

Visión artificial

Robótica

Guillermo Iglesias Hernández y Alberto Díaz Álvarez

Departamento de Sistemas Informáticos - Universidad Politécnica de Madrid

18 de septiembre de 2023

License CC BY-NC-SA 4.0

Introducción

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**

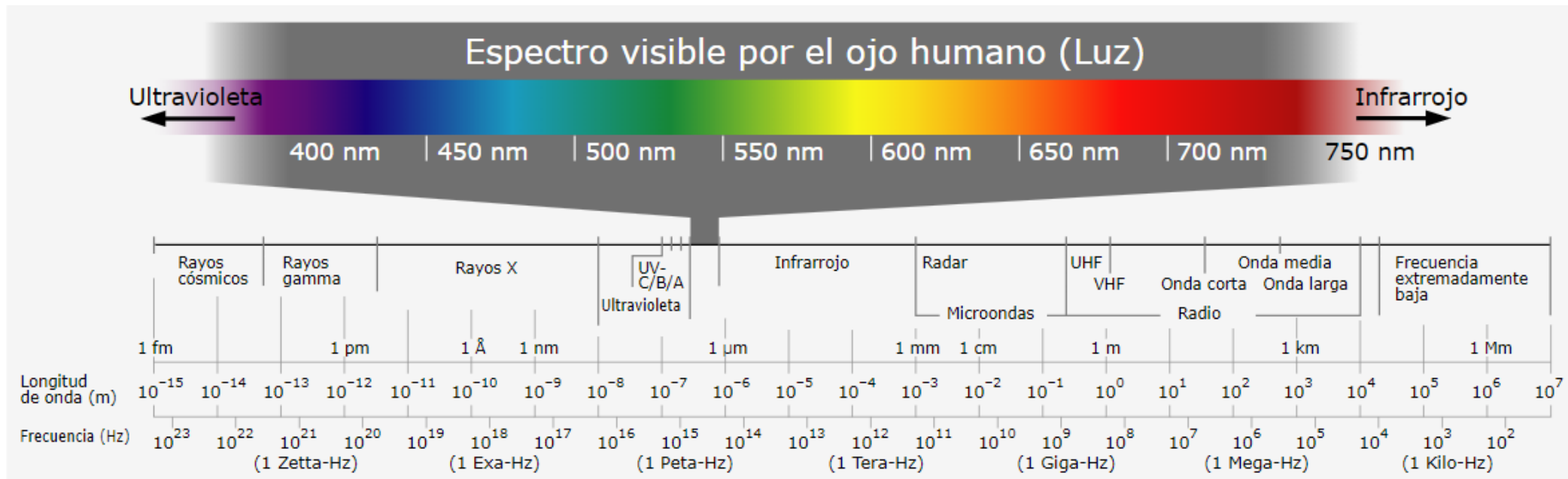


Fig. 2.1 - Espectro visible por el ojo humano. Fuente: [1]

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)

