

GRADO EN INGENIERÍA EN TECNOLOGÍA INDUSTRIAL

## TRABAJO FIN DE GRADO

# GENERACIÓN DE TRAYECTORIA DE UN OBJETO EN UN ENTORNO 3D MEDIANTE VISIÓN ARTIFICIAL

*DOCUMENTO - MEMORIA, METODOLOGÍA, ASPECTOS ECONÓMICOS Y CONCLUSIONES*

**Alumno** Uribe Sáenz de Cámara, Sergio

**Director:** Eguiraun Martinez, Harkaitz

**Curso:** 2017-2018

**Fecha:** 12 de Julio de 2018



- **Resumen:** En el presente trabajo se ha llevado a cabo un estudio acerca del tracking de un objeto en movimiento en un entorno tridimensional. El objetivo del estudio es la búsqueda de un método simple y eficaz de realizar el seguimiento del objeto. El estudio se realizará mediante el procedimiento de visión estereoscópica y para ello se grabará mediante dos cámaras el movimiento del objeto. El resultado de este trabajo permitirá la posibilidad de generar la trayectoria de cualquier objeto en movimiento con un coste computacional mínimo.

**Palabras clave:** visión artificial, visión estereoscópica, generación de trayectoria, calibración estereoscópica, procesamiento de imágenes.

- **Laburpena:** Lan honetan tridimentsionala den ingurune batean higiduretan dagoen objektu baten tracking-ari buruzko ikerketa egin da. Ikerketaren helburua da aipaturiko jarraipena egiteko modu simple eta eraginkorra bilatzea. Ikerketa ikusmen estereoskopikoko prozeduaren bitartez egingo da eta hartarako objektuaren mugimendua bi kamararen bitartez grabatuko da. Lan honen emaitzak higiduraz dagoen edozein objekturen ibilbidearen aukera ahalbidetuko du, kostu konputazionalik txikienarekin.

**Gako-hitzak:** ikusmen artifiziala, ikusmen estereoskopikoa, ibilbide-sorkuntza, kalibratio estereoskopikoa, irudi-prozesatzea.

- **Abstract:** In the present work a study was carried out about the tracking of a moving object in a three-dimensional environment. The objective of the study is the search for a simple and effective method of conducting such tracking. The study will be carried out through the stereoscopic vision procedure and the movement of the object will be recorded by two cameras. The result of this work will allow the possibility of generating the trajectory of any moving object with a minimum computational cost.

**Key words:** artificial vision, stereoscopic vision, generation of a trajectory, stereoscopic calibration, image processing.

## **INDICE DE CONTENIDOS**

1. INTRODUCCIÓN.....	7
2. CONTEXTO .....	9
3. OBJETIVOS Y ALCANCE .....	10
4. BENEFICIOS QUE APORTA EL TRABAJO.....	11
5. ESTADO DEL ARTE .....	12
5.1. INTRODUCCIÓN.....	12
5.2. EL MODELO PINHOLE .....	15
5.3. PASOS EN EL PROCESO DE LA VISIÓN ESTEREOSCÓPICA.....	18
6. APLICACIÓN SELECCIONADA.....	23
7. EQUIPO UTILIZADO PARA EL DESARROLLO DEL ALGORITMO .....	24
8. PASOS SEGUIDOS EN EL DESARROLLO DEL ALGORITMO.....	26
8.1. CALIBRACIÓN ESTEREOSCÓPICA DE LAS CÁMARAS.....	29
8.2. ALGORITMO PARA LA DETECCIÓN DEL OBJETO Y EL TRACKING.....	32
9. ÁLBUM FOTOGRÁFICO DEL VIDEO.....	40
10. RESULTADOS OBTENIDOS .....	42
11. CONCLUSIONES .....	43
12. LIMITACIONES Y PROBLEMAS DEL ESTUDIO. SOLUCIONES ADOPTADAS .....	44
13. PLANIFICACIÓN.....	46
13.1. DESCRIPCIÓN DE TAREAS, FASES Y PROCEDIMIENTOS.....	46
13.2. DIAGRAMA GANTT .....	49
14. PRESUPUESTO .....	50
BIBLIOGRAFÍA .....	51

## **INDICE DE FIGURAS**

Figura 1: Efecto de la separación entre ojos/cámaras.....	13
Figura 2: (a) Sistema de visión humano.; (b) Superposición de las imágenes de ambos ojos .....	13
Figura 3: Modelo PinHole.....	15
Figura 4: (a) Geometría no epipolar; (b) Geometría epipolar.....	16
Figura 5: Geometría epipolar con ejes ópticos paralelos ( $Z_I$ y $Z_D$ ) y ejes de abscisas coincidentes. ....	16
Figura 6: Geometría epipolar en planta.....	20
Figura 7: Modelo de cámara GoPro Hero 3.....	24
Figura 8: Mando inalámbrico.....	24
Figura 9: Pórtatil Dell Inspiron15 7000.....	25
Figura 10: Diagrama de bloques del proceso seguido.....	27
Figura 11: Disposición de las cámaras.....	28
Figura 12: Stereo Camera Calibrator.....	29
Figura 13: Detección de los puntos principales del tablero.....	30
Figura 14: Ángulo máximo de inclinación.....	30
Figura 15: Error medio producido en la calibración.....	31
Figura 16: Diagrama de bloques del 1 <sup>er</sup> script.....	32
Figura 17: Primer frame del video. 1 <sup>a</sup> fase: alejamiento de las cámaras.....	33
Figura 18: Frame intermedio. Cambio de sentido: giro.....	33
Figura 19: Último frame. Final de la 2 <sup>a</sup> fase: acercamiento a las cámaras.....	34
Figura 20: Diagrama de bloques 2 <sup>o</sup> script.....	34
Figura 21: Resta de cada imagen con el fondo.....	35
Figura 22: Resta en blanco y negro.....	36
Figura 23: Imagen con cuadrados de tamaño: 1, 3, 5, 7, 9 y 15 pixeles.....	36
Figura 24: Imagen resultado de aplicar (a) Erosión. (b) Dilatación.....	37
Figura 25: Diagrama de bloques 3er script.....	38
Figura 26: (a) Ejes de MATLAB. (b) Ejes empleados en el algoritmo.....	39
Figura 27: Toma fotográfica del objeto de interés.....	40
Figura 28: Toma fotográfica en blanco y negro.....	41
Figura 29: Reconstrucción 3D.....	42
Figura 30: Reconstrucción 3D en planta.....	42
Figura 31: Diagrama Gantt.....	49

## **ÍNDICE DE TABLAS**

Tabla 1: Presupuesto.....	50
---------------------------	----

# 1. INTRODUCCIÓN

La visión artificial o visión por computador se define como un campo de la inteligencia artificial que tiene como objetivo el programar un ordenador para que interprete un conjunto de imágenes de una determinada escena. Las imágenes que sacamos con las cámaras son bidimensionales, mientras que la escena real tiene una dimensión más, es tridimensional. Por tanto, estamos perdiendo una dimensión en el paso de la realidad a la fotografía. La visión estereoscópica, la cual forma parte de la visión artificial, es un procedimiento, entre otros, para la consecución de dicha dimensión perdida.

Nuestro sistema visual humano es capaz de percibir la realidad en tres dimensiones. Sin embargo, un único ojo por separado tan solo es capaz de percibir una única imagen bidimensional. El cerebro es el encargado de integrar las dos imágenes que se están viendo en una sola, dándole una tercera dimensión a la realidad.

Los sistemas estereoscópicos artificiales, en un intento de imitar el sistema estereoscópico biológico, utilizan al menos dos imágenes distintas de la misma escena. Las imágenes son tomadas por cámaras, mientras que el encargado de procesarlas y analizar dichas imágenes es el computador o el ordenador, además de realizar los cálculos necesarios para el cálculo de esa tercera dimensión y la reconstrucción de la escena 3D. Por tanto, la analogía entre el sistema biológico y el artificial sería la siguiente: nuestros ojos, que son los encargados de tomar las imágenes, serían las cámaras, mientras que el cerebro sería equivalente al computador en el sistema artificial. Eso sí, el sistema artificial podría decirse que se encuentra a años luz del humano y en absoluto consigue los resultados del primero.

Así pues, en este trabajo se hará uso de la visión estereoscópica para obtener de forma sencilla la reconstrucción de una escena real. Para ello, se utilizarán dos cámaras para grabar un objeto en movimiento, en un fondo o background completamente fijo. Se grabará un video de una persona en movimiento y tras obtener las imágenes correspondientes del video, éstas serán analizadas y procesadas como se explicará a lo largo del documento. Con el análisis de dichas imágenes se podrá calcular la dimensión perdida para dar paso a la reconstrucción de la escena y hacer posible el seguimiento de cualquier objeto en movimiento.

Para presentar el contenido que alberga el proyecto, en primer lugar, se pondrá en contexto las aplicaciones de la visión artificial y el entorno tecnológico en el que se encuentra.

Posteriormente, se expondrán los objetivos que se esperan conseguir con el proyecto y el alcance de éste.

A continuación, se pasará al apartado de Estado del Arte, es decir, se explicarán todos los conceptos teóricos para entender posteriormente su implementación en el programa informático MATLAB.

Se explicará el desarrollo del algoritmo generado y se analizará los resultados obtenidos junto a los problemas surgidos y las soluciones adoptadas. Se argumentarán los diversos factores a tener en cuenta para la consecución de los objetivos y las conclusiones del trabajo realizado.

Por último, se proporcionará la planificación y el presupuesto del proyecto.

## 2. CONTEXTO

La introducción en la industria de la visión artificial comenzó a principios de los años 90. Su rápido crecimiento se ha producido de la mano de avances informáticos y tecnológicos desde que comenzó a estudiarse.

Hoy en día, siendo las cámaras como los ordenadores parte de la vida cotidiana de las personas, sobre todo estos últimos, la visión artificial es el campo donde confluyen ambos dispositivos. Se trata de un campo de investigación que cada día va a más, el cual cada vez tiene más aplicaciones en todo tipo de campos, como pueden ser la automoción, el sector farmacéutico, alimenticio, robótica, etc. Por ejemplo, aplicaciones de visión artificial son:

- Uso en los controles de calidad de soldaduras, agujeros, presencia de tornillos, etc.
- Controles dimensionales de cajas de baterías y de pistones en automoción.
- Guiado de robots y recogida de piezas.
- Detección de contaminantes en productos alimenticios.
- Control de la posición de placas solares.
- Detección de objetos abandonados, reconocimiento facial y de conductas sospechosas en aeropuertos.

Éstas son aplicaciones que ya son utilizadas hoy en día, sin embargo, este sector está siendo altamente desarrollado debido al gran avance tecnológico. Como ejemplo, la visión artificial será clave en la consecución del coche autónomo, que, a pesar de su existencia en la actualidad, aún tiene muchísimo margen de mejora. Las funciones, entre otras, de la visión artificial en estos coches son la detección de peatones, ciclistas o cualquier obstáculo que se interponga en la trayectoria del vehículo. Por tanto, se necesita la mejor tecnología para que estos objetivos se consigan con una efectividad total, sin dar posibilidad a percances o accidentes.

Como esta, existen infinitud de aplicaciones futuras en las que únicamente con el desarrollo de la visión artificial será posible su implementación. La reconstrucción de escenas 3D y el tracking o seguimiento de objetos se antoja como claves para el desarrollo de este campo, y es por eso que este proyecto va a tratar de explicar una técnica sencilla para llegar a dicha escena 3D.

### 3. OBJETIVOS Y ALCANCE

Este proyecto tiene como **objetivo principal** desarrollar un algoritmo sencillo con bajo coste computacional que sea capaz de detectar un objeto en movimiento y hacer el tracking del objeto, validando la ecuación del cálculo de la distancia de las cámaras al objeto, de forma que se pueda seguir desarrollando esta tecnología alcanzando niveles más altos y los requerimientos mínimos para poder implantarlos en aplicaciones futuras con alto poder tecnológico.

