



# UNIVERSIDAD DE GRANADA

VISIÓN POR COMPUTADOR  
MÁSTER CIENCIA DE DATOS E INGENIERÍA DE COMPUTADORES

---

## TRABAJO 1

### SISTEMAS DE VISIÓN ARTIFICAL

---

#### Autor

Ignacio Vellido Expósito  
ignaciove@correo.ugr.es



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE  
TELECOMUNICACIÓN

CURSO 2020-2021

## 1. Introducción

### 1.1. Historia

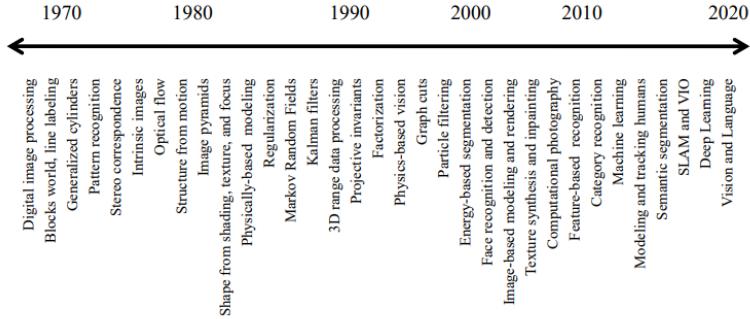
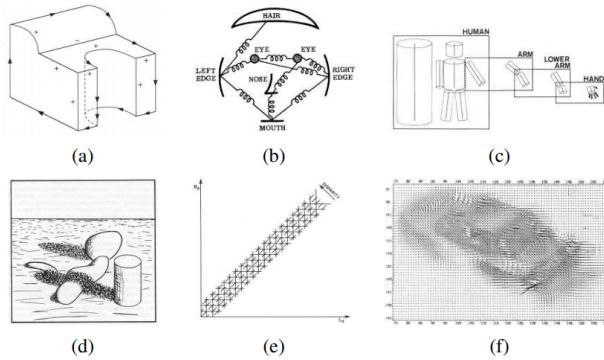


Figura 1: Eventos destacables en la historia de la visión artificial.

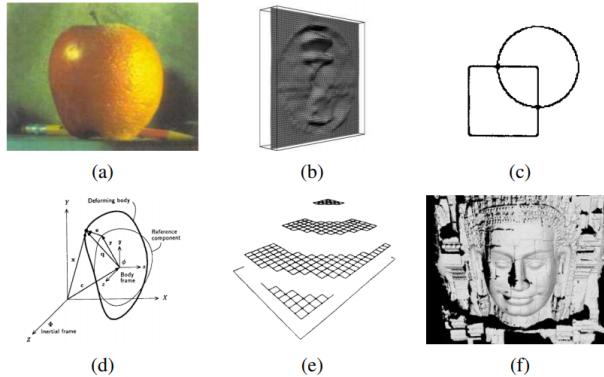
La visión artificial comienza a plantearse en los años 60, después de que Larry Roberts propusiera la extracción de información geométrica a partir de fotografías 2D. Sus primeras aproximaciones trataban tareas específicas como la detección de bordes o la identificación de patrones, basándose en técnicas matemáticas rigurosas. Durante los 80 siguió el énfasis en explorar las distintas técnicas matemáticas y se plantean las primeras ideas de CNN. Entre algunas de las aplicaciones se comienzan a usar las pirámides para la mezcla y para el manejo de imágenes multi-escala. En los 90 surge un mayor esfuerzo en resolver problemas de movimiento. También aumenta el número de aplicaciones relacionadas con gráficos, como la modelización y representación de imágenes 3D.

A comienzos del siglo XXI cambia la aproximación clásica de crear buenos modelos matemáticos con características cuidadosamente extraídas por un enfoque en el aprendizaje de características a partir de los datos bajo la asunción de que múltiples conceptos son compartidos. Entre algunas de las aplicaciones está el primer framework de detección de caras a tiempo real (Viola-Jones) y el inicio de la experimentación con coches autónomos.

