sido tomados por el algoritmo para ajustar el plano, mientras que los *outliers* son los puntos que han quedado fuera del mismo. Para la distancia de 0.5 metros, donde la cámara fotografía la puerta junto con el patrón, el resultado esperable sería un porcentaje de *inliers* mucho mayor que de *outliers*, ya que la superficie plana abarca toda la escena. En cambio, el resultado que se observa es que sólo la cámara ASUS Xtion Pro cumple esta afirmación. También se puede observar en la tabla como, a medida que aumenta la distancia, el porcentaje que representa los *inliers* disminuye, mientras que su homónimo aumenta, esto es debido a que la superficie plana central ocupa menos espacio en la imagen de profundidad conforme escena fotografiada aumenta y al incremento de la variabilidad en el registro de las medidas de los píxeles visto en el Capítulo 3.

Para la distancia de 3 metros se puede observar como en la cámara Kinect V2 no hay datos registrados. Esto es debido la alta dispersión de los puntos tridimensionales en las nubes de puntos. El algoritmo no ha sido capaz de realizar el ajuste del plano sobre la superficie plana central, por lo que el resultado se ha desechado del análisis. Se puede ver el montante del plano ajustado 100 veces sobre una nube de puntos junto con la misma nube de puntos en el par de imágenes correspondiente a la distancia de 3 metros en el Anexo B, (Sección B.3, Figura B.4).

#### 4.3.1. Resultados Cuantitativos

**Resultados del análisis de profundidad.** Como se observa en la Tabla 4.2, de las tres cámaras, la cámara ASUS Xtion Pro ha sido la que mejores valores de media, mediana y desviación típica ha reportado. Estos resultados son congruentes con los obtenidos en el Capítulo 3 (donde esta cámara también reportó los mejores resultados en el análisis de repetibilidad al fotografiar el mismo escenario).

Para las distancias cámara-plano de 0.5 a 2.5 metros las tres cámaras han reportado unos valores de media y mediana similares, puesto que la cámara se encontraba sobre un trípode al realizar las capturas. El

|      | ASUS      |          |            | ORBEC     |          |            | KINECT    |          |            |
|------|-----------|----------|------------|-----------|----------|------------|-----------|----------|------------|
| $D$  | $\bar{D}$ | $P_{50}$ | $\sigma_D$ | $\bar{D}$ | $P_{50}$ | $\sigma_D$ | $\bar{D}$ | $P_{50}$ | $\sigma_D$ |
| 500  | 534       | 534      | 1.99       | 466       | 466      | 0.88       | 519       | 520      | 8.56       |
| 1000 | 990       | 990      | 2.54       | 991       | 991      | 3.41       | 1036      | 1036     | 9.44       |
| 1500 | 1514      | 1515     | 15.59      | 1472      | 1473     | 9.29       | 1526      | 1525     | 21.45      |
| 2000 | 2021      | 2021     | 25.69      | 1986      | 1984     | 14.25      | 2048      | 2052     | 17.51      |
| 2500 | 2509      | 2509     | 29.52      | 2485      | 2486     | 26.99      | 2534      | 2534     | 22.36      |
| 3000 | 2965      | 2964     | 34.65      | 2954      | 2983     | 300.38     | -         | -        | -          |

Tabla 4.2: Resultados del análisis sobre la profundidad de cada plano para todas las cámaras. Para cada distancia ( $D$ ) se han obtenido media ( $\bar{D}$ ), mediana ( $P_{50}$ ) y desviación estándar ( $\sigma_D$ ) del ajuste realizado 100 veces. Todas las unidades están en mm.

hecho de que estos valores sean similares indica que las cámaras ofrecen registros estables de profundidad para superficies planas en este rango de distancias.

A la distancia de 3 metros, la cámara Orbbec Astra contiene unos valores de mediana y desviación típicas desvirtuados. Esto es debido a que una nube de puntos contenía alta dispersión en sus puntos, por lo que el ajuste del plano en esa situación no se ha ajustado a la superficie plana central con el *threshold* estipulado. En el Anexo B, Sección B.3, Figura B.3 (f), se puede ver el resultado completo al ajustar el plano a las 100 nubes de puntos de entrada.

**Resultados del análisis sobre la inclinación del plano.** Los resultados de este análisis, disponibles en la Tabla 4.3, se han obtenido midiendo el ángulo que formaba el plano que se ajustaba a la nube de puntos con el plano ideal de normal  $(0, 0, 1)$ .

|      | ASUS           |          |                   | ORBEC          |          |                   | KINECT         |          |                   |
|------|----------------|----------|-------------------|----------------|----------|-------------------|----------------|----------|-------------------|
| $D$  | $\bar{\theta}$ | $P_{50}$ | $\sigma_{\theta}$ | $\bar{\theta}$ | $P_{50}$ | $\sigma_{\theta}$ | $\bar{\theta}$ | $P_{50}$ | $\sigma_{\theta}$ |
| 500  | 1              | 1        | 0.81              | 1              | 0        | 0.49              | 3              | 2        | 1.83              |
| 1000 | 1              | 1        | 0.97              | 1              | 1        | 0.96              | 3              | 3        | 1.69              |
| 1500 | 2              | 2        | 1.88              | 2              | 2        | 1.54              | 3              | 2        | 2.04              |
| 2000 | 3              | 3        | 2.16              | 3              | 3        | 2.05              | 2              | 2        | 1.24              |
| 2500 | 3              | 3        | 1.83              | 3              | 3        | 1.80              | 1              | 1        | 1.05              |
| 3000 | 4              | 3        | 2.54              | 5              | 3        | 10.55             | -              | -        | -                 |

Tabla 4.3: Resultados del análisis sobre la inclinación de cada plano para todas las cámaras. Para cada distancia ( $D$ ) se han obtenido media ( $\bar{\theta}$ ), mediana ( $P_{50}$ ) y desviación estándar ( $\sigma_{\theta}$ ) del ajuste realizado 100 veces. Las distancias se encuentran en mm mientras que las unidades del resto de valores están en grados.

