Face Detection

Michele Nappi

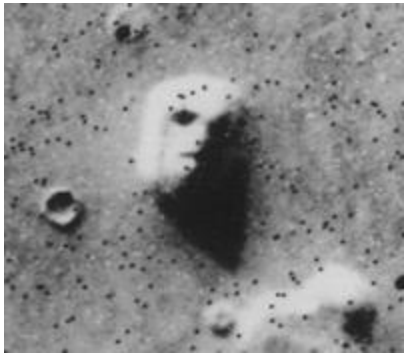
mnappi@unisa.it

biplab.unisa.it

089-963334

Il Volto: Una biometria al limite (?)

La **pareidolia** (dal greco εἴδωλον *èidōlon*, "immagine", col prefisso παρά *parà*, "vicino") è l'illusione subcosciente che tende a ricondurre a forme note oggetti o profili (naturali o artificiali) dalla forma casuale.

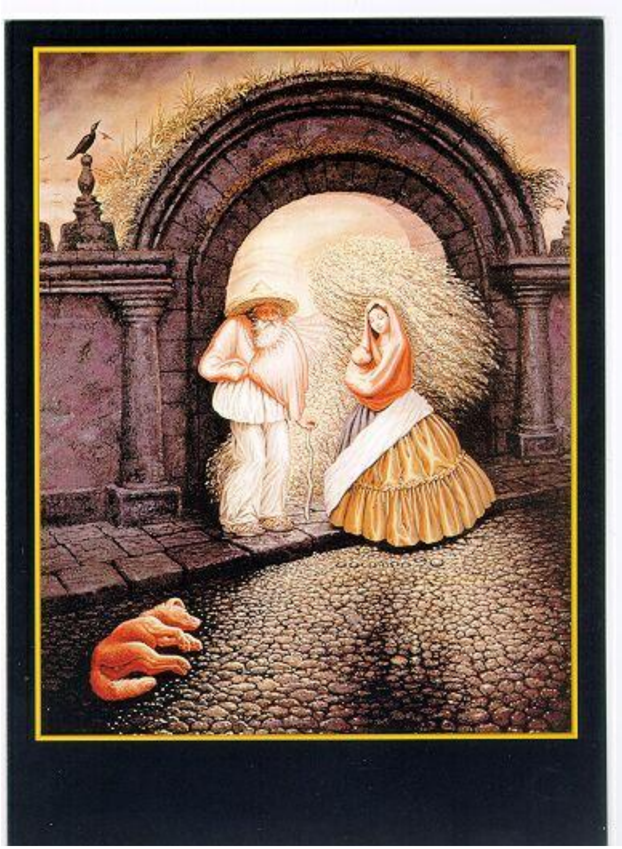


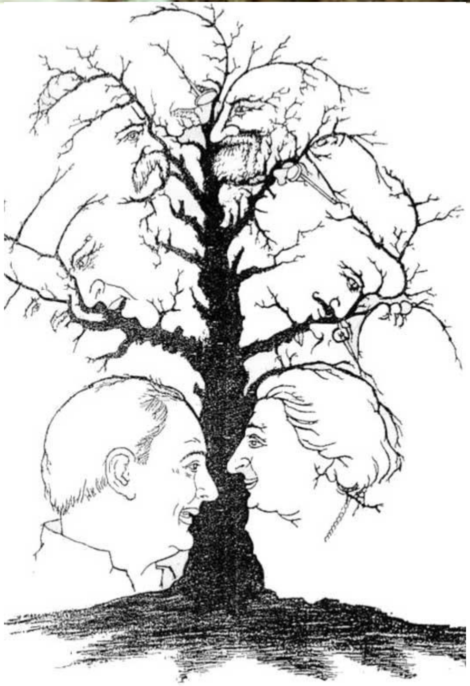
Un celebre caso di pareidolia: il Volto su Marte, una formazione rocciosa marziana ripresa dalla sonda Viking 1, che appare come un volto in particolari condizioni e angolazioni di luce

Il Volto: Una biometria al limite (?)

- La **prosopagnosia** è un deficit percettivo acquisito o congenito del sistema nervoso centrale che impedisce ai soggetti che ne vengono colpiti di riconoscere correttamente i volti delle persone

Esempi di **pareidolia**





Perché il volto

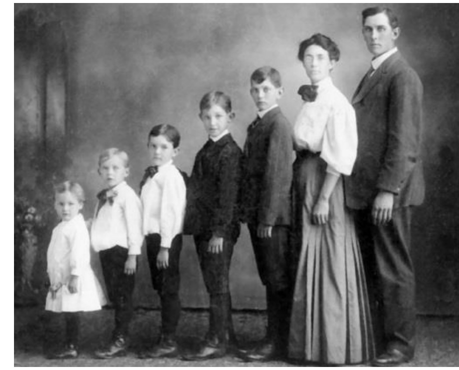
- I due più importanti fattori che sanciscono il successo di una biometria sono Affidabilità e Accettabilità

- Il riconoscimento dell'iride è il sistema più affidabile, ma è anche quello più intrusivo. Le impronte digitali sono più facilmente accettate, ma non applicabili a soggetti non consenzienti.

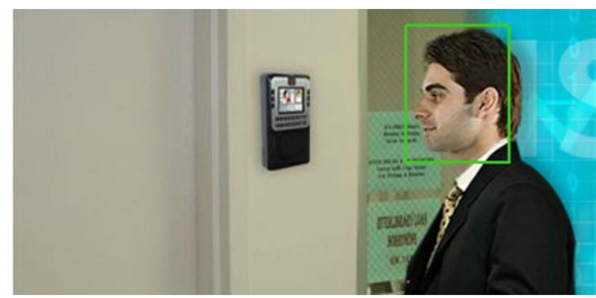
- Il volto ha un'accettabilità molto elevata, mentre l'affidabilità deve essere ancora migliorata

Further advantages …

It is natural to recognize a person from face and it is usual for people to be photographed (e.g., fingerprints may be associated with the bad feeling to be suspected to be a criminal)

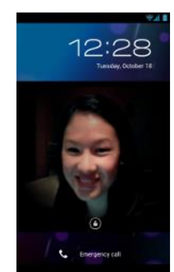


High recognition rate in controlled conditions



Acquisition devices are easy to deploy in the operation environment

It is possible to integrate it in logical access (login)



It is possible to integrate it in remote control applications

Applications

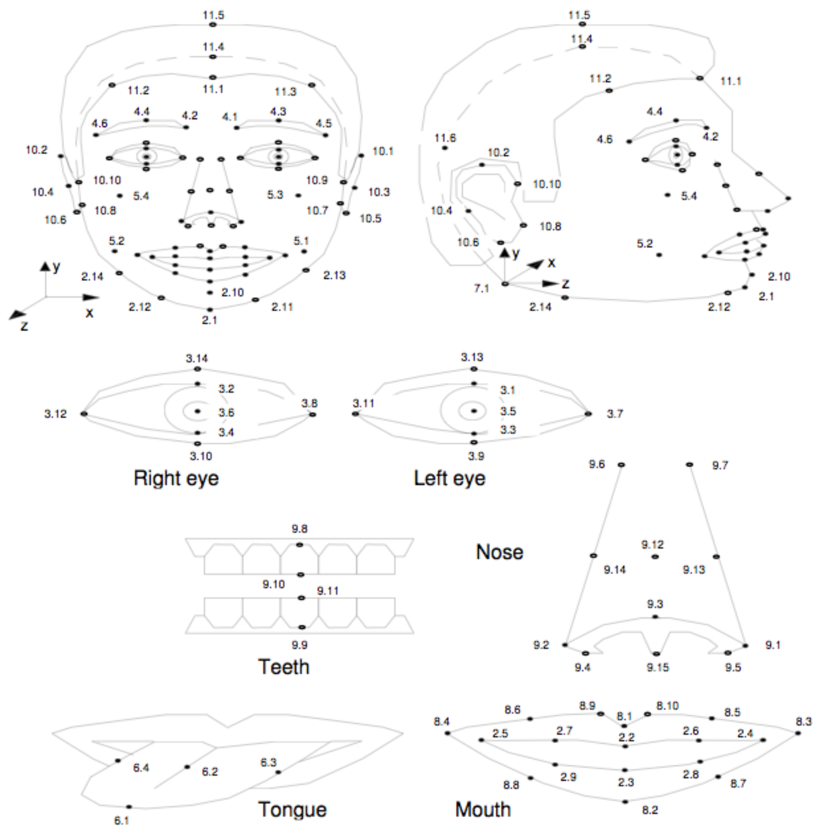
Forensics

iOS face recognition app –

Border control

In a crowd!

… but face is a complex object!



