Visión por Computador

Tema 5. Reconocimiento

EII, GII. *Modesto Castrillón Santana*

Contenidos

- Introducción
- Detección de objetos
- Segmentación semántica
- Clasificación
- Evaluación experimental
- Vídeo
- Visión y lenguaje

Introducción

Tareas VC:

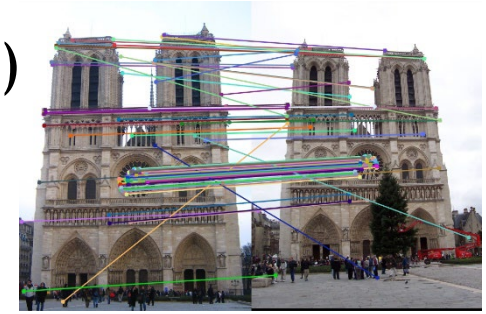
- Clasificación de imagen:
  - ¿Es un gato?
  - ¿Qué clase de escena es esta?

Introducción

Tareas VC:

- Clasificación de imagen:
  - ¿Es un gato?
  - ¿Qué clase de escena es esta?

- Reconocimiento de ejemplar (*instance*)
  - Localiza este edificio
  - Busca a esta persona

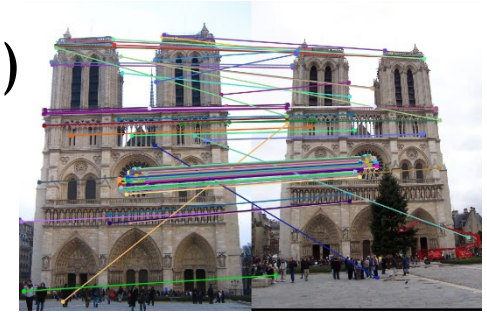


Introducción

Tareas VC:

- Clasificación de imagen:
  - ¿Es un gato?
  - ¿Qué clase de escena es esta?

- Reconocimiento de ejemplar (*instance*)
  - Localiza este edificio
  - Busca a esta persona



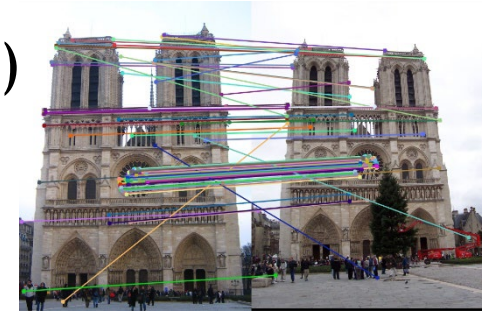
- Detección
  - Localiza coches, personas, etc.

Introducción

Tareas VC:

- Clasificación de imagen:
  - ¿Es un gato?
  - ¿Qué clase de escena es esta?

- Reconocimiento de ejemplar (*instance*)
  - Localiza este edificio
  - Busca a esta persona

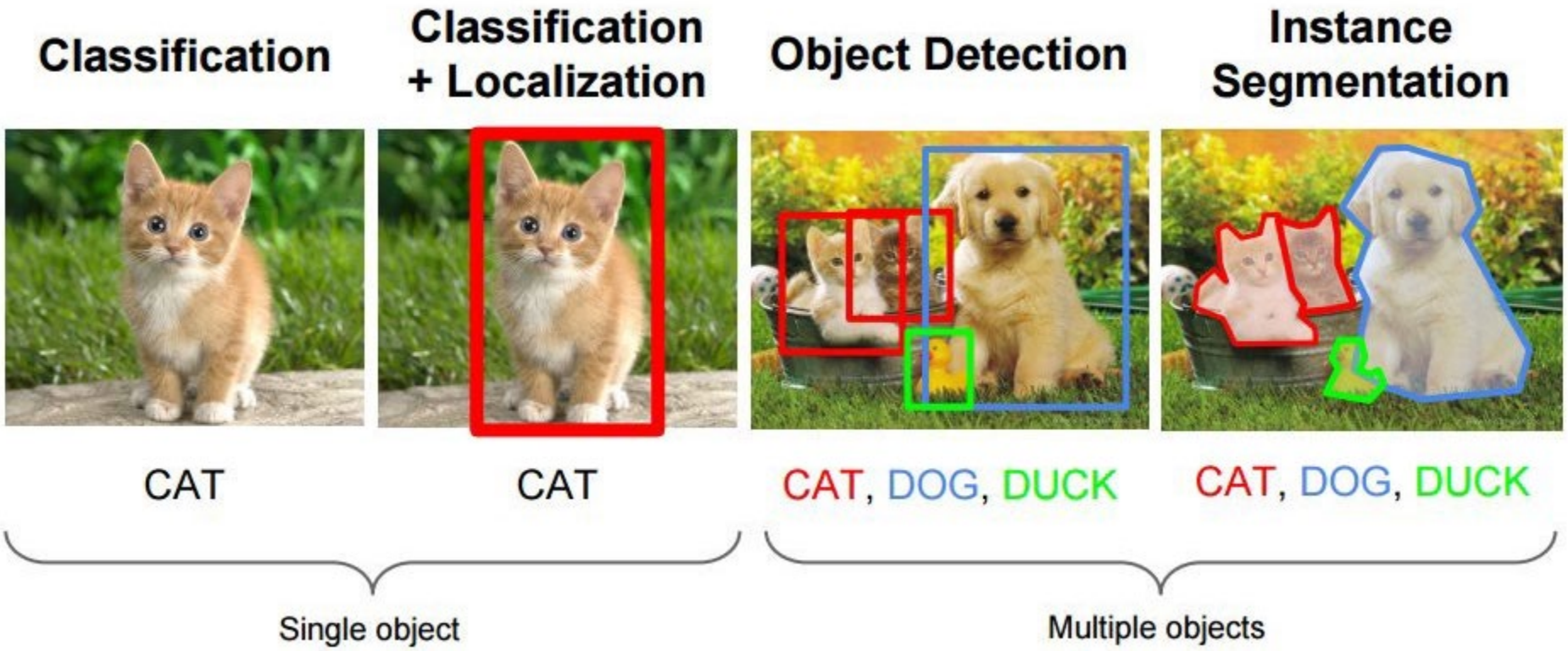


- Detección
  - Localiza coches, personas, etc.

- Comprensión de vídeo
  - ¿Qué está haciendo esta persona?

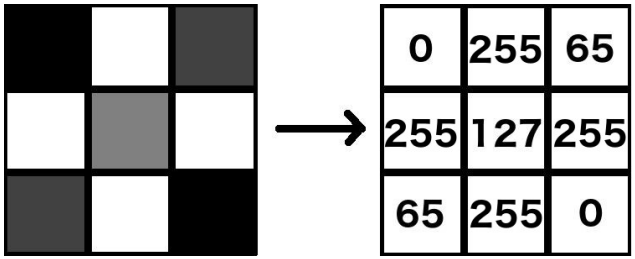
Introducción

Tareas VC:



Introducción

¿Dificilillo?



0	255	65
255	127	255
65	255	0

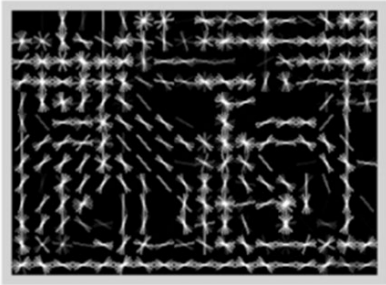
Introducción

Proceso general:

1. Especificar modelo
   - Estadístico
   - Articulado
   - Colección de ejemplares
   - Parametrizable
2. Generar hipótesis
   - Ventana deslizante
   - Puntos de interés
   - Regiones
3. Puntuar hipótesis
   - Confianza
   - Múltiples clasificadores
4. Resolver detecciones
   - Eliminación de no máximos (NMS)
   - Contexto

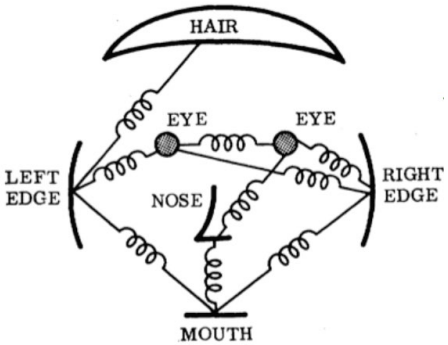
Introducción

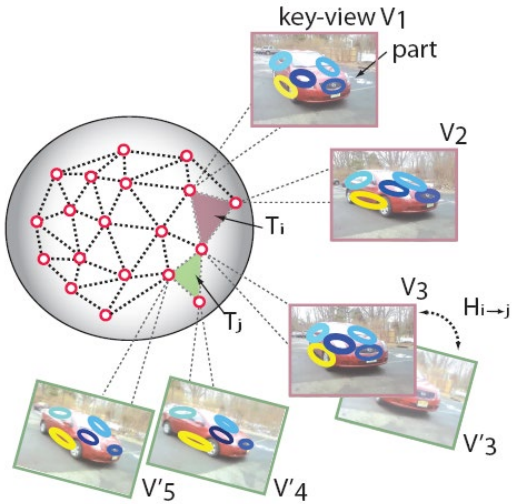
1. Especificar modelo
   - Estadístico
   - Articulado
   - Colección de ejemplares
   - Parametrizable

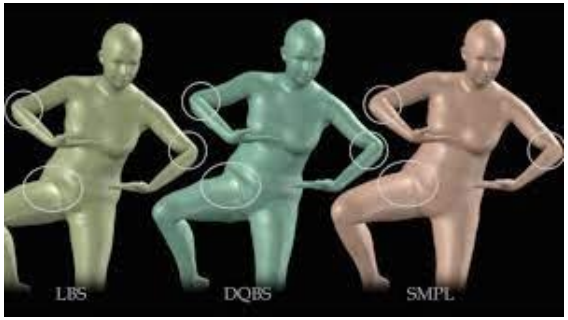


Modelo más complejo no necesariamente mejor

+ complejidad -> + costoso

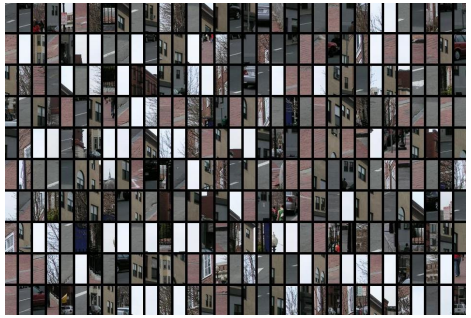






Introducción

2. Generar hipótesis
   - Ventana deslizante
   - Puntos de interés
   - Regiones

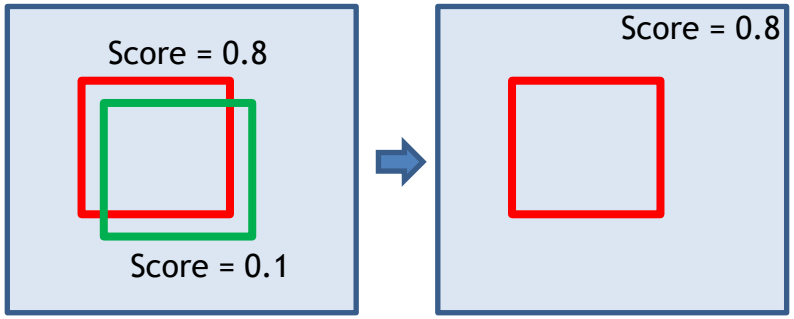


B. Leibe et al.

Endres Hoiem

Introducción

3. Puntuar hipótesis
   - Confianza
   - Múltiples clasificadores
4. Resolver detecciones
   - Eliminación de no máximos (*NMS*)
   - Contexto



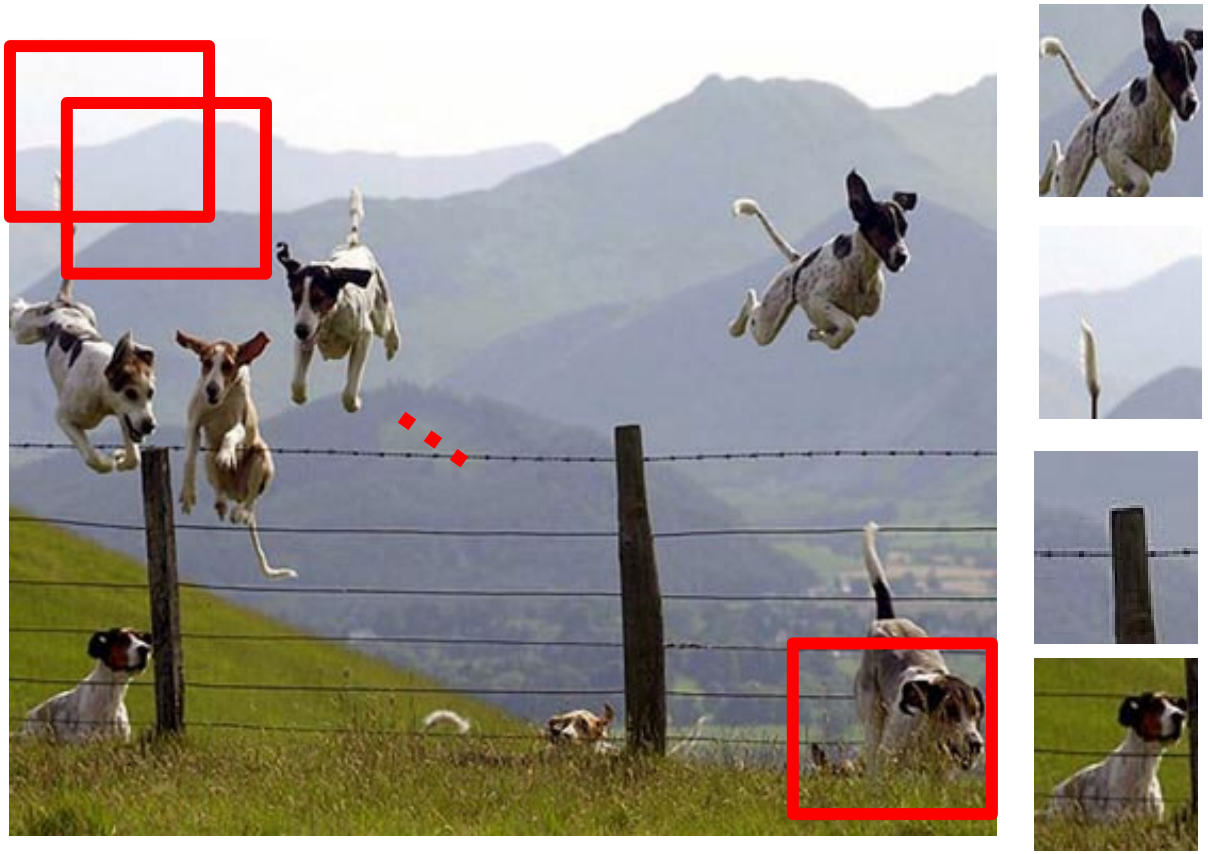
James Hays

(g) Car Detections: Local (h) Ped Detections: Local

Hoiem et al.

