

# Visión por Computadora I

Ing. Maxim Dorogov  
(mdorogov@fi.uba.ar)



Laboratorio de Sistemas Embebidos -FIUBA



# PROGRAMA SUGERIDO

- **Clase 1:** Introducción a imágenes, sistemas de visión y OpenCV
- **Clase 2:** Op. de píxel, histogramas, binarización, coord. cromáticas
- **Clase 3:** Filtros: Lineales, separables, padding, DoG, Fourier, Bordes (Canny)
- **Clase 4:** Detección de objetos. Algoritmo de Harris. Shi-Tomasi. Hough. Pirámides.
- **Clase 5:** Extracción de características. SIFT, SURF, ORB, FAST, HoG, LBP
- **Clase 6:** Segmentación: k-means, watersheed, mean-shift. Procesamiento morfológico.
- **Clase 7:** Procesamiento de video, gstreamer, ffmpeg, optical flow, sustracción de fondo tracking.
- **Clase 8:** Examen + Teórica opcional.



# RÉGIMEN DE APROBACIÓN

- Trabajos prácticos correspondientes a cada unidad temática.
- Examen teórico.

Dinámica esperada para las clases:

- 90 minutos de teoría
- 10 minutos de descanso
- 80 minutos de práctica



# HERRAMIENTAS PARA LA CURSADA

- Lenguaje de programación
  - Python > 3.6
- Bibliotecas de código
  - Numpy
  - OpenCV/OpenCV-contrib > 3.4
  - Matplotlib
  - Entorno de programación:
    - Jupyter notebook/Google Colab
  - Gestión de entornos: Conda/Miniconda

<https://www.anaconda.com/distribution/>



# BIBLIOGRAFÍA SUGERIDA

- La bibliografía es de referencia y no será obligatorio el uso de la misma.

→ ▪ Computer Vision: Algorithms and Applications | Szeliski | Springer

▪ <http://szeliski.org/Book>

▪ Computer Vision: A Modern Approach | Forsyth, Ponce | Pearson

▪ Computer Vision | Shapiro | Pearson

[ ▪ Learning OpenCV | Bradski, Kaehler | O'Reilly



# DESAFÍOS

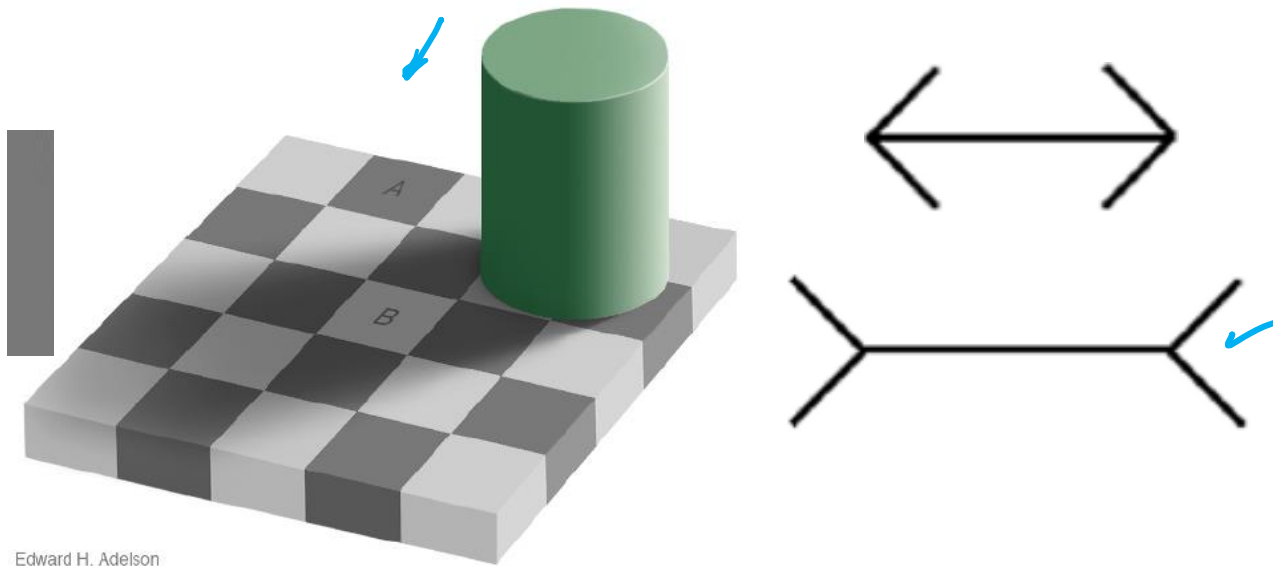
- ¿Por qué puede la visión por computadora resultar compleja?
- Algunas tareas que nuestros ojos y cerebro realizan con suma facilidad en tiempo real:
  - Reconocemos todo tipo de objetos, incluso aquellos que nunca vimos antes.
  - Podemos mirar una foto y nombrar a las personas que reconocemos. RECONOCIMIENTO FACIAL
  - Intuimos emociones a partir de las expresiones
  - Podemos seguir objetos que se mueven en entornos complejos, incluso con oclusión. TRACKING ✓
  - Podemos reproducir imágenes a partir una descripción. DALL-E
  - Visualizar y crear caras u objetos que nunca existieron. GANs
  - Predecir la trayectoria viendo como se mueve un objeto en segundos. KALMAN
  - Podemos extraer y reconocer características de objetos a diferente escala y orientación. SIFT ✓
  - Tenemos noción de profundidad incluso con un solo ojo. VISION 3D-SFM