- Feature points affected by FAPs (Facial Animation Parameters)
- Other feature points

MPEG-4 feature points

Comparisons

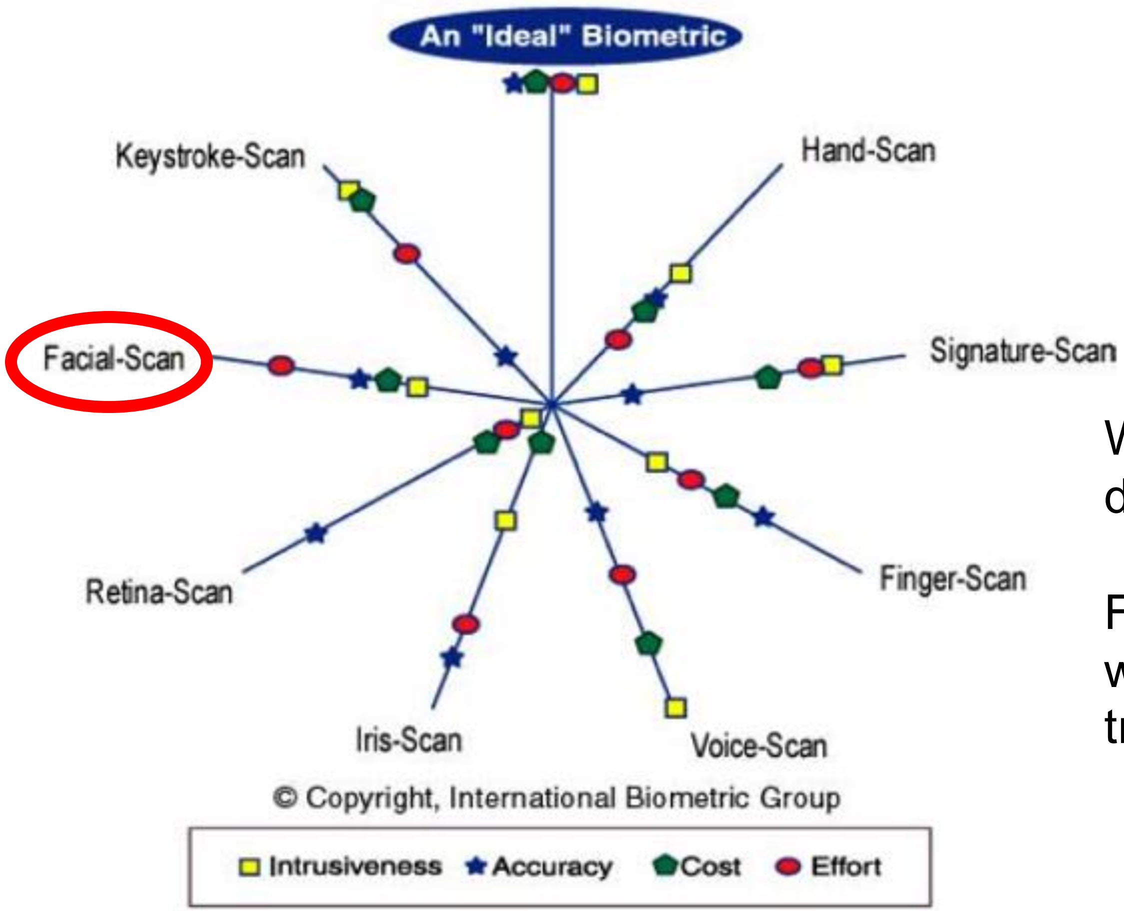
Biometrics	Univer-sality	Unique-ness	Perma-nence	Collect-ability	Perfor-mance	Accept-ability	Circum-vention
Face	H	L	M	H	L	H	L
Fingerprint	M	H	H	M	H	M	H
Hand Geometry	M	M	M	H	M	M	M
Keystroke Dynamics	L	L	L	M	L	M	M
Hand vein	M	M	M	M	M	M	H
Iris	H	H	H	M	H	L	H
Retina	H	H	M	L	H	L	H
Signature	L	L	L	H	L	H	L
Voice	M	L	L	M	L	H	L
Facial Thermogram	H	H	L	H	M	H	H
DNA	H	H	H	L	H	L	L

H=High, M=Medium, L=Low

Classical properties

From http://www.cs.bham.ac.uk/~mdr/teaching/module/security/lectures/biometric.html

Comparisons

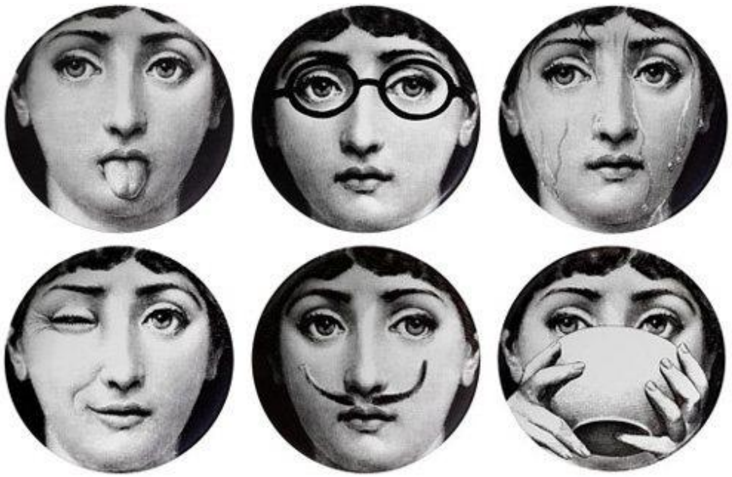


Source: Zephyr Analysis

Which is the most important dimensions?

Face is in acceptable position when compared with other traits

Problems



Intra-personal variations. Piero Fornasetti's template of Cavalieri's face gave birth to hundreds of variations.

http://www.artistdaily.com/blogs/artistdaily/archive/2013/04/22/350-ways-of-drawing-faces.aspx

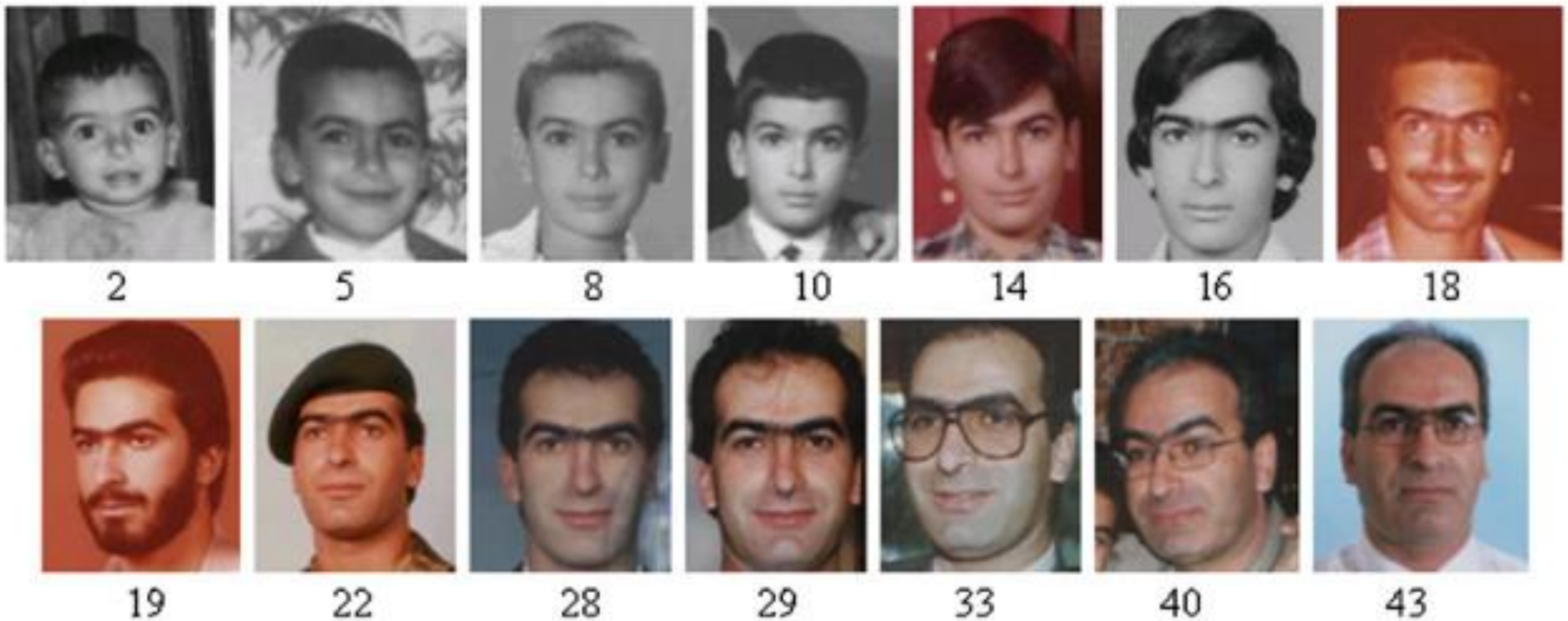
Inter-personal similarities. Bradley Cooper & Hrithik Roshan

http://www.mensxp.com/entertainment/gossip/7356-have-we-met-before-bollywoodhollywood-lookalikes.html

In detail: PIE (Pose, Illumination, Expression) variations

http://www.csee.wvu.edu/~gidoretto/courses/2011-fall-cp/assignments/final_project/results/cao_chen/index.html