Detección

Localizar objeto en la imagen
Modelar el objeto

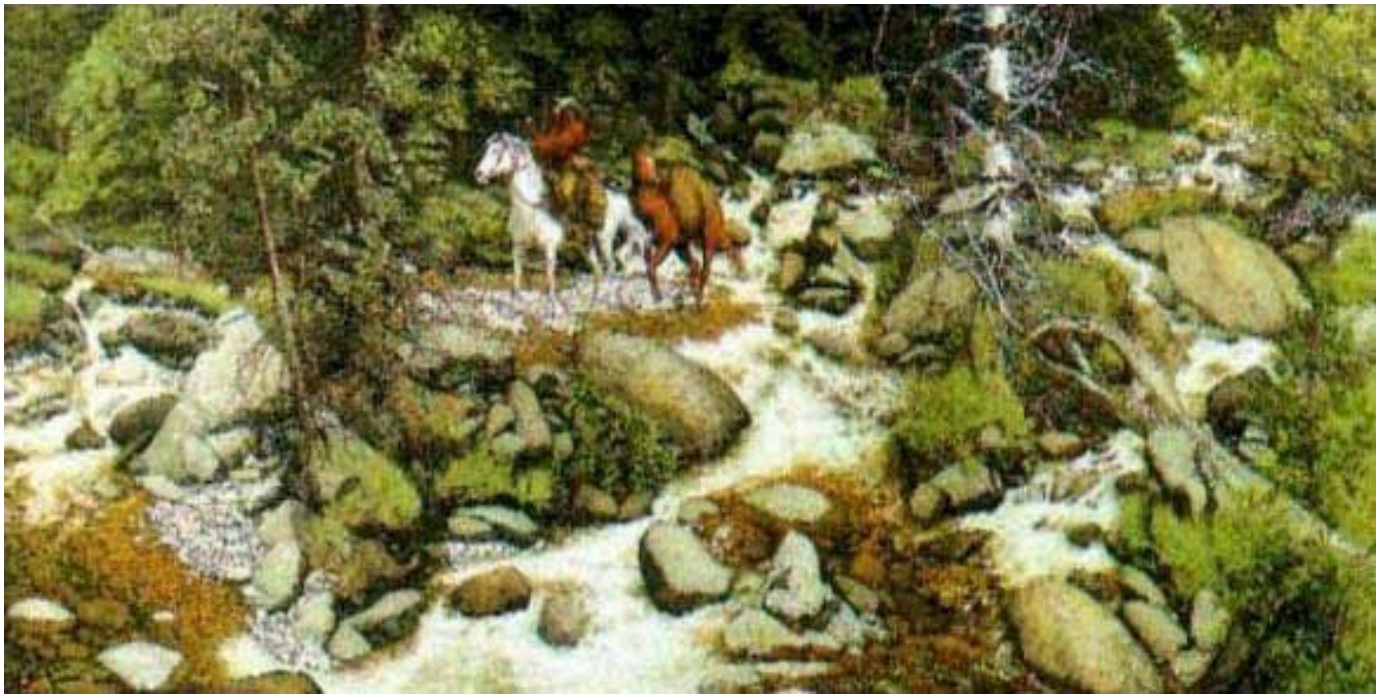


¿Ser o no ser?

James Hays

Detección

Comprobar multitud de ventanas/zonas y escalas
Selección de características
Medidas de encaje/similitud
Cambios de puntos de vista

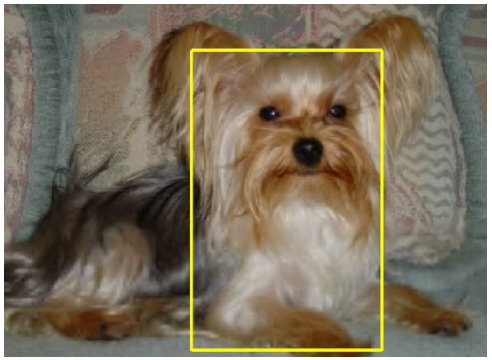


Detección

Modelar el objeto

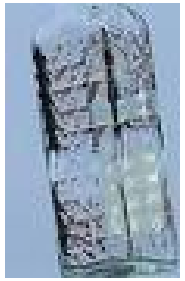
TP

Mala localización



Similares

Diferentes



James Hays

Detección

Retos

Iluminación

Pose

Fondo

Oclusiones

Apariencia intraclase

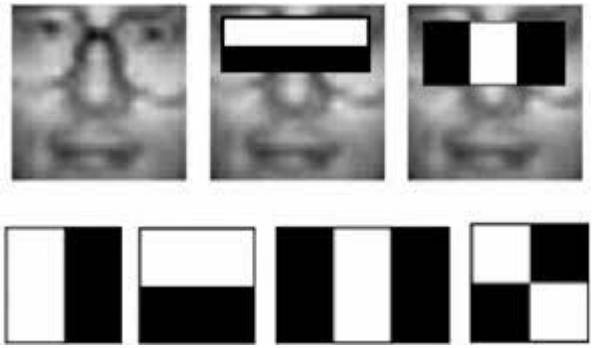
Punto de vista

K. Grauman, B. Leibe

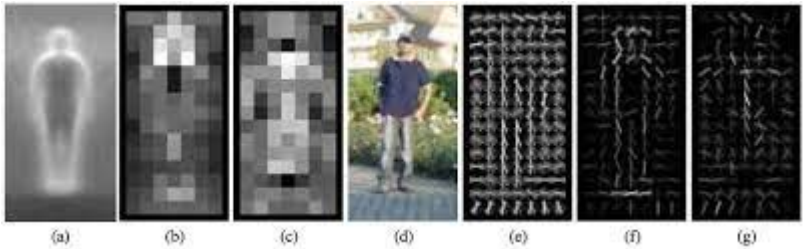
Detección

Enfoques previos a DL:

- Viola y Jones (caras)
  - *Boosting*
  - Árboles de decisión



- Dalal y Triggs (personas)
  - HOG
  - SVM

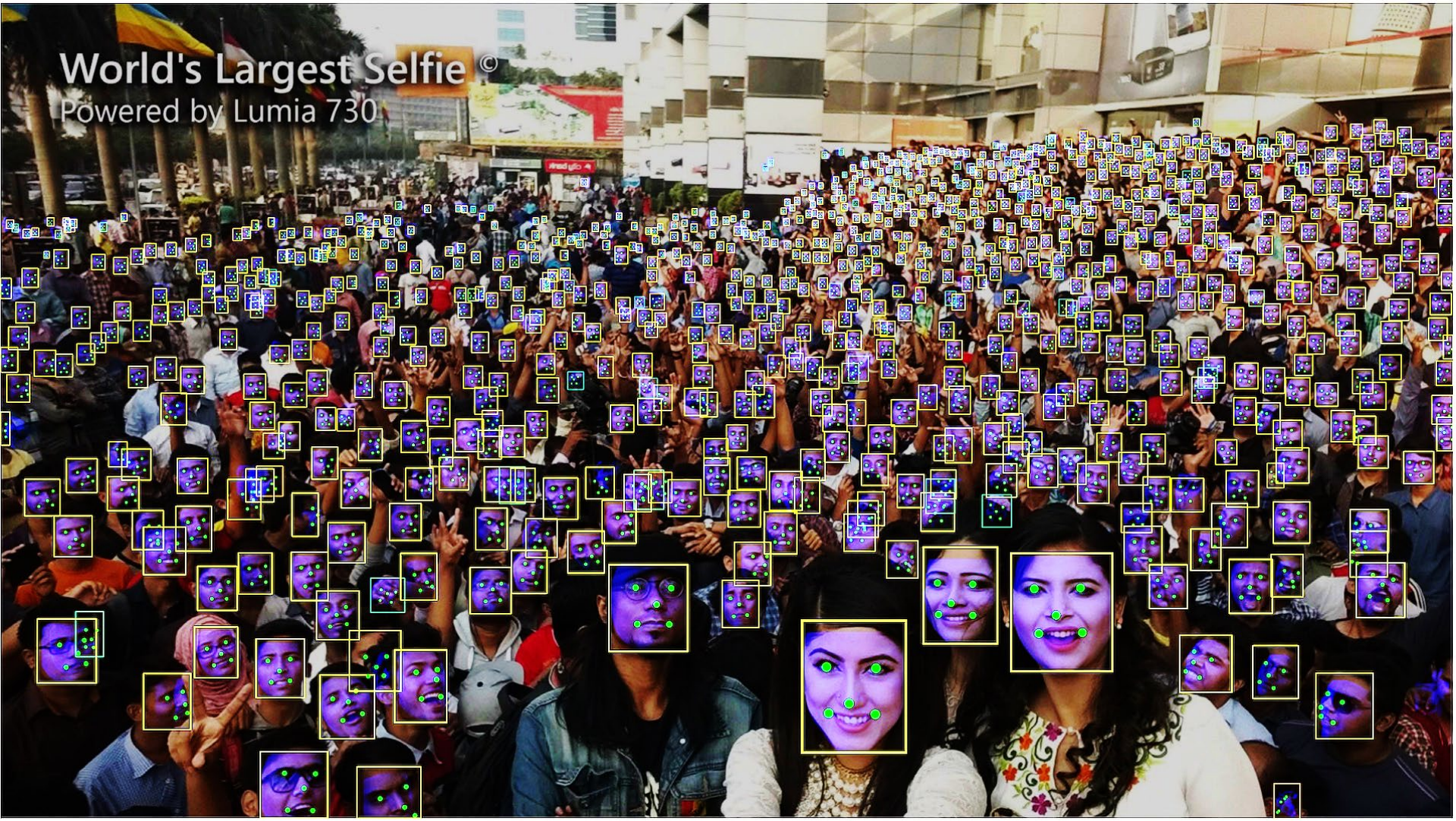


Aplicado tb a otras entidades

World's Largest Selfie ©
Powered by Lumia 730

P. Viola, M. Jones, M. *Rapid object detection using a boosted cascade of simple features* CVPR 2001

V. Kazemi and J. Sullivan. *One Millisecond Face Alignment with an Ensemble of Regression Trees*. CVPR14



J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou. 2019. *RetinaFace: Single-stageDense Face Localisation in the Wild*. 2019

V. Albiero, X. Chen, X. Yin, G. Pang, T. Hassner, *img2pose: Face Alignment and Detection via 6DoF, Face Pose Estimation*, 2020

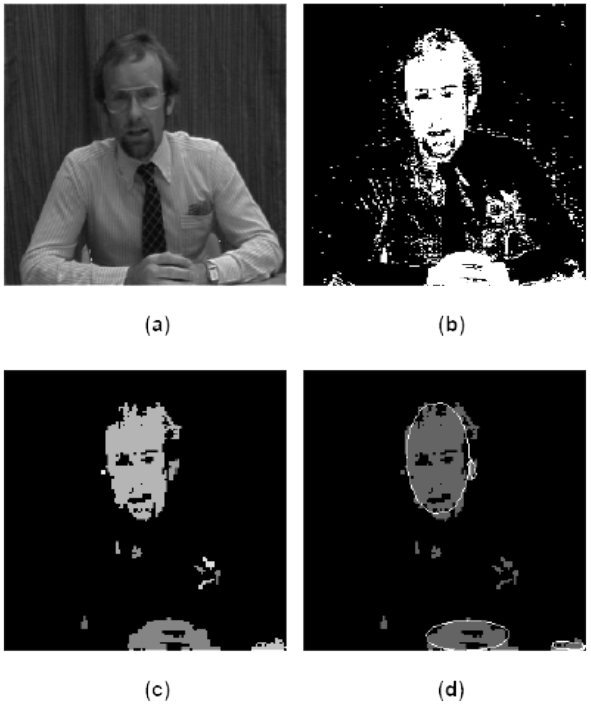
Caras

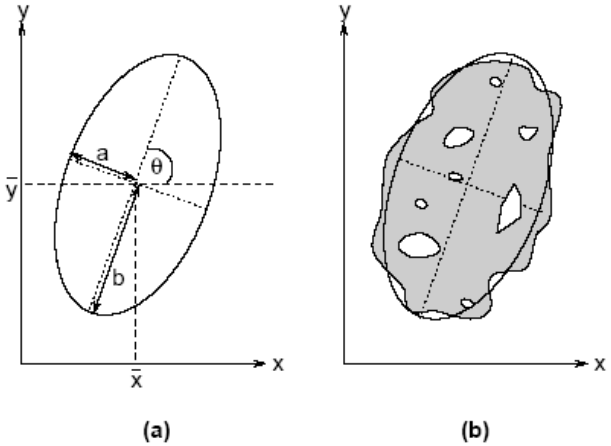
s. XX: heurísticas como color y movimiento, restricciones

Sobottka et al.

Color zonas similares a la piel humana (*inRange*)

Componentes conectadas con **restricciones geométricas**





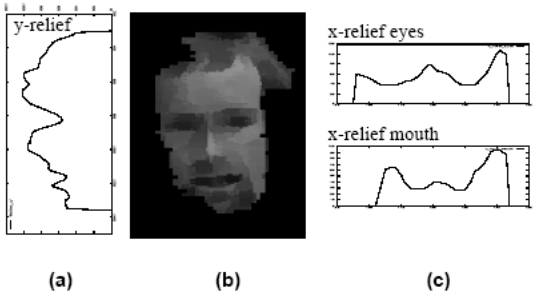
K. Sobottka,I. Pitas. A novel method for automatic face segmentation, face feature extraction and tracking. Signal Processing: Image Communication, 1998.

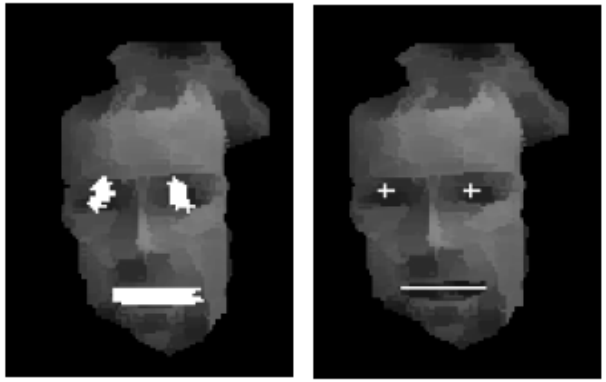
Caras

s. XX: heurísticas como color y movimiento, restricciones

Sobottka et al.

