

Máster Universitario en
Nuevas Tecnologías en Informática

Asignatura “Visión Artificial”

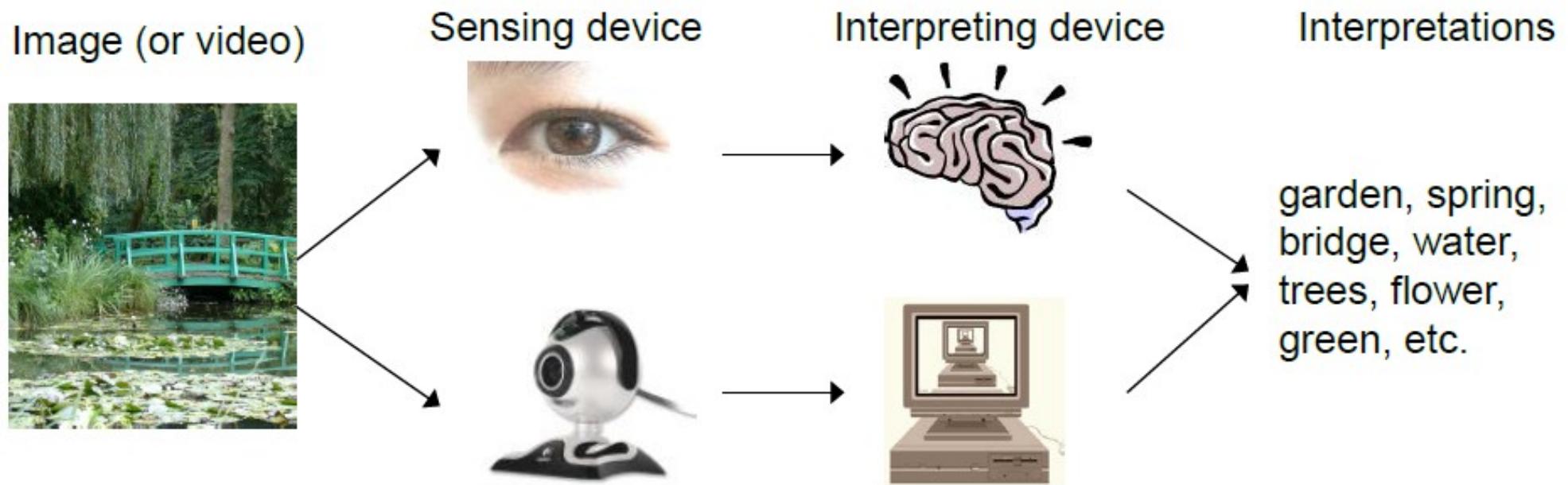
Introducción a la visión por computador

Facultad de Informática
Universidad de Murcia
Curso 2018/19

Introducción

- Introducción a la visión por computador
- Visión general del curso
 - Profesorado
 - Contenidos teóricos
 - Contenidos prácticos
 - Evaluación
 - Bibliografía

¿Qué es la visión por computador?



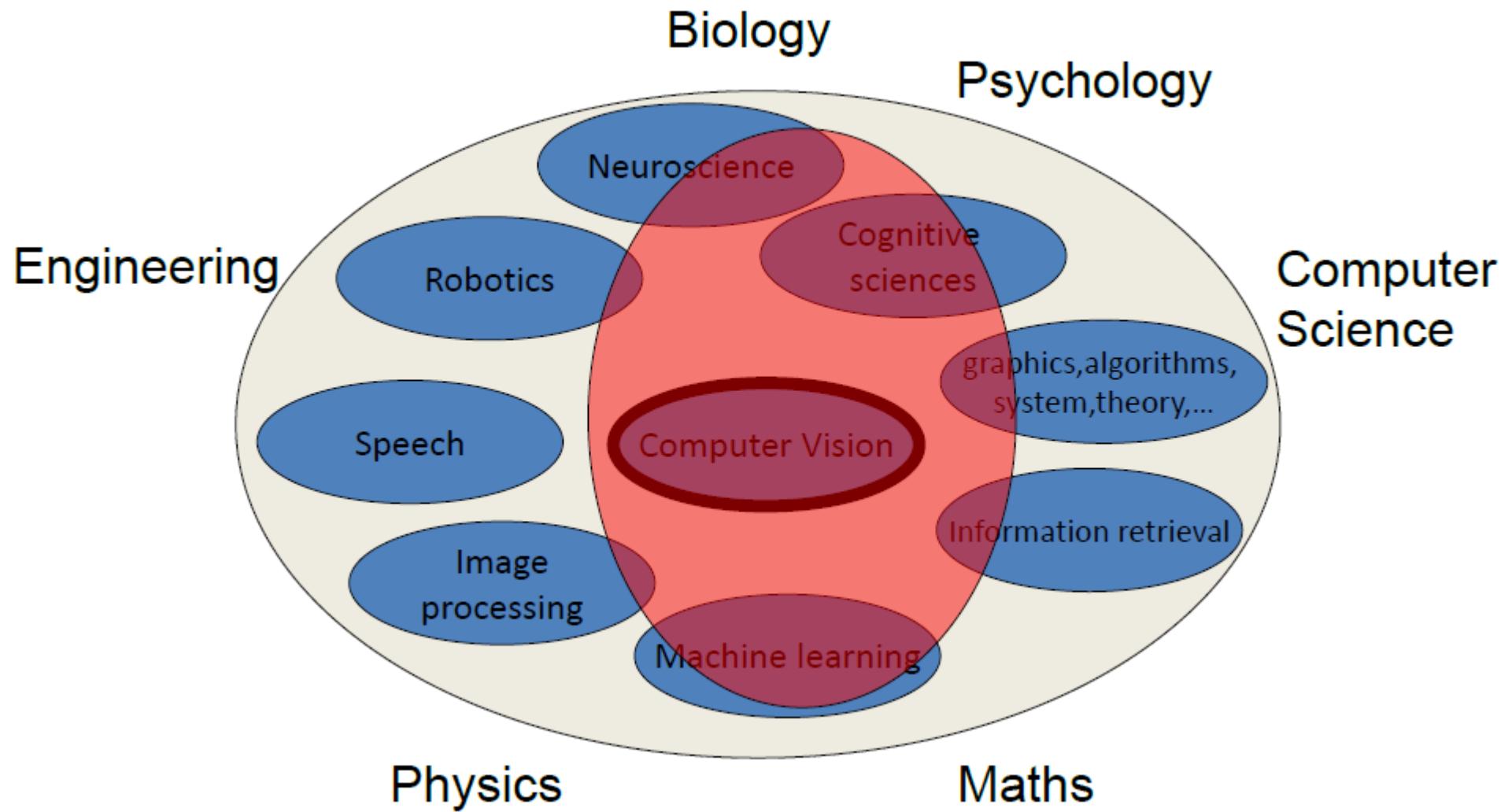
¿Qué es la visión por computador?

- Interpretación de una imagen / vídeo por parte de un computador:



- ¿Qué tipo de escena es?
- ¿Dónde están los coches?
- ¿A qué distancia está el edificio?

¿Con qué está relacionada?



El objetivo de la visión por computador

- Conseguir realizar el salto desde los “píxeles” hasta el “significado”



0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

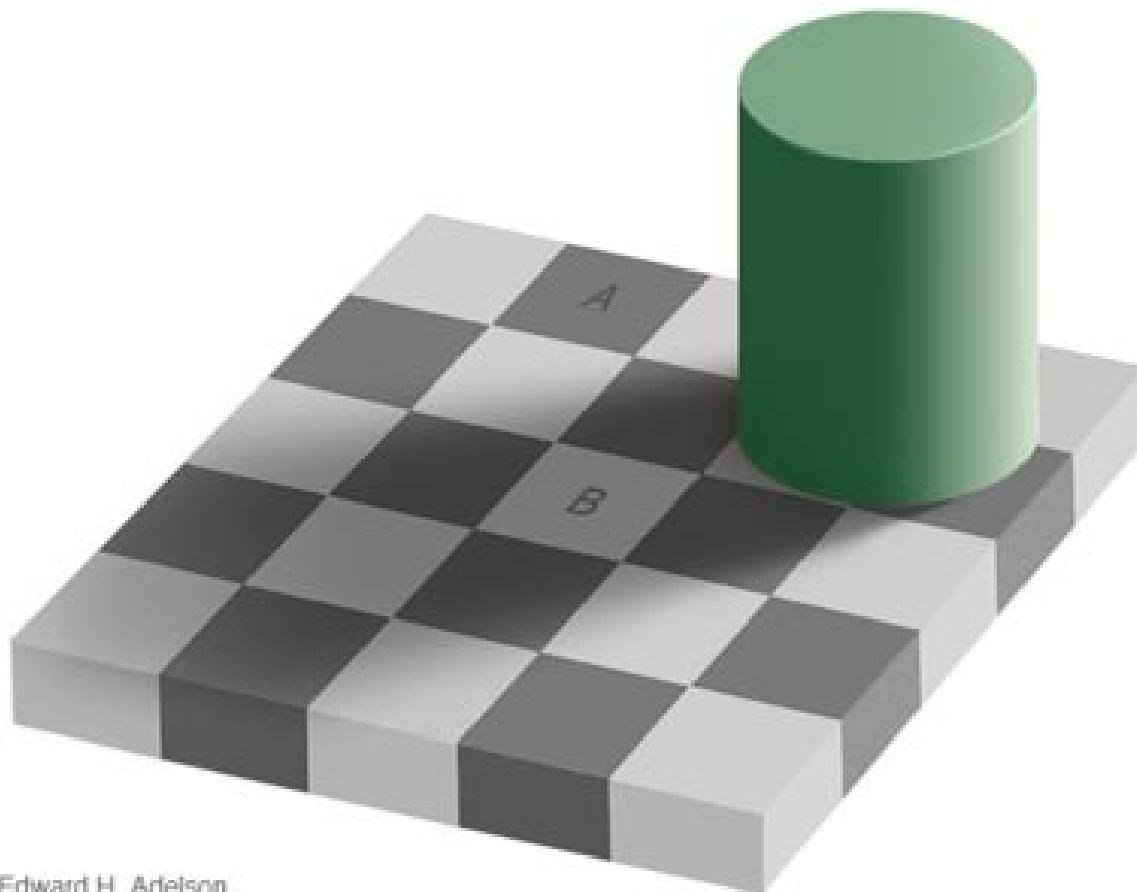
Segmentación



Reconocimiento (clasificación)

