

## Capítulo 1

# Introducción

El ser humano no es consciente del proceso neuronal que tiene lugar en nuestro cerebro con el simple hecho de andar o coger un objeto. Se podría decir que tenemos un super ordenador conectado a los órganos sensoriales capaces de recoger muchísima información y procesarla en un tiempo récord.

Desde la antiguedad ya se estuvo pensando en reproducir las habilidades humanas en algún tipo de máquina, la noción de concebir la mente humana como algún tipo de mecanismo no es reciente es referida en célebres filósofos, sin embargo, no es hasta 1950 y con la noción de la computación cuando se introduce la IA (Inteligencia Artificial) por el científico Alan Turing en su artículo *Maquinaria Computacional e Inteligencia* y donde se empieza a coger interés por este campo que será el precursor de una gran cantidad de desarrollos e innovaciones.

### 1.1. Visión artificial

Dentro del campo de la inteligencia artificial se puede definir visión artificial como la disciplina científica que incluye métodos para adquirir, procesar y analizar imágenes con el fin de producir información que pueda ser tratada por una máquina ofreciendo soluciones a problemas del mundo real.

Una manera simple de comprender este sistema es basarnos en nuestra propia experiencia. Los humanos usamos nuestros sentidos, en este contexto el ojo, para comprender el mundo que nos rodea, y la visión artificial busca producir ese mismo efecto en máquinas.

Cada vez son más los dispositivos electrónicos que llevan incorporada al menos una cámara; *smartphones*, ordenadores portátiles, *tablets*, consolas de videojuegos... Debido a la gran cantidad de información que se puede extraer de las imágenes, el bajo coste, el reducido tamaño de las cámaras y el aumento de capacidad de cómputo de los dispositivos, es un área que ha suscitado el interés por los investigadores, ha crecido enormemente en los últimos años y está cogiendo cada vez más fuerza.

Podemos ver cada vez más como los dispositivos electrónicos disponen de alguna nueva funcionalidad relacionada con el procesamiento de imágenes (Figura 1.1), como puede ser el reconocimiento facial que incorporan algunos smartphones o tablets para desbloquear el dispositivo o procesado automático de fotos realizadas por la cámara como la que incluye el terminal chino **Meitu T8** que incorpora un software llamado *AI Beautification* para embellecer las imágenes.<sup>1</sup>

---

<sup>1</sup><https://www.cnet.com/products/meitu-t8/preview/>



FIGURA 1.1: Embellecimiento de fotos (Meitu T8) (a). Reconocimiento facial (b).

Sin ir más lejos, la reciente aplicación que ha desatado el revuelo en las diferentes redes sociales; FaceApp<sup>2</sup>. La aplicación disponible tanto para Android e iOS es capaz de añadir sonrisas a las fotos, cambiar de edad o transformar el género de la persona que ha sido fotografiada.

El procesado de imágenes puede llegar a resultar muy útil en otros ámbitos como el de la medicina. Un ejemplo es la radiografía de la Figura 1.2 que partiendo de una imagen de muy baja calidad se pretende extraer información sobre las manchas blancas que aparecen en la misma.

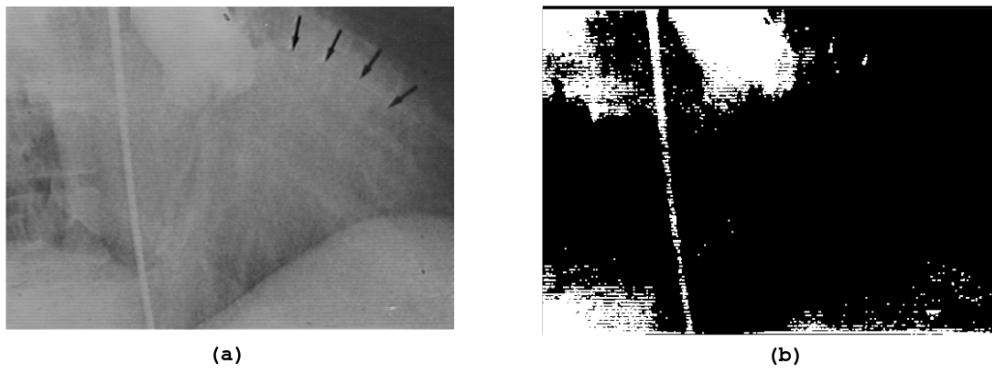


FIGURA 1.2: Radiografía inicial, con los puntos a analizar (a), imagen final procesada (b).

Son numerosas las aplicaciones de visión industrial relacionadas con el entorno de la alimentación. Permiten automatizar el control de calidad para tomar la decisión si un determinado producto cumple el estándar de calidad o no. Un ejemplo es el sistema EggInspector<sup>3</sup> por la empresa Moba que se utiliza para clasificar huevos de gallina de forma automática. El sistema está compuesto por 6 cámaras suspendidas por encima de la cinta transportadora que con unos complejos algoritmos no solo pueden comprobar si los huevos están rotos o sucios, sino que son capaces de determinar el tipo de rotura y suciedad, una vez determinada la calidad, los que no corresponden a los estándares mínimos, son separados de la línea por un robot.

<sup>2</sup><https://www.faceapp.com/>

<sup>3</sup><http://www.moba.net/page/es/Grading/Moba-Grader-Options/Detection-Systems/Egg-inspector>

Siguiendo en la línea de la industria la inspección de embalajes se ha incrementado enormemente con la automatización del proceso y la visión artificial facilitando tareas como la detección del correcto nivel de llenado, verificación de tapones, control de calidad de sellado, lectura de óptica de caracteres (OCR), códigos de barras, verificación de posición, calidad de impresión de las etiquetas, conteo de productos en cajas o *palets*. Algunas de las aplicaciones típicas de la industria del *packaging* están representadas en la Figura 1.3.



FIGURA 1.3: Presencia, aplicación e integridad de etiquetas (a), códigos 2D (b), códigos de barras (c), validación de lote, fecha y código (d), orientación de piezas montadas (e), correcto sellado (f), calidad de impresión (g), presencia y cierre de tapones (h).

En los deportes quizás la aplicación más conocida sea el Ojo de Halcón (Figura 1.4), que se utiliza en los torneos de tenis de alto nivel para determinar la trayectoria de la pelota y saber si entró o no en el campo contrario, pero la visión artificial se usa en numerosos deportes sobretodo en estudios estadísticos post-partido para averiguar el tiempo de posesión del balón en los partidos de fútbol o los kilómetros hechos por cada jugador en el terreno de juego, entre muchos otros.

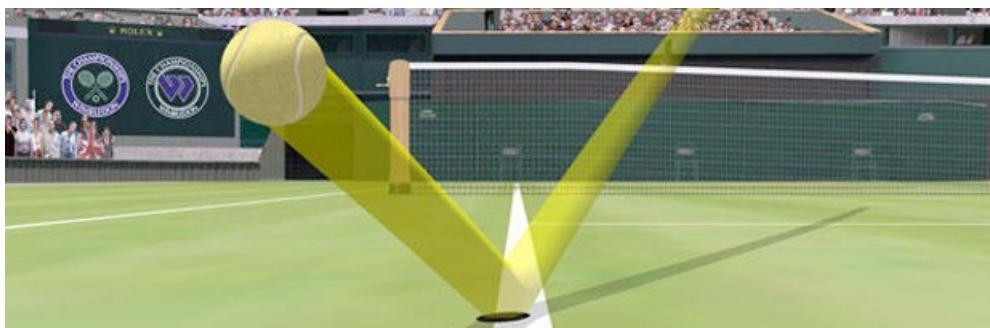


FIGURA 1.4: Ojo de Halcón en tenis

## 1.2. Autolocalización visual

La autolocalización visual consiste en conocer la localización de la cámara en todo momento simplemente con las imágenes capturadas sin disponer de ninguna información extra. Debido al gran abanico de posibilidades que propone este sistema, es uno de los retos más importantes dentro del campo de la robótica.

La autolocalización visual se plantea en los sistemas de navegación automáticos náuticos, terrestres y aéreos. Actualmente numerosas empresas están invirtiendo en este tipo de sistemas en el que apuestan por una navegación total o parcial.

Actualmente se pueden encontrar sistemas de asistencia y seguridad en los vehículos más modernos como; sistemas de frenado automático de emergencia, asistente de mantenimiento de carril o aparcado automático. Aunque estos sistemas se entienden como asistentes o ayudas a la conducción, el conductor sigue tomando la gran responsabilidad de la navegación.

La empresa israelí **Mobileye**<sup>4</sup> presentará su primer modelo de vehículo completamente autónomo, junto a Intel y BMW, en 2021. El cerebro de la máquina se basa en un sensor, que identifica lo que ocurre a su alrededor al instante: los carriles, las señales de tráfico, otros automóviles, motos, bicicletas e incluso a los peatones. En la Figura 1.5 se puede ver una captura de la vista del coche antes de parar en un semáforo.



FIGURA 1.5: Vista del coche antes de parar en un semáforo.

### 1.2.1. Realidad aumentada

La realidad aumentada es el término que se usa para definir una visión directa o indirecta de un entorno físico del mundo real, cuyos elementos se combinan con elementos virtuales generados por ordenador para la creación de una realidad mixta en tiempo real.

<sup>4</sup><https://www.mobileye.com/>

Aunque se ha popularizado con el juego de **Pokémon Go**<sup>5</sup>, cada vez son más los gigantes tecnológicos que se interesan por ella. La empresa sueca Ikea ya cuenta con una aplicación móvil que permite ver su catálogo en realidad aumentada (Figura 1.6).