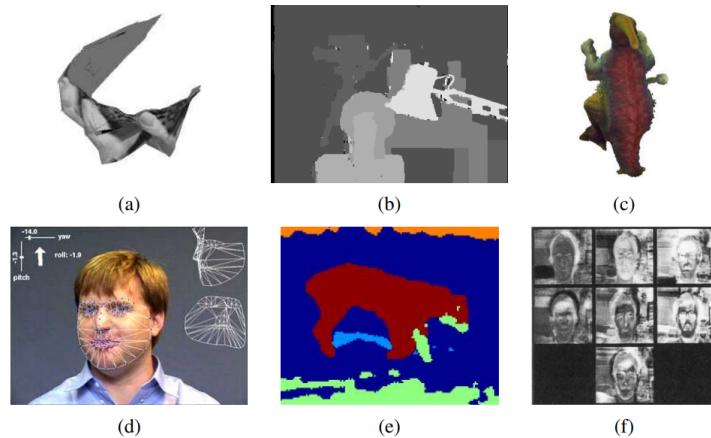
Durante la década pasada sigue la tendencia de usar grandes bases de datos para desarrollar algoritmos de aprendizaje, y en 2012 comienza la expansión del aprendizaje profundo cuando la arquitectura AlexNet gana la competición ImageNet. A partir de entonces se diversifica el uso de la visión artificial en múltiples campos diferentes, la mayoría basándose en el potencial de estas nuevas técnicas.



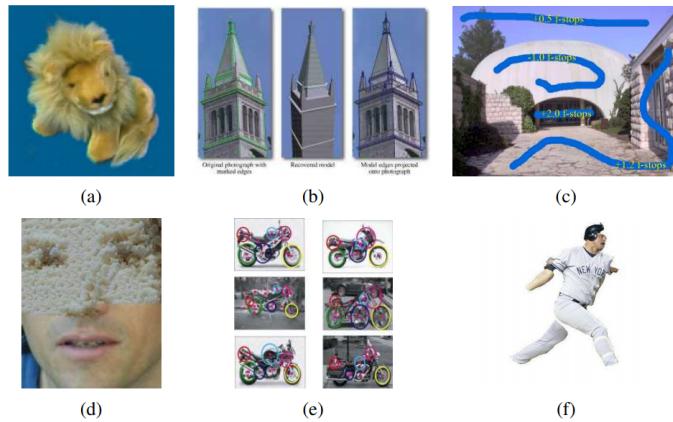
**Figure 1.7** Some early (1970s) examples of computer vision algorithms: (a) line labeling (Nalwa 1993) © 1993 Addison-Wesley, (b) pictorial structures (Fischler and Elschlager 1973) © 1973 IEEE, (c) articulated body model (Marr 1982) © 1982 David Marr, (d) intrinsic images (Barrow and Tenenbaum 1981) © 1973 IEEE, (e) stereo correspondence (Marr 1982) © 1982 David Marr, (f) optical flow (Nagel and Enkelmann 1986) © 1986 IEEE.



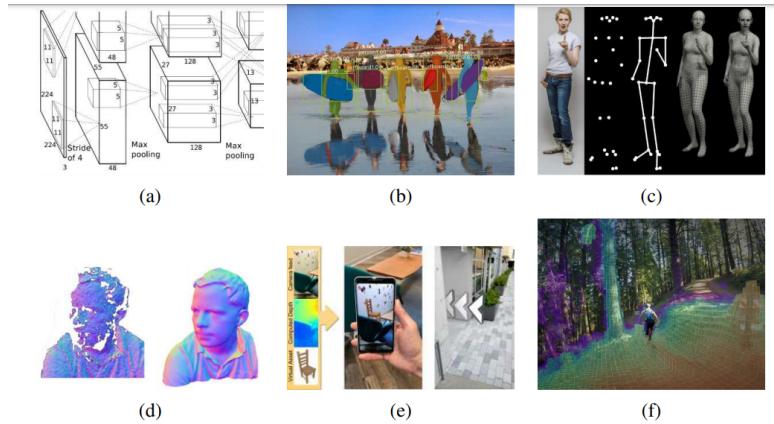
**Figure 1.8** Examples of computer vision algorithms from the 1980s: (a) pyramid blending (Burt and Adelson 1983b) © 1983 ACM, (b) shape from shading (Freeman and Adelson 1991) © 1991 IEEE, (c) edge detection (Freeman and Adelson 1991) © 1991 IEEE, (d) physically based models (Terzopoulos and Witkin 1988) © 1988 IEEE, (e) regularization-based surface reconstruction (Terzopoulos 1988) © 1988 IEEE, (f) range data acquisition and merging (Banno, Masuda et al. 2008) © 2008 Springer.