Los resultados obtenidos evidencian una dispersión mínima en los puntos tridimensionales que conforman la superficie plana dentro de las nubes de puntos, encontrándose para las tres cámaras una inclinación de los ángulos formados entre las normales del plano ajustado y el ideal siempre menor de  $6^\circ$ . Teniendo en cuenta que las cámaras en el momento de capturar las imágenes se encontraban colocadas sobre trípodes (pudiendo tener una ligera inclinación con respecto los  $90^\circ$  que formaban con la puerta) los valores de desviación típica obtenidos en las tres cámaras, menores de  $2^\circ$  para la mayoría de las distancias

cámara-plano corroboran la buena reconstrucción de superficies planas.

### 4.3.2. Resultados Cualitativos

En la Figura 4.2 se visualizan los resultados del ajuste del plano a las nubes de puntos realizado para una distancia considerada como cercana (0.5 metros) y otra lejana (2.5 metros). Los diferentes planos solapados se encuentran coloreados de manera que el plano ajustado es verde si supera la media de *inliers* registrada como porcentual sobre el total, en la Tabla 4.1, y rojo si no es así.

Destacan las nubes de puntos obtenidas para la cámara Orbbec Astra y Kinect V2 a la distancia cercana, teniendo zonas sin puntos (cuyos píxeles fueron incapaces de registrar medidas de distancia), destacando una vez más la sensibilidad que tienen los sensores de estas cámaras a la textura del material que fotografían.

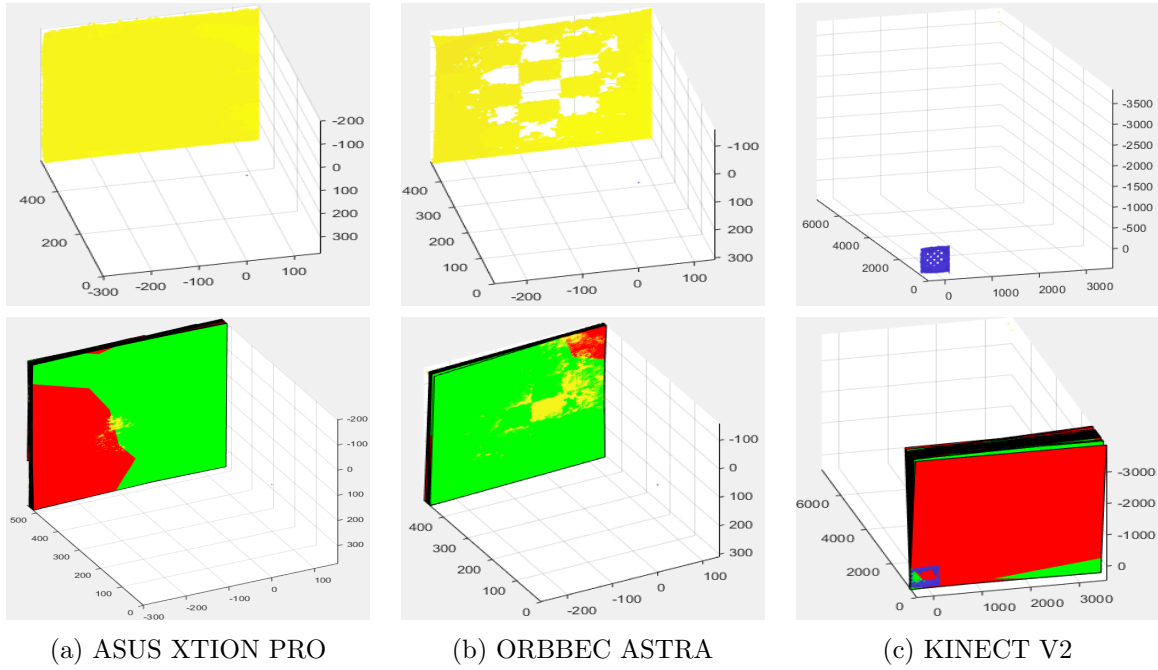
La cámara Kinect V2 es la que más volatilidad presenta en los puntos que conforman sus nubes de puntos al fotografiar el escenario del experimento. Se puede ver como para la distancia de 0.5 metros la nube de puntos reconstruida se encuentra en la esquina inferior izquierda, cuando lo normal sería que ocupara el fondo de la escena como pasa con las otras dos cámaras. Esto es debido al auto-ajuste de los ejes que realiza Matlab, dando cuenta de que hay puntos tridimensionales que han sido situados a distancias muy superiores a la distancia de estudio, o lo que es lo mismo, la cámara ha registrado en las imágenes de profundidad de esta distancia, valores nulos y muy superiores 0.5 metros. Lo que denota una vez más la influencia sobre la tecnología que utiliza esta cámara para medir la profundidad de los materiales y las superficies de los escenarios que fotografían.

El resto de pares de imágenes de las diferentes distancias se encuentran agrupados por las diferentes cámaras en el Anexo B, Sección B.3.