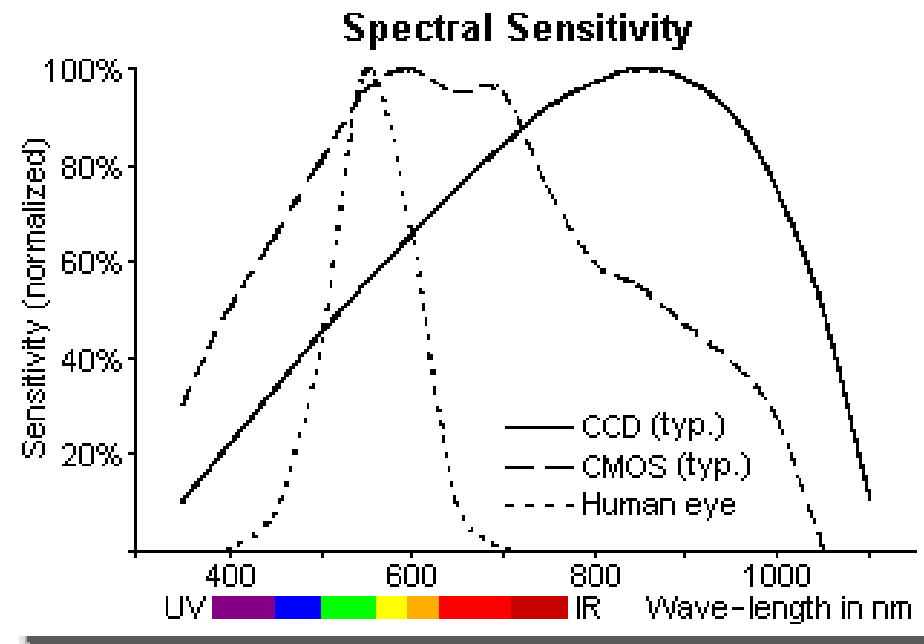


Fig. 2.2 - Sensibilidad a diferentes frecuencias de luz dependiendo del sensor que las captura. Fuente: [2]

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**¹

Bastones

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

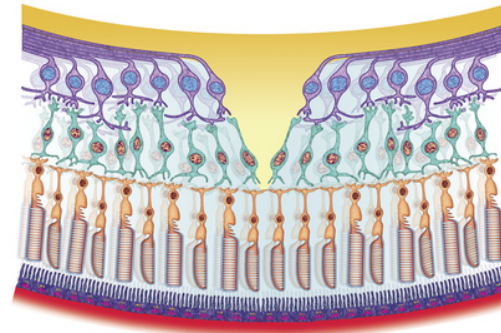


Fig. 2.3 - Fragmento de la retina. Fuente: [3]

¹ 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**.

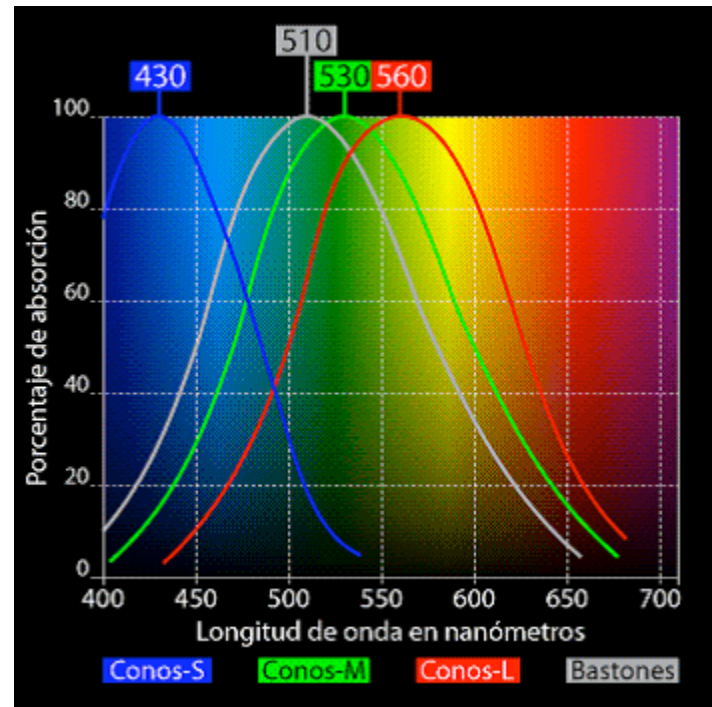


Fig. 2.4 - Longitudes de onda capturadas por cada tipo de célula fotosensible del ojo. Fuente: [6]

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**

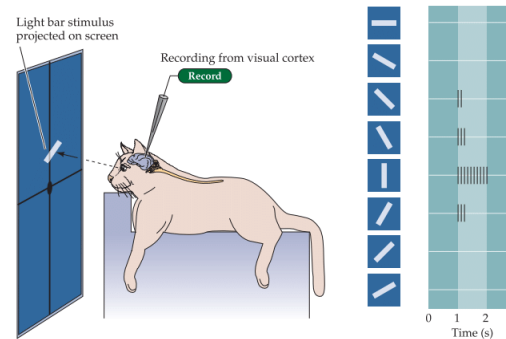


Fig. 2.5 - La información visual se procesa en un *pipeline* de menor a mayor complejidad. Fuente: [9]