Por tanto, para lograr este objetivo principal, se llevarán a cabo los siguientes objetivos secundarios:

- Determinar la disposición de las cámaras.
- Determinar la calibración que se debe realizar en la escena para poner en sintonía las dos cámaras que van a ser utilizadas.
- Determinar los procesos morfológicos a realizar en las imágenes para eliminar la información innecesaria y quedarnos únicamente con nuestro objeto de interés.
- Obtener y comprender las fórmulas teóricas para la determinación de la distancia de un objeto a las cámaras y las fórmulas para obtener las coordenadas de un punto cualquiera de una escena 3D.
- Determinar la distancia de las cámaras al objeto.
- Determinar la posición exacta del objeto en un sistema de ejes coordenados y representar la trayectoria o el tracking que éste realiza.

## **4. BENEFICIOS QUE APORTA EL TRABAJO**

Este proyecto tendrá como una posible aplicación el monitoreo de organismos acuáticos sujetos a experimentación en la Estación Marina de Plentzia (PIE-UPV/EHU). Por tanto, es necesario llegar a unos resultados satisfactorios para que esta tarea se pueda realizar con garantía.

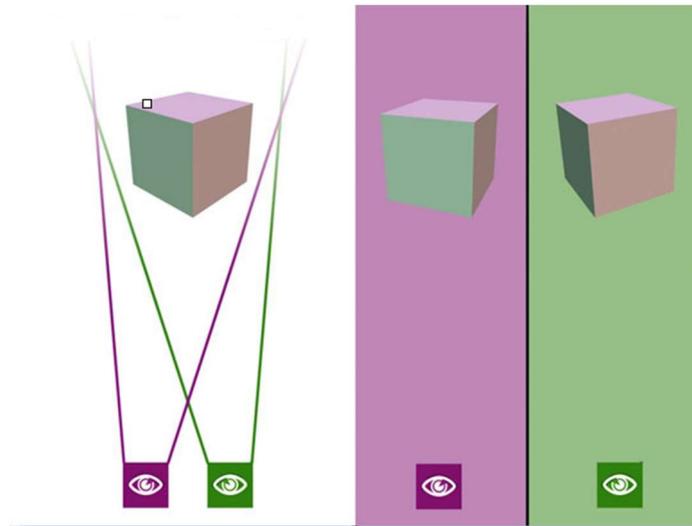
## 5. ESTADO DEL ARTE

### 5.1. INTRODUCCIÓN

En este capítulo, se realizará una breve revisión bibliográfica de los sistemas de visión artificial, visión estéreo y tracking. A la par de la revisión bibliográfica, se proporcionarán las ecuaciones teóricas que serán posteriormente utilizadas para la implementación de los diferentes algoritmos necesarios para alcanzar los diferentes objetivos del proyecto.

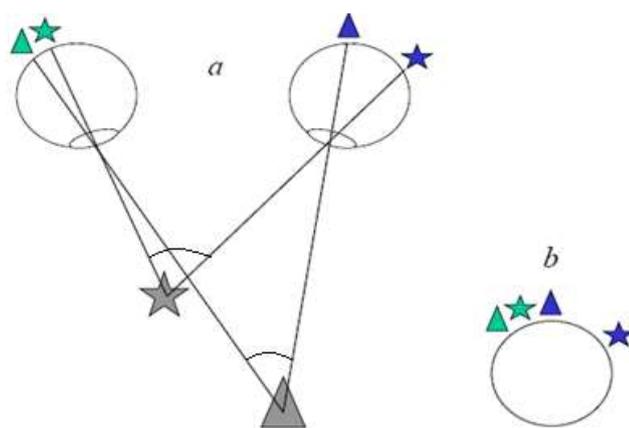
Existen dos tipos de métodos para determinar la **composición de una escena**: activos y pasivos. Los métodos activos son aquellos en los que se interviene desde el exterior en la escena, iluminándola o enviando un haz energético. Un ejemplo de método activo puede ser el telémetro láser, que es un aparato que utiliza un rayo láser para determinar la distancia hasta un objeto. Por el contrario, los pasivos no actúan sobre la escena, siendo la visión estereoscópica el principal método [1].

Los sistemas de visión estereoscópica utilizan dos o más cámaras, pero la mayor parte de ellos utilizan dos, tomando como referencia el propio modelo biológico, donde debido a que los ojos están separados, se puede determinar la tercera dimensión, es decir, la distancia a la que se encuentra el objeto. El hecho de que los ojos estén separados tiene el siguiente efecto: imágenes desplazadas de la misma escena en sendos ojos, es decir, las imágenes de ambos ojos son prácticamente la misma, pero tienen un desplazamiento relativo. Las instantáneas que perciben los ojos son la misma escena, pero debido a que estos están separados una distancia de 55mm, las perspectivas difieren y estas instantáneas son diferentes, como se puede ver en la Figura 1.



**Figura 1: Efecto de la separación entre ojos/cámaras. El ojo izquierdo ve sobre todo la cara verde del cubo, mientras que el ojo derecho ve en mayor medida la cara marrón.**

Para analizar que produce este efecto observamos la Figura 2 (a). Sean dos objetos, uno más cercano, el triángulo y el otro más lejano, la estrella. En función de donde se encuentre el objeto, éste se verá bajo un determinado ángulo por ambos ojos. En la Figura 2 (b) se observa que al solapar las imágenes que se obtienen en cada ojo en la figura (a), la separación relativa entre las imágenes de los objetos lejanos, es decir, la separación entre el triángulo verde y el azul es menor que la separación relativa entre las imágenes de las estrellas (objeto cercano).



**Figura 2: (a) Sistema de visión humano.; (b) Superposición de las imágenes de ambos ojos.**

Estas separaciones de los objetos existentes entre cada imagen del ojo izquierdo y derecho es lo que se denomina **disparidad**. La disparidad es el parámetro que relaciona el mismo objeto en ambas imágenes. Posteriormente se verá cómo se realiza su cálculo.

Desde el punto de vista artificial, básicamente existen dos formas para **obtener las imágenes** desplazadas que permitan reconstruir la escena:

1. Utilizar dos o más cámaras alineadas y separadas una cierta distancia. Esta distancia es totalmente aleatoria, pero dependerá de la aplicación. La única condición es que esta distancia debe ser conocida con exactitud.
2. Utilizar únicamente una cámara móvil que sea capaz de desplazarse en línea recta. La condición que debe cumplir este método es que la cámara debe ser capaz de tomar imágenes mientras realiza este desplazamiento [1].

La **geometría o la estructura de las cámaras** también es algo importante a considerar. Así, en los sistemas de visión estereoscópicos, los ejes pueden convergen o pueden ser paralelos. El sistema visual humano trabaja fundamentalmente con ejes ópticos convergentes, aunque se puede hacer la siguiente apreciación: cuando nuestros ojos visualizan un objeto cercano, los ejes ópticos aproximadamente convergen en el objeto. Sin embargo, cuando se trata de un objeto lejano, se puede decir que los ejes prácticamente no convergen y se comportan como ejes ópticos paralelos. Es por eso, que en este último caso, si se enfoca en un objeto lejano, pero en la escena existen otros objetos de interés más cercanos, estos últimos pueden salir borrosos, con una gran pérdida de información. Lo mismo ocurre en la situación inversa. Sin embargo, el sistema de ejes ópticos paralelos es el más utilizado en los sistemas de visión artificial [2].

## 5.2. EL MODELO PINHOLE

El modelo de cámara PinHole es el más utilizado para modelar la formación de imágenes de una cámara. Se trata de una simplificación relativamente sencilla comparado con el modelo físico de la cámara, ya que nos proporciona la siguiente ventaja: las ecuaciones resultantes son mucho más sencillas de aplicar. El modelo PinHole considera una cámara como una caja que tiene un pequeño orificio en un plano y la luz únicamente puede pasar por dicho agujero. Sin embargo, este modelo no tiene en cuenta distorsiones geométricas o el desenfoque [3]. Por tanto, el modelo PinHole solo puede usarse como una aproximación y nunca como un método de gran precisión.

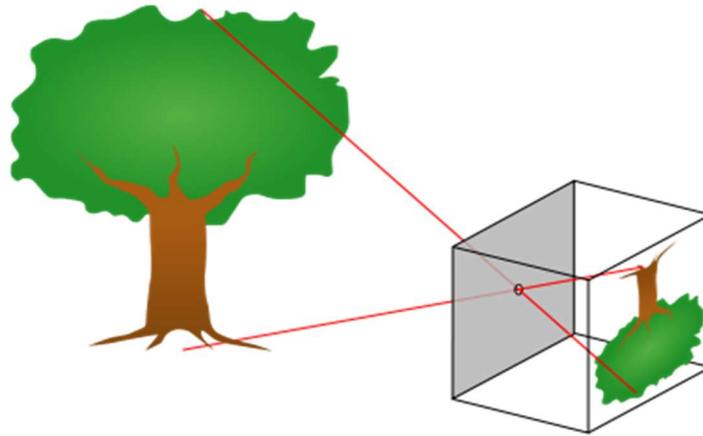
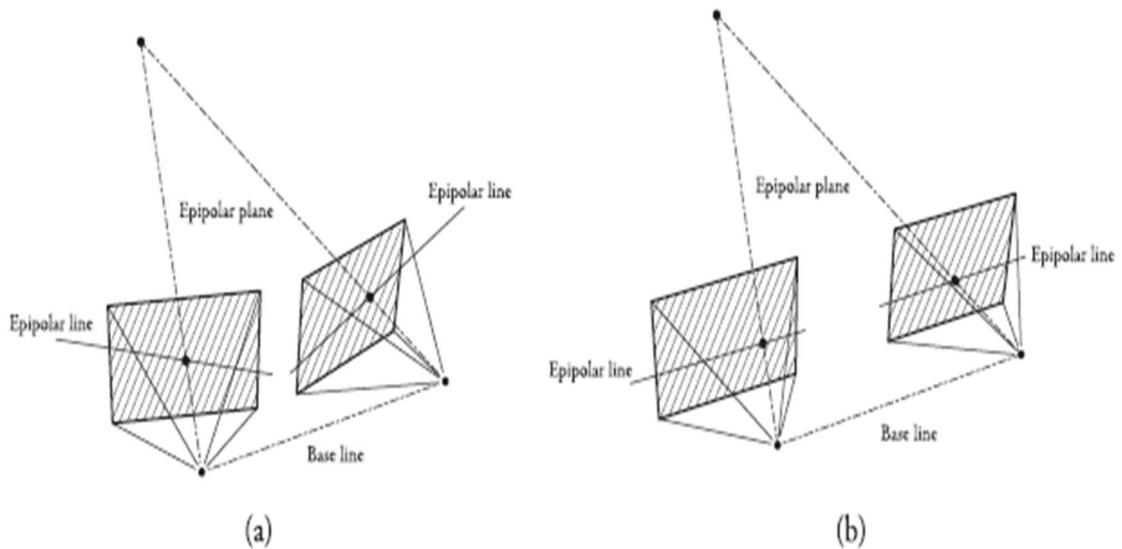


Figura 3: Modelo PinHole [3].

Con este supuesto, la geometría que se genera se denomina geometría epipolar. La **línea epipolar** es la línea que une la imagen izquierda y derecha de un mismo punto, mientras que la **línea base** es la distancia que separa ambas cámaras.