**Figure 1.9** Examples of computer vision algorithms from the 1990s: (a) factorization-based structure from motion (Tomasi and Kanade 1992) © 1992 Springer, (b) dense stereo matching (Boykov, Veksler, and Zabih 2001), (c) multi-view reconstruction (Seitz and Dyer 1999) © 1999 Springer, (d) face tracking (Matthews, Xiao, and Baker 2007), (e) image segmentation (Belongie, Fowlkes et al. 2002) © 2002 Springer, (f) face recognition (Turk and Pentland 1991).



**Figure 1.10** Examples of computer vision algorithms from the 2000s: (a) image-based rendering (Gortler, Grzeszczuk et al. 1996), (b) image-based modeling (Debevec, Taylor, and Malik 1996) © 1996 ACM, (c) interactive tone mapping (Lischinski, Farbman et al. 2006) (d) texture synthesis (Efros and Freeman 2001), (e) feature-based recognition (Fergus, Perona, and Zisserman 2007), (f) region-based recognition (Mori, Ren et al. 2004) © 2004 IEEE.



**Figure 1.11** Examples of computer vision algorithms from the 2010s: (a) the SuperVision deep neural network © Krizhevsky, Sutskever, and Hinton (2012); (b) object instance segmentation (He, Gkioxari et al. 2017) © 2017 IEEE; (c) whole body, expression, and gesture fitting from a single image (Pavlakos, Choutas et al. 2019) © 2019 IEEE; (d) fusing multiple color depth images using the KinectFusion real-time system (Newcombe, Izadi et al. 2011) © 2011 ACM; (e) smartphone augmented reality with real-time depth occlusion effects (Valentin, Kowdle et al. 2018) © 2018 ACM; (f) 3D map computed in real-time on a fully autonomous Skydio R1 drone (Cross 2019)

## 1.2. Elementos

Podríamos decir que los sistemas de visión por computador tienen en común: Un sistema de captación y/o almacenamiento de imágenes, ya sea un físico (ej: cámara, escáner) o virtual (ej: recuperación de imágenes de internet). Un sistema de preprocesamiento y adecuación de las imágenes para su uso. La aplicación de técnicas para la extracción de información a partir de las imágenes. Realizar un proceso o tarea a partir de la información extraída, ya sea razonar o actuar (ej: vehículo autónomo), o generar una salida (ej: detección de objetos).

## 1.3. Inspiración humana

En términos generales el ojo y el cerebro humano son las fuentes principales de inspiración en la visión artificial, pero al haber aún mucho desconocimiento sobre el funcionamiento de este último siguen utilizándose técnicas matemáticas basadas en los pixeles de la imagen.

Existe también inspiración humana en la funcionamiento de las CNN. El ojo humano detecta características geométricas de bajo nivel (líneas, bordes, esquinas) con las que agrupando jerárquicamente va identificando el objeto. Esto mismo se ha visto modelado en las capas de una CNN a diferentes profundidades. Además, estas características básicas son comunes entre objetos, y son las diferentes formas de combinarse las que los distinguen los unos de los otros.

## 2. Aplicaciones

Ejemplos por campos de aplicación:

■ Medicina

- Extracción de información a partir de imágenes médicas para el descubrimiento o tratamiento de enfermedades.
- Realzado de imágenes ruidosas para su interpretación por humanos.
- En robot terapéuticos, detección de caras y poses, clasificación de emociones, localización de puntos de mira.

■ Robótica

- Ayuda a la maquinaria en la fabricación, mediante inspección de los productos o la localización de elementos utilizados en una cadena de montaje.
- Para la navegación y el esquivo de obstáculos.

■ Seguridad

- Reconocimiento de personas mediante huellas dactilares.
- Identificación de caras.
- Detección de armas, video vigilancia.

■ Gráficos

- Reconstrucción de espacios 3D.
- Match moving: Juntar CGI con imágenes reales.
- Captura de formas y movimientos para efectos especiales.