---

Cerca = 0.5m

---




---

Lejos = 2.5m

---

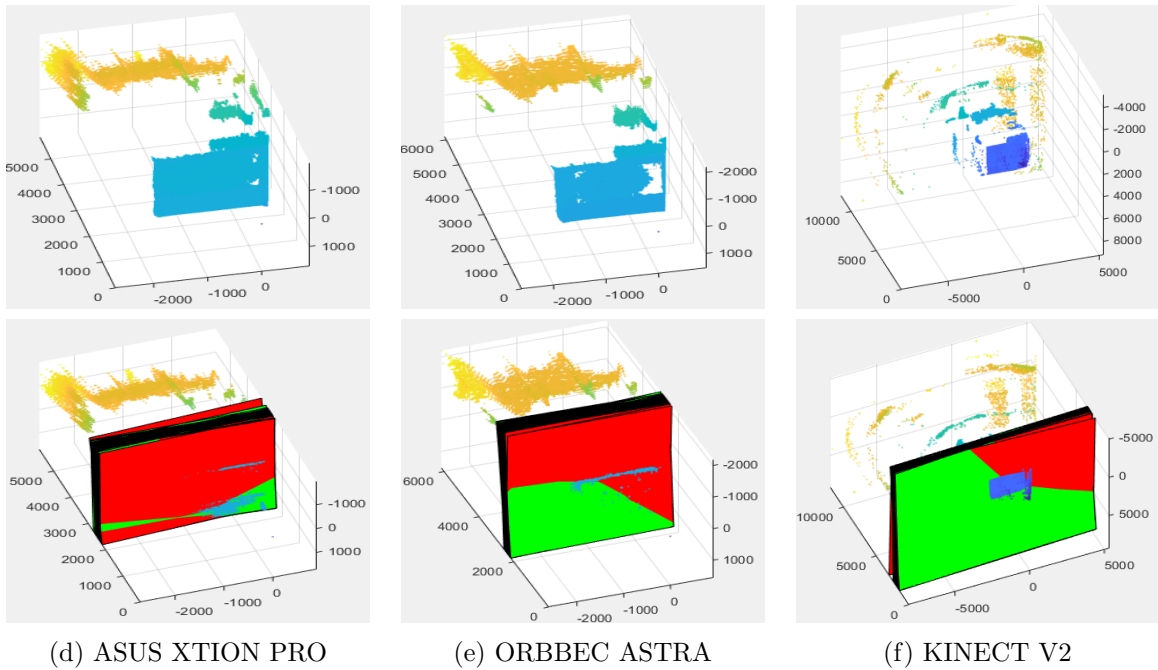


Figura 4.2: Resultados obtenidos en el análisis de las nubes de puntos de las tres cámaras para las distancias de 0.5 y 2.5 metros. Pares de imágenes: (arriba) Nube de puntos extraída de una imagen de profundidad para la distancia de estudio y, (abajo) el montante de los cien planos ajustados a lo largo de las cien nubes de puntos.

## Capítulo 5

# Conclusiones

En este TFG se ha realizado un estudio para evaluar la capacidad de las cámaras de visión y profundidad, RGB-d, para ser utilizadas en aplicaciones de Metrología Industrial. El estudio realizado cubre satisfactoriamente los objetivos planteados: estudio bibliográfico sobre las distintas tecnologías empleadas por las cámaras RGB-d para la medida de la profundidad, análisis formal sobre la repetibilidad que ofrecen los sensores y evaluación de la calidad de las nubes de puntos extraídas de las imágenes de profundidad capturadas por las cámaras, y recopila una serie de resultados interesantes que se discuten a continuación.

### 5.1. Discusión de resultados.

**Discusión sobre tecnologías y calibración.** La primera parte del trabajo ha consistido en estudiar las principales tecnologías usadas por las cámaras para medir la profundidad, encontrándose todas ellas dentro de dos tecnologías: luz estructurada y tiempo de vuelo. Una vez analizadas las tecnologías, se han realizado los procesos de calibración de las distintas cámaras disponibles: ASUS Xtion Pro, Orbbec Astra y Kinect V2. Los métodos estándar estudiados han permitido obtener la calibración sin problemas, con la única restricción de que necesitan una cuidadosa adquisición de los datos utilizados para la calibración. Esta calibración nos ha permitido poder obtener las medidas más precisas posibles para cada sensor, y construir nubes de puntos en 3D a partir de dichas medidas.

**Discusión sobre resultados de la caracterización.** La caracterización de estos sensores ha consistido en un estudio dividido en dos bloques. Puesto que la aplicación potencial de las cámaras sería a través de visión tridimensional, primero se ha realizado un análisis del *hardware*, estudiándose la respuesta que ofrecen los píxeles que conforman los mapas (imágenes) de profundidad que registran, seguido de un análisis a alto nivel de las nubes de puntos que se obtienen de las mismas imágenes.

El estudio de *repetibilidad* ha consistido en un análisis formal del comportamiento que han ofrecido las distintas cámaras píxel a píxel en un escenario controlado, dominado por un plano con orientación perpendicular a la cámara. Se ha constatado un comportamiento de los píxeles altamente ligado a las texturas y a las formas de los objetos que fotografían. Se ha visto que las cámaras pueden contener zonas donde no son capaces de medir ningún valor válido de profundidad independientemente de la escena que estén fotografiando. Por último, se ha corroborado un mejor comportamiento de las cámaras en cuanto a la estabilidad de las medidas que registraban a cortas distancias frente a distancias largas.

Para la realización del segundo bloque del estudio, basado en las **nubes de puntos en 3D** que se obtienen a partir de las imágenes de profundidad, primero se ha estudiado la metodología a seguir para poder reconstruir puntos tridimensionales a partir de los parámetros intrínsecos de las cámaras y las propias imágenes de profundidad. Aprovechando el escenario utilizado para el estudio de repetibilidad se han reconstruido las nubes de puntos de las imágenes de profundidad, ajustando para cada nube la ecuación del plano dominante de la misma. Para lograrlo, se ha estudiado tanto la profundidad como la inclinación del plano ajustado dependiendo de la distancia del sensor al plano dominante. Con ello, se ha

conseguido una visión global acerca del comportamiento ofrecido por las cámaras ante medidas obtenidas a diferentes distancias de una misma escena.

