Visión Artificial

Juan Diego Gallego Nicolás
jdiego.gallego@um.es

23/03/2025

Entrega Parcial

Índice

Índice de figuras

1. Introducción

Este documento recoge explicaciones y resultados de los ejercicios realizados durante el desarrollo de la asignatura siguiendo las instrucciones del repositorio de GitHub albertoruiz/umucv/blob/master/notebooks/ejercicios.ipynb (revisado el 20/03/2025). Cada ejercicio se corresponde con la sección del documento con el mismo nombre. A su vez, cada sección se divide en los subapartados: Marco Teórico, Resultados y Comentarios.

La sección Marco Teórico servirá como resumen de los contenidos teóricos desarrollados en la teoría necesarios para la realización de la práctica. A continuación, en el apartado de Resultados se hablará del trabajo realizado (código y pruebas) y de cómo se han aplicado los conocimientos teóricos. Finalmente, la sección Comentarios se reserva para realizar alguna reflexión y plasmar las conclusiones y opiniones sobre el ejercio.

Las imágenes incluidas son capturas de pantalla de Geogebra clásico (www.geogebra.org/classic?la
y Geogebra 3D (www.geogebra.org/3d?lang=es), material de la asignatura, gráficas de matplotlib e imágenes tomadas con un Google Pixel 8 Pro.

A lo largo del documento se hará mención a una serie de scripts de python desarrollados por el profesor Alberto Ruiz. Estos scripts se encuentran en el repositorio de GitHub albertoruiz/umucv/ (revisado el 20/03/2025).

2. Calibración

2.1. Calibración: Marco Teórico

Una cámara es un dispositivo compuesto esencialmente por una lente que enfoca luz en un plano de proyección. Ese plano de proyección físico se sitúa detraś de la lente y de forma invertida. Tras deshacer la inversión, los dispositivos de visualización presentan la imagen en la orientación correcta. Nótese que el proceso de reconstrucción de la imagen genera un plano de proyección virtual con las mismas dimensiones y a la misma distancia a la lente que el plano real 1.

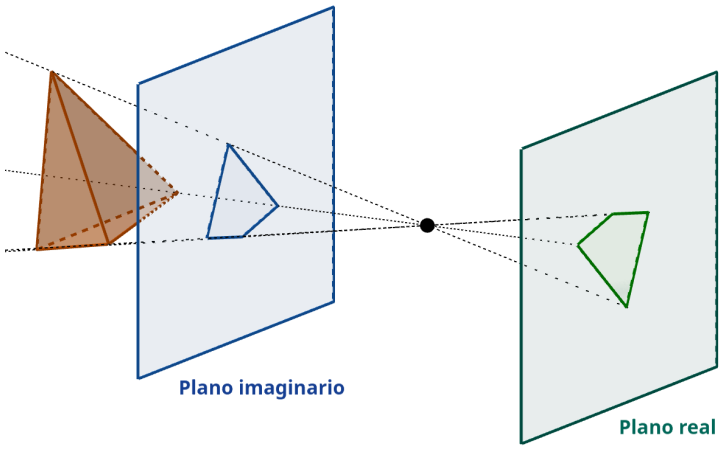

Figura 1: Modelo de una cámara.

Hay dos parámetros esenciales que describen las propiedades de una cámara: su resolución y su distancia focal. La resolución hace referencia a las dimensiones del plano de proyección. Esta se mide en píxeles (unidad mínima de representación de color) y se puede presentar como el par $W \times H$ o como el número de total píxeles en el plano. La distancia focal es la distancia entre el foco de la lente y el plano de proyección. También se mide en píxeles.

Calibrar la cámara consiste en aproximar lo máximo posible estos parámetros. Una vez conocidos, se pueden hacer mediciones en el mundo real de distancias, dimensiones y ángulos como veremos más adelante. En el caso de esta práctica, usaremos un patrón cuadriculado para este proceso 2, aprovechando las distorsiones de tamaños y de los ángulos rectos.

