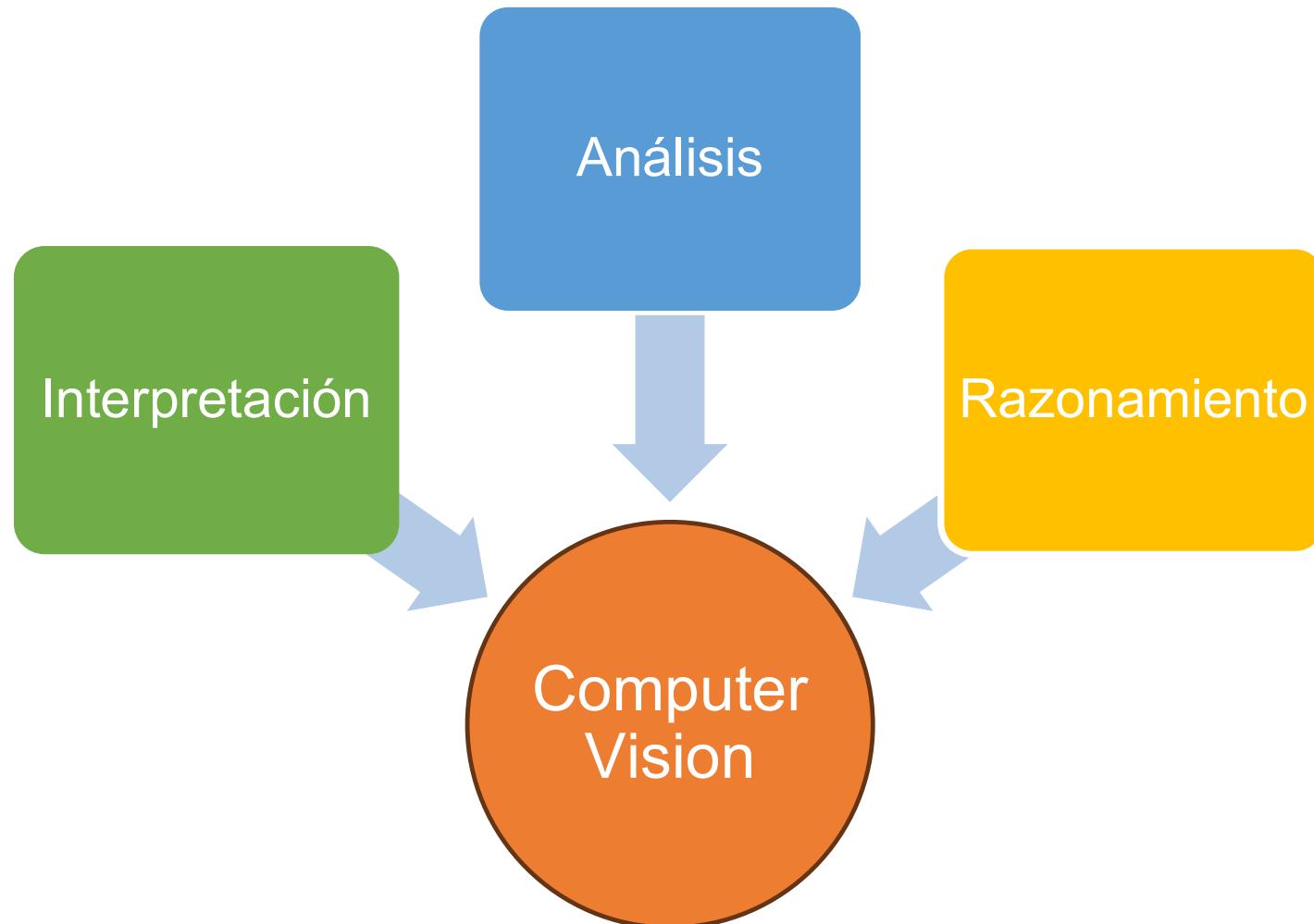
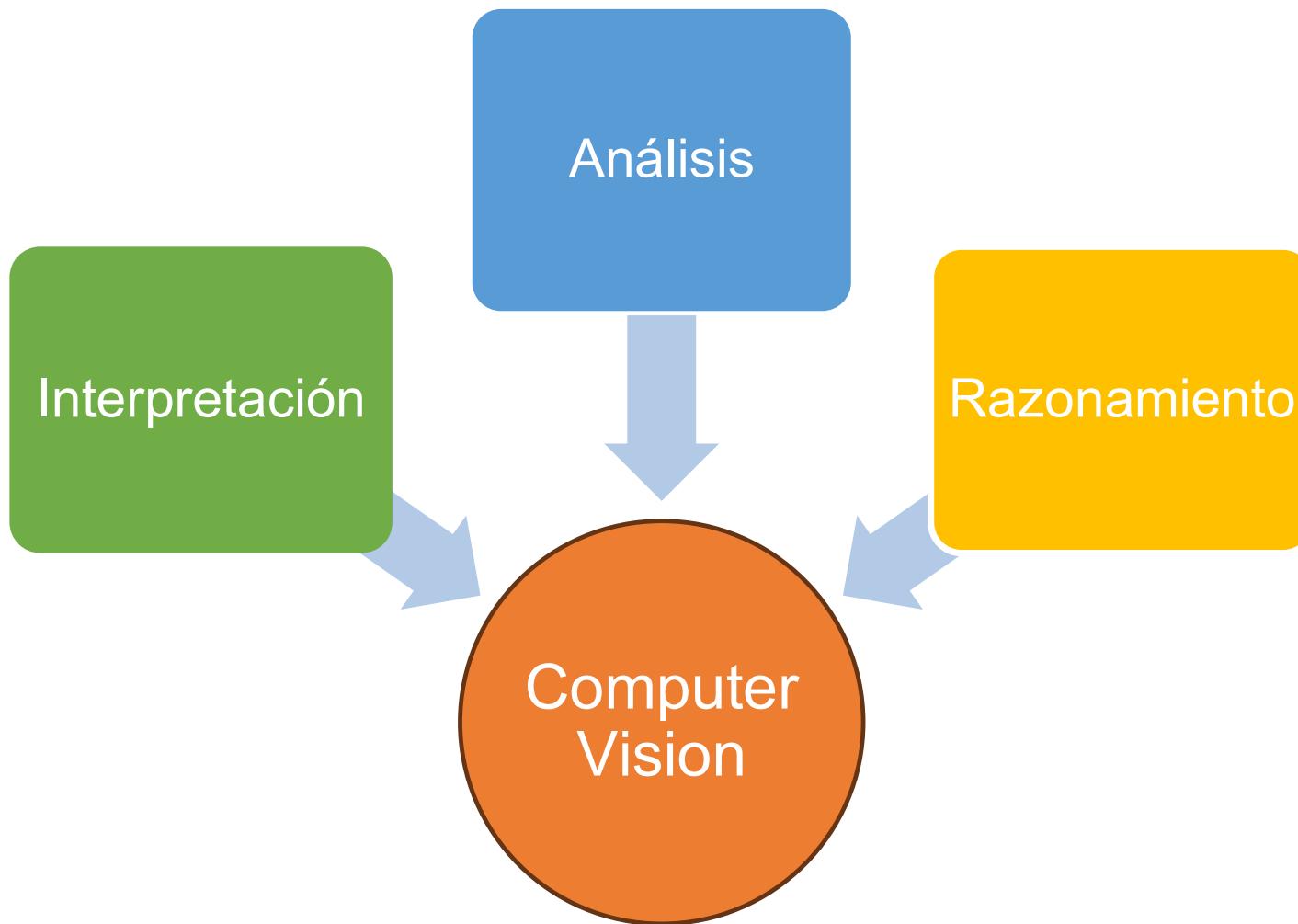


# Visión Computacional

Ivan Sipiran

# ¿Qué es visión computacional?



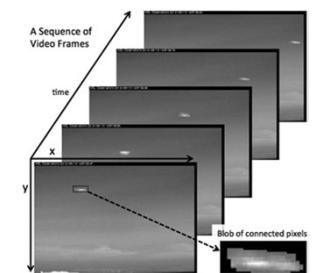


## Inputs Visuales

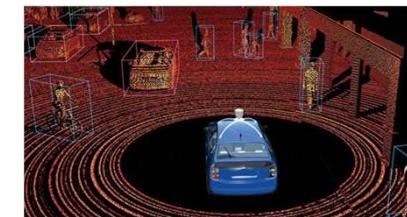
Imágenes



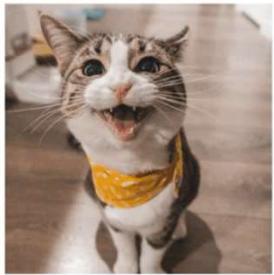
Videos



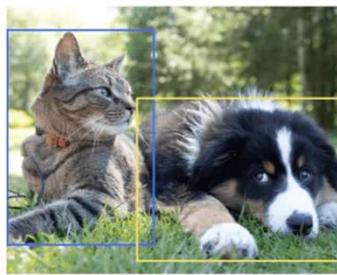
3D



## Interpretación



Clasificación



Detección



Segmentación

## Razonamiento



Cuántos tigres hay en la foto?