FIGURA 1.6: Catálogo Ikea con realidad aumentada.

Puesto que la realidad virtual es una experiencia ficticia, tiene gran potencial en el mundo de los videojuegos. Pero no es el único. También puede tener aplicaciones en medicina, la industria del cine, la moda, los deportes o la publicidad.

### 1.3. Técnicas de autolocalización visual

Las técnicas de autolocalización ha suscitado gran interés por los investigadores en los últimos años. El problema ha sido abordado por dos comunidades distintas, por un lado la de visión artificial que denominó al problema como **structure from motion (SfM)**, donde la información es procesada por lotes, capaz de representar un objeto 2D a 3D con solo unas cuantas imágenes desde diferentes puntos de vista. Y por otro lado la comunidad robótica denominó al problema **SLAM (Simultaneous Localization and Mapping)** que trata de resolver el problema de una manera más compleja adaptando el funcionamiento de los sistemas en tiempo real.

#### 1.3.1. Structure from Motion (SfM)

Las técnicas SfM se analizan generalmente de forma offline, las escenas se graban a través de un conjunto de imágenes y luego se procesa, lo que permite realizar optimizaciones para el cálculo de la trayectoria, como por ejemplo el llamado ajuste de haces.

Existen aplicaciones comerciales que utilizan estas técnicas como es el caso de la aplicación PhotoTourism (Noah Snavely y Szeliski, 2006) desarrollada por Microsoft. Que consiste en el cálculo de la posición 3D en la que fueron captadas las imágenes,

<sup>5</sup><http://www.pokemongo.com/es-es/>

por ejemplo de un monumento, para después extraer el modelo 3D con el que el usuario puede interactuar libremente (Figura 1.7).

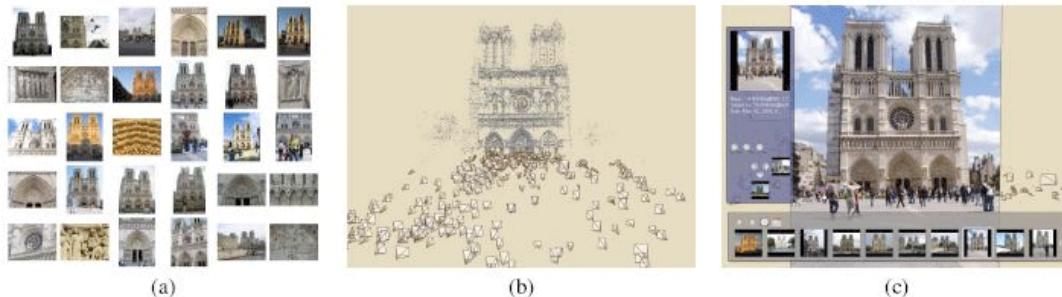


FIGURA 1.7: PhotoTourism: Se recogen una gran colección de imágenes (a), se reconstruyen los puntos 3D y los puntos de vista (b), por último la interfaz permite al usuario interactuar moviéndose a través del espacio 3D mediante la transición entre fotografías.

### 1.3.2. Visual SLAM

En el problema conocido como *Simultaneous Localization and Mapping* (SLAM) busca resolver los problemas que plantea colocar un robot móvil en un entorno y una posición desconocidas, y que él mismo se encuentre capaz de construir incrementalmente un mapa de su entorno consistente y a la vez utilizar dicho mapa para determinar su propia localización.

La solución a este problema conseguiría hacer sistemas de robots completamente autónomos que junto con un mecanismo de navegación el sistema se encontrará con la capacidad para saber a dónde desplazarse, ser capaz de encontrar obstáculos y reaccionar ante ellos de manera inteligente.

La resolución al problema SLAM ha suscitado un gran interés en el campo de la robótica y ha sido resuelto teóricamente de diversas formas como es el caso del artículo (Durrant-Whyte y Bailey, 2006). Y aunque algunas de ellas han obtenido buenos resultados en la práctica siguen surgiendo problemas a la hora de buscar el método más rápido o el que genere un mejor resultado con menos índice de fallo. La búsqueda de algoritmos y métodos que resuelvan estos problemas sigue siendo una tarea pendiente.

### Odometría visual

Dentro de las familias de técnicas pertenecientes a las de Visual SLAM se encuentra la de odometría visual, que es la que abordaremos en este trabajo. Consiste en la estimación del movimiento de la cámara en tiempo real. Es decir, el cálculo de la rotación y traslación de la cámara a partir de dos imágenes simultáneas. Se trata de una técnica incremental ya que se basa en la posición anterior para calcular la nueva.

Este tipo de algoritmos se suelen utilizar técnicas de extracción de puntos de interés, cálculos de descriptores y algoritmos para el emparejamiento. Normalmente el proceso es el mismo, una vez calculados los puntos emparejados se calcula la matriz fundamental o esencial y descomponerlas mediante SVD para obtener la matriz

de rotación y translación (RT) (Scaramuzza y Fraundorfer, 2011. Fraundorfer y Scaramuzza, 2012).

Uno de los trabajos más importantes en el ámbito es el de monoSLAM de Davison<sup>6</sup> (Andrew J. Davison y Stasse, 2007) que propone resolver este problema con una única cámara RGB como sensor y realizar el mapeado y la localización simultáneamente. El algoritmo propuesto por Davison utiliza un filtro extendido de Kalman para estimar la posición y la orientación de la cámara, así como la posición de una serie de puntos en el espacio 3D. Para determinar la posición inicial de la cámara es necesario a priori dotar de información con la posición 3D de por lo menos 3 puntos. Después el algoritmo es capaz de situar la cámara en el espacio tridimensional y de generar nuevos puntos para crear el mapa y servir como apoyo a la propia localización de la cámara. En la Figura 1.8 se pueden ver unas capturas de pantalla sobre uno de los experimentos realizados.

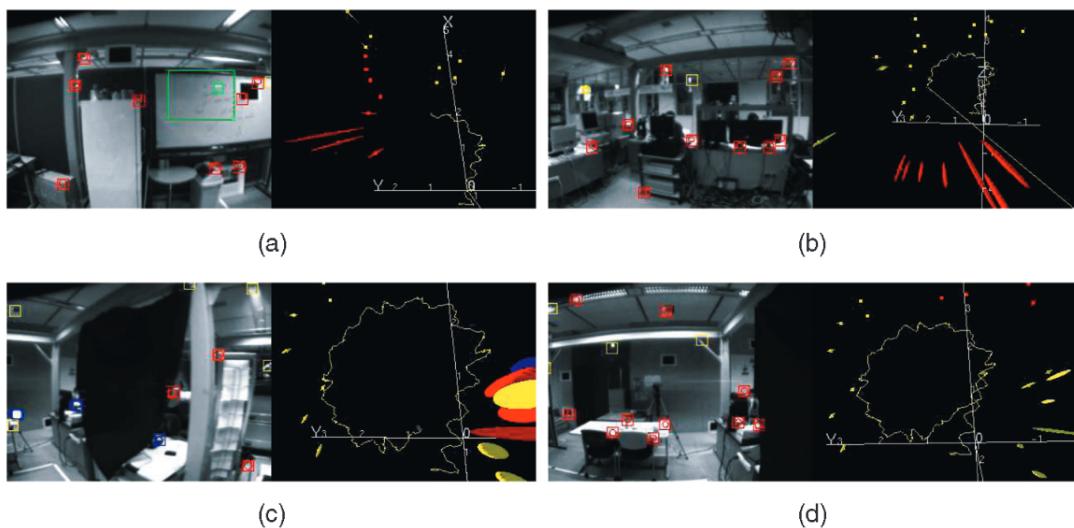
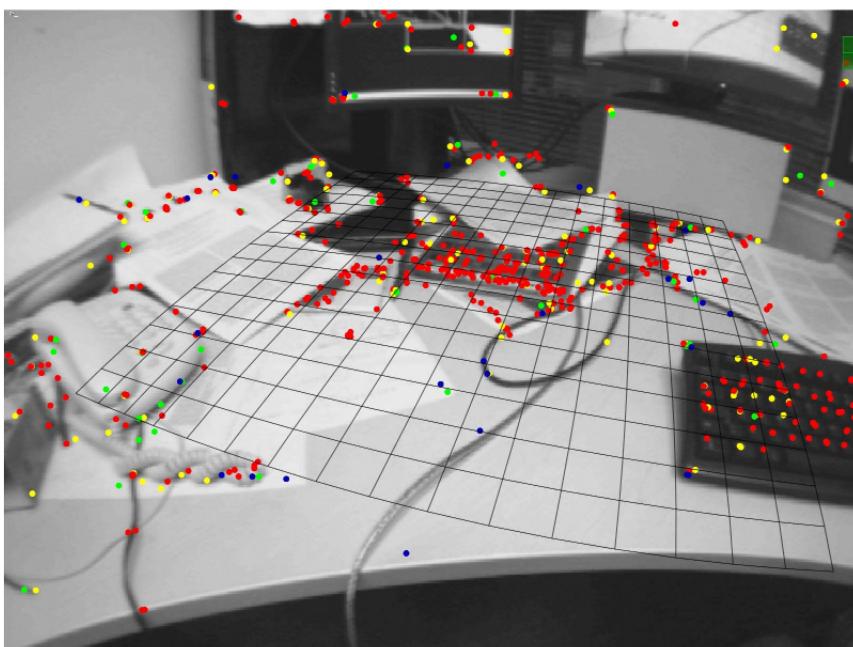


FIGURA 1.8: MonoSLAM: Un robot humanoide camina en una trayectoria circular de radio 0.75m. La estela amarilla muestra la trayectoria estimada del robot, y las elipses muestran los errores de localización.