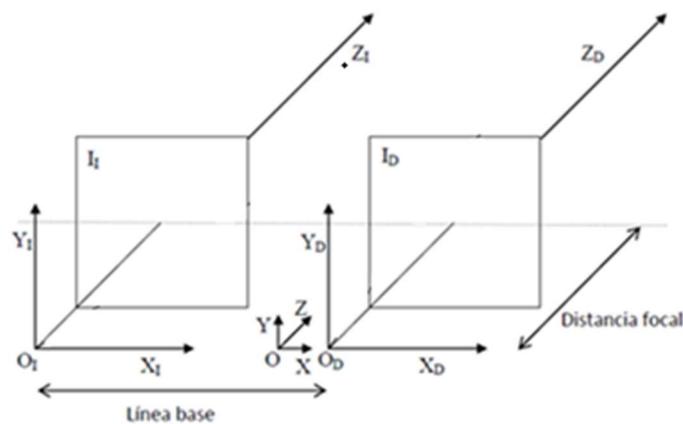
Si nuestro sistema fuera convergente, esta línea epipolar no sería paralela a la línea base como se ve en la Figura 4 (a). En cambio, considerando ejes ópticos paralelos nos queda la geometría que podemos ver en la Figura 4 (b), en la que se observa que la línea epipolar es paralela a la línea base. Esto presenta una ventaja muy clara: cualquier punto contenido en la línea epipolar de la imagen izquierda, su proyección derecha debe estar en la misma línea epipolar de la imagen derecha [4]. La correspondencia entre estos dos puntos es considerada el mayor problema de la visión estereoscópica. Al realizar la correspondencia con esta

geometría en lugar de analizar toda la imagen, tan solo es suficiente analizar la línea epipolar para encontrar el punto correspondiente.



**Figura 4: (a) Geometría no epipolar; (b) Geometría epipolar [5].**

Por tanto, el objetivo de la geometría epipolar es que ambas líneas epipolares queden paralelas a la línea base. Para ello, con situar y orientar las cámaras de forma que sólo exista un desplazamiento horizontal entre ellas, teóricamente, debería ser suficiente, quedando los ejes ópticos totalmente paralelos y los ejes de abscisas (eje X) de cada una de las cámaras coincidentes como sucede en la Figura 5.



**Figura 5: Geometría epipolar con ejes ópticos paralelos ( $Z_1$  y  $Z_D$ ) y ejes de abscisas coincidentes.**

Sin embargo, estas condiciones son normalmente difíciles de materializar. Es realmente complicado que ambos ejes ópticos queden totalmente paralelos, ya que en la adquisición de

imágenes se introducen deformaciones, como la distorsión radial y tangencial, que degradarán la precisión del cálculo de la distancia del objeto a las cámaras.

Para darle solución a este problema se realiza un rectificado de imágenes. El **rectificado** es un proceso que consiste en transformar ambas imágenes en una pareja de imágenes estéreo de forma que sus líneas epipolares se convierten colineales y paralelas a la línea base.

### 5.3. PASOS EN EL PROCESO DE LA VISIÓN ESTEREOSCÓPICA

Los pasos de naturaleza secuencial que debería resolver un algoritmo estereoscópico son, a grandes rasgos, los siguientes:

1. Calibración estereoscópica de las cámaras.
2. Adquisición de imágenes.
3. Correspondencia entre imágenes.
4. Determinación de la distancia al objeto.
5. Reconstrucción de la escena 3D.

#### 1) Calibración estereoscópica de las cámaras.

Para realizar una correcta correspondencia entre puntos, primeramente, es necesario conocer las características del conjunto de cámaras a utilizar. El proceso de calibración estereoscópica consiste en determinar los parámetros intrínsecos (distancia focal, punto principal, distorsión de las lentes, entre otros) como extrínsecos (posición y orientación) de las cámaras. Es decir, al utilizar dos cámaras, es necesario obtener una relación entre ellas. [6]

Si se asume que no se tiene ningún tipo de información sobre la relación entre las cámaras, se podría encontrar un punto de interés en la imagen tomada por la cámara izquierda, pero no se sabría identificar ese mismo punto en la derecha. Se tendría que recorrer toda la imagen pixel a pixel para identificar dicho punto. Sin embargo, si se conoce la posición y orientación de las cámaras además de los parámetros intrínsecos, y aplicando la geometría epipolar, el problema de la correspondencia se reduce considerablemente como se ha visto anteriormente.

Una situación importante que debe existir en la calibración y que en algunos casos se considera obvia, es que el punto a analizar se debe encontrar en las 2 imágenes y saber que es el mismo punto en el espacio.

Para la calibración, existen multitud de métodos y algoritmos, siendo los más conocidos los métodos de Fougeras [7] y de Zhang [8]. Sin embargo, en la actualidad los diferentes programas usados en el procesado de imágenes como MATLAB o las diferentes bibliotecas de visión artificial como OpenCV [9] traen ya de por sí los algoritmos de calibración implementados, por lo que el análisis de estos métodos no será algo que se recoja en este proyecto.

## 2) Adquisición de imágenes

Existe una gran variedad de formas para llevar a cabo este paso. Por ejemplo, éstas pueden ser tomadas desde posiciones y direcciones ligeramente distintas o radicalmente opuestas. Pueden ser tomadas simultáneamente en el tiempo o en intervalos de una duración determinada. Si esta duración temporal es grande, habrá que considerar aspectos como el cambio de luz, cambios en las condiciones atmosféricas o cualquier otro cambio que se produzca en la escena.

En el caso de utilizar dos o más cámaras, es de vital importancia que todas ellas puedan comenzar a grabar en el mismo instante, para que cada imagen de una cámara coincida con la imagen de la otra cámara. Si este paso no se pudiera cumplir, comenzaríamos a grabar con una cámara, y en otro instante, comenzaríamos con la otra cámara, por lo que la posición del objeto sería distinta para el primer frame del vídeo.

Son las aplicaciones a las que se destine el sistema las que marcan el tipo de escena y la forma de adquisición de imágenes. Cambia drásticamente una aplicación para un vehículo aéreo con fines cartográficos, donde las imágenes son de baja resolución, que una aplicación para un vehículo autónomo, en los que se necesita gran precisión para evitar obstáculos.

## 3) Correspondencia entre imágenes.

Este paso trata de obtener elementos identificativos de la imagen, y de estos elementos habrá que obtener ciertos atributos, para poder hacer la correspondencia entre imágenes. Existen dos técnicas para llevar a cabo la correspondencia: las técnicas basadas en el área y las técnicas basadas en las características.

Las primeras emplean la intensidad correspondiente de cada pixel y su vecindario para buscar los valores parejos de intensidad del vecindario en la otra imagen. En cambio, en las técnicas basadas en características, lo que se busca son propiedades simbólicas como pueden ser puntos de borde aislados, cadenas de puntos o regiones delimitados por bordes [1].

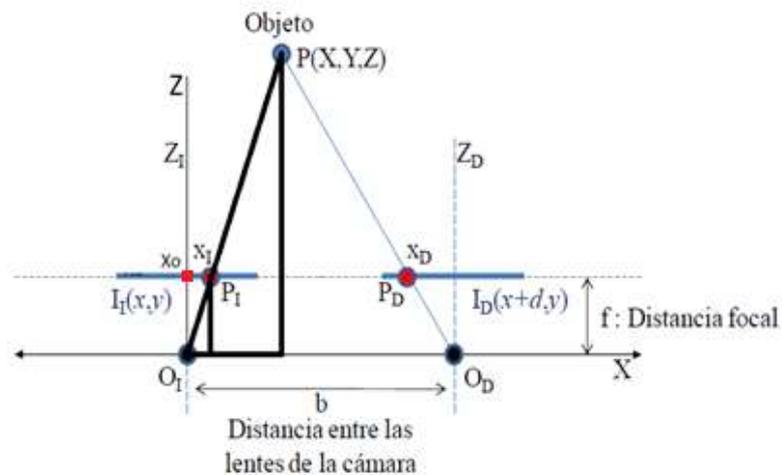
Dependiendo de la precisión que necesitemos y del método que se emplee, puede ser necesario un paso de segmentación adicional, en la que se extraerá información adicional de las características ya definidas.

Es considerado el paso con mayor dificultad del proceso estereoscópico, sin embargo, este proyecto recoge una alternativa más sencilla para evitar este paso.

4) Determinación de la distancia al objeto.

Como se ha visto en la Figura 2, cuánto más lejos se halle un objeto, la distancia relativa entre las dos imágenes, es decir, la disparidad, será menor que cuando el objeto se encuentre en una posición más cercana. Es por eso que en el cálculo de la distancia  $Z$  al objeto, ésta será inversamente proporcional a la disparidad.

Considerando la geometría epipolar, se va a obtener la fórmula que nos proporciona la distancia al objeto. Para tal propósito nos apoyaremos en la Figura 6:



**Figura 6: Geometría epipolar vista en planta.**

Sea  $P$  un punto de la escena real con coordenadas  $(X, Y, Z)$  según el sistema de coordenadas global, cuyo origen lo colocaremos coincidente con el punto  $O_I$ . El subíndice  $I$  hará referencia a la cámara izquierda mientras que el subíndice  $D$  a la cámara derecha.

Este sistema de coordenadas global puede situarse en cualquier punto del sistema, pero una forma sencilla de visualizar la triangulación que se va a realizar es colocarlo coincidente en dicho punto o en el punto medio entre  $O_I$  y  $O_D$ , siendo estos los centros de proyección, puntos teóricos por el que pasan todos los rayos de luz en la cámara izquierda y derecha respectivamente. En dichos puntos se considera un eje de coordenadas relativo para cada cámara.

Mediante semejanza de triángulos, marcada en la Figura 6, la cual es posible gracias a la simplificación de modelo PinHole, obtenemos,

$$\begin{array}{l}
 \text{O}_I: \quad \frac{X}{Z} = \frac{x_I}{f} \quad \rightarrow \quad x_I = f \frac{X}{Z} \\
 \\
 \text{O}_D: \quad -\frac{b-X}{Z} = \frac{x_D}{f} \quad \rightarrow \quad x_D = f \frac{X-b}{Z}
 \end{array}
 \left. \vphantom{\begin{array}{l} \text{O}_I \\ \text{O}_D \end{array}} \right\}
 \quad d = x_I - x_D = f \frac{b}{Z} \quad \rightarrow \quad \boxed{Z = f \frac{b}{d}} \quad (1)$$

Por tanto, para obtener la distancia  $Z$  de las cámaras al objeto, primeramente tenemos que calcular la disparidad de la fórmula, que viene dado como  $d = x_I - x_D$ , siendo  $x_I$  las coordenadas del objeto en el eje  $X$  de la cámara izquierda y  $x_D$  de la cámara derecha.

En cuanto al numerador, comentar que  $b$  es la distancia entre las cámaras, valor que debe conocerse con precisión y que se suele dar en centímetros. La  $f$  hace referencia a la distancia focal, la cual nos viene dada en la calibración con unidades pixel, aunque también viene dada en las propiedades de las imágenes. En la implementación de la fórmula, habrá que tener cuidado con las unidades de cada parámetro, que hagan corresponder a las unidades de la distancia al objeto.

Por tanto, se ha demostrado que la distancia al objeto es inversamente proporcional a la disparidad y directamente proporcional a la distancia entre cámaras (línea base) y a la distancia focal. Estas dos últimas se consideran constantes una vez elegida la estructura de las cámaras.

Como ya se ha mencionado, al hacer la simplificación de modelo de PinHole, la fórmula dada en (1) sólo puede ser utilizada para dicha hipótesis. Y así, aplicando dicha fórmula a cada pixel, obtenemos el mapa de profundidades, que será la representación de la distancia en cada punto de la imagen.

### 5) Reconstrucción de la escena 3D.

La reconstrucción de la escena real trata de establecer coordenadas 3D a partir de las proyecciones 2D de los emparejamientos.