En la imagen hay **un tigre**. Es un tigre blanco y está ubicado en la parte derecha, cerca del centro. 🐯

Cuántas aves hay en la foto?

## Análisis



Reconstrucción 3D



Pose estimation

En la imagen hay **13 aves**. Aquí está el desglose:

1. Flamenco (izquierda)
2. Ave blanca y negra (arriba, cerca del centro)
3. Ave blanca volando (centro derecha)
4. Ave oscura (cerca del conejo blanco)
5. Dos patos anaranjados (abajo del ave blanca volando)
6. Ave marrón con cuello blanco (debajo del lobo)
7. Ganso (cerca de los búhos)
8. Dos búhos blancos (parte derecha inferior)
9. Pavo real (parte inferior izquierda)
10. Ave con plumas rojas y negras (al lado del pavo real)
11. Urraca blanca y negra (cerca del centro inferior)

✓ Total: **13 aves**.

# El curso

- Unidad I: Fundamentos
  - Imágenes digitales, formación, representación y transformaciones
  - Procesamiento de imágenes: filtros y convolución
  - Operaciones de caracterización
  - Cámaras, calibración y modelos de proyección
- Unidad II: Reconstrucción y Geometría
  - Estructura desde movimiento
  - Epipolaridad
  - Visión estéreo
  - Slam

# El curso

- Unidad III: Deep Learning para Visión
  - Redes convolucionales
  - Arquitecturas: VGG, ResNet, EffNet
  - Transfer Learning
- Unidad IV: Tópicos avanzados
  - Detección y segmentación
  - Transformers visuales (ViT, DETR)
  - Modelos generativos: GAN, VAE, Diffusion
  - Modelos multimodales
  - Neural Radiance Fields (NeRF)

# El curso

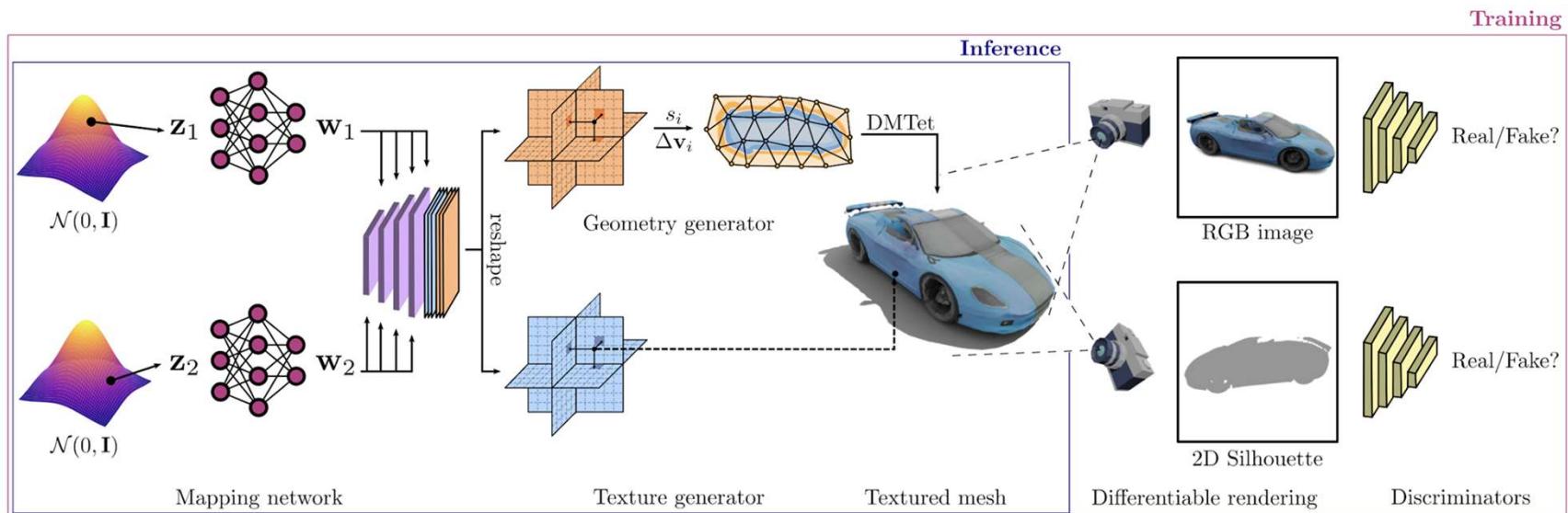
- Cinco tareas individuales (TA)
- Cinco cuestionarios cortos (CC)
- Proyecto grupal (PR)
- Nota final:  
$$0.4 \times TA + 0.4 \times PR + 0.2 \times CC$$

# El curso - Libros

- Computer Vision: Algorithms and Applications - Szeliski
  - <https://szeliski.org/Book/>
- Deep Learning – Goodfellow, Bengio, Courville
  - <https://www.deeplearningbook.org/>
- Dive into Deep Learning – Zhang, Lipton, Li, Smola
  - <https://d2l.ai/>