Es importante destacar también la trascendencia que ha tenido el trabajo PTAM (Klein y Murray, 2007) que viene a solucionar uno de los principales problemas que tienen los algoritmos monoSLAM; el tiempo de cómputo, ya que aumenta exponencialmente con el número de puntos (Figura 1.9). Para ello se aborda el problema separando el mapeado de la localización, de tal modo que solo la localización deba funcionar en tiempo real, dejando así que el mapeado trabaje de una manera asíncrona. Este algoritmo parte de la idea de que solo la localización es necesaria que funcione en tiempo real. PTAM hace uso de *keyframes*, es decir, fotogramas clave que se utilizan tanto para la localización como para el mapeado y también de una técnica de optimización mediante ajuste de haces, como en SfM.

<sup>6</sup><http://www.doc.ic.ac.uk/~ajd/>



---

FIGURA 1.9: PTAM: Funcionamiento típico del sistema sobre un escritorio.

## Capítulo 2

# Objetivos

Una vez presentado el contexto, en este capítulo se abordarán los objetivos que se pretenden alcanzar. Tras una descripción del problema y unos requisitos obligatorios, se detallará la metodología y la planificación que se ha llevado a cabo en la elaboración de este trabajo.

### 2.1. Descripción del problema

El objetivo principal de este trabajo es desarrollar una solución al problema SLAM.

A través de técnicas de odometría visual incrementales, se ha construido un sistema capaz de averiguar la posición y orientación de un sensor RGBD que se mueve libremente por el espacio y que va alimentando al sistema con las imágenes obtenidas en tiempo real.

A fin de abordar el objetivo principal de la manera más simple se han propuesto los siguientes subobjetivos:

1. Actualización de los datos del sensor. Creación de un módulo encargado de recoger de manera dinámica las imágenes RGB y DEPTH del sensor, y actualizarlas en una memoria compartida.
2. Detección de puntos de interés. Creación de un componente capaz de recoger a través de diferentes técnicas (SIFT, SURF) puntos de interés de una imagen con la opción de añadir un filtro frontera para desechar los puntos cercanos a bordes o zonas con mucha profundidad.
3. Emparejamiento de puntos. Desarrollo de una funcionalidad encargada de relacionar los puntos de interés previamente obtenidos, a través de técnicas como Fuerza bruta o FLANN y la posibilidad de añadir un filtro de sobresaliente.
4. Estimación de matriz. Implementación de la matemática necesaria para una vez tenidos los puntos emparejados calcular la matriz de rotación y traslación (RT) que se necesitará para calcular en ese instante el movimiento de la cámara.
5. Pruebas y experimentos. Ejecución de la aplicación con objetivo de ajustar, pulir y validar las funciones y algoritmos realizados.

## 2.2. Requisitos

A parte de los requisitos mencionados en el apartado anterior. La implementación y solución final del proyecto debe satisfacer además los siguientes puntos:

- Desarrollo del proyecto haciendo uso de la plataforma **JDeRobot 5.4.0**, que ha resultado de mucha utilidad y ha permitido un ahorro significante de tiempo, ya que permite abstraerse de algunas de las funcionalidades de más bajo nivel como puede ser la captura de información del sensor o el protocolo de comunicaciones. Permite llevar un desarrollo modular y el aprovechamiento de componentes ya implementados en la plataforma. Tanto las ventajas como el uso de JDeRobot se tratarán en detalle en el siguiente capítulo (3).
- Como la mayoría de componentes están en C++, el trabajo también se ha desarrollado utilizando el mismo lenguaje de programación.
- Funcional bajo sistema operativo linux, en este caso Ubuntu 14.04 LTS.
- Uso exclusivo de un único sensor RGBD.
- Implementación de una interfaz de usuario clara donde se pueda apreciar el proceso de estimación de posición. Incluyendo un entorno visual 3D donde se refleje la posición y movimiento de la cámara en tiempo real.

## 2.3. Metodología

Para abordar un proyecto de tal envergadura es necesaria una metodología de desarrollo para ir progresando de una manera ordenada y efectiva. En este caso se ha optado por el modelo en espiral basado en prototipos propuesto por B. Boehm en 1986 (Boehm, 1986), ya que permite el desarrollo de una manera progresiva e incremental.

Esta metodología permite el desarrollo de implementaciones parciales que van siendo probadas a medida que después de cada ciclo se va generando un prototipo más completo de lo que incorpora el ciclo anterior. Por lo tanto, en cada ciclo o iteración se va añadiendo complejidad a la vez que se van generando funcionalidades nuevas.

Este modelo, utilizado en el trabajo, ha servido de gran ayuda, ya que permite ir avanzando de menos a más, con unos requisitos dependientes de los anteriores y a la vez diferentes para cada iteración. Además permite ir evaluando y adaptando la evolución del desarrollo a nuestros intereses, algo que suele ocurrir normalmente en los proyectos de investigación.

En la Figura 2.1 se puede observar el ciclo completo de desarrollo de software en el modelo en espiral. Cada etapa o ciclo completo está compuesto por cuatro fases:

- Identificación de objetivos. En esta primera fase se deciden y se planifican los objetivos a alcanzar en la siguiente iteración partiendo de lo realizado en el ciclo anterior. En caso de la primera iteración se definen los objetivos iniciales.
- Evaluación alternativa. Aquí se definen requisitos y se estudian las distintas maneras de abordar los objetivos marcados de la etapa anterior. Se estudian los riesgos y se evalúan de manera que se puedan reducir lo máximo posible. Se debe tener un prototipo antes de la siguiente etapa.

- Desarrollo del producto. En esta fase se diseña y se implementa el producto en base a lo planteado en las anteriores fases. Por último, se verifica y se prueba.
- Planificación de la siguiente fase. Considerando el resultado de la fase anterior, se planifica la siguiente considerando los errores cometidos y los resultados esperados, comenzando así una nueva iteración.

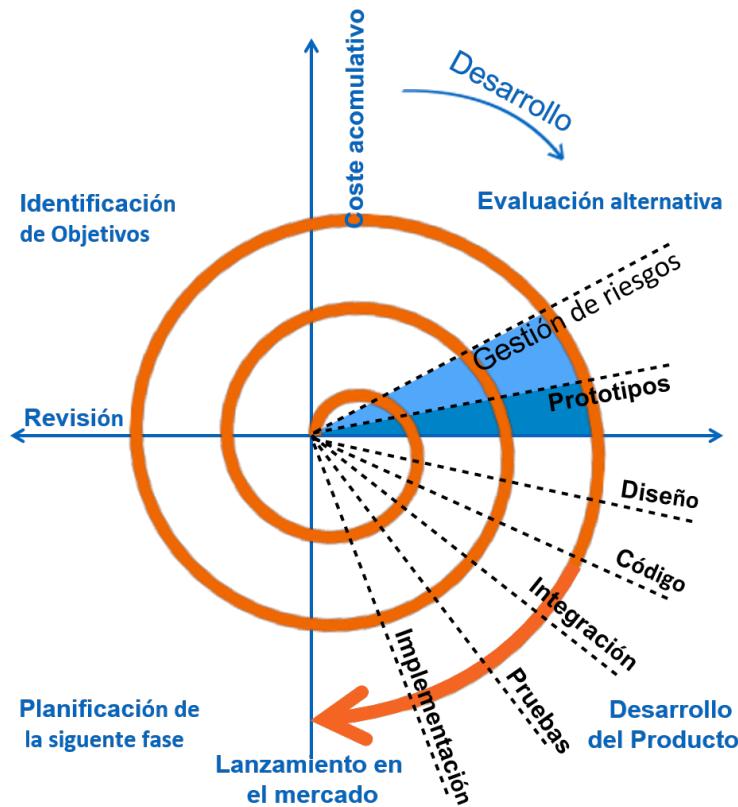


FIGURA 2.1: Ciclo de vida del desarrollo del *software* en el modelo espiral.

En la siguiente sección se detallarán los ciclos que han seguido a lo largo de este proyecto. Para las fases de planificación y análisis se han mantenido reuniones semanales con el tutor, con intención de revisar, resolver problemas y encarar los nuevos objetivos establecidos.

A fin de documentar y guardar los hitos realizados en el desarrollo del proyecto, así como los errores cometidos y su posible solución, se ha llevado un seguimiento en mediawiki<sup>1</sup> con los detalles de las diferentes iteraciones, ayudadas a veces por imágenes y/o videos.

Para la gestión de código se han usado herramientas *software* de control de versiones, primeramente con Subversion (SVN)<sup>2</sup> y finalmente con GIT en un repositorio de GitHub<sup>3</sup>.

<sup>1</sup><http://jderobot.org/J.benitod.tfg>

<sup>2</sup><http://svn.jderobot.org/users/j.benitod/pfc>

<sup>3</sup><https://github.com/RoboticsURJC-students/2014-pfc-Javier-Benito>

## 2.4. Planificación

A lo largo del trabajo se han ido proponiendo etapas asesoradas y con supervisión del tutor. Las más importantes caben destacar:

### 1. Familiarización de la herramienta JDeRobot.

Esta etapa consistió en la instalación y el estudio de la plataforma, profundizando en el uso de algún componente con un objetivo muy concreto y sencillo.