Volviendo a la Figura 6, y de nuevo mediante semejanza, obtenemos:

$$\frac{X}{Z} = \frac{x_I - x_o}{f} \quad \rightarrow \quad \boxed{X = \frac{Z}{f} (x_I - x_o)} \quad (2)$$

Esta ecuación representa la coordenada X del punto P en las coordenadas globales. El único parámetro que no se había mencionado hasta ahora es  $x_o$ , que es la coordenada x del punto principal de la imagen, punto determinado por la proyección ortogonal del centro de proyección en el plano imagen. De igual forma obtenemos la coordenada Y del punto P, que viene dada por:

$$Y = \frac{z}{f}(y_I - y_o) \quad (3)$$

Donde  $y_I$  es la coordenada y del centroide en la imagen izquierda (se realiza con los de la imagen izquierda porque los ejes han sido colocados en dicha cámara, pero perfectamente podrían haber sido colocados en la cámara derecha) e  $y_o$  es la coordenada y del punto principal de la imagen.

Con estas fórmulas ya podemos calcular las coordenadas de todos los puntos de la escena tridimensional, es decir, en cada instante de tiempo, tendríamos las coordenadas del objeto completamente definidas. Es decir, el movimiento del objeto sería conocido. El tracking del objeto, por tanto, se podría realizar representando cada punto en un gráfico.

## 6. APLICACIÓN SELECCIONADA

Como ya se ha comentado, este proyecto consiste en el desarrollo de un algoritmo con bajo coste computacional que pueda que sea capaz de detectar un objeto en movimiento y que trace la trayectoria de dicho objeto en una escena tridimensional.

Para conseguir dicho objetivo, dentro de los métodos pasivos se ha elegido el procedimiento de visión estereoscópica como el método a seguir para el desarrollo del algoritmo. Se considerará un sistema de cámaras de ejes ópticos paralelos, donde se dará la geometría epipolar vista en el apartado de Estado del Arte, por lo que las ecuaciones demostradas anteriormente podrán ser utilizadas.

Además, en el proyecto se decidió evitar las sombras y las oclusiones entre objetos y demás condicionantes que dificultarían el procesamiento de las imágenes y que aumentarían el coste computacional del algoritmo.

Asimismo, se propuso que el algoritmo fuera implementado con varios objetos para comprobar que su funcionamiento con distintos objetos de distinta naturaleza, entre ellos, un balón de fútbol y una persona. Tras comprobar que funciona de igual forma para ambos objetos, sólo se mostrarán los resultados para el caso de una persona, por ser éste un caso de mayor importancia para la visión artificial.

Por último, decir que se decidió realizar este trabajo en un aula de la Escuela de Ingeniería de Bilbao. Se trata de un aula bastante amplia, con espacio suficiente para realizar el movimiento del objeto en diferentes direcciones. En cuanto a la iluminación, algo crítico en el procesado de imagen, cuenta con 2 ventanales translúcidos, los cuales dejan pasar la luz en cierta medida pero cuya influencia se considera despreciable por encontrarse estos a una gran distancia de la posición de las cámaras, y un ventanal totalmente transparente que junto a la iluminación artificial, lámparas fluorescentes, tienen todo el peso en la problemática de la iluminación.

## 7. EQUIPO UTILIZADO PARA EL DESARROLLO DEL ALGORITMO

- Dos cámaras **GoPro Hero 3 White Edition**. Presentes en el mercado desde 2012, cuyas características generales son: videocámara de 12 Mp, video Full HD 1080p y como gran ventaja y la razón de su compra que cuentan con WIFI para conectarlas simultáneamente a un mando inalámbrico. Su principal desventaja es que no trae pantalla para visualizar lo que se está grabando, sino que habrá que descargarse la aplicación de GoPro en el móvil y visualizarlo en dicho aparato.



Figura 7: Modelo de cámara GoPro Hero 3.

- **Mandos inalámbricos**. Ambas cámaras pueden ser conectadas a un mando, por lo que su gran ventaja es que se puede comenzar a grabar el vídeo exactamente en el mismo preciso instante, algo vital para la buena realización del proyecto.



Figura 8: Mando inalámbrico.

- **MPEG Streamclip.** Se trata de un potente convertidor de vídeo gratuito, reproductor y editor, cuyo uso es sencillo y eficaz. En este proyecto, en primera instancia se utilizó para transformar el video en frames o imágenes, aunque posteriormente este proceso se realizó con MATLAB, debido a que cuenta con ciertas limitaciones a la hora de definir el tamaño de los frames.
- **MATLAB.** Es una herramienta de software matemático con un lenguaje de programación propio. Es ideal para trabajar con aplicaciones de visión artificial, ya que procesa las imágenes como matrices en el que en cada celda almacena un pixel, por lo que la operativa entre imágenes realmente se convierte en operativa entre matrices. Ha sido el programa elegido para la implementación del algoritmo.
- **Portátil Dell Inspiron 15 7000,** con tarjeta gráfica NVIDIA GeForce 940MX, suficiente para aguantar la implementación del algoritmo mediante MATLAB.



**Figura 9: Portátil Dell Inspiron15 7000.**

## **8. PASOS SEGUIDOS EN EL DESARROLLO DEL ALGORITMO**

El algoritmo que se ha realizado en este proyecto trata de ser sencillo y eficaz. La teoría seguida es la anteriormente expuesta, sin embargo, se han realizado ciertas simplificaciones. Así, por ejemplo, recordando los pasos en el proceso de la visión estereoscópica (apartado 5.3), la correspondencia entre imágenes no se ha realizado. En su lugar, se ha obtenido el flujo óptico y tras el correspondiente proceso morfológico, se ha extraído el centroide del objeto en cada imagen, como se explica posteriormente en detalle.

A continuación, en la Figura 10 se muestra el diagrama de bloques del proceso que se ha seguido.

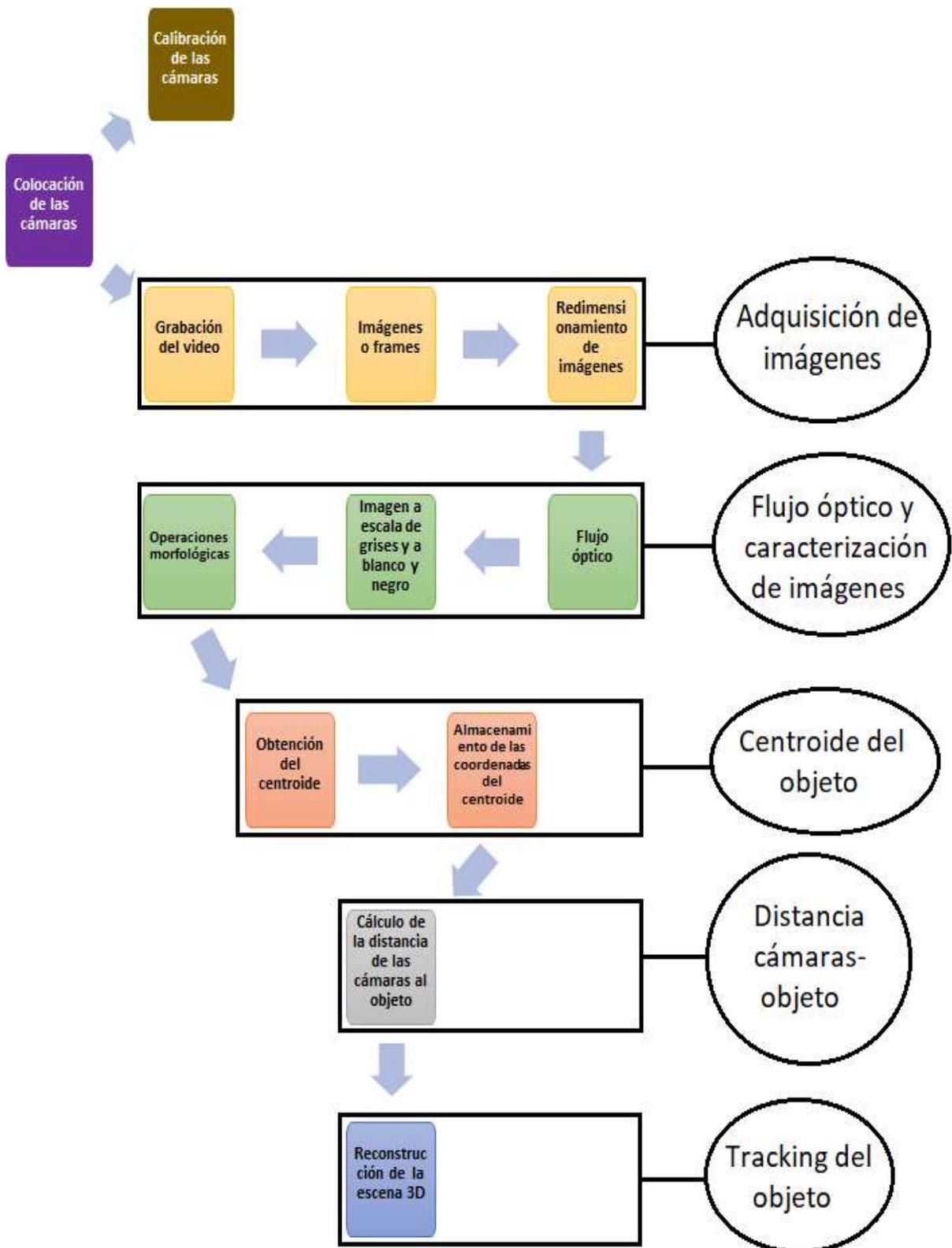


Figura 10: Diagrama de bloques del proceso seguido. Se observan también los diferentes scripts en los que se ha realizado el trabajo.

Como se observa en el diagrama de bloques, el primer paso a realizar es determinar la colocación o la disposición de las cámaras. Siendo una aplicación de ejes ópticos paralelos, las cámaras deberán quedar con sus ejes ópticos lo más paralelo posible. Además, habrá que tener en cuenta la distancia a la que se encuentre el objeto, de forma que éste pueda ser grabado en ambas cámaras, y así determinar la distancia entre las 2 cámaras, parámetro que necesitaremos para la implementación de la ecuación (1). Una disposición simple podría ser la siguiente:



**Figura 11: Disposición de las cámaras. Colocadas en una mesa plana de forma que sus ejes queden paralelos.**

Determinada la disposición de las cámaras y la distancia entre ellas, hay dos posibles caminos a tomar. El primero corresponde a la calibración estereoscópica entre ellas y el segundo camino trata del desarrollo del algoritmo para la detección del objeto en movimiento y el tracking.

## 8.1. CALIBRACIÓN ESTEREOSCÓPICA DE LAS CÁMARAS.

De forma totalmente independiente al resto del proceso se realizó la calibración de las cámaras mediante una aplicación incorporada en MATLAB, denominada “Stereo Camera Calibrator”.