# Algunos proyectos

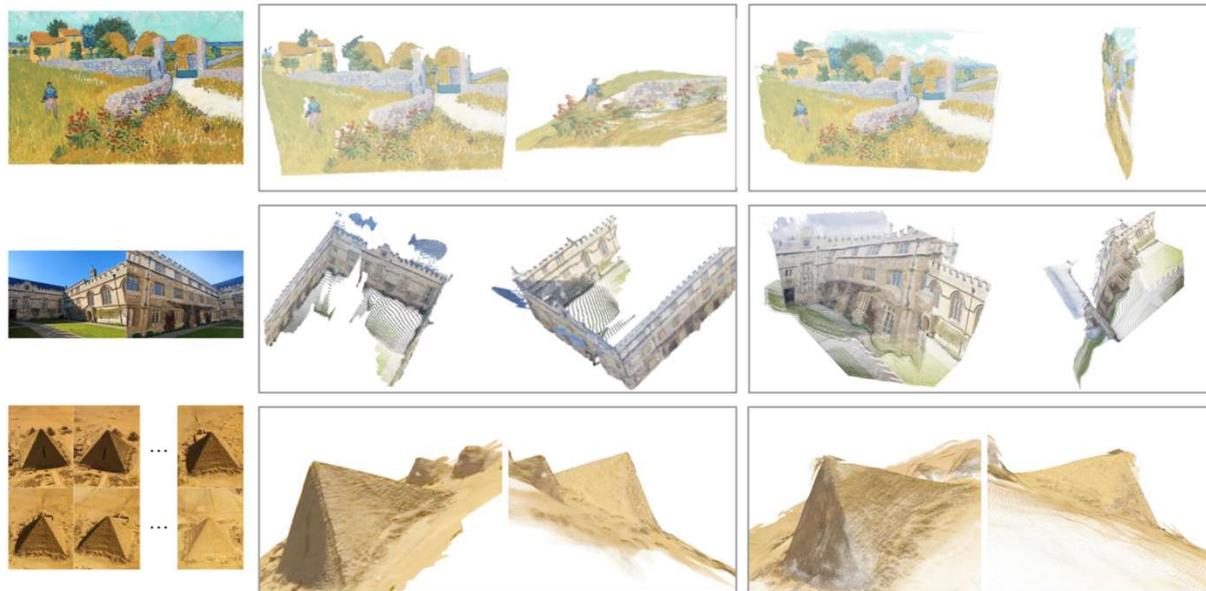
- Modelos generativos 3D



Nicolás Caytiro – doctorado DCC  
Cristóbal Alvarez – memoria DCC

# Algunos proyectos

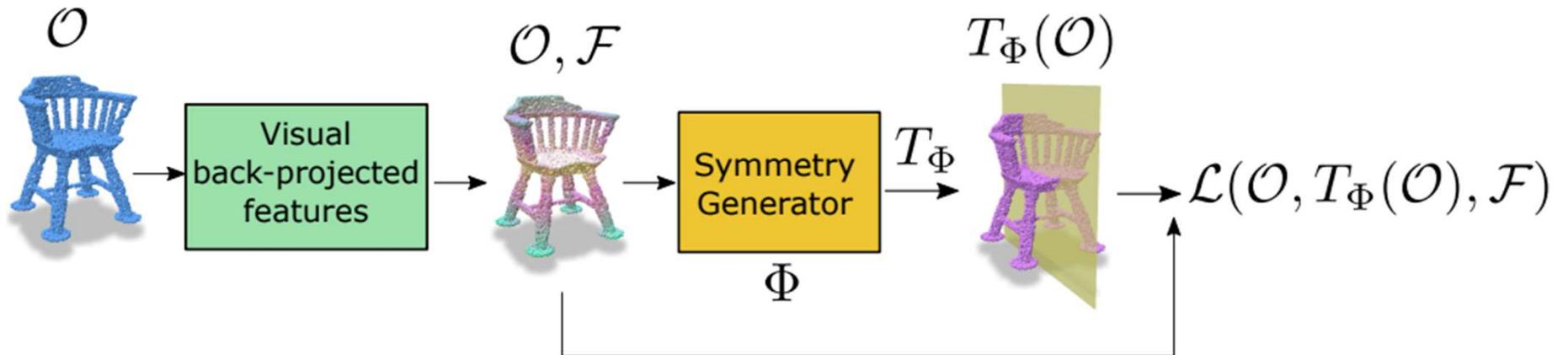
- Reconstrucción 3D



Cristian Llull – doctorado DCC

# Algunos proyectos

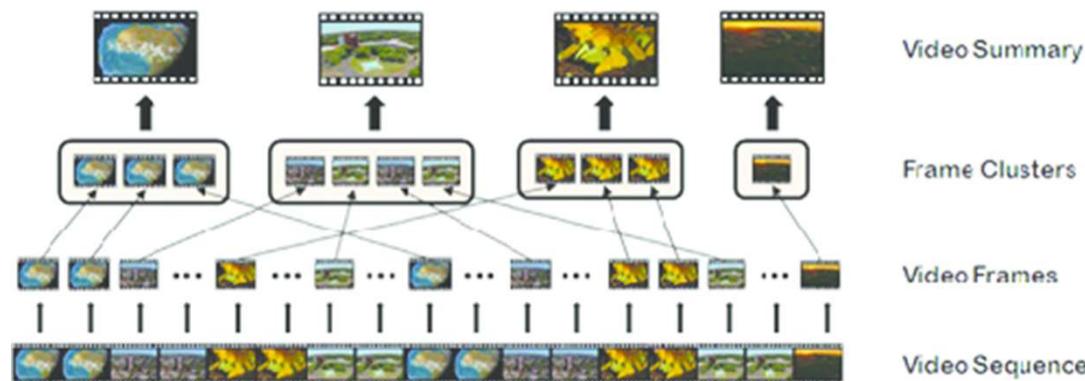
- Análisis 3D



Isaac Aguirre – magíster DCC  
Vicente Muñoz – magíster DCC

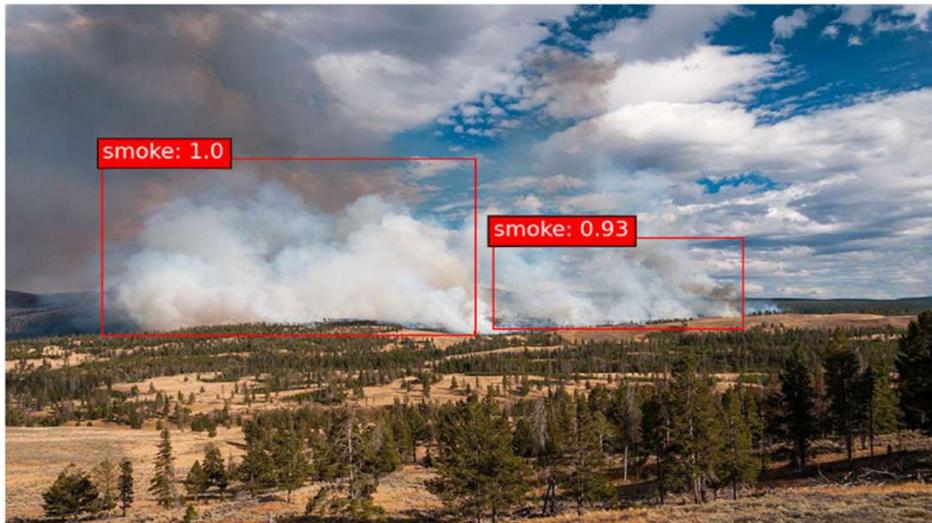
# Algunos proyectos

- Video summarization



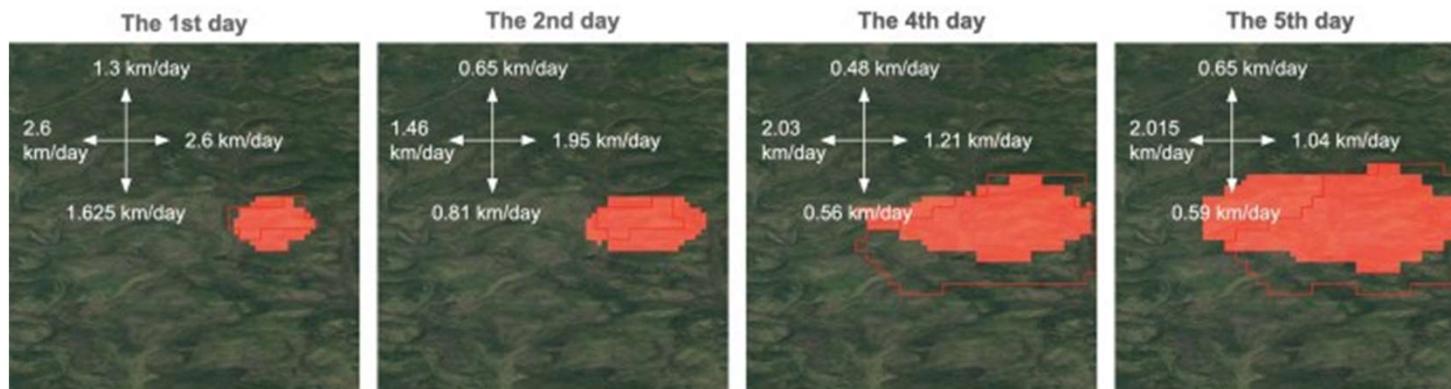
# Algunos proyectos

- Smoke column detection



# Algunos proyectos

- Wildfire spread prediction