In detail: PIE (Pose, Illumination, Expression) variations + Ageing = A-PIE

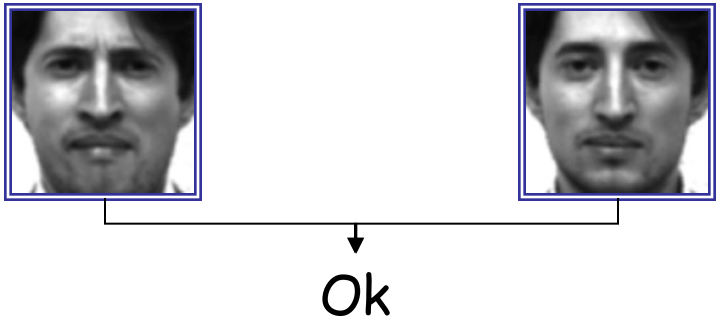


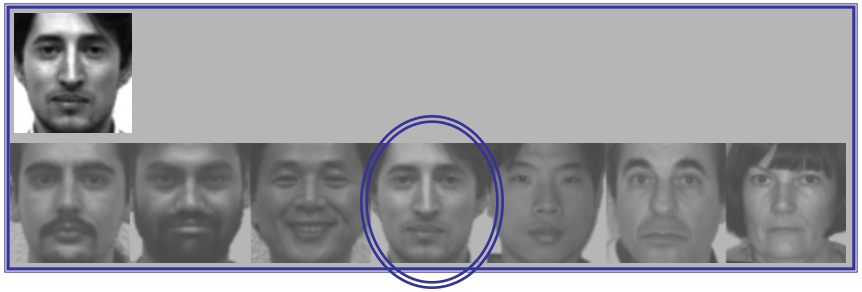
(a) FG-NET

(b) MORPH

Verifica e Riconoscimento

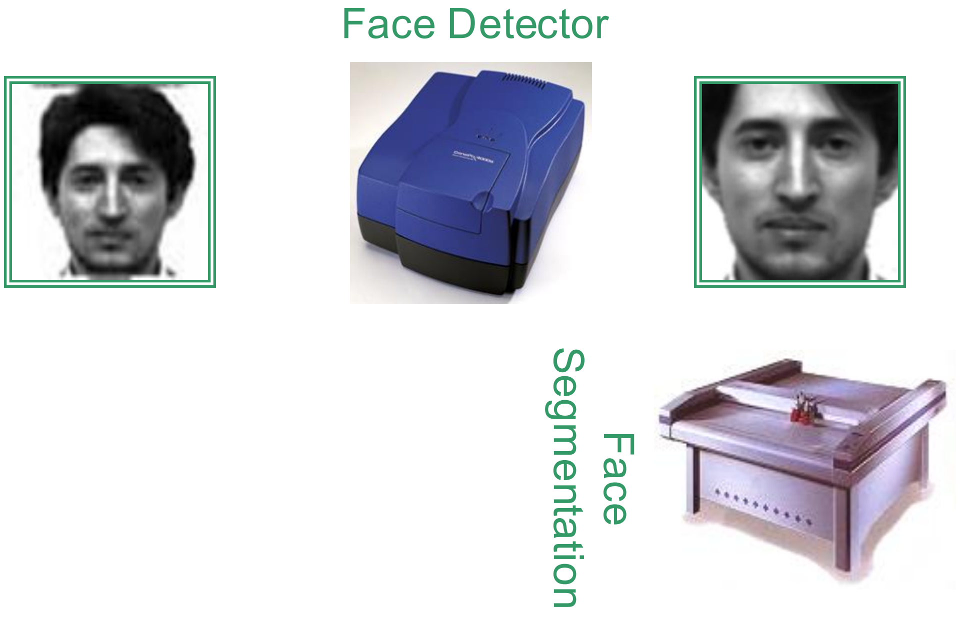
- Verifica: Confronto Uno a Uno. Conferma l'identità dichiarata da un individuo
  - Dichiarata mediante: carta di identità, codice utente, ...
- Riconoscimento: Confronto Uno a Molti. Stabilisce l'identità di un soggetto a partire da un insieme di persone registrate





Struttura di un Riconoscitore Facciale

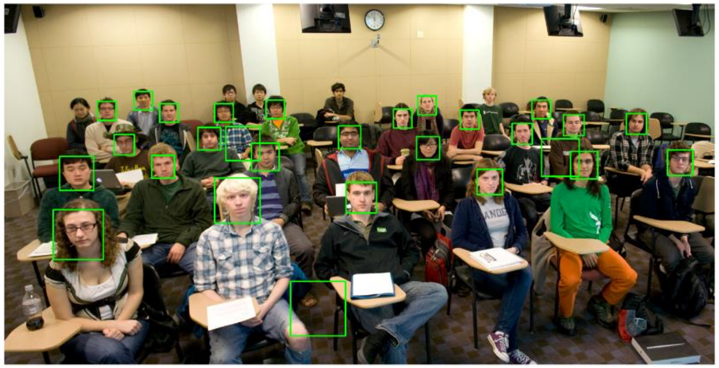
- Il primo step di un sistema automatico è un face detector.
- A partire dall'immagine, si estrae la regione contenente unicamente il volto.
- Infine la regione estratta viene segmentata, associando un significato semantico a ciascuna zona del volto.

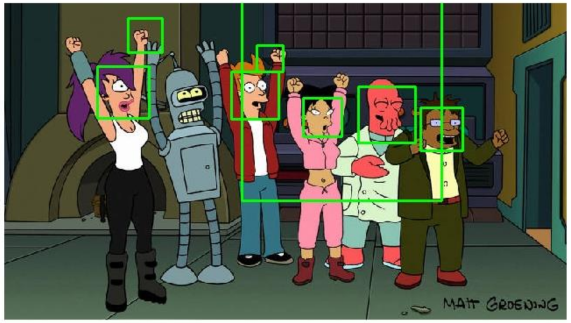


Face localization

• **Problem**: given a single image or a video sequence, detect the presence of one or more faces and locate their position within the single image.

• **Requirements**: it is necessary to be independent with respect to position, orientation (pose), scale, expression, possibly DIFFERENT for different subjects in the image, as well as to illumination or cluttered background.



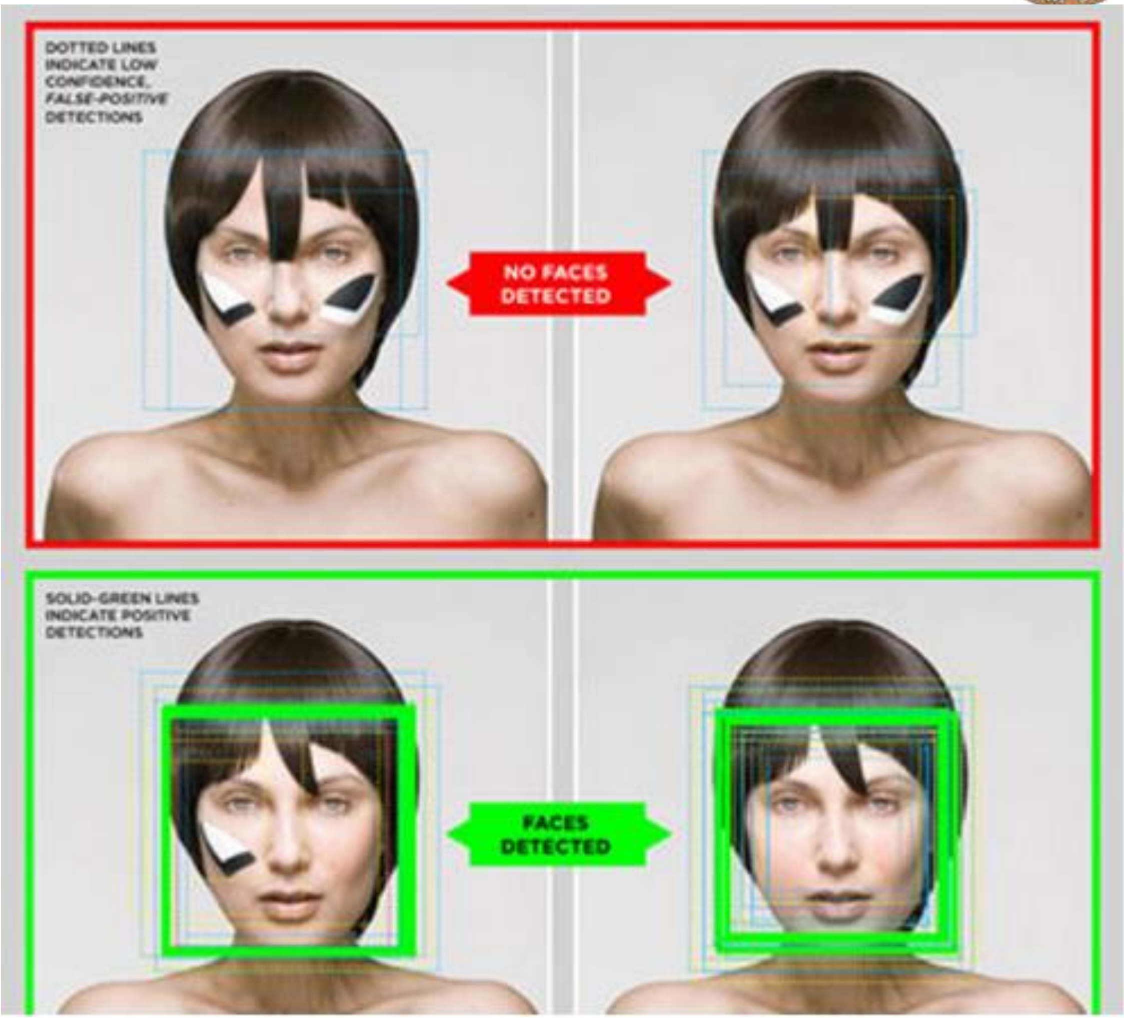


What to do here???

Face localization … possible to hide from it?

