

Indexación, búsqueda y análisis en repositorios multimedia: imagen-video

Juan Carlos San Miguel

Profesorado (sólo parte imagen-video)



Juan C. San Miguel Avedillo
Profesor Contratado Doctor
juancarlos.sanmiguel@uam.es



Video Processing
and Understanding
Lab

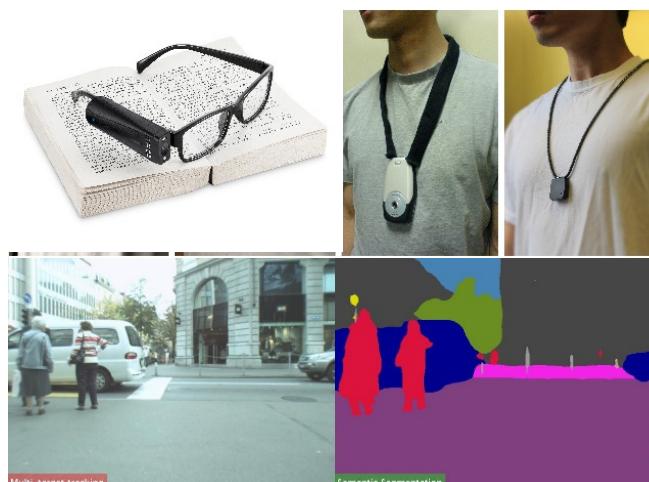
Profesorado (sólo parte imagen-video)

- Investigación en tratamiento de video en **tiempo real** con **múltiples cámaras**

Control de vehículos y personas desde cámaras móviles

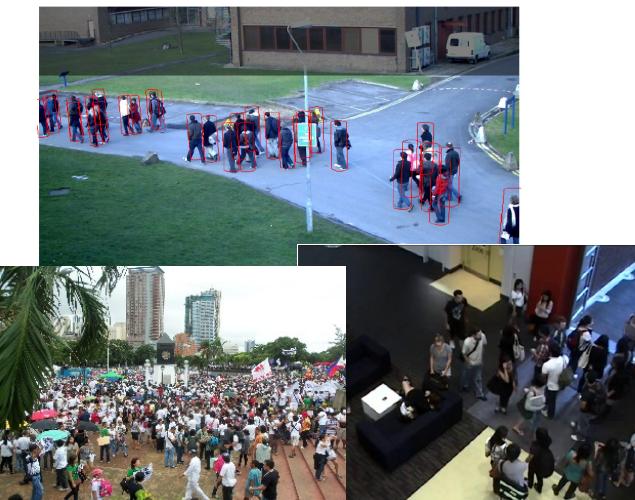


Tecnologías de asistencia para personas ciegas



Video Processing
and Understanding
Lab

Análisis de multitudes con múltiples cámaras



<http://www-vpu.eps.uam.es/>

Agenda: Multimedia (imagen, video)

Bloque 1: Introducción

- Sesión 1 (02/09/22): Aspectos generales sobre visión por computador
- Sesión 2 (15/09/22): Estrategias de indexación y búsqueda

Bloque 2: Indexación basada en Deep learning

- Sesión 3 (16/09/22): Redes convolucionales: fundamentos
- Sesión 4 (22/09/22): Redes convolucionales: entrenamiento 1/2
- Sesión 5 (01/10/22): Redes convolucionales: entrenamiento 2/2
+ Presentación práctica P1 (entrega 09/10/22)
- Sesión 6 (01/10/22): Temas avanzados

Agenda: Multimedia (imagen, video)

Aspectos generales

- Multimedia
- Información de interés
- Ejemplos
- Áreas de conocimiento
- Principales problemas
- Relación con Big Data
- Imágenes
 - Formación y representación
 - Tratamiento de imágenes
 - Componentes
 - Operadores
- Video
 - Estructura espacio-temporal
 - Tratamiento de video
 - Segmentación y seguimiento de objetos
 - Reconocimiento de actividades

Agenda: Multimedia (imagen, video)

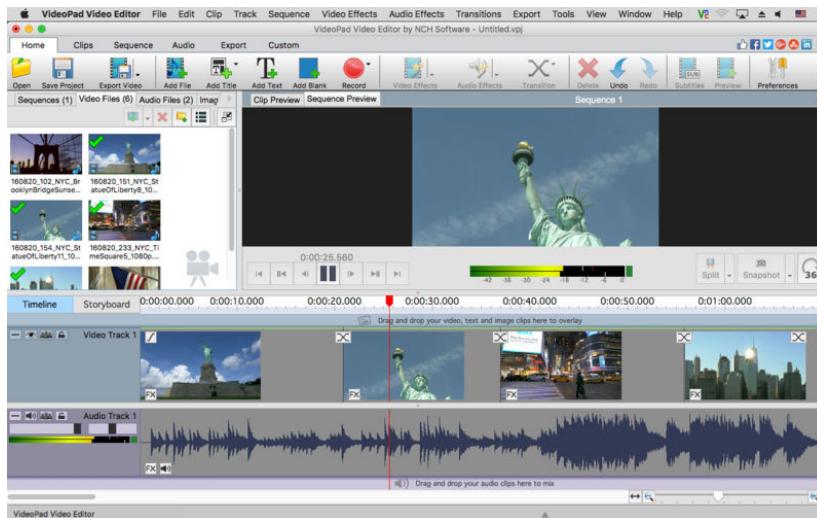
Aspectos generales

- **Multimedia**
- **Información de interés**
- **Ejemplos**
- **Áreas de conocimiento**
- **Principales problemas**
- **Relación con Big Data**
- **Imágenes**
 - Formación y representación
 - Tratamiento de imágenes
 - Componentes
 - Operadores
- **Video**
 - Estructura espacio-temporal
 - Tratamiento de video
 - Segmentación y Seguimiento de objetos
 - Reconocimiento de actividades

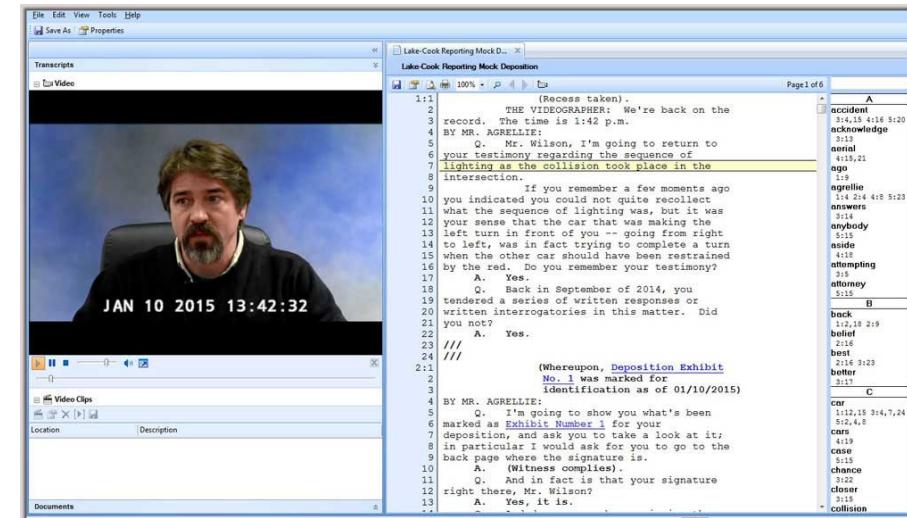
¿Qué es multimedia?

Mezcla coherente de tipos de datos de distinta naturaleza (texto, imágenes, audio, video)

Definición del ACM Special Interest Group on Multimedia Retreat 2003



imágenes+audio



video+texto

Indexación, búsqueda y análisis
en repositorios multimedia

Información de interés en imágenes y video...



¿Qué tipo de escena?

¿Qué objetos hay en la escena?

¿Dónde están los coches?

¿A qué distancia está el edificio?

...

Computer Vision es una disciplina que trata de resolver estas preguntas mediante la comprensión automática de imágenes y videos

Ejemplos



<https://youtu.be/NrmMk1Myrxc>

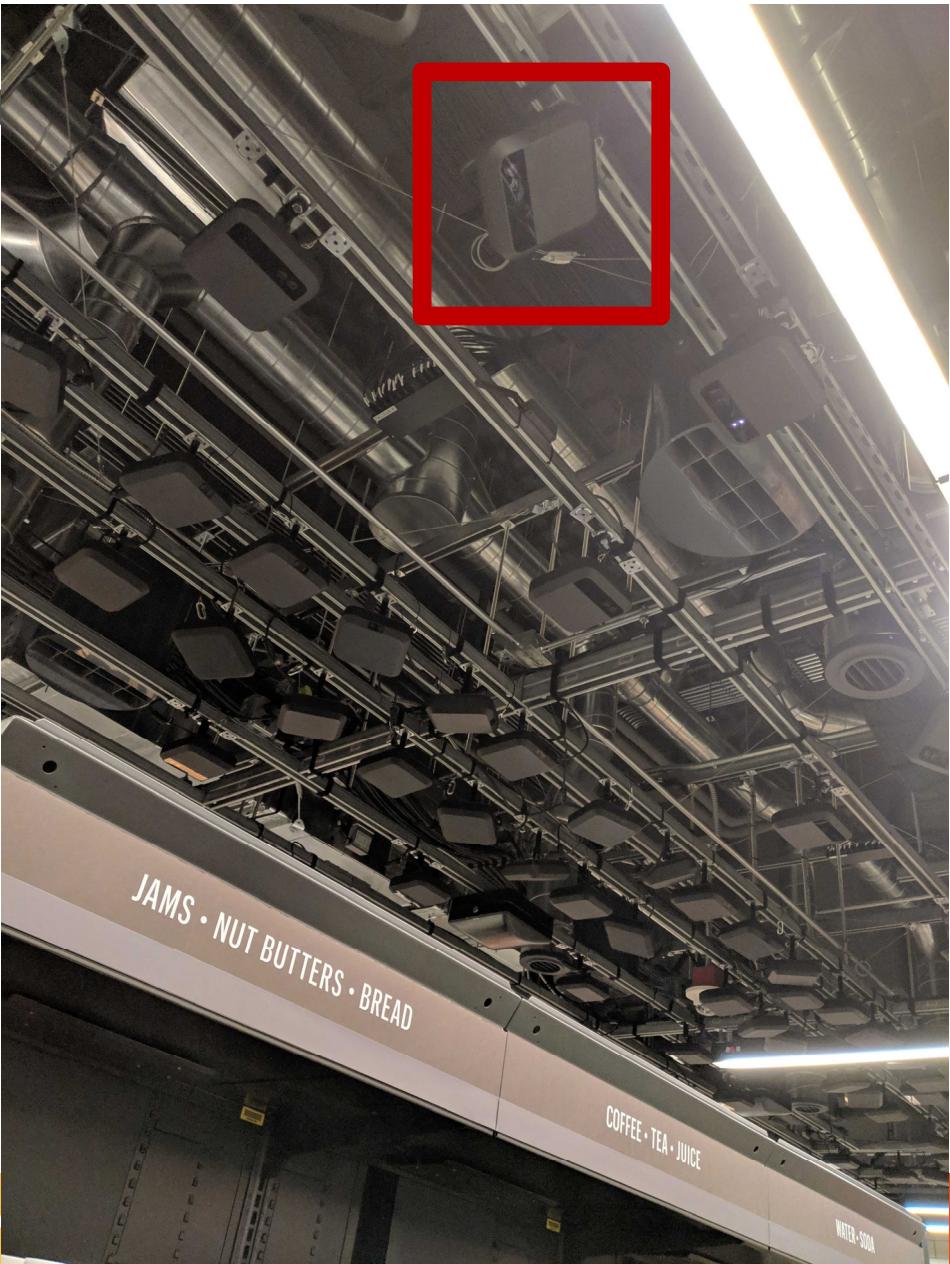


1:00

¿Cómo crees que funciona?