Después, y para entender el funcionamiento de algunos de los componentes a bajo nivel más importantes para el trabajo, se propuso el desarrollo de algunos de ellos en otros lenguajes de programación tales como Java o Python.

### 2. Aprendizaje de las herramientas específicas.

Aquí, a través de prácticas muy concretas se entendió el funcionamiento de algunas de las librerías esenciales para la práctica final.

- Se realizó un componente utilizando **PCL** para el cálculo de planos desde nubes de puntos.
- Se utilizó **Eigen** para otra práctica en el cálculo de sistemas sobredimensionados de ecuaciones, con descomposición QR y en valores singulares (SVD) resolviendo en diferentes casos una recta de regresión para distintos escenarios.
- **GSL**, para el cálculo de un componente rectificador de imágenes (Figura 2.1).

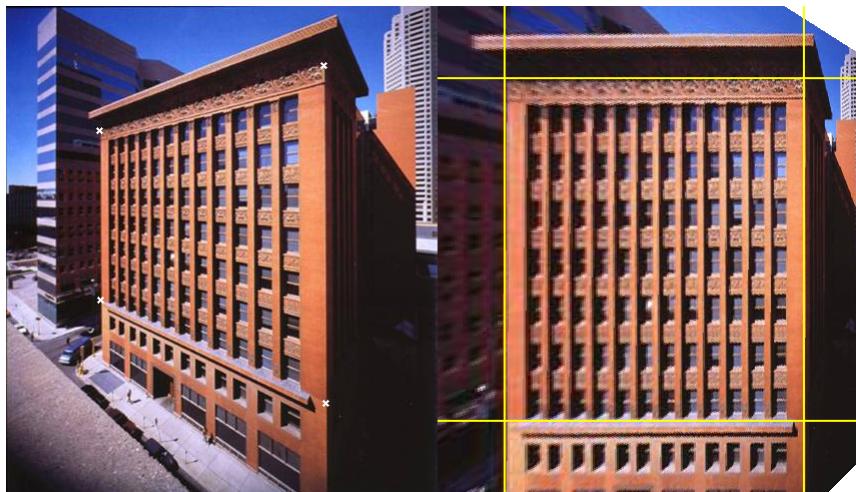


FIGURA 2.2: Componente rectificador de imágenes.

### 3. Implementación de un algoritmo para el cálculo de la matriz RT.

En esta parte se propuso una práctica que sirvió de base para la parte final en la que desde una nube de puntos y otra multiplicada por una matriz RT inventada, se sacaría a través de SVD dicha matriz a partir de las nubes de puntos iniciales y con la opción también de añadir ruido gaussiano a una de ellas. En la Figura 2.3 se encuentra el esquema realizado.

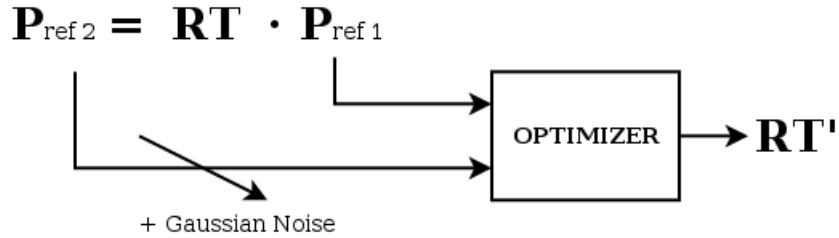


FIGURA 2.3: Esquema de la práctica de cálculo de matriz RT.

#### 4. Creación de un componente para la extracción de puntos de interés y emparejamiento.

Aquí se empezó a desarrollar la práctica final, en donde se comenzó por la extracción de puntos de interés con SIFT de las imágenes y los diferentes algoritmos para los emparejamientos. Como esta parte corresponde al desarrollo de la práctica final será detallada en el capítulo 4.

#### 5. Integración y pruebas experimentales.

Esta fue la etapa más larga y costosa, ya que el objetivo era dar una solución funcional al problema propuesto. Surgieron multitud de errores que se tuvieron que ir depurando con cada iteración, proponiendo soluciones y alternativas. Fueron necesarias numerosas pruebas para conseguir un desarrollo capaz de aportar una solución estable. Esta etapa también será detallada en los siguientes capítulos, en concreto en los capítulos 4 y 5.

## Capítulo 3

# Infraestructura

En este capítulo se detallarán las herramientas base empleadas en la realización de este trabajo.

### 3.1. Sensores RGBD

Los sensores RGBD son capaces de captar a parte de las componentes roja, verde y azul de la luz, información de profundidad (o "D" depth en inglés). Es decir, por cada píxel asocia la información de color con su correspondiente componente de profundidad. Esta tecnología fue desarrollada por la empresa israelí **PrimeSense**. El sensor Kinect dispone también de un micrófono multiarray con el cual puede predecir de dónde proviene el sonido.

En el 2010 Microsoft sacó al mercado el sensor Kinect (Figura 3.2) para la consola de juegos Xbox 360 y Xbox One. Pronto se convirtió en uno de los dispositivos electrónicos más vendidos en todo el mundo después de su lanzamiento.



FIGURA 3.1: Sensor Microsoft Kinect

Este sensor salió al mercado a un precio mucho más reducido que algunos que existían antes que él por lo que el interés por este tipo de sensores se disparó y comenzaron a aparecer en diferentes áreas de la tecnología, como interfaces naturales de usuario (en inglés natural user interface, NUI), reconstrucción y realidad virtual o cartografía 3D.

CUADRO 3.1: Especificaciones técnicas del Asus Xtion PRO LIVE

Campo de visión:	58° H, 45° V, 70° D
Distancia de uso:	Entre 0.8m y 3.5m
Tamaño de la imagen de profundidad:	VGA (640x480) : 30 fps QVGA (320x240): 60 fps
Resolución:	SXGA (1280*1024)

El sensor utilizado en este trabajo es el **Asus Xtion PRO LIVE** que dispone de la misma tecnología comercializado por Asus, que proporciona profundidad, color y audio (utilizando un micrófono multiarray como el sensor Kinect). <sup>1</sup>



FIGURA 3.2: Asus Xtion PRO LIVE

Las especificaciones técnicas de este sensor se encuentran recogidas en la tabla 3.1

### 3.2. JDeRobot

JDeRobot es un proyecto desarrollado por el grupo de robótica de la Universidad Rey Juan Carlos <sup>2</sup>. Consiste en una plataforma de desarrollo de aplicaciones robóticas y de visión artificial. Está en su mayoría escrito en C++, donde disponen de una colección de componentes capaces de comunicarse a través de ICE middleware <sup>3</sup>, los componentes pueden ejecutarse en diferentes ordenadores y pueden ser programados en diferentes lenguajes.

JdeRobot incluye numerosas herramientas, drivers, interfaces, librerías y tipos. Es software libre con licencia GPL y LGPL. También utiliza software de terceros como Gazebo, ROS, OpenGL, GTK y Eigen entre otros.

La versión de JdeRobot empleada ha sido la versión 5.4.0. A continuación se detallarán los componentes de JdeRobot que han sido de utilidad para la realización de este proyecto.

<sup>1</sup>[https://www.asus.com/3D-Sensor/Xtion\\_PRO\\_LIVE/](https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/)

<sup>2</sup><http://jderobot.org>

<sup>3</sup><https://zeroc.com/products/ice>

### 3.2.1. Biblioteca Progeo

Es una biblioteca de geometría proyectiva incluida en JdeRobot, que proporciona funciones muy útiles que relacionan puntos en dos y tres dimensiones.

Ha sido realmente útil en este trabajo para que a partir de puntos en dos dimensiones (pixeles) y su correspondiente información de profundidad (distancia), sacar los puntos relativos de la cámara en tres dimensiones.

Progeo usa el modelo de cámara **Pinhole**, en la Figura 3.3 se puede observar la representación geométrica de la retroproyección y la proyección. Este modelo es definido por unos parámetros intrínsecos y extrínsecos que definen la composición de todos los parámetros iniciales de configuración de la cámara. Los parámetros extrínsecos que establecen la posición 3D, foco de atención (foa) y roll, mientras que los parámetros intrínsecos determinan la distancia focal y el centro óptico o píxel central.