“According to Adam Harvey, the key part of the face that computers can read is the "nose bridge," or the area between the eyes. If you can obscure that, you have a good chance of tricking computers into thinking you don't have a face, he said. Another technique is to create an "anit-face," which is less terrifying than it sounds since it just means inverting your face's color scheme. So the black-and-white triangles on the cheeks aim to achieve this effect.” (from CNN - http://whatsnext.blogs.cnn.com/2012/04/29/how-to-hide-from-face-detection-technology/)

Adam Harvey proposes CV Dazzle (Computer vision Dazzle) to avoid one’s face detection (http://ahprojects.com/projects/)



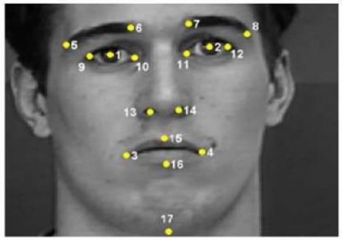
Face localization: Some Approaches

Feature-based techniques

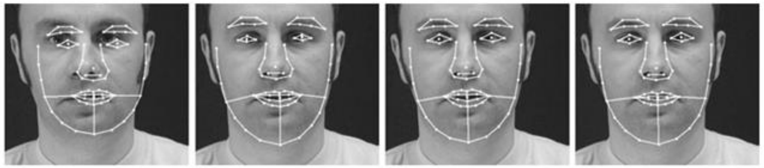
These techniques make explicit use of the knowledge about the expected face appearance, which is characterized by a set of features at different levels (below, they are from lower to higher)

- Pixel properties:
  - Edges
  - Skin color

- Face geometry properties (Informazioni sulla geometria del volto)
  - Constellation
  - Feature searching



- Template matching against a standard model (Modello standard di un volto definito manualmente o descritto da una funzione)
  - Correlation
  - Snakes
  - Active Shape Models



Initial After 5 it.s 10 it.s Converged (18 it.s)

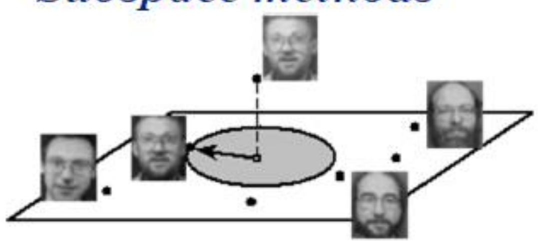
Face localization: some approaches

Image-based techniques

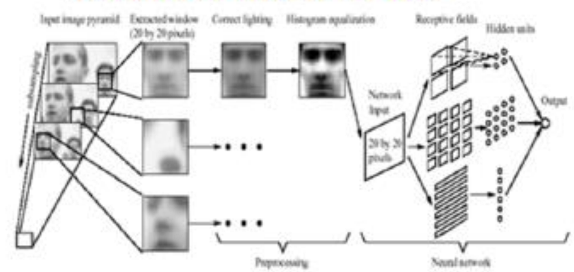
These techniques address the localization problem as a generic pattern recognition problem (class of faces).

The goal is to learn to recognize a face image according to a number of examples.

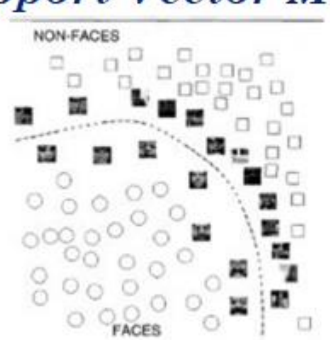
Subspace methods



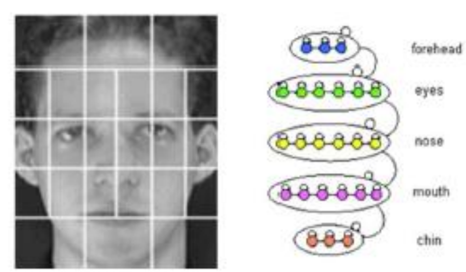
Neural networks



Support Vector Machine



Hidden Markov Model

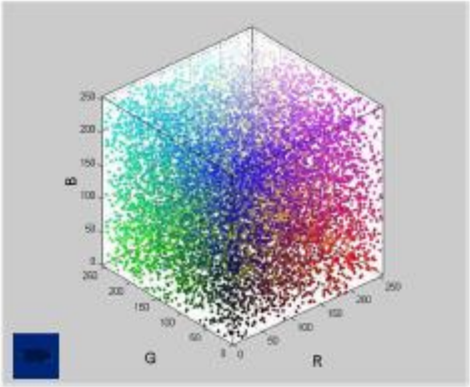


Face localization: recent advances

Localizing Parts of Faces Using a Consensus of Exemplars
from: http://neerajkumar.org/projects/face-parts/

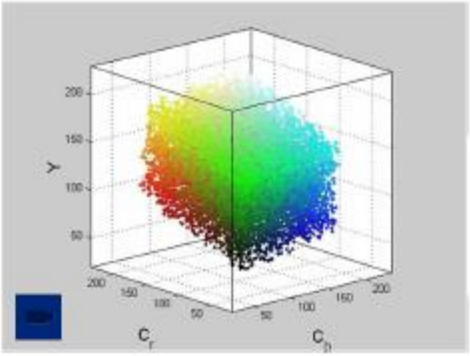
Algorithm A: color space transformation

- RGB is not a perceptually uniform space = colors which are close to each other in RGB space may not be perceived as similar
- Skin model: a set of close colors (cluster) within the color space
- It is advisable to perform the detection in a different space



RGB color space

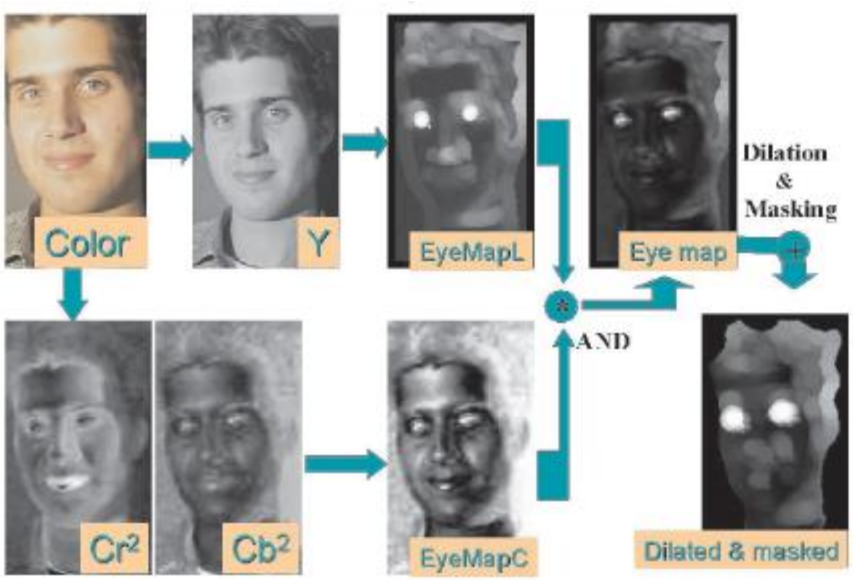
$$\begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256}\begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \bullet \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

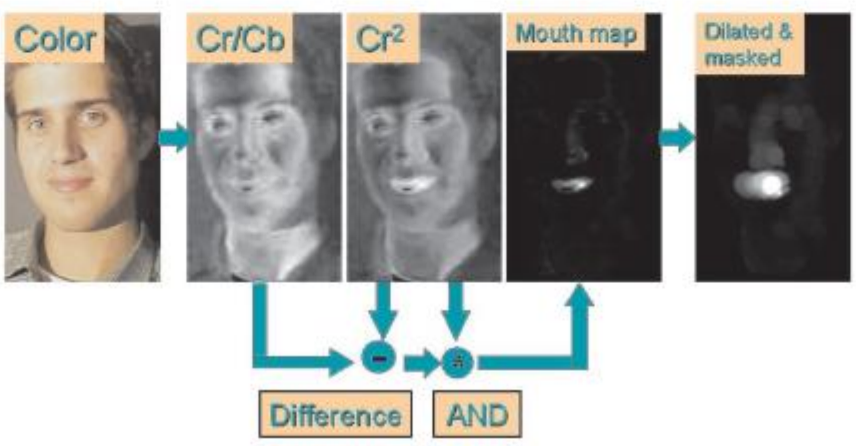


YC_BC_R color space

Eye/Mouth localization

- Chroma map is enhanced by histogram equalization
- The two maps are combined through AND operator
- The resulting map undergoes dilation, masking and normalization to discard the other face regions and brighten eyes.
- Further operations allow to refine this map.





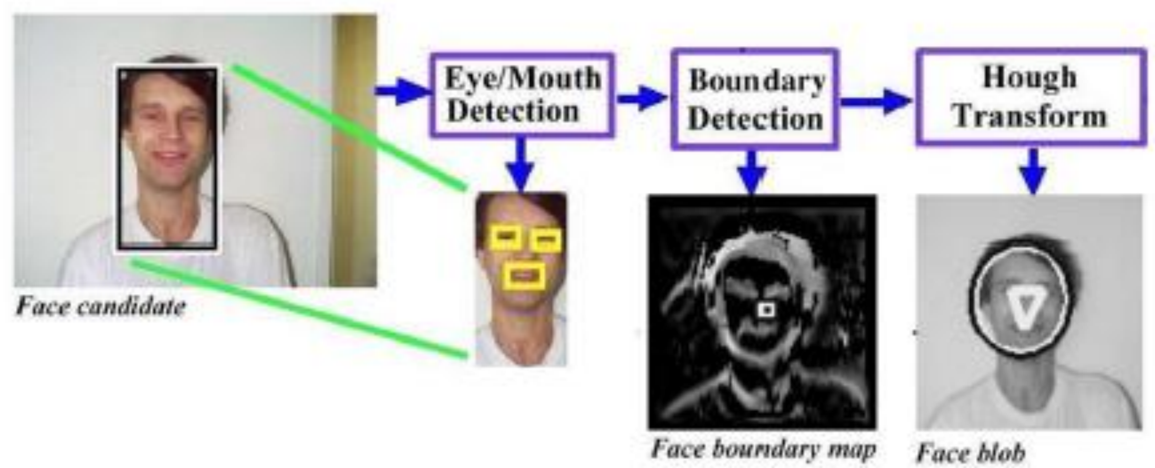
Face contour

The algorithm analyzes all the triangles composed by two candidate eyes and a candidate mouth.