**Visión por computadora:** Replicar el ojo humano + Replicar el cerebro humano

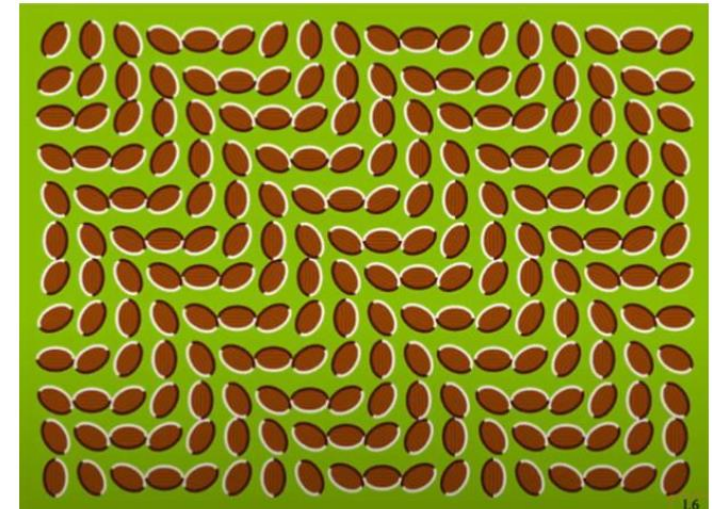
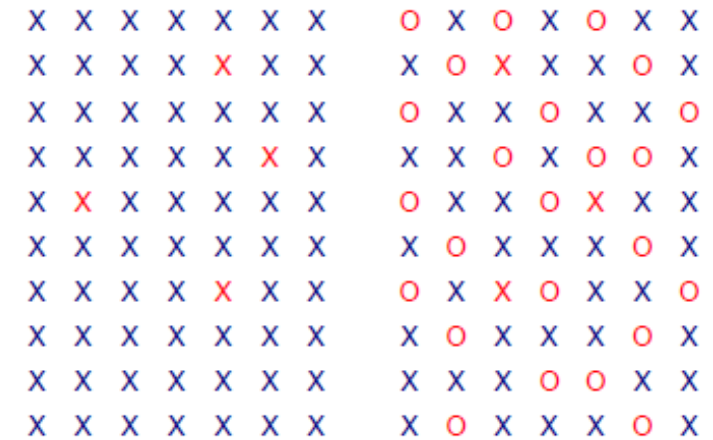


# DESAFÍOS

- A su vez nuestro sistema óptico es muy fácil de engañar...

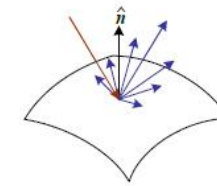
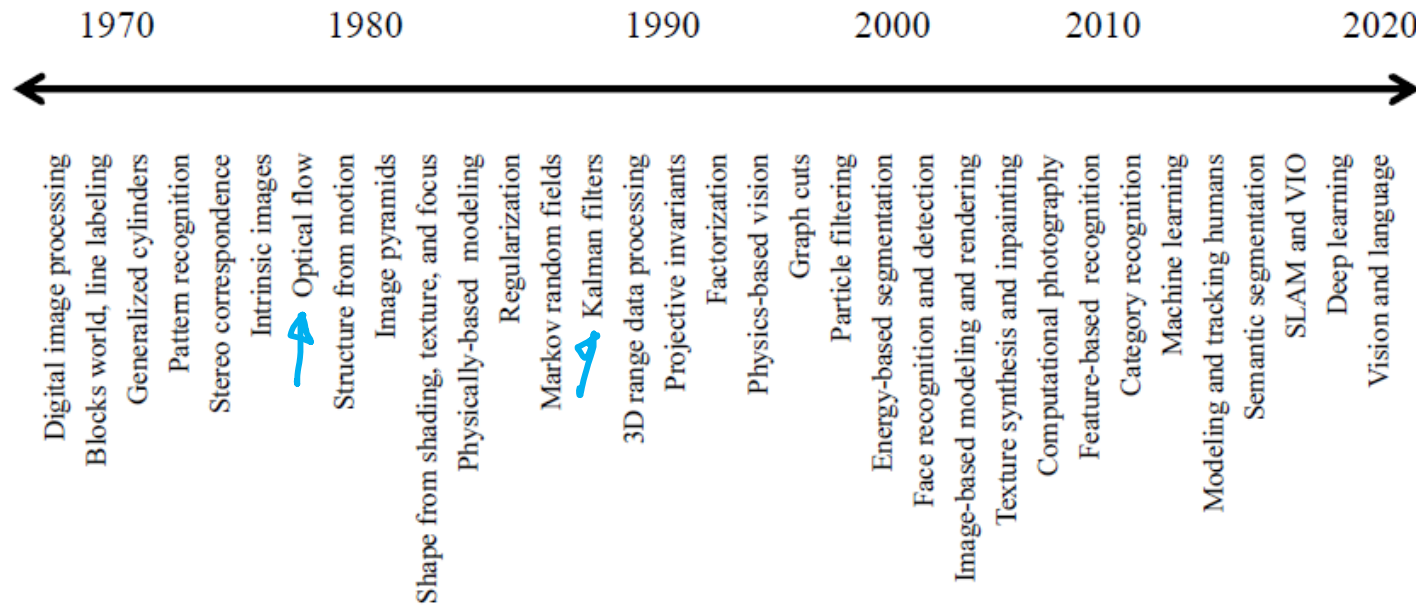


Para un sistema de visión artificial estos efectos ópticos no serían un problema en absoluto.