Ejemplo



© Vivek Ramanujan

Indexación, búsqueda y análisis
en repositorios multimedia

Ejemplos

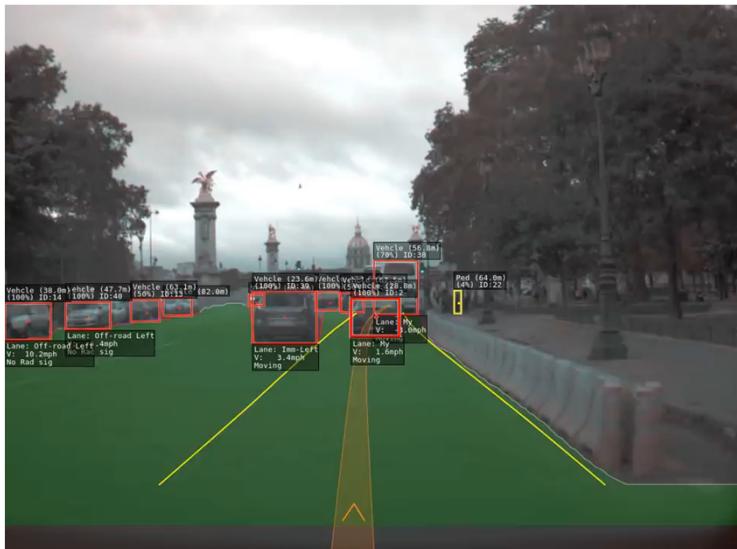
➤ Vehículo autónomo: Tesla Autopilot



<https://electrek.co/2018/06/11/tesla-ai-director-insights-autopilot-computer-vision-neural-net/>

Ejemplos

➤ Tesla Autopilot



https://youtu.be/_1MHGUC_BzQ



<https://youtu.be/24dRkHdpEPo>

Ejemplos

- Vehículo autónomo



<https://www.instagram.com/p/B6aBUxYlsZ0>

Ejemplos

□ Realidad Aumentada/Mixta



<https://youtu.be/Lum4tTsqKbc>

Demostrador para comercio minorista
Proyecto FUAM 2019/0165
Desarrollo de aplicaciones sobre dispositivo Magic Leap One

 Fundación de la Universidad Autónoma de Madrid

 Conexionio

 Cincuenta Aniversario UAM Universidad Autónoma de Madrid

 Escuela Politécnica Superior

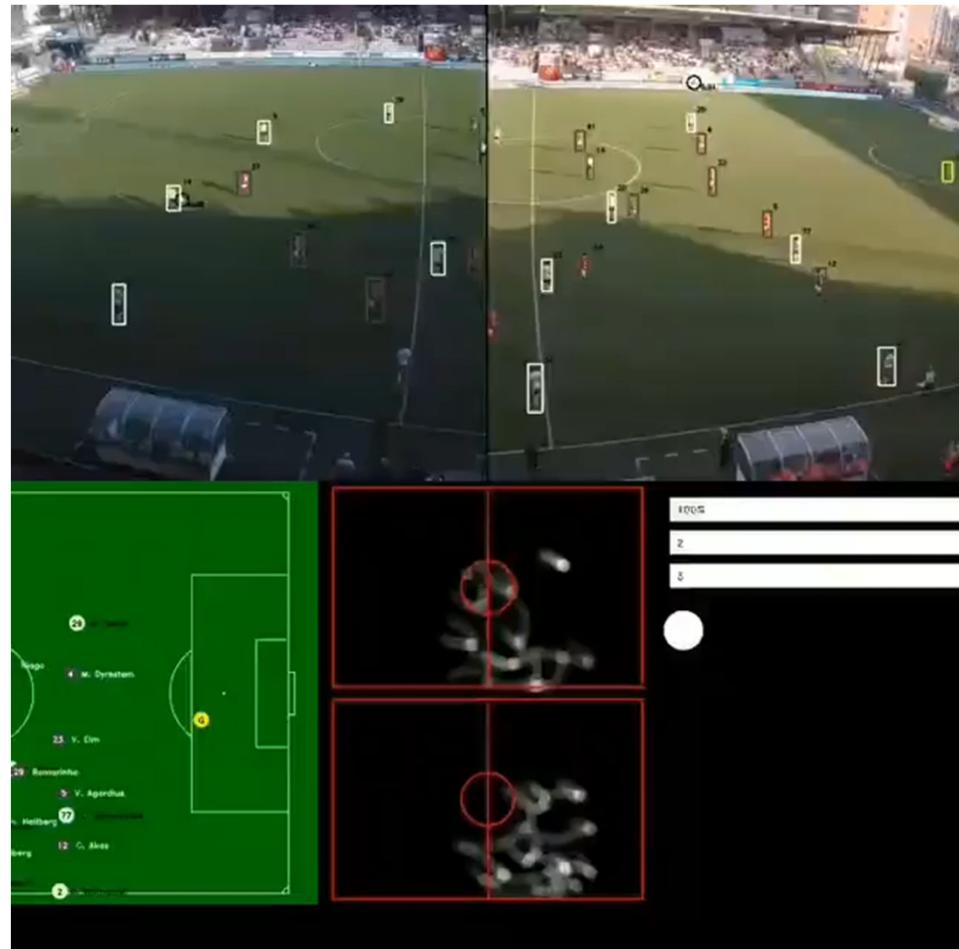
 Video Processing and Understanding Lab

<https://youtu.be/VJsX6p6cs2c>

Indexación, búsqueda y análisis
en repositorios multimedia

Ejemplos

➤ Deportes



<https://www.instagram.com/p/B69Om2JJtIB/>

Ejemplos

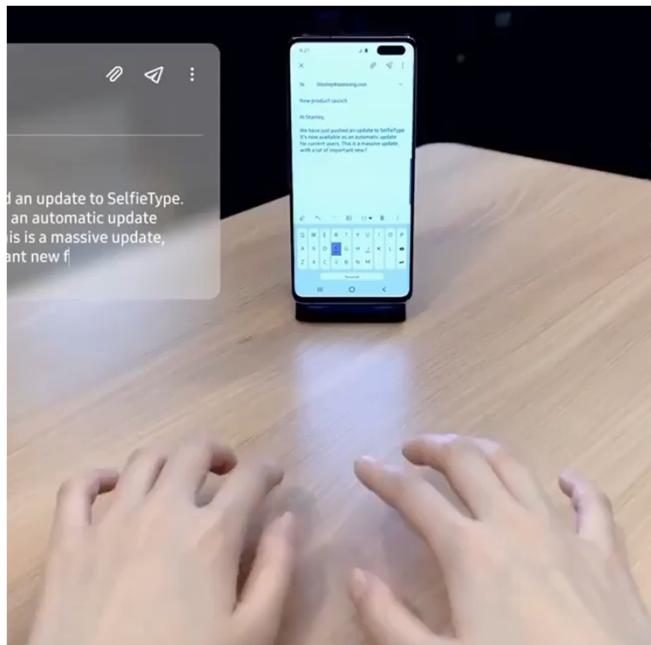
- Clasificación de basura y recolección automática



https://www.instagram.com/p/B6_A8hJJ2-V

Ejemplos

- Visto recientemente...



<https://www.instagram.com/p/B7AtaOsAXIO>



<https://www.instagram.com/p/B7BsSq4FqLp>



Indexación, búsqueda y análisis
en repositorios multimedia

Ejemplos



IBM

Using Watson in
the **battle against**
skin cancer.

Client Success Field Notes

A close-up image showing a person's arm with a dark mole. A smartphone is held over the mole, displaying a digital interface from the Watson system for skin cancer analysis.

https://youtu.be/n3_oGnXkMAE

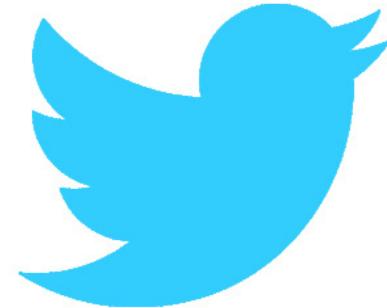
Ejemplos

- ¿Más?



Instagram

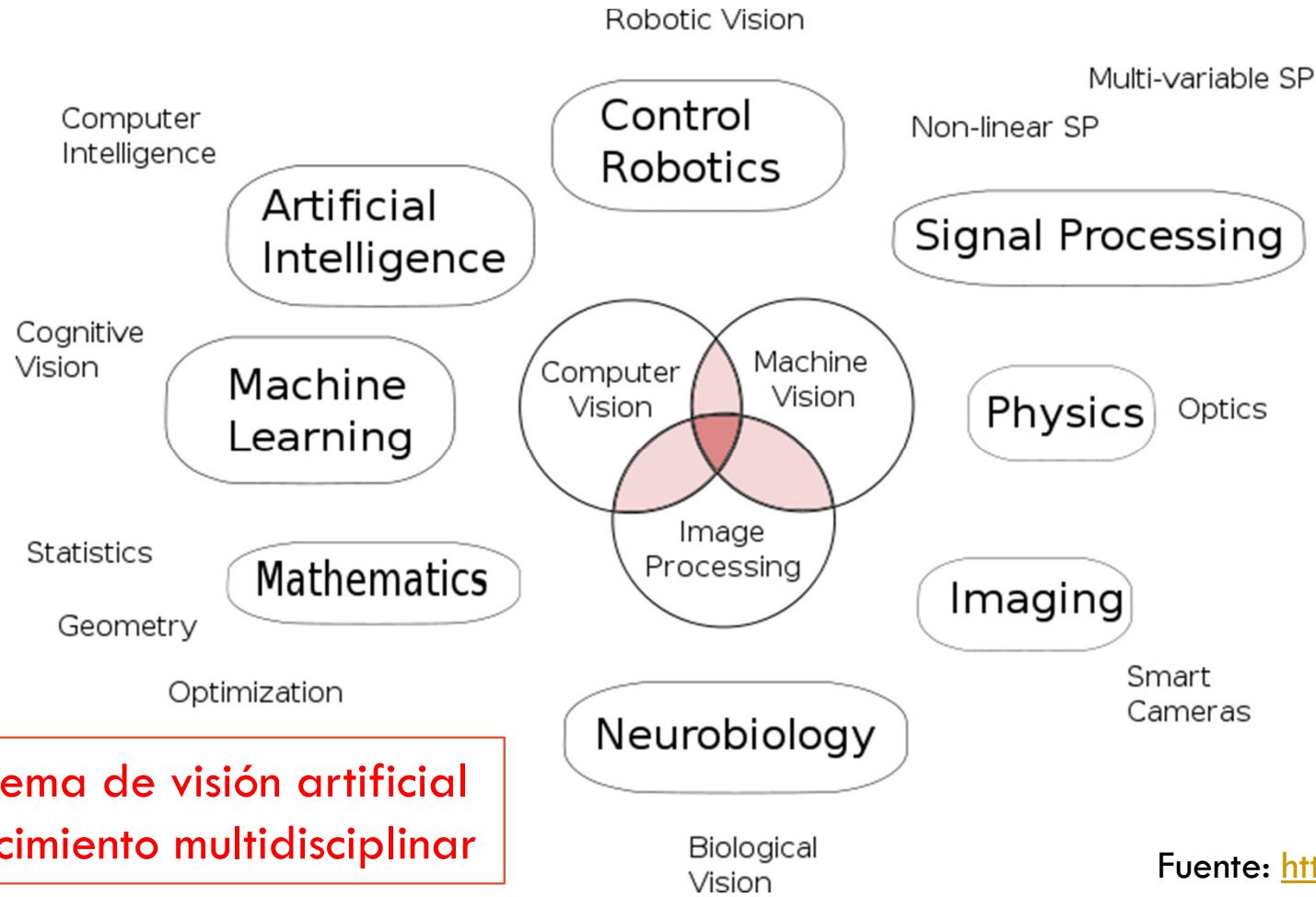
[https://www.instagram.com/explore
/tags/computervision/?hl=es](https://www.instagram.com/explore/?tags/computervision/?hl=es)



twitter

[https://twitter.com/search?
q=%23ComputerVision](https://twitter.com/search?q=%23ComputerVision)

Áreas de conocimiento relacionadas



Problemas principales (1/5)



Punto
de vista



Illuminación



Escala

Problemas principales (2/5)



Variación
intraclasa



Movimiento
de objetos



Distractores (clutter)



Oclusiones

Problemas principales (3/5)

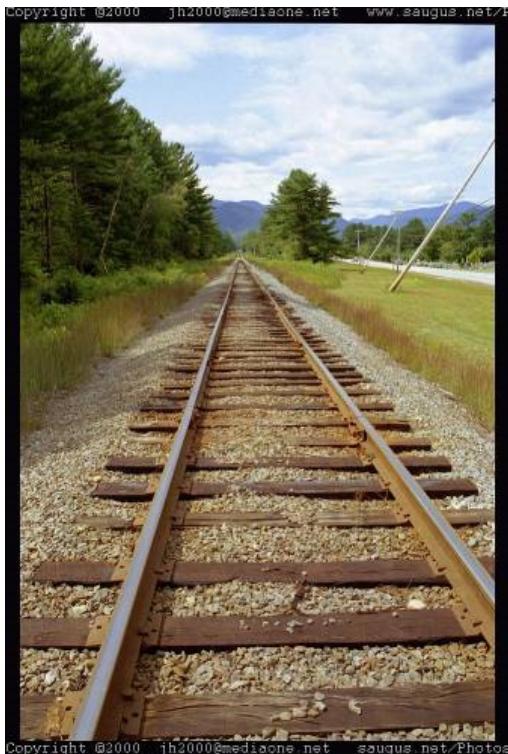
- Punto de vista + proyección



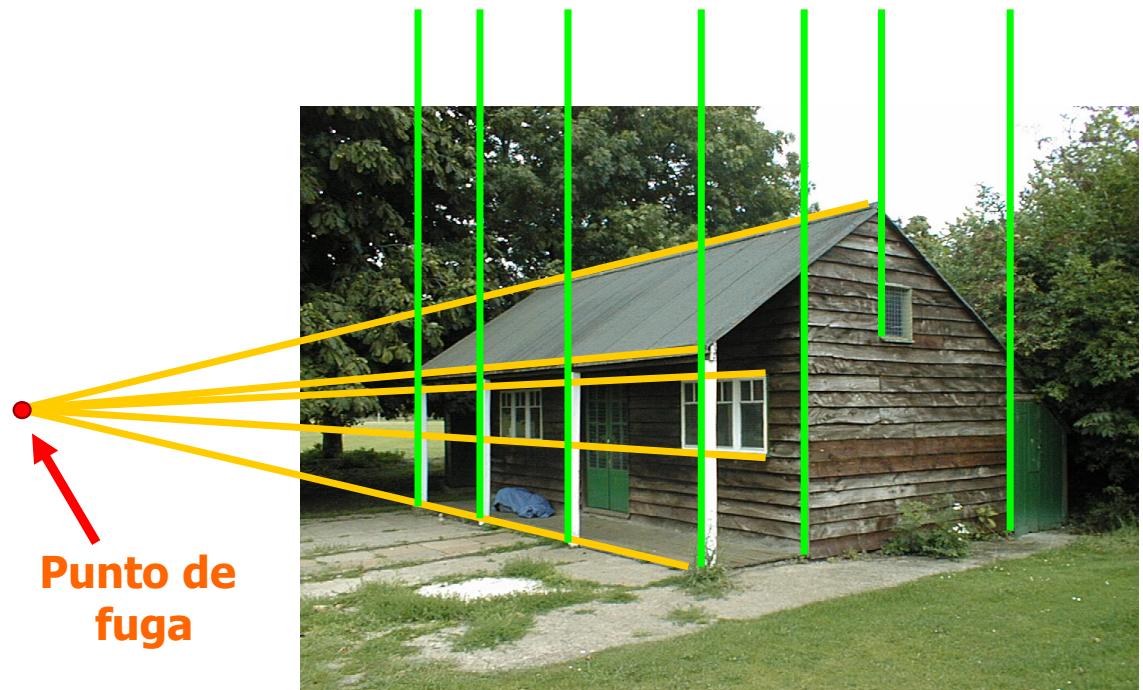
Fuente: <http://www.coolopticalillusions.com>

Problemas principales (4/5)

Líneas paralelas en la realidad
pueden no serlo en la imagen...