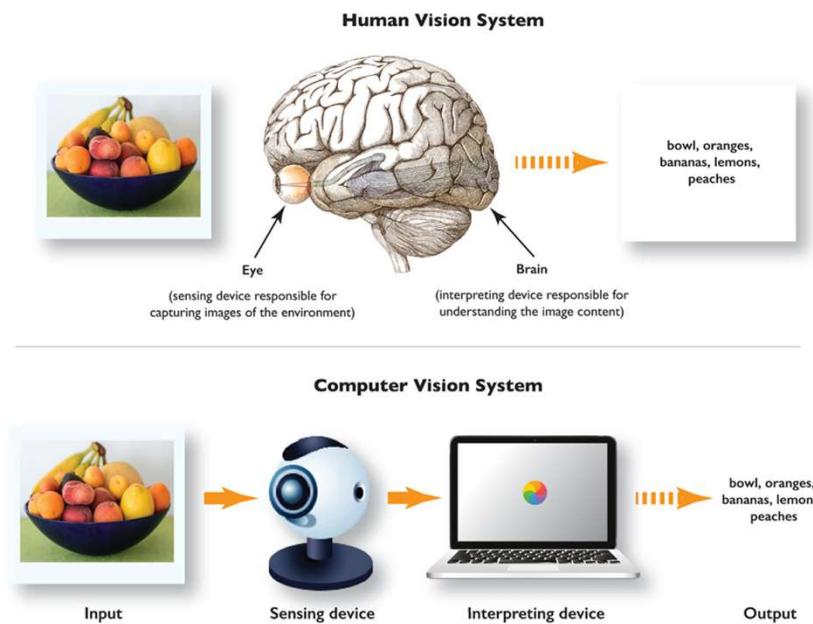
<https://shapevision.dcc.uchile.cl/>



# Fundamentos de Visión Computacional

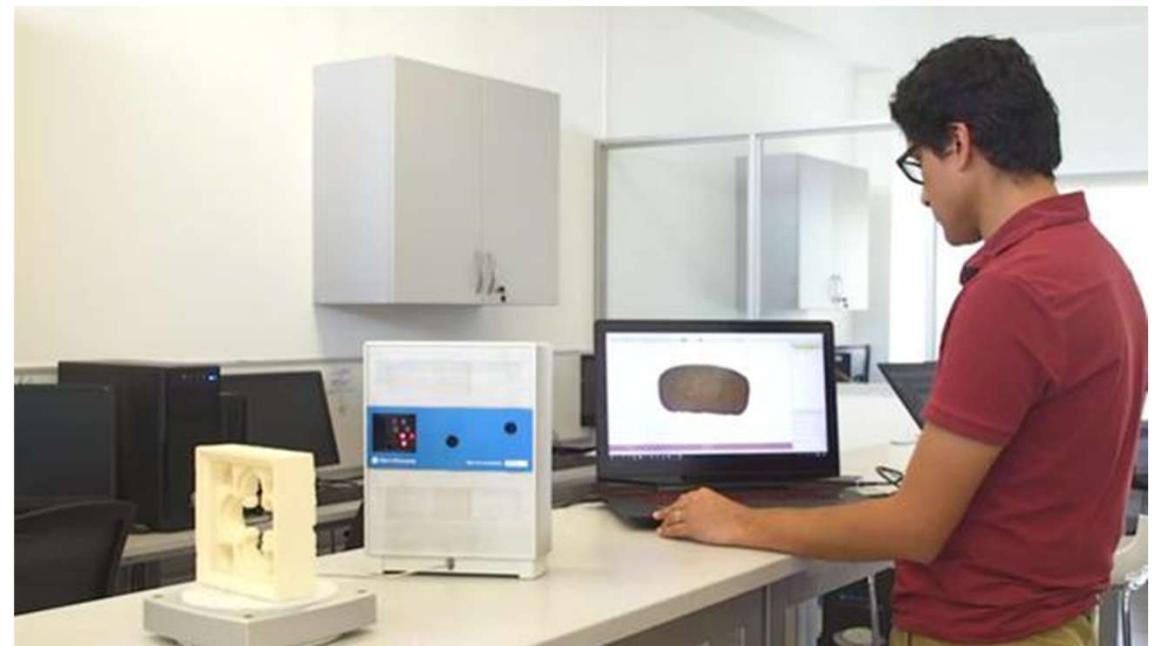
# Visión Computacional

- La Visión Computacional es una subdisciplina de la inteligencia artificial que estudia cómo hacer que las máquinas interpreten, analicen y razonen a partir de imágenes y videos, de manera similar —o incluso superior— a los humanos.



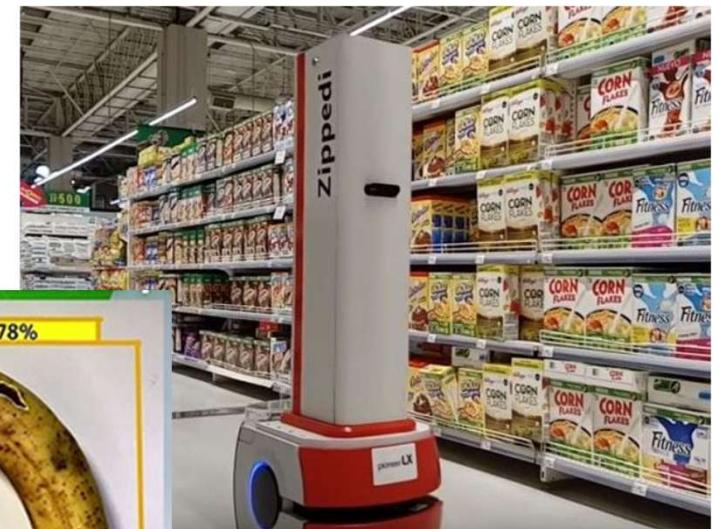
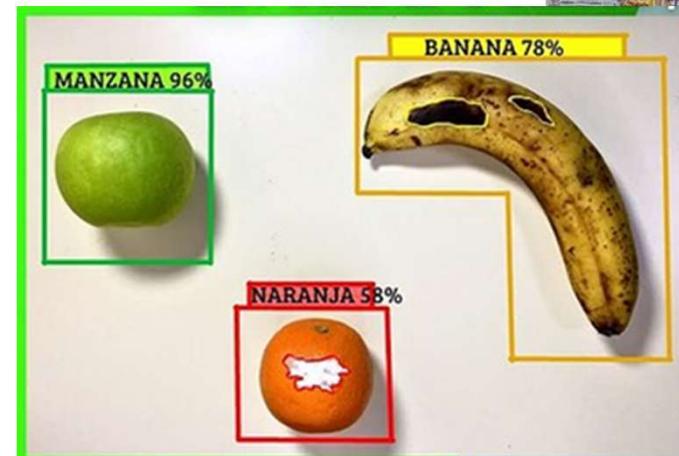
# Sistema de Visión Computacional

- Etapas:
  - Adquisición
  - Procesamiento
  - Análisis y Comprensión
  - Toma de decisiones