Como resumen de los resultados, cabe destacar que pese a los precios similares, hay gran diferencia en la robustez de las medidas de los distintos sensores estudiados. Se ha comprobado que se pueden alcanzar medidas de repetibilidad y precisión de hasta 0.4 mm al fotografiar un escenario con un elemento central plano para una distancia de posicionamiento de la cámaras (en el caso de la ASUS y la Orbbec) de 0.5 metros con respecto al escenario.

## **5.2. Trabajo futuro**

Este TFG ha abordado una primera visión acerca del desempeño de las cámaras RGB-d ante un tipo de escenario donde el principal elemento era una superficie plana. Una línea futura de trabajo sería continuar con el análisis a un siguiente nivel, evaluando objetos de mayor complejidad estructural, con distintas formas geométricas, como esferas o cubos. Analizando el comportamiento que ofrecen las mismas, así como la capacidad para reconstruir de manera fidedigna las nubes de puntos. Otra línea de trabajo futuro podría ir encaminada a un estudio acerca del comportamiento de las cámaras ante diferentes texturas y superficies, pudiendo discretizar de manera más fehaciente hasta que punto son capaces de llegar en base a la tecnología sobre la que están construidas.



## Apéndice A

# Error de reproducción en la calibración de cámaras

Cada barra del histograma representa los errores de reproyección medios obtenidos en esa imagen. A su vez, el histograma muestra el error de reproyección medio obtenido entre todas las imágenes.

**Cálculo del error de reproyección.** El algoritmo del programa primero detecta los puntos en el patrón (esquinas del tablero de ajedrez). Posteriormente estima los parámetros de la cámara y en base a estos, reproyecta los puntos.

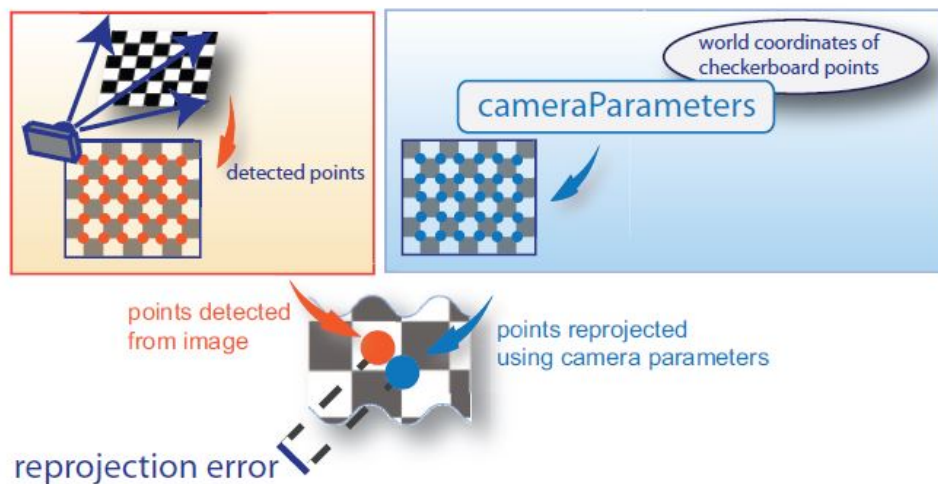
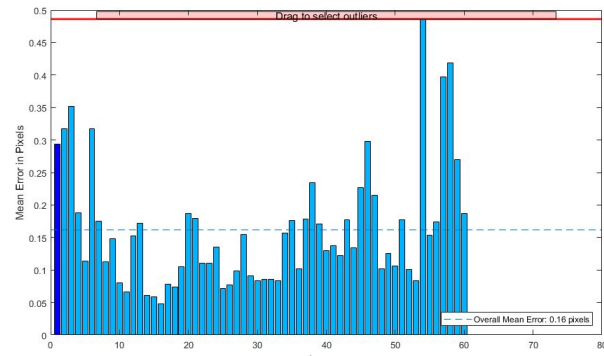
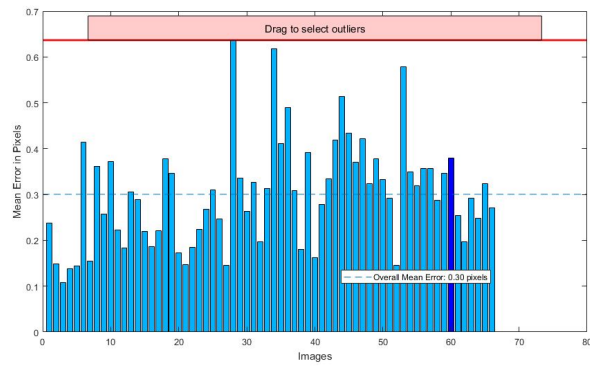


Figura A.1: Detalle del Error de Reproyeccion teórico. Imagen extraída de [5].

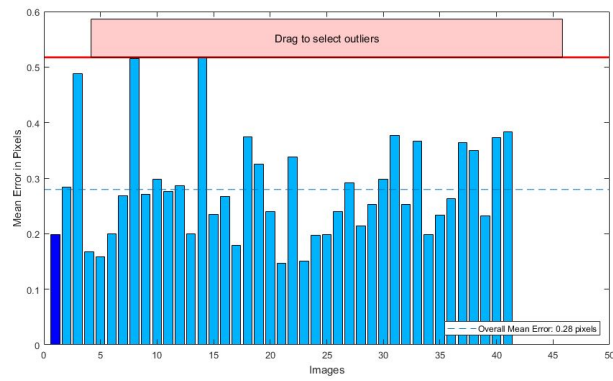
La diferencia entre los puntos detectados de la imagen real y los puntos proyectados a partir de los parámetros calculados resulta en un error de proyección para cada punto. En la figura A.1 se puede ver un esquema en detalle.