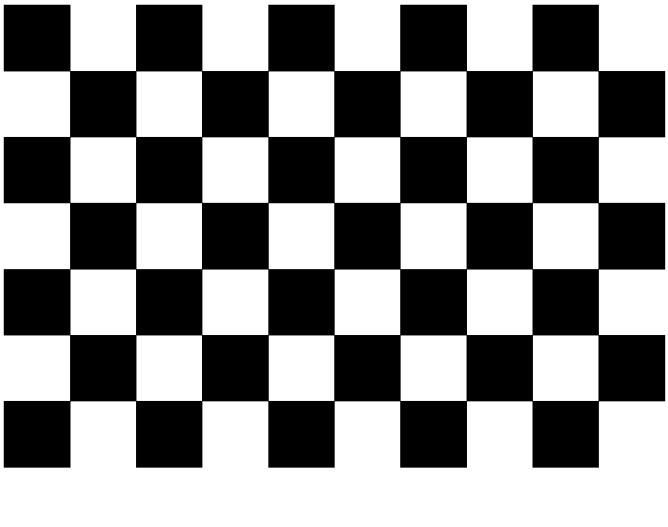

Figura 2: Patrón para la calibración de la cámara.

El proceso de calibración da como resultado una matriz $K \in \mathcal{M}_{3\times3}(\mathbb{R})$:

$$K = \begin{bmatrix} f & 0 & o_x \\ 0 & f_r & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

donde:

- f, f_r es la distancia focal y un valor cercano
- (o_x, o_y) son las coordenadas en las que se sitúa la proyección del foco de la cámara (normalmente $(W/2, H/2)$)

Una propiedad de la cámara interesante, derivada de los parámetros anteriores, es el campo de visión o FOV (Field Of View). Es una medida angular que representa las amplitudes horizontal y veltical máximas reconocibles por la cámara 3. Aplicando trigonometría básica se derivan las siguientes fórmulas para el FOV horizontal y vertical:

$$FOV_H = 2 * arctan\left(\frac{w}{2f}\right);\ FOV_V = 2 * arctan\left(\frac{h}{2f}\right)$$

También, a partir del FOV se deduce la unidad angular mínima medible como $\frac{FOV_H}{W} = \frac{FOV_V}{H}$.

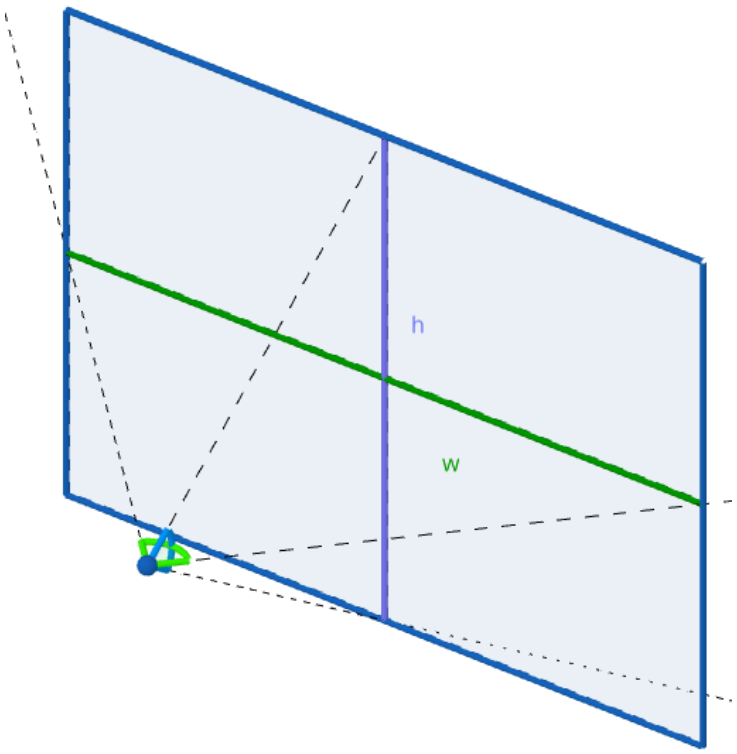

Figura 3: FOV

La última propiedad que necesitamos para la realización de la práctica es la relación entre tamaño, tamaño aparente, distancia a la lente y distancia focal. Esta la podemos obtener aplicando el Teorema de Tales aplicado a los triángulos formados por el foco de la cámara y los pares de extremos reales y virtuales del objeto proyectado 4.

Claro está que para que esto funcione el objeto a medir ha de estar lo más centrado posible, ya que en los extremos de la imagen se produce distorsión. En las imagenes tomadas para este documento se ha intentado centrar lo máximo posible las imagenes evitando así dicho efecto.