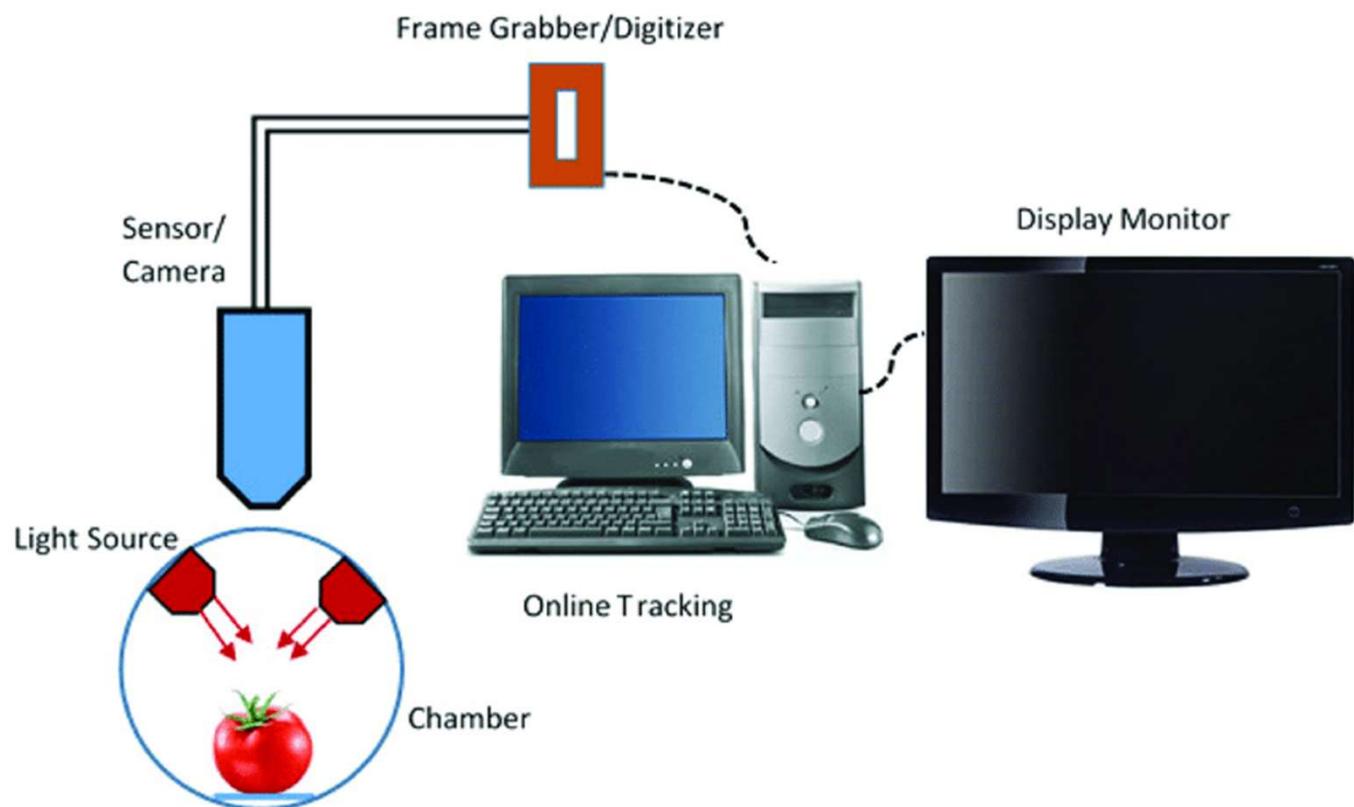


# Sistema de Visión Computacional

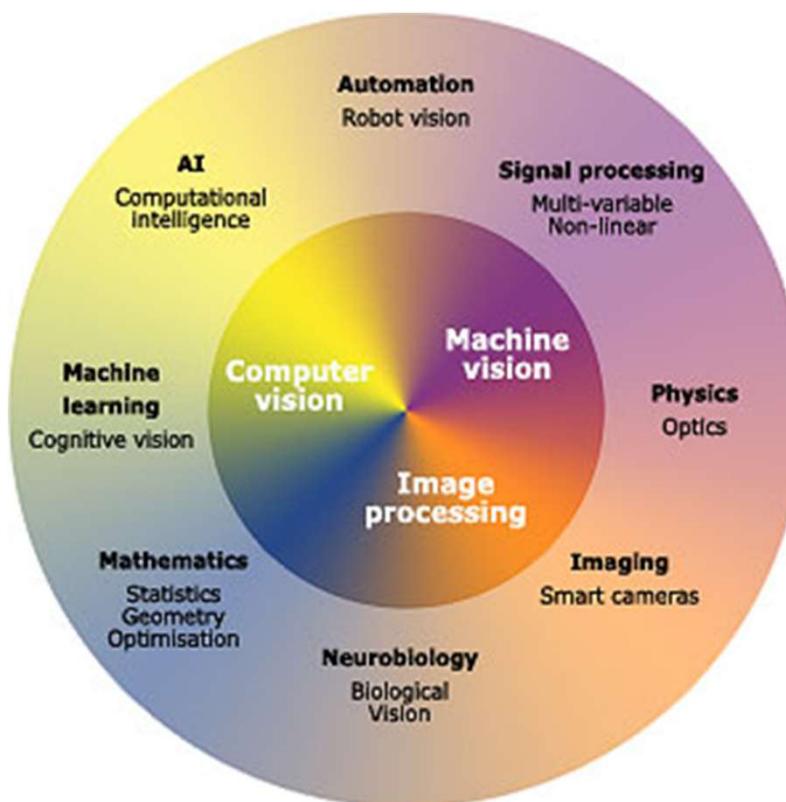
- Etapas:
  - Adquisición
  - Procesamiento
  - Análisis y Comprensión
  - Toma de decisiones



# Componentes de Sistema Visión Computacional



# Background computacional



# Aplicaciones de la Visión Computacional

<https://www.youtube.com/watch?v=2KvEM9JzLMs>

# Introducción

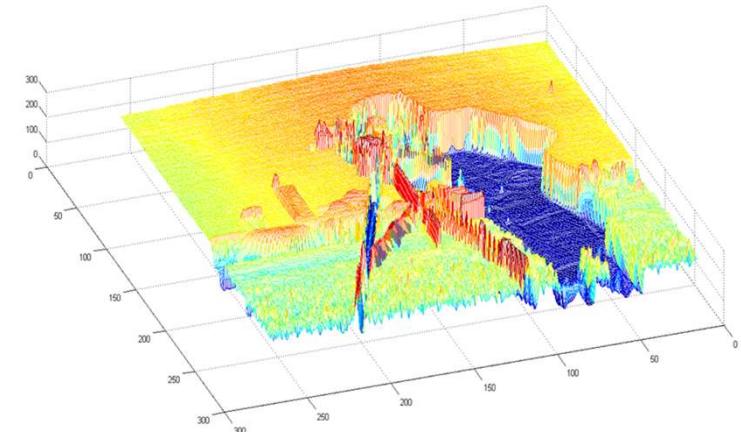


[Albrecht Dürer, 1525]

“Una imagen vale más que mil palabras”

# Introducción

- Imagen es una función  $f(x,y)$ 
  - Coordenadas espaciales:  $x,y$
  - Intensidad:  $f(x,y)$
- En DIP las coordenadas espaciales y intensidad son valores discretos
- El elemento más pequeño de representación es un punto y su intensidad : píxel



# Introducción

- Representación de la imagen  $f$  como un arreglo 2-D o matriz :



MxN

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,N-1) \\ f(1,0) & f(1,1) & \dots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(M-1,0) & f(M-1,1) & \dots & f(M-1,N-1) \end{bmatrix}$$

# Adquisición

- Proceso por el cual una escena es capturada por un sensor y convertida a una entidad manejable

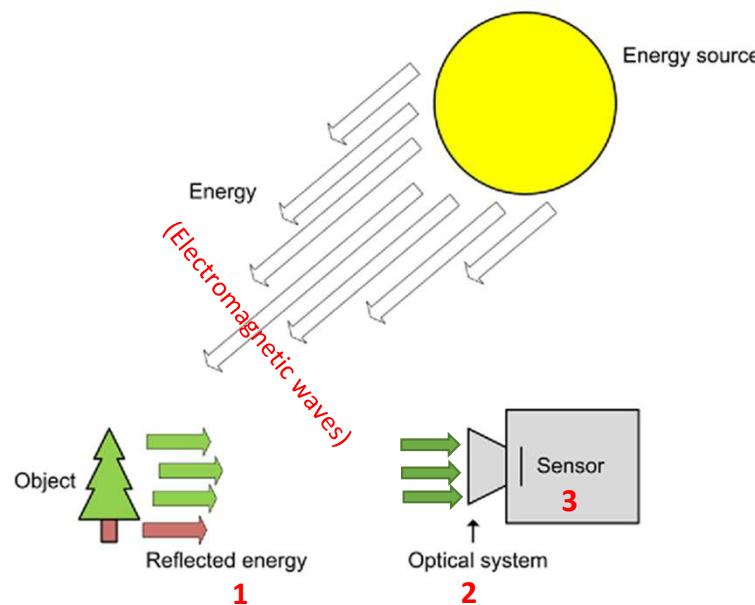
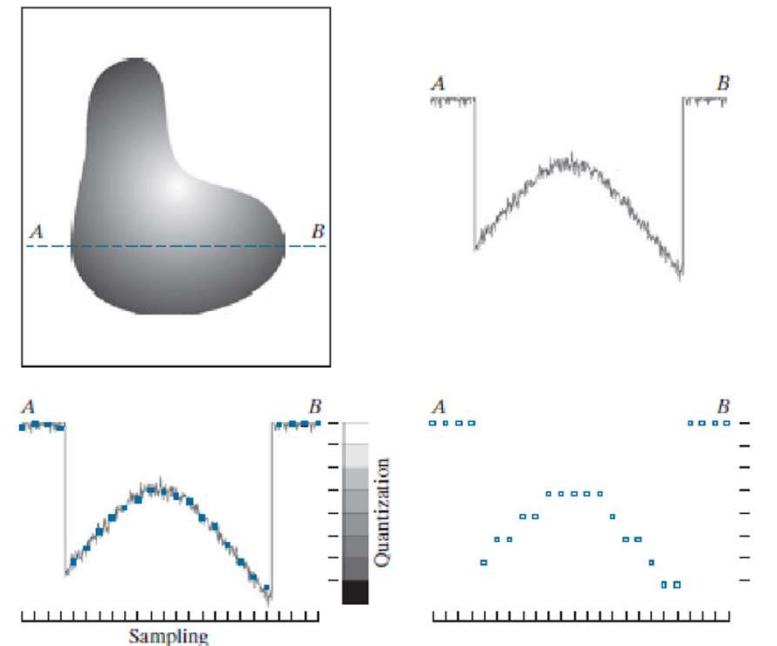


Figura. Muestra un ejemplo de la captación de una imagen usando una cámara convencional