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

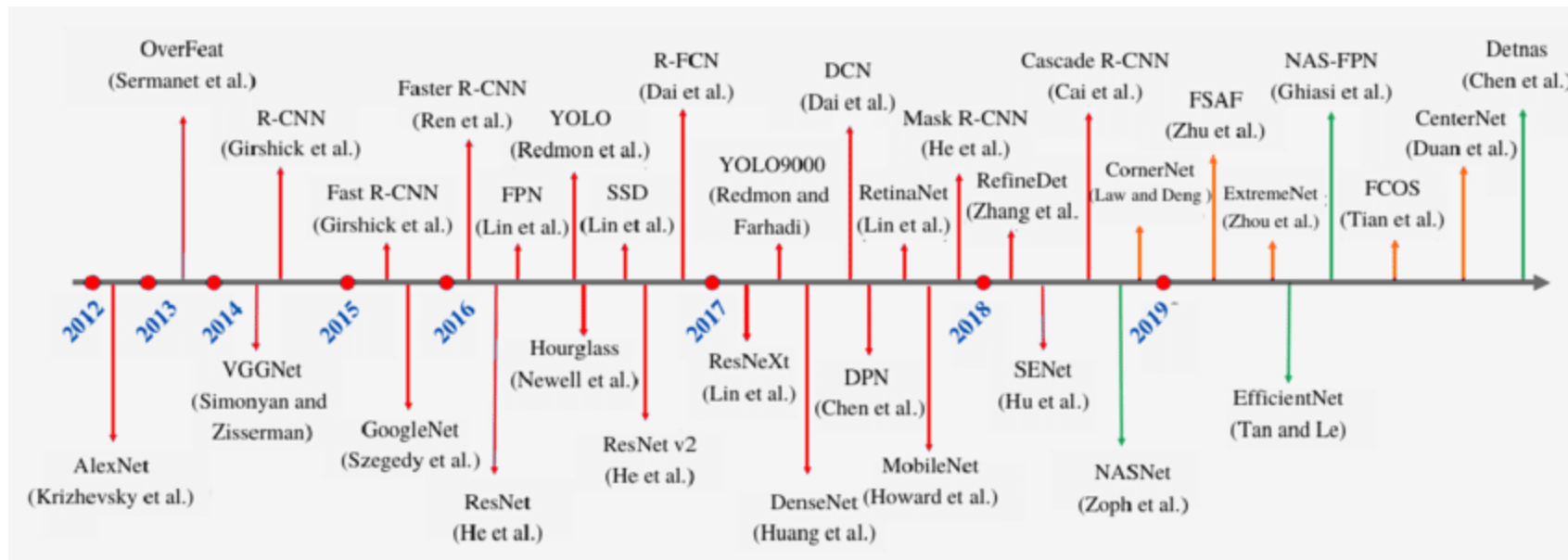


Fig. 2.6 - Evolución del campo de la visión por computador durante los últimos años. Fuente: [11]

Algunas de sus aplicaciones

- **Classificación** 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

El auge de la \alert{tecnología} de los últimos siglos ha posibilitado \alert{simular} el sentido de la vista haciendo uso de herramientas creadas por los humanos.

La idea de capturar imágenes de la realidad haciendo uso de la \alert{proyección de la luz} es un invento de Leonardo da Vinci. Todos estos precursores de la imagen digital permitían capturar porciones de la realidad, pero no su digitalización.

```
\begin{figure}  
\centering  
\includegraphics[width=0.7\textwidth]{figures/Tema 1/Camara_Oscura.jpeg}  
\end{figure}
```

La cámara \textit{pinhole}}

El modelo de cámara \textit{pinhole} permite la formación de imágenes a través del paso por una apertura \textit{minúscula} de la luz, proyectándose sobre un plano de imagen.

```
\begin{figure}  
\centering  
\includegraphics[width=0.8\textwidth]{figures/Tema 1/Pinhole_Camera.png}  
\end{figure}
```