(a)



(b)



(c)

Figura A.2: Errores de Reproyección obtenidos durante la calibración. (a): Cámara Kinect V2. (b): Cámara ASUS Xtion PRO. (c): Cámara Orbbec Astra.

## Apéndice B

# Nubes de puntos y ajuste del plano

El siguiente anexo está dividido en varias secciones, incluyendo cada una información adicional o resultados obtenidos durante el desarrollo del Capítulo 4.

### B.1. Del modelo de la cámara a la nube de puntos

Para obtener la nube de puntos de una imagen de profundidad, primero se debe considerar un sistema de referencia tridimensional (con ejes  $x$ ,  $y$ ,  $z$ ), llamado sistema de coordenadas de la cámara (*Camera Coordinates System, CCS*). El sistema de coordenadas de la cámara está formado por un origen  $O$ , centro de proyección, y un plano paralelo al plano formado por los ejes  $x$  e  $y$ , que intersecta en el eje  $z$  en la coordenada negativa  $f$ , llamado plano de la imagen, Figura B.1 (a).

Considerando ahora un sistema de referencia 2D asociado al plano de la imagen y formado por las ecuaciones (B.1) y (B.2):

$$u = x + c_x, \quad (\text{B.1})$$

$$v = y + c_y. \quad (\text{B.2})$$

La intersección  $c$  del eje  $z$  con el plano de la imagen tiene coordenadas  $\mathbf{c} = [u = c_x, v = c_y]^T$ . Por tanto el conjunto de puntos (píxeles),  $p$ , que conforman el plano de la imagen, con coordenadas  $\mathbf{p} = [u, v]^T$  obtenidas con la intersección de los rayos que conectan el centro de proyección  $O$  con todos los puntos tridimensionales de la escena,  $P$ , con coordenadas  $\mathbf{P} = [x, y, z]^T$ , representan la escena tridimensional en el plano de la imagen.

La relación entre  $P$  y  $p$ , llamada proyección central o proyección perspectiva se puede demostrar con facilidad a través de la similitud de triángulos, resultando las ecuaciones contenidas en (B.3) y (B.4) [11]:

$$u - c_x = f \frac{x}{z}, \quad (\text{B.3})$$

$$v - c_y = f \frac{y}{z}. \quad (\text{B.4})$$

Donde  $f$  representa la distancia entre el centro de proyección  $O$  y el centro del plano de la imagen. Esta distancia es la denominada *Distancia Focal*. Si descomponemos la distancia focal en los ejes  $x$  e  $y$ ,  $(f_x, f_y)$ , para una imagen de profundidad, con valores de profundidad registrados en sus píxeles, las ecuaciones para la obtención de las coordenadas tridimensionales  $[X, Y, Z]^T$  de cada píxel  $(i, j)$  del plano de la imagen son:

$$X = \frac{(i - c_x) \times z}{f_x}, \quad (\text{B.5})$$

$$Y = \frac{(j - c_y) \times z}{f_y}, \quad (\text{B.6})$$

$$Z = z. \quad (\text{B.7})$$

Con  $z$  igual a la profundidad registrada por los píxeles de la imagen de profundidad.

Para obtener finalmente la nube de puntos a partir de la imagen de profundidad, es necesario recorrer todos los píxeles que conforman el plano de la imagen utilizando las ecuaciones (B.5), (B.6) y (B.7).

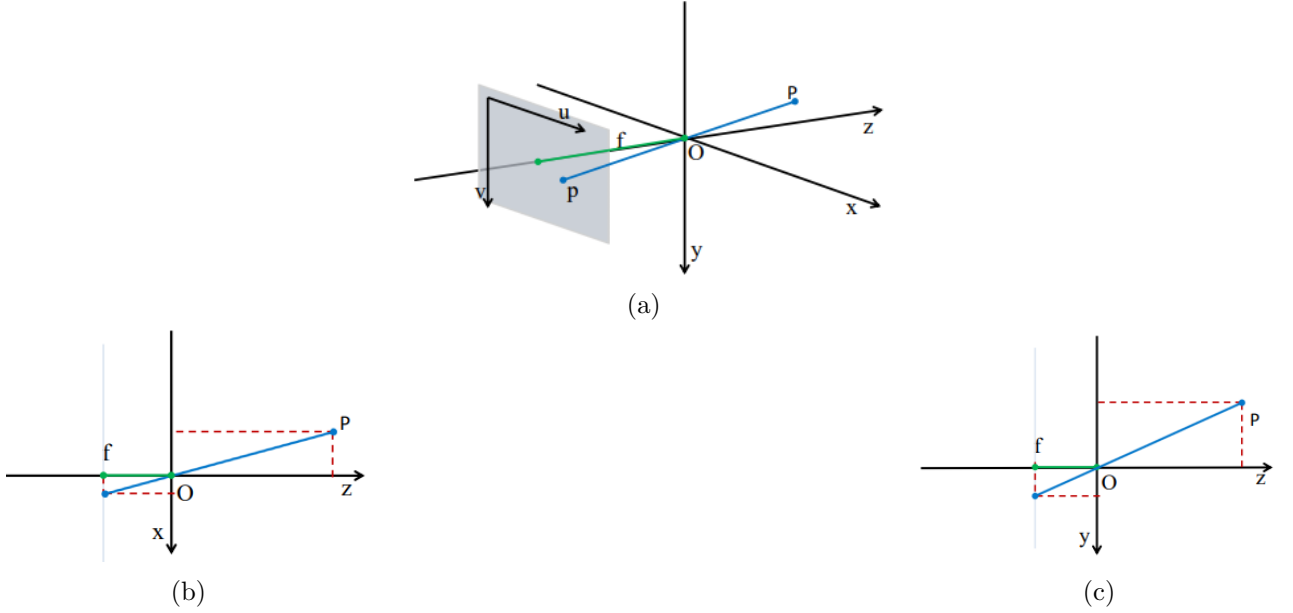


Figura B.1: Geometría proyectiva: (a): Un punto de la escena,  $P$  se proyecta en el plano de la imagen  $p$ . (b): sección horizontal de (a), (c): sección vertical de (a). Imágenes extraídas de: [11].