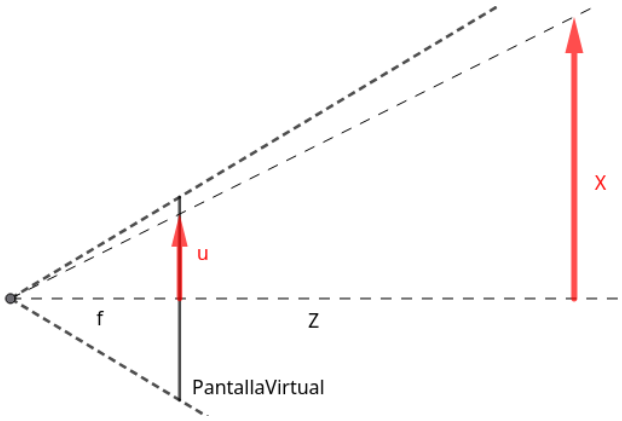

Figura 4: Medidas sobre la imagen.

2.2. Calibración: Resultados

La realización de los ejercicios se puede ver en el notebook de python 'calibracion.ipynb'.

El primer paso para la realización de la práctica es la calibración de la cámara. Para comenzar, he utilizado el programa 'stream.py' para tomar imágenes del patrón 2 desde una variedad de posiciones y ángulos 5.

Figura 5: Fotografía tomada para calibración.

Es importante que exista una variedad en los ángulos de captura ya que la funcionalidad findChessboardCorners de OpenCV utiliza la distorsión en los ángulos de las esquinas de los cuadrados para calcular los parámetros que buscamos.

Una vez tenemos las imágenes, usamos el programa 'calibracion.py' para calibrar la cámara, dando como resultado la matriz de calibración:

$$K = \begin{bmatrix} 561{,}11 & 0{,}0000 & 317{,}80 \\ 0{,}0000 & 561{,}09 & 234{,}53 \\ 0{,}0000 & 0{,}0000 & 1{,}0000 \end{bmatrix}$$

para una resolución de 640×480. Con los programas 'triangulate.py' y 'verify.py' podemos comprobar que el proceso se ha completado con éxito.

Con la distancia focal obtenida podemos calcular el FOV de la cámara:

$$FOV_H = 2 * arctan\left(\frac{w}{2f}\right) = 59{,}39^\circ$$

$$FOV_V = 2 * arctan\left(\frac{h}{2f}\right) = 46{,}32^\circ$$

Con este valor, podemos calcular la altura a la que tendríamos que situar la cámara para capturar por completo una pista de baloncesto. A partir de sus dimensiones (28×15 metros):

$$tan\left(\frac{FOV_H}{2}\right) = \frac{28/2}{h_1} \Rightarrow h_1 = \frac{14}{tan(FOV_H/2)} = 24{,}5485625m$$

$$tan\left(\frac{FOV_V}{2}\right) = \frac{15/2}{h_2} \Rightarrow h_2 = \frac{7{,}5}{tan(FOV_V/2)} = 17{,}5346875m$$

De entre los dos resultados tenemos que tomar el mayor, pues necesitamos que la pista se vea tanto a lo largo como a lo ancho.

Buscando el tamaño del resto de líneas de la pista podemos hacer una representación en 3D que nos hace a la idea de la situación que experimentaría la cámara en estas condiciones 6:

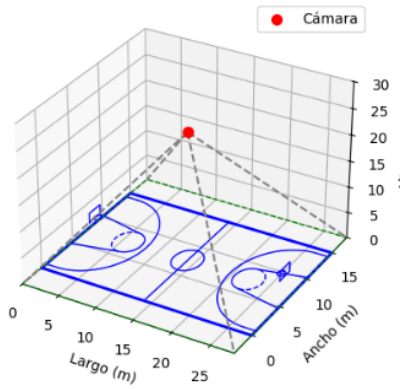

Figura 6: Cámara sobre la pista de baloncesto.

Como se puede observar, sobra espacio mas a allá bandas debido a que la altura necesaria para ver la distancia entre los laterales era menor. También se intuye que al estar a una determinada altura las canastas no son entran dentro de la pirámide que constituye el espacio de visión de la cámara. Aun así, posteriormente supondremos que las canastas se encuentran a nivel de suelo.