...pero su proyección “infinita” puede interseccionar
(puntos de fuga o *vanishing points*)



© Prof. Efros

Problemas principales (5/5)

➤ Conocimiento a priori



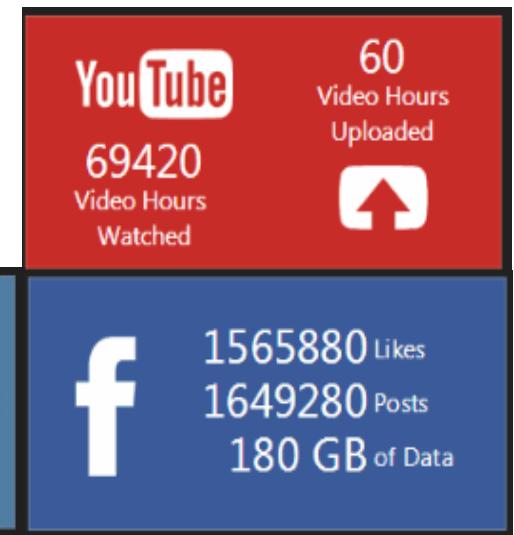
Big Data: imágenes y video (1/3)

- Billones de imágenes/vídeos cada día...



+170M cámaras en China!!
(fuente <http://bit.ly/30etH5n>)

Internet en 30
segundos...



<https://visual.ly/community/infographic/how/internet-real-time>

Indexación, búsqueda y análisis
en repositorios multimedia

Big Data: imágenes y video (2/3)

ImageNet Large Scale Visual Recognition Challenge

- 1K categorías / +1.2M imágenes

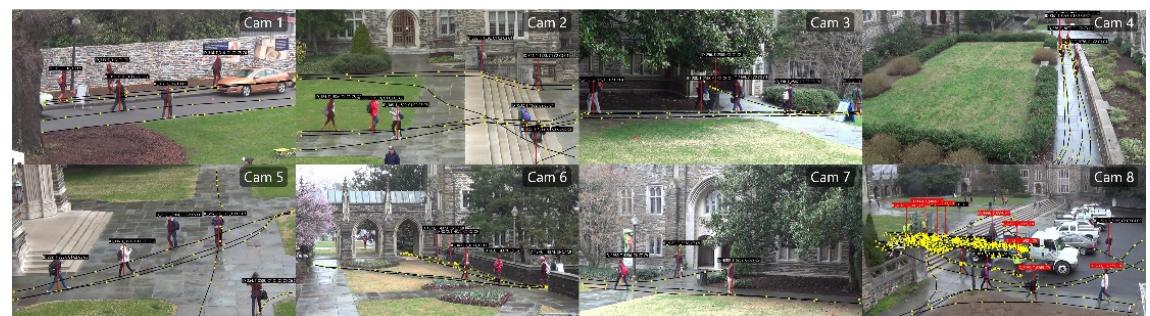


<http://image-net.org/challenges/LSVRC/2017/>

MCMT Multi-camera Multi-target dataset

- 8 cámaras
- 85 minutos @1080p 60fps
- +2M imágenes, +2K personas

<http://vision.cs.duke.edu/DukeMTMC/>

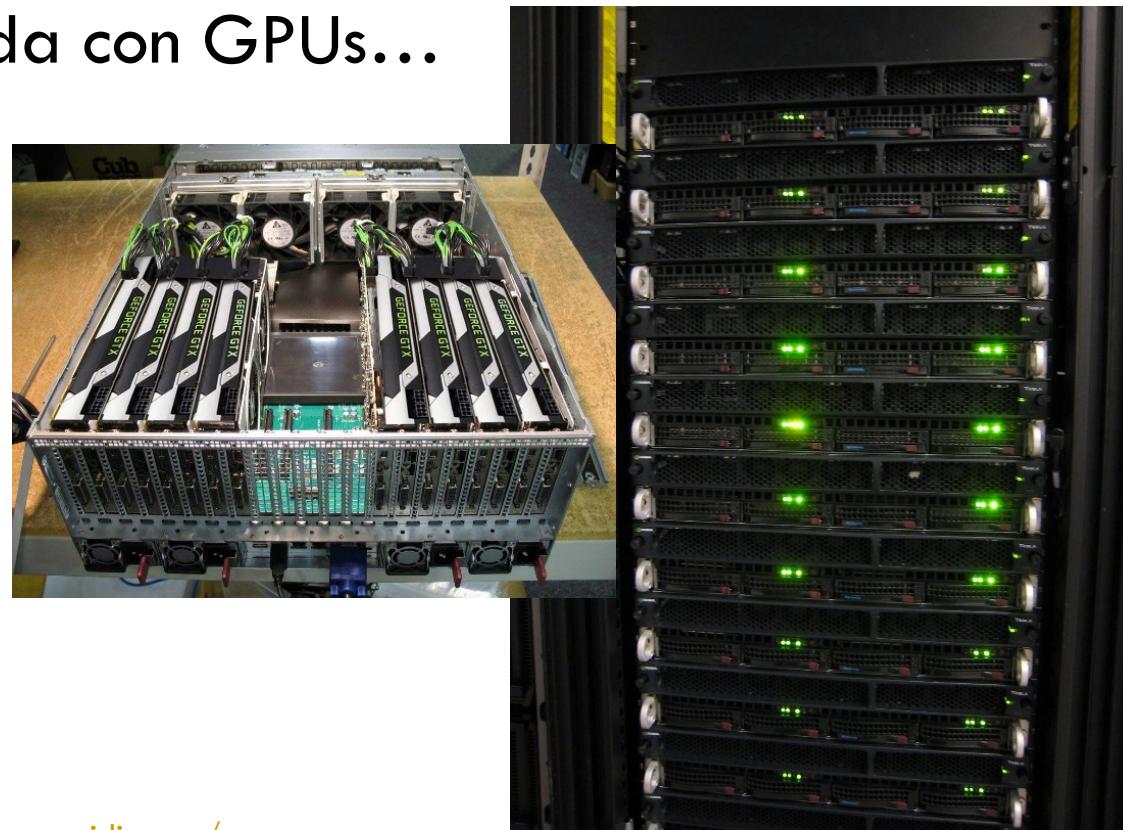
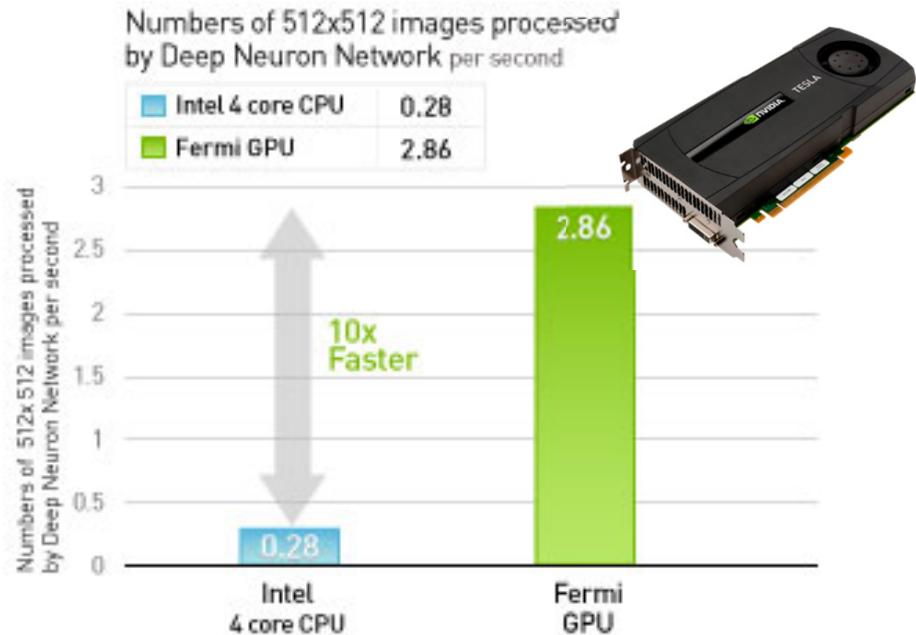


Más en <https://computervisiononline.com/datasets>

Indexación, búsqueda y análisis
en repositorios multimedia

Big Data: imágenes y video (3/3)

- Necesarias altas capacidades de computación
- Boom sólo en la última década con GPUs...



Fuente: <http://www.nvidia.com/>

Agenda: Multimedia (imagen, video)

Aspectos generales

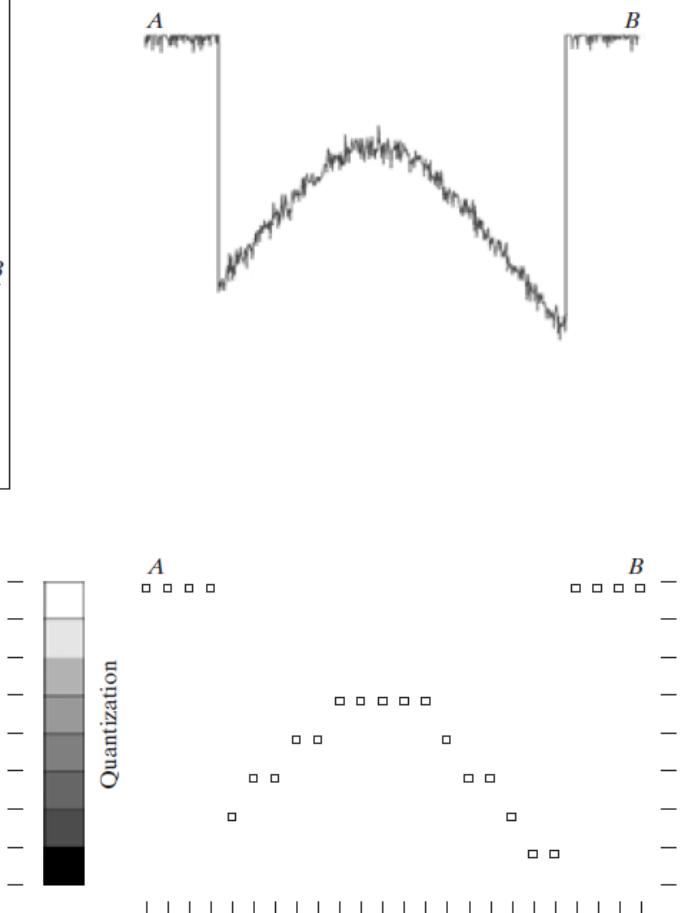
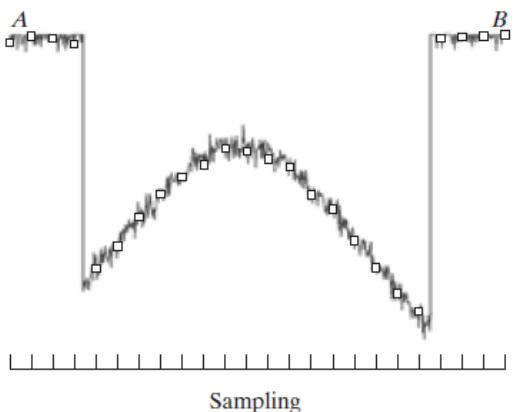
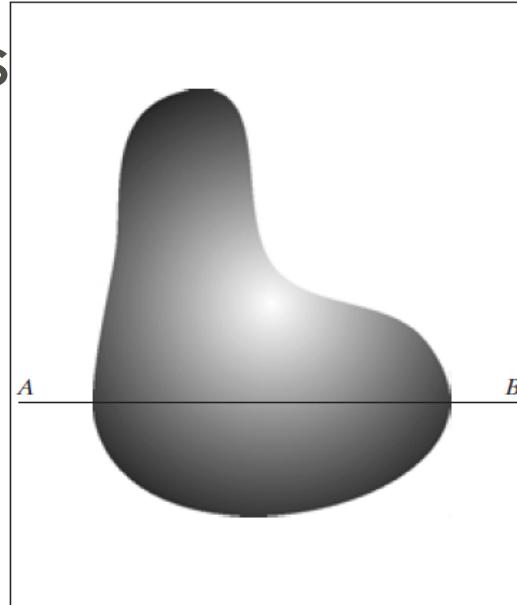
- Multimedia
- Información de interés
- Ejemplos
- Áreas de conocimiento
- Principales problemas
- Relación con Big Data
- Imágenes
 - Formación y representación
 - Tratamiento de imágenes
 - Componentes
 - Operadores puntuales, locales y globales
- Video
 - Estructura espacio-temporal
 - Tratamiento de video
 - Segmentación de objetos
 - Seguimiento de objetos
 - Reconocimiento de actividades

Formación de imágenes

➤ La imagen digital

Señales en dos dimensiones 2D
muestreadas y cuantificadas:

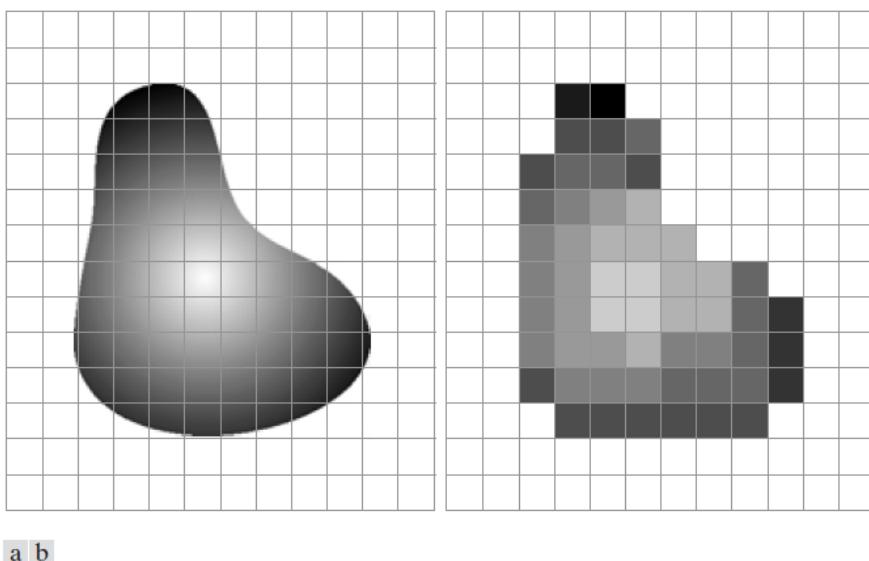
- definidas en un rango limitado y discreto de posiciones
- con un rango limitado y discreto de amplitudes



Gonzalez, Digital Image Processing, 2002

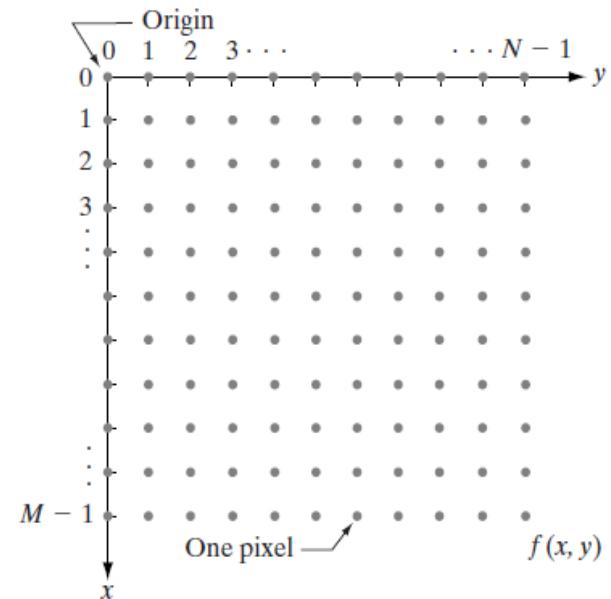
Formación de imágenes

- **Pixel:** la menor unidad espacial homogénea que forma la imagen digital y describe el color capturado por la cámara



a b

FIGURE 2.17 (a) Continuos image projected onto a sensor array. (b) Result of image sampling and quantization.



Fuente: "Digital Image Processing", R. Gonzalez and Goods, Prentice Hall, 2002

Representación directa (true color)

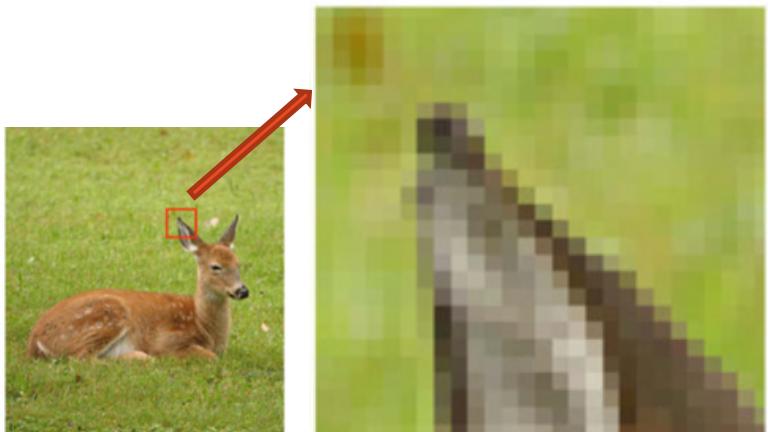
- Imagen monocroma (1 canal) con 8 bits/píxel (Valores: [0, 255])



62	62	63	64	65	66	67	67	69	70	71	72	72	73	73	73	72	72	71	70	69	67	66	66	65	63	62	61	60	6			
61	62	63	64	66	66	67	68	68	69	70	71	71	72	72	73	72	72	71	71	70	69	68	66	66	65	65	63	62	61	60	6	
61	62	63	64	66	66	68	68	69	70	70	71	72	73	73	73	72	72	71	71	69	68	67	66	66	65	65	64	63	62	61	6	
61	63	64	64	66	67	68	68	68	69	70	71	71	73	73	74	73	73	73	71	70	69	68	66	66	65	64	63	62	61	61	6	
61	63	64	65	67	68	68	68	69	70	70	71	71	72	55	53	69	72	72	71	71	70	69	68	67	66	65	64	63	62	60	60	6
63	64	65	66	67	68	68	69	69	70	70	71	72	42	4	5	11	48	72	71	71	69	69	68	67	66	65	64	62	62	60	59	5
63	65	66	66	68	68	69	70	71	71	71	72	18	4	4	7	8	66	71	70	69	68	68	67	66	65	64	63	61	59	59	5	
63	65	67	67	68	68	69	70	71	71	72	64	4	27	24	54	33	29	52	64	68	68	67	66	65	64	63	62	61	59	58	5	
64	65	66	66	68	68	69	70	71	41	24	24	12	17	24	48	60	37	43	36	52	66	68	67	66	65	64	63	61	60	59	58	5
65	66	67	67	68	68	69	71	48	6	6	6	5	34	36	12	47	34	17	29	54	43	63	67	66	65	64	63	62	60	59	58	5
64	65	66	66	68	68	69	71	38	6	6	5	7	16	19	4	47	44	27	40	40	67	66	66	65	65	64	63	61	60	59	58	5
63	64	65	65	67	60	6	6	5	5	5	6	8	9	20	27	51	78	41	44	66	65	65	65	64	63	62	60	59	58	5		
63	64	65	65	64	34	5	5	5	5	5	5	4	19	6	7	54	64	20	59	65	65	64	64	64	63	62	61	60	59	57	5	
63	64	64	65	14	5	6	5	5	4	5	4	18	7	5	4	19	10	11	65	64	64	64	63	61	66	62	61	60	59	58	5	
63	64	64	65	53	7	4	5	6	6	7	10	6	5	4	21	24	18	64	64	64	63	62	64	65	62	62	60	59	58	5		
64	64	64	64	65	50	4	4	4	5	11	16	6	6	4	35	16	26	66	64	64	63	61	72	67	63	62	61	59	58	5		
64	64	64	64	65	46	4	4	4	5	6	9	8	5	29	10	43	56	29	57	64	64	63	61	70	67	62	64	65	59	59	5	
64	64	64	65	66	27	5	4	4	5	6	6	6	18	66	20	57	60	46	34	75	70	62	61	70	67	62	61	60	59	58	5	
49	50	62	65	57	5	5	6	5	6	6	6	6	41	59	28	60	58	44	22	63	71	72	60	69	68	61	60	58	59	59	5	
42	52	57	52	26	5	5	5	5	5	5	5	5	70	50	43	61	62	64	39	42	64	60	62	56	63	65	65	67	61	53	5	
32	32	32	33	6	5	5	5	5	6	6	11	39	21	33	51	50	45	46	18	32	36	33	23	44	70	71	51	42	27	3		
50	50	51	39	5	5	5	6	5	6	6	42	69	28	34	42	39	43	37	26	29	40	26	29	26	35	42	35	33	18	1		
52	53	51	22	5	5	5	6	5	6	5	44	56	17	51	54	53	54	56	51	22	54	54	55	55	54	53	53	52	5			
54	54	53	8	5	5	5	6	5	6	13	52	42	21	51	54	51	49	49	50	22	41	45	42	42	41	40	41	44	43	4		
52	52	54	38	8	5	5	6	5	6	28	55	32	32	54	53	51	51	51	51	44	25	51	51	49	49	50	49	48	46	4		
54	54	52	53	30	7	5	6	6	5	6	40	54	29	52	51	53	56	55	52	52	51	38	52	52	50	49	46	46	45	46	4	
51	52	51	53	27	14	5	4	4	7	47	51	21	39	49	47	49	52	52	52	49	35	31	48	46	47	47	47	46	46	4		
48	50	51	53	25	14	17	8	4	4	17	46	40	18	43	47	46	49	52	54	53	53	54	18	50	49	46	47	47	47	4		
49	49	49	49	22	12	20	24	6	14	35	51	39	48	48	50	51	51	49	51	51	52	50	41	58	48	47	47	47	45	4		
51	49	50	50	22	13	19	38	13	12	42	50	40	73	50	50	50	49	48	49	49	48	49	45	51	46	44	44	42	45	4		
47	49	49	47	20	16	26	39	21	15	36	48	42	61	47	48	51	47	50	51	51	51	49	47	47	52	47	47	44	43	45	4	

Representación directa (true color)

- Imagen de color (3 canales)
 - Cada pixel 24 bits/píxel
 - Rojo/Red – 8 bit (256 valores)
 - Verde/Green – 8 bit (256 valores)
 - Azul/Blue – 8 bit (256 valores)



49	55	56	57	52	53
58	60	60	58	55	57
58	58	54	53	55	56
83	78	72	69	68	69
88	91	91	84	83	82
69	76	83	78	76	75
61	69	73	78	76	76

Red

64	76	82	79	78	78
93	93	91	91	86	86
88	82	88	90	88	89
125	119	113	108	111	110
137	136	132	128	126	120
105	108	114	114	118	113
96	103	112	108	111	107

Green

66	80	77	80	87	77
81	93	96	99	86	85
83	83	91	94	92	88
135	128	126	112	107	106
141	129	129	117	115	101
95	99	109	108	112	109
84	93	107	101	105	102

Blue

Fuente: An introduction to Digital Image Processing with MATLAB, , A. McAndrew, Victoria University of Technology (Australia), 2004.

Representación indirecta

➤ Imagen de color – VLT (Video Lookup Table)

Imagen tamaño $M \times N$



Matriz tamaño $M \times N$

4	5	5	5	5	5
5	4	5	5	6	6
5	5	5	0	8	9
5	5	5	5	11	11
5	5	5	8	16	20
8	11	11	26	33	20
11	20	33	33	58	37

Indices

Número de colores representable (índices)

VLT tamaño $K \times 3$

0.1211 0.1211 0.1416

0.1807 0.2549 0.1729

0.2197 0.3447 0.1807

0.1611 0.1768 0.1924

0.2432 0.2471 0.1924

0.2119 0.1963 0.2002

0.2627 0.2588 0.2549

0.2197 0.2432 0.2588

:

:

:

Color RGB

Colour map

Fuente: An introduction to Digital Image Processing with MATLAB, , A. McAndrew, Victoria University of Technology (Australia), 2004.



3:00

- ¿Qué representación ocupa más espacio para una imagen en escala de grises (con 256 niveles)?



¿ Directa (true color)
ó Indirecta (VLT) ?



3:00

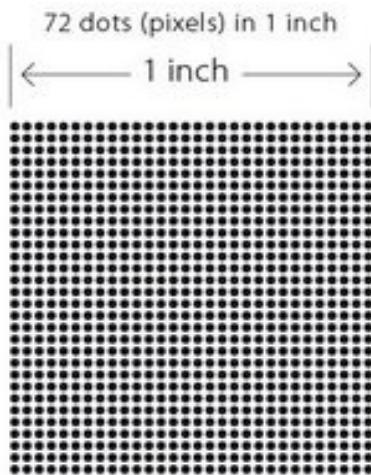
- ¿Qué representación ocupa más espacio para una imagen RGB que permite representar colores?



¿ Directa (true color) con 256 niveles para cada valor
ó
Indirecta (VLT) con 256 colores distintos en la paleta?