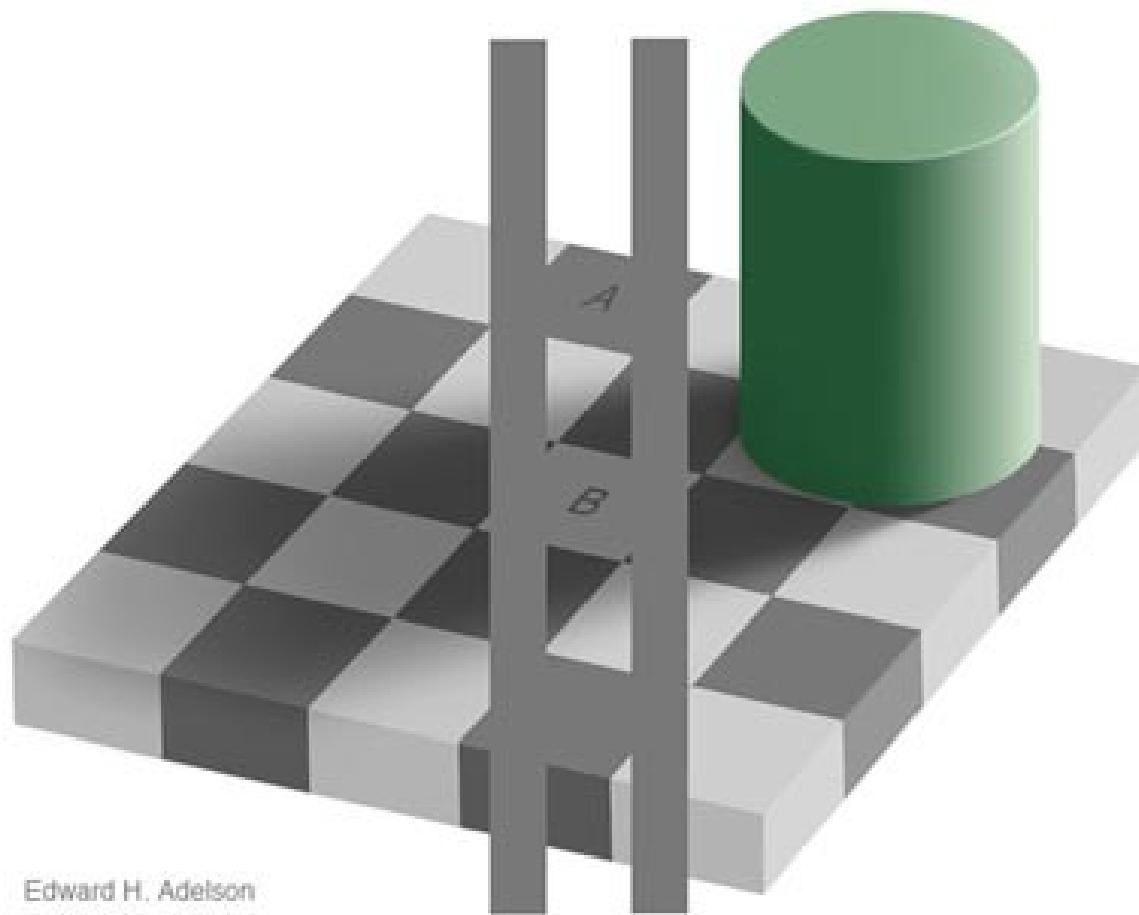


Dificultades...



Edward H. Adelson

Dificultades...



Edward H. Adelson

Reconocimiento de “categorías”



Origins of computer vision: an MIT undergraduate summer project

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

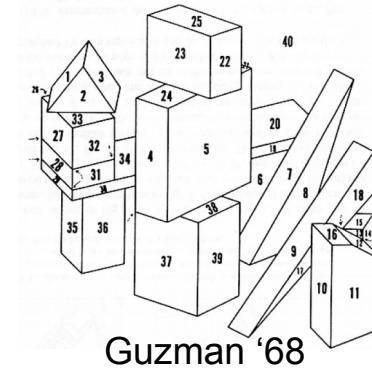
THE SUMMER VISION PROJECT

Seymour Papert

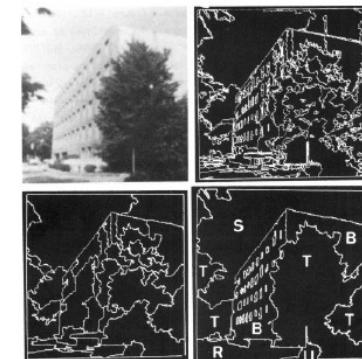
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Historia (ridículamente breve) de la visión por computador

- 1966: Minsky asigna la visión como “proyecto de verano”
- 1960's: Interpretación de mundos sintéticos
- 1970's: Algunos progresos en dominios restringidos
- 1980's: Primeras redes neuronales. Desplazamiento hacia la geometría y el rigor matemático
- 1990's: Reconocimiento facial, auge métodos estadísticos
- 2000's: Dominios más complejos; grandes *datasets* disponibles; procesamiento de vídeo; geometría visual muy madura
- 2010's: Deep Learning



Guzman '68



Ohta Kanade '78

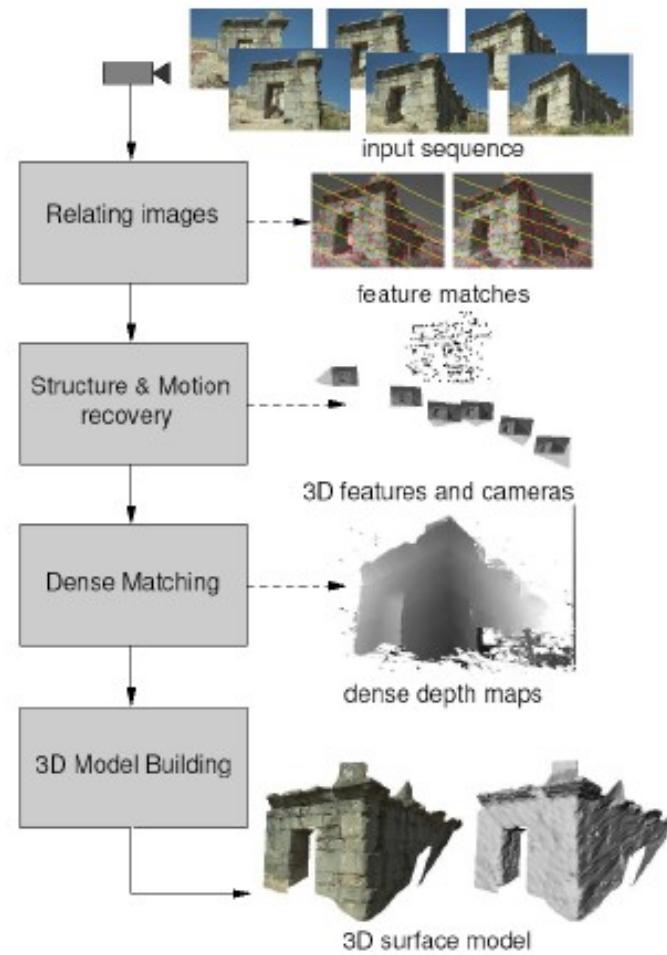


Turk and Pentland '91

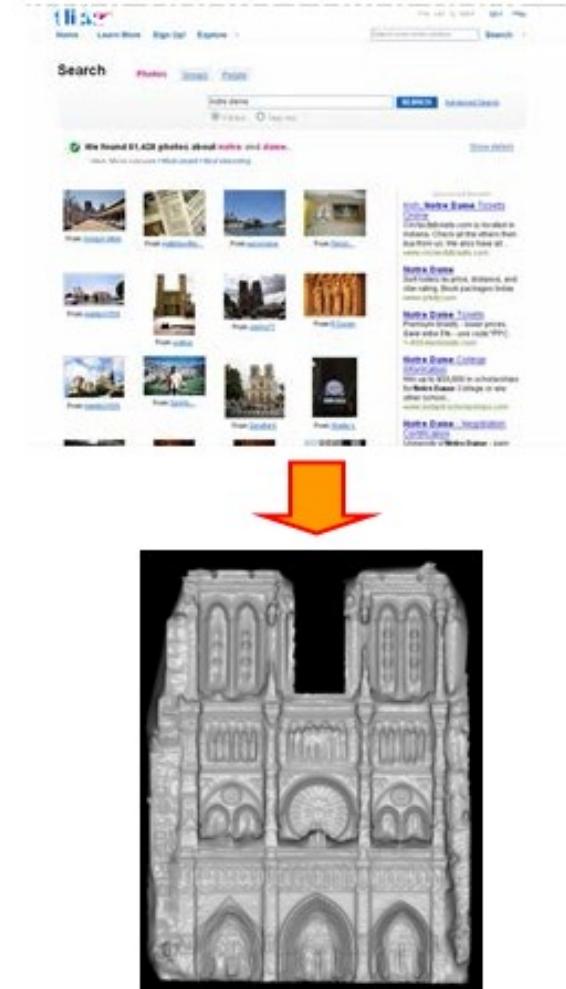
¿Qué tipo de información extraer de la(s) imagen(es)?

- Información métrica:
 - 2D
 - 3D
 - ...
- Información semántica:
 - Reconocimiento de objetos
 - Reconocimiento de categorías de objetos
 - ...