# HISTORIA Y APLICACIONES



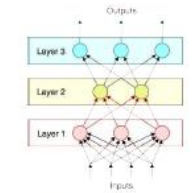
2. Image formation



3. Image processing



4. Optimization



5. Deep learning



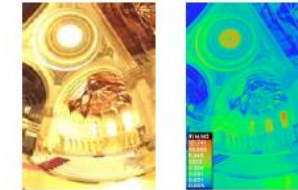
6. Recognition



7-8. Features & alignment



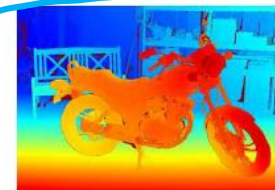
9. Motion estimation



10. Computational Photography



11. Structure from motion



12. Depth estimation



13. 3D reconstruction



14. Image-based Rendering

**Visión por computadora: Historia y campos de aplicación**





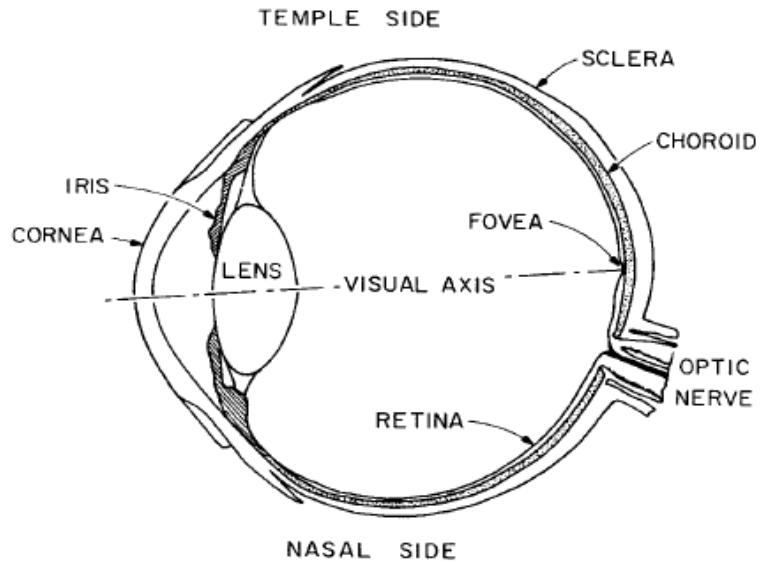
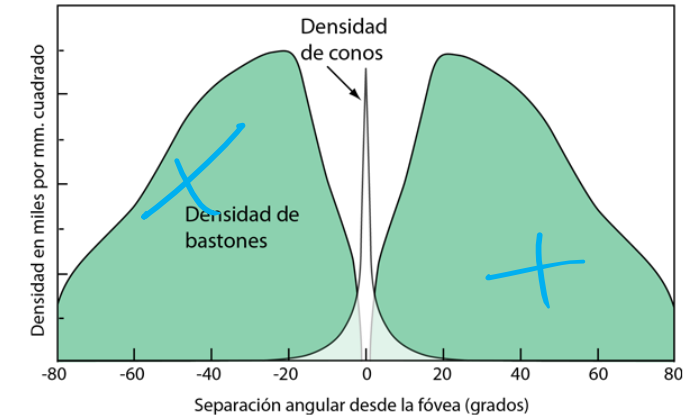
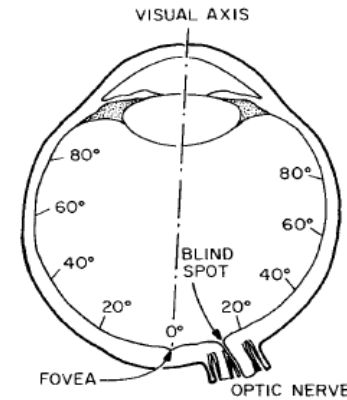


FIGURE 2.2-1. Eye cross section.



# PERCEPCIÓN DE LA LUZ POR EL OJO

- **Cornea:** Superficie transparente que envuelve al ojo
- **Sclera:** Lo que resta de la cobertura exterior
- **Coroide:** Capa que contiene capilares sanguíneos, dentro de esta capa está la retina
- **Retina:** Contiene dos tipos de células receptoras de luz, conos y bastones
- **Nervio óptico:** Conjunto de nervios que provienen de la retina
- **Cristalino:** Lente que cambia de forma mediante control muscular



# PERCEPCIÓN DE LA LUZ POR EL OJO

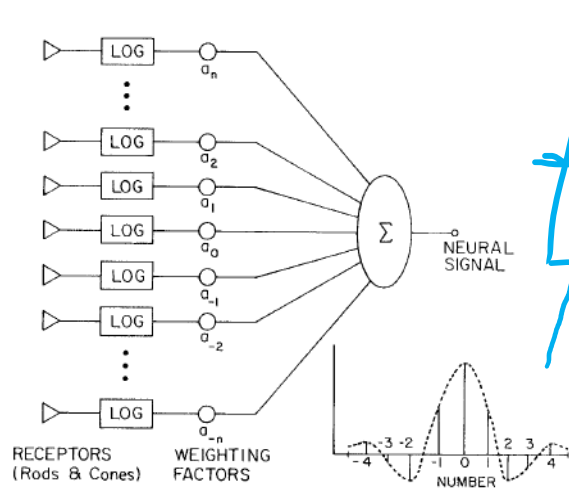
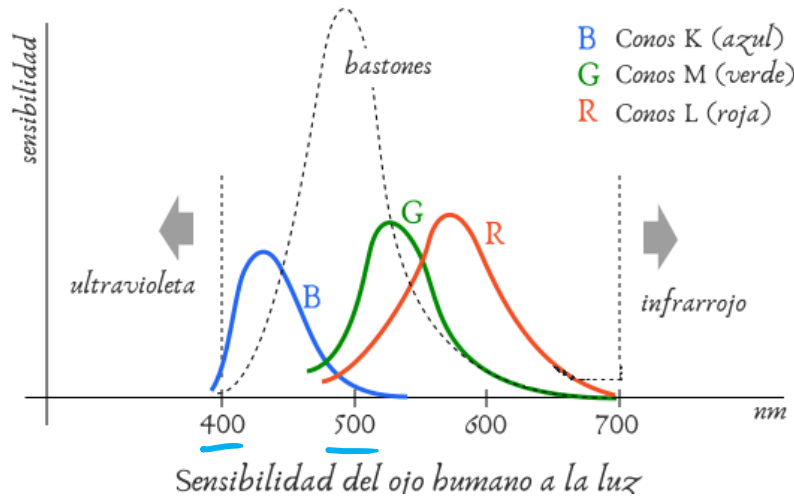
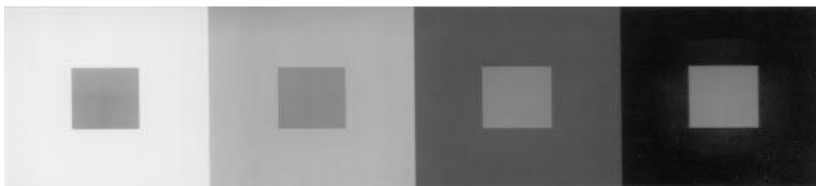


FIGURE 2.4-6. Lateral inhibition effect.



