

# Visión artificial

---

## Robótica

Alberto Díaz Álvarez y Guillermo Iglesias Hernández

Departamento de Sistemas Informáticos - Universidad Politécnica de Madrid

18 de septiembre de 2023

---

License CC BY-NC-SA 4.0

# La visión física y biológica

# ¿Qué es la visión?

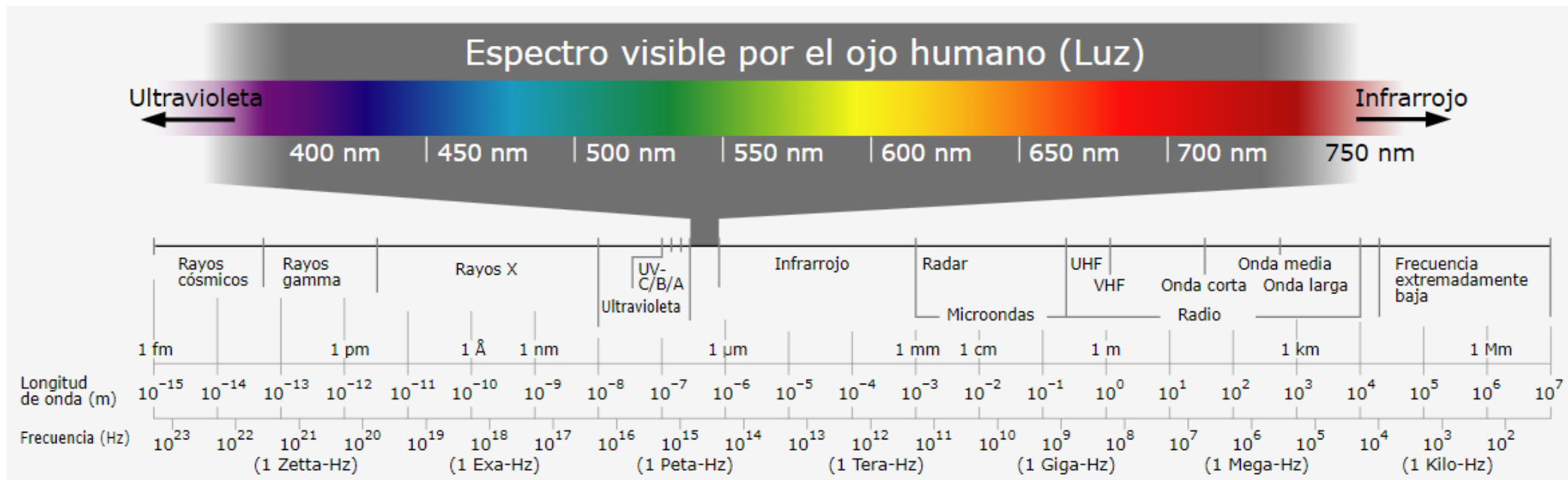
---

Capacidad de captar el entorno a través de la **captación** de **rayos de luz**

- Sentido casi esencial para la relación entre ser humano y su entorno
- Tanto es así que se estima que al menos **un tercio** del cerebro se dedica a procesar el sentido de la vista
- El procesamiento cerebral de la información visual juega un papel **crucial** en la tarea de la visión.

# El espectro electromagnético visible (I)

El ser humano solo es **capaz** de ver las longitudes de onda entre **380 y 750 nm**



**Fig. 2.1** - Espectro visible por el ojo humano. Fuente: [1]

Esto es gracias a su mecanismo de captura de luz, pero **no es el único**

# El espectro electromagnético visible (y II)

Ejemplo: El **sensor digital** de una cámara fotográfica puede **capturar frecuencias fuera del espectro visible** (especialmente en el rango infrarrojo)



**Fig. 2.2** - Sensibilidad a diferentes frecuencias de luz dependiendo del sensor que las captura. Fuente: [2]

# La visión en el ser humano

El ojo **procesa** la luz a través de células sensibles a este espectro electromagnético

## Conos

- Visión de **alta intensidad** (diurna)
- Tres tipos: **L**, **M** y **S**<sup>1</sup>

## Bastones

- Visión de **baja intensidad** (nocturna)
- Mayor cantidad que los conos



**Fig. 2.3** - Fragmento de la retina. Fuente: [3]

<sup>1</sup> Su ausencia o mal funcionamiento llevan a problemas de percepción de colores como la protanopía (**L**), deuteranopía (**M**), tritanopía (**S**) o acromatopsia (**L**, **M** y **S**)

# Rango del color percibido

No todos los colores se captan con la misma **intensidad**.