Asumiendo **cara frontal**, estima **ojos** por proyecciones



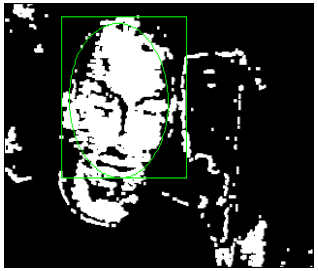


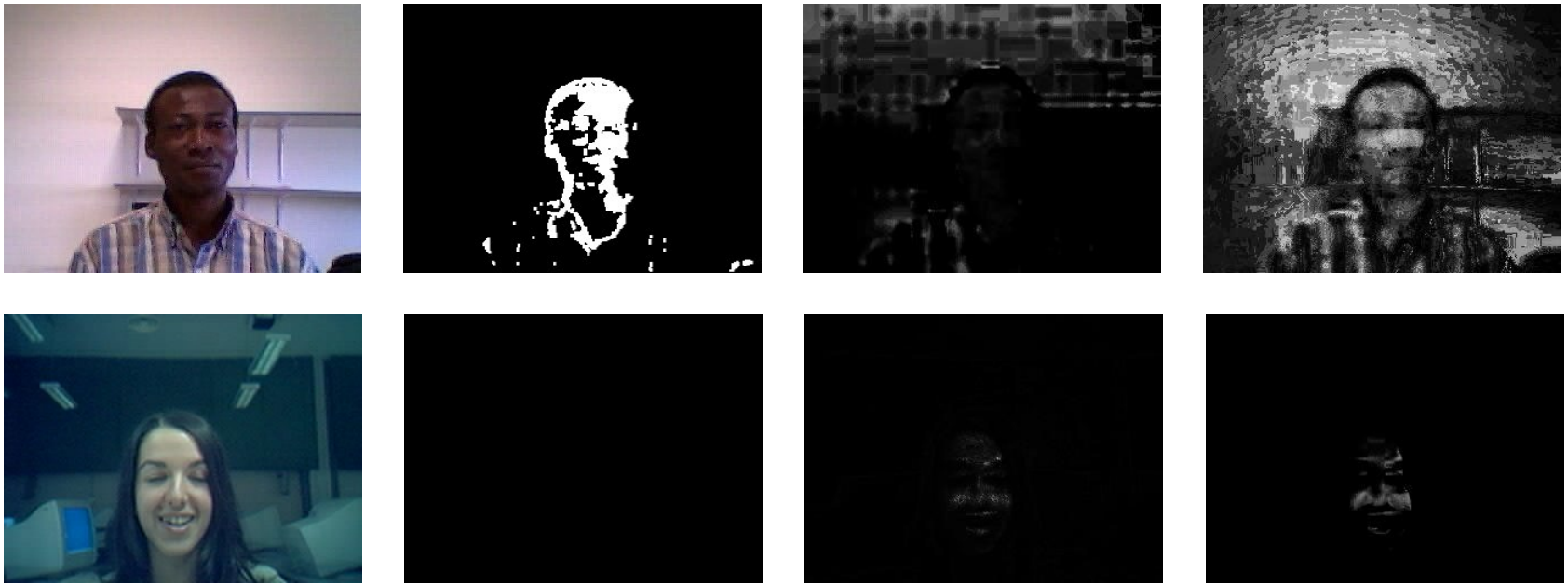
K. Sobottka,I. Pitas. A novel method for automatic face segmentation, face feature extraction and tracking. Signal Processing: Image Communication, 1998.

Caras

s. XX: color y movimiento, restricciones

Técnicas eficientes
No robustas





Esquema simple

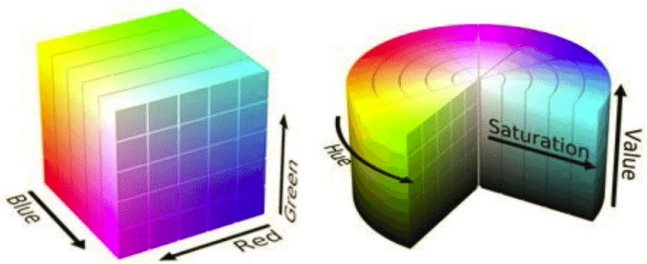
Jones and Rehg [1]

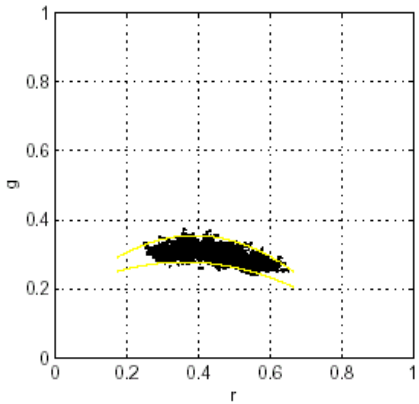
Kruppa [2]

Caras

s. XX: heurísticas como color y movimiento, restricciones

No robustas
Sin modelo general para el color piel

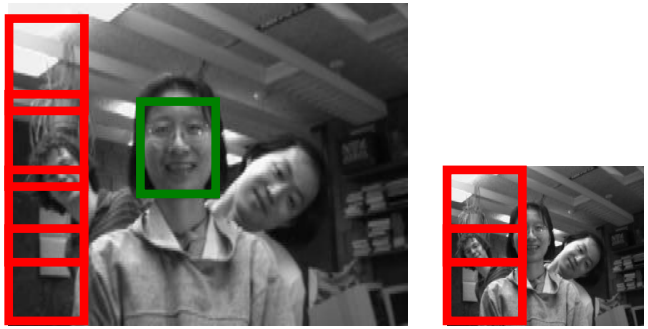




M. Störring et al.. Physics-based modelling of human skin colour under mixed illuminants. Robotics and Autonomous Systems, 2001.

Caras

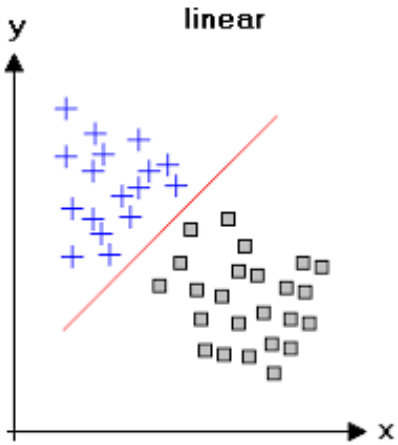
Aprendizaje automático
No heurísticas
Ventana deslizante
Coste temporal del clasificador



Caras

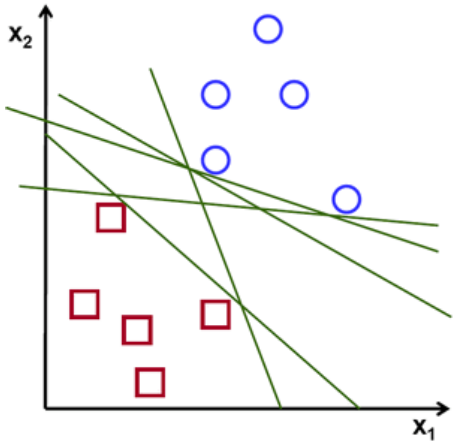
Clasificar cada ventana

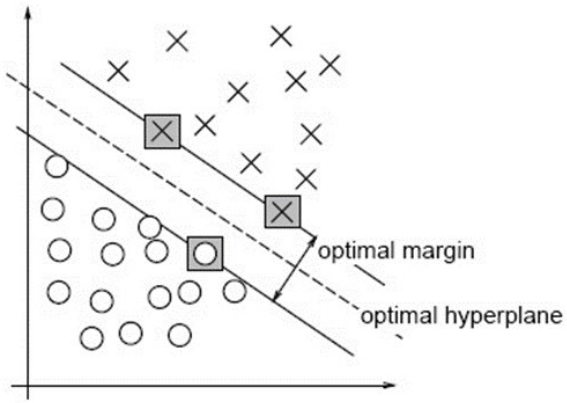
$$x \in \Re^n \,, \quad y \in \{\pm 1\}$$

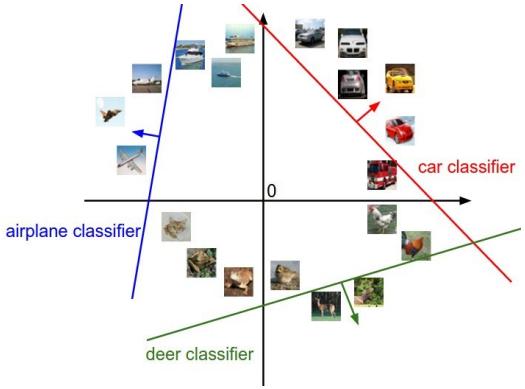


Caras

Clasificar cada ventana







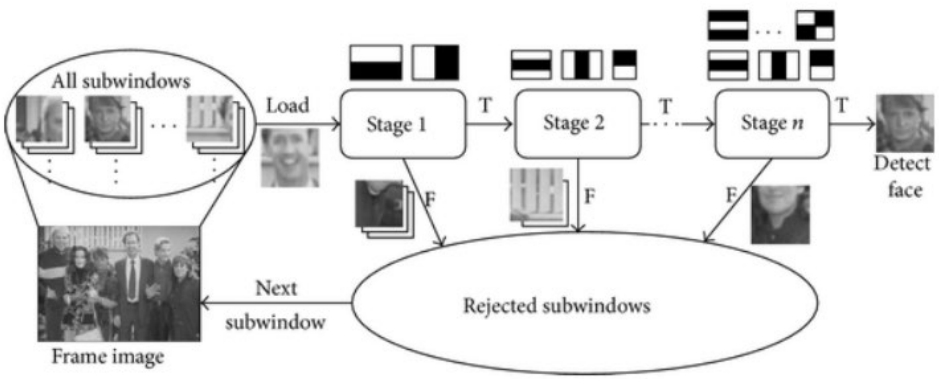
Caras

Rowley et al.
Supervisado
Modela clases **cara** y **no cara** en zonas próximas
Negativas con *bootstrap*
Define mejor la frontera entre ambas clases
Clasificador basado en redes neuronales

Henry A. Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 1998.

Caras

Viola y Jones (>44k citas)
Hito en detección en "*tiempo real*"
Clasificador en cascada, clasificadores débiles
Reduce tiempo en zonas poco prometedoras
Características de cómputo rápido
Entrenamiento supervisado lento
Detección rápida
Imagen integral
Boosting



Cascade structure for Haar classifiers.

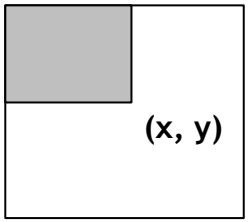
P. Viola and M. J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. In Computer Vision and Pattern Recognition, 2001.

Caras

Viola y Jones

Imagen integral

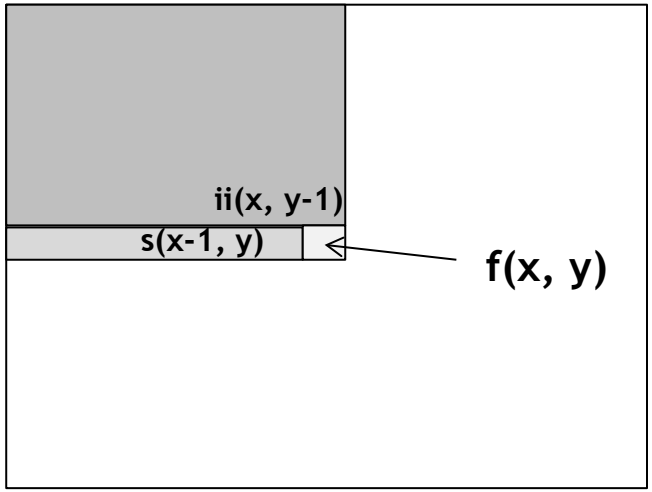
Valor del píxel, suma situados encima y a la izquierda



Cálculo en una pasada:

$$s(x,y) = s(x-1,y) + f(x,y)$$

$$ii(x,y) = ii(x,y-1) + s(x-1,y) + f(x,y)$$



Caras

Viola y Jones

Imagen integral

Cálculo de suma en zona rectangular

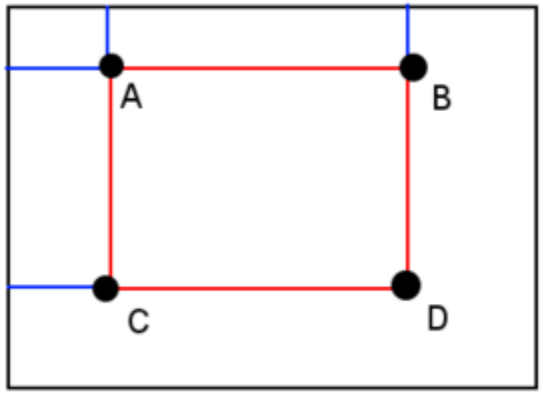
$$R = D - B - C + A$$

Tres sumas

31	2	4	33	5	36
12	26	9	10	29	25
13	17	21	22	20	18
24	23	15	16	14	19
30	8	28	27	11	7
1	35	34	3	32	6

31	33	37	70	75	111
43	71	84	127	161	222
56	101	135	200	254	333
80	148	197	278	346	444
110	186	263	371	450	555
111	222	333	444	555	666

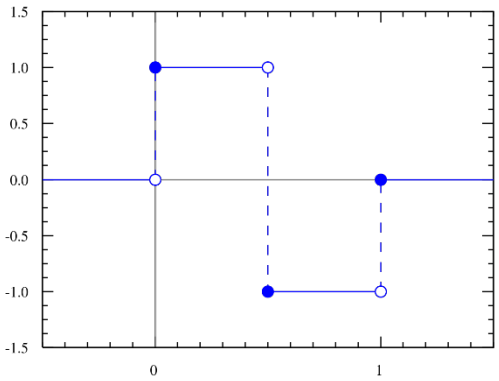
15 + 16 + 14 + 28 + 27 + 11 =
101 + 450 – 254 – 186 = 111



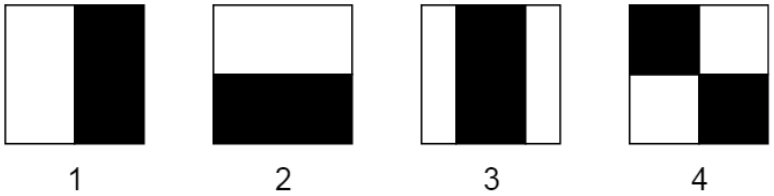
Sum = D - B - C + A

Caras

Viola y Jones
Características
Basadas en Haar *wavelet*



Diferencias de sumas de intensidad en zonas rectangulares
Zonas blancas suman, zonas negras restan
Variaciones tamaño y posición



Caras

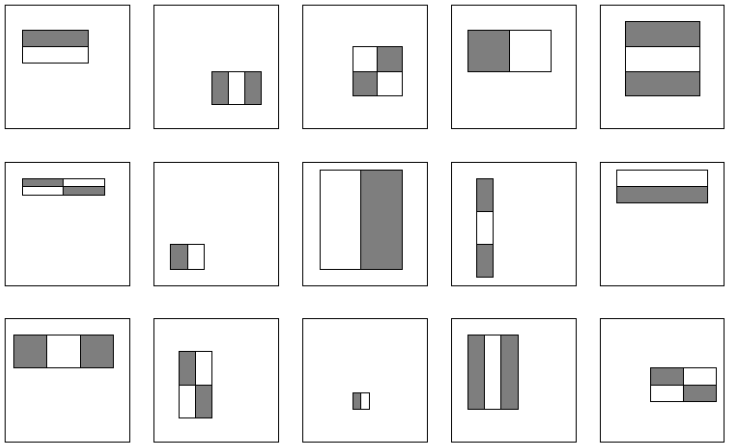
Viola y Jones

Características

Evaluando miles de combinaciones

Para imágenes 24x24 ~160000 combinaciones

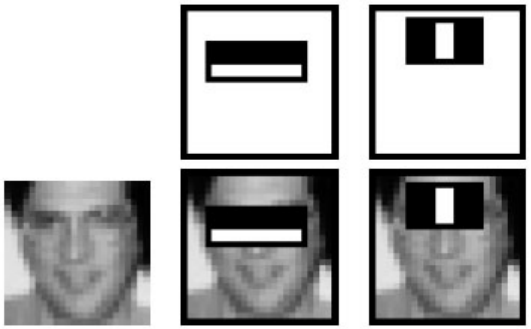
¿Qué característica individual acierta más con las muestras?