■ Bastones: Intensidad (responden poco al rojo). 120 millones

■ Conos: Color (concentrados en la mácula). 6~7 millones

■ Rojos – 64%

■ Verdes – 32 %

■ Azules – 2 %

■ Respuesta logarítmica. Fracción de Weber (0.02)

■ Daltonismo (8% hombres / 1% mujeres)



# MODELOS DE COLOR

- RGB: Commission Internationale d'Eclairage (CIE) en 1930

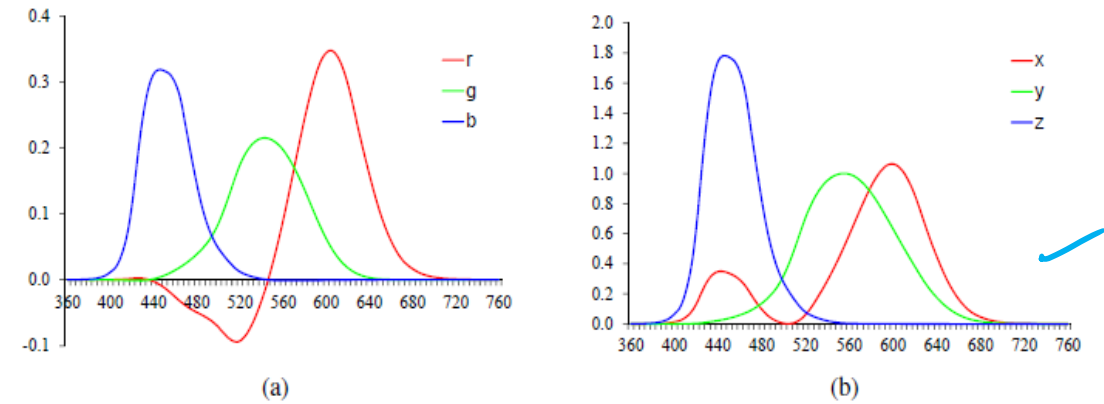
- Rojo: 700 nm
- Verde: 546,1 nm
- Azul: 435,8nm

- XYZ: Resuelve el color negativo

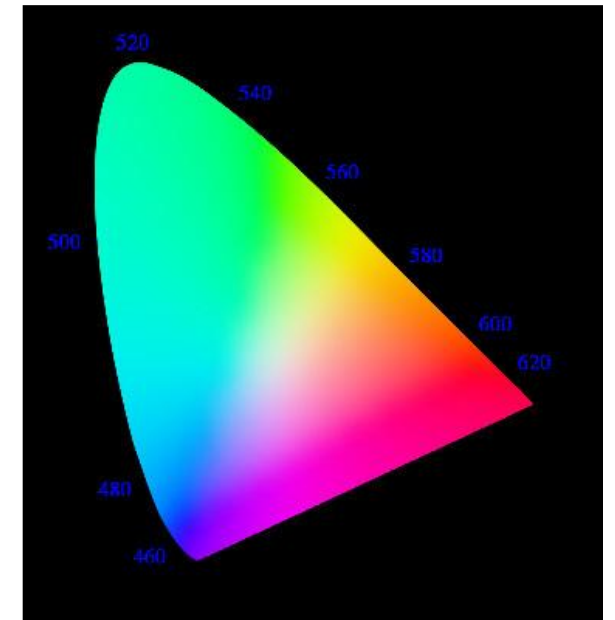
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0,17697} \begin{bmatrix} 0,49 & 0,31 & 0,20 \\ 0,17697 & 0,81240 & 0,01063 \\ 0,00 & 0,01 & 0,99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Además, permite separar crominancia de luminancia

$$x = \frac{X}{X+Y+Z} \quad y = \frac{Y}{X+Y+Z} \quad z = \frac{Z}{X+Y+Z}$$

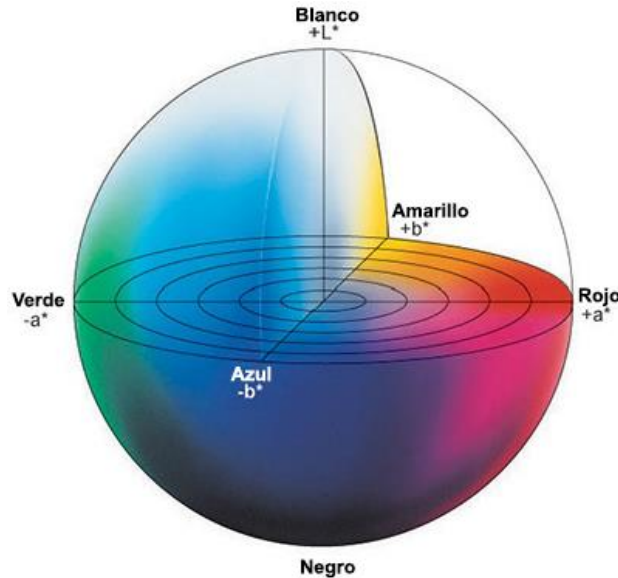


**Figure 2.28** Standard CIE color matching functions: (a)  $\bar{r}(\lambda)$ ,  $\bar{g}(\lambda)$ ,  $\bar{b}(\lambda)$  color spectra obtained from matching pure colors to the R=700.0nm, G=546.1nm, and B=435.8nm primaries; (b)  $\bar{x}(\lambda)$ ,  $\bar{y}(\lambda)$ ,  $\bar{z}(\lambda)$  color matching functions, which are linear combinations of the  $(\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda))$  spectra.



# OTROS ESPACIOS DE COLOR

## ■ CIELAB (L\*a\*b)



YUV

$$L^* = 116 f\left(\frac{Y}{Y_n}\right); \quad a^* = 500 \left[ f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right) \right]; \quad b^* = 200 \left[ f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right) \right]$$

$$f(t) = \begin{cases} t^{1/3} & \text{si } t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{2\delta}{3} & \text{otro caso} \end{cases}$$

## ■ HSV

- **Hue:** Dirección alrededor de la rueda de color, en grados  $\in [0, 360]$
- **Saturation:** Distancia escalada desde la diagonal  $\in [0, 1]$
- **Value:** Promedio o máximo valor de color  $\in [0, 1]$