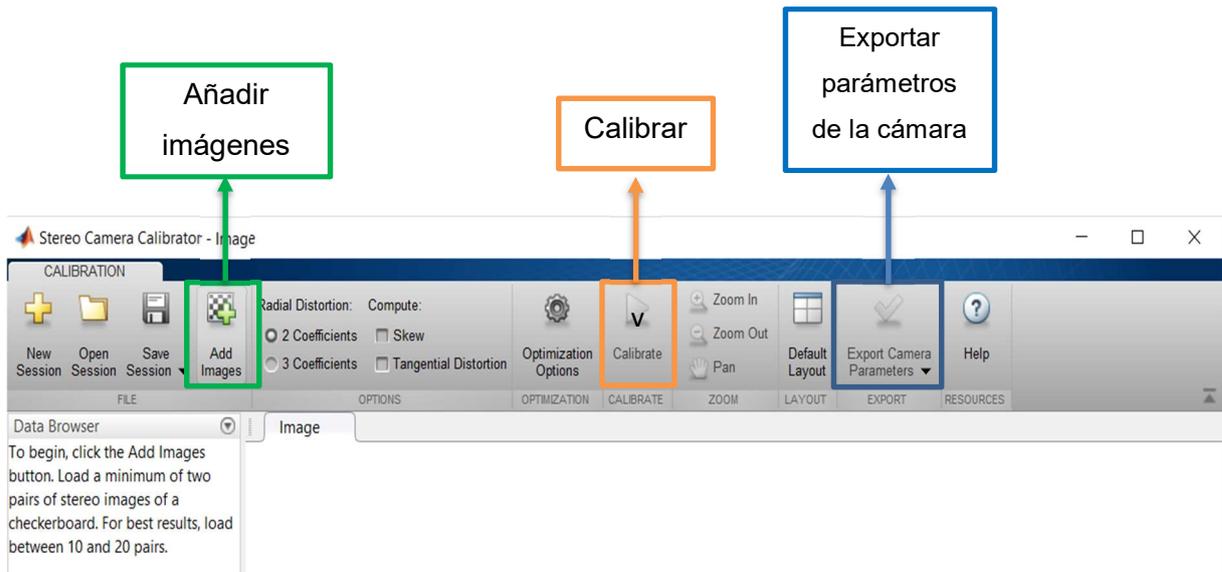


Figura 12: Stereo Camera Calibrator

Esta aplicación trata de insertar imágenes de un patrón de tablero de ajedrez obtenidas con ambas cámaras. La propia aplicación encuentra los puntos principales del tablero, es decir, las esquinas de cada cuadradito del tablero de ajedrez, como se ve en la Figura 13. En definitiva, lo que se está haciendo es relacionar un mismo punto real en ambas imágenes.

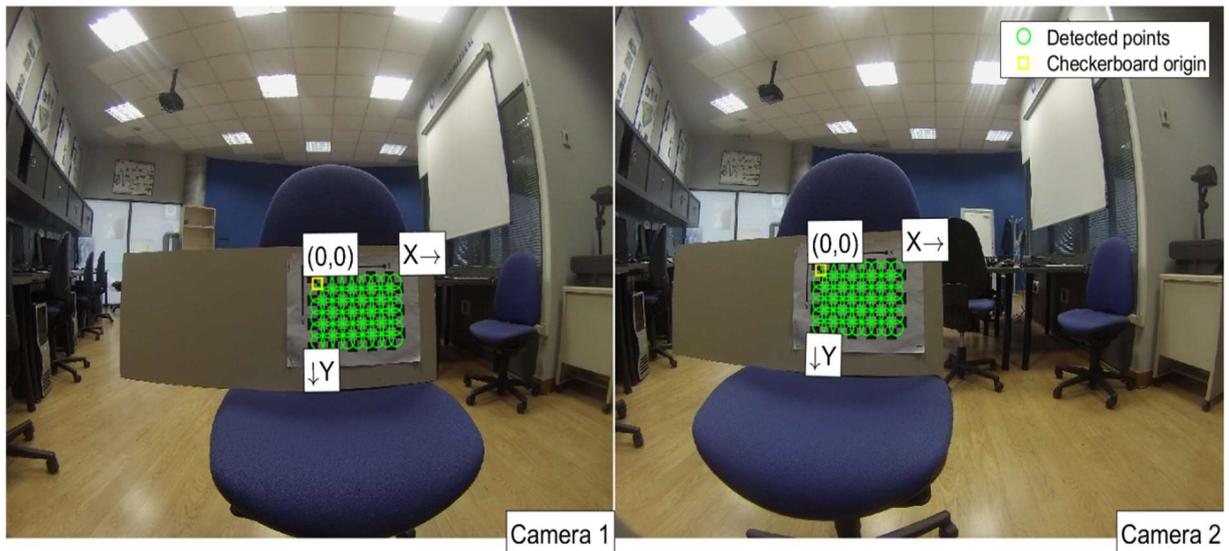


Figura 13: Detección de los puntos principales del tablero.

Una vez identificados los puntos principales en ambas imágenes, que serán correspondientes con el mismo punto real, la aplicación podrá realizar la calibración, con lo que podrá hacer los cálculos de los parámetros intrínsecos (distancia focal, punto principal y distorsión de las lentes) y extrínsecos (posición y orientación) de las cámaras. Además, nos proporcionará el error cometido en la calibración según la calidad de las imágenes y el cumplimiento de las siguientes condiciones:

- El patrón de tablero de ajedrez es el proporcionado por MATLAB. Salirse a otros patrones de tablero de ajedrez no dado por MATLAB puede significar una calibración con resultados negativos o que simplemente nos muestre error.
- Se requiere entre 10 y 20 pares de imágenes tomadas a una distancia aproximadamente a la que se encontrará nuestro objeto de interés.
- El tablero no puede colocarse con un ángulo mayor de  $45^\circ$  respecto al plano de la cámara.

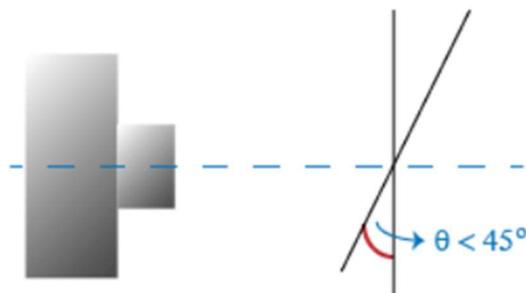


Figura 14: Ángulo máximo de inclinación.

Se aconseja que el error medio en la calibración sea menor de 0.5 píxeles. En caso de sea mayor, habrá que tomar nuevas imágenes del tablero de ajedrez o eliminar el par de imágenes que introduce un error elevado. Así, por ejemplo, en la calibración realizada para este proyecto el error medio fue de 0.13 píxeles, menor que el máximo recomendado, como se observa en la Figura 15. En el caso de que error medio fuera mayor a 0.5 píxeles, se podrían tomar nuevas imágenes o eliminar el par de imágenes 2 o 4 que son las que mayor error introducen.

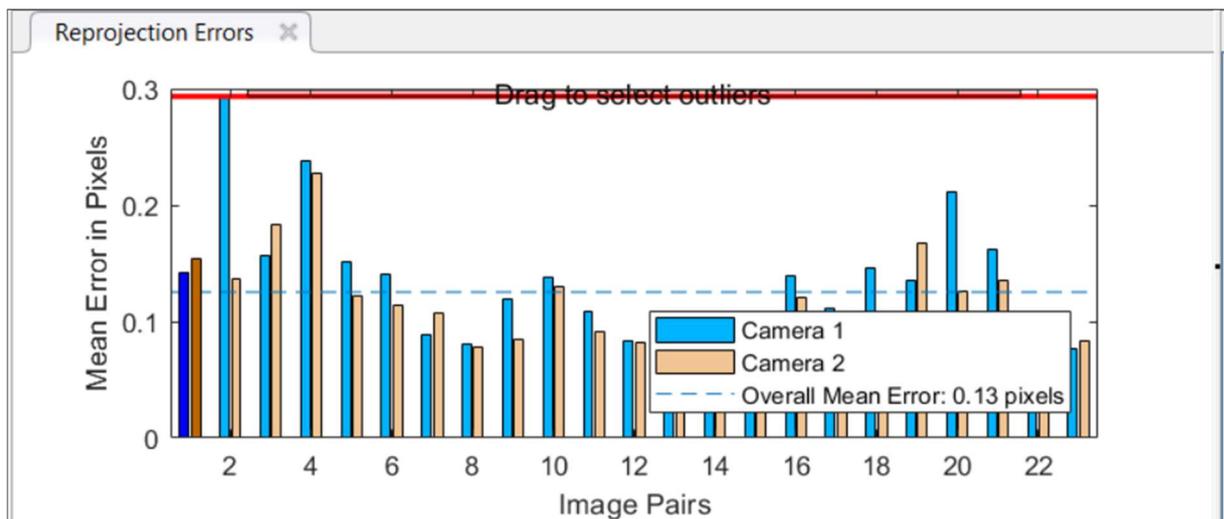


Figura 15: Error medio producido en la calibración.

Además, esta aplicación nos permite eliminar la distorsión de las imágenes, de tal forma que con una imagen podemos calcular la relación existente entre los centímetros y los píxeles. Dados dos puntos reales, podemos medir la distancia entre ellos, y de igual forma podemos hacerlo en una imagen sin la distorsión de dicha realidad. Así es como podemos obtener tal relación.

Sin embargo, los parámetros más interesantes que obtenemos de la calibración son la distancia focal, que vendrá dado en píxeles, y el punto principal de la imagen, también en píxeles. Estos parámetros serán necesarios para la implementación de las fórmulas vistas anteriormente en el apartado del Estado del Arte.

## 8.2. ALGORITMO PARA LA DETECCIÓN DEL OBJETO Y EL TRACKING.

Este apartado corresponde al segundo camino posterior a determinar la colocación de las cámaras. Para su realización en Matlab, este paso se ha dividido en 5 pasos, implementados cada uno de ellos en un “script” o archivo de órdenes diferente.

### 1) Adquisición de imágenes o frames



Figura 16: Diagrama de bloques del 1er script.

El video grabado corresponde a un video de una duración aproximada de 6 segundos, el cual grabado a 25 frames por segundo, da lugar a un total de unas 150 imágenes a procesar, las cuales se han de extraer del video.

Las cámaras presentaban la siguiente limitación: las **imágenes para la calibración** solo se pueden tomar a un tamaño de 2592x1944 pixeles. Sin embargo, las **imágenes obtenidas del video** mediante el programa MPEG Streamclip no permite un tamaño de imagen tan grande. Por tanto, se presentaba una gran incompatibilidad entre las imágenes a procesar sacadas del video y las imágenes para la calibración, lo cual se traducía en parámetros no útiles obtenidos en la calibración. Por tanto, la solución adoptada fue la siguiente: mediante MATLAB obtener los frames del video y convertirlos en el mismo tamaño que las de calibración mediante un “resize” o redimensionamiento.

En el video que se ha grabado en este proyecto, el movimiento realizado por el objeto de interés (persona) consta de 2 fases: en la primera de ellas, la persona se aleja de las cámaras. Posteriormente, hace un giro de algo menos de 180°, es decir, da la vuelta y vuelve hacia las cámaras. A continuación, junto a la descripción del movimiento, se muestran 3 imágenes del video, con las que se pretende que el lector sea capaz de hacerse una idea del movimiento realizado por la persona, de forma que pueda entender el proceso realizado. Además, en el apartado 10 se incluye un pequeño álbum fotográfico más detallado del movimiento de la persona.



Figura 17: Primer frame del video. 1ª fase: alejamiento de las cámaras



Figura 18: Frame intermedio. Cambio de sentido: giro.



Figura 19: Último frame. Final de la 2ª fase: acercamiento a las cámaras.

2) Flujo óptico y caracterización de imágenes.



Figura 20: Diagrama de bloques 2º script.

En este script, lo primero que se ha realizado es el cálculo del flujo óptico. Una forma de hacerlo para obtener el movimiento que realiza el objeto de interés es hacer la resta entre imágenes correspondientes a instantes sucesivos. Así, por ejemplo, podríamos restar la imagen correspondiente al instante 10 a la imagen correspondiente al instante 1, con lo que nos quedaríamos con el movimiento realizado entre esos dos instantes.

Sin embargo, en este proyecto se ha seguido otra técnica, conocida como “**background subtraction**” o resta de la imagen de fondo. En lo que consiste es en restar las imágenes en cada instante a una imagen sin el objeto de interés, únicamente con el fondo. Así, al hacer la resta entre imágenes, nos quedamos únicamente con la posición del objeto de interés en dicho instante. (Ver Figura 20).

La primera técnica no ha podido ser utilizada en este proyecto ya que se trata de un video de poca duración con un gran movimiento entre ciertos instantes, y muy poco movimiento entre otros instantes, por lo que la técnica no resulta tan precisa.