Each triangle is verified by cheking:

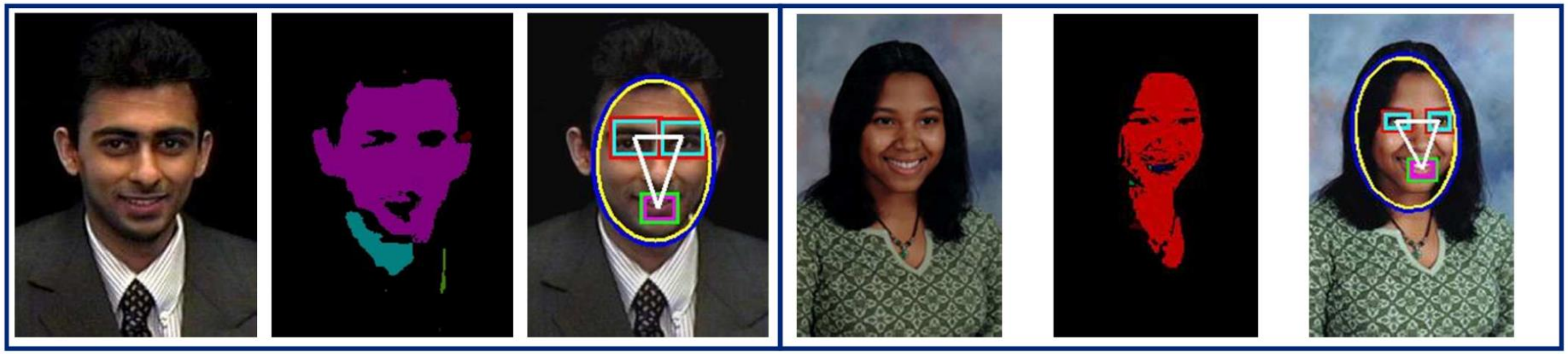
- Luma variations and average of the orientation gradient of the blobs containing eyes and mouth
- Geometry and orientation of the triangle
- Presence of a face contour around the triangle

A score is assigned to each triangle satisfying the conditions, which also considers preference for upright orientation and simmetry.



The triangle with highests score (above a threshold) selected.

Face contour



Some results

Face Detection: Viola-Jones

- Paul Viola and Michael Jones proposed one of the most successful (so far) approaches to object localization (included in OpenCV).
- The algorithm is image based and can be applied to face detection (but also to eye and mouth detection in a hierarchical strategy).
- The algorithm requires to create a classifier that is initially trained using multiple instances of the class to identify (positive examples), and several instances of images that do not contain any object of the class but may cause an error (negative examples).
- Training is designed to extract several features from the examples and to select the most discriminating ones. The statistical model which is built incrementally contains such information.
- Misses (a present object is not detected) or false alarms (an object is detected but it is not present) can be decreased by retraining adding new suited examples (positive or negative).

Face Detection: Viola-Jones (II)

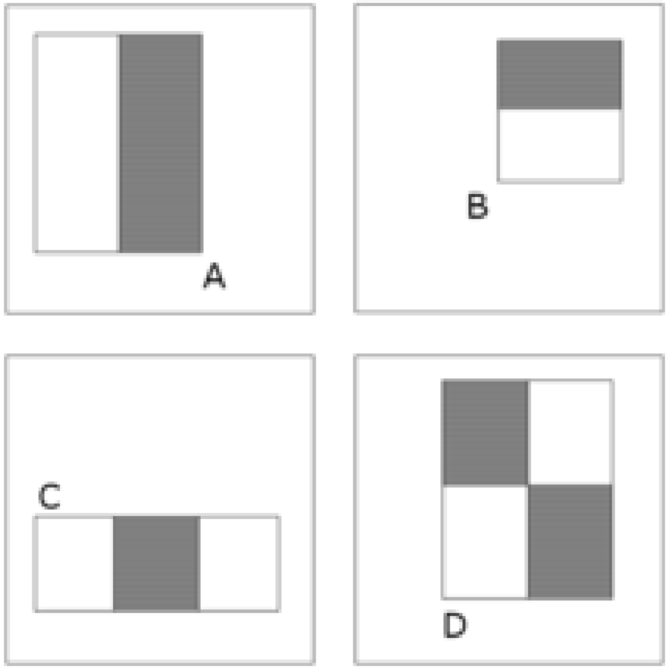
- Haar Features – All human faces share some similar properties. This knowledge is used to construct certain features known as **Haar Features**.
- The properties that are similar for a human face are:
  - The eyes region is darker than the upper-cheeks.
  - The nose bridge region is brighter than the eyes.
- That is useful domain knowledge:
  - Location - Size: eyes & nose bridge region
  - Value: darker / brighter

Haar Feature that looks similar to the bridge of the nose is applied onto the face

Haar Feature that looks similar to the eye region which is darker than the upper cheeks is applied onto a face

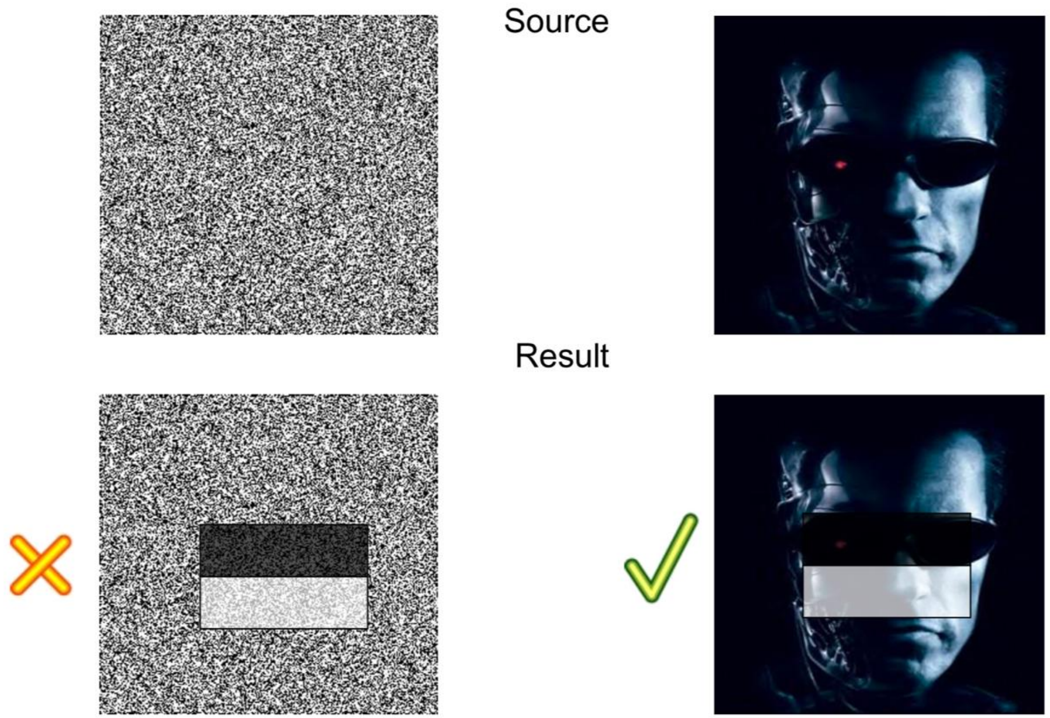
Face Detection: Viola-Jones (III)

- The value of any given feature is always simply the sum of the pixels within clear rectangles subtracted from the sum of the pixels within shaded rectangles.
- Rectangle features:
  - Value = Σ (pixels in black area) - Σ (pixels in white area)
- Three types: two-, three-, four-rectangles, Viola & Jones used two-rectangle features
- For example: the difference in brightness between the white &black rectangles over a specific area
- Each feature is related to a special location in the sub-window

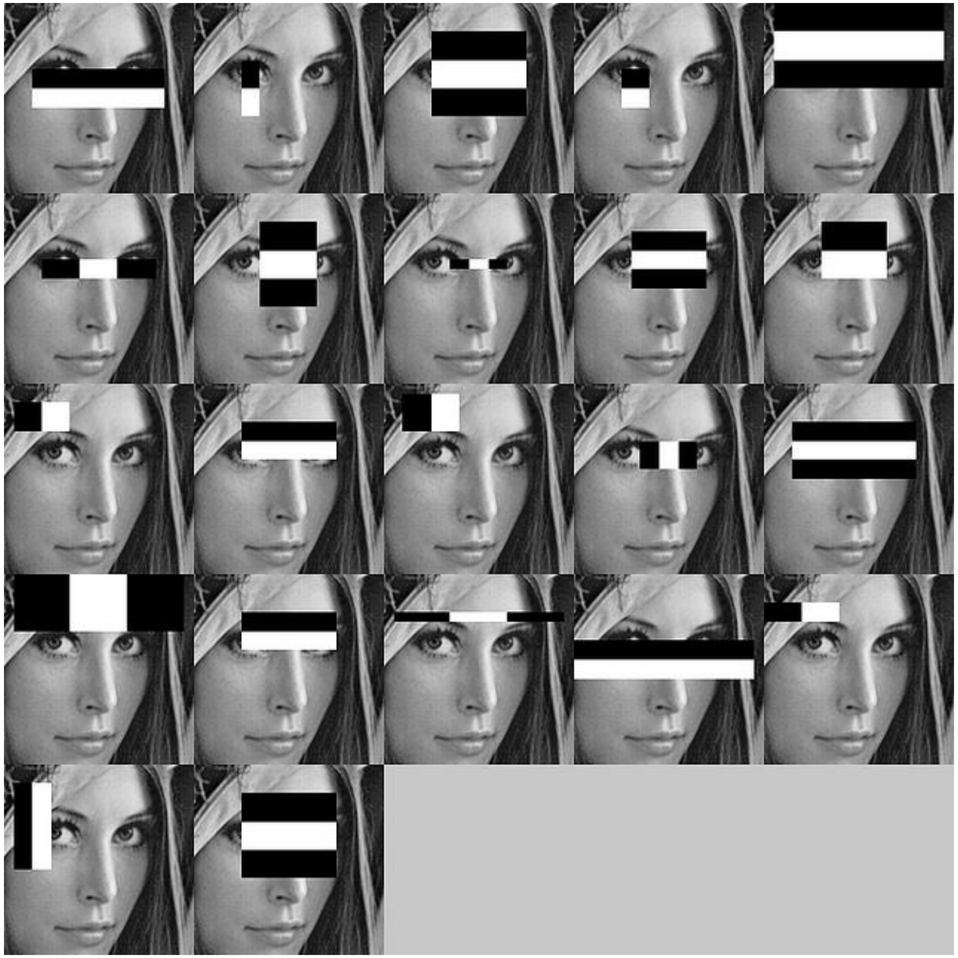


Example use of rectangular features

Example



Example use of rectangular features



An example of an early stage in the Haar cascade. Each black and white patch represents a feature that the algorithm hunts for in the image.

https://vimeo.com/12774628

Viola-Jones: L’algoritmo

- Utilizzo delle Haar features in combinazione con una nuova rappresentazione dell'immagine detta *Integral Image*. Le features hanno basso costo computazionale e la nuova struttura dati permette di effettuare l'analisi in tempo costante indipendentemente dalla dimensione delle regioni analizzate.