Caras

Viola y Jones
Características
Construir un clasificador con un subconjunto

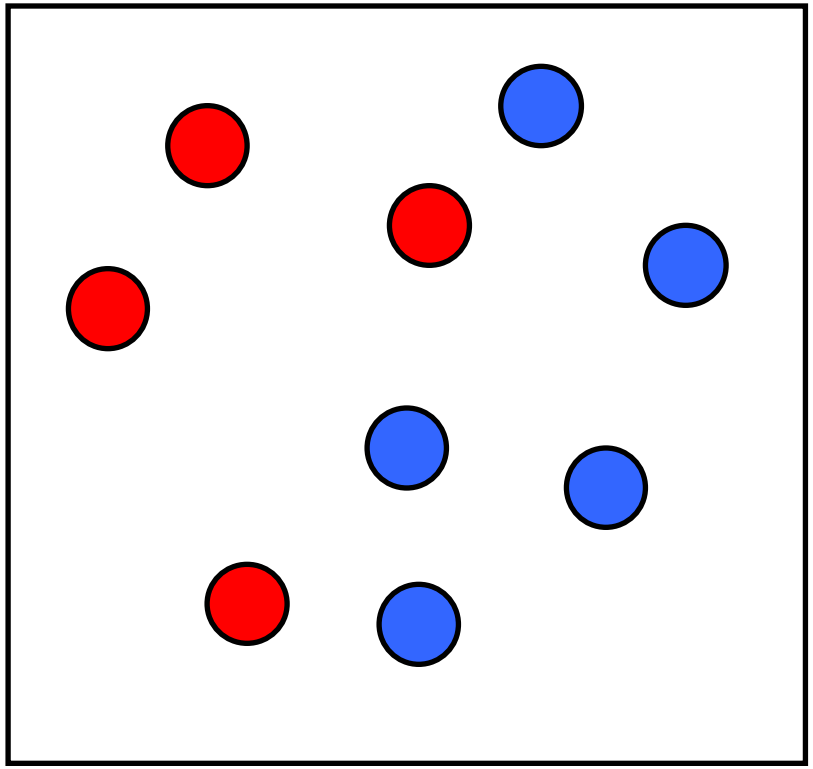
Clasificador débil con mejores características
¿Cuáles funcionan mejor para mi conjunto de datos?
Aplica *boosting*
Iterar modificando peso de muestras mal clasificadas



Caras

Viola y Jones
Boosting

Inicio, peso homogéneo
para todas las muestras



Paul Viola

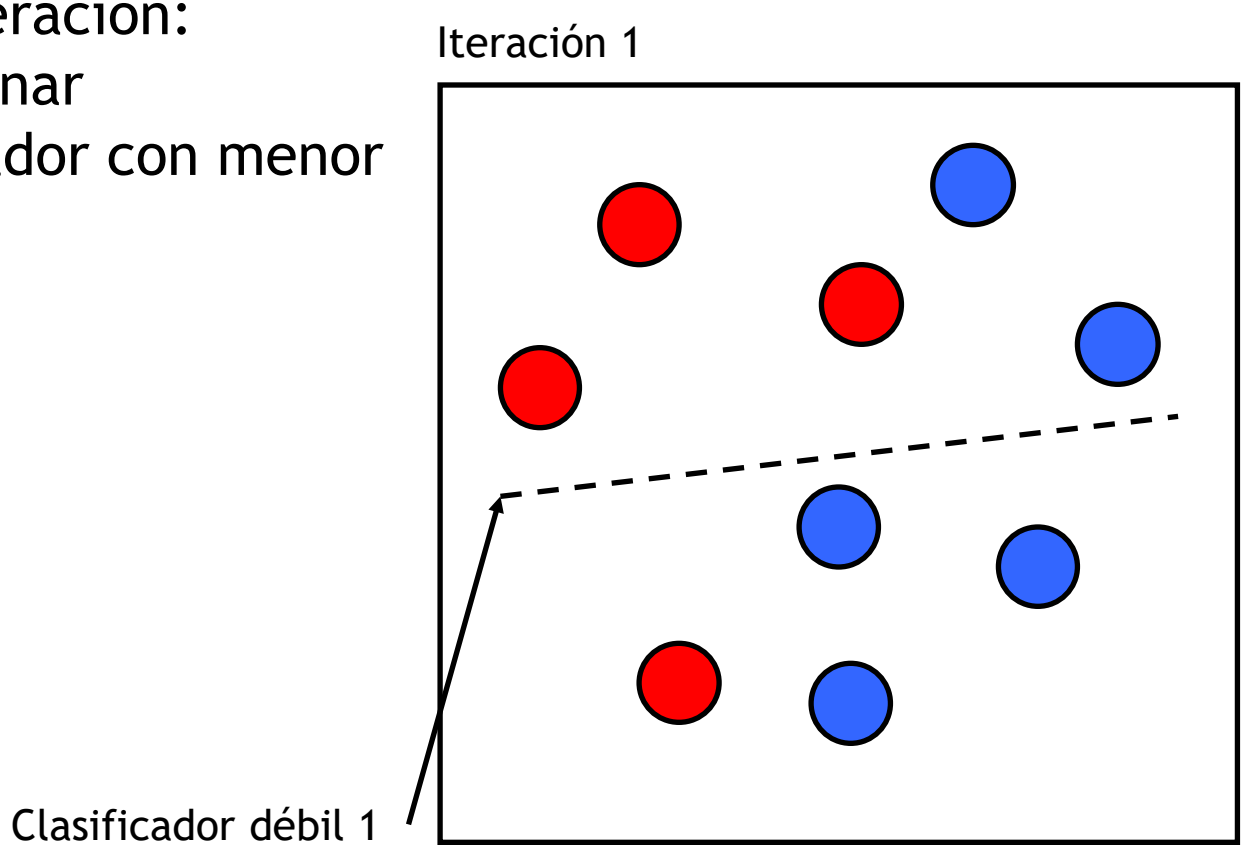
Caras

Viola y Jones

Boosting

En cada iteración:

- Determinar clasificador con menor error



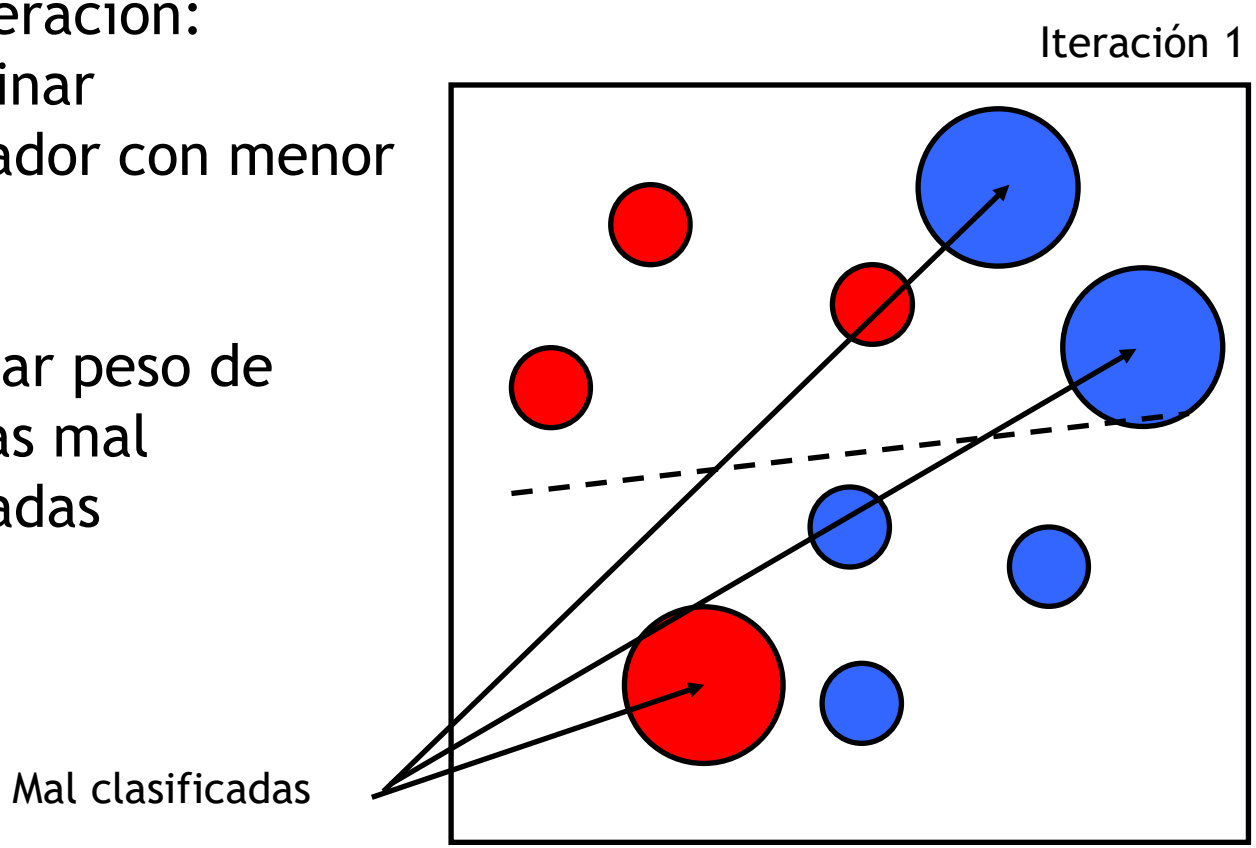
Paul Viola

Caras

Viola y Jones
Boosting

En cada iteración:

- Determinar clasificador con menor error
- Aumentar peso de muestras mal clasificadas



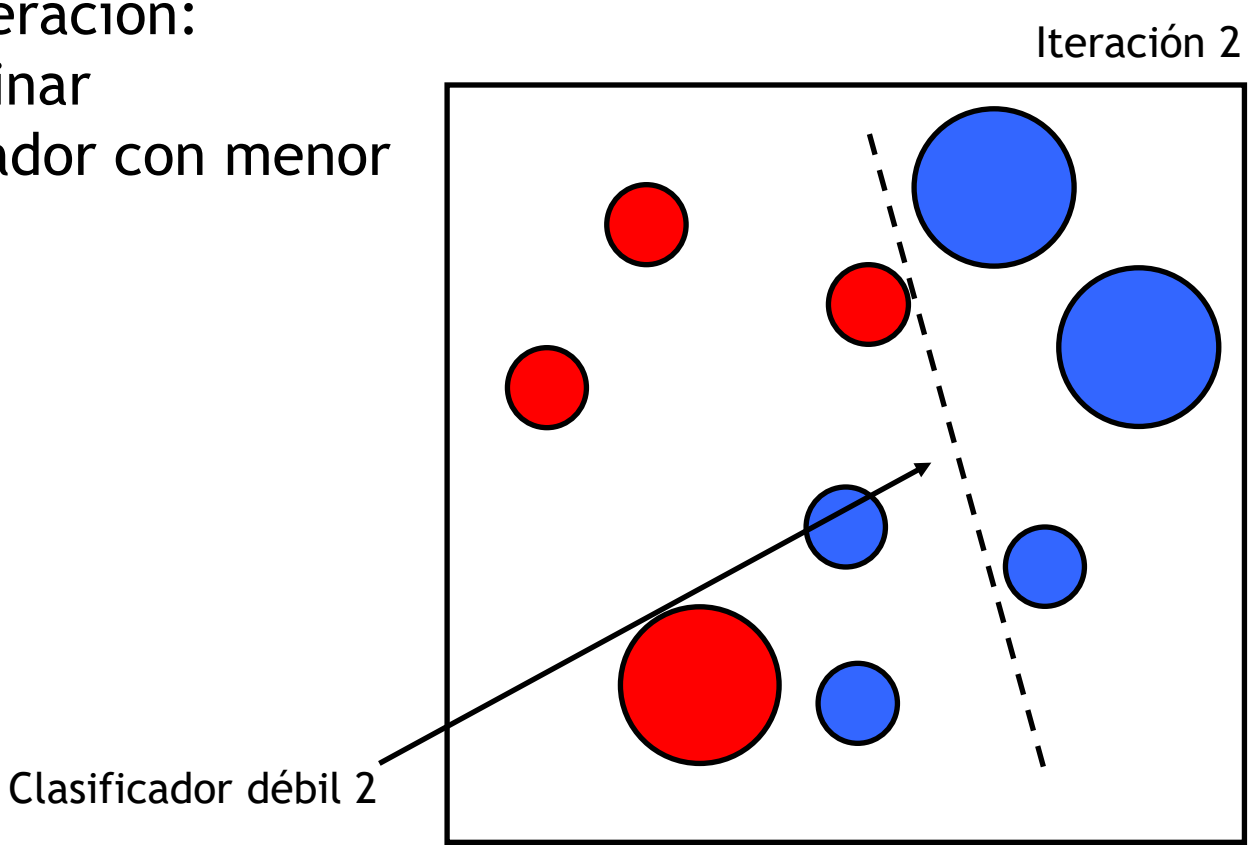
Paul Viola

Caras

Viola y Jones
Boosting

En cada iteración:

- Determinar clasificador con menor error



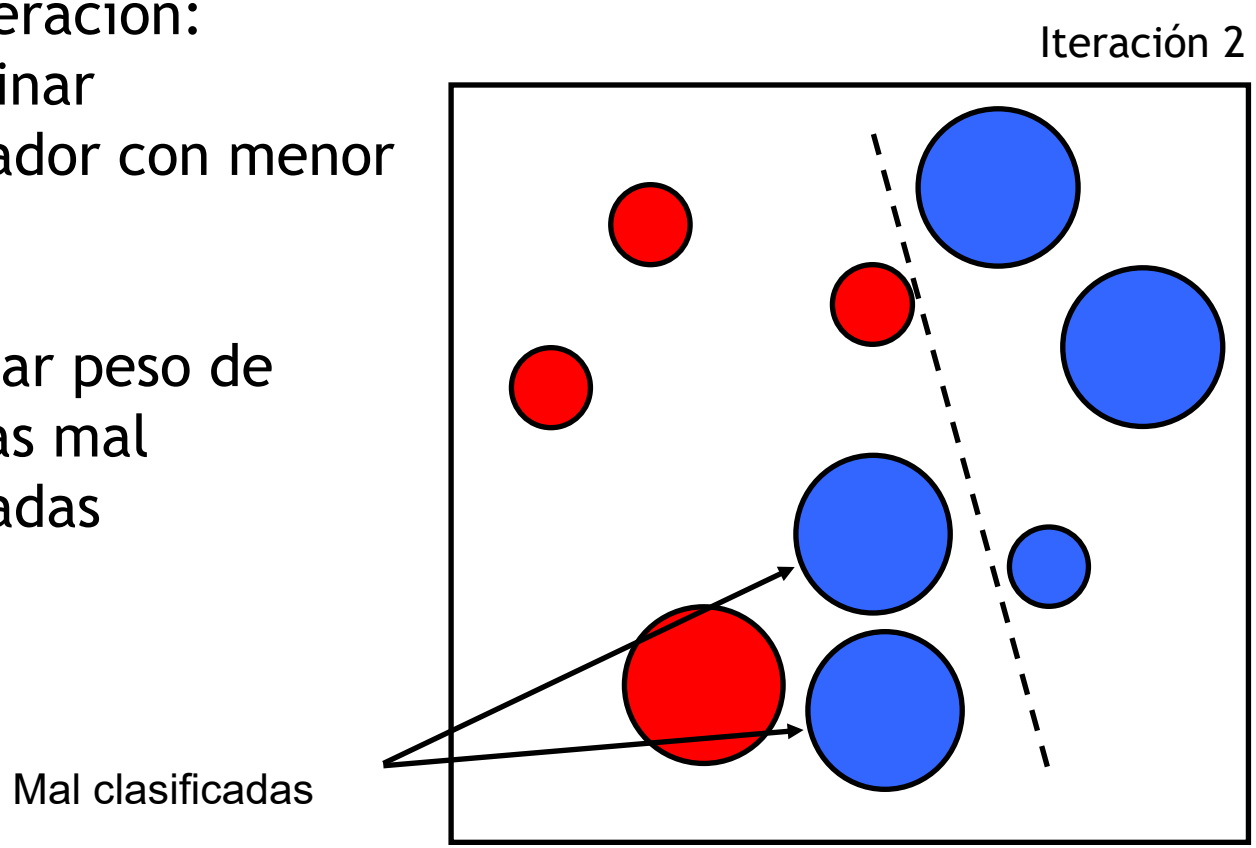
Paul Viola

Caras

Viola y Jones
Boosting

En cada iteración:

- Determinar clasificador con menor error
- Aumentar peso de muestras mal clasificadas



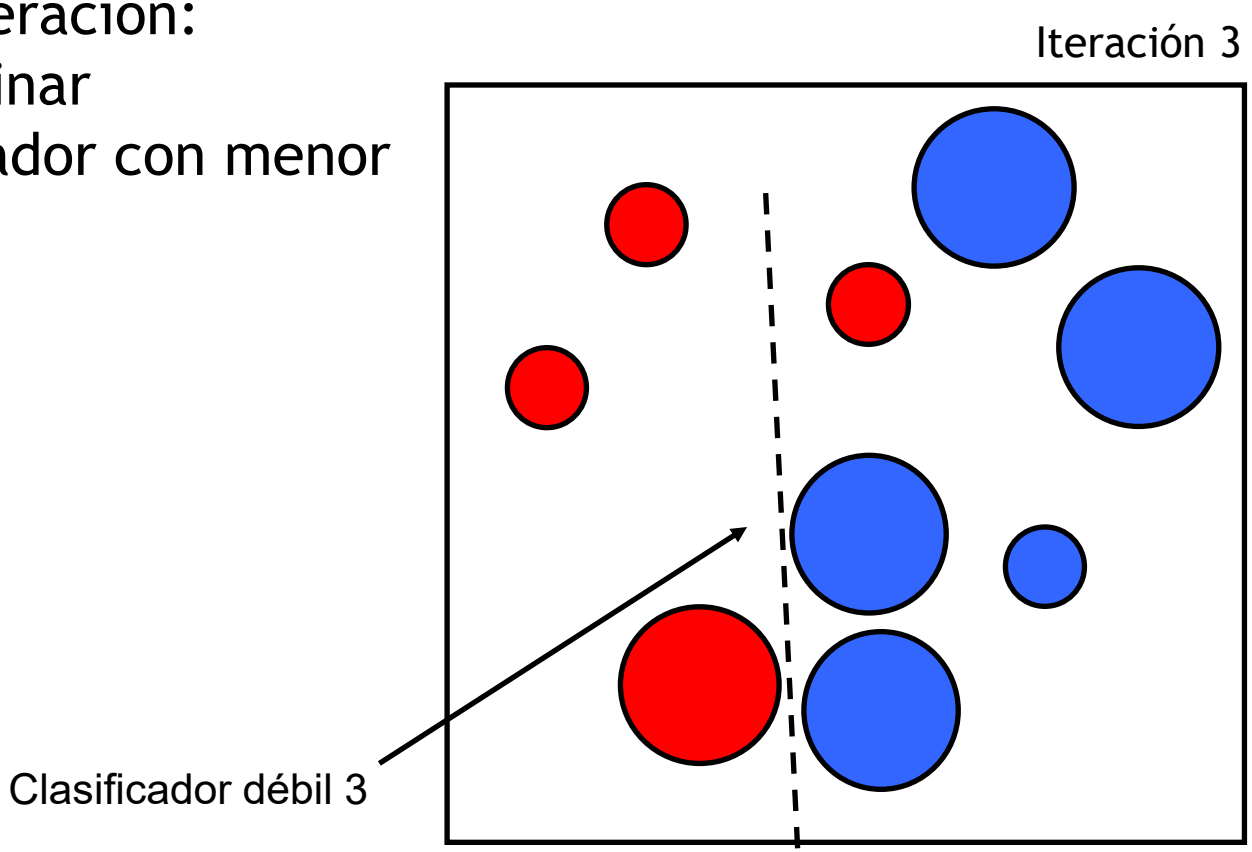
Paul Viola

Caras

Viola y Jones
Boosting

En cada iteración:

- Determinar clasificador con menor error
- …



Paul Viola

Caras

Viola y Jones

Boosting

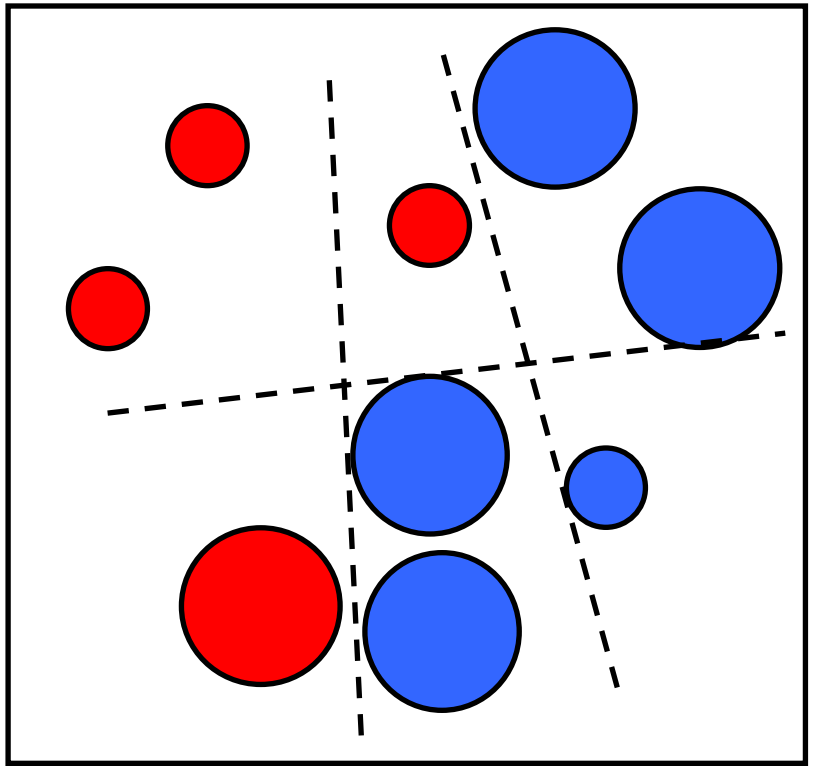
Tras n iteraciones

Combinación lineal (hay variantes) de clasificadores débiles (h_i)

$$h(x) = sgn\left(\sum_{j=1}^{M} \propto_j h_j(x)\right)$$

Cada clasificador débil:

$$h_j(x) = \begin{cases} -s_j & si \quad f_j < \theta_j \\ s_j & en \ \ otro \ caso \end{cases}$$



Paul Viola

Caras

Viola y Jones
Boosting vs. clasificador fuerte (p.e. SVM)

<u>Ventajas</u>
Integra entrenamiento y selección de características
Evaluación rápida
Flexibilidad en elección de esquema de *boosting*

<u>Desventajas</u>
Entrenamiento lento
Necesita muchas muestras
A menudo no tan bueno como SVM (multiclase)

Caras

Viola y Jones

Entrenamiento

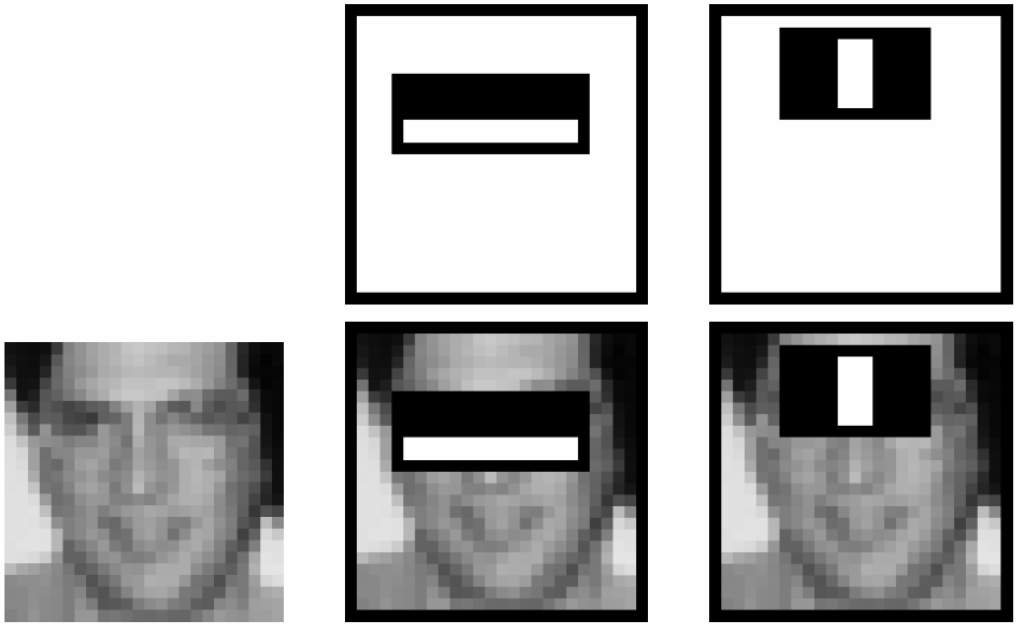
Realiza *boosting* añadiendo clasificadores débiles

Buscando cierta tasa de TPR y FPR en cada etapa

Muestras negativas >> positivas

Caras frontales

100% *recall* y 50% FPR



Caras

Viola y Jones

No basta, FPR alta

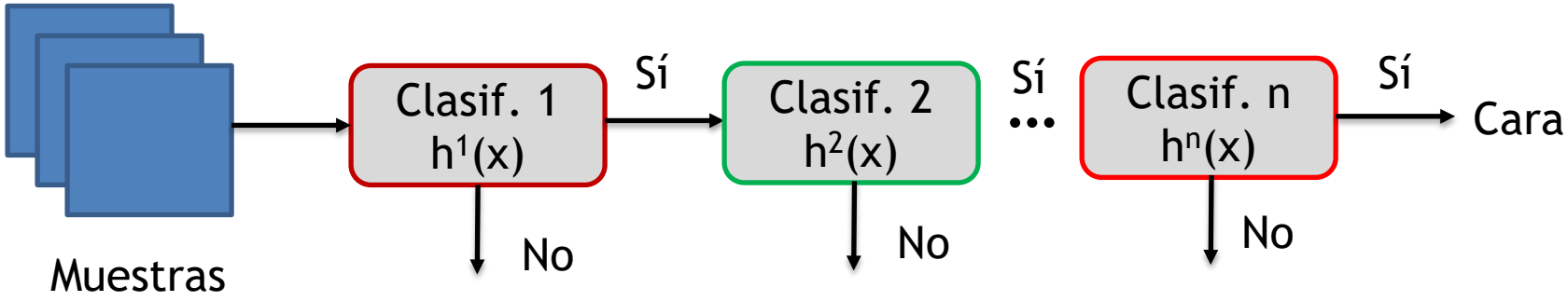
Ventana deslizante -> muchas hipótesis

No hay tantas caras

¿Cómo reducir el tiempo de zonas *fáciles*?

Atención, clasificador en cascada

Alcanzar en cada etapa una cierta tasa de FPR



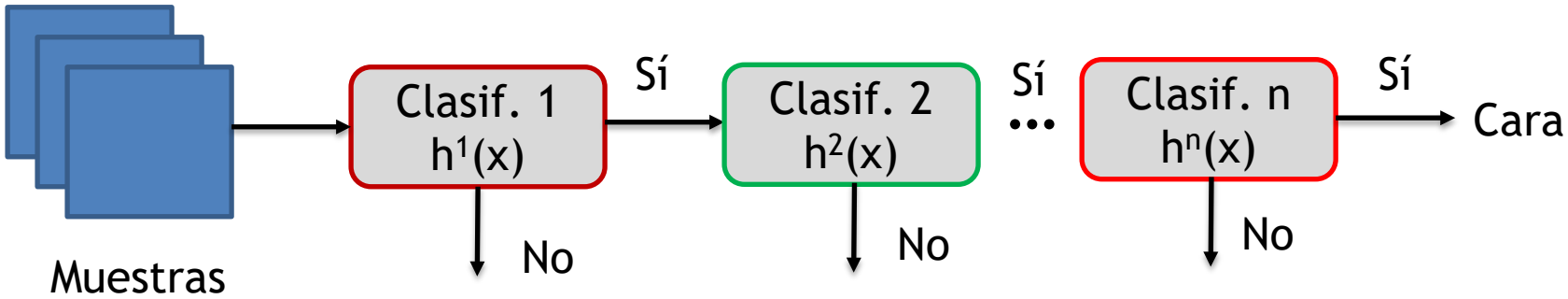
Caras

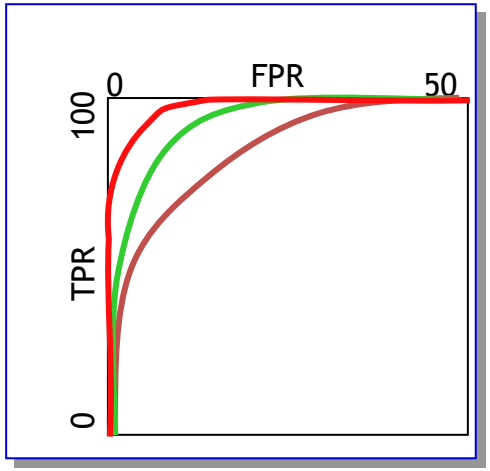
Viola y Jones

Clasificador en cascada

Siguiente nivel entrena con muestras aceptadas por anterior

Cada nivel clasificador más complejo





Caras

Viola y Jones

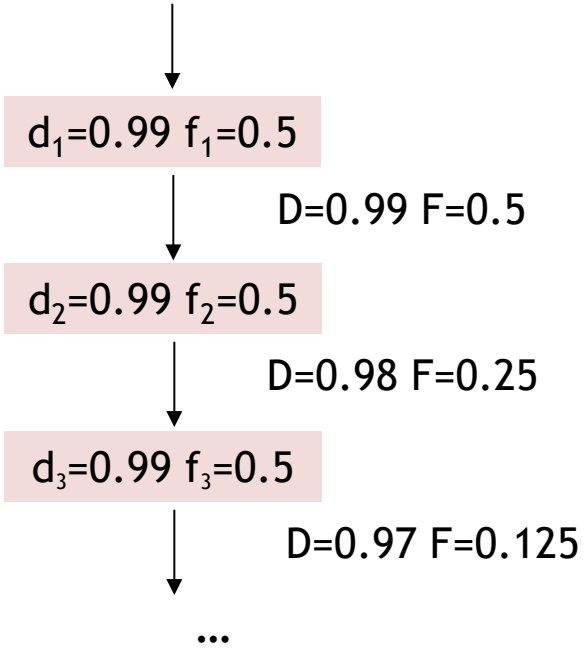
Clasificador en cascada

Siguiente nivel entrena con muestras aceptadas por anterior

Cada nivel clasificador más complejo

Ratios resultantes de detecciones, D, y de falsas detecciones, F, es la combinación de los de cada etapa:

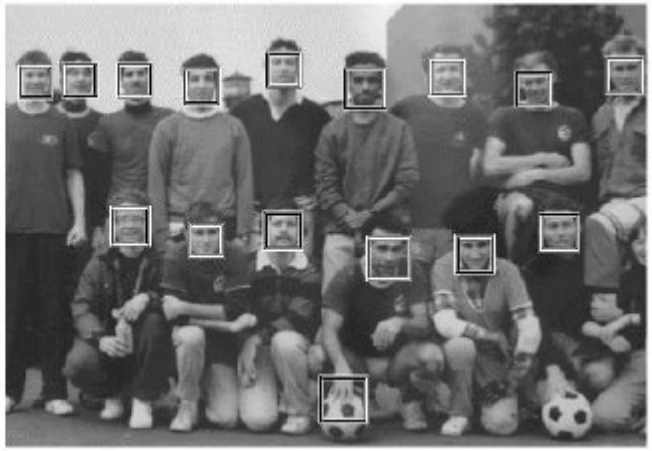
Detector de **20 etapas**, etapas que rechazan un 50% de patrones no faciales, y sólo rechaza 0.1% de caras. Rendimiento general: tasa de **detección de 0.999^{20}**, tasa de **falsas detecciones de 0.5^{20}**



Caras

Viola y Jones

Resultados



Detector \ False detections	10	31	50	65	78	95	167
Viola-Jones	76.1%	88.4%	91.4%	92.0%	92.1%	92.9%	93.9%
Viola-Jones (voting)	81.1%	89.7%	92.1%	93.1%	93.1%	93.2 %	93.7%
Rowley-Baluja-Kanade	83.2%	86.0%	-	-	-	89.2%	90.1%
Schneiderman-Kanade	-	-	-	94.4%	-	-	-
Roth-Yang-Ahuja	-	-	-	-	(94.8%)	-	-

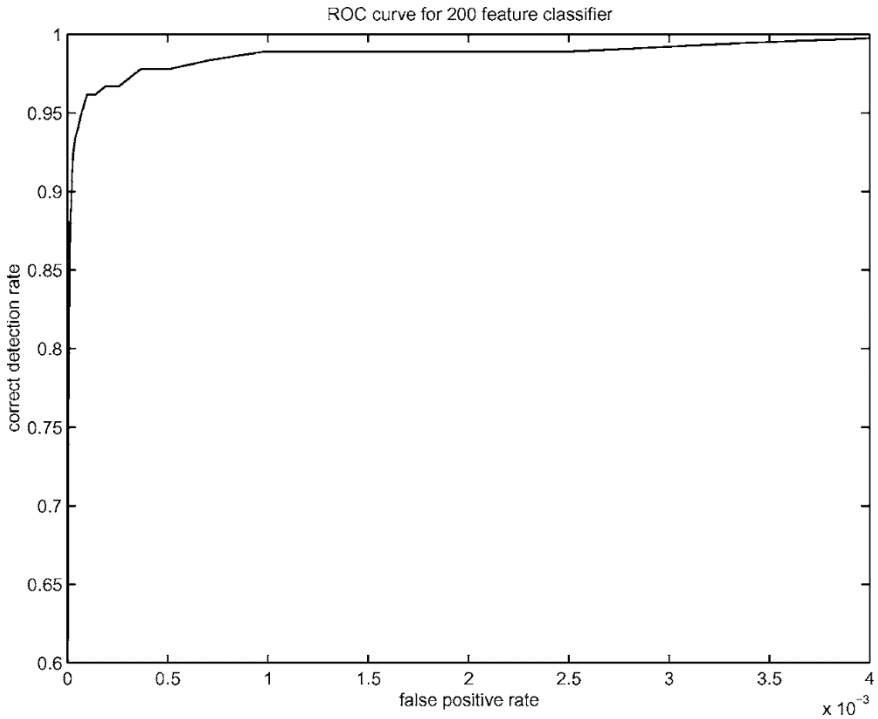
MIT + CMU face dataset

Caras

Viola y Jones

Modelos disponibles en OpenCV desde 2002 (R. Leinhart)

Democratizó acceso a detectores faciales



Caras

Viola y Jones

No suficiente

Falsas detecciones

Caras

Viola y Jones

No suficiente

Insuficientes detecciones

No sólo hay caras frontales, las caras pueden estar rotadas:

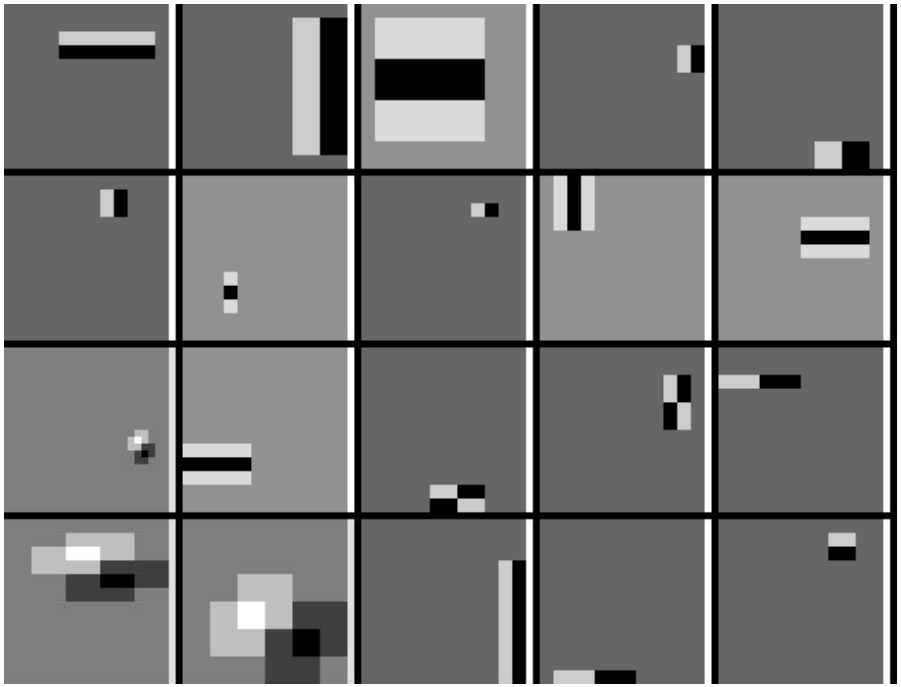
- Rotación fuera del plano (*out-of-plane*)
- Rotación en el plano (*in-plane*)
- Rotación arriba y abajo

Rowley ya proponía usar + clasificadores

Caras

Viola y Jones

Perfil



Caras

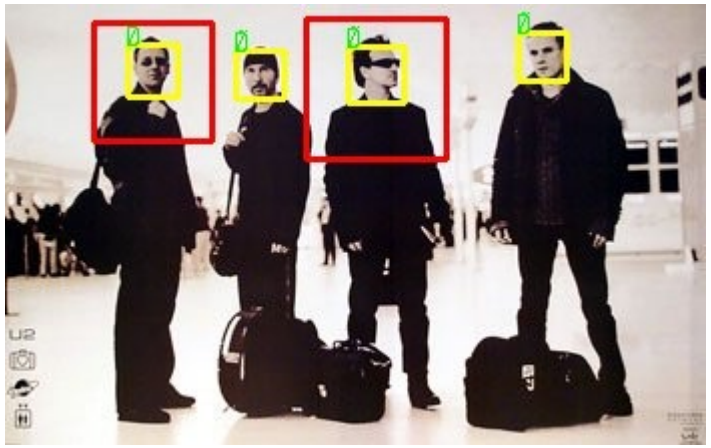
Viola y Jones

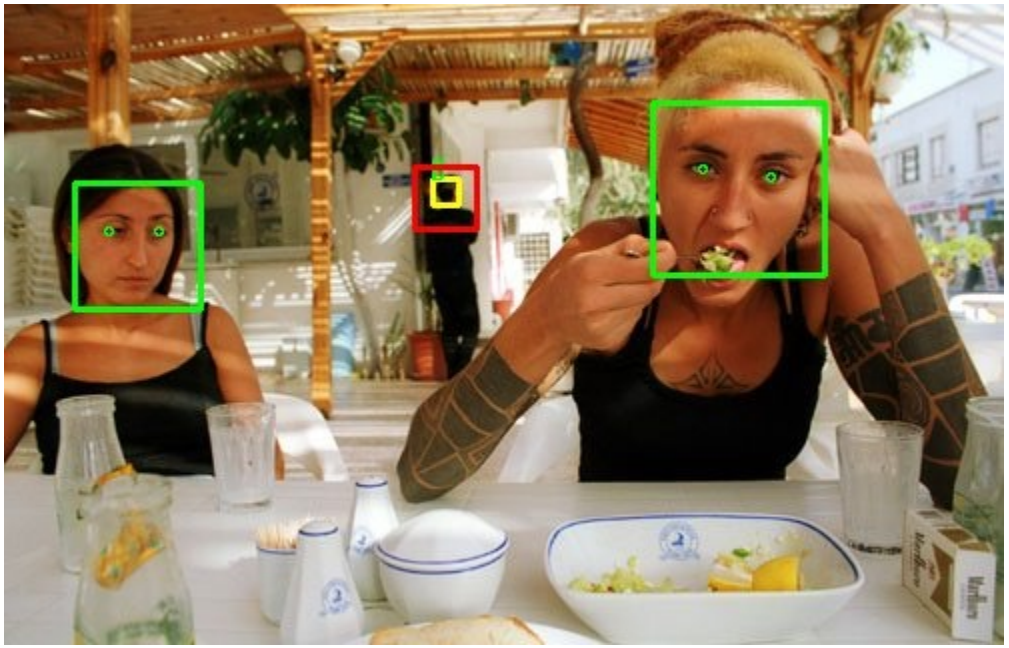
Cabeza y hombros

H. Kruppa, M. Castrillon-Santana, B. Schiele, Fast and Robust Face Finding via Local Context, PETS 2003

Caras

Viola y Jones





Caras

[Batallita]
Tesis sobre detección y seguimiento de caras
Dependencia del color piel
Visito el ETH en 2003 para conocer a Hannes Kruppa

Caras

[Batallita]
Desestimamos color
Viola Jones disponible en OpenCV desde 2002
Aprendemos a entrenar detectores
Incluidos en OpenCV desde 2010 a 2021

- The reference manual has been converted from HTML to LaTeX (by *James Bowman and Caroline Pantofaru*), so there is now:
  - opencv.pdf for reading offline
  - and the online up-to-date documentation (as the result of LaTeX→Sphinx→HTML conversion) available at
    - http://opencv.willowgarage.com/documentation/index.html
- **Samples, misc.:**
- Better eye detector has been contributed by *Shiqi Yu*, **see opencv/data/haarcascades/**[lefteye|righteye]*.xml
- sample LBP (Local Binary Pattern) cascade for the frontal face detection has been created by ***Maria Dimashova_**, see opencv/data/lbpcascades/lbpcascade*frontalface.xml
- Several high-quality body parts and facial feature detectors have been
  - *contributed by Modesto Castrillon-Santana*,
  - see opencv/data/haarcascades/haarcascade_mcs*.xml

Tutorial en pdf (esp)

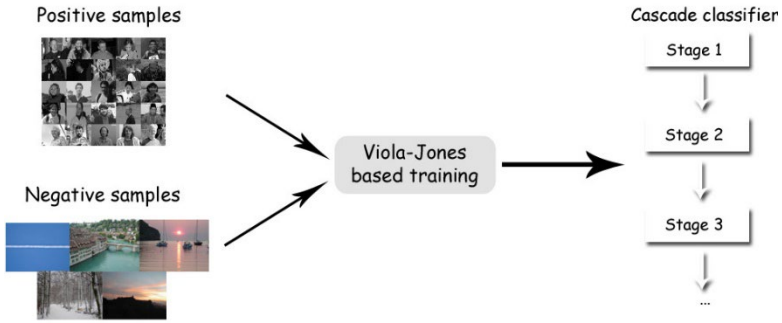
Tutorial haartraining (eng)

Nuestros Clasificadores para OpenCV

Rainer Lienhart and Jochen Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. ICIP2002

Caras

[Batallita]
Entrenando un clasificador VJ
Descargar binarios OpenCV



Dos conjuntos:

1. **Muestras positivas:**
   - ***opencv_createsamples*** genera entrenamiento en formato OpenCV
   - ***vec*** fichero resultante
   - Número de muestras positivas se indica con ***num***
   - ***info*** especifica fichero de texto con imágenes del patrón de interés
   - Muestras positivas, no necesariamente del mismo tamaño, ***opencv_createsamples*** escala en base a valores de ***w*** y ***h***

opencv_createsamples -info positives/trainpos.txt -vec data/positives.vec -num 5000 -w 22 -h 20

Caras

[Batallita]

Entrenando un clasificador VJ

Cada fila de trainpos.txt indica: 1) ruta del fichero, 2) número de objetos en la imagen y 3) sus contenedores

Contendedor: Coordenadas x e y de la esquina superior izquierda, seguidas del ancho y alto del contenedor.