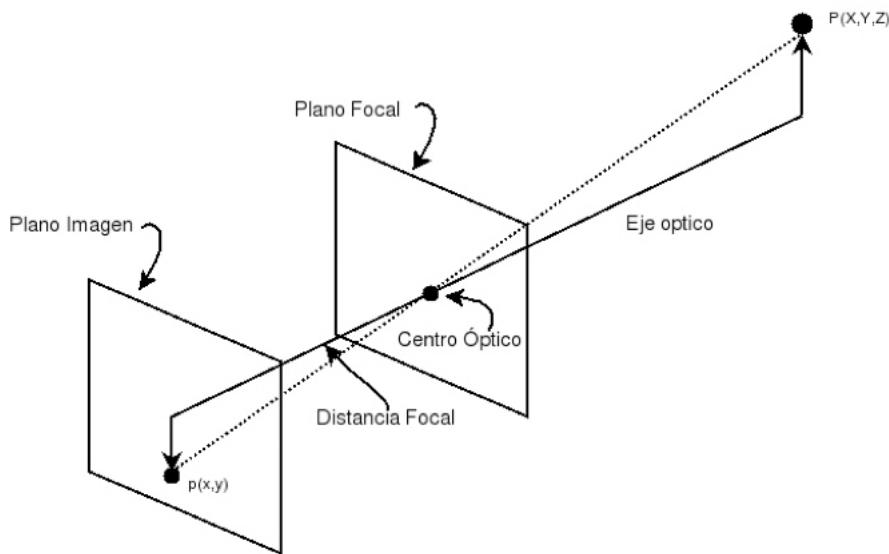


FIGURA 3.3: Modelo de cámara Pinhole

Las funciones que proporciona esta biblioteca son las siguientes:

- **Project:** Esta función permite proyectar un punto 3D del mundo al correspondiente pixel en 2D de la imagen de la cámara.
- **Backproject:** Esta función es capaz de a partir de las coordenadas de un píxel en 2D, obtener la línea de proyección que conecta la cámara y el foco con el rayo 3D que proyecta dicho píxel en el plano imagen. Con esto y conociendo la distancia real del punto 3D a calcular, se calcula las coordenadas reales del punto 3D.
- **DisplayLine:** Esta función permite conocer si una línea definida por dos puntos en 2D es visible dentro del plano imagen.
- **Display\_info:** Esta función muestra toda la información sobre la cámara utilizada.

### 3.2.2. Biblioteca parallelIce

Es otra librería incluída en JdeRobot, que soluciona el problema de latencia de información proveniente de los diferentes drivers, evitando la espera y proviniendo de un acceso asíncrono a una copia en local de las interfaces con muy bajo tiempo de procesado.

### 3.2.3. Servidor OpenniServer

OpenniServer es un driver que se comporta como un servidor y es capaz de proporcionar con un sensor RGBD (Kinect o Xtion), imágenes de color, de profundidad o nubes de puntos que son enviados a través de la interfaz ICE a un puerto específico, donde se pueden escuchar los datos. Este driver es el que se necesita para el funcionamiento de este trabajo ya que es desde donde se recogen tanto las imágenes de color (RGB) como las de profundidad (Depth) para su posterior procesado.

### 3.2.4. Herramienta RGBDViewer

Es una herramienta que permite enseñar la información proveniente de los sensores RGBD con openniServer como forma de visualización de los datos; imágenes RGB, DEPTH o nubes de puntos.

El funcionamiento corresponde a un hilo de ejecución llamado Control que se encarga de recolectar las imágenes provenientes del driver, una clase Shared para guardar y recoger los datos, y por último una clase Gui que se encargará de coger los datos (imágenes y nubes de puntos) guardados en Shared y mostrarlas.

Esta herramienta a servido como referencia para la realización de este trabajo. En la Figura 3.4 podemos ver una captura de pantalla con las diferentes visualizaciones.

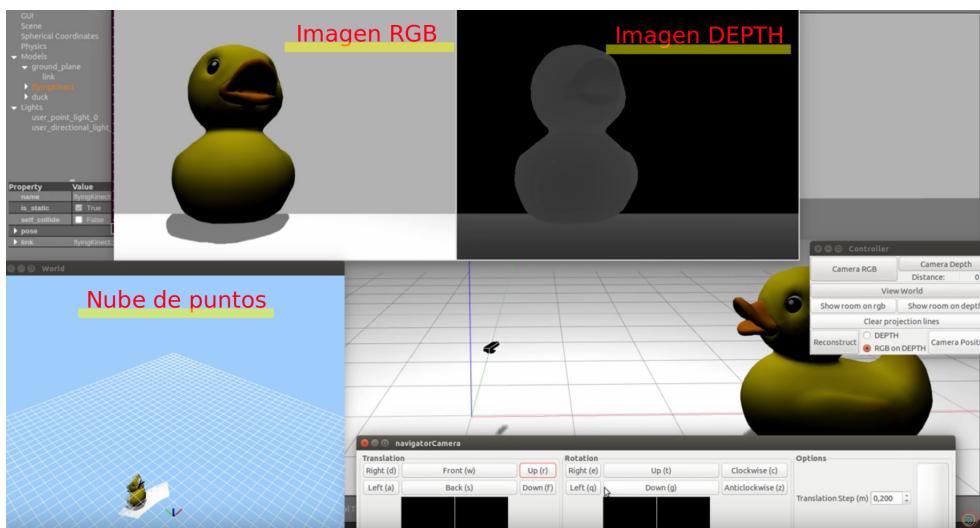


FIGURA 3.4: RGBDViewer: Captura de pantalla con las tres vistas de los diferentes datos; imagen de color, de profundidad y nube de puntos.

### 3.2.5. Pose3D

Es una interfaz que define una posición en tres dimensiones ( $x, y, z, h$ ) y una orientación con un cuaternión ( $q_0, q_1, q_2, q_3$ ).

## 3.3. Biblioteca ICE de comunicaciones

ICE (Internet Communications Engine) es un RPC framework desarrollado por Zeroc con soporte en C++, C#, Java, JavaScript y Python entre otros. Se encuentra bajo doble licencia GNU GPL y código cerrado. Actúa como plataforma de comunicaciones y funciona bajo TCP/IP.<sup>4</sup>

En JdeRobot la podemos encontrar como librería y es utilizada como protocolo de comunicaciones entre los diferentes componentes de JdeRobot. En nuestro trabajo se ha usado la versión 3.5.1 y nos ha servido para establecer la comunicación entre el componente y el driver del sensor, recogiendo las imágenes de éste.

## 3.4. Biblioteca Point Cloud Library (PCL)

PCL es una librería desarrollada en C++ para el procesamiento de imágenes 2D/3D y nubes de puntos. Está publicada con licencia BSD y libre bajo usos comerciales y de investigación. Está financialmente soportada por un consorcio de compañías comerciales y su propia organización sin ánimo de lucro, **Open Perception**. A parte de los donadores y contribuidores individuales que aportan al proyecto.<sup>5</sup>

Para simplificar el uso y el desarrollo, esta librería se encuentra dividida en módulos individuales de los que destacan el filtrado de puntos *outliers* o de ruido, estructuras de datos, estimación 3D, algoritmos para la detección de puntos de interés, combinación, segmentación, algoritmos para el reconocimiento de objetos.

En su página web disponen de mucha información y ejemplos prácticos que ayudan mucho a la compresión de todas las funcionalidades de esta librería. PCL también dispone de una librería i/o de entrada y salida para leer o crear nubes de puntos a partir de diferentes dispositivos, así como visualizadores 3D.

## 3.5. Biblioteca OpenCV

OpenCV (Open Source Computer Vision Library) es una librería de código abierto que fue desarrollada para proporcionar una infraestructura común en aplicaciones de visión artificial y facilitar la inteligencia máquina, con mecanismos de aprendizaje y de interpretación de datos. Con licencia BSD da facilidades para su uso y su modificación bajo fines comerciales.<sup>6</sup>

---

<sup>4</sup><https://zeroc.com/products/ice>

<sup>5</sup><http://pointclouds.org/>

<sup>6</sup><http://opencv.org/>

La librería contiene más de 2500 algoritmos. Estos algoritmos pueden ser usados para detectar y reconocer rostros, identificar objetos, clasificar acciones humanas determinadas en videos, seguimiento del movimiento de cámaras, seguimiento de objetos, extraer modelos de objetos 3D, producir nubes de puntos a partir de cámaras, encontrar imágenes similares de un conjunto, juntar trozos de imágenes para producir una imagen final con más resolución, etc... OpenCV tiene más de 47 miles de usuarios en la comunidad, excediendo los 14 millones de descargas, la librería es usada ampliamente en empresas, grupos de investigación y organismos gubernamentales.

OpenCV a sido diseñada de forma eficiente y con un fuerte enfoque en aplicaciones de tiempo real. Escrita en C/C++, la librería obtiene las ventajas del procesamiento multi-núcleo. Dispone de interfaces en C++, C, Python, Java y MATLAB y es soportada por diferentes sistemas operativos como Windows, Linux, Android y Mac OS.

En este trabajo se ha utilizado la versión 2.4.8 y se ha usado a la hora de identificar puntos de interés y para los distintos métodos de emparejamiento. En la Figura 3.5 se puede apreciar un ejemplo de su uso en la detección de puntos de interés sobre un entorno real.