**Figura 21: Resta de cada imagen con el fondo.**

A simple vista puede parecer que en la resta se pierde algo de información, como la cabeza de la persona y los pies. Para evitar esto, dicha imagen se convierte en una imagen a escala de grises y posteriormente a blanco y negro (BW), como la de la Figura 22.

En realidad, dichas partes, que a simple vista parecen no verse, tienen un tono gris no tan negro como el resto de la imagen y por tanto, ajustando el **umbral de binarización**, es decir, ajustando el valor de grises adecuado para la conversión de la imagen a blanco y negro, podremos recuperar parte de esa información que parecía perdida. La binarización se define como la reducción de información a verdadero, designado por el número 1, y falso, designado por 0, como únicos valores posibles. Por tanto, partiendo de una imagen en escala de grises, si la tonalidad de gris de un determinado pixel supera el umbral de binarización, dicho pixel será verdadero, y mostrará un 1 o el color blanco. Es posible realizar procesos previos para la selección del umbral de binarización, pero en este trabajo la elección del umbral se hará experimentalmente.

Sin embargo, este proceso de binarización no nos deja únicamente con el objeto de interés, sino que habitualmente aparecen bordes imprecisos y regiones solapadas que pueden no pertenecer a nuestro objeto. Por tanto, para quedarnos únicamente con el objeto de interés,

como el de la Figura 22, y que éste tenga la forma del objeto real, habrá que aplicar diferentes filtros.

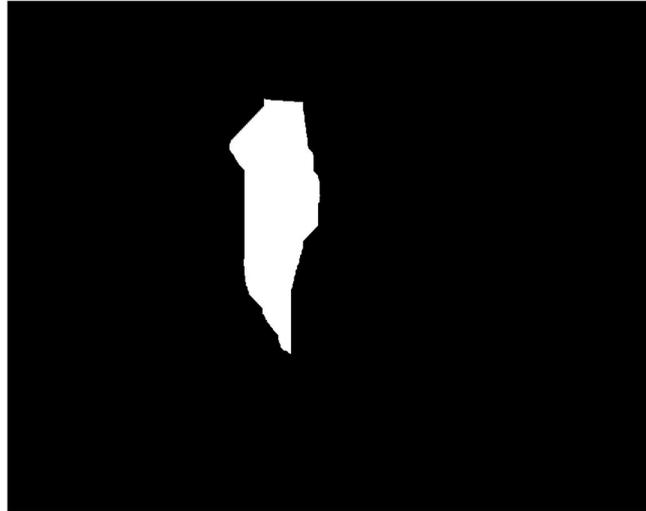


Figura 22: Resta en blanco y negro.

La aplicación de diferentes filtros es lo que se denomina **procesamiento morfológico**. Se trata de una técnica que se caracteriza por realzar la geometría y la forma de los objetos. Para una explicación sencilla del procesamiento morfológico, tomamos la siguiente figura con cuadrados de tamaños: 1, 3, 5, 7, 9 y 15 píxeles, y aplicamos sobre ella los dos filtros utilizados:

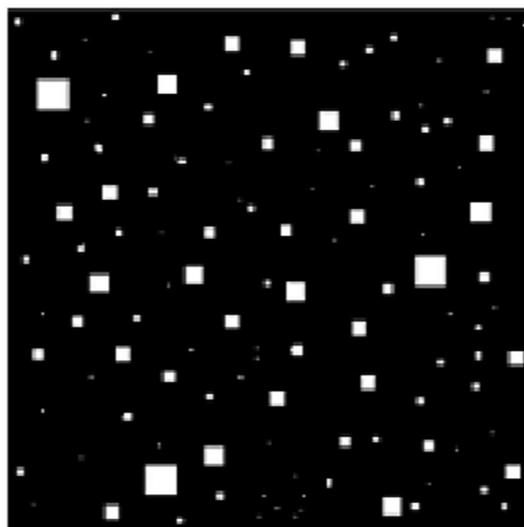
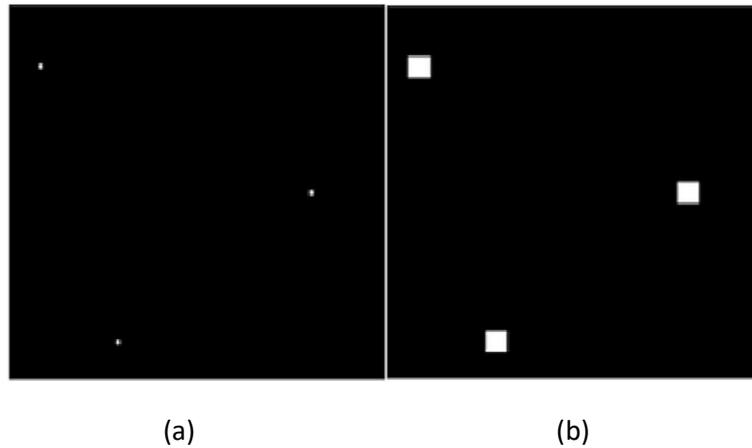


Figura 23: Imagen con cuadrados de tamaño: 1, 3, 5, 7, 9 y 15 píxeles.

- Erosión binaria: al realizar el proceso de erosión a la anterior figura con un elemento cuadrado de 13 pixeles, el resultado se aprecia en la Figura 24 (a). La erosión nos elimina los cuadrados menores de 13 pixeles, y nos quedamos únicamente con aquellos cuya área es mayor de 13x13 pixeles. Además, nos convierte esos cuadrados de 15 pixeles en cuadrados más pequeños de 2 pixeles.
- Dilatación: aplicando el proceso de dilatación, lo que se consigue es retomar el tamaño anterior. Si lo aplicaríamos a la Figura 24 (a), el resultado final sería una imagen únicamente con los elementos de interés en su tamaño real como la de la Figura 24 (b).



**Figura 24: Imagen resultado de aplicar (a) Erosión. (b) Dilatación.**

Los filtros habrá que utilizarlos con especial atención de forma que consigamos con ellos el mayor acercamiento posible al objeto real. Realizar incorrectamente el procesamiento morfológico puede significar la aparición de errores considerables.

Por tanto, de este script, obtenemos las imágenes en blanco y negro en las que únicamente tenemos nuestro objeto de interés.

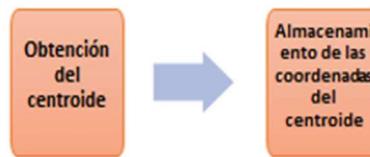
3) Centroide del objeto.

Figura 25: Diagrama de bloques 3er script.

Con dichas imágenes en blanco y negro como entrada, ya se estará en disposición de obtener el centroide del área blanco, que es nuestro objeto de interés. Es muy posible que al realizar este paso, nos detecte más de un área en blanco, debido a pequeños ruidos en la imagen, normalmente áreas de un único pixel. Sin embargo, nuestro objeto de interés va a ser de un área mucho mayor.

Para la obtención del centroide, se empleará una función de MATLAB que obtiene tanto el área del objeto de interés, como el centroide y el rectángulo más pequeño que lo contiene, también conocido como “Bounding Box”.

Sin embargo, en nuestro algoritmo sólo haremos uso del centroide y será el dato que almacenemos en 4 matrices, dos para la cámara izquierda (una del eje X y otra del eje Y) y las otras dos para la cámara derecha. Se almacenarán para cada imagen o instante de tiempo.

4) Determinación de la distancia de las cámaras al objeto.

En este script, únicamente se realiza el cálculo de la distancia de las cámaras al objeto.

La entrada a este script serán las matrices de coordenadas del eje X obtenidas en el paso anterior, y habrá que implementar la fórmula de la distancia al objeto demostrada en el apartado de Estado del Arte y que aquí recordamos:

$$Z = f \frac{b}{d}$$

Primeramente, habrá que calcular la disparidad de la fórmula, que viene dado como  $d = x_l - x_D$ , siendo  $x_l$  y  $x_D$  las coordenadas del objeto en el eje X de la imagen izquierda y derecha respectivamente, almacenadas en dichas matrices de entrada.

Siendo  $b$  la distancia entre cámaras, totalmente conocido, y  $f$  la distancia focal dada en la calibración, podemos obtener la distancia del objeto de interés a las cámaras, valor que se almacena para cada instante en una variable Z, que será el parámetro de salida.

### 5) Reconstrucción de la escena 3D.

Finalmente, sólo queda la implementación de las fórmulas (2) y (3), que nos darán las otras coordenadas del objeto.

$$X = \frac{Z}{f}(x_I - x_o) \qquad Y = \frac{Z}{f}(y_I - y_o)$$

En este script tendremos como parámetros de entrada la variable  $Z$ , la distancia focal  $f$ , y los parámetros correspondientes al punto principal  $(x_o, y_o)$  y las coordenadas del objeto de interés en cada instante de tiempo en la cámara izquierda  $(x_I, y_I)$ .

Calculadas las coordenadas  $X$  e  $Y$  para cada instante, y con la coordenada  $Z$  calculada anteriormente también para cada instante, ya se puede realizar la reconstrucción de la escena 3D. En este caso, habrá que considerar los ejes que utiliza MATLAB para la representación de gráficos, los cuales se ven en la Figura 26 (a).

Sin embargo, como veremos en el apartado de resultados, nuestro parámetro  $Z$  por nomenclatura es la distancia al objeto y no corresponde con el eje  $Z$  de la representación gráfica de Matlab. El eje  $Z$  lo colocaremos en el eje  $X$  de MATLAB. Nuestro eje  $X$  será el eje  $Y$  de MATLAB y nuestro eje  $Y$  será la altura del cubo.

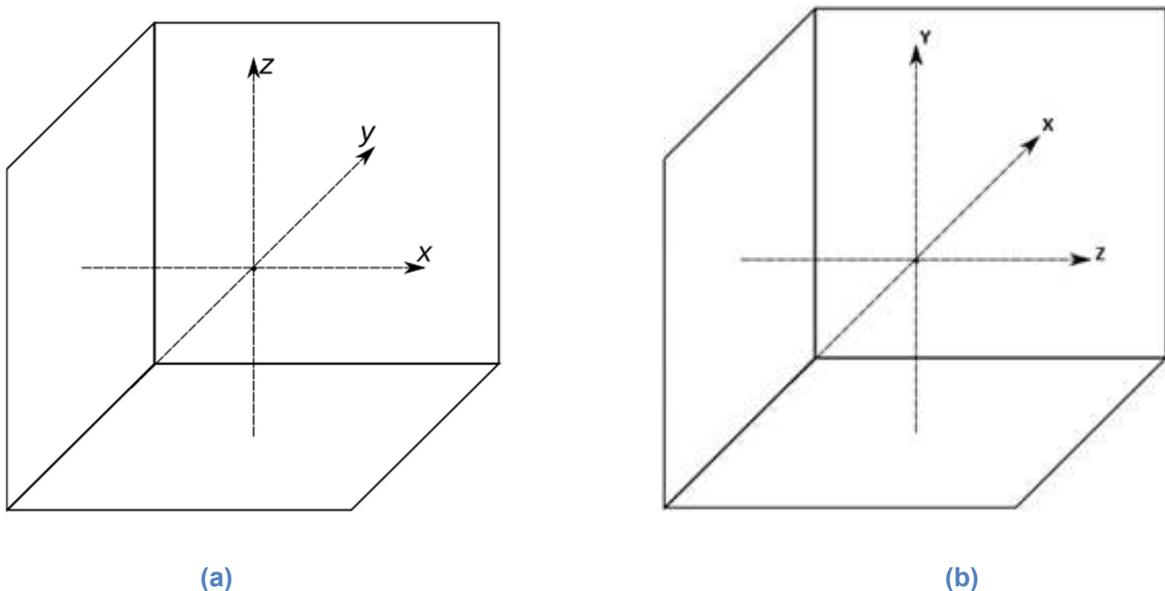


Figura 26: (a) Ejes de MATLAB. (b) Ejes empleados en el algoritmo.