La visión como dispositivo de medida

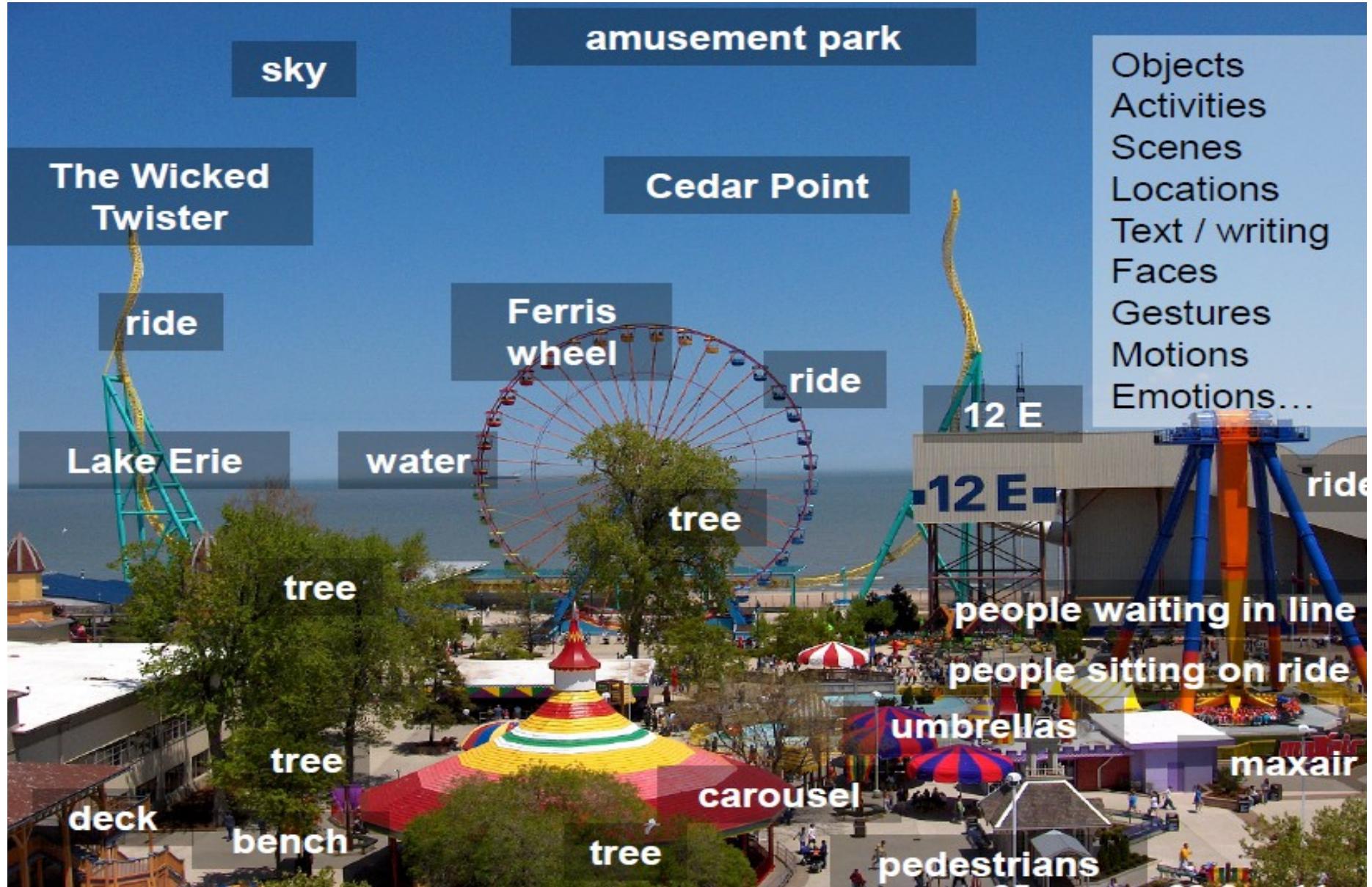


Pollefeys et al.



Goesele et al.

La visión como fuente de información semántica



La visión como fuente de información semántica

The screenshot shows the Clarifai website interface. At the top, there is a navigation bar with links: ABOUT, TECHNOLOGY, API, NEWS, JOBS, and CONTACT. The Clarifai logo is circled in red. Below the navigation bar, the text "Recognize, Localize & Search by Concept" is displayed, followed by "Classify, Localize & Search by Similarity". A descriptive text states: "Our algorithms enable fine grained classification, localization of objects in images and similarity search based on the combination of semantic and visual properties." The main content area displays five image cards, each with a caption and a set of semantic tags:

- Coffee and Croissant:** coffee, croissant, beverage, morning, breakfast, food.
- Brooklyn Bridge at Night:** night, bridge, city, suspension bridge, river.
- Siberian Husky in Snow:** winter, snow, cold, mammal, dog, arctic.
- Music Performance:** (no explicit tags shown)
- Train Station Platform:** (no explicit tags shown)

Below the image cards, there is a large photograph of a canal in Amsterdam with buildings on either side.

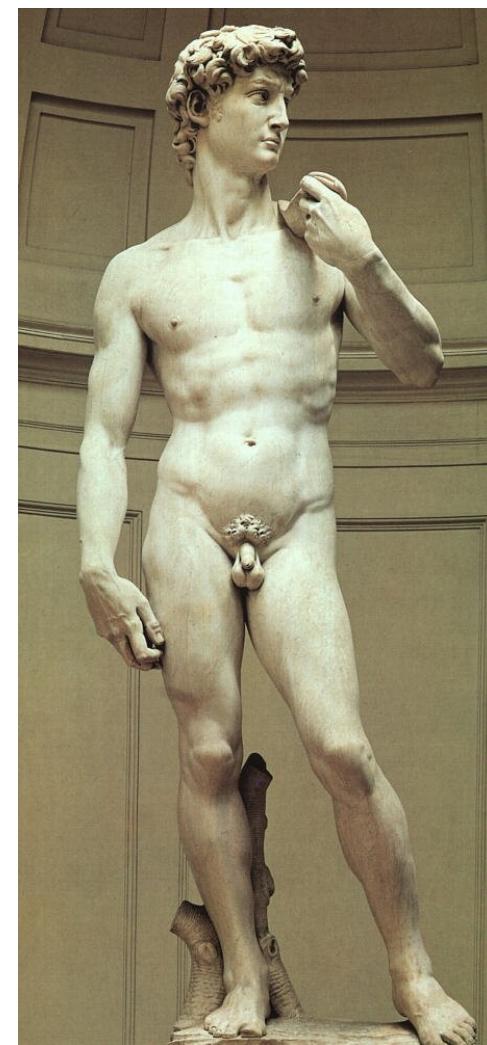
<https://www.clarifai.com/demo>

¿Por qué estudiar Visión por Computador?

- La visión es útil.
- La visión es interesante.
- La visión es difícil.
 - ¡La mitad del córtex de los primates está dedicada al procesamiento visual!
 - Alcanzar una percepción visual al nivel de un humano es un problema AI-completo.

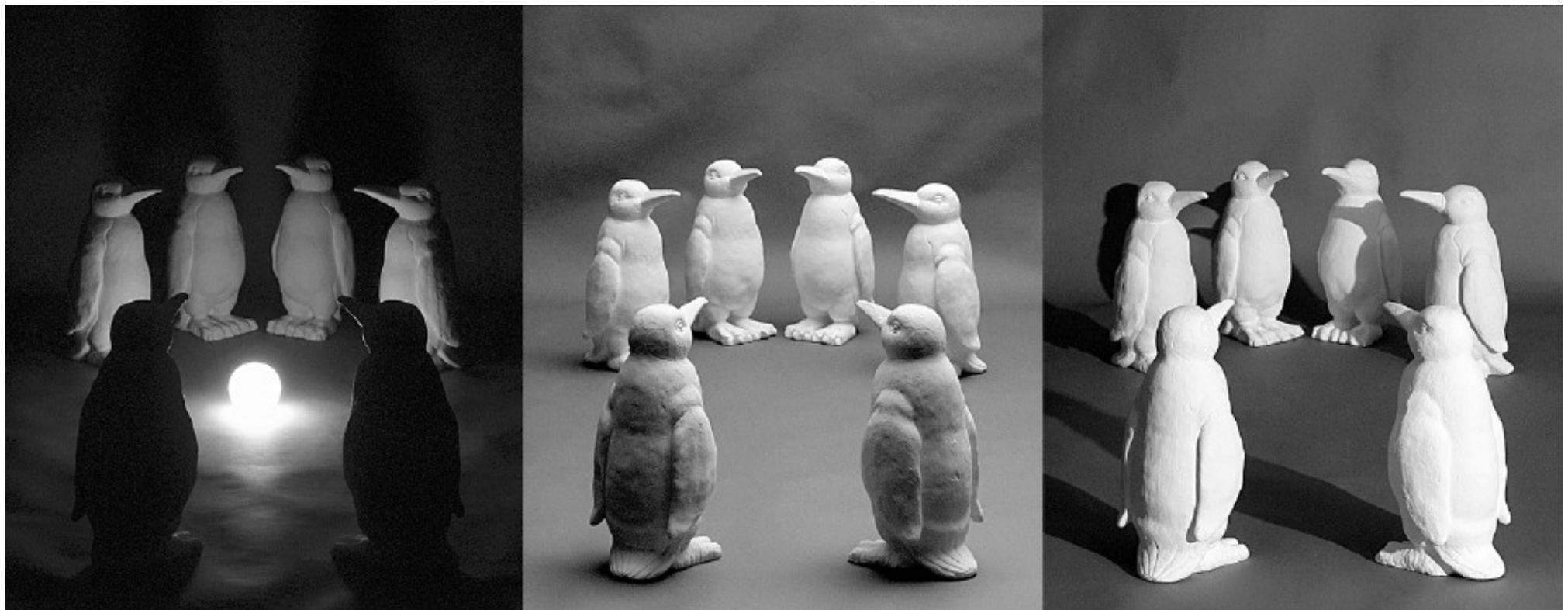
¿Por qué es difícil?

- Desafíos: cambios de punto de vista, color, ...



¿Por qué es difícil?