**Fig. 2.4** - Longitudes de onda capturadas por cada tipo de célula fotosensible del ojo. Fuente: [6]

# La visión neuronal



# El cerebro en la visión

Tras captar la luz, el **cerebro da sentido a los estímulos** capturados por el ojo

- El **hardware** se encarga de **captar la señal** y el **software** la **procesa adecuadamente**



**Fig. 2.5** - La información visual se procesa en un *pipeline* de menor a mayor complejidad. Fuente: [9]

La **corteza visual primaria** es la **primera capa que procesa** la información visual

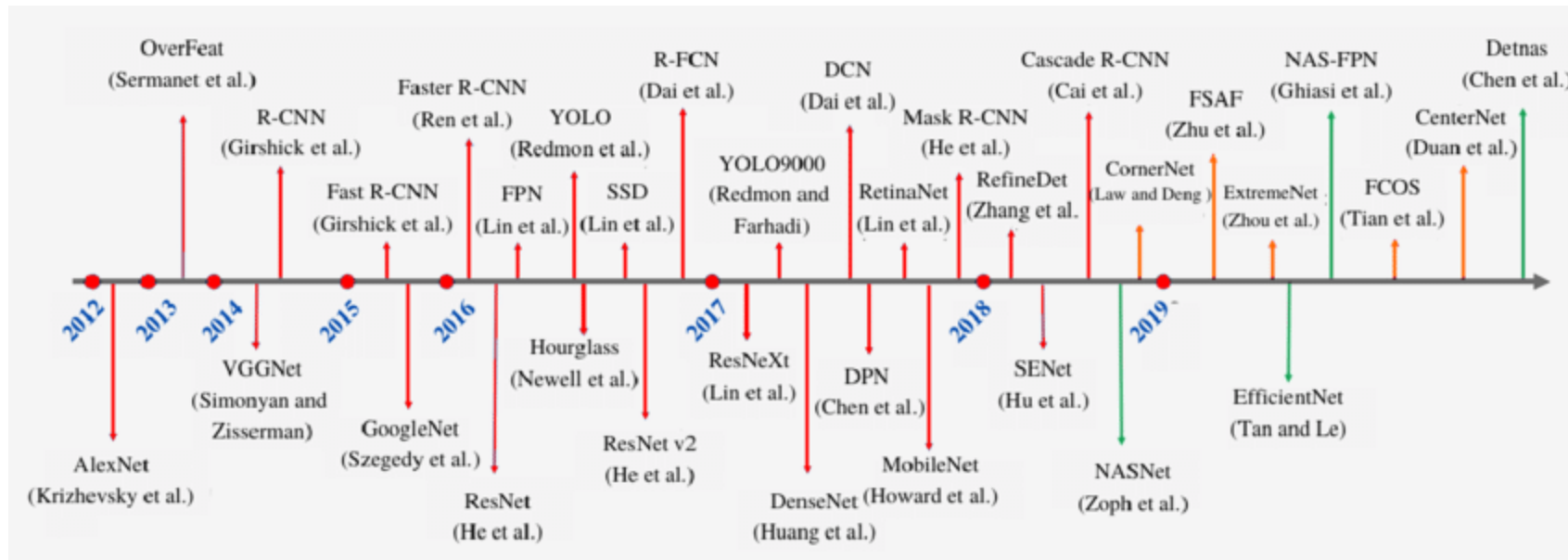
- Trabaja con formas sencillas, dando información procesada a capas superiores
- Según avanza en el procesamiento, se reconocen formas más complejas

# Introducción a la visión por computador

# ¿Qué es la visión por computador?

Rama de la inteligencia artificial encargada del procesamiento de imagen digital

- No dejan de producirse grandes avances gracias a la mejora en la técnica
- Las redes neuronales tienen **mucho** (en realidad **todo**) que ver



**Fig. 2.6** - Evolución del campo de la visión por computador durante los últimos años. Fuente: [11]

# Algunas de sus aplicaciones

---

- **Clasificación** de imágenes
- **Detección y reconocimiento** de objetos
- **Segmentación** de objetos
- **Generación** de imágenes
- **Domain-to-domain translation**
- **Text-to-image translation**
- **Superresolution**

# Captura de imagen tradicional

# De la necesidad de capturar el mundo

La tecnología ha permitido capturar el mundo de formas cada vez más precisas

- Una de estas formas ha sido **simular** el sentido de **la vista**

Leonardo da Vinci fue el precursor en capturar imágenes a través de la **proyección de la luz**

- Éste y otros permitían **capturar** parte de la realidad, pero **no su digitalización**



**Fig.1 - Camera Obscura.** Fuente: [Wikimedia Commons](#)

# La cámara *pinhole* (I)

El modelo de cámara **pinhole** permite la formación de imágenes a través del paso por una abertura **minúscula** de la luz, proyectándose sobre un plano de imagen.