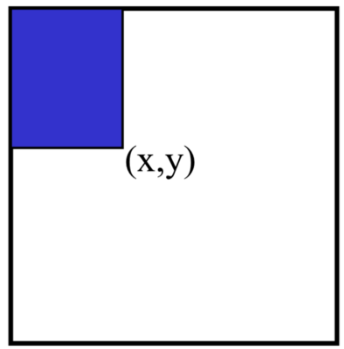
- Viene introdotto un metodo di selezione di feature di Haar attraverso l'algoritmo AdaBoost di Freud Shapire (1995). Questa strategia permette di eliminare in addestramento la maggior parte delle feature di scarsa capacità discriminante e selezionare solo quelle più efficaci per il problema.

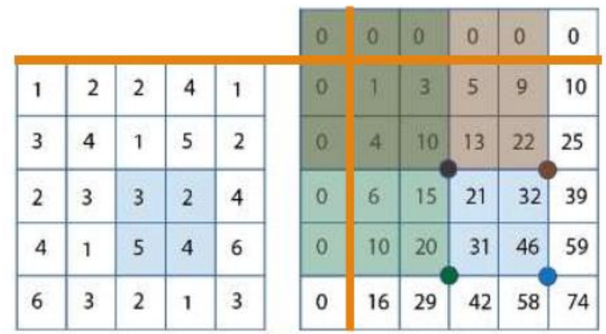
Integral Images

Fast computation with integral images

$$II(x_0, y_0) = \sum_{x=0}^{x_0} \sum_{y=0}^{y_0} I(x, y);$$

- The *integral image* computes a value at each pixel (*x*,*y*) that is the sum of the pixel values above and to the left of (*x*,*y*), inclusive
- This can quickly be computed in one pass through the image





1	2	2	4	1
3	4	1	5	2
2	3	3	2	4
4	1	5	4	6
6	3	2	1	3

input image

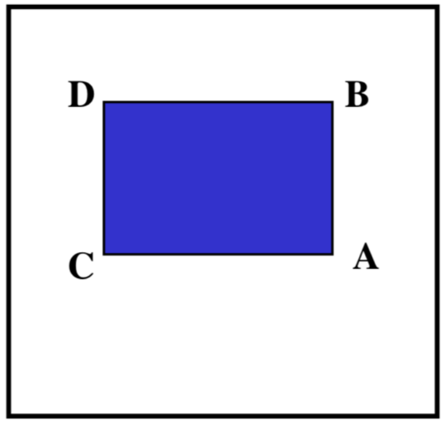
0	0	0	0	0	0
0	1	3	5	9	10
0	4	10	13	22	25
0	6	15	21	32	39
0	10	20	31	46	59
0	16	29	42	58	74

integral image

Computing sum within a rectangle

- Let A,B,C,D be the values of the integral image at the corners of a rectangle
- Then the sum of original image values within the rectangle can be computed as:

  sum = A – B – C + D
- Only 3 additions are required for any size of rectangle!
  - This is now used in many areas of computer vision



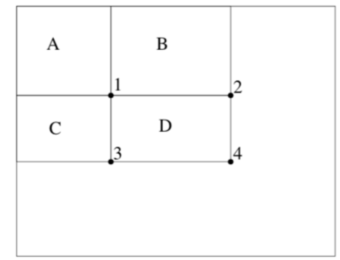


Figure 2: The sum of the pixels within rectangle D can be computed with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is $A+B$, at location 3 is $A+C$, and at location 4 is $A+B+C+D$. The sum within D can be computed as $4+1-(2+3)$.

$$II(x_0, y_0) = \sum_{x=0}^{x_0} \sum_{y=0}^{y_0} I(x, y);$$

1	2	2	4	1
3	4	1	5	2
2	3	3	2	4
4	1	5	4	6
6	3	2	1	3

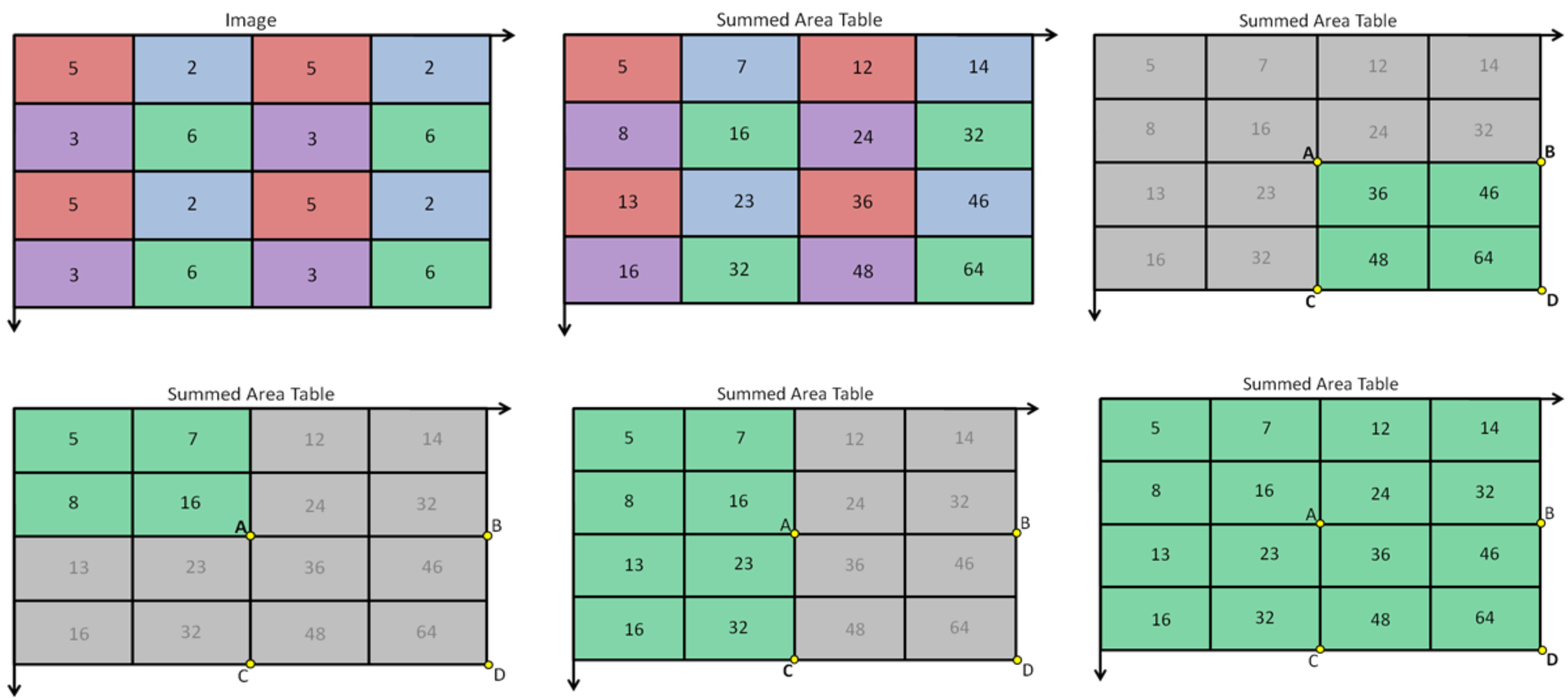
input image

0	0	0	0	0	0
0	1	3	5	9	10
0	4	10	13	22	25
0	6	15	21	32	39
0	10	20	31	46	59
0	16	29	42	58	74

integral image

Integral Images

Once the integral image is obtained, the sum of pixels in any rectangular region can be obtained in constant time (O(1) time complexity) by the following expression:
Sum = Bottom right + top left – top right – bottom left



A B

```
[[1, 0, 0, 1, 0],        [[ 1,  1,  1,  2,  2],
 [0, 0, 0, 1, 0],         [ 1,  1,  1,  3,  3],
 [0, 1, 0, 1, 0],   ==>   [ 1,  2,  2,  5,  5],
 [1, 1, 0, 1, 0],         [ 2,  4,  4,  8,  8],
 [1, 0, 0, 0, 1]]         [ 3,  5,  5,  9, 10]]
```