- Desafíos: cambios de iluminación



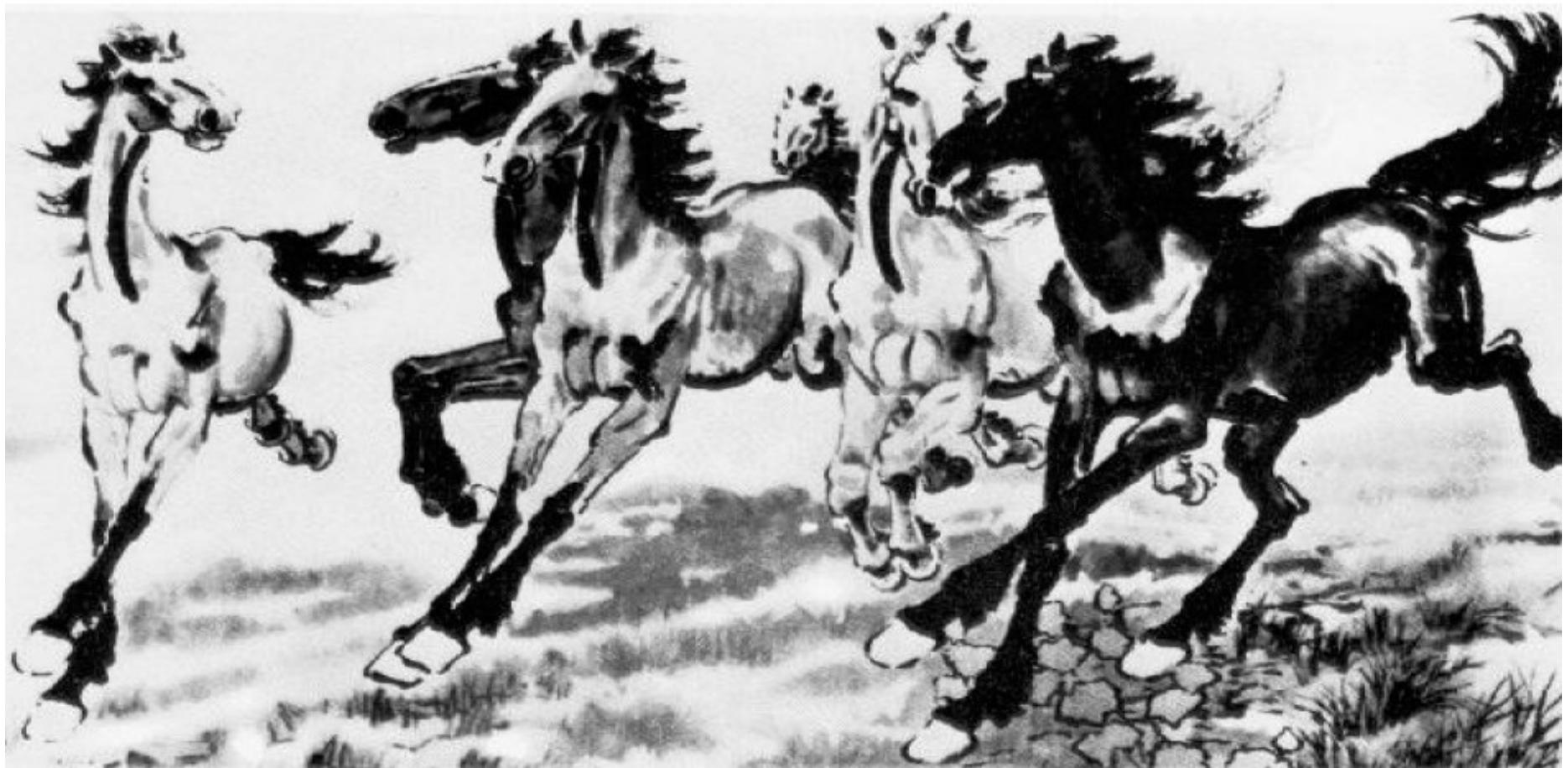
¿Por qué es difícil?

- Desafíos: cambios de escala



¿Por qué es difícil?

- Desafíos: deformaciones



¿Por qué es difícil?

- Desafíos: occlusiones



¿Por qué es difícil?

- Desafíos: *background clutter*



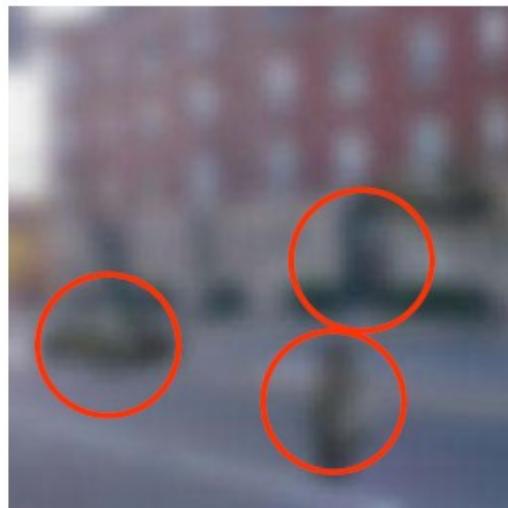
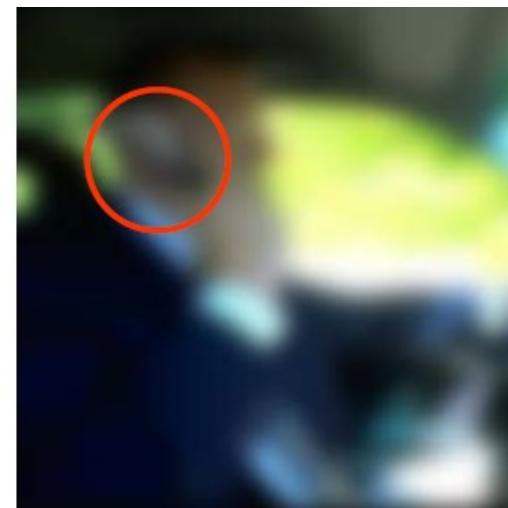
¿Por qué es difícil?

- Desafíos: variación *intraclase*



¿Por qué es difícil?

- Desafíos: ambigüedad local



¿Desafíos u oportunidades?

- Los datos “crudos” de las imágenes pueden confundir, pero...
- ... también revelan la estructura del mundo a través de numerosas pistas.
- ¡Nuestro trabajo es interpretarlas!



Ejemplo pistas: profundidad (*depth*) perspectiva lineal



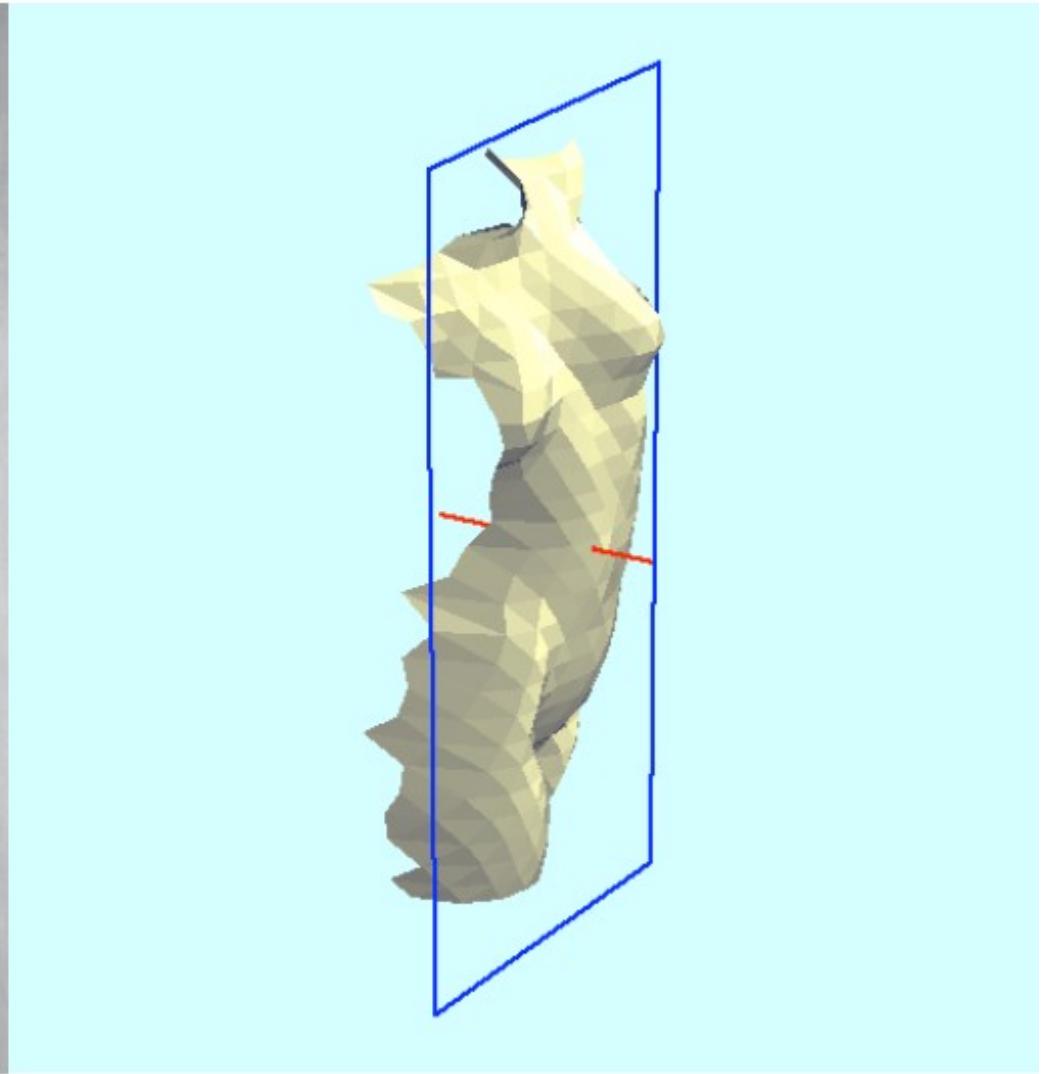
Ejemplo pistas: profundidad (*depth*) enfoque