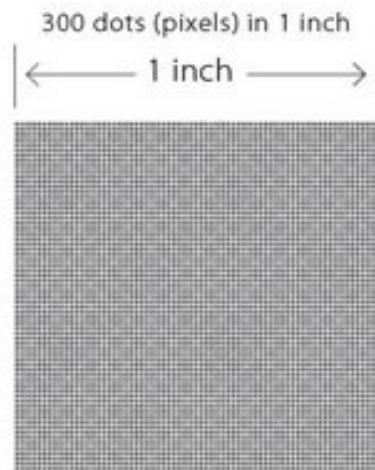
Representación: Resolución espacial (dpi)

➤ Resolución vs definición



72 dpi

72 dots per-inch



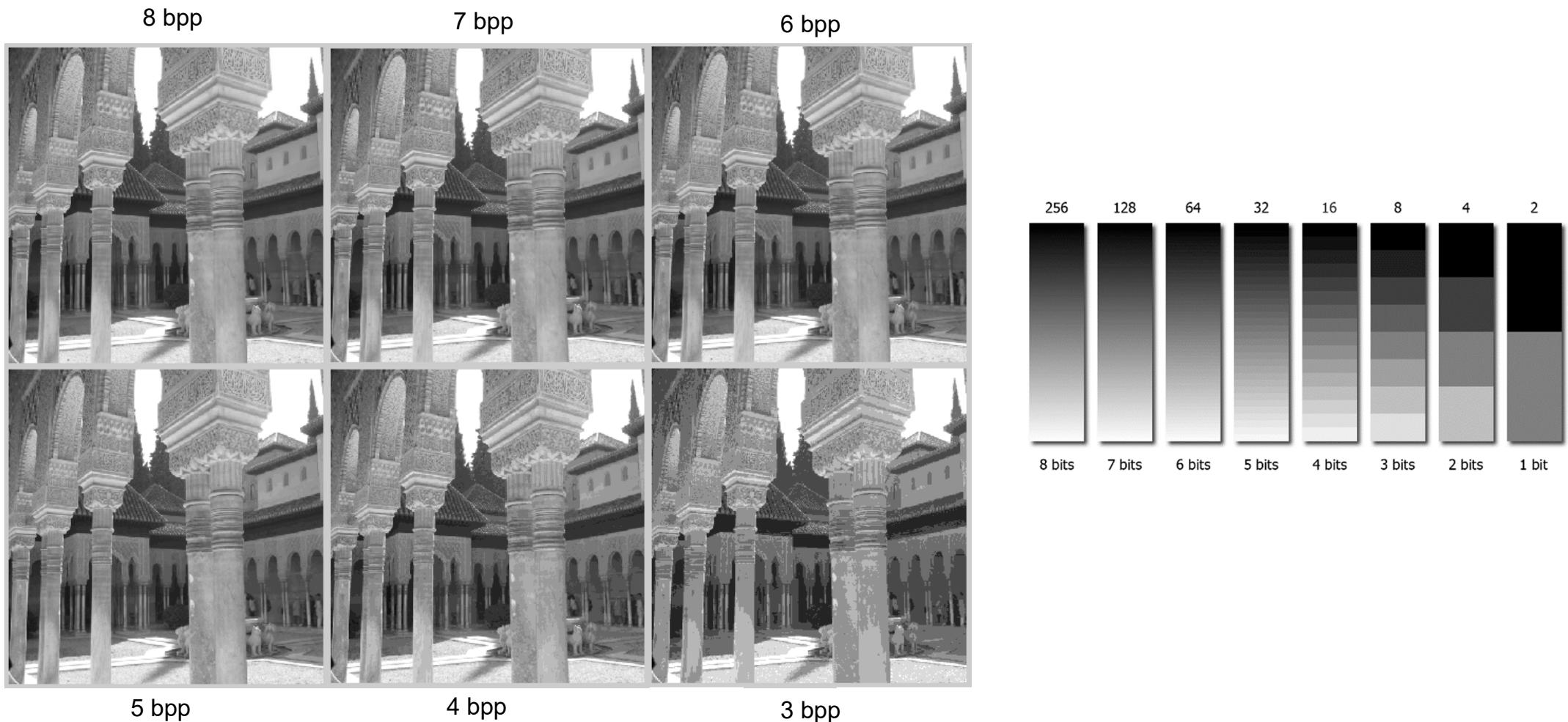
300 dpi

300 dots per-inch

Fuente: hp.com.



Representación: profundidad (Bits por pixel, bpp)



Representación: compresión

- Formatos de imágenes:
 - Sin compresión: BMP, RAW, PPM
 - Con compresión (sin pérdidas): PNG, GIF, TIFF
 - Con compresión (con pérdidas): JPEG, TIFF



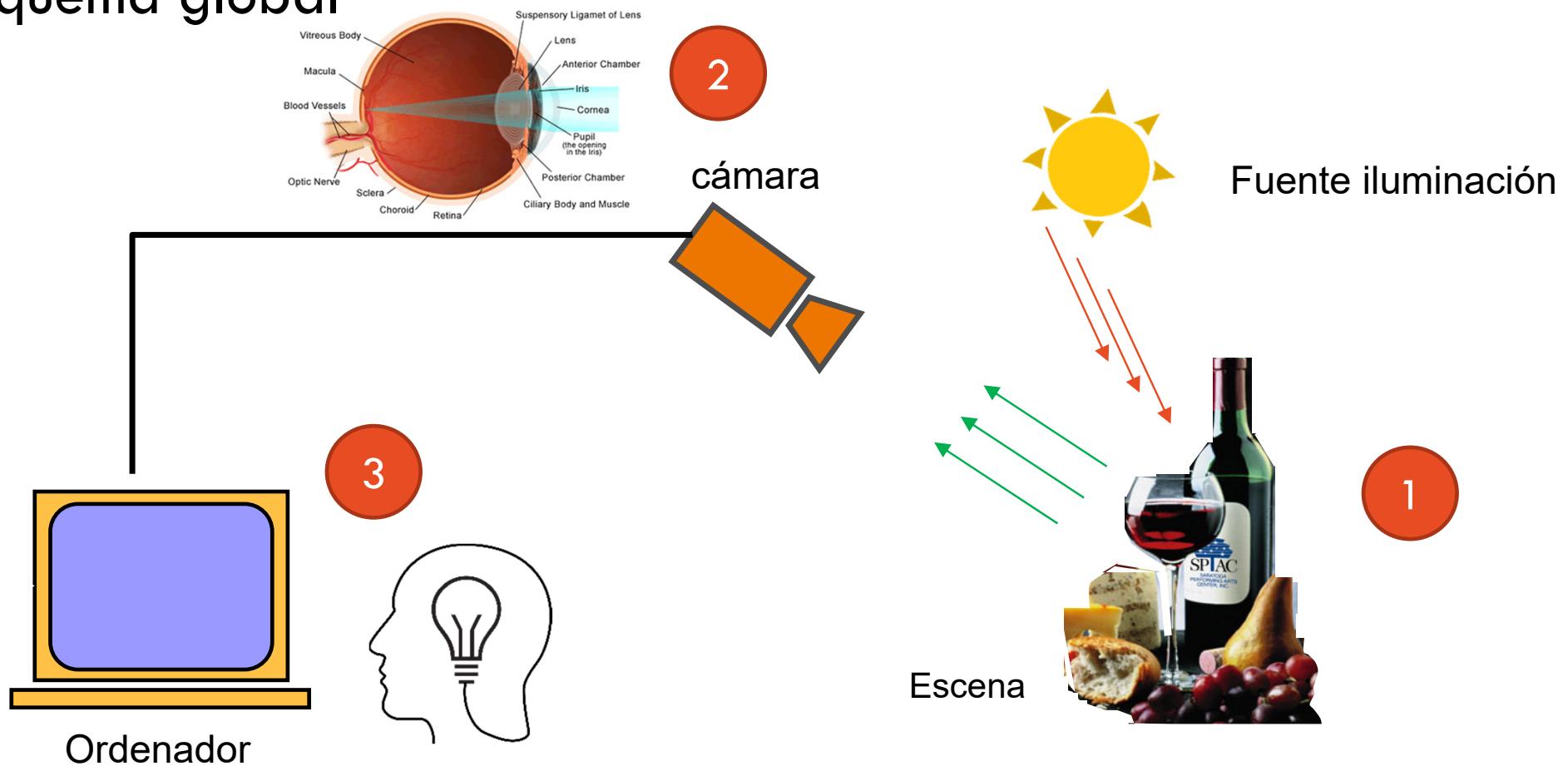
Mo

JPEG 90% - 5.93KB

SIN COMPRIMIR - 57.5KB

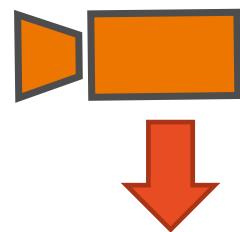
Componentes de un sistema de tratamiento de imágenes

➤ Esquema global

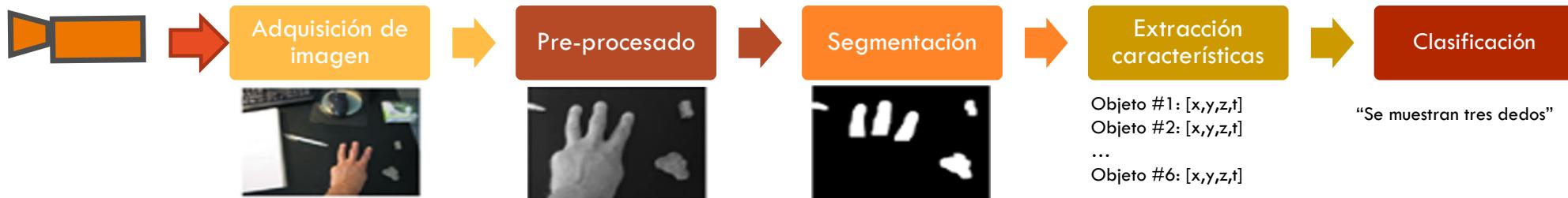


Componentes de un sistema de tratamiento de imágenes

➤ Cadena de procesamiento



Tipos de operadores



Operador puntual: cada píxel de la imagen resultante es función del píxel homólogo de la imagen original

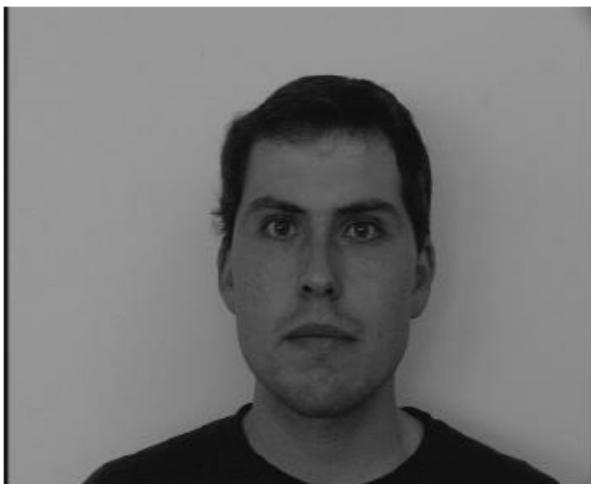
Operador local: cada píxel de la imagen resultante es función de un entorno de píxeles del píxel homólogo de la imagen original

Operador global: cada píxel de la imagen resultante es función de todos los píxeles de la imagen original

Los tipos de operadores se pueden aplicar en cualquier etapa de procesado

Operadores puntuales: preprocessamiento

➤ Escalado de imagen



$$\psi[n, m]$$



$$\theta[n, m] = a \cdot \psi[n, m]$$

Fuente: *Introduction to Digital Image Processing* (EE368, <http://www.stanford.edu/class/ee368/>)

Operadores puntuales: preprocessamiento

➤ Umbralización

Imagen original

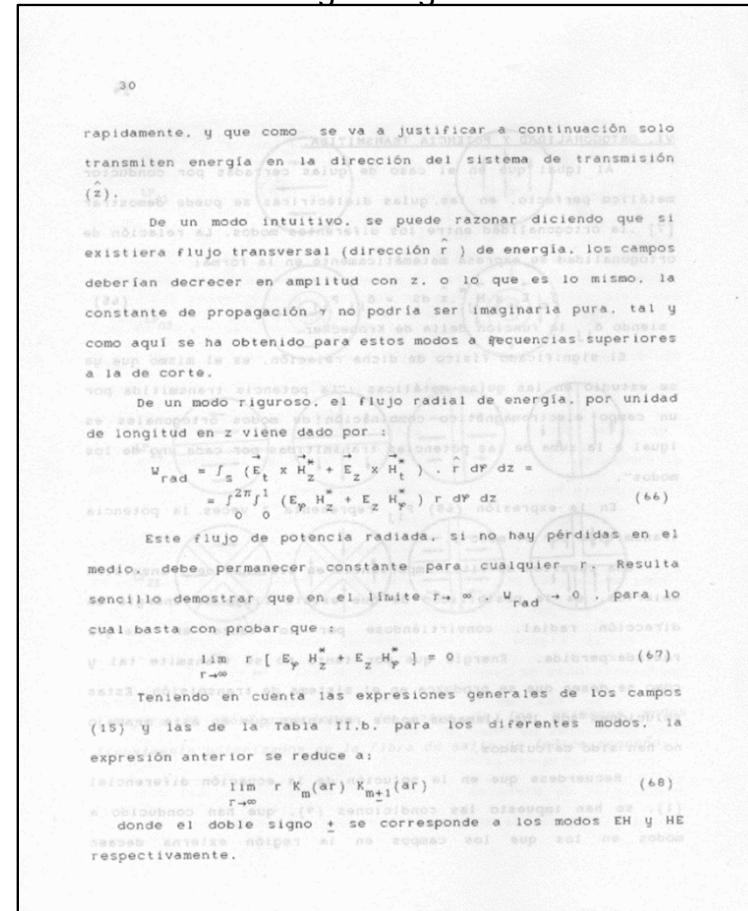
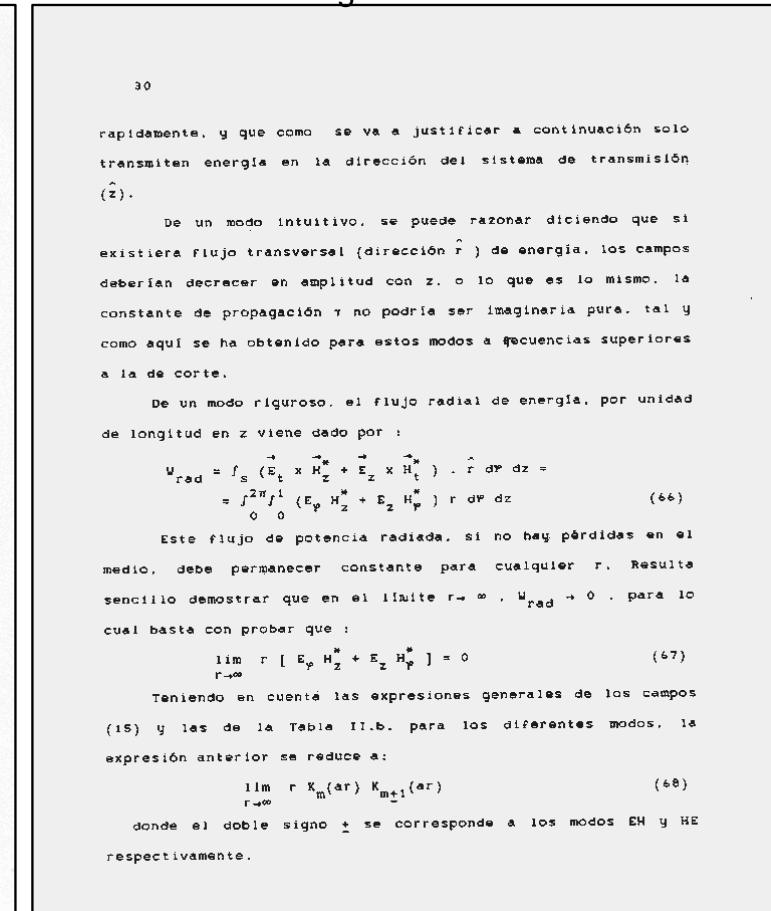