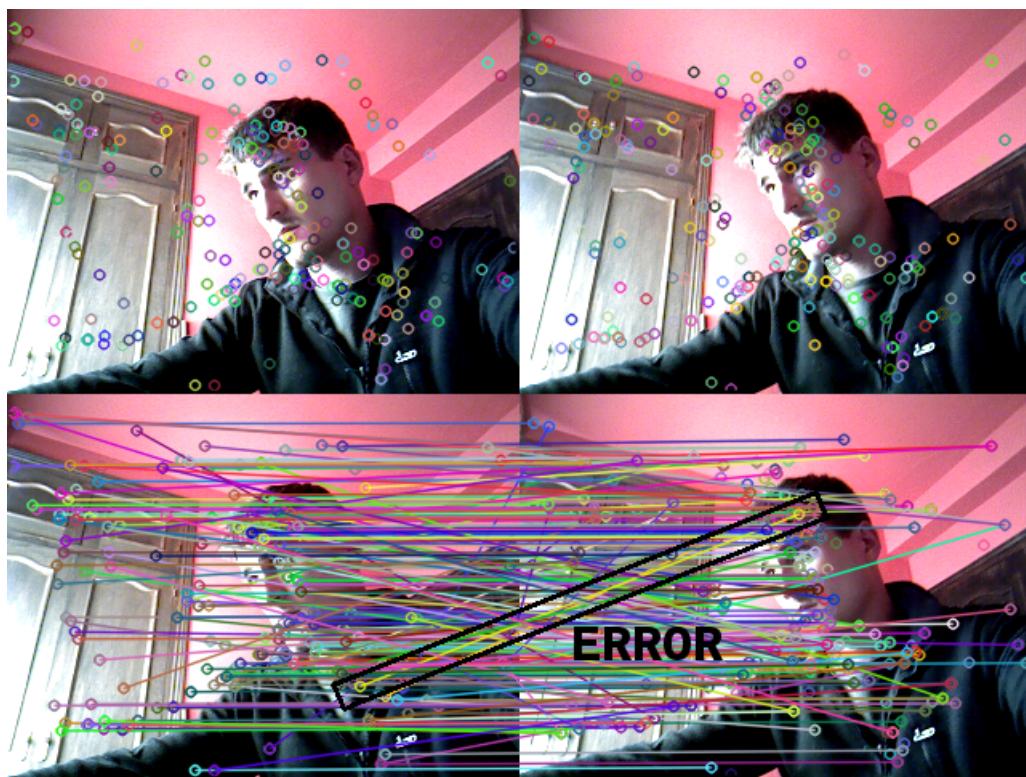


FIGURA 3.5: Detección y emparejamiento de puntos de interés con OpenCV usando SIFT.

### 3.6. Biblioteca Eigen

Eigen es una librería de álgebra lineal que permite hacer operaciones aritméticas con matrices y vectores, a través de los operadores comunes de C++, tales como +, -, \*,

o a través de métodos especiales tales como `dot()`, `cross()`, etc... Para la clase `Matrix` (matrices y vectores) los operadores solo soportan operaciones de álgebra lineal. En Figura 3.6 se puede ver un ejemplo de como es simple hacer una multiplicación y una división por un escalar.<sup>7</sup>

Example:	Output:
<pre>#include &lt;iostream&gt; #include &lt;Eigen/Dense&gt;  using namespace Eigen;  int main() {     Matrix2d a;     a &lt;&lt; 1, 2,         3, 4;     Vector3d v(1,2,3);     std::cout &lt;&lt; "a * 2.5 =\n" &lt;&lt; a * 2.5 &lt;&lt; std::endl;     std::cout &lt;&lt; "0.1 * v =\n" &lt;&lt; 0.1 * v &lt;&lt; std::endl;     std::cout &lt;&lt; "Doing v *= 2;" &lt;&lt; std::endl;     v *= 2;     std::cout &lt;&lt; "Now v =\n" &lt;&lt; v &lt;&lt; std::endl; }</pre>	<pre>a * 2.5 = 2.5   5 7.5   10 0.1 * v = 0.1 0.2 0.3 Doing v *= 2; Now v = 2 4 6</pre>

FIGURA 3.6: Ejemplo de una multiplicación y división por un escalar con Eigen.

Eigen es *software* libre y desde la versión 3.1.1 tiene licencia MPL2 (GPL3+ para las anteriores versiones). Se ha usado la versión 3.2.0 y ha sido de utilidad en este proyecto para realizar los cálculos de la matriz RT a través de los vectores de puntos 3D ya emparejados.

### 3.7. Biblioteca de interfaz gráfica GTK+

GTK+, o the GIMP Toolkit es una herramienta multiplataforma de creación de interfaces gráficas. Es multiplataforma y está escrito en C, pero a sido diseñado para tener soporte para un gran rango de lenguajes, tales como Perl y Python. GTK++ tiene una gran colección de *widgets* y interfaces para usar en la aplicación, tales como ventanas, botones, selectores, cajas de texto, etc.

La versión utilizada ha sido la 3.10.8. Es *software* libre y parte del proyecto GNU. Con licencia LGPL, permite que sea utilizado por todos los desarrolladores, incluyendo aquellos que están desarrollando un *software* privativo. GTK+ ha sido utilizado en muchos proyectos y en grandes plataformas.<sup>8</sup>

#### 3.7.1. Glade

Glade es una *RAD tool* (Rapid Application Development Tool) que permite desarrollar de manera fácil y rápida interfaces de usuario en GTK+ para el entorno de escritorio GNOME. La interfaz gráfica diseñada en Glade es guardada en un XML

<sup>7</sup><http://eigen.tuxfamily.org/>

<sup>8</sup><https://www.gtk.org/>

que usando los objetos GTK+ de **GtkBuilder** pueden ser cargados y utilizados por aplicaciones de forma dinámica como se ha hecho en este trabajo.<sup>9</sup>

### 3.8. OpenGL

OpenGL es el principal entorno para el desarrollo de aplicaciones gráficas 2D y 3D interactivas. Desde 1992, OpenGL se ha convertido en la interfaz de aplicaciones gráficas más utilizada y soportada en la industria 2D y 3D, con miles de aplicaciones disponibles en diferentes plataformas. OpenGL ayuda al desarrollo de aplicaciones al incorporar un amplio conjunto de renderizado, mapeo de texturas, efectos especiales y otras potentes funciones de visualización. Se puede usar OpenGL en la mayoría de entornos de escritorio y diferentes plataformas. Es muy utilizada y conocida en la industria de los videojuegos.

Algunas de las ventajas de las que presume OpenGL son; que es un estándar de la industria, con soporte, multiplataforma y el único libre. Es estable, dispone de compatibilidad hacia atrás, escalable, fácil de usar y bien documentado.<sup>10</sup>

Se ha usado la librería **Mesa 3D Graphics** en linux que es una implementación de la especificación de OpenGL con código abierto<sup>11</sup>.

OpenGL en este trabajo se ha usado para visualizar la posición de la cámara, su estela y la colección de nubes de puntos obtenida y procesada de las imágenes RGB y de profundidad. En la Figura 3.7 se puede apreciar una captura de pantalla con la posición de la cámara dibujada en el espacio tridimensional con el visualizador utilizado con OpenGL.

---

<sup>9</sup><https://glade.gnome.org/>

<sup>10</sup><https://www.opengl.org/>

<sup>11</sup><https://www.mesa3d.org/>

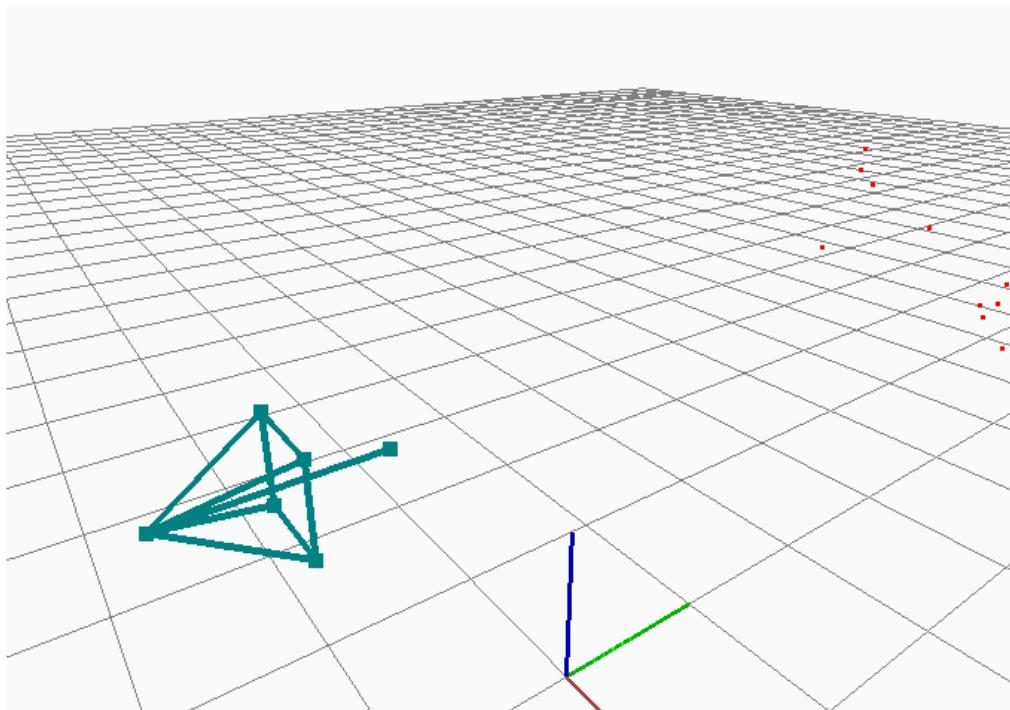


FIGURA 3.7: Captura de la posición de la cámara en el visualizador 3D con OpenGL.

## Capítulo 4

# Desarrollo

Una vez presentado el contexto, los objetivos, así como las herramientas empleadas y los fundamentos teóricos, en este capítulo se detallará la solución software final desarrollada. Primero se presenta el diseño global utilizado y después se analizará en detalle el componente en cuestión realizado con una visión profunda del desarrollo por bloques y su funcionamiento.

### 4.1. Diseño

El trabajo se basa principalmente en dos componentes; un componente de JDeRobot (**OpenniServer**) que funciona como driver del sensor y proporciona las imágenes obtenidas por éste y el componente realizado (**RealRTEstimator**) que se encargará, una vez recogidas las imágenes, de toda la lógica restante.

El objetivo del componente, como ya se ha comentado, consiste en analizar en tiempo real la posición y movimiento del sensor, por lo que el componente deberá dar una estimación en todo momento.