Ejemplo pistas: profundidad (*depth*) gradiente de textura



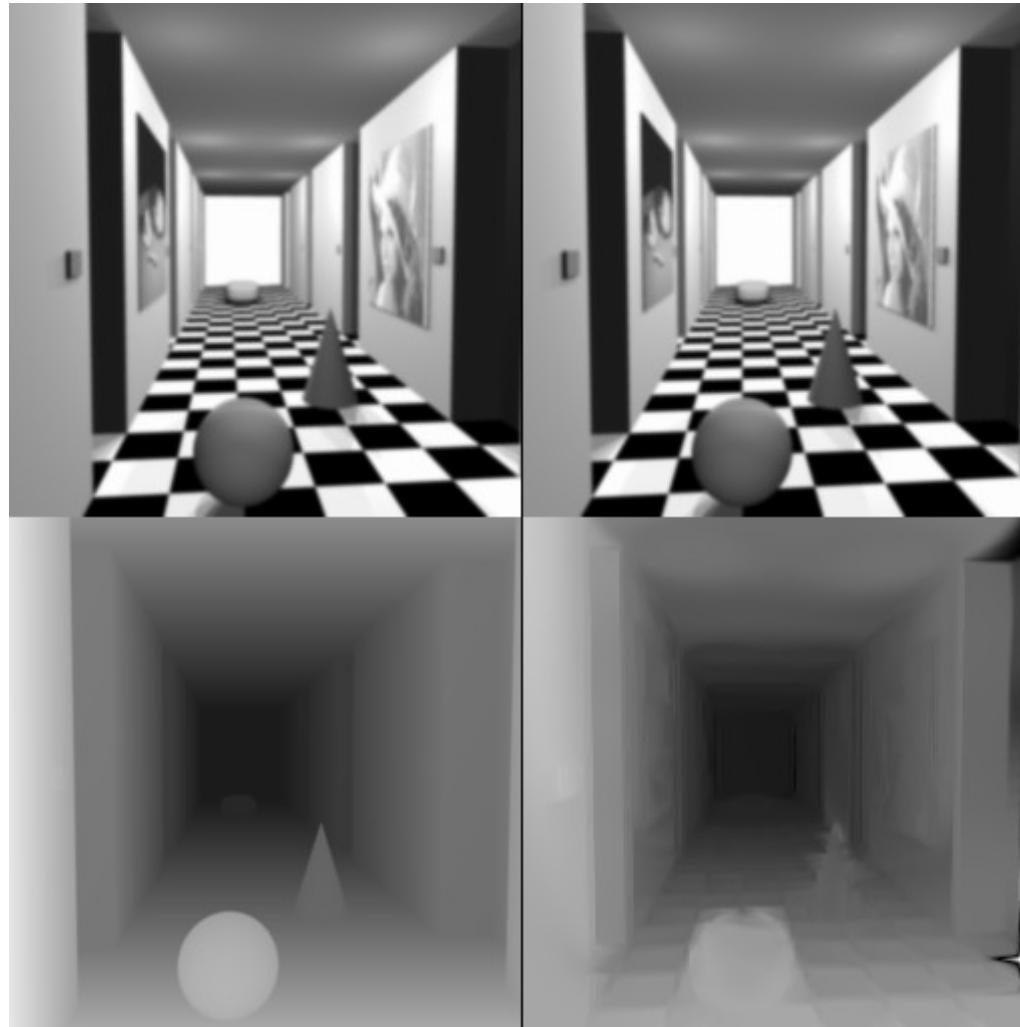
Ejemplo pistas: profundidad (*depth*) sombreado (*shading*)



Ejemplo pistas: profundidad (*depth*) pistas de interpretación (p.e. iluminación)



Ejemplo pistas: profundidad (*depth*) múltiples vistas (*multiple view geometry*)



Ejemplo pistas: agrupamiento color, textura, proximidad...



Lo más importante

- La percepción visual es inherentemente ambigua:
 - Muchas escenas diferentes podrían dar lugar a una misma imagen.



Ambigüedad

- Posibles soluciones a la ambigüedad 2D→3D:
 - Añadir más restricciones (más imágenes)
 - Usar información a priori sobre la estructura del mundo (conocimiento del dominio)
 - Combinación de ambos métodos



Aplicaciones

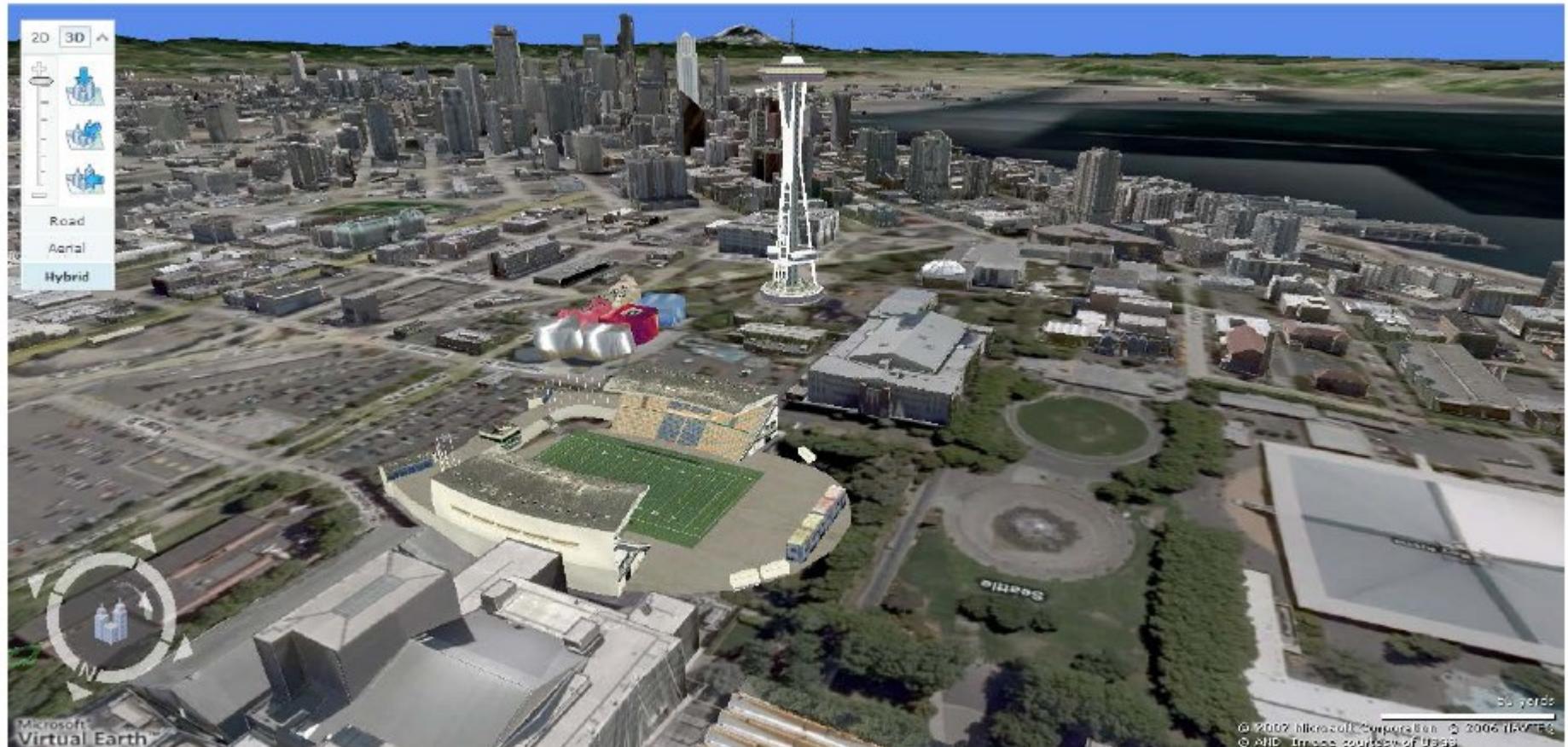
Efectos especiales (captura de forma / movimiento / realidad aumentada)