Imagen umbralizada



Operadores puntuales: preprocessamiento

- **Promediado de imágenes: eliminar ruido incorrelado**

Imagen original

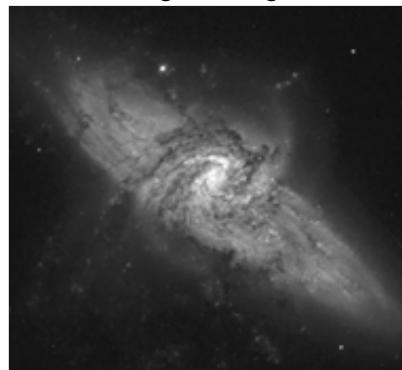
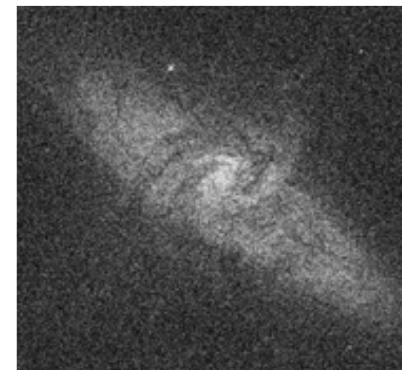
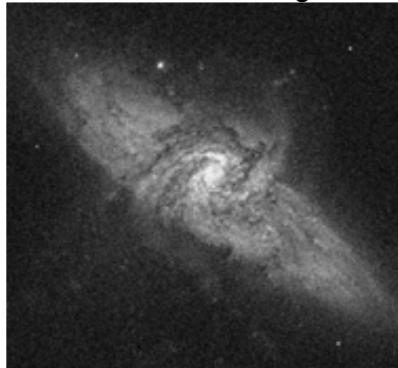


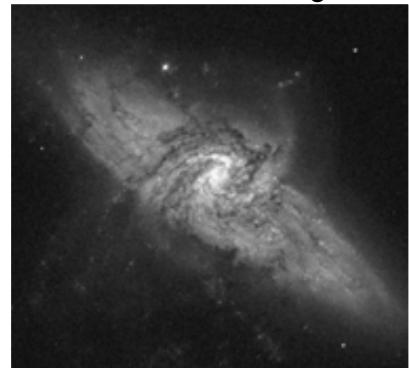
Imagen original con ruido gaussiano



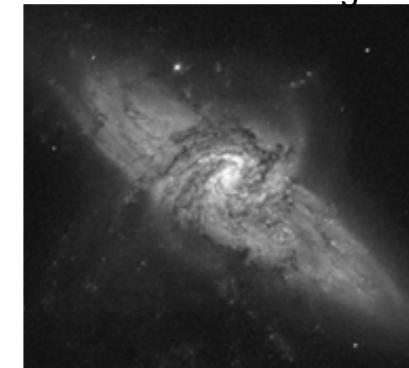
Promediado 8 imágenes



Promediado 64 imágenes



Promediado 128 imágenes



Fuente: Gonzalez, Digital Image Processing, 2002

Indexación, búsqueda y análisis
en repositorios multimedia

Operadores locales

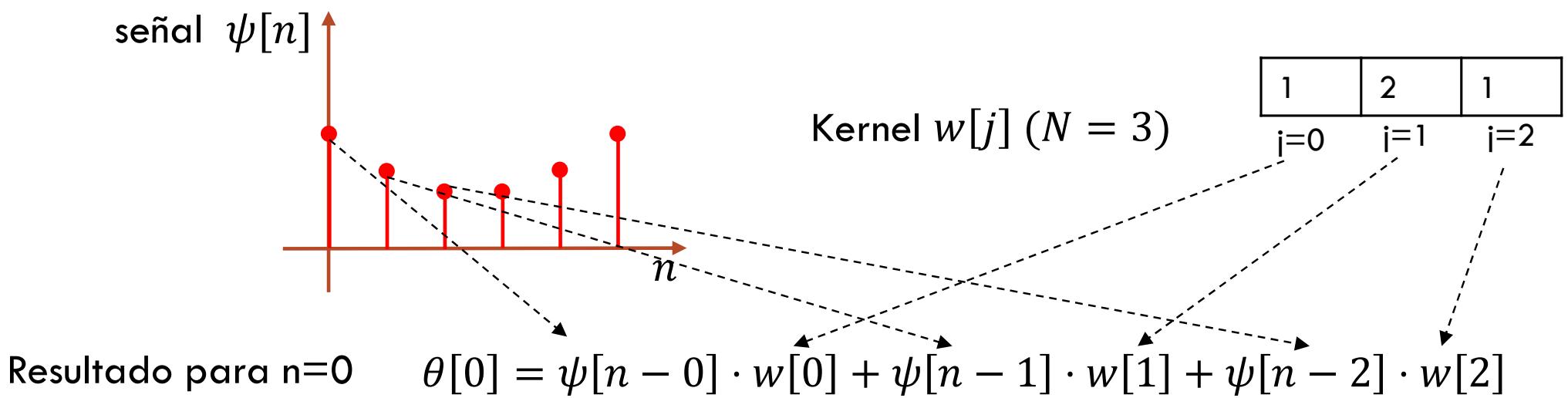
- Considera un vecindario del pixel donde se aplica (filtrado espacial)
 - Uso intensivo en tratamiento de imagen/vídeo
 - Deep learning
 - Morfología matemática
 - ...
 - Aplicaciones
 - Eliminación de ruido/suavizado
 - Transformaciones geométricas
 - Extracción de características
 - ...
-
- The diagram shows a 5x9 grid of numerical values representing an image. A 3x5 kernel is applied to the center pixel (value 113). The kernel is highlighted with a black border, and its center pixel is shaded gray. Arrows point from the labels 'Pixel de aplicación' and 'Vecindario' to their respective elements. The grid values are:
- | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|---------------------|
| 48 | 219 | 168 | 145 | 244 | 188 | 120 | 58 |
| 49 | 218 | 87 | 94 | 133 | 35 | 17 | 148 |
| 174 | 151 | 74 | 179 | 224 | 3 | 2 | Pixel de aplicación |
| 77 | 127 | 87 | 139 | 44 | 228 | 1 | aplicación |
| 138 | 229 | 136 | 113 | 250 | 51 | 108 | 163 |
| 38 | 210 | 185 | 177 | 69 | 76 | 131 | 53 |
| 178 | 164 | 79 | 158 | 64 | 169 | 8 | Vecindario |
| 96 | 209 | 214 | 203 | 223 | 73 | 11 | 3x5 |

Operadores locales: filtrado espacial

➤ Filtrado lineal 1D

- Núcleo (kernel) w define la operación a realizar
- Combinación lineal de valores en el vecindario
- Resultado señal filtrada $\theta[n]$

$$\theta[n] = \sum_{j=0}^{N-1} \psi[n-j] \cdot w[j]$$



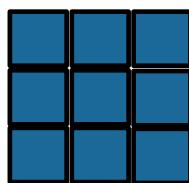
Operadores locales: filtrado espacial

➤ Filtrado lineal / no lineal (2D)

$$\psi[n, m] \xrightarrow{T} \theta[n, m] /$$

$$\theta[n, m] = T(\text{entorno de } \psi[n, m])$$

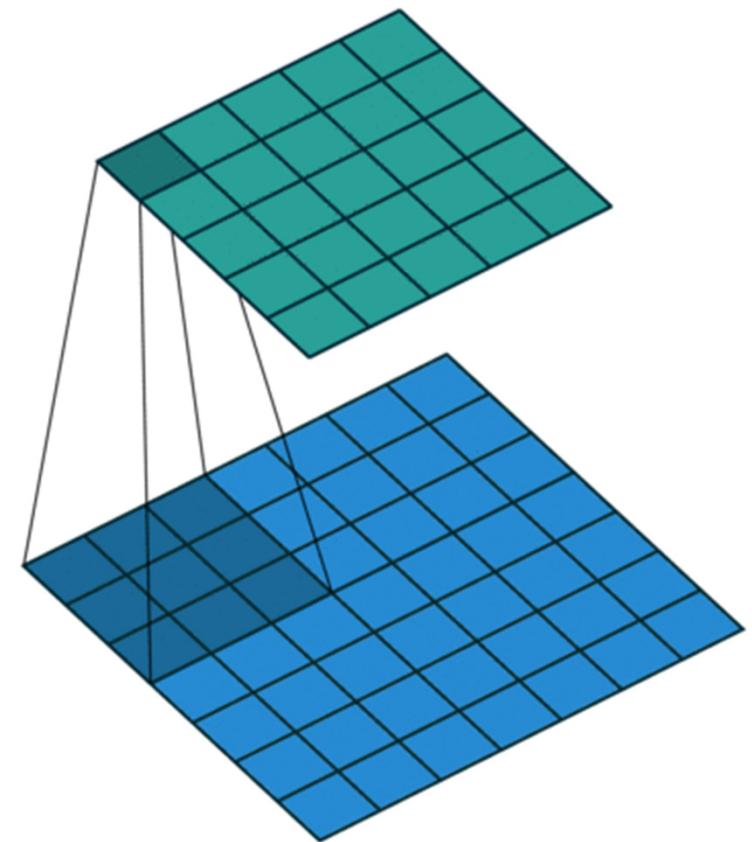
- Kernel definido por el operador T
- Aplicación mediante desplazamiento del kernel sobre la señal original



Datos utilizados
por el kernel



Resultado



Operadores locales: filtrado espacial

- Kernels típicos para eliminación ruido

Lineal: media 3x3

$$S_n = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$



Imagen original



Resultado

No-lineal: Mediana 5x5

A 5x5 matrix representing a median filter kernel. The central element is 1, and all other elements are 0. A black border surrounds the matrix.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Filtro 5x5



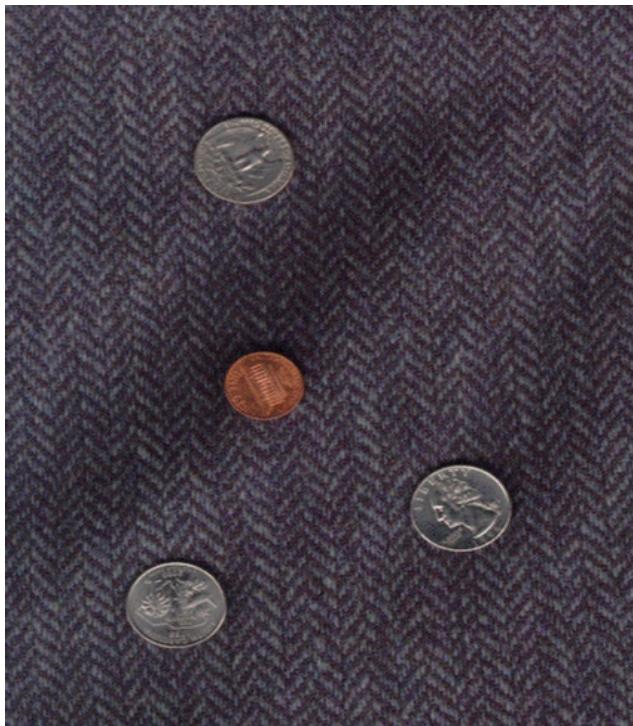
Imagen original



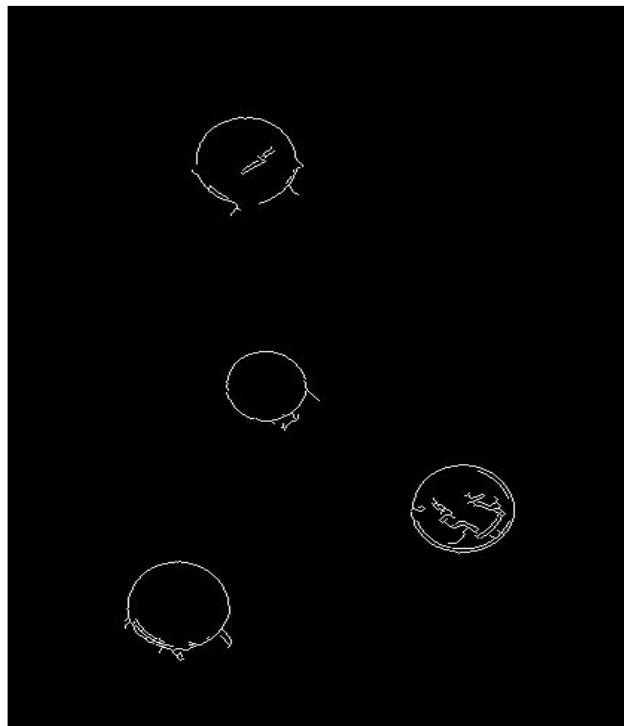
Resultado

Operadores globales: transformada Hough

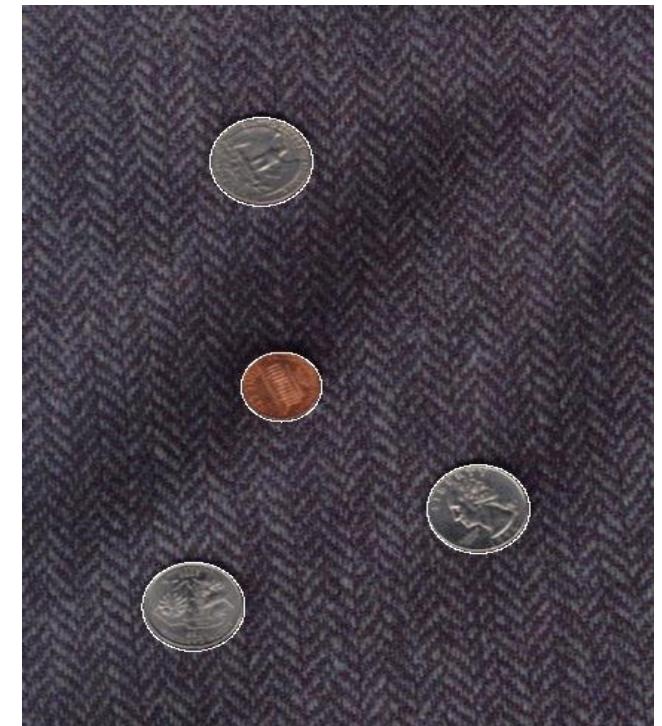
- Detección de formas en imágenes (entrenamiento previo)