## 9. ÁLBUM FOTOGRÁFICO DEL VIDEO

Previamente a la visualización de los resultados, se va a proporcionar un álbum fotográfico con el movimiento de la persona, y las imágenes en blanco y negro que se han ido obteniendo para ciertos instantes.



Frame 1



Frame 10



Frame 30



Frame 50



Frame 70



Frame 90



Frame 110



Frame 130

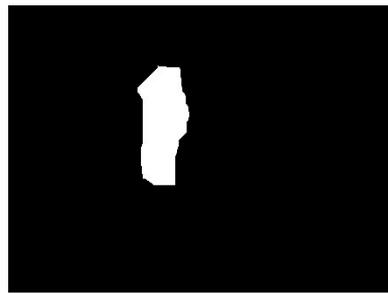


Frame 159

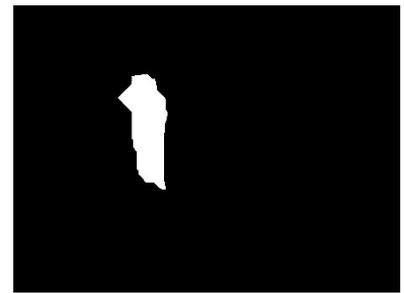
**Figura 27: Toma fotográfica del objeto de interés**



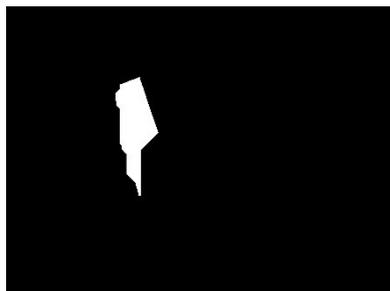
Frame 1



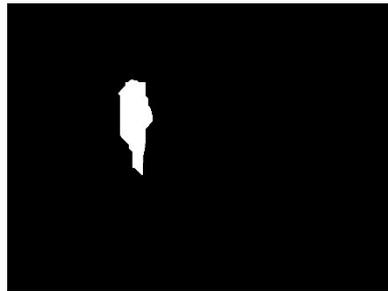
Frame 10



Frame 30



Frame 50



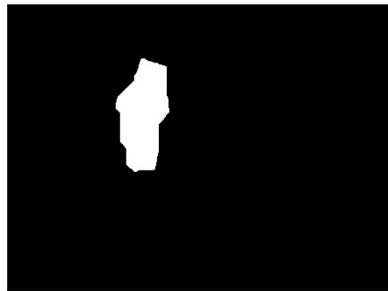
Frame 70



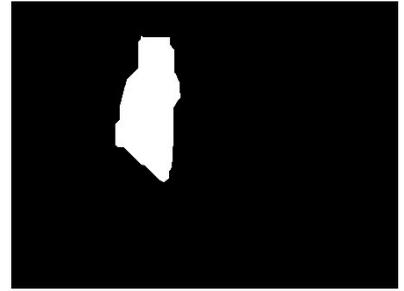
Frame 90



Frame 110



Frame 130



Frame 159

**Figura 28: Toma fotográfica en blanco y negro.**

## 10. RESULTADOS OBTENIDOS

La reconstrucción 3D obtenida se puede ver en la Figura 29. Las dimensiones están dadas en centímetros y están acorde al aula donde ha sido grabado el video. También se aporta la reconstrucción de la escena vista en planta (Figura 30), debido a que la variación en el eje y apenas varía (entre 40 y 60 centímetros), variación producida debido a que a medida que la persona se va alejando, el objeto de interés se va haciendo más pequeño y el centroide queda colocado con más altura en la escena 3D.

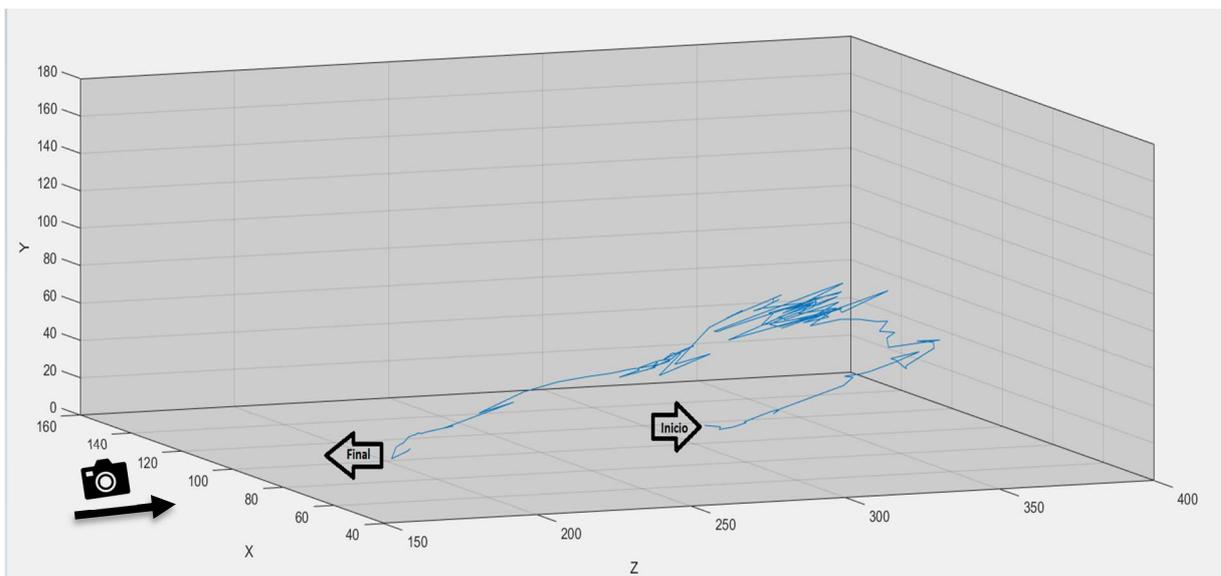


Figura 29: Reconstrucción 3D.

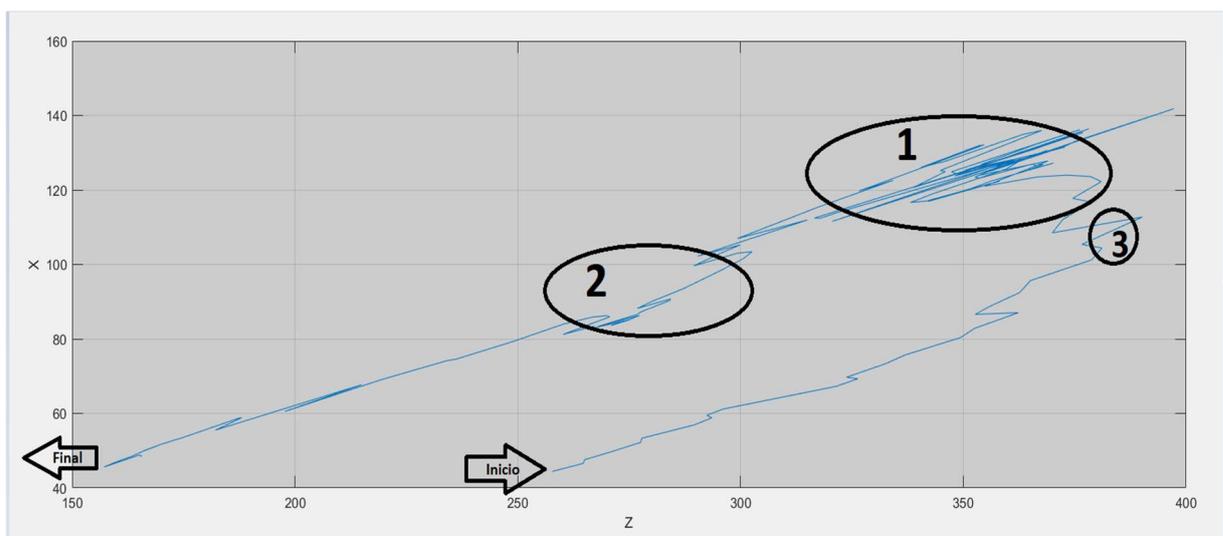


Figura 30: Reconstrucción 3D en planta.

## 11. CONCLUSIONES

- El tracking realizado corresponde en gran medida con el movimiento desarrollado por el objeto de interés. Visualizando la reconstrucción de la escena en su conjunto, la silueta generada se asemeja a la trayectoria que describe nuestro objeto de interés.  
Sin embargo, se produce un error considerable en la zona de giro (círculo numerado como 1), donde la persona da la vuelta para comenzar la fase 2: el acercamiento hacia las cámaras. Los círculos numerados como 2 y 3 también representan un cierto error, aunque bastante menores que los del círculo 1.  
Por tanto, este algoritmo nos permite un tracking sencillo y eficaz con bajo coste computacional para aplicaciones no muy exigentes en cuanto a precisión.
- La teoría existente es totalmente válida, sin embargo, habrá que determinar si las simplificaciones consideradas se pueden realizar en cualquier otro sistema.
- Una buena calibración estereoscópica, cumpliendo los requisitos, se traduce en errores mucho menores.
- Habrá que elegir convenientemente el método de cálculo del flujo óptico, así como el umbral de binarización.
- El uso de los filtros para quedarnos con la mayor aproximación posible de nuestro objeto de interés debe realizarse con especial atención, ya que puede marcar el éxito del trabajo.
- El desarrollo del proyecto puede ser de larga duración debido a los numerosos pasos que se deben realizar, los cuales deben ser implementados de forma óptima.

## 12. LIMITACIONES Y PROBLEMAS DEL ESTUDIO. SOLUCIONES ADOPTADAS

Durante el desarrollo del algoritmo, ha sido necesario ir realizando muchos cambios. Estas son las principales limitaciones y problemas surgidos durante el desarrollo del algoritmo a los que se ha tenido que dar solución:

- Como se ve en las imágenes, la persona se mueve con un cartón blanco. La explicación de este hecho reside en que con ropa oscura, al hacer el procesado de imagen, se perdía mucha información. Al realizar el cálculo del flujo óptico (resta de imágenes), con una camiseta oscura, la escala de grises de esa parte del cuerpo, se acercaba en mayor medida al negro, y por este motivo, se hacía realmente difícil segmentar en buenas condiciones el objeto de interés en su totalidad. Por eso, se decidió usar dicho cartón blanco para tapar la camiseta.

Se considera éste el mayor de los problemas en la zona de giro (círculo 1), ya que en esa zona el cartón blanco debe darse la vuelta rápidamente para ponerlo por la parte de adelante del cuerpo. Por este motivo, en las imágenes de esa zona, el cartón blanco se coloca de canto o en malas condiciones, y esto supone que la segmentación debe hacerse de la camiseta, de forma que en esa zona se pierde una gran parte de información y el error se acentúa.

- La calibración, como hemos visto, ya nos introduce error, es decir, desde el principio ya contamos con un cierto error, que aunque pequeño, va sumando. Por este motivo es por lo que se ha hecho hincapié en la importancia de reducir el error de calibración.
- Experimentalmente, casi nunca quedan los ejes ópticos paralelos entre sí, por lo que es una fuente más a añadir en el error. Para minimizarlo, habrá que colocar las cámaras minuciosamente y no mover ni tocar la base donde se sustentan las cámaras en la adquisición del video y de las imágenes de calibración.
- La iluminación también juega un papel importante siempre que existen cámaras de por medio. Anteriormente se ha mencionado que el ventanal transparente y la iluminación artificial abarcaban todo el peso de la problemática de la iluminación. Es por ello que se ha

tratado que las cámaras quedasen colocadas simétricamente tanto al ventanal como a la iluminación artificial, de forma que llegase la misma intensidad de luz a ambas cámaras.