Aplicaciones

Modelado 2D/3D de ciudades

- Google Maps / street view / Bing maps



Aplicaciones

Organización de *datasets* de fotografías

- Microsoft Photosynth

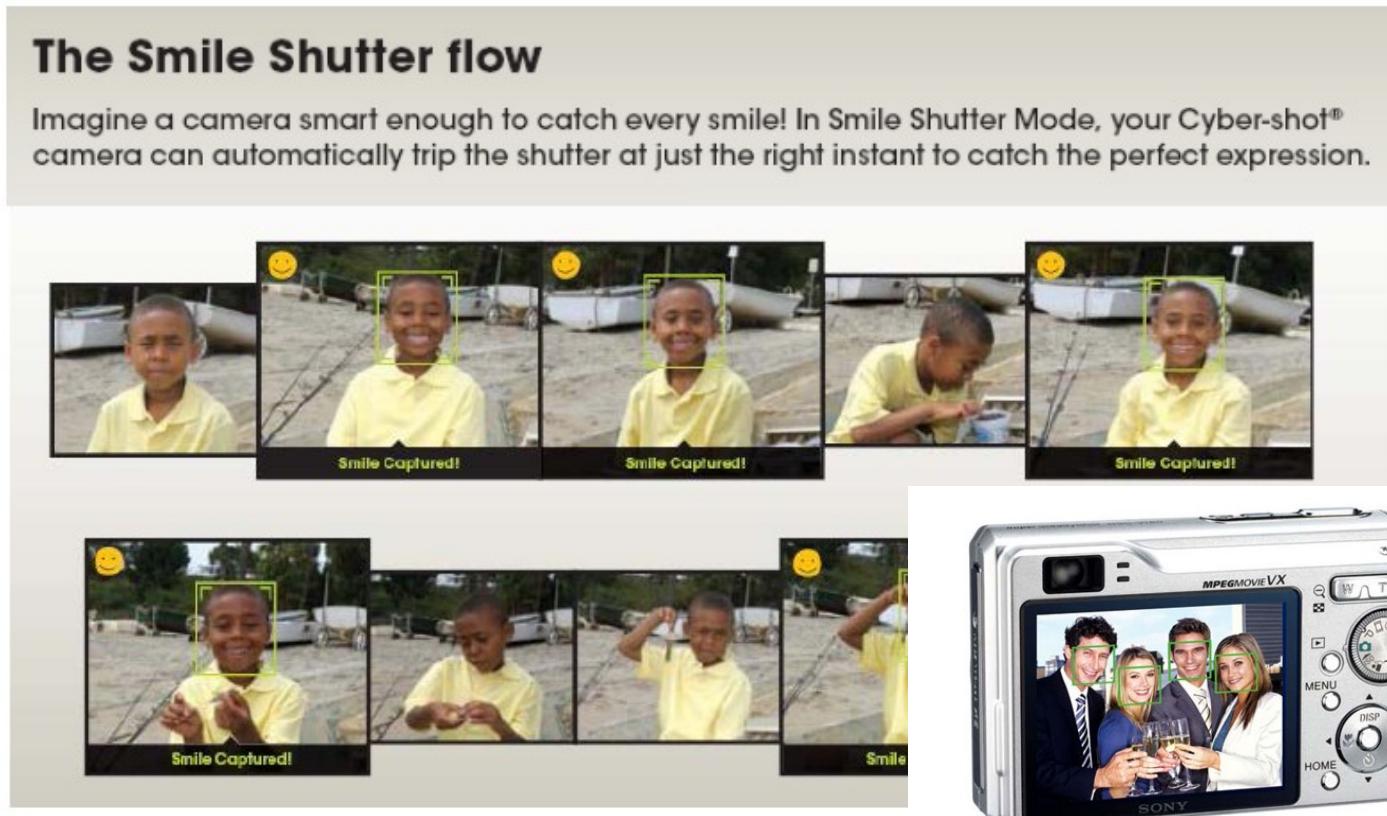


<https://www.youtube.com/watch?v=p16frKJLVi0>

Aplicaciones

Detección/interpretación de caras

- Cámaras digitales Sony, Fuji, Canon...

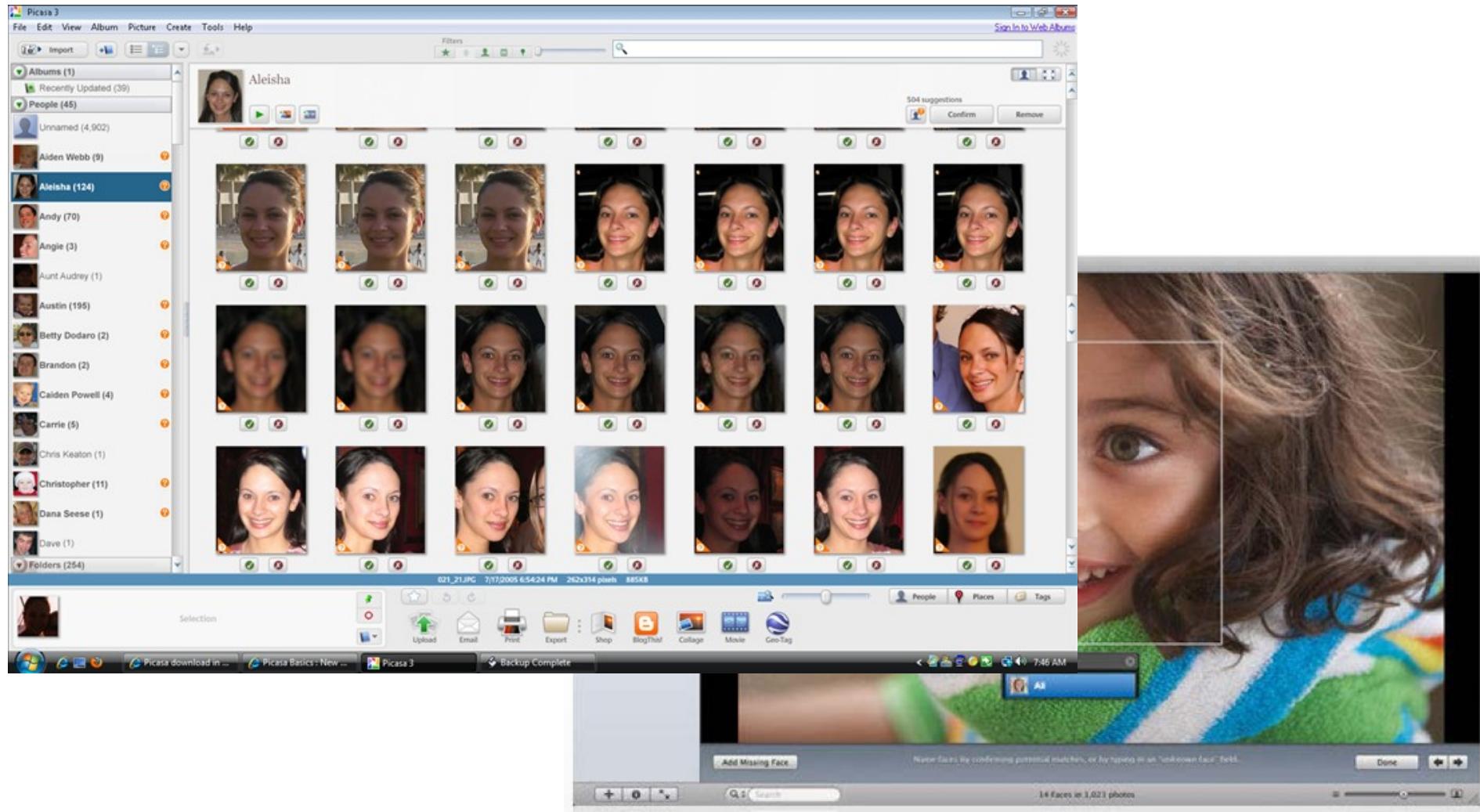


Demo sencilla operativa en OpenCV para Python

Aplicaciones

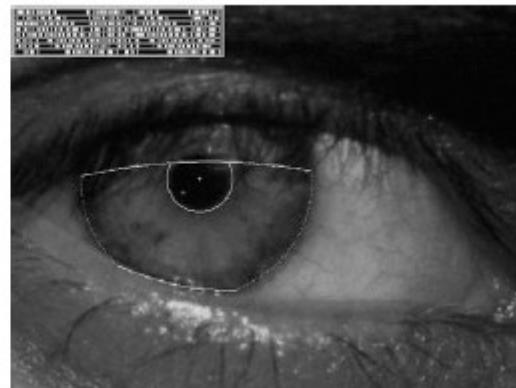
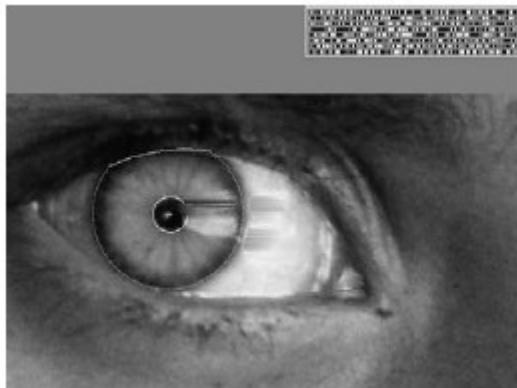
Clasificación de caras

- Google Picassa, Apple iPhoto



Aplicaciones Biometría

- Reconocimiento iris, huellas digitales, caras, ...

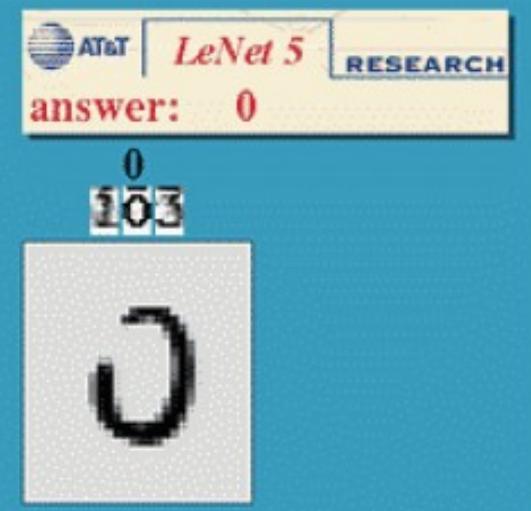


Aplicaciones

Reconocimiento óptico de caracteres (OCR)

- Extraer datos textuales a partir de imágenes

Quizá sea éste uno de los artículos más ilusos que uno pueda escribir hoy en día en una sociedad tan autocomplaciente y autoindulgente como la española actual, y eso que tengo conciencia de haber ya publicado unos cuantos de esa índole -ilusa, quiero decir-. Porque si



Aplicaciones

Robots, juguetes, ...



Aplicaciones

Búsqueda visual en movilidad

Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.



Aplicaciones

Seguridad en automóviles

- Mobileye (BMW, Volvo, ...)