■ Otros

- Soporte al cumplimiento de reglas en deportes.
- Modelado de terrenos y creación de imágenes panorámicas en misiones espaciales.
- Identificación forense.
- En el campo militar, guiado de misiles.

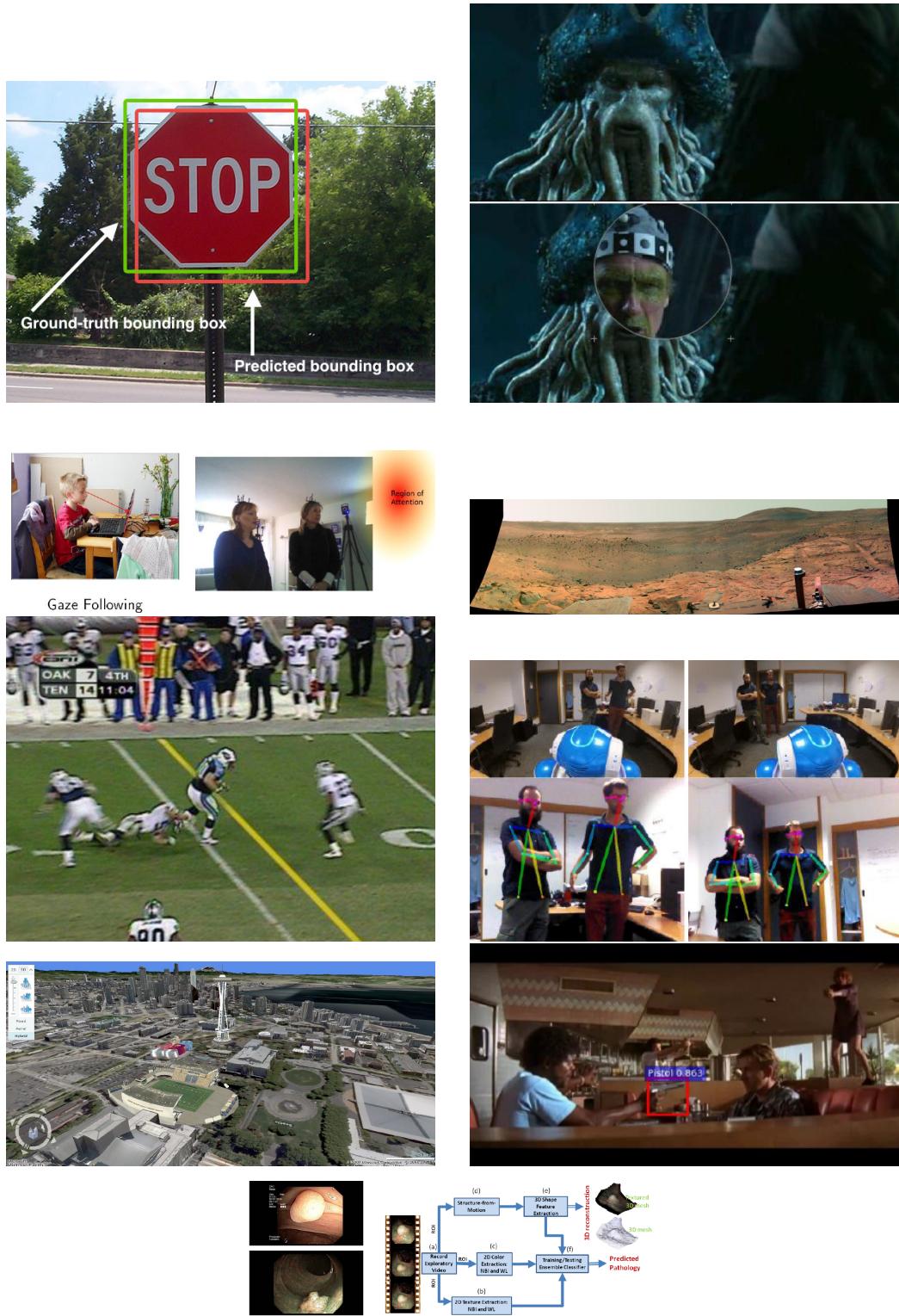


Figura 2: Aplicaciones de la visión artificial.