Como se observa, los objetos capturados son \textit{invertidos} respecto su posición real

capturar el contrario, un agujero demasiado pequeño puede causar
\alert{difracciones} en la luz\cite{hecht1997}.

```
\begin{columns}[T]
\begin{column}{.48\textwidth}
\begin{figure}
\centering
\includegraphics[width=0.7\textwidth]{figures/Tema 1/Pinhole_Photo1.jpeg}
\caption{\cite{Pinhole_1}}
\end{figure}
\end{column}
\hfill
```

```
\begin{column}{.48\textwidth}
\begin{figure}
\centering
\includegraphics[width=0.9\textwidth]{figures/Tema 1/Pinhole_Photo2.jpg}
\caption{\cite{Pinhole_2}}
\end{figure}
\begin{figure}
\centering
```

De la cámara pinhole al modelo de lente fina}

Para evitar la difuminación de las imágenes de la cámara pinhole se inventó la óptica. Esta permite:

`\begin{itemize}`

`\item` Aumentar el tamaño de la apertura de la cámara.

`\item` Disminuir el desenfoque producido.

`\item` Orientar los ejes de luz.

`\end{itemize}`

`\begin{figure}`

`\centering`

`\includegraphics[width=0.7\textwidth]{figures/Tema 1/Lente_Fina.png}`

`\end{figure}`

Calibración de cámara}

Dependiendo de cada \alert{cámara de fotos} se pueden producir \alert{deformaciones} en la captura de la escena.

La \alert{calibración} es una rama de la visión por computador importante, esta consiste en la \alert{rectificación} de estos \alert{errores} para que la escena aparezca representada como en la realidad.

```
\begin{figure}  
\centering  
\includegraphics[width=0.4\textwidth]{figures/Tema 1/Calibration.jpg}  
\caption{\cite{Calibration}}  
\end{figure}
```

La cámara digital}

Una cámara está compuesta por miles de sensores, que son los que se encargan de captar la intensidad de luz para cada uno de los puntos capturados. De esta manera se captura una imagen digital.

Esta intensidad ha de ser transformada para pasar de un valor analógico a uno digital (voltaje), formando así la imagen. Dependiendo del tipo de tecnología las cámaras pueden ser:

\begin{itemize}

\item Cámaras CCD.

\item Cámaras CMOS.

\end{itemize}

```
\begin{columns}[T]
```

```
\begin{column}{.48\textwidth}
```

Una cámara \alert{CCD} realiza la conversión de photon (luz) y electrón (voltaje) a través de un chip externo a los sensores

```
\begin{figure}
```

```
\centering
```

```
\includegraphics[width=\textwidth]{figures/Tema 1/CCD.png}
```

```
\end{figure}
```

```
\end{column}
```

```
\hfill
```

```
\begin{column}{.48\textwidth}
```

Una cámara \alert{CMOS} realiza la conversión de photon (luz) y electrón (voltaje) en cada uno de los sensores de la cámara.

```
\begin{figure}
```

```
\centering
```

```
\includegraphics[width=\textwidth]{figures/Tema 1/CMOS.png}
```

```
\end{figure}
```

```
\end{column}
```

¿Cómo capturar color?

Hasta este punto se ha explicado cómo capturar los niveles de luz de una escena, sin embargo las imágenes capturadas de esta manera no son capaces de diferenciar colores.

Existen dos soluciones distintas para capturar color:

```
\begin{columns}[T]
\begin{column}{.48\textwidth}
\begin{itemize}
\item Triple CCD.
\item Matriz de Filtros Bayer.
\end{itemize}
\end{column}
\hfill
```

```
\begin{column}{.48\textwidth}
\end{column}
\end{columns}
```

¿Cómo capturar color?

Hasta este punto se ha explicado cómo capturar los niveles de luz de una escena, sin embargo las imágenes capturadas de esta manera no son capaces de \alert{diferenciar colores}.