**Fig.1** - Esquema de cámara tipo *pinhole*. Fuente: [Wikimedia Commons](#)

Los objetos capturados **se invierten** respecto su posición real

# La cámara *pinhole* (II)

La principal desventaja tiene que ver justo con la abertura por la que pasa la luz:

1. Si es demasiado **grande**, la imagen se **difumina**
2. Si es demasiado **pequeña**, se producen **difracciones**

De los intentos de evitar esta difuminación nace el campo de la **óptica**



**Fig.1** - Diferentes aperturas. Fuente: Ni idea.



# Modelo de lente fina

Las lentes son el principal logro de la óptica. Éstas permiten:

- Aumentar el tamaño de la apertura de la cámara
- Disminuir el desenfoque producido
- **Orientar** los rayos de luz
- **Corregir** las aberraciones producidas por otras lentes



**Fig.1** - Diagrama de rayos de una lente convexa. Fuente: [FisicaLab](#)

# Calibración en una cámara

Las diferentes lentes de una cámara producen **distorsiones** en la imagen

- También se denominan **aberraciones** y son inevitables

La **calibración** es un proceso que permite **corregir** estas distorsiones

- Se realiza a través de **algoritmos** que permiten **rectificar** la imagen
- Forma parte del campo de la **visión por computador**

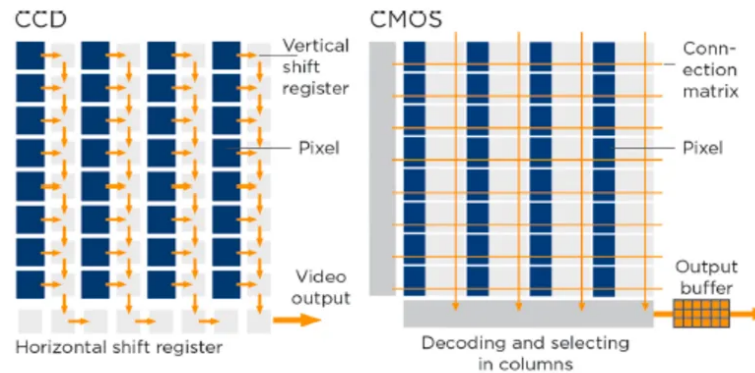


**Fig.1** - Aberración de tipo esférica. Antes y después de aplicar el algoritmo **COSTAR**. Fuente: [Revista Astronomía](#).

# La cámara digital

Dispositivo que permite capturar imágenes a través de un sensor digital

- Suelen contener un sensor con miles (millones) de receptores de intensidad de luz
- Dos tipos principales, dependiendo de cómo se convierte de fotón a electrón:
  - **CCD** (**Charge-Coupled Device**): A través de un chip externo al sensor
  - **CMOS** (**Complementary Metal-Oxide Semiconductor**): En el propio sensor



**Fig.1** - Diferencia entre sensor CCD y sensor CMOS. Fuente: [Tech Briefs](#).

Capturar sólo niveles de luz **no nos permite diferenciar colores**

Para la captura de colores existen principalmente dos soluciones:

## Triple CCD

- Un prisma divide la luz en tres haces
- Cada haz a un sensor digital distinto



**Fig.1** - Esquema de Triple CCD. Fuente: [Wikipedia](#).

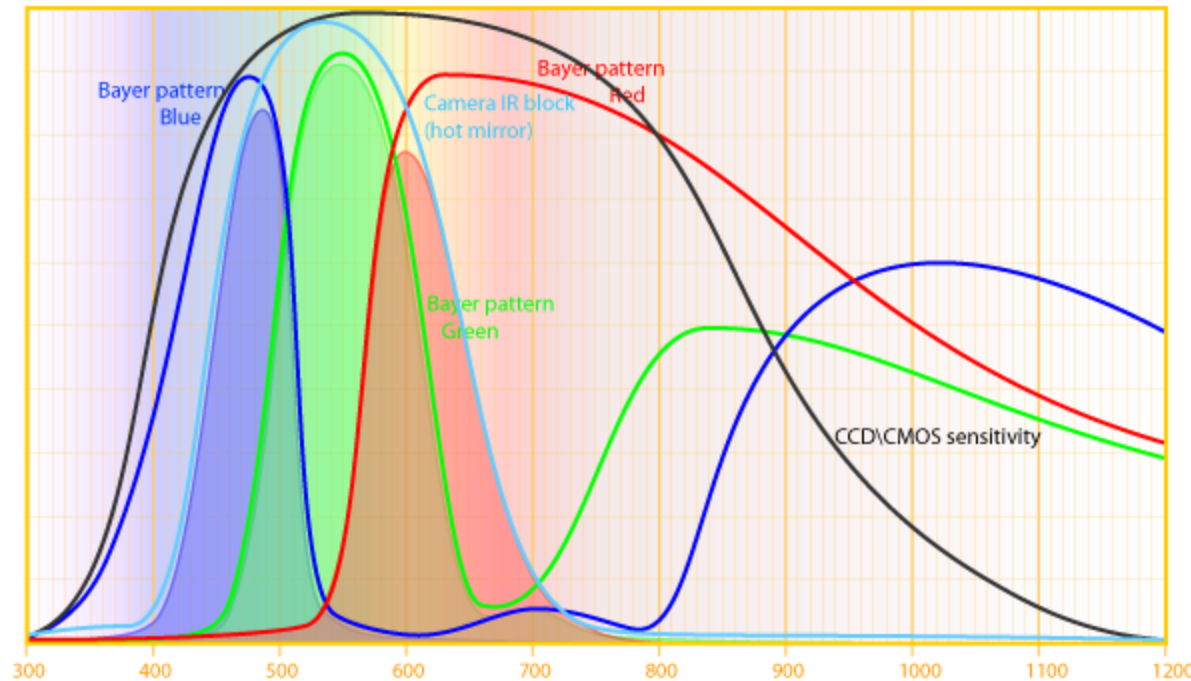
## Mosaico de Bayer

- Un filtro en cada receptor de luz
- 50% verdes, 25% rojos y 25% azules



# Limitados al rango de luz visible

Los sensores están limitados para capturar **lo que ve el ojo humano**



**Fig.1** - Esquema estándar de sensibilidad de la cámara. Fuente: [IR Photo](#).

El rango **infrarrojo** se bloquea a través del uso de **filtros**

- Estos impiden que ciertas frecuencias atraviesen la lente y lleguen al sensor

**Imagen digital**

# ¿Qué es una imagen digital?

Tras conocer cómo se captura una escena de la realidad y esta es digitalizada, ahora es vital conocer en profundidad las características de las imágenes digitales.



**Fig.1** - Ejemplo de una imagen digital.

Una imagen digital está formada por **píxeles**, los cuales corresponden con la intensidad de color para cierto punto de la imagen

# Elementos de una imagen digital

---

Una imagen digital se compone de píxeles, sin embargo existen distintas características que definen a dicha imagen:

- Tamaño de la imagen
- Número de canales
- Codificación de los canales
- Profundidad de color



# Tamaño de la imagen

El tamaño de la imagen o **resolución** corresponde con el **número de píxeles** que forman dicha imagen.

- Dependiendo de la imagen, su dimensión o **relación de aspecto** puede variar



**Fig.1** - Tamaño de una imagen digital

Conjunto de píxeles que definen la **intensidad para cierto rango de frecuencias** de una imagen

- Los distintos canales de una imagen se **combinan** para formar la imagen final
- Dependiendo del **tipo de codificación**, el **número** y **configuración** de canales puede variar



**Fig.1** - Diferentes canales de una imagen que conforman la imagen final

# Profundidad de color

---

La información de cada píxel **debe estar normalizada** en un rango de valores

- Éste indica la **precisión** a la cual fue capturada la intensidad de luz en ese punto

La **profundidad de color** indica cuánta intensidad de luz representa el valor de cada píxel

# La imagen en la informática

A la hora de programar, una imagen se representa con una **matriz numérica** de valores, en la que cada posición corresponde a un **píxel**

- A través de bibliotecas como numpy se pueden realizar transformaciones en la composición de las imágenes



Ejercicio: [Imagen\\_Digital.ipynb](#)<sup>1</sup>

[https://colab.research.google.com/drive/1k2vqQOiMMxOSOsZqHdhZ1Z-ebLhtvVvV?usp=drive\\_link](https://colab.research.google.com/drive/1k2vqQOiMMxOSOsZqHdhZ1Z-ebLhtvVvV?usp=drive_link)

**¡GRACIAS!**