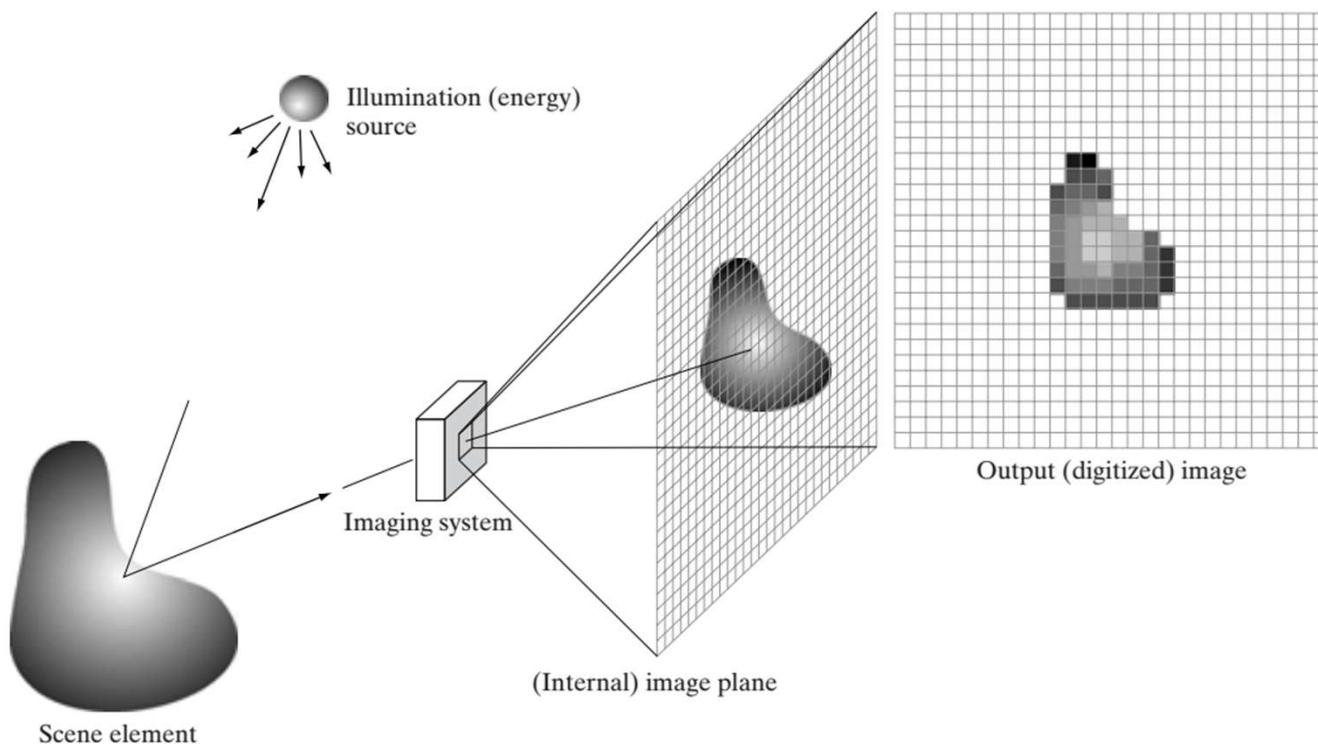
# Adquisición

- La salida de los sensores produce una imagen continua que necesita Digitalización.
- Digitalización es un proceso:
  - Muestreo
  - Cuantización



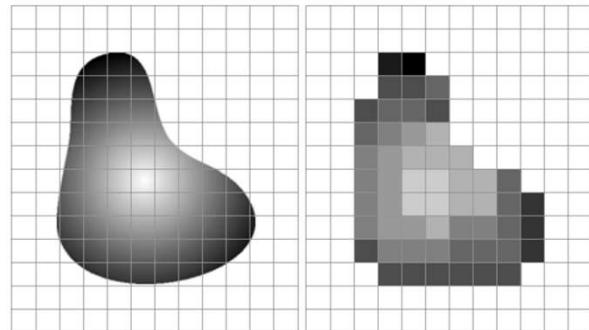
a) Imagen continua, b) Escaneo de la Intensidad de la línea AB, c) Muestreo y Cuantización d) Escaneo digital línea AB

# Adquisición



# Adquisición

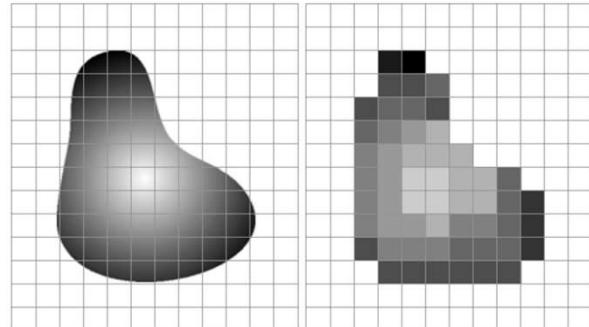
- El resultado es un arreglo (array) de valores



¿Cómo podemos representar este arreglo de valores ?

# Adquisición

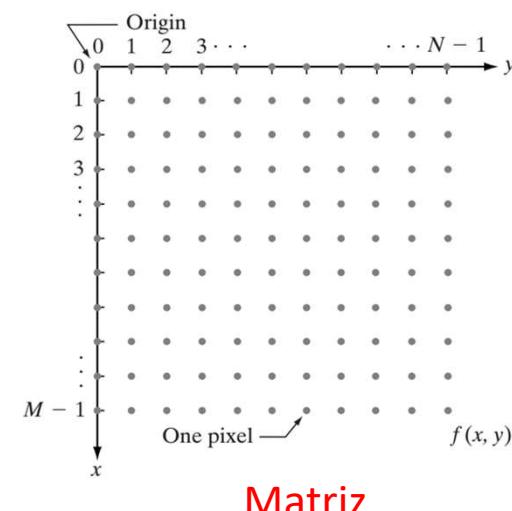
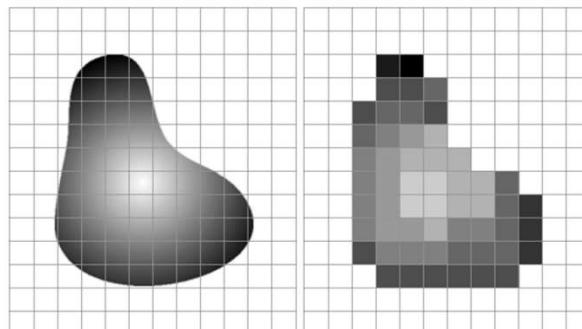
- El resultado es un arreglo (array) de valores



¿Cómo podemos representar este arreglo de valores ?

# Adquisición

- El resultado es un arreglo (array) de valores que pueden ser representados en una estrutura como una Matriz



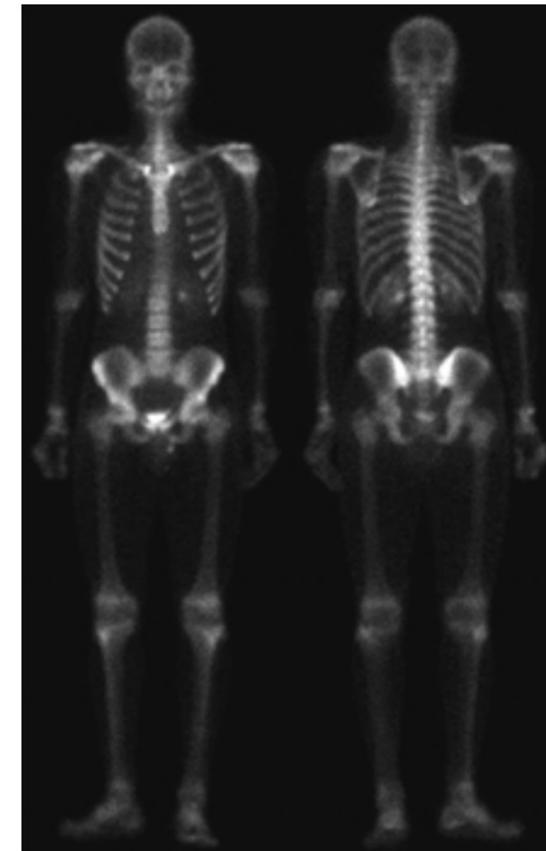
Matriz

# Adquisición

<https://youtu.be/t48UUwDvBv8>

# Aplicaciones

- Rayos Gamma
  - Detectar malformaciones óseas y tumores



# Aplicaciones

- Rayos X
  - Detectar objetos peligrosos en imágenes de rayos x de múltiples vistas

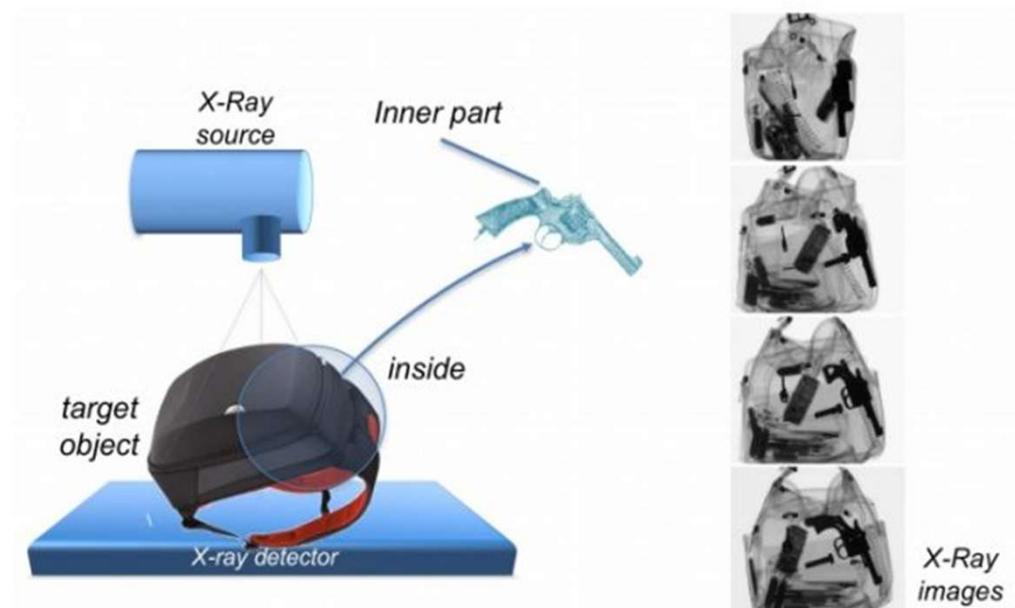
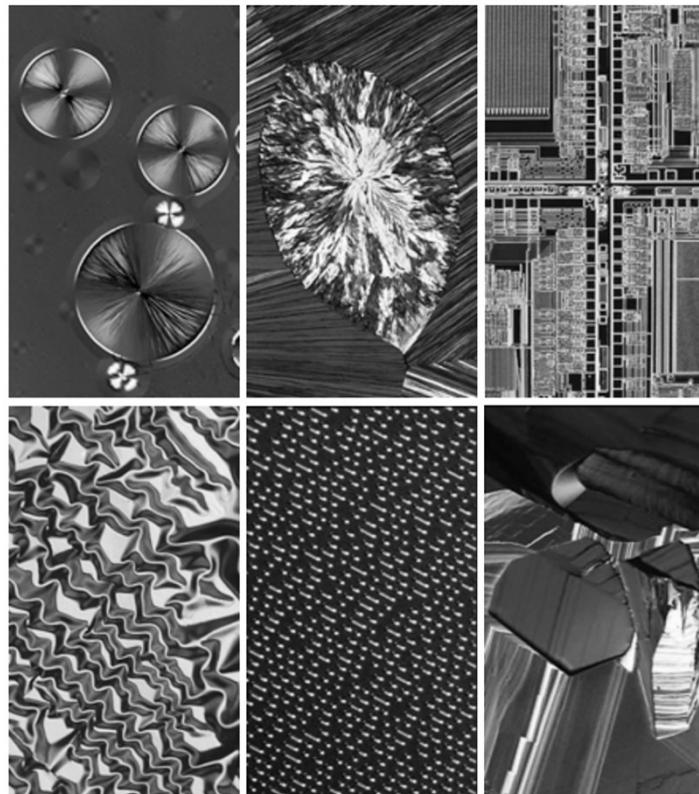