Conversión RGB  $\rightarrow$  HSV

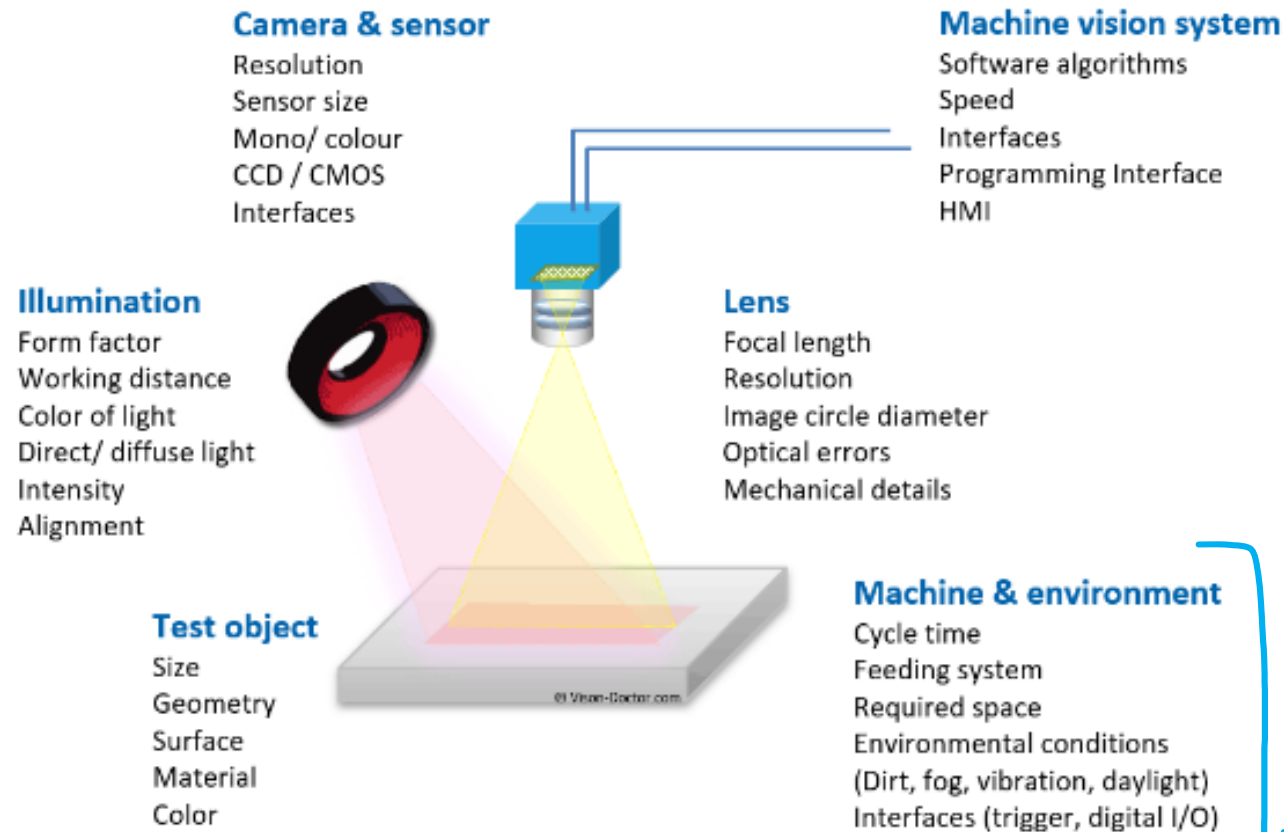
- $V = M = \max(R, G, B); m = \min(R, G, B)$
- $S = (M - m)/M$  ( $S = 0$ , si  $V = 0$ )
- $H = 60 \times \begin{cases} 0 & , \text{si } (M - m) = 0 \\ 0 + (G - B)/(M - m), & \text{si } \max = R \\ 2 + (B - R)/(M - m), & \text{si } \max = G \\ 4 + (R - G)/(M - m), & \text{si } \max = B \end{cases}$

$$H = H + 360, \text{ si } H < 0$$



# SISTEMAS DE VISIÓN

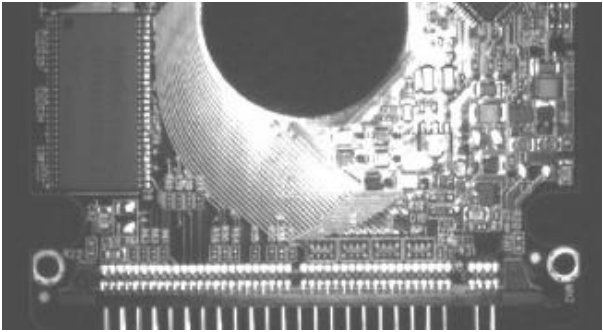
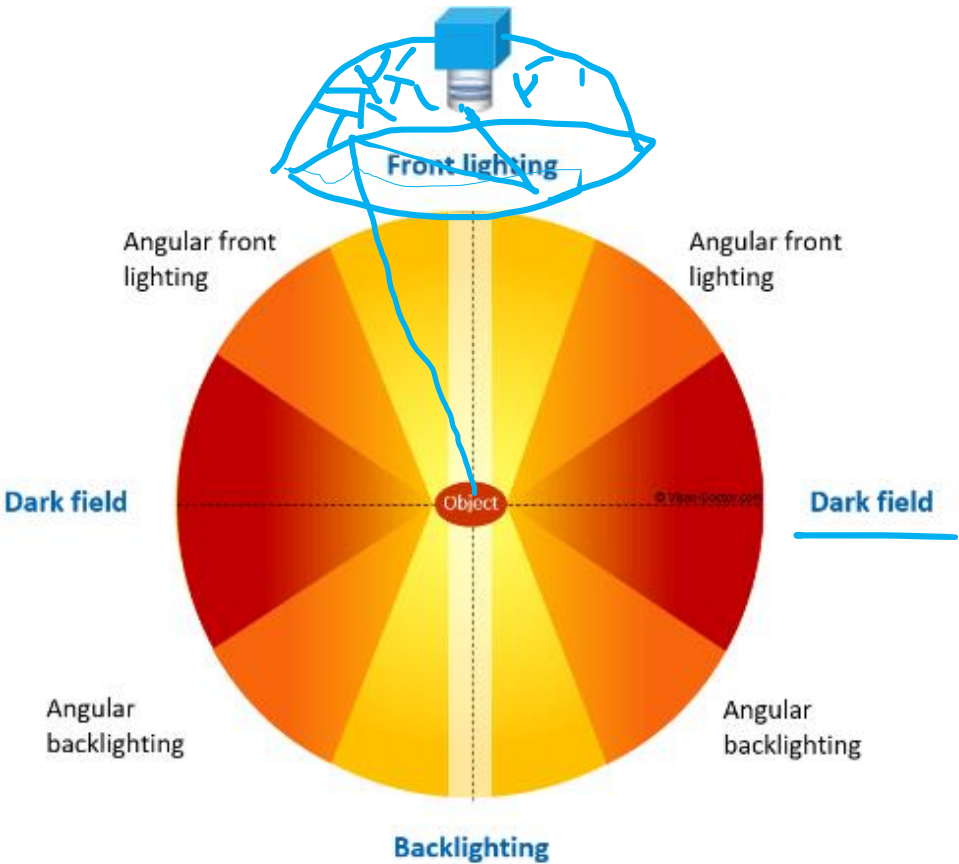
BASLER