▷▶ manufacturer products consumer products ◀◀

Our Vision. Your Safety.

rear looking camera forward looking camera side looking camera

› **EyeQ** Vision on a Chip

> read more

› **Vision Applications**

Road, Vehicle, Pedestrian Protection and more

> read more

› **AWS** Advance Warning System

News

› Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System

› Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end

> all news

Events

› Mobileye at Equip Auto, Paris, France

› Mobileye at SEMA, Las Vegas, NV

> read more

Aplicaciones

Navegación visual

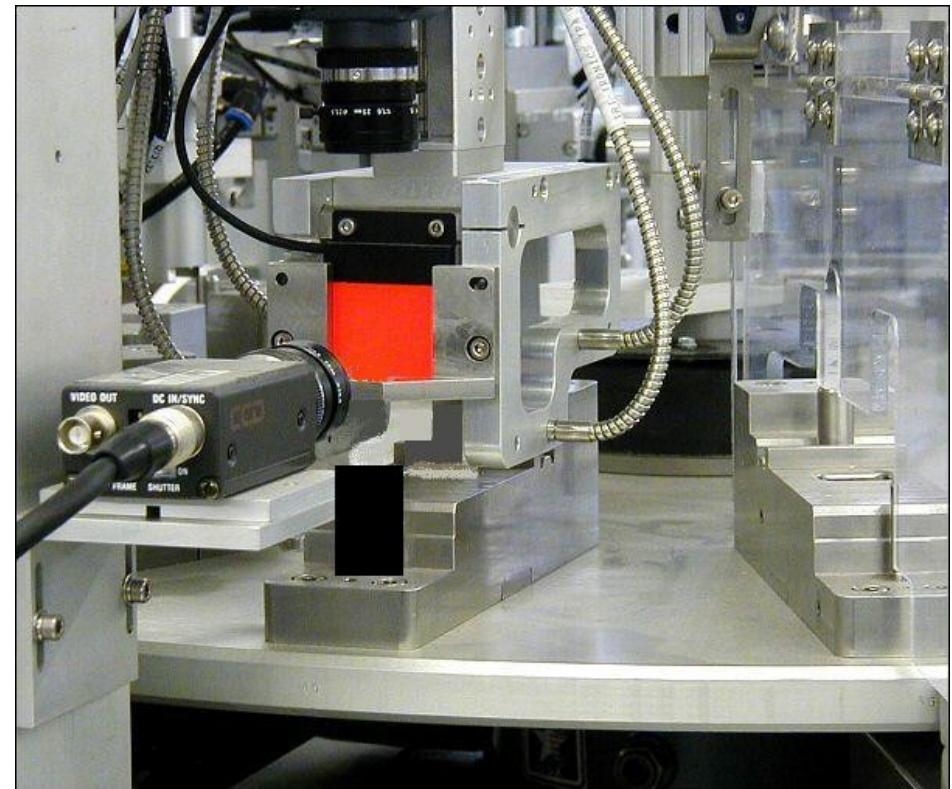
- Google self-driving car:



Aplicaciones

Aplicaciones en la industria

- Control de calidad, metrología, control de brazos industriales ...



Aplicaciones

Aplicaciones en supermercados

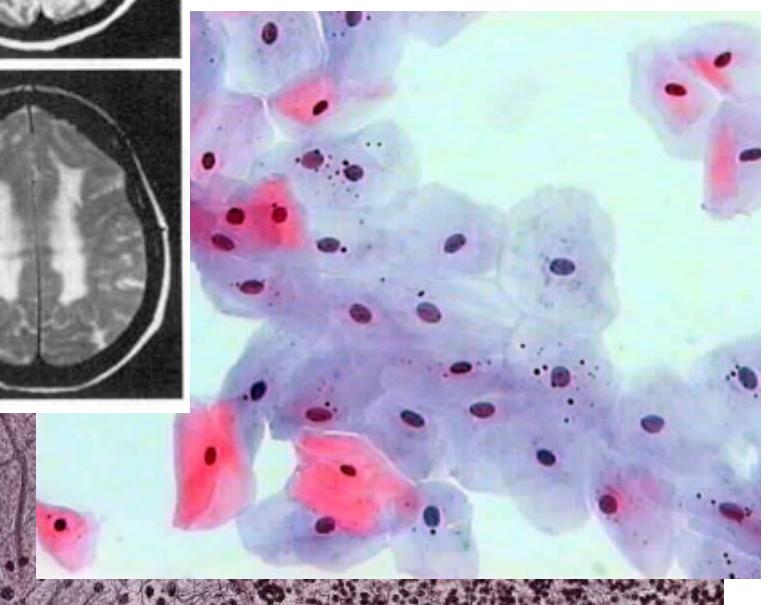
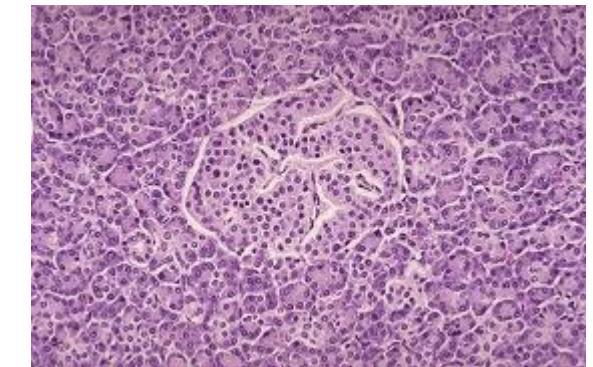
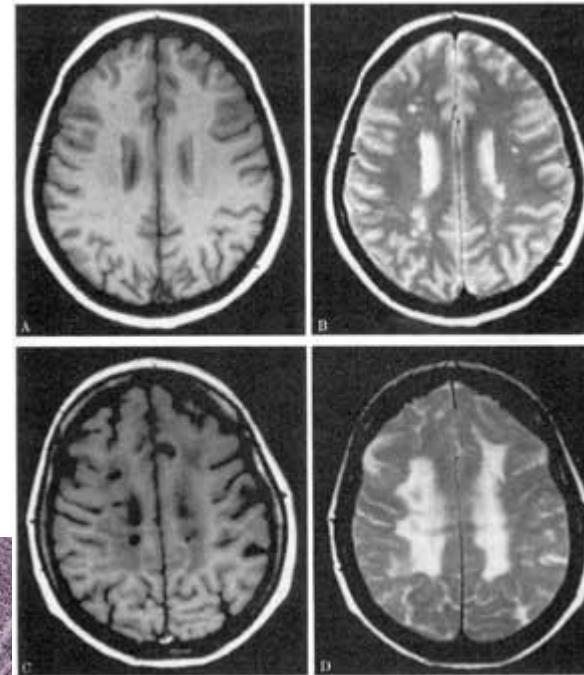
- Lanehawk



Aplicaciones

Aplicaciones en biología/medicina

- Análisis de imágenes (radiología, microscopía...)



Aplicaciones

Interacción basada en visión

- Videojuegos, asistencia a discapacitados...



Microsoft's Kinect



Sony EyeToy



Assistive technologies

Aplicaciones

Image mosaicing

- Robótica, exploración espacial



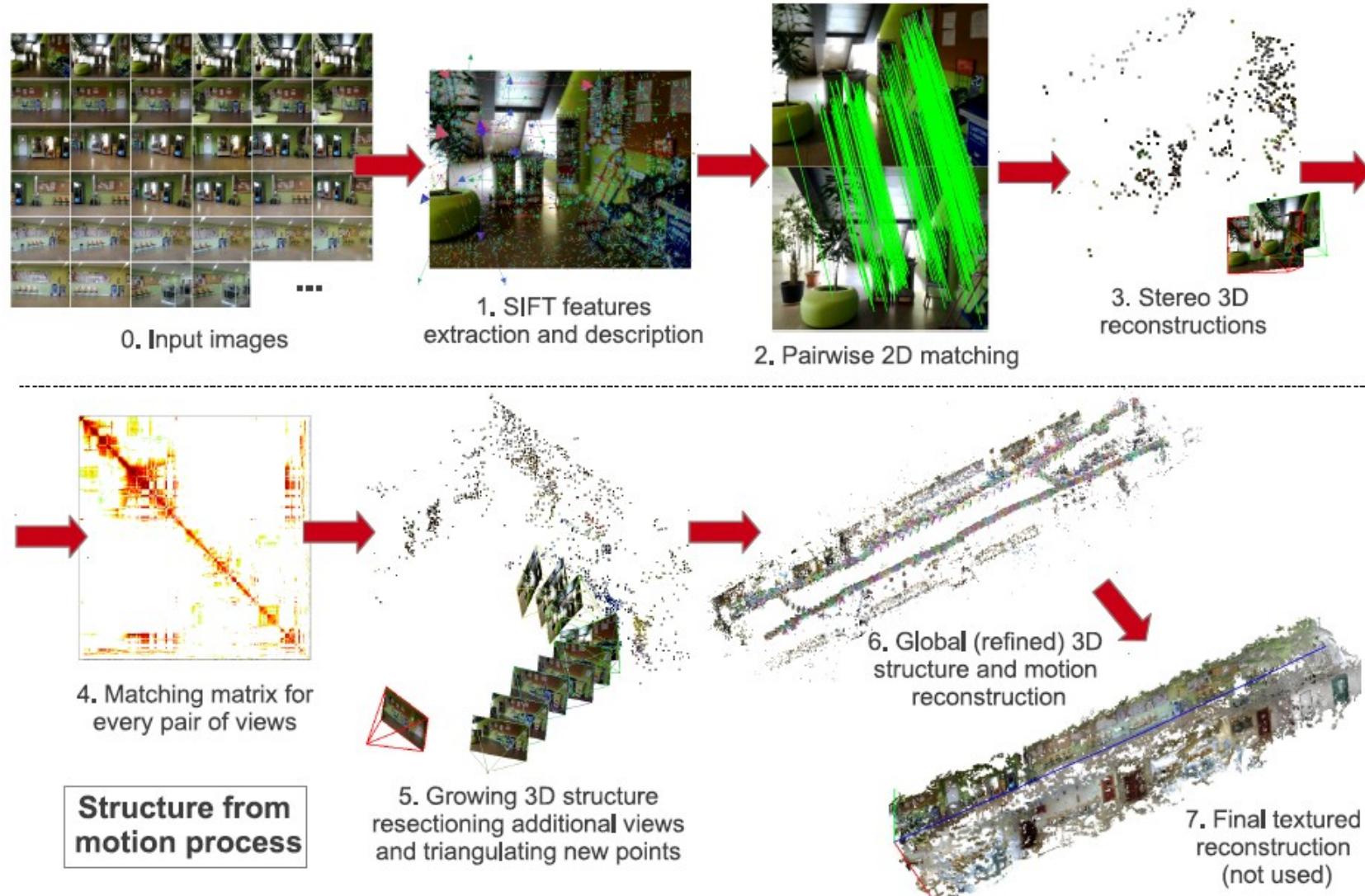
- Adobe Photoshop



Aplicaciones

Realidad aumentada basada en localización

- Proyecto LOCUM-AR:



Aplicaciones

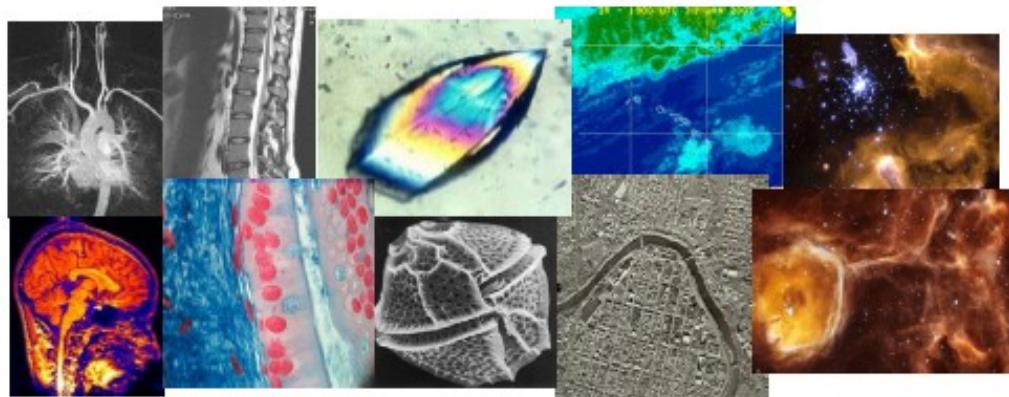
Realidad aumentada basada en localización

- Proyecto LOCUM-AR:
 - Objetivo: realidad aumentada precisa en dispositivos móviles, también en interiores



<https://www.youtube.com/watch?v=qycdlROrtXE&t=15s>

La industria de la visión por computador

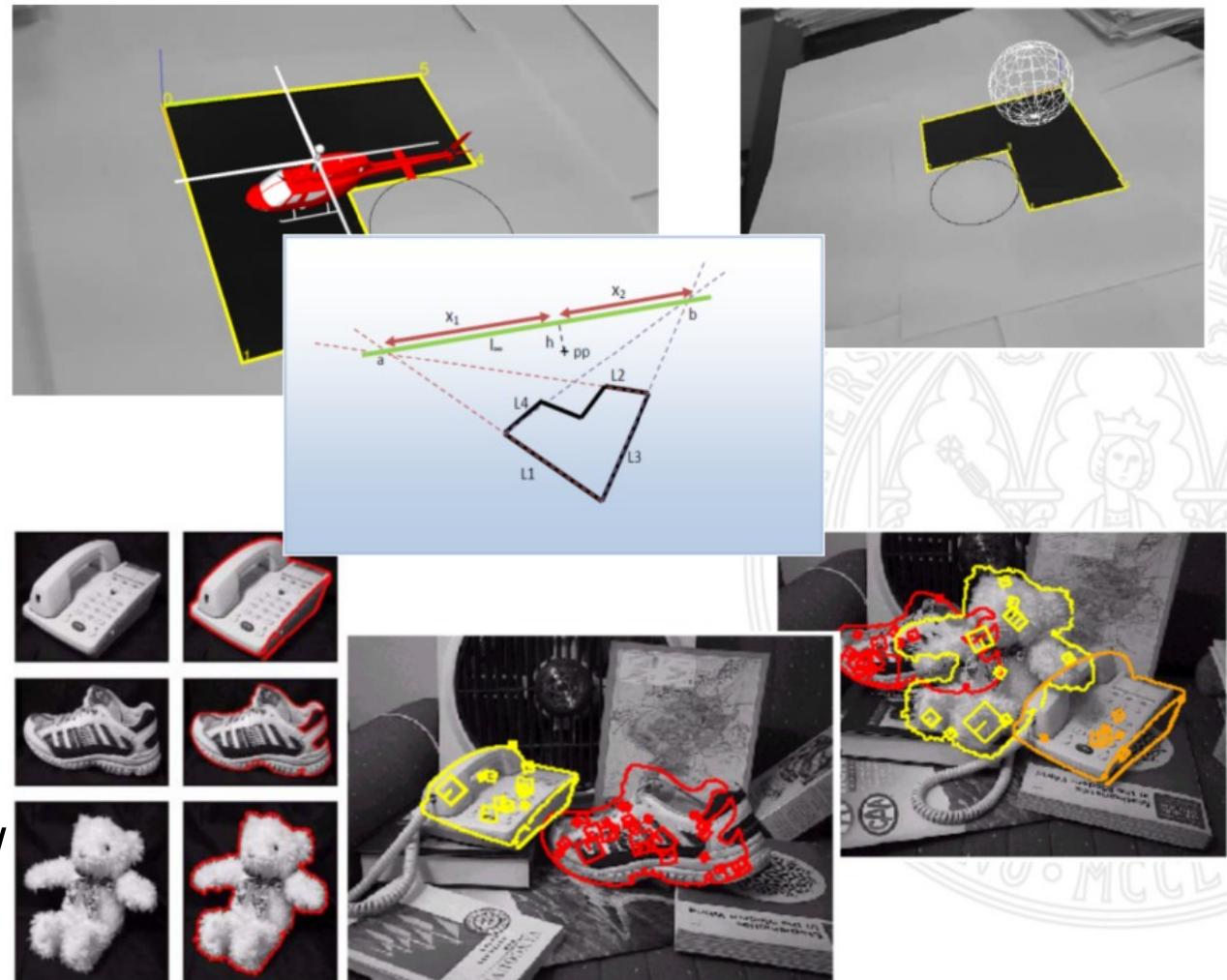


- Una lista de compañías aquí:

<http://www.cs.ubc.ca/~lowe/vision.html>

¿Qué cubriremos en la asignatura?

- Procesamiento de imagen orientado a visión artificial
- Reconocimiento de patrones visuales (identificación de objetos)
- Aplicaciones de geometría proyectiva (realidad aumentada, reconstrucción 3D, seguimiento cámaras/escena)



Profesorado

- Pedro Enrique López de Teruel Alcolea
 - Despacho 3.43 de la Facultad de Informática
 - Tutorías: jueves de 12:30 a 14:00 y viernes de 9:30 a 11:00
 - También tutoría electrónica a través del aula virtual (Herramienta “*Mensajes privados*” en <http://aulavirtual.um.es>)

Objetivos

- Revisión estado del arte en visión por computador:
 - Filtros de imagen, extracción de características (detectores/descriptores)
 - Técnicas estadísticas (robustez, clasificación, *clustering*, ...)
 - Reconocimiento de objetos. Invarianza a transformaciones (escala, rotación, iluminación)
 - Geometría de múltiples vistas (SfM)
 - ¿Otros? ...

Contenidos teóricos (I)

- Bloque 1: Introducción
 - **Tema 1: Introducción a la visión por computador**
 - Introducción a la disciplina de la visión por computador
 - Organización de la asignatura
 - **Tema 2: Introducción al software de prácticas**
 - Lenguaje Python; NumPy, SciPy, Matplotlib
 - Librería OpenCV para procesamiento de imagen y visión por computador en entorno de programación Python

Contenidos teóricos (II)

- Bloque 2: Procesamiento de imagen y extracción de características
 - **Tema 3: Procesamiento de imagen aplicado a visión por computador**
 - Transformaciones de píxel, basadas en vecinos
 - Filtros lineales y no lineales
 - Transformaciones geométricas, morfológicas, de distancia
 - Componentes conexas
 - Pirámides de imagen

Contenidos teóricos (III)

- Bloque 2: Procesamiento de imagen y extracción de características (cont.)
 - **Tema 4: Extracción de características I**
 - Introducción a las *features*, ejemplos de aplicación, repetibilidad, discernibilidad, eficiencia
 - Técnicas de extracción de puntos y bordes, gradiente, *hipass*, Canny, Harris, Hessian, Shi-Tomasi, multiescala, SUSAN, FAST

Contenidos teóricos (IV)

- Bloque 2: Procesamiento de imagen y extracción de características (cont.)
 - **Tema 5: Extracción de características II**
 - Introducción a técnicas estadísticas robustas, RANSAC, *Hough transform*
 - Detección y extracción de *features* invariantes a escala, SIFT, MSER, *component trees*
 - Aplicación al reconocimiento de objetos

Contenidos teóricos (V)

- Bloque 3: Geometría proyectiva
 - **Tema 6: Geometría proyectiva I**
 - Introducción a la geometría visual, y el *Structure from Motion*
 - Modelos de formación de imágenes
 - Transformaciones de perspectiva, calibración, homografías, rectificación de planos
 - Aplicaciones en realidad aumentada

Contenidos teóricos (VI)

- Bloque 3: Geometría proyectiva (cont.)
 - **Tema 7: Geometría proyectiva II**
 - Geometría epipolar, matrices fundamental y esencial, algoritmo de los 8 puntos
 - Obtención simultánea de estructura y movimiento
 - Autocalibración
 - *Bundle Adjustment*
 - Reconstrucción densa

Contenidos prácticos

- Aplicaciones de visión artificial desarrolladas sobre Python + OpenCV:
 - Miniproyectos prácticos relacionados con los bloques de la asignatura:
 - Extracción de primitivas
 - Calibración
 - Realidad aumentada
 - Reconocimiento de objetos
 - Se partirá de códigos base sobre los que:
 - Habrá que añadir ciertas funcionalidades obligatorias
 - Podrán realizarse aportaciones voluntarias (propuestas por el profesor, o propias)

Evaluación

- Miniproyectos de programación **individual o por parejas**
 - Corrección y eficiencia de aportaciones obligatorias
 - Corrección, eficiencia, dificultad y originalidad de aportaciones voluntarias
- **Entrevista final** sobre miniproyectos de programación
- Asistencia a clases de teoría / prácticas

Bibliografía

- Básica:
 - Transparencias teoría, basadas en *Stanford University Open Computer Vision course, 2014* (<http://aulavirtual.um.es>).
 - Documentación Python (Numpy + Scipy + Matplotlib) + OpenCV (ver bibliografía tema 2).
 - Apuntes de Sistemas de Percepción y Visión por computador (<http://dis.um.es/~alberto/material/percep.pdf>).
- Complementaria
 - "Multiple view geometry" (2nd ed.), R. Hartley, A. Zisserman (2003), Cambridge University Press.
 - "Computer Vision: Algorithms and Applications", R. Szeliski (2010), Springer (<http://szeliski.org/Book/>).