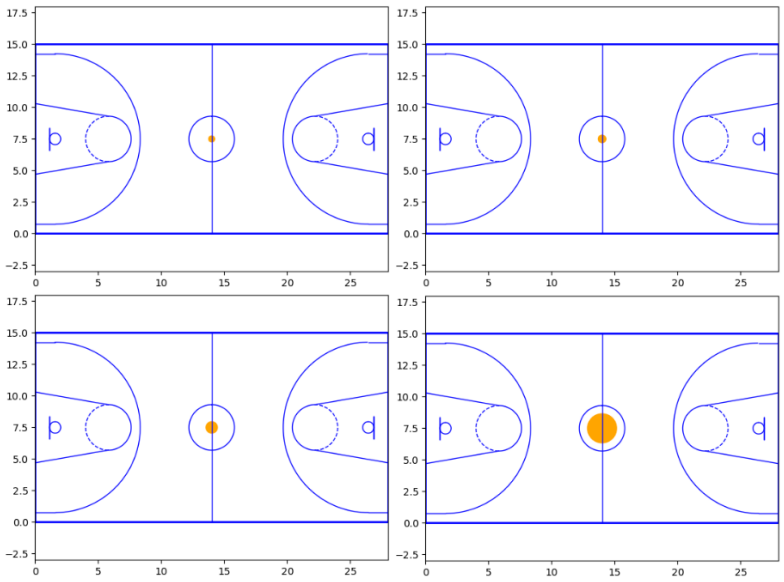

Figura 7: Pelota al 0 %, 25 %, 50 % y 80 % de la altura de la cámara.

En la imagen anterior 7 aprovechamos el modelo de la pista para simular el tamaño de una pelota de baloncesto situada a diferentes alturas del círculo central. Para la primera instancia, el tamaño de la pelota ($48cm$ de radio) es muy pequeño en comparación de la pista. En el tramo de 0 % a 50 % el tamaño aparente solo se duplica ya que esta magnitud es inversamente proporcional a la distancia a la cámara. Al 80 % de la distancia al suelo el tamaño se ha multiplicado por 5. Para valores más altos la pelota empieza a ser demasiado grande como para caber en el espacio visual de la cámara.

Cambiando de tema, vamos a explorar ahora las posibilidades de la relación

$$\frac{u}{f} = \frac{X}{Z}$$

utilizando la herramienta 'mouse.py' desarrollada en las prácticas que nos permite medir distancias en una imagen conociendo la matriz de calibración.

En primer lugar, podemos predecir la altura en píxeles que tendrá una persona a una cierta distancia de la cámara. Por ejemplo, en la imagen 8 aparezco yo mismo situado a 7,2 metros de la cámara. Sabiendo que mi altura es de 1,74 metros mi tamaño aparente debería de ser:

$$u = \frac{X}{Z} \times f = \frac{1{,}74}{7{,}2} \times 561{,}11 = 135{,}60px$$

Por tanto, la altura medida entra dentro de la precisión esperada.

Figura 8: Altura de una persona

Por otro lado, vamos a intentar medir un objeto directamente con la cámara. Para ello, situamos un muñeco a 19,5cm de la cámara y utlizamos el mismo programa para medir su tamaño aparente 9.

Figura 9: Tamaño de un objeto

Con el dato de 279px utilizamos la misma relación para obtener el tamaño del muñeco:

$$X = \frac{u}{f} \times Z = \frac{279}{561{,}11} \times 19{,}5 = 9{,}7cm$$

lo que coincide con el tamaño real con una precisión milimétrica.

Lo último que podemos hacer directamente con esta relación es calcular la distancia a la que se encuentra un objeto cuyas dimensiones conocemos. Por ejemplo, un coche que mida 4 metros de largo y que ocupe 20 píxeles en pantalla estará a una distancia de:

$$Z = \frac{561{,}11}{20} * 4 = 112{,}2m$$

El programa 'angle.py' es una variación de 'mouse.py' que permite medir ángulos sobre una imagen. Esta medición se calcula sobre el segmento que une dos puntos seleccionados. Para ello, se completan las coordenadas sobre la imagen (con respecto al centro calculado en la matriz de calibración) con una tercera componente dada por la distancia focal. Después, el ángulo entre

los dos puntos se deriva de la fórmula:

$$\langle \vec{u}, \vec{v} \rangle = |\vec{u}||\vec{v}| \cos(\widehat{\vec{u}\vec{v}})$$

Con la capacidad de calcular ángulos sobre una imagen vamos a intentar triangular la posición desde la cual se ha tomado una fotografía 10.

Figura 10: Paisaje sobre el que triangular.