En la Figura 4.1 se puede apreciar el diagrama global de funcionamiento del componente desarrollado y su conexión con otros componentes para los diferentes datos de entrada.

OpenniServer se encarga de preparar y enviar las imágenes del sensor. El componente recoge las imágenes a través de ICE y éste es el encargado de procesarlas. También recibe los datos de los parámetros intrínsecos de la cámara así como algunos parámetros de configuración, como pueden ser la activación/desactivación de la interfaz de usuario o algunos parámetros configurables de algunos de los algoritmos internos. A su salida entrega una matriz RT que describe la posición y orientación absolutas en ese preciso instante de tiempo.

Respecto al funcionamiento interno del componente se puede ver a grandes rasgos el diagrama en la Figura 4.2. Se observa el diseño implementado así como sus bloques funcionales:

- Extracción de puntos de interés (análisis 2D) del fotograma actual.
- Emparejamientos de puntos de interés en  $t$  con respecto a los puntos extraídos en el instante anterior ( $t-1$ ).
- Transformación de puntos (pixeles) en 2D más imagen de profundidad a nube de puntos en 3D.

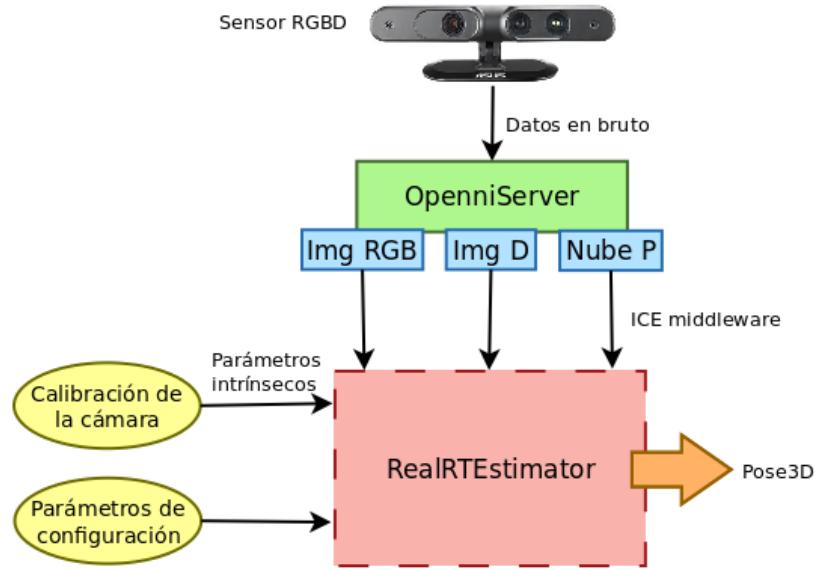


FIGURA 4.1: Esquema global de funcionamiento.

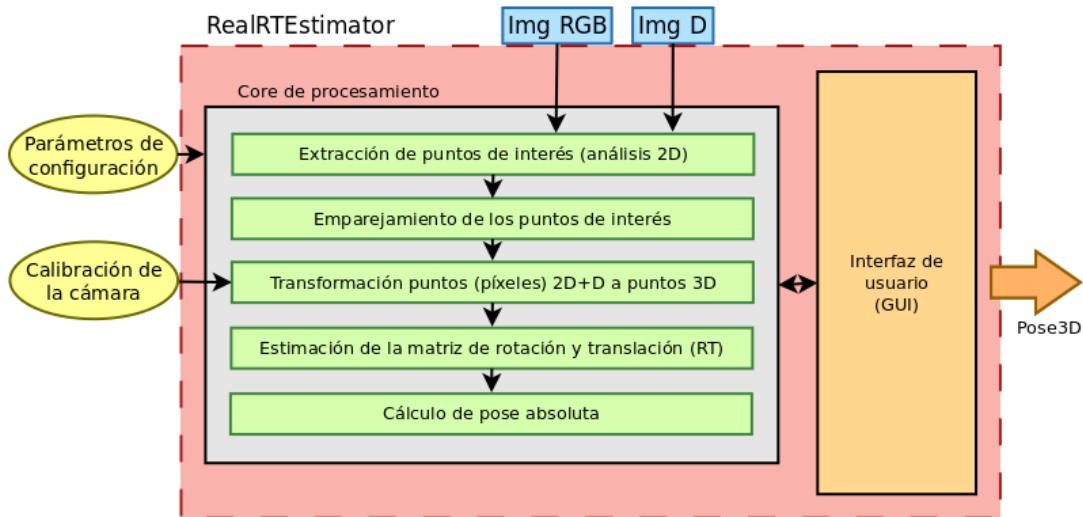


FIGURA 4.2: Diagrama del componente RealRTEstimator.

- Cálculo de movimiento. Es decir, estimación de la matriz de rotación y translación (Matriz RT).
- Cálculo de pose 3D absoluta.

En las siguientes secciones desglosaremos el funcionamiento de estos diferentes bloques funcionales.

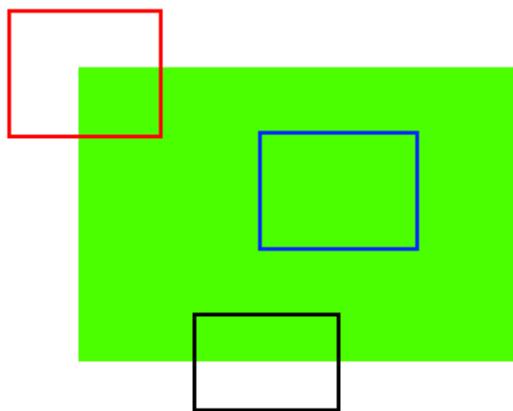
## 4.2. Análisis 2D

El primer bloque del componente RealRTEstimator es el de análisis 2D. A partir de dos imágenes; la imagen de color y la de profundidad, se procede a la extracción de puntos de interés.

### 4.2.1. Detección de puntos de interés

El término puntos de interés o detección de características (*Feature Detection* en inglés) hace referencia a la tarea de localizar en una imagen puntos relevantes o característicos. Estos puntos suelen ser comunes y son fáciles de seguir de fotograma en fotograma.

Para entender cuales son estos puntos característicos podemos observar un ejemplo sencillo en la Figura 4.3. El cuadrado azul se encuentra en una área plana, y es difícil de seguir o encontrar. En cualquier lugar por donde se desplace parecerá que es el mismo. Para el cuadrado negro, que es un borde, igual para el desplazamiento lateral, sin embargo, para el desplazamiento vertical el punto ya cambia. Por último está el cuadrado rojo, que es una esquina. Para cualquier desplazamiento de esta figura, el punto ya es diferente, lo que significa que ese punto en la figura es único y por lo tanto vamos a poder identificarlo o seguirlo en diferentes imágenes. Así pues, las esquinas suelen ser candidatos idóneos para la detección puntos de características en una imagen (en algunos casos las manchas también pueden ser consideradas buenas zonas).




---

FIGURA 4.3: Ejemplo sencillo de puntos característicos.

Una vez entendido el concepto, el siguiente paso consiste, en averiguar cómo encontrar estos puntos de interés en una imagen real. Por ejemplo, una manera sencilla de hacerlo es buscar las regiones en las imágenes que contienen una gran variabilidad cuando son desplazadas (una pequeña distancia) hacia todas las direcciones de los alrededores.

Existen multitud de implementaciones para calcular estas características en las imágenes. Uno de los primeros intentos en encontrar estas esquinas fue hecho por Chris Harris y Mike Stephens (Harris y Stephens, 1988). El método, llamado *Harris Corner*

*Detector* transforma la simple idea a una fórmula matemática (4.1) que básicamente encuentra la diferencia en intensidad por un desplazamiento  $(u,v)$  en todas las direcciones.

$$E(u, v) = \sum_{x,y} w(x, y) [I(x + u, y + v) - I(x, y)]^2 \quad (4.1)$$

Donde  $w(x,y)$  es una ventana rectangular o gaussiana e  $I(x,y)$  corresponde a la intensidad. Aplicando algunos cálculos matemáticos que no vamos a entrar en detalle podemos llegar a la ecuación (4.1) básica que determina si una ventana contiene una esquina o no.

$$\begin{aligned} R &= \det(m) - k(\text{trace}(M))^2 \\ R &= \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \end{aligned} \quad (4.2)$$

$\lambda_1$  y  $\lambda_2$  son los autovalores de la matriz  $M$ , que determinarán si una región es esquina, borde o zona plana.

- Cuando  $|R|$  es pequeño, que sucede cuando  $\lambda_1$  y  $\lambda_2$  son pequeños, la región es plana.
- Cuando  $R < 0$ , que sucede cuando  $\lambda_1 \gg \lambda_2$  o viceversa, la región es un borde.
- Cuando  $R$  es grande, que sucede cuando  $\lambda_1$  y  $\lambda_2$  son grandes y más o menos iguales, la sección es una esquina.

Más tarde, J. Shi y C. Tomasi hicieron una pequeña modificación que obtuvo mejores resultados comparados con los obtenidos en el detector de Harris (Shi y Tomasi, 1994). El resultado del detector *Shi-Tomasi Corner Detector* se puede ver en la ecuación (4.3)

$$R = \min(\lambda_1, \lambda_2) \quad (4.3)$$

Si  $R$  es mayor que un determinado umbral, o dicho de otro modo; solo cuando  $\lambda_1$  o  $\lambda_2$  se encuentran por encima de un valor mínimo  $\lambda_{\min}$ , se considera que cierta región es esquina. En la Figura 4.4 se puede observar el resultado de aplicar dicho algoritmo en una imagen.