Imagen



Detector bordes Canny



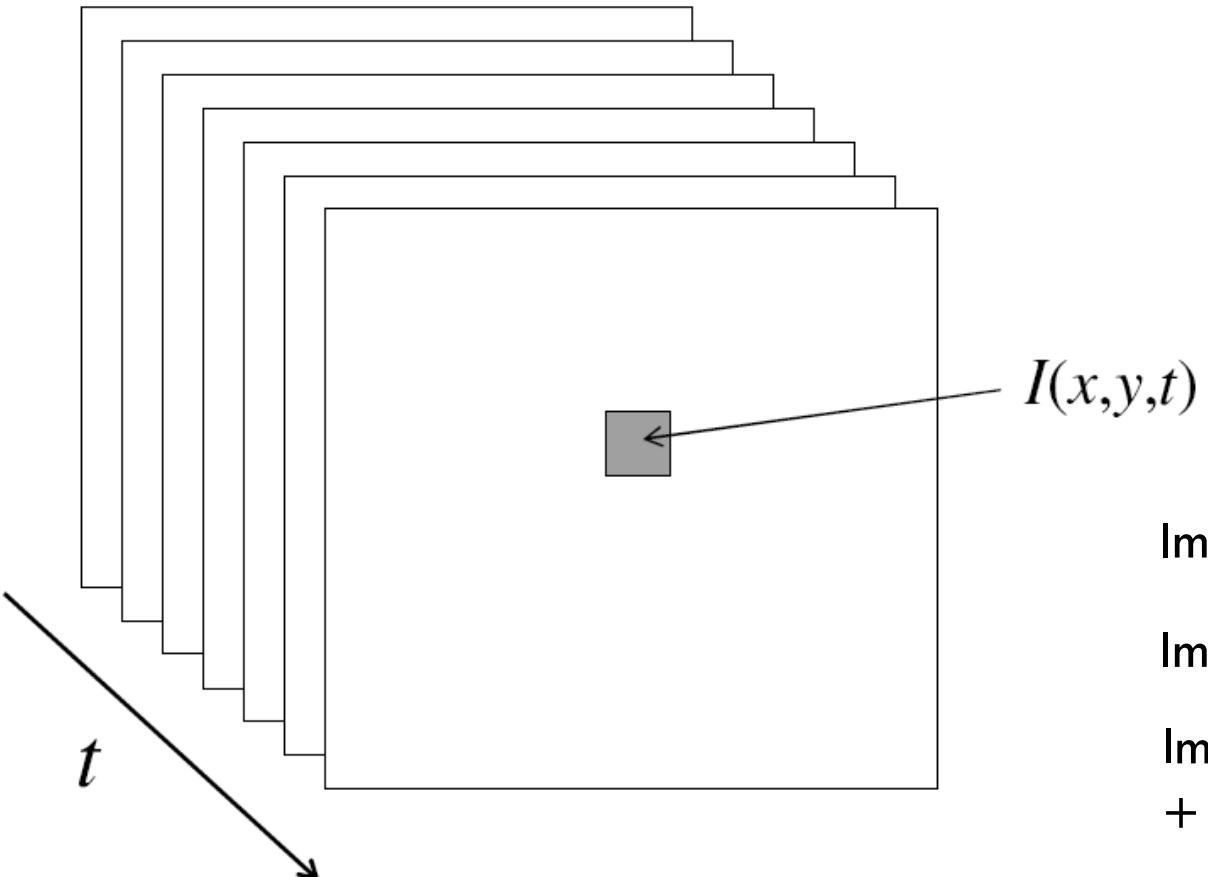
Círculos mediante
Transformada de Hough

Agenda: Multimedia (imagen, video)

Aspectos generales

- Multimedia
- Relación con Big Data
- Información de interés
- Áreas de conocimiento
- Principales problemas
- Ejemplos
- Imágenes
 - Formación y representación
 - Tratamiento de imágenes
 - Componentes
 - Operadores puntuales, locales y globales
- Video
 - Estructura espacio-temporal
 - Tratamiento de video
 - Segmentación de objetos
 - Seguimiento de objetos
 - Reconocimiento de actividades

Video: señal multidimensional



Imágenes gris (2D) + Tiempo → 3D

Imágenes color (3D) + Tiempo → ¿4D?

Imágenes color (3D) + Profundidad (1D)
+ Tiempo → ¿5D?

Tratamiento de video: estructura espacio-temporal



Tratamiento de video: estructura espacio-temporal



Video

Escenas



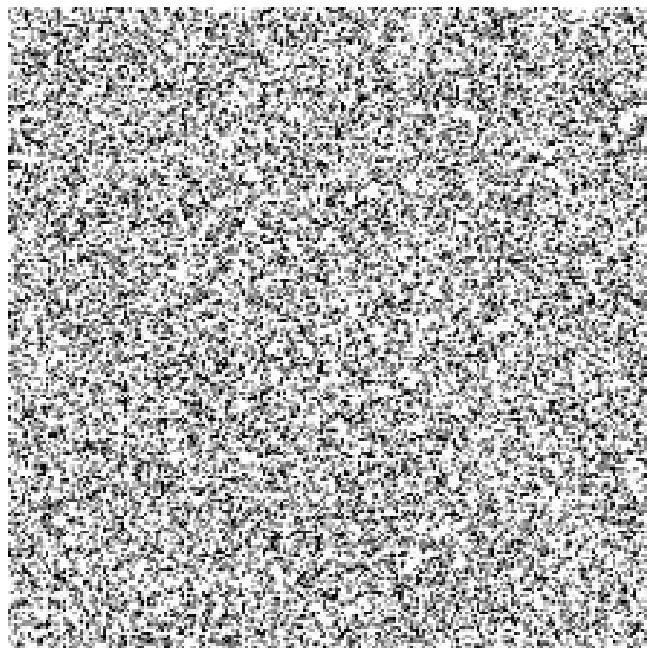
- Varios objetos
- Cámara estática

- Un objeto
- Cámara móvil

- Un objeto
- Cámara vibrando (jitter)

Tratamiento de video: estimación de movimiento

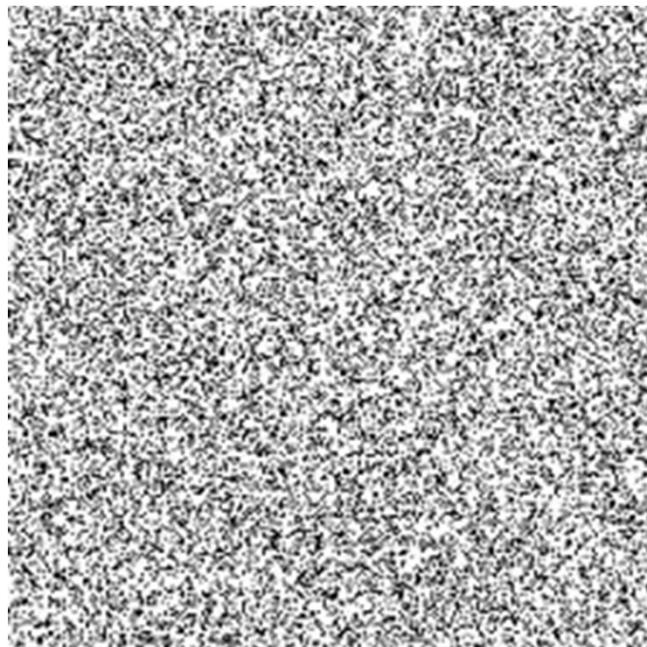
- La importancia del movimiento y organización perceptual



¿Se observa algún objeto en la imagen?

Tratamiento de video: estimación de movimiento

- La importancia del movimiento y organización perceptual
 - Si añadimos movimiento...



En ciertos casos,
el movimiento es la
característica más fiable...

<https://youtu.be/aEoxORdGhE>

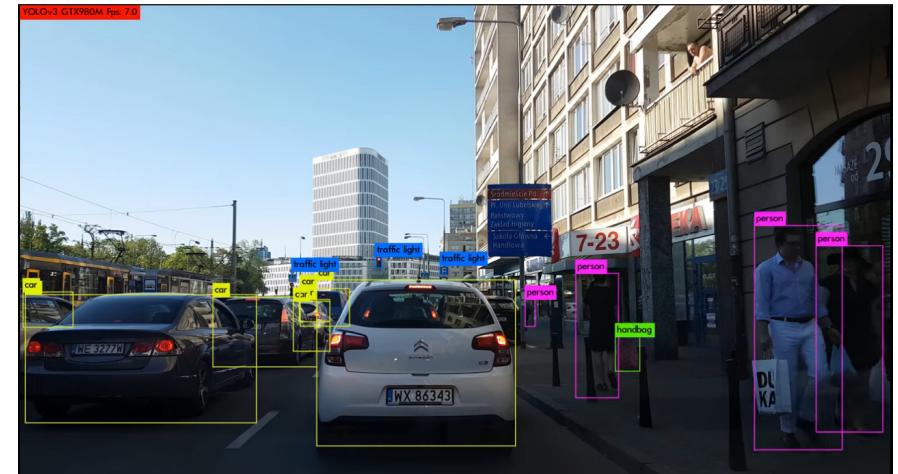
Tratamiento de video: segmentación de objetos

Objetos de interés: **segmentación frente-fondo**



Tratamiento de video: segmentación de objetos

- Con cámaras móviles → clasificación directa en imagen

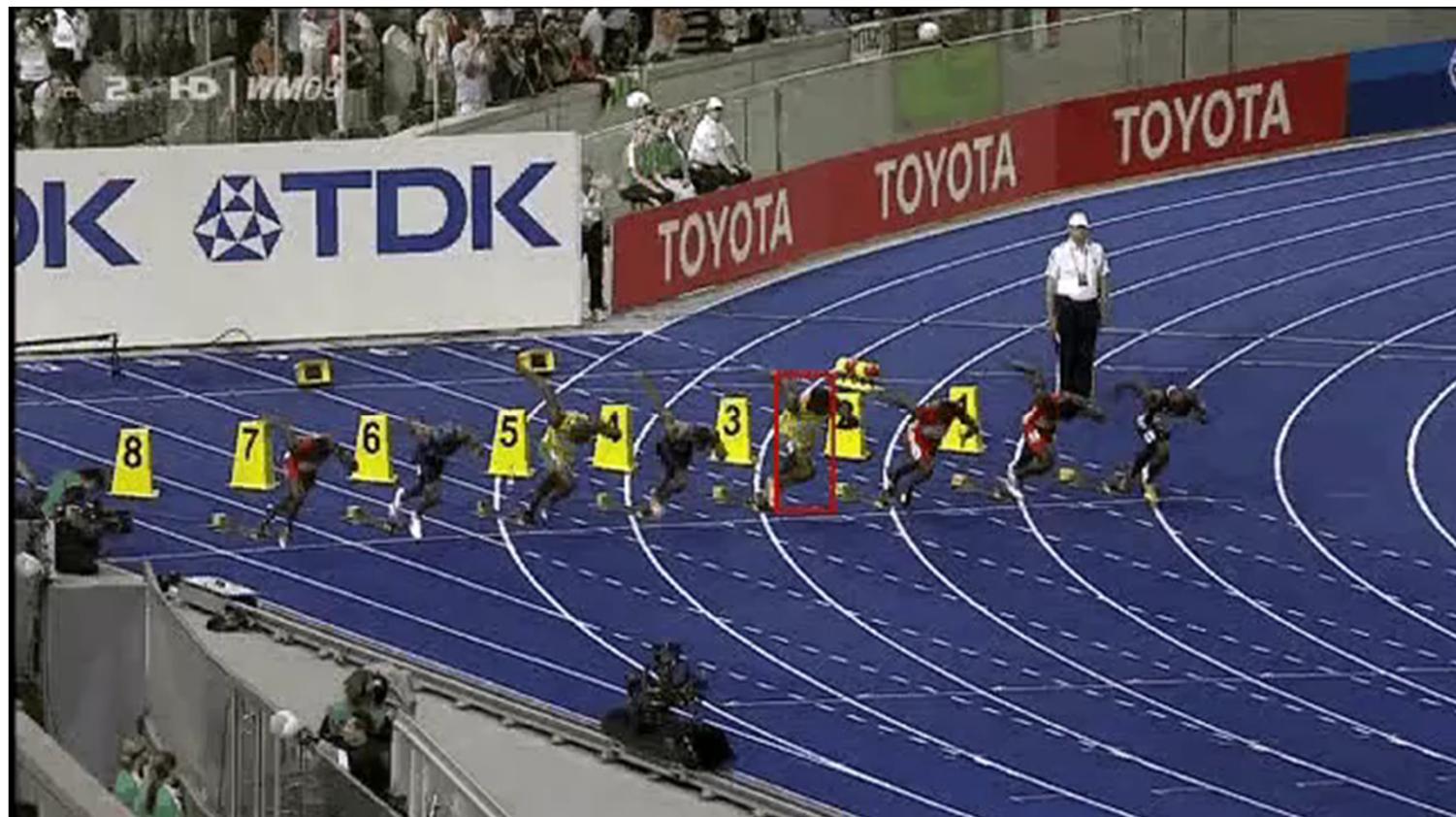


https://youtu.be/8jfscFuP_9k

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. IEEE CVPR 2016 pp. 779-788.