## B.2. Función *pcfitplane* de Matlab

La función *pcfitplane* de Matlab ajusta un plano a una nube de puntos fijando un límite de distancia para los puntos de entrada al plano. La función devuelve un modelo geométrico que describe un plano. Usa el algoritmo *M-estimator Sample Consensus* (MSAC) para encontrar el plano. El algoritmo MSAC es una variante del algoritmo *RANdom SAMple Consensus* (RANSAC). El consenso de muestreo aleatorio, (*RANdom SAMple Consensus*, es un algoritmo iterativo para estimar los parámetros de un modelo matemático a partir de un conjunto de datos observados que contiene valores atípicos, cuando no se debe otorgar ninguna influencia a los valores atípicos en los valores de las estimaciones. Los datos que se ajustan al modelo matemático se denominan *inliers*, mientras que los valores atípicos que quedan fuera del modelo son denominados *outliers*.

El algoritmo que implementa el método RANSAC puede ser sensible a la elección del umbral de ruido correcto que define que datos se ajustan a un modelo instanciado con un cierto conjunto de parámetros. Si dicho umbral es demasiado grande, entonces todas las hipótesis tienden a ser clasificadas por igual (dadas por buenas). En cambio, cuando el umbral de ruido es demasiado pequeño, los parámetros estimados tienden a ser inestables (añadiendo o eliminando un dato al conjunto de *inliers*, la estimación de los parámetros puede fluctuar).

La función se implementa como

$$model = pcfitplane(ptCloudIn, maxDistance), \quad (B.8)$$

donde *model* es un objeto tipo *planeModel* que contiene el modelo geométrico del plano ajustado, *ptCloudIn* es la nube de puntos de entrada a la función y, *maxDistance* es la distancia máxima que la función toma para encontrar posibles puntos del plano, descartando aquellos que se encuentren a mayor distancia.

### **B.3. Resultados: ajuste de un plano a la nube de puntos**

Pares de imágenes con los resultados obtenidos al ajustar los planos a las diferentes nubes de puntos de cada distancia. Las pares de imágenes se componen: arriba, una de las 100 nubes de puntos reconstruidas, abajo, el conjunto de los 100 planos ajustados.