Existen varias implementaciones para el cálculo de características de una imagen. A parte de las mencionadas, OpenCV proporciona entre otras **SIFT** y **SURF** que son las que hemos usado para el trabajo ya que permiten además de la detección de puntos de interés, el cálculo de descriptores.

#### 4.2.2. Cálculo de descriptores

Una vez que se conoce el punto de interés, necesitamos asignarle una huella, algo característico que nos permita encontrar el mismo en otra imagen. Para ello, se procede al cálculo de descriptores (*Feature Description* en inglés).

Consiste en definir la región alrededor del punto de interés para poder buscar el punto con la misma región en otra imagen. Es decir, se guarda una descripción de

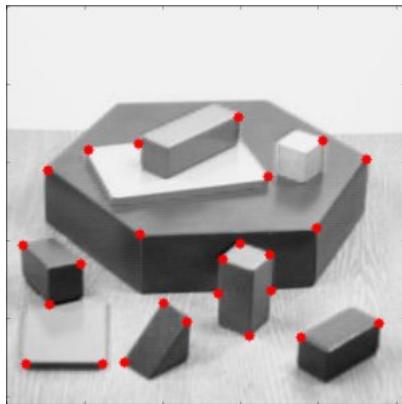


FIGURA 4.4: Resultado de encontrar las mejores 25 esquinas de la imagen con *Shi-Tomasi Corner Detector*.

la región del punto dado y se busca el mismo (o el que más se parezca) a otro punto perteneciente a otra imagen.

Una vez localizado el punto se podrá llevar un seguimiento de dónde está ese punto en otra imagen. No en todos los casos, se va a encontrar un descriptor perfecto para un cierto punto, por lo que al estudiar los emparejamientos se evaluará cuanto se parecen los descriptores entre sí. Esto lo veremos con detalle en la siguiente sección.

### SIFT

SIFT (Scale-Invariant Feature Transform) soluciona uno de los problemas que nos encontrábamos en los métodos anteriormente mencionados. Los métodos hasta ahora vistos para el cálculo de puntos de interés o esquinas, se suponen invariantes a la rotación, es decir, incluso si la imagen es rotada es posible encontrar las mismas esquinas. Esto es así porque una esquina sigue siendo una esquina si la imagen a sido rotada. Sin embargo, no contemplan los cambios de escala, un esquina no puede ser una esquina si la imagen ha sido escalada. En la Figura 4.5 podemos ver un ejemplo de este hecho; una esquina en una pequeña imagen con una ventana no lo es cuando la imagen se amplia y se usa la misma ventana.

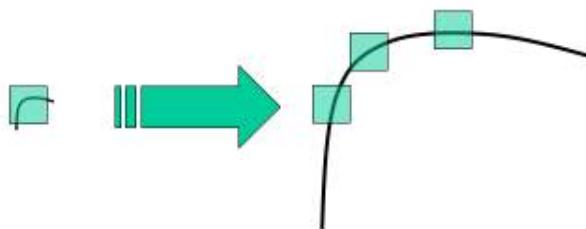


FIGURA 4.5: Ejemplo de diferentes puntos con escala

SIFT nace para proveer esta carencia de mano de D. Lowe (D.Lowe, 2004). Un algoritmo escalarmente invariante que localiza puntos de interés y calcula descriptores. Hay principalmente 4 etapas básicas en el algoritmo de SIFT:

1. Extrema detección en espacio-escala

Para poder detectar características en diferentes escalas es necesario poder variar el tamaño de la ventana a ampliar. Para ello se utiliza un filtro de espacio-escala; un filtro LoG (Laplaciana de una Gaussiana) que con diferentes valores de  $\sigma$  es capaz de detectar puntos de interés para diferentes escalas.  $\sigma$  actúa como un parámetro de escala, por ejemplo como se puede ver en la FIGURA, para bajos niveles de  $\sigma$  la gaussiana devuelve altos valores para las pequeñas esquinas, sin embargo, altos valores de  $\sigma$  encajan bien para grandes esquinas.

Por lo tanto se busca a lo largo de la imagen y en diferentes escalas para encontrar el punto

Así pues a lo largo de la imagen y en diferentes escalas tenemos una lista de  $(x, y, \sigma)$  valores, donde  $(x, y)$  representa el espacio y  $\sigma$  la escala.

Sin embargo, el filtro LoG es muy costoso por lo que SIFT calcula una aproximación; la diferencia de gausianas con diferente  $\sigma$  y el proceso se repetirá para diferentes octavas ( $k\sigma$ ) como se puede apreciar en la Figura 4.6.

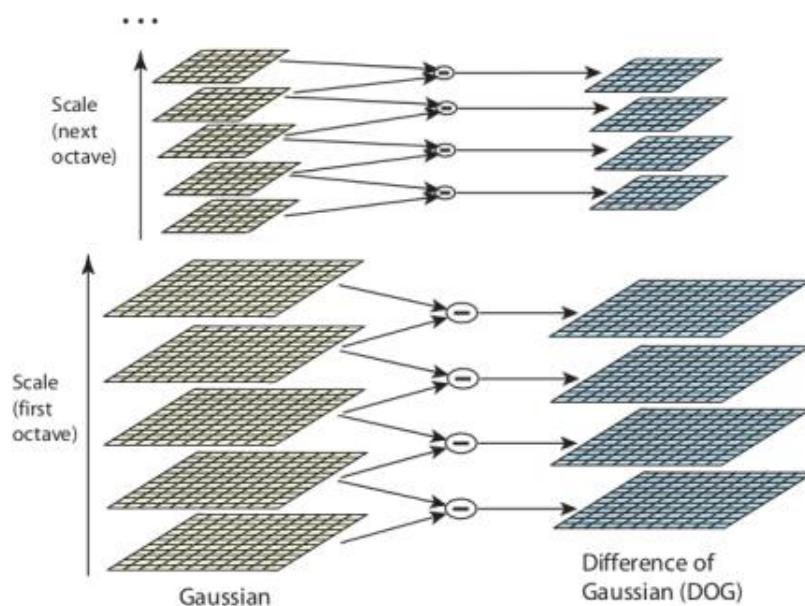


FIGURA 4.6: Proceso del cálculo de la diferencia de Gaussianas para diferentes octavas.

Una vez obtenida la diferencia de gaussianas (DOG) se calcula el local-extrema, por ejemplo, un pixel en una imagen es comparado con sus 8 vecinos y también con los 9 píxeles en la escala anterior y la posterior

2. Localización de puntos de interés  
bla bla balb alba
3. Asignación de orientación
4. Descriptor del punto de interés
5. Emparejamiento de puntos

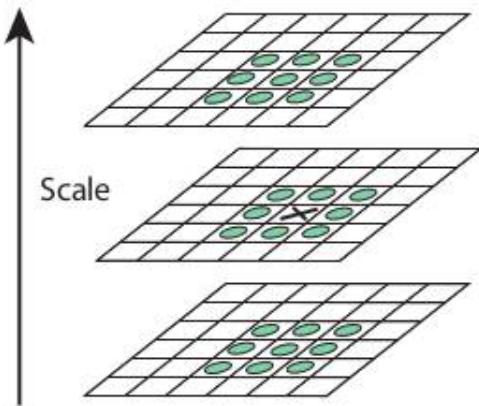


FIGURA 4.7: Proceso del cálculo de la diferencia de Gaussianas para diferentes octavas.

**SURF**

#### 4.3. Emparejamiento (*matching*)

#### 4.4. Cálculo de movimiento

##### 4.4.1. Matriz RT

#### 4.5. Interfaz gráfica

## Capítulo 5

# Experimentos

# Bibliografía

- Andrew J. Davison Ian D. Reid, Nicholas D. Molton y Olivier Stasse (2007). «MonoSLAM: Real-Time Single Camera SLAM». En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(6), págs. 1052-1067.
- Boehm, B. (1986). «A spiral model of software development and enhancement». En: *ACM SIGSOFT Software Engineering Notes* 11(4), págs. 14-24.
- D.Lowe (2004). «Distinctive Image Features from Scale-Invariant Keypoints». En: *University of British Columbia*.
- Durrant-Whyte, H. y T. Bailey (2006). «Simultaneous localization and mapping: part I». En: *IEEE Robotics Automation Magazine* 13, issue 2, págs. 99-110.
- Fraundorfer, Friedrich y Davide Scaramuzza (2012). «Visual odometry: Part ii: Matching, robustness, optimization, and applications». En: *Robotics Automation Magazine, IEEE* 19(2), págs. 78-90.
- Harris, Chris y Mike Stephens (1988). «A Combined Corner and Edge Detector». En: *Alvey vision conference*. URL: <http://courses.daiict.ac.in/>.
- Klein, G. y D. Murray (2007). «Parallel Tracking and Mapping for Small AR Workspaces». En: *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, págs. 225-234.
- Noah Snavely, Steven M. Seitz y Richard Szeliski (2006). «Photo tourism: Exploring photo collections in 3d». En: *SIGGRAPH Conference Proceedings*, págs. 835-846.
- Scaramuzza, Davide y Friedrich Fraundorfer (2011). «Visual odometry [tutorial]». En: *Robotics Automation Magazine, IEEE* 18(4), págs. 80-92.
- Shi, J. y C. Tomasi (1994). «Good Features to Track». En: *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*.