Tratamiento de vídeo: seguimiento de objetos

- Objetivo: localización espacio-temporal de un objeto (*iniciado manualmente*)



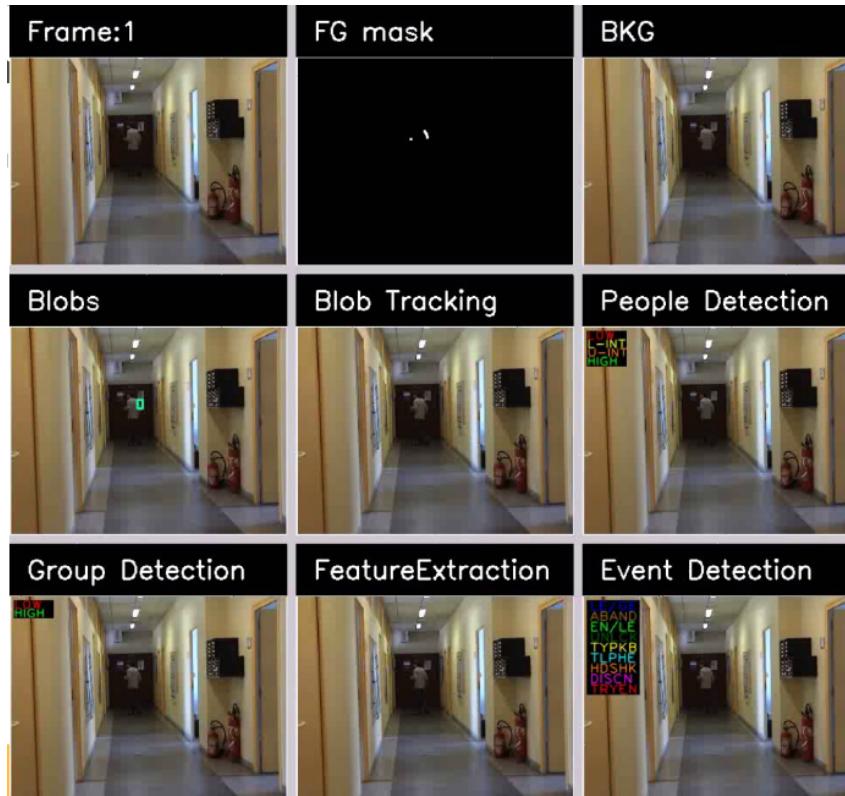
Tratamiento de vídeo: seguimiento de objetos

- Objetivo: localización espacio-temporal de un objeto (*iniciado automáticamente*)



Tratamiento de vídeo: reconocimiento actividades

- Objetivo: localización espacio-temporal de acciones o interacciones entre objetos

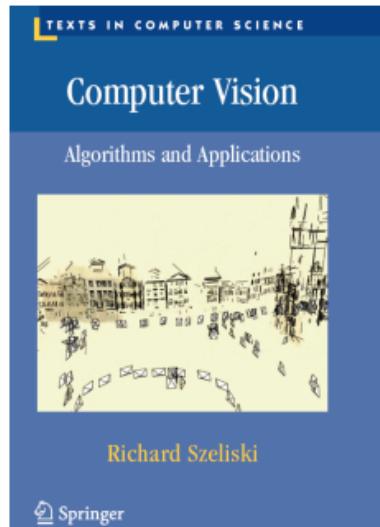


<https://youtu.be/EeNyyOnrRKE>

Bibliografía

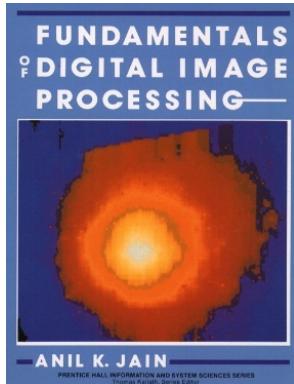
Computer Vision: Algorithms and Applications

© 2010 [Richard Szeliski](#), Microsoft Research

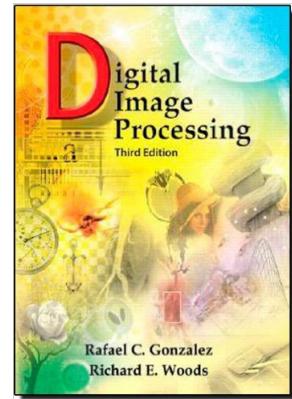


<http://szeliski.org/Book/>

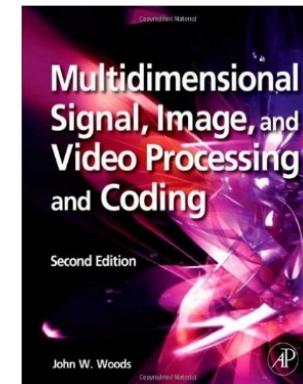
Bibliografía



A.K. Jain, Fundamentals of Digital Image Processing", Prentice Hall, 2005



R. Gonzalez, R. Woods,
"Digital Image Processing",
2^a Ed, Prentice Hall, 2002.



J. Woods, "Multidimensional signal, image & video processing", Ass. Press , 2006

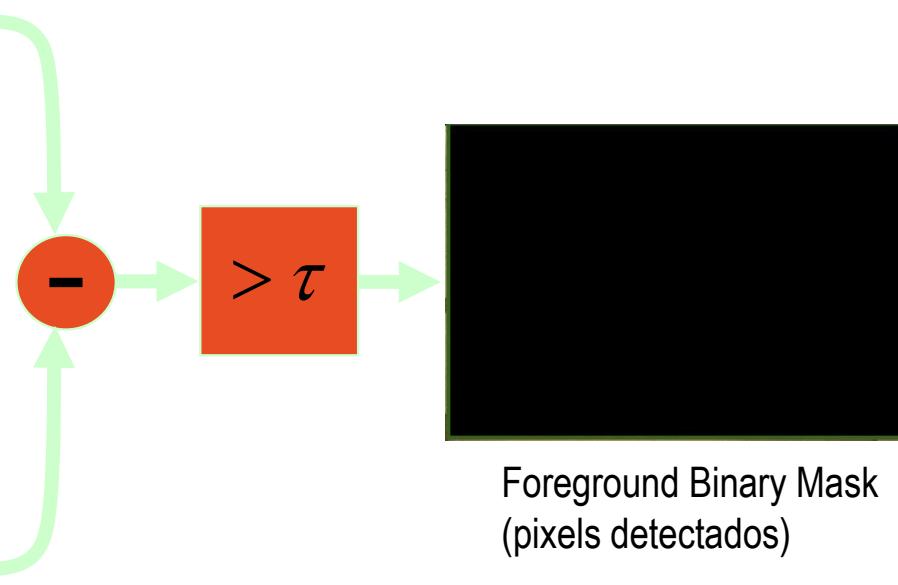
Tratamiento de video: segmentación de objetos

- Con cámaras estáticas...
 - Sustracción de fondo (*background subtraction*)

Imagen de la secuencia



Modelo de Fondo de escena estático

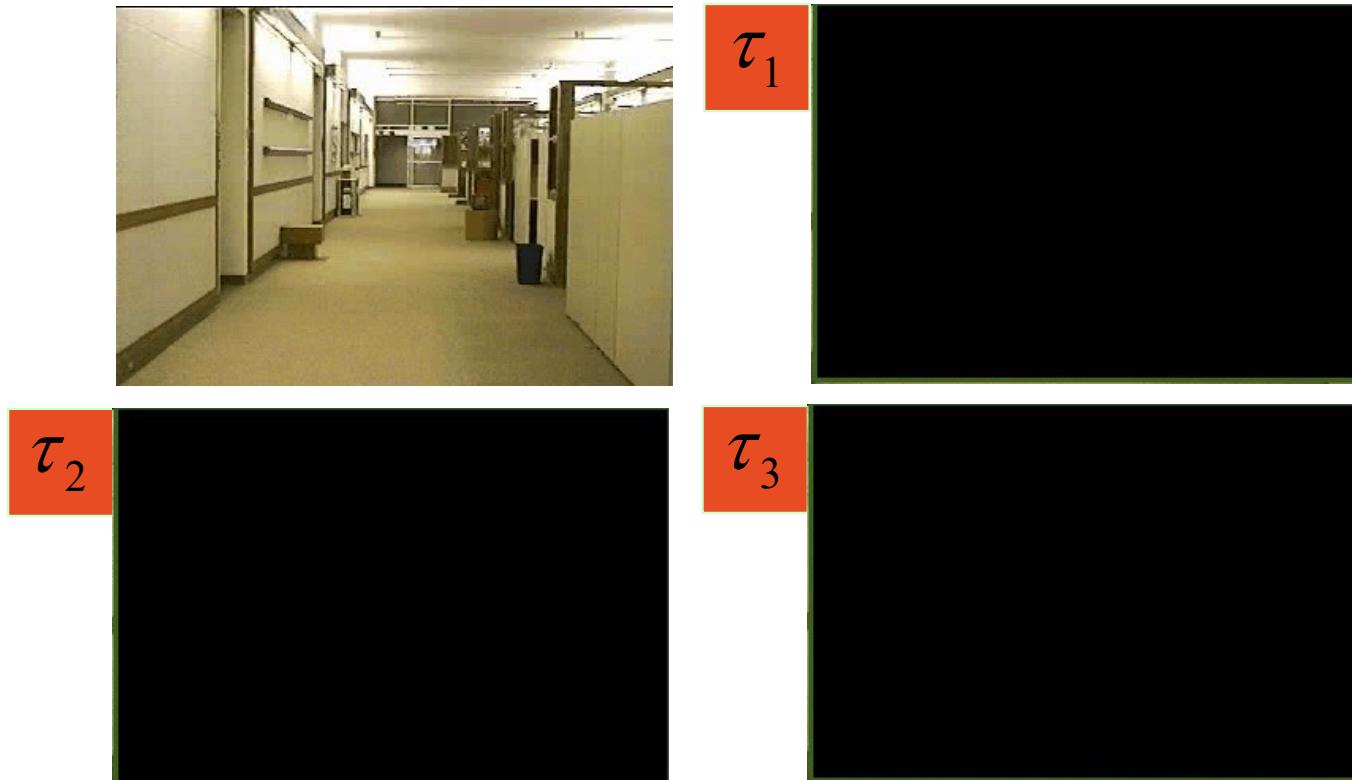


Tratamiento de video: segmentación de objetos

- Con cámaras estáticas...

- Sustracción de fondo (*background subtraction*)

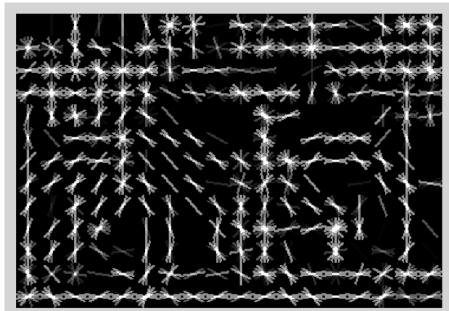
$$\tau_1 > \tau_2 > \tau_3$$



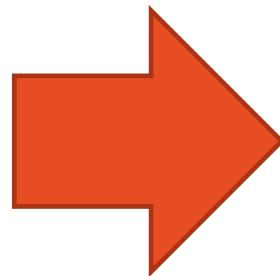
Tratamiento de video: segmentación de objetos

- Con cámaras móviles → clasificación directa en imagen
Partes

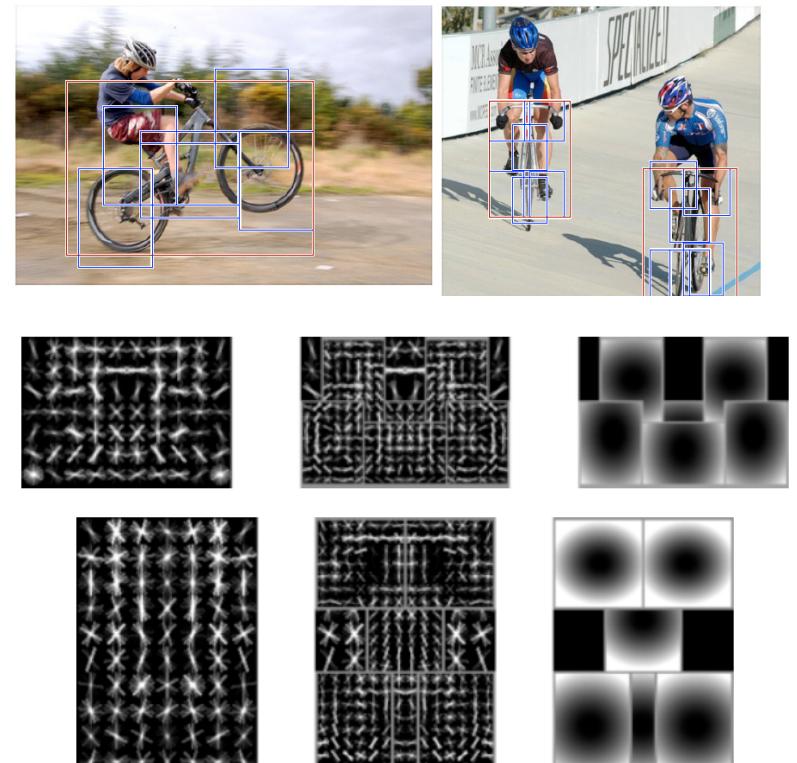
Holístico



Detecciones



Modelos



root filters
coarse resolution

part filters
finer resolution

deformation
models