Fig. 1: X-ray images of a backpack from different points of view containing a handgun.

Detección objetos en rayos X usando Múltiples Vistas  
por Vladimir Riffo, Domingo Mery, Irene Zuccar y Cristhian Pieringer

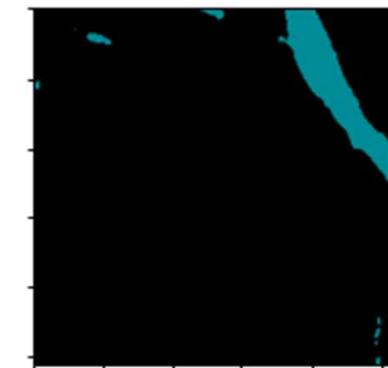
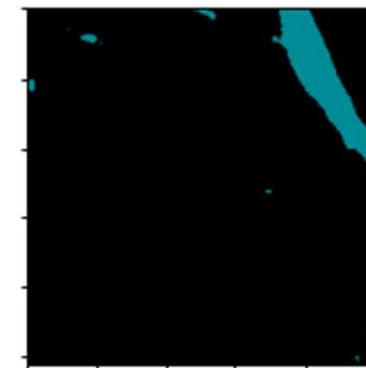
# Aplicaciones

- Microscópicas



# Aplicaciones

- Satelitales  
Objetos de 17x17 son tratados como 1 pixel



Detección de cuerpos de Agua del PeruSat-1

# Aplicaciones

- Seguridad



# Aplicaciones

- Control de calidad





# La imagen Digital

- Una imagen monocroma es almacenada como un arreglo 2D de píxeles
- Cada píxel necesitan 8 bits para representar una tonalidad de gris o intensidad



a) Imagen monocroma (en escala de grises)

# La imagen Digital

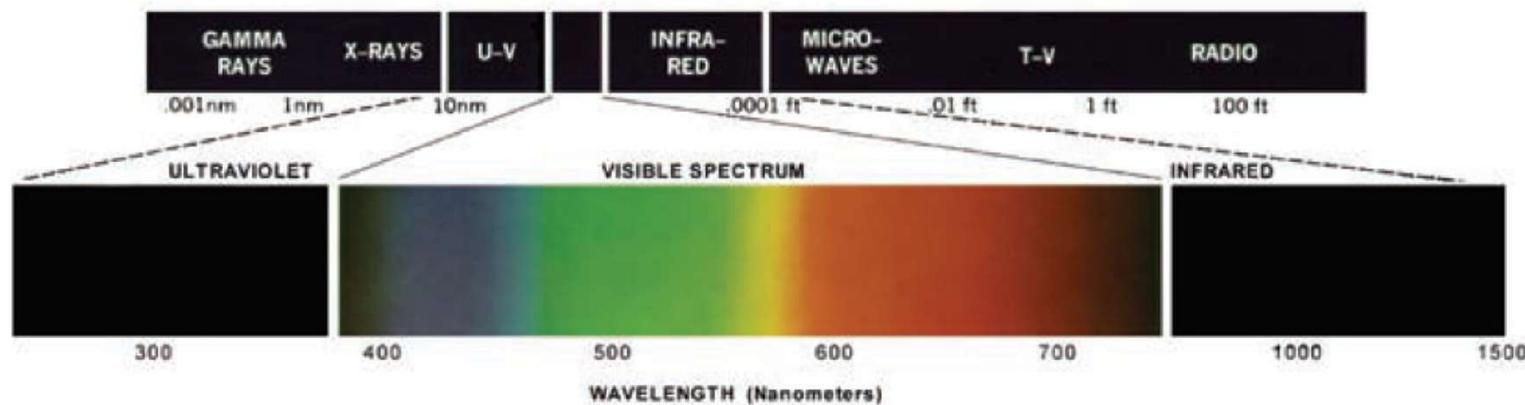
- Una imagen monocroma es almacenada como un arreglo 2D de píxeles
  - Cada píxel necesitan 8 bits para representar una tonalidad de gris o intensidad
  - Cómo se compone una imagen a color?
  - Cómo almacenamos una imagen a color?



- a) Imagen original RGB
- b) Imagen R,
- c) Imagen G
- d) Imagen B

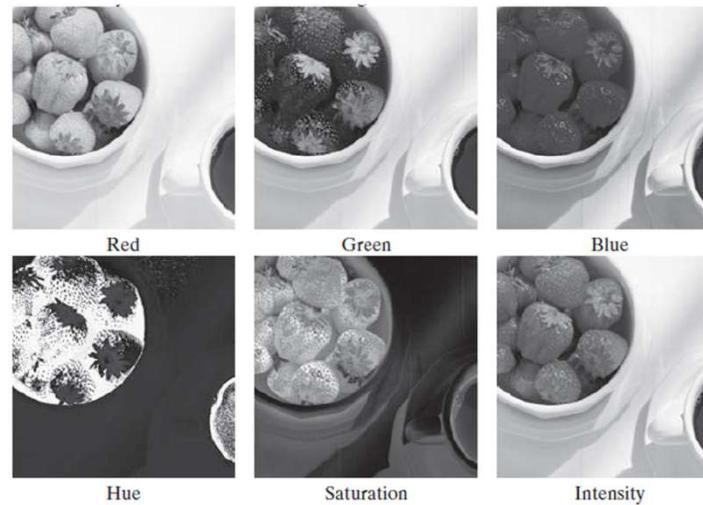
# Espacios de Color

- Humano percibe el color a través de sensores llamados conos (retina)
- El ojo humano absorbe el color como una combinación variable de RGB
- Espectro visible para humanos es 400nm (azul) a 700 nm(rojo)

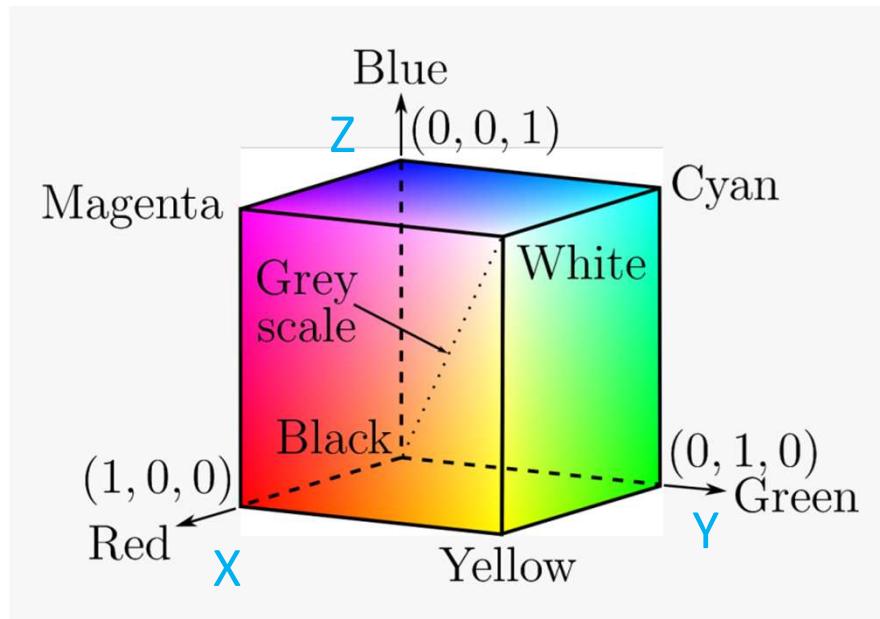


# Color en DIP

- En DIP el color es una propiedad del píxel útil para resolver problemas de reconocimiento
- Existen diversos espacios de color: RGB, HSV (humano), CMY, CMYK , LAB