### 3. Desarrollo de una aplicación

Vamos a describir el proceso de desarrollo del sistema de visión de un supuesto robot terapéutico, capaz de detectar y hablar a la cara de una persona. Adicionalmente buscamos poder identificar personas o clasificar expresiones faciales.

1. **Adquisición:** El sistema de detección de caras debe incluir una cámara de vídeo de calidad suficiente para distinguir objetos en la escena. Puesto que vamos a aplicar técnicas de aprendizaje sobre las caras encontradas, necesitamos además una resolución buena y una buena capacidad de zoom (dependiente de la altura del robot).

Adicionalmente, puesto que no se espera que los objetos en la imagen se muevan a alta velocidad, no es necesario contar con un framerate alto.

Otra posible opción sería acompañar la cámara de vídeo con otro de tipos de cámaras no independientes de la luminosidad. Esto nos forzaría a luego traspasar las regiones detectadas a la imagen normal, y por simplicidad vamos a suponer en este caso que contamos con una única cámara.

2. **Preprocesamiento:** Para facilitar la tarea de aprendizaje sería conveniente eliminar ruido de la imagen y ajustar el contraste de la imagen en base a las condiciones lumínicas de la escena. Una forma simple con la que empezar podría ser un equalización de histograma, pero se podrían utilizar otras técnicas más complejas y adaptativas, por ejemplo haciendo uso de otros sensores que pudiera contar el robot.

Según el tipo de aprendizaje que vayamos a utilizar podría sernos útil una representación solo de bordes de la imagen. En este caso vamos a suponer que tenemos datos suficientes y podemos utilizar técnicas de deep learning donde los detalles y colores de la imagen nos son útiles.

3. **Segmentación:** Podría ser interesante poder segmentar los objetos del fondo, de cara a facilitar la tarea de aprendizaje.

4. **Extracción de características:** Podemos utilizar redes profundas como YOLO, donde no es necesario extraer características previas de la imagen. Por contrapartida esta red nos fuerza a que una vez tengamos las regiones de interés donde se encuentran las caras debamos comparar con frames anteriores para detectar cuál de estas regiones corresponde a la que estaba centrada el robot, de cara a no perder el “interés” de una cara por otra.

Otra forma sería extraer las regiones aparte y luego aplicar un entrenamiento sobre cada una de ellas. En este caso sería importante no perder la localización espacial de la región y que la técnica de aprendizaje sea flexible en el tamaño de estas.

También se podrían considerar usar descriptores de características como histogramas de gradientes orientados.

5. **Entrenamiento:** En cualquiera de los casos tratamos con técnicas de aprendizaje supervisado, por lo que necesitamos datos lo más representativos posibles.

Una vez obtenidas las regiones de interés podríamos llevar más allá el sistema realizando aprendizajes adicionales sobre ella, como identificación de personas o clasificación de sentimientos. En este caso probablemente la primera opción a considerar sería una red convolucional pre-entrenada del estado del arte, sobre la que podríamos aplicar transfer learning y fine tuning.

Por último, el sistema de tracking debe poder retroalimentar los actuadores del robot para volver a centrar la cara. Esto se podría realizar fácilmente segmentando la imagen en forma de brújula e indicando en qué sector se ubica la cara.

Bibliografía:

- Computer Vision: Algorithms and Applications, 2nd ed. 2021 Richard Szeliski
- Computer Vision: Evolution and Promise. T. S. Huang
- [https://en.wikipedia.org/wiki/Computer\\_vision](https://en.wikipedia.org/wiki/Computer_vision)
- <https://www.verdict.co.uk/computer-vision-timeline/>
- [https://en.wikipedia.org/wiki/Pyramid\\_\(image\\_processing\)](https://en.wikipedia.org/wiki/Pyramid_(image_processing))
- Presentaciones de la asignatura: Visión por Computador. Nicolas Perez de la Blanca Capilla
- Presentaciones de la asignatura: Aplicaciones de Ciencia de Datos. Pablo Mesejo
- [https://en.wikipedia.org/wiki/Match\\_moving](https://en.wikipedia.org/wiki/Match_moving)
- <https://www.cs.ubc.ca/~lowe/vision.html>