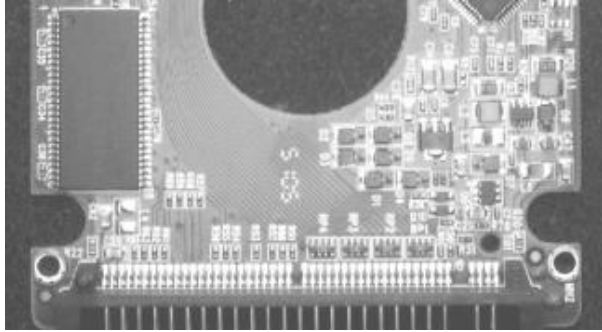
- Cámara: ✓
  - Sensores CMOS o CCD.
  - Tipo de shutter
  - Tipo de interfaz (USB, Ethernet, etc...)
  - LUTs ✓
  - Depth Camera ✓
  - Smart ✓
  - Etc...
- Lente:
  - Telecéntrica —
  - Entocéntrica —
- Iluminación:
  - Campo oscuro (Dark field)
  - Luz directa
  - Luz difusa (Domo)
  - Backlight
  - Color
- Procesamiento ✓
  - Cloud
  - Edge



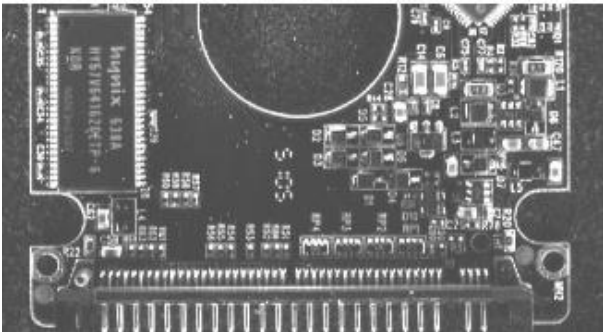
# ILUMINACIÓN



Luz directa



Luz con difusor



Dark Field



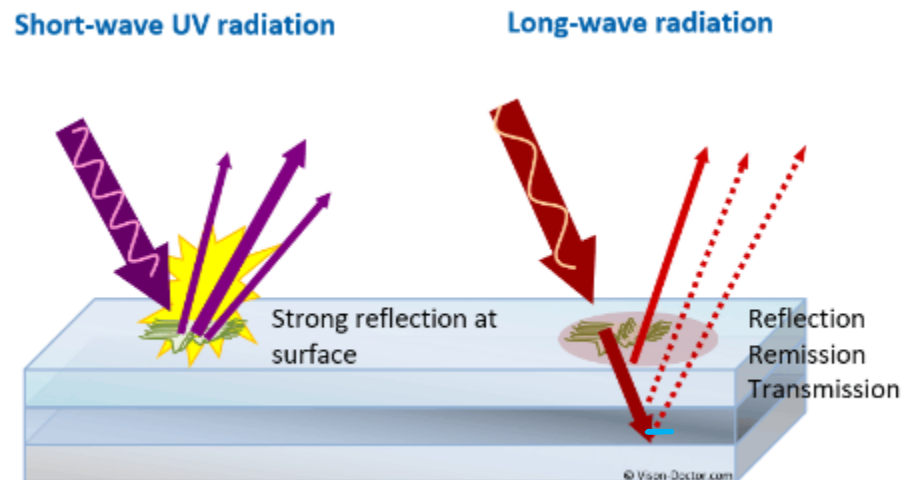
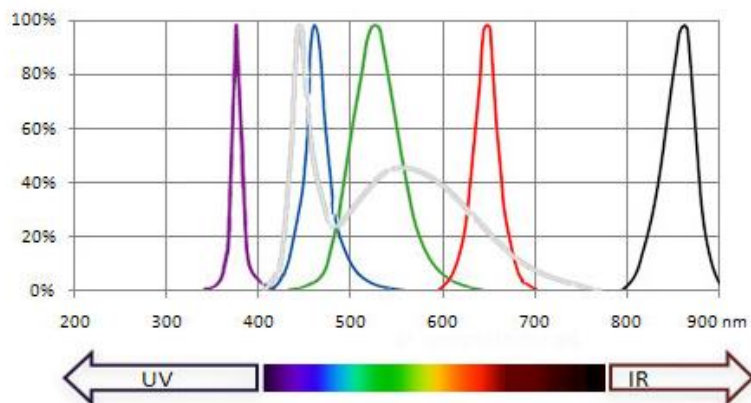
Backlight



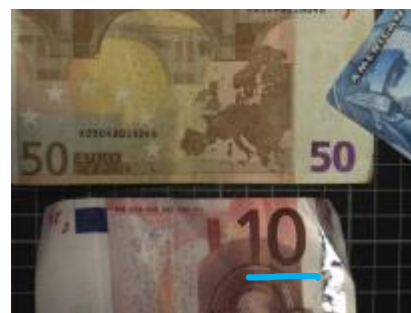


# ILUMINACIÓN

Iluminación por encima (y debajo) del espectro visible:



Luz ultravioleta: Permite ver capas internas del material a analizar

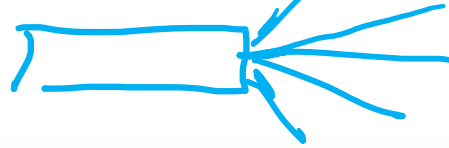


Algo similar sucede cuando se utiliza luz infrarroja



# LENTE

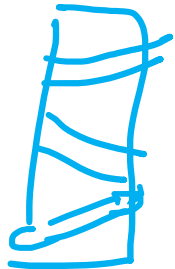
Principalmente se dividen en etnocéntrica y telecéntrica, las primeras producen un error de perspectiva y son las lentes mas comunes usadas en fotografía, cine, y aplicaciones de consumo masivo.



Lens class	Aperture angle
Tele lens	$< 20^\circ$
Long focal length lens	$20-40^\circ$
Normal lens	$40-55^\circ$
Wide angle lens	$> 55^\circ$
Super wide angle lens	$\sim 110^\circ$
Fisheye lens	$\sim 180^\circ$

La lente telecéntrica corrige los errores de perspectiva ya que los rayos de luz inciden paralelos al eje óptico de la cámara. Son usadas principalmente en metrología y no tienen aplicaciones por fuera del ambiente industrial.