El primer paso para hacer la triangulación es elegir tres puntos distinguidos. En mi caso y debido a lo sencillos que son de reconocer serán:

NC: El poste del centro comercial Nueva Condomina

TH: El psote del centro comercial Thader

MA: El Cristo de Monteagudo

Para medir su distancia angular de la forma más precisa posible primero debemos fijar una línea recta sobre la cual tomar las medidas. En mi caso tomaré la línea del horizonte 11

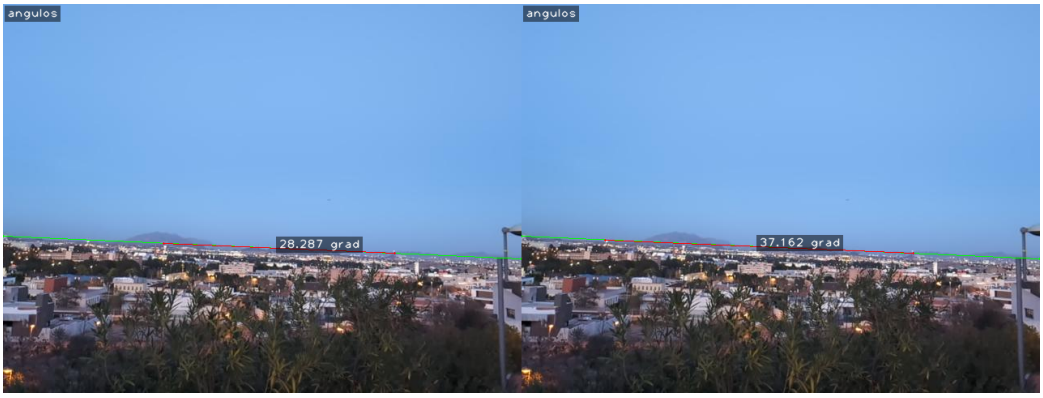

Figura 11: Midiendo ángulos.

Eligiendo los puntos de esta línea (verde) más cercanos a los puntos seleccionados podemos determinar que el ángulo $NC - MA$ es de 37,162° y el ángulo $TH - MA$ es de 28,287°. Con acceso a las distancias relativas entre estos puntos y conociendo desde dónde he tomado esta fotografía podemos ver en la imagen 12 que el error absoluto no llega a los 0,6°.

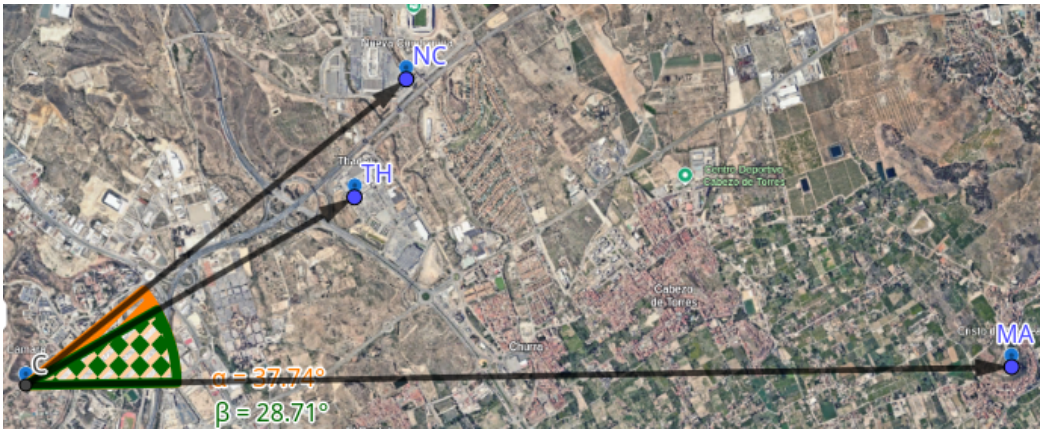

Figura 12: Ángulos reales.

3. Filtros

En esta sección se describen los diferentes filtros utilizados en el proceso de procesamiento de imágenes. Los filtros son herramientas esenciales para mejorar la calidad de las imágenes antes de aplicar técnicas de visión artificial más avanzadas.

4. Clasificador

Aquí se hablará sobre los clasificadores utilizados para la tarea específica. Se puede incluir la descripción de los algoritmos empleados, su rendimiento y los datos con los que se entrenaron.

Figura 13: Una imagen de ejemplo.

Como se puede ver en la 13, esta es una imagen de ejemplo que se utiliza en el documento.