# Espacio de color RGB

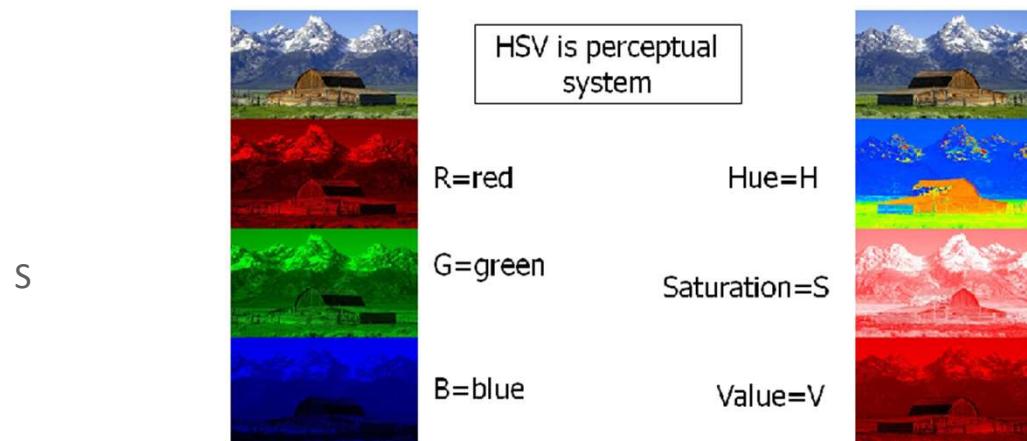


Cubo de Color RGB

- Valores RGB normalizados a 0 – 1
- Colores “puros” en las esquinas
- Gris son valores en la diagonal del blanco al negro
- Que significa el negro?

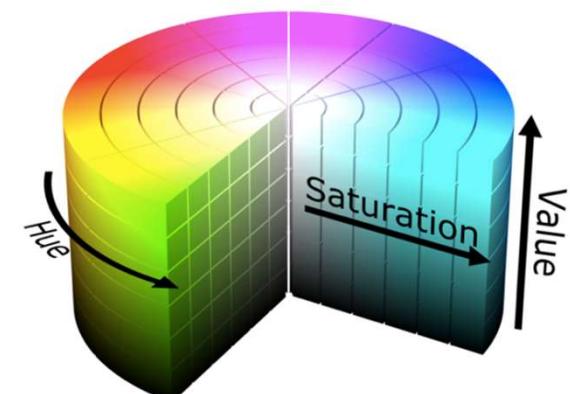
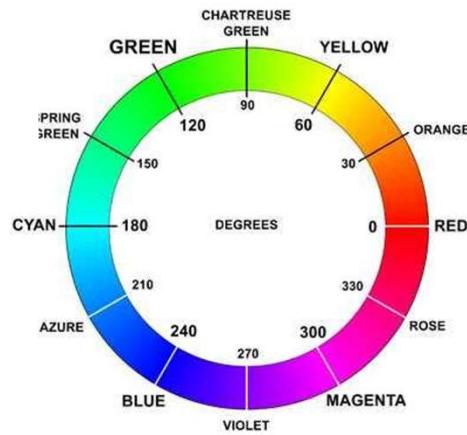
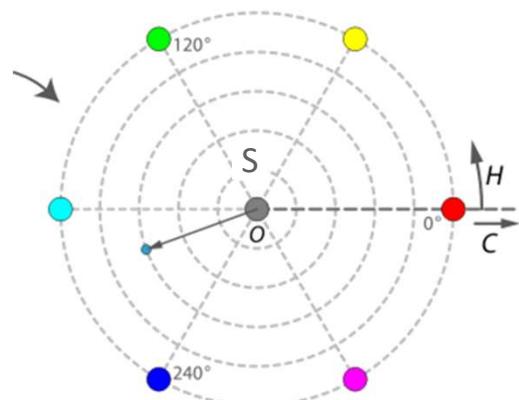
# Espacio de color HSV

- Un color se caracteriza por sus componentes HSV
- Separa la intensidad de la codificación
- Dos valores para codificar cromaticidad (H y S)
- Conveniente para diseñar colores



# Espacio de color HSV

- Componentes HSV :
  - Matiz (Hue), es el color elegido codificado como un ángulo de 0 a 360
  - Saturación, intensidad cromática, es decir, claridad-oscuridad del color (centro círculo al borde 0-100)
  - Luminosidad (Value), todas las versiones de brillantez del color elegido (down 0 – top 100)



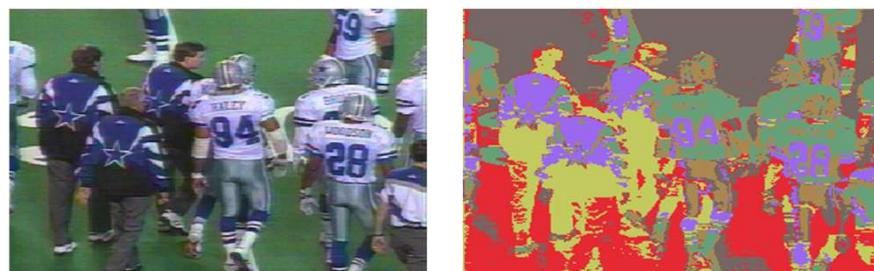
## Ejemplo :



Imagen tomada con cámara digital (izquierda ), al centro saturación de cada pixel decrementado en 20% y a la derecha saturación de cada pixel aumentado en 40%.

# Aplicaciones

- Segmentación a color:

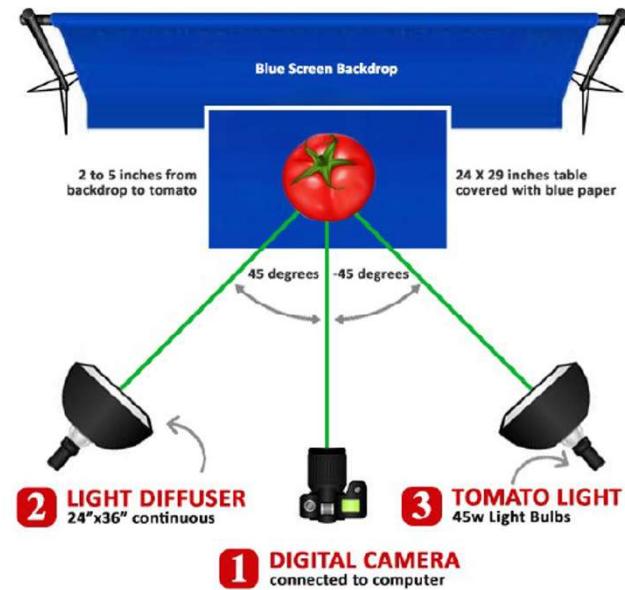
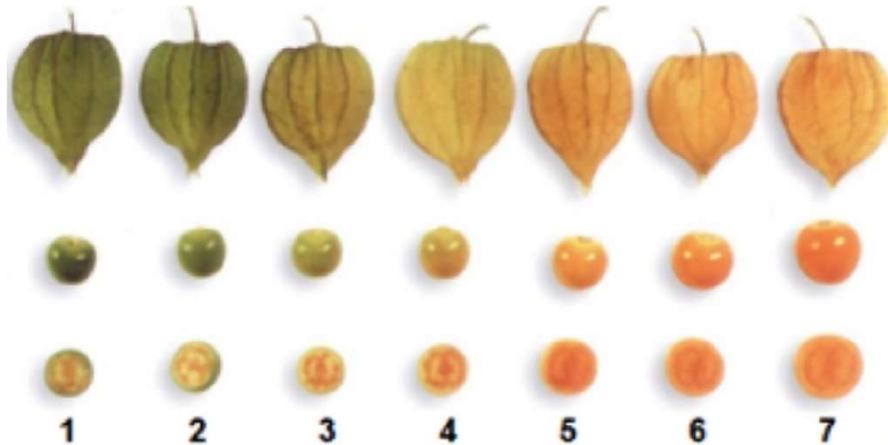


- Detección de Rostros



# Aplicaciones

- Análisis automático de comida:



Identificación del grado de madurez de aguaymanto (izquierda) y tomates (derecha ) usando espacios de color y técnicas de machine learning.