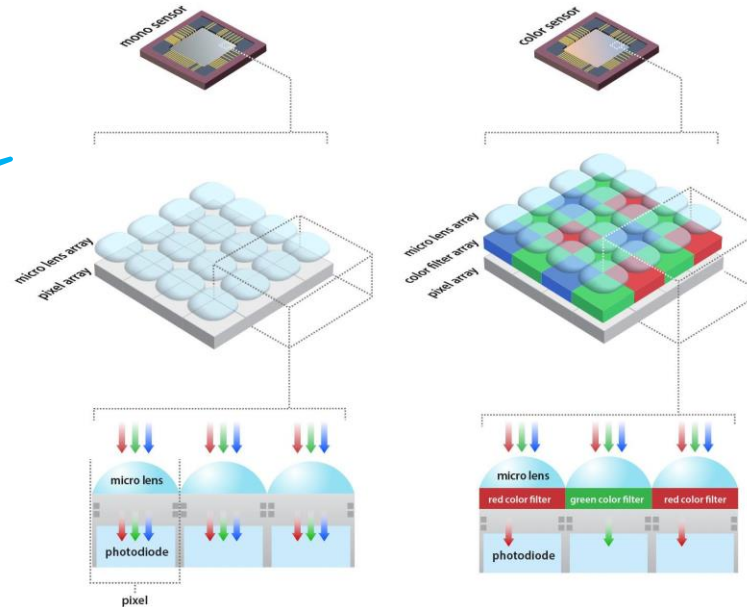
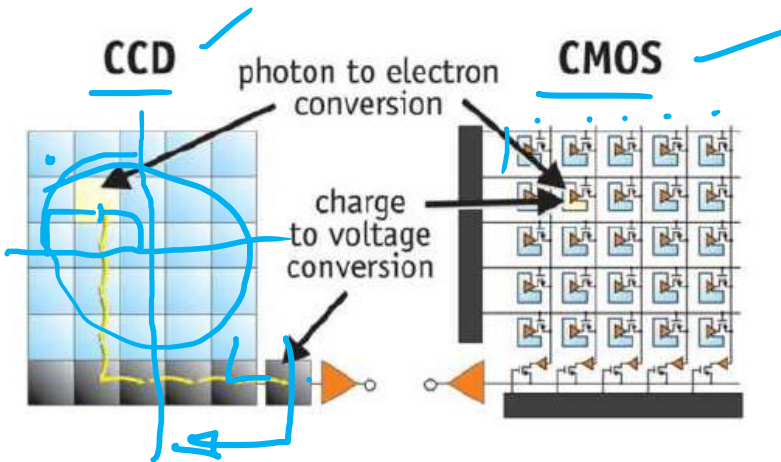
Tipos de lente etnocéntrica



Lente etnocentrica, (izquierda) y telecéntrica (derecha)



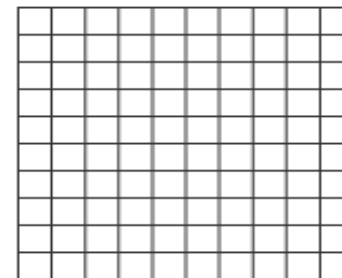
# TIPOS DE SENSORES



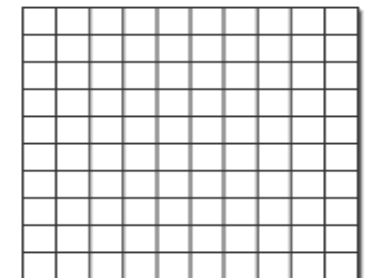
- **CCD** (charge-couple device):  
Willard Boyle y George E. Smith  
– Laboratorios Bell – 1969
  - A/D central 20 a 75MHz
  - Blooming/ Smearing
  - Mayor sensibilidad ✓
- **CMOS** (complementary metal oxide semiconductor)
  - Conversión en el fotosito
  - Rolling shutter
  - Microlentes