- La mala selección del umbral de binarización y la mala aplicación de los filtros puede suponer una pérdida de información importante capaz de alterar en gran medida el valor de la disparidad y por tanto, el valor de la distancia al objeto. Por ello, se ha intentado no trabajar con un número excesivo de imágenes, para poder ir comprobando que éste proceso se hacía de una forma apropiada.
- Para reducir el error que nos aportaban estos problemas, se decidió hacer el rectificando de imágenes, pero éste no aportaba nada, ya que en este algoritmo no se hace la correspondencia de puntos, y no es necesario que los pares de puntos estén situados en la misma línea epipolar.

## 13. PLANIFICACIÓN

### 13.1. DESCRIPCIÓN DE TAREAS, FASES Y PROCEDIMIENTOS

#### Tarea 1: Formación (28 días)

- Descripción: familiarizarse con MATLAB, con las cámaras GoPro y con los mandos inalámbricos. Además, habrá que conocer la teoría a implementar en el desarrollo del algoritmo.
- Hitos: conocer MATLAB y determinar las condiciones que deben cumplir la calibración y el algoritmo de detección del objeto y su seguimiento.
- Entregables: informe relativo al apartado de Estado del Arte, el cual se incluye en este trabajo fin de grado.

#### Tarea 2: Calibración estereoscópica (13 días)

- Descripción: realizar la calibración de ambas cámaras para relacionarlas entre sí.
- Hitos: obtener los parámetros intrínsecos y extrínsecos de las cámaras.
- Entregables: explicación del proceso de calibración en el apartado 8.1.

#### Tarea 3: Adquisición de imágenes (12 días)

- Descripción: en esta tarea se incluye la determinación de la disposición de las cámaras, así como grabar los videos y convertirlos a frames en el tamaño adecuado.
- Hitos: obtener las imágenes redimensionadas en el tamaño adecuado.
- Entregables: se incluye un álbum fotográfico en el proyecto.

#### Tarea 4: Caracterización de imágenes (16 días)

- Descripción: obtener el flujo óptico y aplicar el proceso de binarización y el procesamiento morfológico.

- Hitos: obtener la forma más aproximada del objeto de interés en las imágenes en blanco y negro.
- Entregables: código matemático en MATLAB, el cual se incluye en el anexo.

#### **Tarea 5: Obtención del centroide (6 días)**

- Descripción: coordenadas del centroide del objeto de interés.
- Hitos: matriz de coordenadas del centroide.
- Entregables: código matemático en MATLAB, el cual se incluye en el anexo.

#### **Tarea 6: Determinación de la distancia al objeto (11 días)**

- Descripción: aplicación de la fórmula teórica para obtener la distancia de las cámaras al objeto.
- Hitos: matriz de la distancia Z.
- Entregables: código matemático en MATLAB, el cual se incluye en el anexo.

#### **Tarea 7: Reconstrucción de la escena 3D (7 días)**

- Descripción: obtener una representación de la escena 3D.
- Hitos: tracking del objeto en movimiento.
- Entregables: código matemático en MATLAB, el cual se incluye en el anexo, además del apartado de análisis de resultados del presente documento.

#### **Tarea 8: Análisis de resultados y conclusiones (10 días)**

- Descripción: evaluar si se han logrado los objetivos analizando los resultados y sacando conclusiones.
- Hitos: conclusiones en relación a los resultados.
- Entregables: apartados referidos al análisis de resultados y conclusiones del presente documento.

### **Tarea 9: Documentación (14 días)**

- Descripción: escritura del documento. Se realiza una vez se haya desarrollado el algoritmo y obtenidas las conclusiones.
- Hitos: documentar el proyecto.
- Entregables: documento de trabajo de fin de grado.

### 13.2. DIAGRAMA GANTT

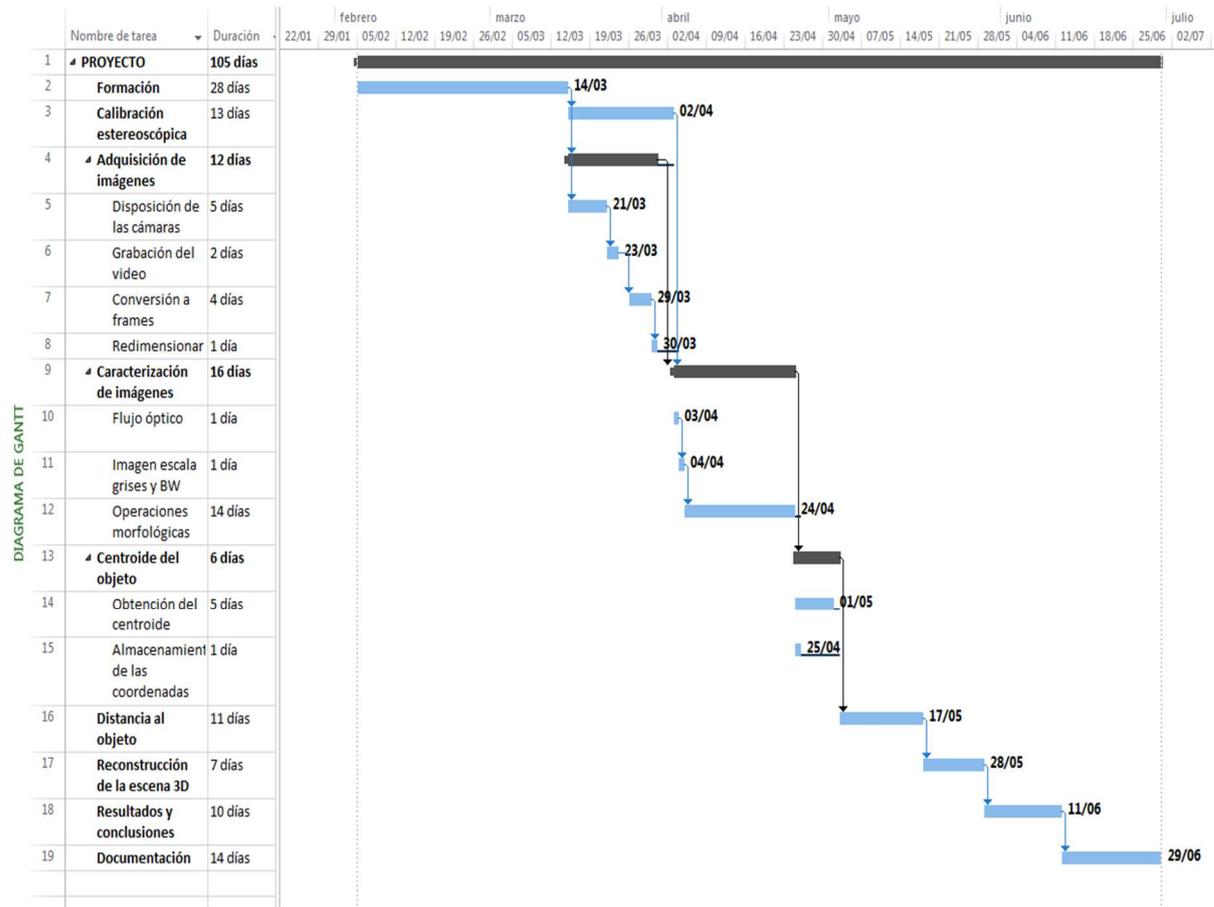


Figura 31: Diagrama Gantt

La duración total del proyecto es de 105 días, con fecha de comienzo el día lunes 5 de Febrero de 2018 y con fecha de finalización el día viernes 29 de Junio de 2018, es decir, alrededor de 5 meses naturales.

## 14. PRESUPUESTO

HORAS INTERNAS			
	Horas empleadas (h)	Coste unitario (€/h)	Coste (€)
Ingeniero Junior	400 h	25 €/h	10.000,00 €
Director del Proyecto	50 h	60 €/h	3.000,00 €
<b>TOTAL</b>			<b>13.000,00 €</b>

AMORTIZACIONES				
	Precio inicial	Vida útil	Utilización en el proyecto	Amortización
Ordenador personal	1.050,00 €	5 años	5 meses	308,82 €
MATLAB	2.000,00 €	3 años	4 meses	222,22 €
Cámaras GoPro Hero 3	310,00 €	3 años	5 meses	43,06 €
Mandos inalámbricos	120,00 €	3 años	5 meses	16,67 €
Word	115,00 €	1 año	1 mes	9,58 €
<b>TOTAL</b>				<b>600,35 €</b>

OTROS GASTOS			
	Coste kWh	Horas	Coste
Alquiler Aula Universidad			1.000,00 €
Conexión a Internet			50,00 €
Electricidad	0,115 kWh	400 h	46,00 €
<b>TOTAL</b>			<b>1.096,00 €</b>

IMPREVISTOS (5%)	734,82 €
<b>SUBTOTAL</b>	<b>14.830,82 €</b>
IMPUESTOS (21%)	3.114,47 €
<b>TOTAL</b>	<b>17.945,29 €</b>

Tabla 1: Presupuesto

## BIBLIOGRAFÍA

1. **G. Pajares, JM. Guerrero, M. Guijarro.** Técnicas de procesamiento de imágenes estereoscópicas.
2. **M. Montalvo.** Técnicas de visión estereoscópicas para determinar la estructura tridimensional de la escena.
3. Modelo PinHole. [Fecha acceso: 07/07/2018].  
[https://en.wikipedia.org/wiki/Pinhole\\_camera\\_model](https://en.wikipedia.org/wiki/Pinhole_camera_model).
4. University of Toronto – Epipolar geometry [Fecha acceso: 10/07/2018]  
<http://www.cs.toronto.edu/~jepson/csc420/notes/epiPolarGeom.pdf>
5. **J. Borg.** Detecting and tracking players in football using stereo vision (2007).
6. **D. Perdomo, J.Alonso, C. Travieso, M. Ferrer.** Automatic scene calibration for detecting and tracking people using a single camera. Volume 26.
7. **O. D. Faugueras, G. Toscani.** The Calibration Problem for Stereoscopic Vision.
8. **Z. Zhang.** A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, 1998.
9. OpenCV 2.4.13.7 documentation. Camera calibration with OpenCV. [Fecha acceso: 15/06/2018].
10. Lecture Stereo Vision. [Fecha acceso: 10/03/2018].
11. Aplicaciones de visión artificial. [Fecha acceso 6/06/2018].  
<http://www.bcnvision.es/aplicaciones-vision-artificial>.  
<https://www.aisvision.com/es/noticias/vision-artificial-coche-autonomo/>.
12. **D. Armando, F. Usama, JF. Florez-Marulanda.** 3D Reconstruction of Scenes by Means of a Steoroscopic Vision System, Based on Feature Extraction and Developed in OpenCV.
13. **I. Danilo García.** Visión Artificial y Procesamiento Digital de Imágenes usando Matlab. Ibarra-Ecuador 2008.
14. **I. Dario.** Estimación de un mapa disperso de profundidad usando visión estereo.  
<https://www.youtube.com/watch?v=hab07nMeUzA>.
15. Manual GoPro White Edition [Fecha acceso: 09/07/2018].  
<https://www.manualslib.com/manual/512995/GoPro-Hero-3-White-Edition.html?page=26#manual>.
16. MATLAB Help. [Fecha acceso: 11/07/2018].

17. Precios mandos a distancia. [Fecha acceso: 13/07/2018].  
<https://www.pccomponentes.com/mando-a-distancia-inalambrico-para-gopro-hero-3-3-4>.
18. Universidad Nacional de San Luis. Procesamiento Morfológico. [Fecha acceso: 11/07/2018] <http://dsp1.materia.unsl.edu.ar/Procesamiento%20Morfologico.pdf>.