Existen dos soluciones distintas para capturar color:

```
\begin{itemize}
\item \alert{Triple CCD.}
\item Matriz de Filtros Bayer.
\end{itemize}
```

```
\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figures/Tema 1/TripleCCD.png}
\end{figure}
```

A través de un \alert{prisma} se \alert{divide} la luz capturada en 3 haces de luz, cada uno capturado por un sensor digital distinto

¿Cómo capturar color?

Hasta este punto se ha explicado cómo capturar los niveles de luz de una escena, sin embargo las imágenes capturadas de esta manera no son capaces de diferenciar colores.

Existen dos soluciones distintas para capturar color:

```
\begin{itemize}
\item Triple CCD.
\item Matriz de Filtros Bayer.
\end{itemize}
```

```
\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figures/Tema 1/Matriz_Bayer.png}
\end{figure}
```

Se sitúa un filtro delante de cada sensor de luz, de tal manera que cada sensor captura únicamente la intensidad de la luz para cierto rango de frecuencias

¿Cómo capturar color?

Los sensores digitales son capaces de capturar parte del espectro infrarrojo. Sin embargo, esto se limita a la captura del rango de luz visible por el ser humano.

```
\begin{figure}  
\centering  
\includegraphics[width=0.8\textwidth]{figures/Tema 1/CCD_Spectrum.png}  
\caption{\cite{Bayer_Spectrum}}  
\end{figure}
```

¿Cómo capturar color?

El rango \alert{infrarrojo} se bloquea a través del uso de \alert{filtros} que impiden que los rayos de luz \alert{atraviesen} la lente.

```
\begin{figure}  
\centering  
\includegraphics[width=0.8\textwidth]{figures/Tema 1/IR_Filter.jpg}  
\caption{\cite{IR_Filter}}  
\end{figure}
```

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.

Una imagen digital está formada por `\alert{píxeles}`, los cuales corresponden con la intensidad de color para cierto punto de la imagen.

```
\begin{figure}  
\centering  
\includegraphics[width=0.8\textwidth]{figures/Tema 1/ImagenDigital.png}  
\end{figure}
```

Elementos de una imagen digital}

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

\begin{itemize}

\item Tamaño de la imagen.

\item Número de canales.

\item Codificación de los canales.

\item Profundidad de color.

\end{itemize}

Tamaño de la imagen}

El tamaño de la imagen o `\alert{resolución}` corresponde con el `\alert{número de píxeles}` que forman dicha imagen.

Dependiendo de la imagen en concreto, las `\alert{dimensiones o relación de aspecto}` de una imagen puede variar.

```
\begin{figure}  
\centering  
\includegraphics[width=0.6\textwidth]{figures/Tema 1/TamanolImagen.png}  
\end{figure}
```

Número de canales}

Un canal es un conjunto de píxeles de la resolución de la imagen, de tal manera que definen la \alert{intensidad} de la luz para cierto \alert{rango} de frecuencias de la imagen.

Los distintos canales de una imagen son \alert{combinados} para formar la imagen final, haciendo uso de la \alert{información} captada a distintos rangos de intensidad.

Dependiendo del tipo de codificación, el número de canales puede variar.

```
\begin{figure}  
\centering  
\includegraphics[width=\textwidth]{figures/Tema 1/CanalesImagen.png}  
\end{figure}
```

Codificación de los canales}

La codificación de los canales de una imagen define cómo estos deben combinarse para formar la imagen final. Esta información permite conocer qué información está almacenada en cada píxel

Profundidad de color}

La información de cada píxel \alert{debe estar normalizada} dentro de un rango de valores. Este rango es el que indica la \alert{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 \alert{matriz numérica} de valores, en la que cada posición corresponde a un \alert{píxel}.

A través de librerías como \alert{numpy} se podrán realizar transformaciones en la composición de las imágenes.

```
\begin{figure}
\centering
\includegraphics[width=0.4\textwidth]{figures/GoogleColab.png}
\end{figure}
```

```
\begin{itemize}
\centering
\item {\Large
\href{https://colab.research.google.com/drive/1k2vqQOiMMxOSOsZqHdhZ1Z-
eblhtvvVV?usp=drive_link}{01-Imagen_Digital.ipynb}}
\end{itemize}
```

¡GRACIAS!