### B.3.1. ASUS XTION PRO

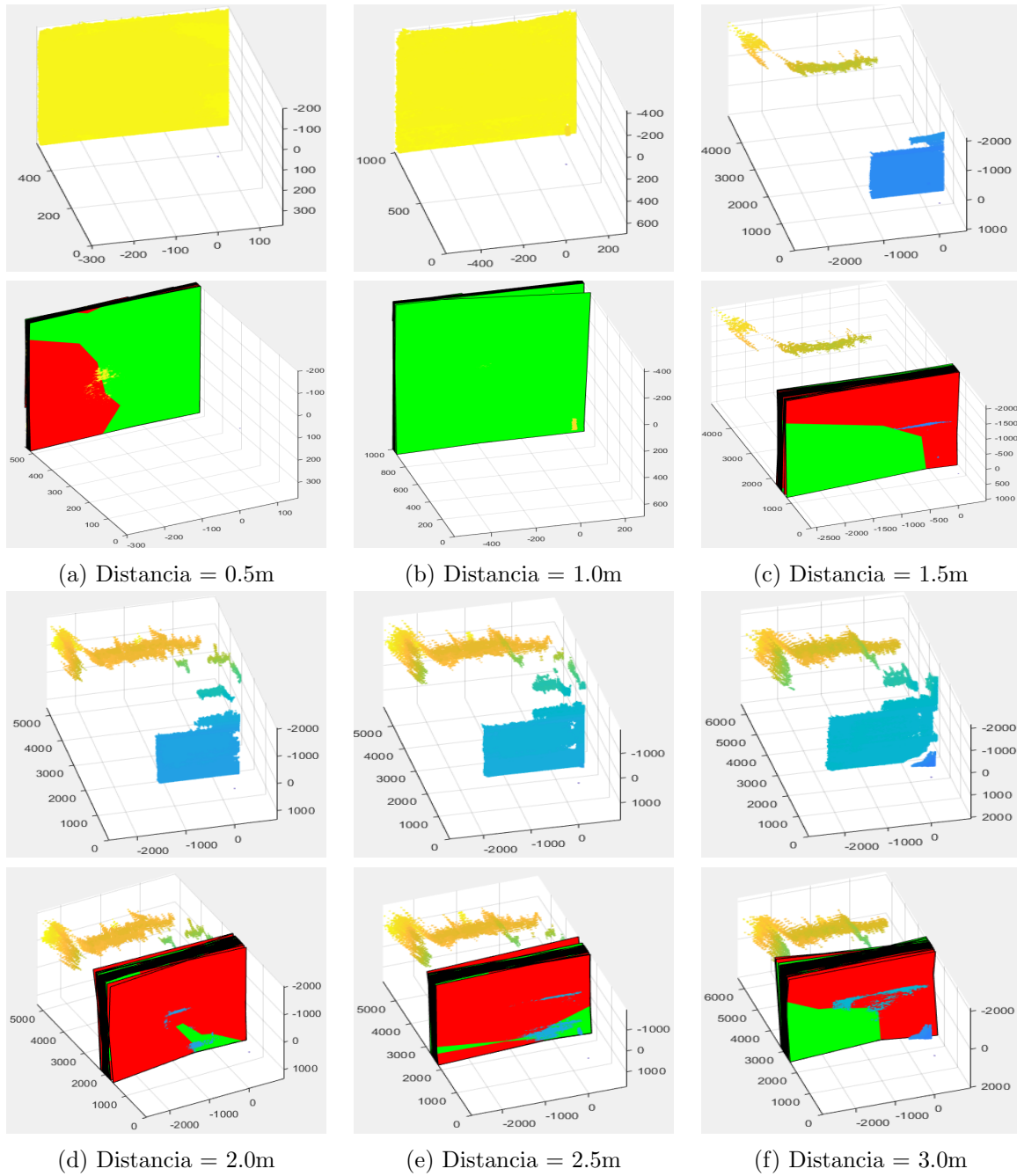


Figura B.2: Cámara Asus Xtion Pro.

### B.3.2. ORBBEC ASTRA

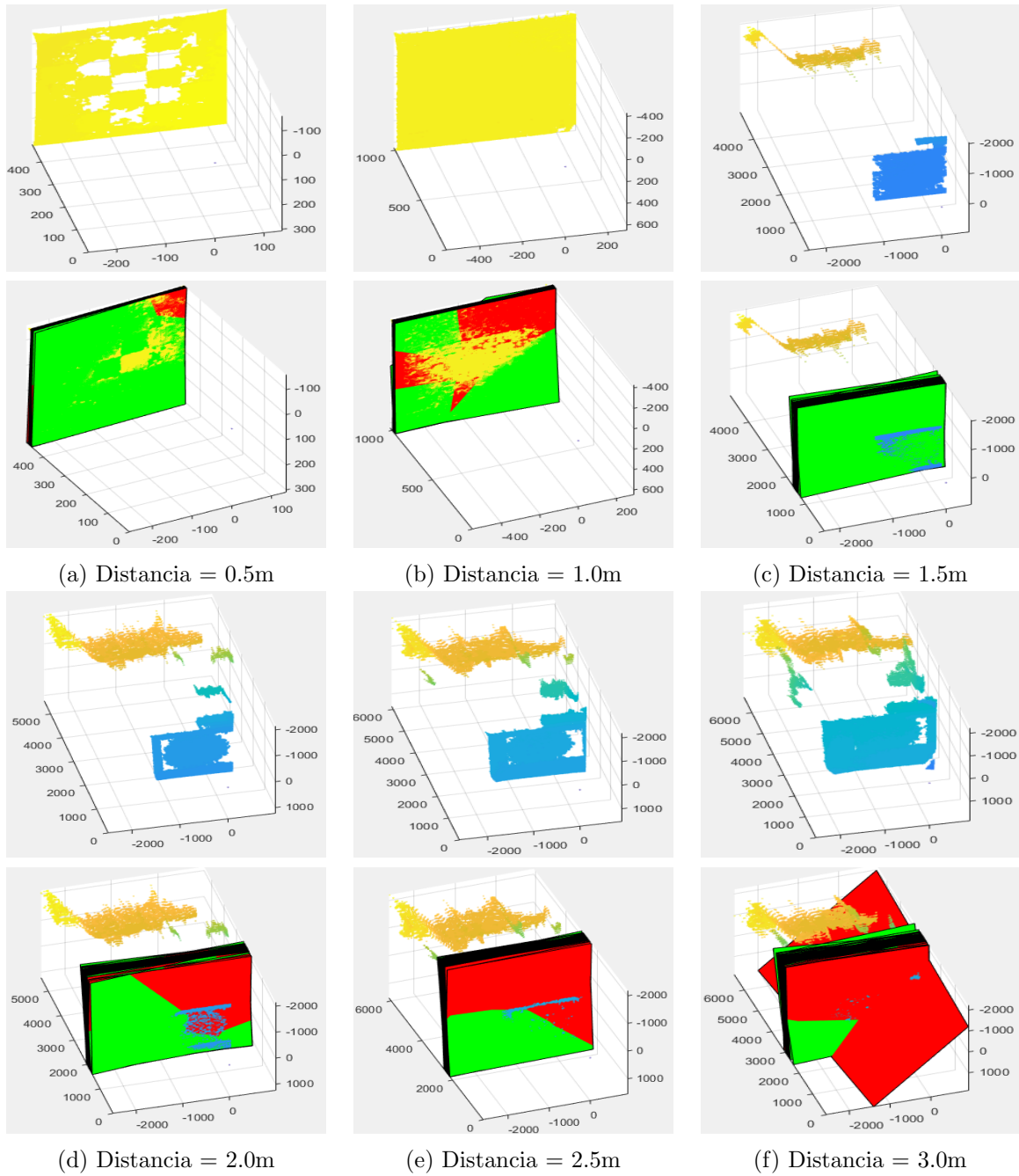


Figura B.3: Cámara Orbbec Astra.