Samples/subject01_centerlight.png 1 126 88 121 137

Samples/subject01_glasses.png 1 116 80 121 137

Caras

[Batallita]
Entrenando un clasificador VJ

Dos conjuntos:

2. **Muestras negativas**:
   - Fichero texto
   - Cada fila del fichero una imagen sin presencia del patrón
   - Requiere ruta absoluta
   - Imágenes grandes estilo fondo de escritorio

C:\Users\usuario\...\samples\000001.jpg

C:\Users\usuario\...\ \samples\01y_fenster.JPG

Caras

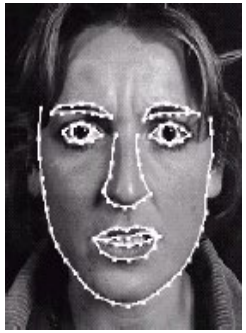
[Batallita]
Código Python

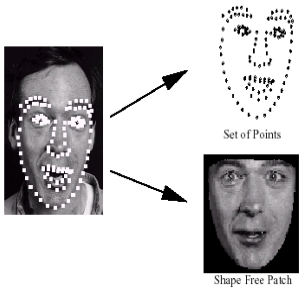
```
import cv2 # OpenCV Library

# Carga clasificador
Cascade = cv2.CascadeClassifier('./Cascades/classifier.xml’)

# captura de la cámara
video_capture = cv2.VideoCapture(0)
while True:
	# Obtiene fotograma
	ret, frame = video_capture.read()
	gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)

	# Detecta objetos y dibuja contenedores
	objects = Cascade.detectMultiScale(gray)
	for (x, y, w, h) in objects:
		container = cv2.rectangle(frame,(x,y),(x+w,y+h),(255,0,0),2)

	# Muestra resultados
	cv2.imshow('Video', frame)
	if cv2.waitKey(1) & 0xFF == ord('q’):
		break

# Libera
video_capture.release()
cv2.destroyAllWindows()
```

Caras

Active Shape/Appearance Models (ASM/AAM)

Modelo estadístico para la forma (contornos) y apariencia





Actualizado: **STASM**

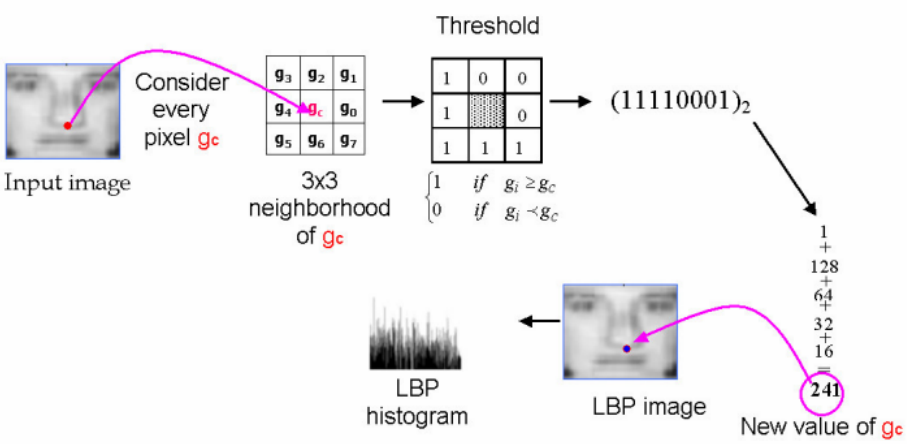
T. Cootes and C. Taylor. Statistical Models of Appearance for Computer Vision, University of Manchester. Report University of Manchester, 2000.
S. Milborrow and F. Nicolls. Active Shape Models with SIFT Descriptors and MARS. VISAPP, 2014

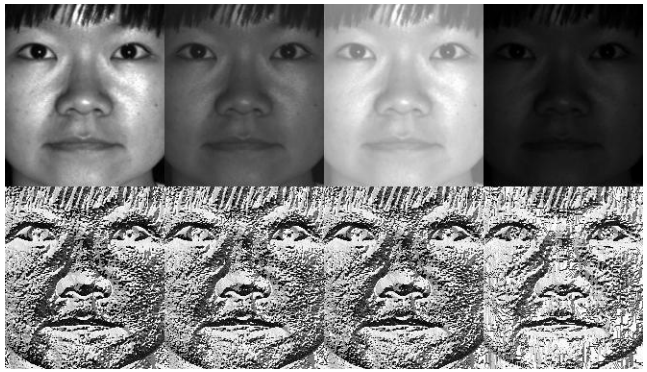
Caras

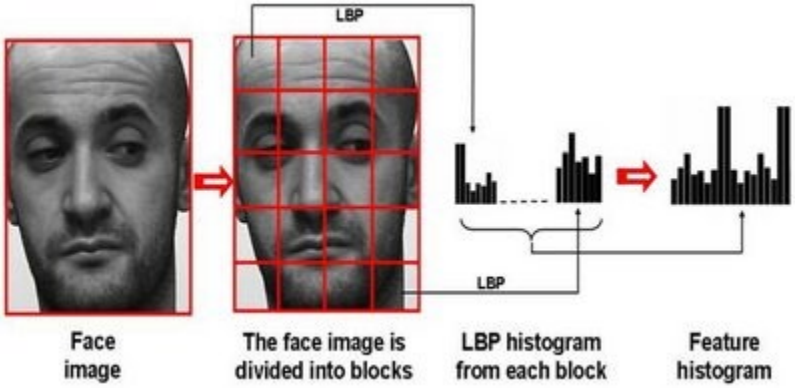
Textura

Local Binary Patterns (LBP)

Descriptores locales invariantes a cambios intensidad



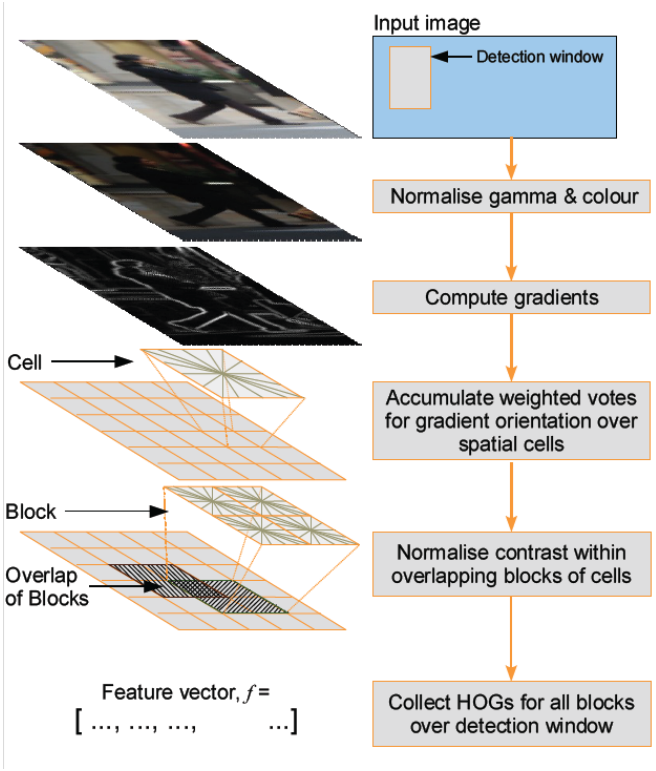


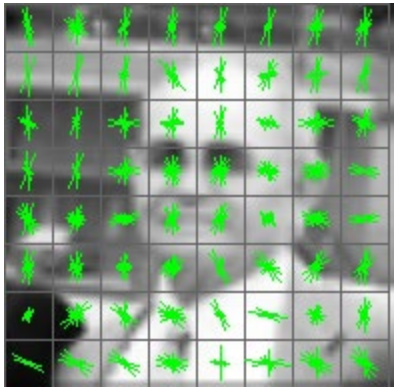


Caras

Textura

Histogram of Oriented Gradients (HOG)





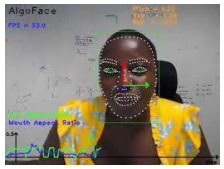
Caras

HOG + SVM

Elementos del rostro

Implementación en dlib

ALGOFACE



V. Kazemi and J. Sullivan. *One Millisecond Face Alignment with an Ensemble of Regression Trees*. CVPR, 2014

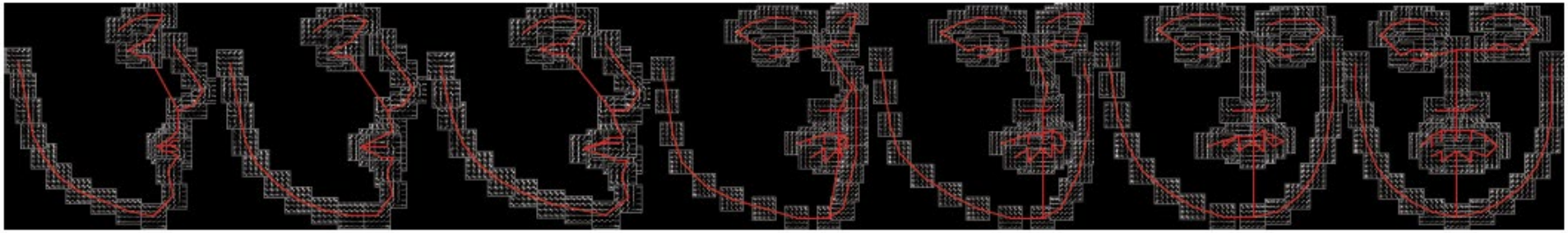
Caras

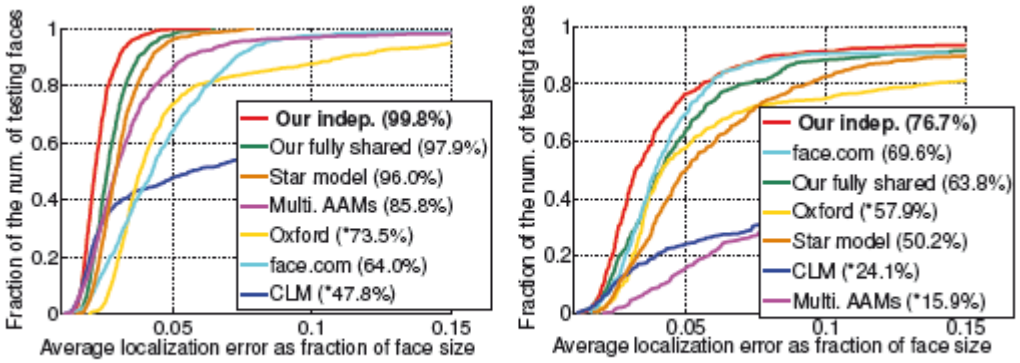
In the wild

Combinaciones de árboles

Muestras anotadas manualmente con todos los puntos

Caras de cierto tamaño

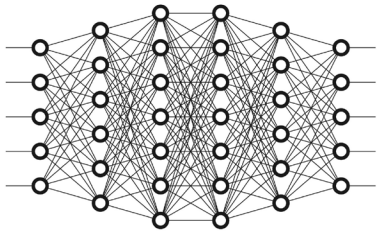




(a) MultiPIE (b) AFW

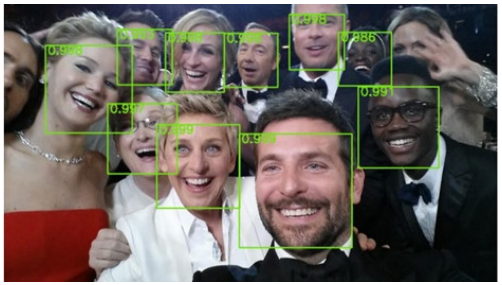
X. Zhu, D. Ramanan, "**Face Detection, Pose Estimation, and Landmark Localization in the Wild**", *Computer Vision and Pattern Recognition*, 2012

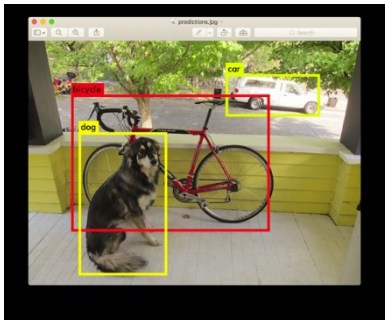
Caras



Convolutional Neural Networks (CNNs)

GPUs



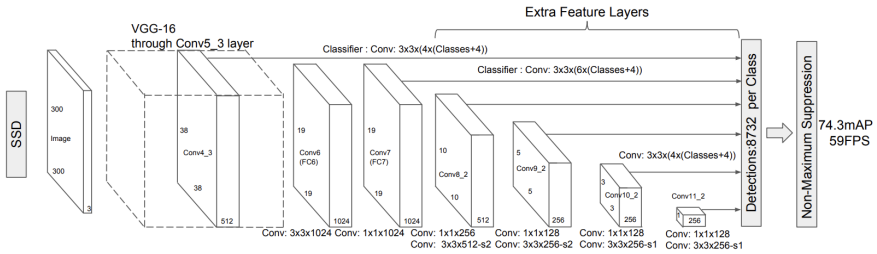


Y. Sun, X. Wang, and X. Tang. Deep Convolutional Network Cascade for Facial Point Detection. *CVPR* 2013
Zhang, K. et al. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 2016

Caras

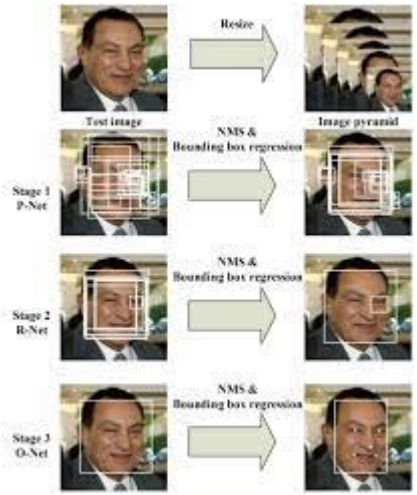
DL con código disponible:

OpenCV basado en SSD [1]



Dlib incluye un detector MMOD [2]

MTCNN [3]

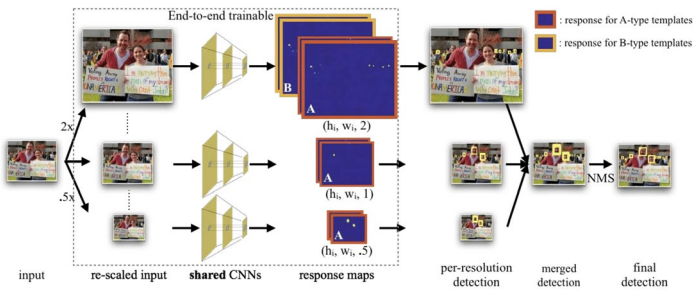


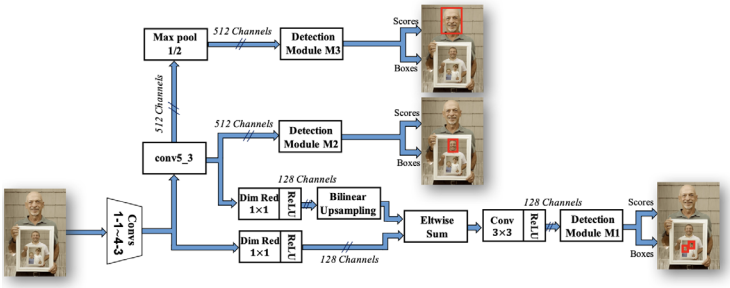
[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg. SSD: Single Shot MultiBox Detector. ECCV, 2016
[2] Davis E. King. Max-Margin Object Detection, arXiv:1502.00046, 2015
[3] Zhang, K. et al. Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 2016

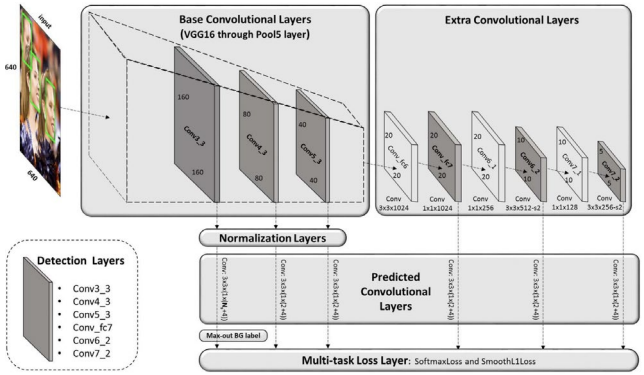
Caras

DL con código disponible:

Tiny face detector [4]
SSH [5]
FaceKit [6]
S3FD [7]
RetinaFace [8]
Img2face [9]







[4] P. Hu and D. Ramanan. Finding tiny faces. CVPR, 2017
[5] M. Najibi, P. Samangouei, R. Chellappa, Larry S. Davis SSH: Single Stage Headless Face Detector ICCV, 2017
[6] X. Shi, S. Shan, M. Kan, S. Wu, X. Chen, Real-Time Rotation-Invariant Face Detection with Progressive Calibration Networks. 2018
[7] Shifeng Zhang, Xiangyu Zhu, Zhen Lei, Hailin Shi, Xiaobo Wang, Stan Z. Li. S3FD: Single Shot Scale-invariant Face Detector ICCV, 2017
[8] J. Deng, J. Guo, Y. Zhou, J. Yu, Irene Kotsia, Stefanos Zafeiriou. RetinaFace: Single-stage Dense Face Localisation in the Wild. arXiv:1905.00641, 2019
[9] . Albiero, X. Chen, X. Yin, G. Pang, T. Hassner, *img2pose: Face Alignment and Detection via 6DoF, Face Pose Estimation*, 2020

Caras

Retinaface

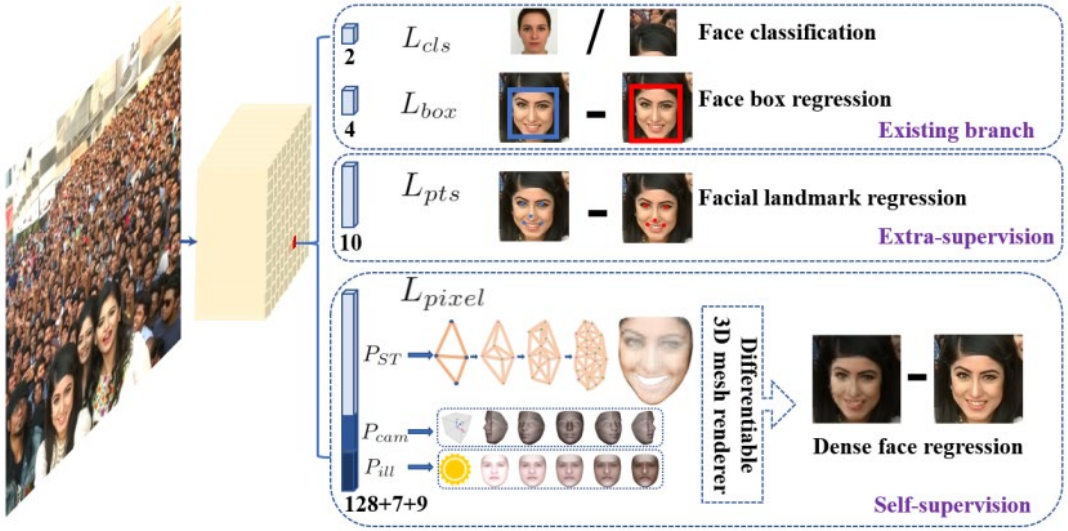
Pirámide: características (ResNet) a cinco escalas

Contexto de la cabeza

Regresión para estimar el contenedor

Multi tarea:

- Detección
- Máscara 2D
- Pose 3D



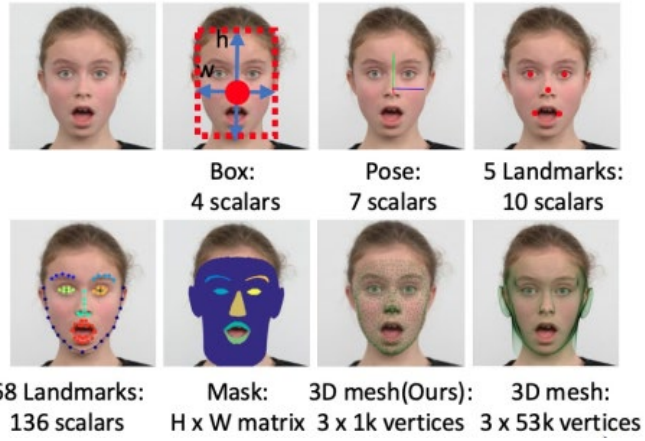
J. Deng, J. Guo, Y. Zhou, J. Yu, I. Kotsia, and S. Zafeiriou. 2019. *RetinaFace: Single-stageDense Face Localisation in the Wild*. 2019

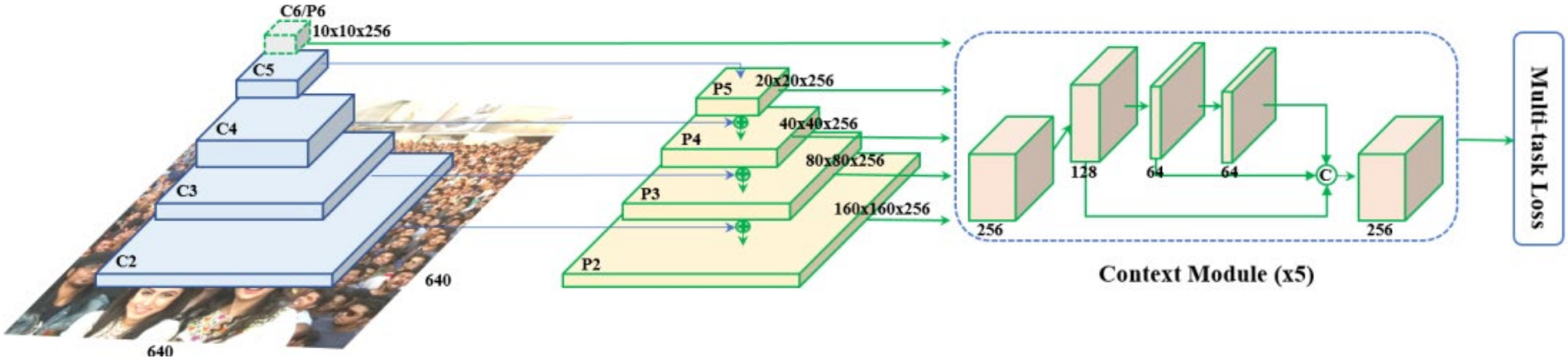
Caras

Retinaface

Multi tarea:

- Detección
- Máscara 2D
- Pose 3D





Caras

Benchmarks

[Face Detection Data Set (FDDB)](#)

[WiderFace](#)

[Unconstrained Face Detection Dataset](#) (UFDD)

[VGGFace2](#)

Caras

Benchmarks

TABLE 1
The performance of face detection models on different versions of Wider-Face dataset.

Method	Easy	Medium	Hard
Faceness [64]	71.3	63.4	34.5
Multiscale Cascade CNN [6]	69.1	66.4	42.4
CMS-RCNN [34]	90.2	87.4	64.3
LFFD [68]	89.6	86.5	77.0
img2pose [84]	90.0	89.1	83.9
S3FD [40]	92.8	91.3	84.4
EXTD [62]	91.2	90.3	85.0
FACE R-FCN [35]	94.3	93.1	87.6
SRN [50]	95.9	94.8	89.6
FDNet [37]	95.0	93.9	89.6
DSFD [51]	96.0	95.3	90.0
PyramidBox [49]	95.6	94.6	90.0
AInnoFace [69]	96.5	95.7	91.2
RetinaFace [53]	-	-	91.4
TinaFace [54]	-	-	92.4

TABLE 2
The performance of face detection models on the FDDB dataset.

Method	AP
CascadeCNN [25]	85.7
Joint-Cascade [28]	86.3
HyperFace [63]	90.1
Faceness [64]	90.3
DP2MFD [50]	90.3
UnitBox [70]	95.1
FaceBoxes [41]	96.0
Faster R-CNN [17]	96.1
DPSSD [85]	96.1
LFFD [68]	97.3
S3FD [40]	98.3
PyramidBox [49]	98.7
SRN [50]	98.8
FACE R-FCN [35]	99.0
DSFD [51]	99.1

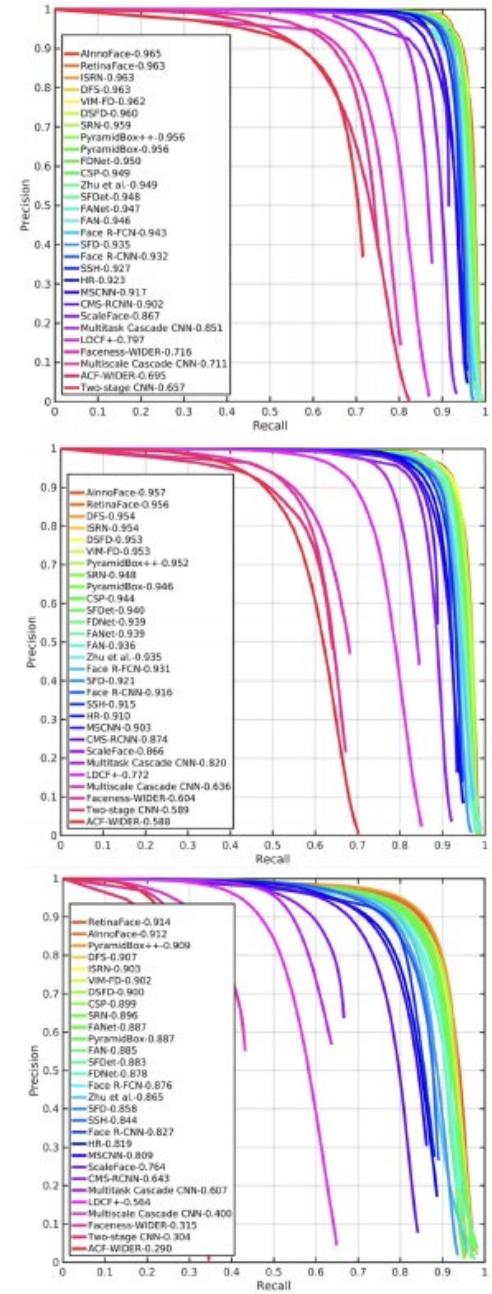
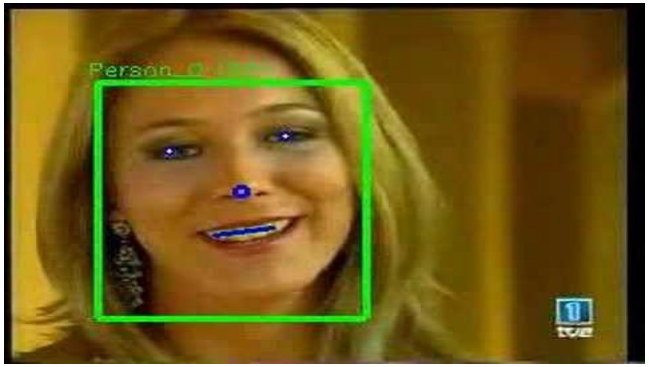


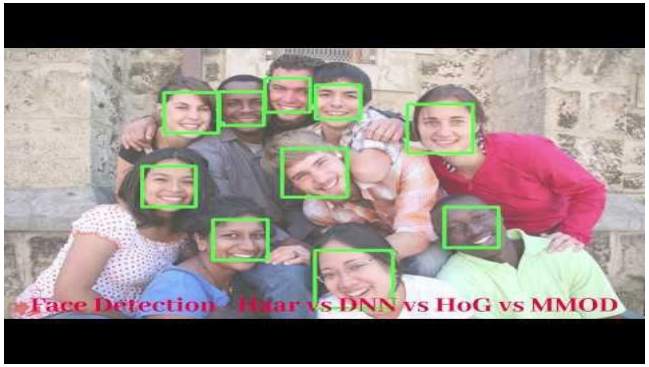
Fig. 35. The top, middle, and bottom figures present the Precision-Recall curves of promising face-detection models on WiderFace Easy/Medium/Hard test sets, respectively. Courtesy of [43].

Going Deeper Into Face Detection: A Survey, 2021

Caras

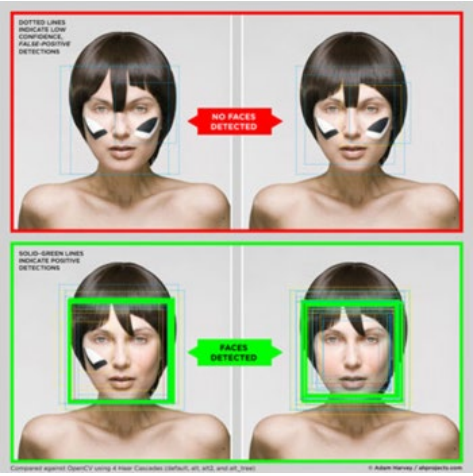
Comparativas de detectores
Detección en vídeo
Tracking by detection

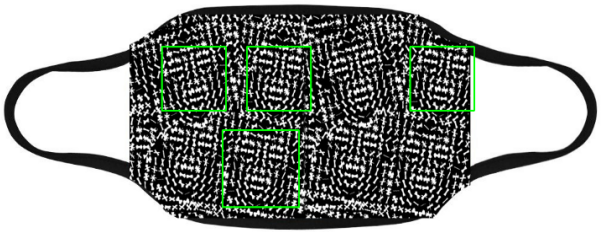




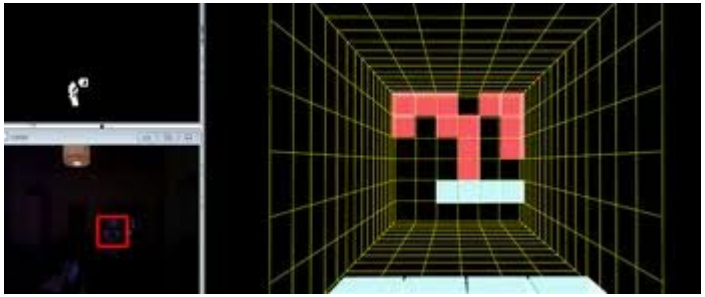
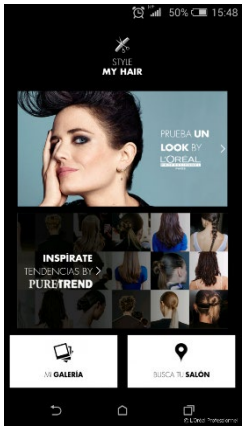
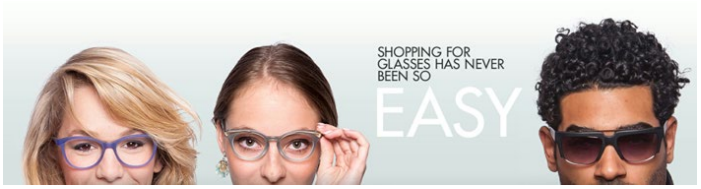
Caras

Confundir al detector
Privacidad





Aplicaciones



NISSAN