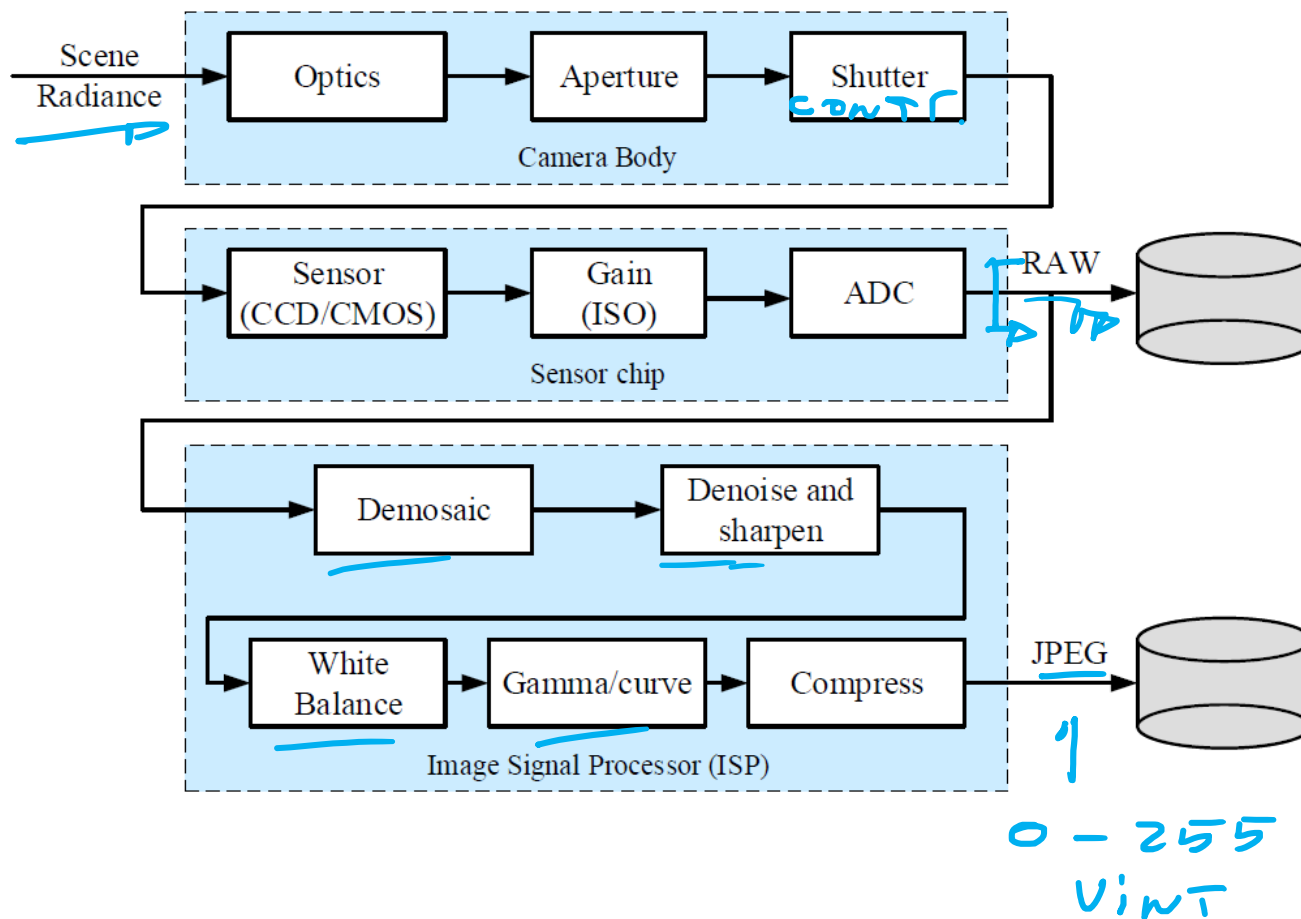
Rolling Shutter



Total Shutter ✓



# CÁMARA DIGITAL



Las cámaras profesionales y de aplicaciones específicas permiten, además de una imagen, obtener los datos en formato RAW. (Raster en aplicaciones espaciales)

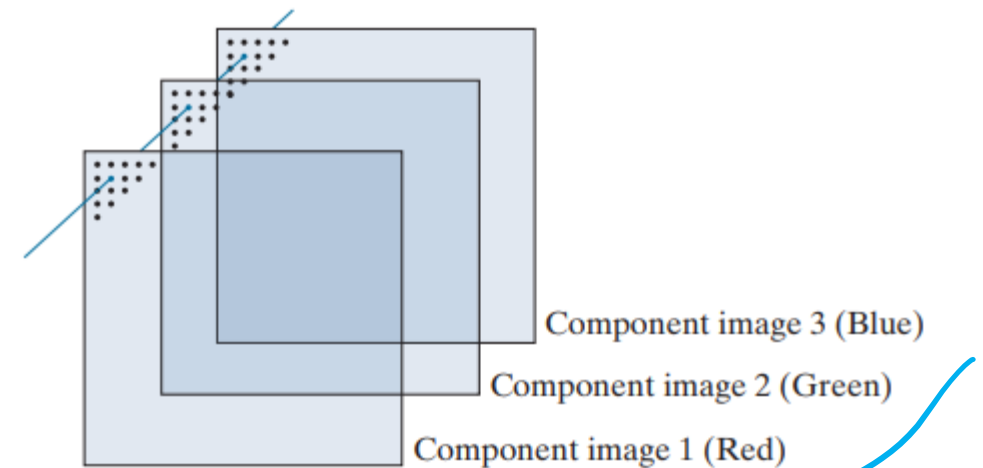
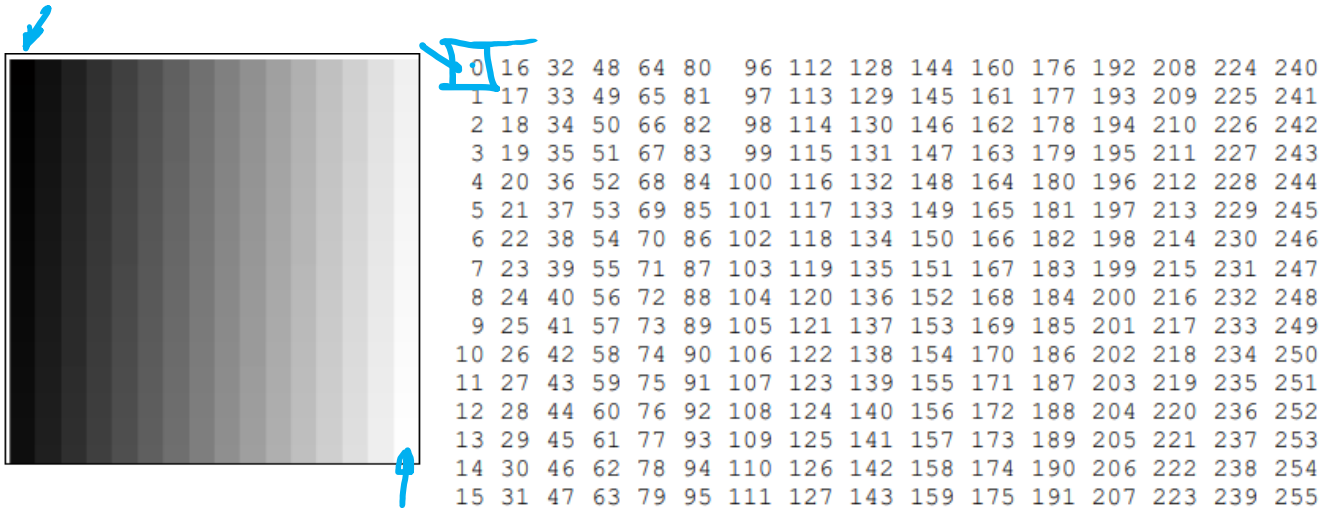
- No es un “formato” estandarizado
- Algunos formatos de datos RAW: TIFF, CR2, 3FR, NRW.
- Ocupan 3 veces (o mas!) de espacio que un archivo de imagen.

Muchas veces en la cámara se efectúa un post procesamiento digital para compensar las falencias del sistema óptico (lente-sensor), algo típico en teléfonos celulares o cámaras hogareñas.



# REPRESENTACIÓN DE IMÁGENES

- $N \times M \times 1$  si la imagen esta en escala de grises ✓
- $N \times M \times 3$  para imágenes a color
- Rango: 0 (pixel apagado) – 255 (max. Intensidad) con uint8 para indicar la intensidad de cada pixel ✓
- También existen imágenes binarias, se utilizan como mascaras 0/255
- En OpenCV y Numpy se indexa por [N-fil, N-col, N-canal]



Se toma la esquina superior izquierda como origen (y=0, x=0) de coordenadas