### B.3.3. KINECT V2

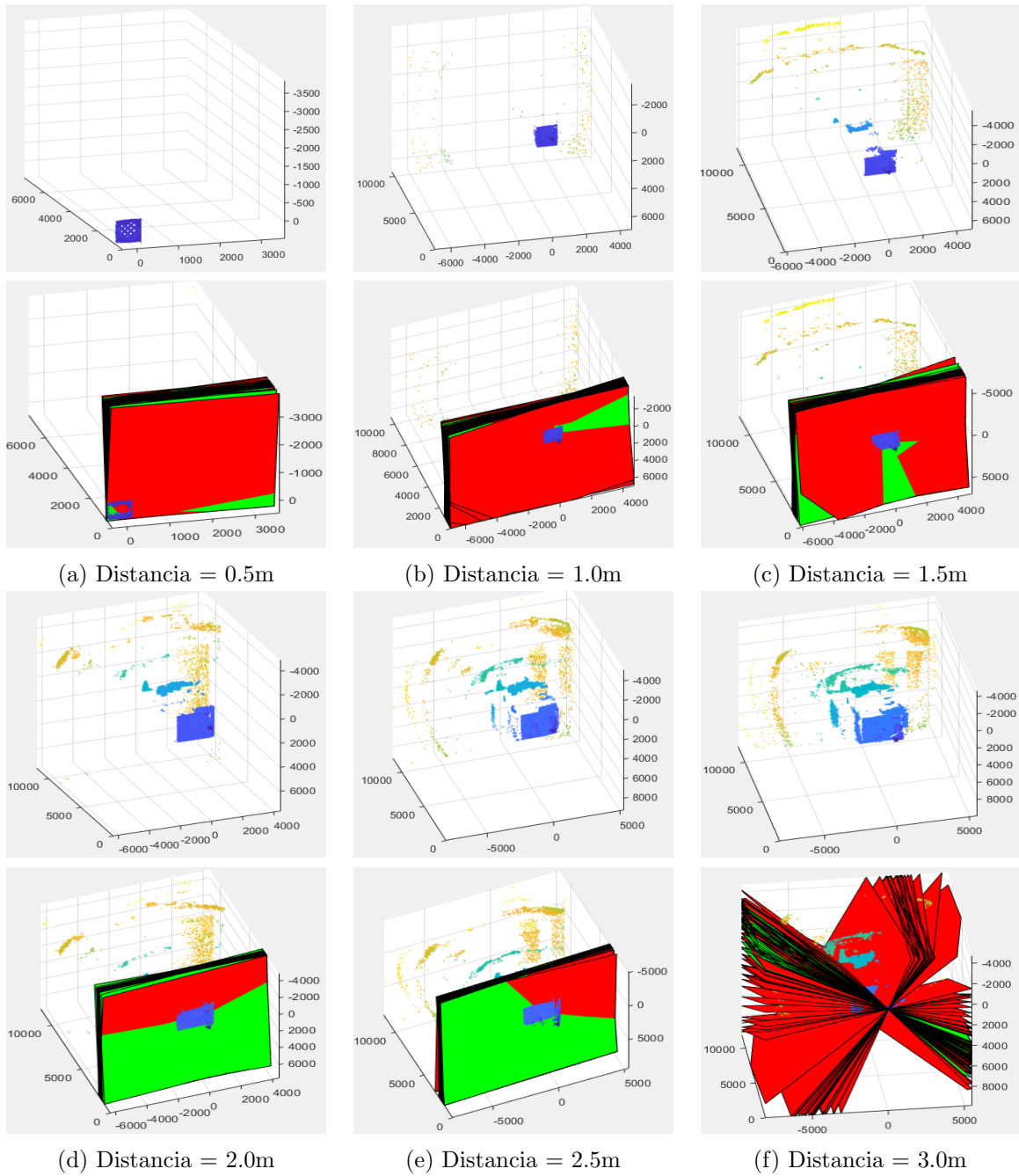


Figura B.4: Cámara Kinect V2.



# Bibliografía

- [1] Kourosh Khoshelham and Sander Oude Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, 2012.
- [2] Peter Henry, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. Rgb-d mapping: Using kinect-style depth cameras for dense 3d modeling of indoor environments. *The International Journal of Robotics Research*, 31(5):647–663, 2012.
- [3] Walid Darwish, Shenjun Tang, Wenbin Li, and Wu Chen. A new calibration method for commercial rgb-d sensors. *Sensors*, 17(6):1204, 2017.
- [4] Elise Lachat, Hélène Macher, Tania Landes, and Pierre Grussenmeyer. Assessment and calibration of a rgb-d camera (kinect v2 sensor) towards a potential use for close-range 3d modeling. *Remote Sensing*, 7(10):13070–13097, 2015.
- [5] MATLAB. *Computer Vision System Toolbox User’s Guide*. The MathWorks Inc., Natick, Massachusetts, R2017a.
- [6] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.
- [7] Giovanna Sansoni, Marco Trebeschi, and Franco Docchio. State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, 9(1):568–601, 2009.
- [8] ASUS Xtion Pro. Página oficial. Accedido el 04/11/2017: [https://www.asus.com/es/3D-Sensor/Xtion\\_PRO/](https://www.asus.com/es/3D-Sensor/Xtion_PRO/).
- [9] Orbbec Astra. Página oficial. Accedido el 04/11/2017: <https://orbbec3d.com/product-astra/>.
- [10] Structure Sensor. Página oficial. Accedido el 04/11/2017: <https://structure.io/>.
- [11] Carlo Dal Mutto, Pietro Zanuttigh, and Guido M Cortelazzo. *Time-of-flight cameras and microsoft kinect (TM)*. Springer Publishing Company, Incorporated, 2012.
- [12] Windows Dev Center. Página oficial. Accedido el 04/11/2017: <https://developer.microsoft.com/en-us/windows/kinect/hardware>.
- [13] Jean-Yves Bouguet. Matlab camera calibration toolbox. *Caltech Technical Report*, 2000.