Original image   Integral image

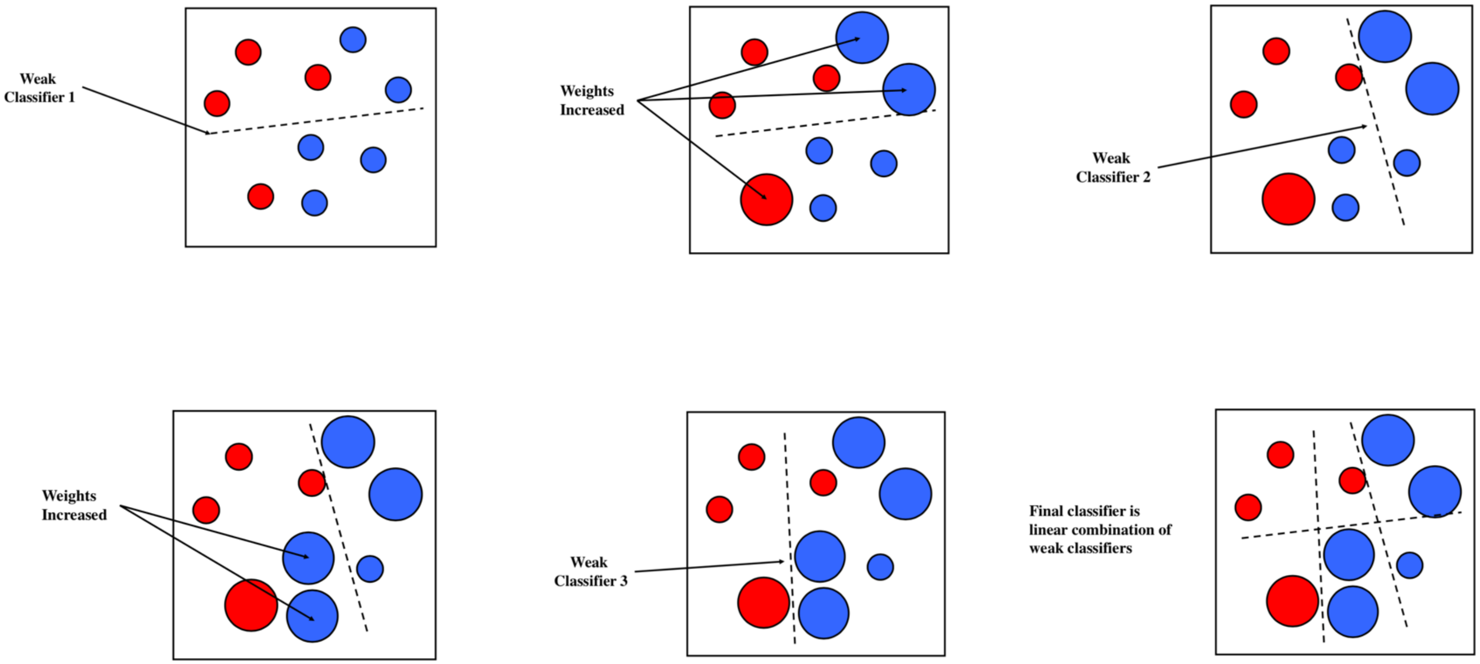
Sum = 8 + 1 - 2 - 2 = 5

```
[[1, 0, 0, 1, 0],        [[ 1,  1,  1,  2,  2],
 [0, 0, 0, 1, 0],         [ 1,  1,  1,  3,  3],
 [0, 1, 0, 1, 0],   ==>   [ 1,  2,  2,  5,  5],
 [1, 1, 0, 1, 0],         [ 2,  4,  4,  8,  8],
 [1, 0, 0, 0, 1]]         [ 3,  5,  5,  9, 10]]
```

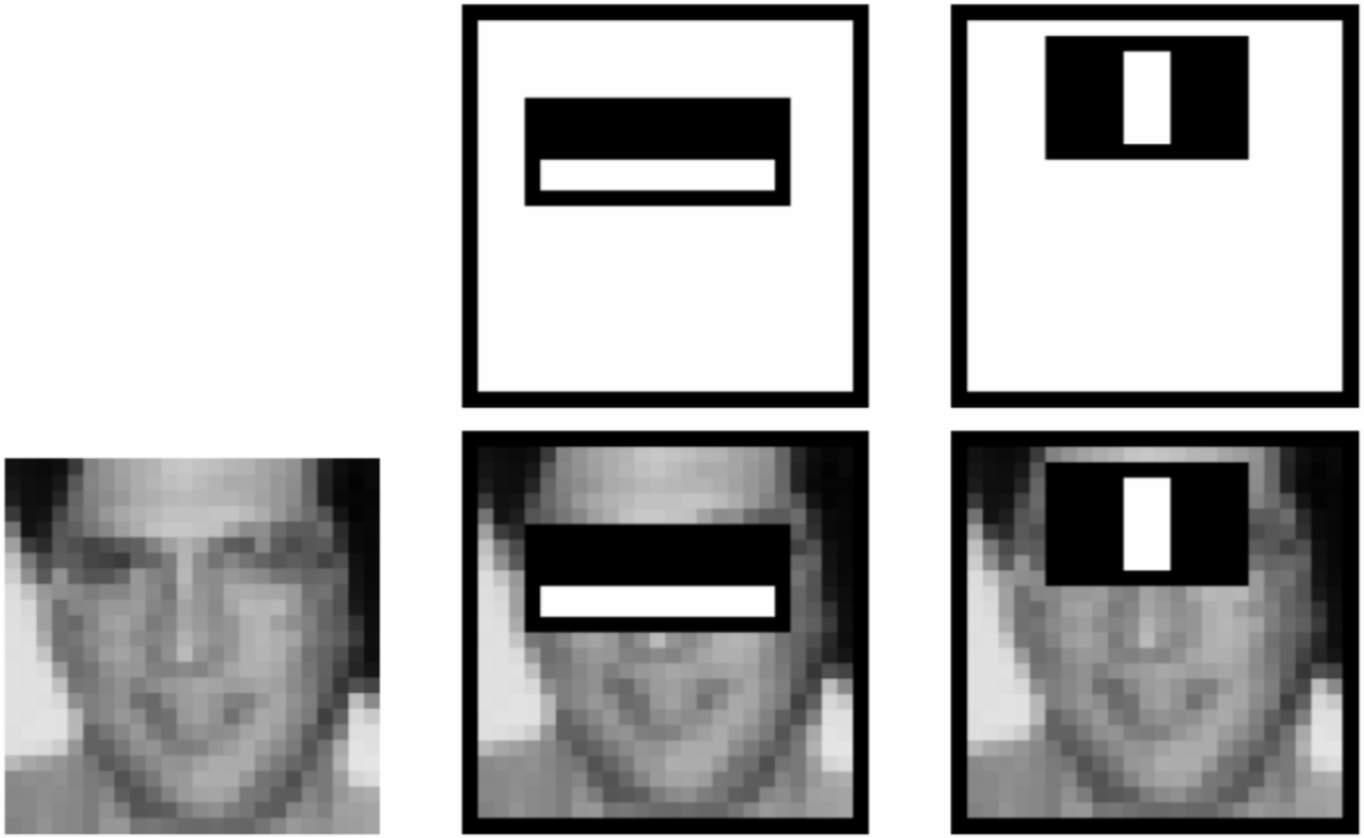
Original image   Integral image

Sum = 8 + 1 - 2 - 2 = 5

Boosting


Weak
Classifier 1
Weights
Increased
Weak
Classifier 2
Weights
Increased
Weak
Classifier 3
Final classifier is
linear combination of
weak classifiers

First two features selected by boosting



Boosting for face detection

For each round of boosting:

- Evaluate each rectangle filter on each example
- Select best threshold for each filter
- Select best filter/threshold combination
- Reweight examples

- Computational complexity of learning: O(MNT)
  - M filters, N examples, T thresholds

Popular databases

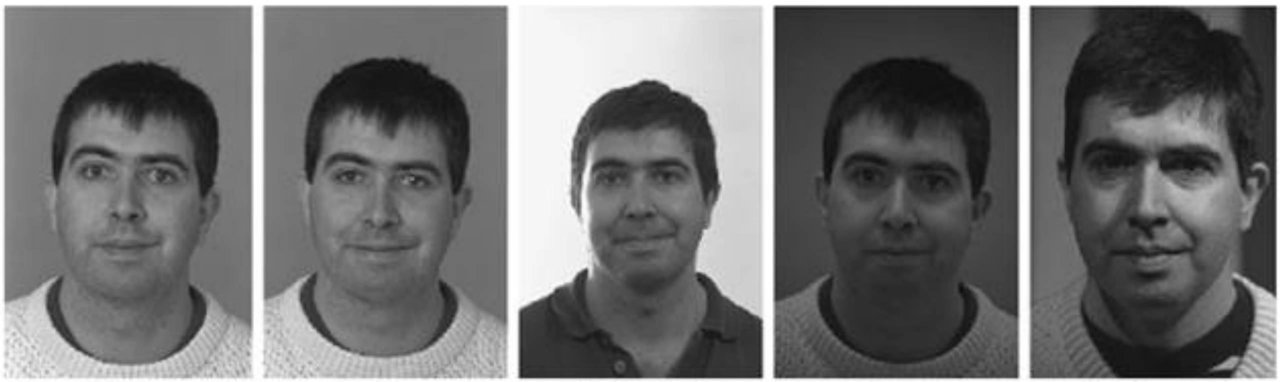
Very popular public face databases:

- AR-Faces
- FERET
- MIT
- ORL
- Harward
- MIT/CMU
- CMU test set II
- …

Popular databases

Name	RGB/Gray	Image Size	Number of people	Pictures / person	Number of conditions	Available	Web Address
AR Face Database*	RGB	576 × 768	126 70 Male 56 Female	26	i, e, o, t	yes	http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html
Richard's MIT database	RGB	480 × 640	154 82 Male 74 Female	6	p, o	yes	
CVL Database	RGB	640 × 480	114 108 Male 6 Female	7	p, e	yes	http://www.lrv.fri.uni-lj.si/facedb.html
The Yale Face Database B*	Gray Scale	640 × 480	10	576	p, i	yes	http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html
The Yale Face Database*	Gray Scale	320 × 243	15 14 Male 1 Female	11	i, e	yes	http://cvc.yale.edu/projects/yalefaces/yalefaces.html
PIE Database*	RGB	640 × 486	68	~ 608	p, i, e	yes	http://www.ri.cmu.edu/projects/project_418.html
The UMIST Face Database	Gray	220 × 220	20	19 to 36	p	yes	http://images.ee.umist.ac.uk/danny/database.html
Olivetti Att - ORL*	Gray	92 × 112	40	10		yes	http://www.uk.research.att.com/facedatabase.html
(JAFFE) Database	Gray	256 × 256	10	7	e	yes	http://www.mis.atr.co.jp/~mlyons/jaffe.html
The Human Scan Database	Gray	384 × 286	23	~66		yes	http://www.humanscan.de/support/downloads/facedb.php
The University of Oulu Physics-Based Face Database	Gray	428 × 569	125	16	i	Cost $50	http://www.ee.oulu.fi/research/imag/color/pbfd.html
XM2VTSDB	RGB	576 × 720	295		p	Frontal $153 Side $229.5	http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb/
FERET*	Gray RGB	256 × 384	30.000		p, i, e, i/o, t	yes	http://www.itl.nist.gov/iad/humanid/feret/

Popular databases: FERET

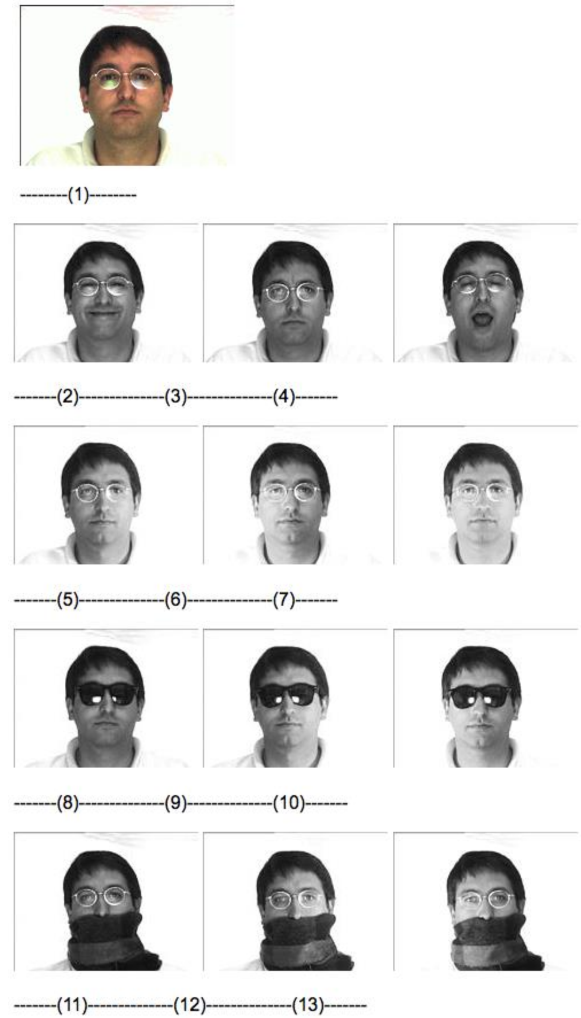


fa fb duplicate I fc duplicate II

Images from the FERET dataset. The fa and fb were taken with the same lighting condition with different expressions. The fc image has a different lighting condition than the fa and fb images. The duplicate I image was taken within one year of the fa image and the duplicate II and fa image were taken at least one year apart

Popular databases: AR-Faces (1998)

1 : Neutral expression
2 : Smile
3 : Anger
4 : Scream
5 : left light on
6 : right light on
7 : all side lights on
8 : wearing sun glasses
9 : wearing sun glasses and left light on
10 : wearing sun glasses and right light on
11 : wearing scarf
12 : wearing scarf and left light on
13 : wearing scarf and right light on
14 to 26 : second session (same as 1 to 13)

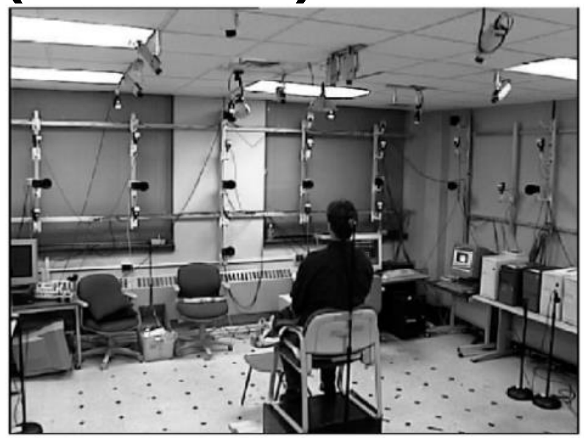


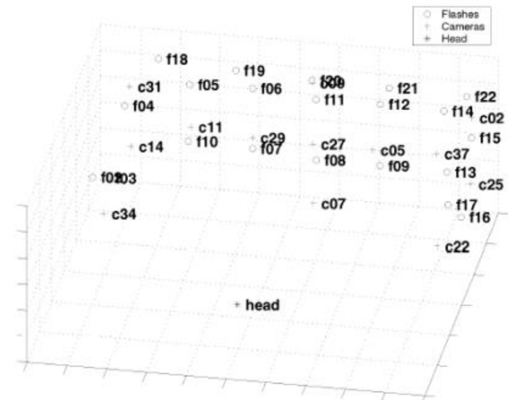
126 people (over 4,000 color images).
Different facial expressions, illumination conditions and occlusions.
Two sessions per person (2 different days).

http://wwwsipl.technion.ac.il/new/DataBases/Aleix%20Face%20Database.htm

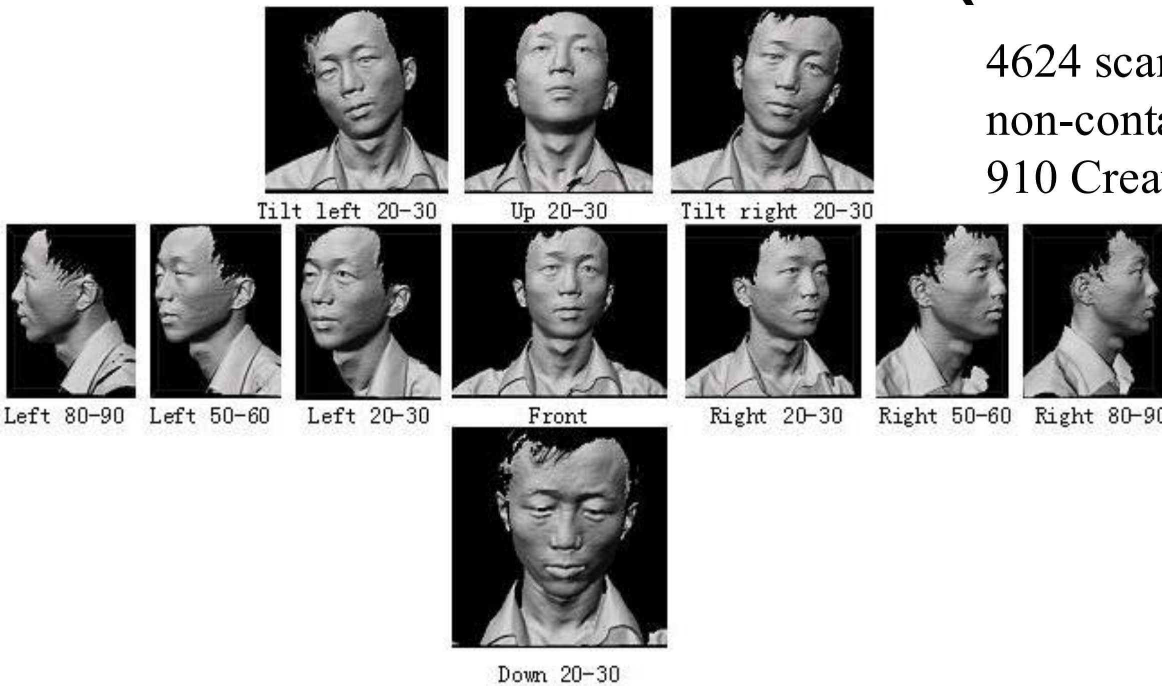
Popular databases : CMU-PIE (2000)

- CMU Pose, Illumination, and Expression (PIE) database.
- 68 subjects
- About 608 color pictures per subject (resolution 640×486 pixels) =41,368 images
- Each person is imaged under 13 different poses, 43 different illumination conditions, and with 4 different expressions.

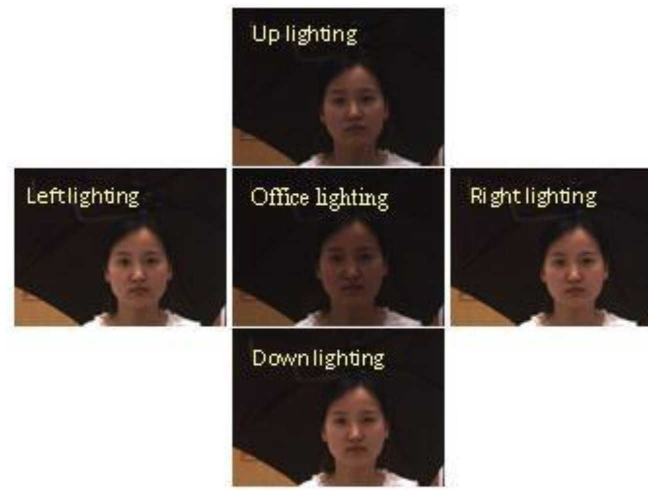




Popular databases: CASIA 3D Face V1 (2004)

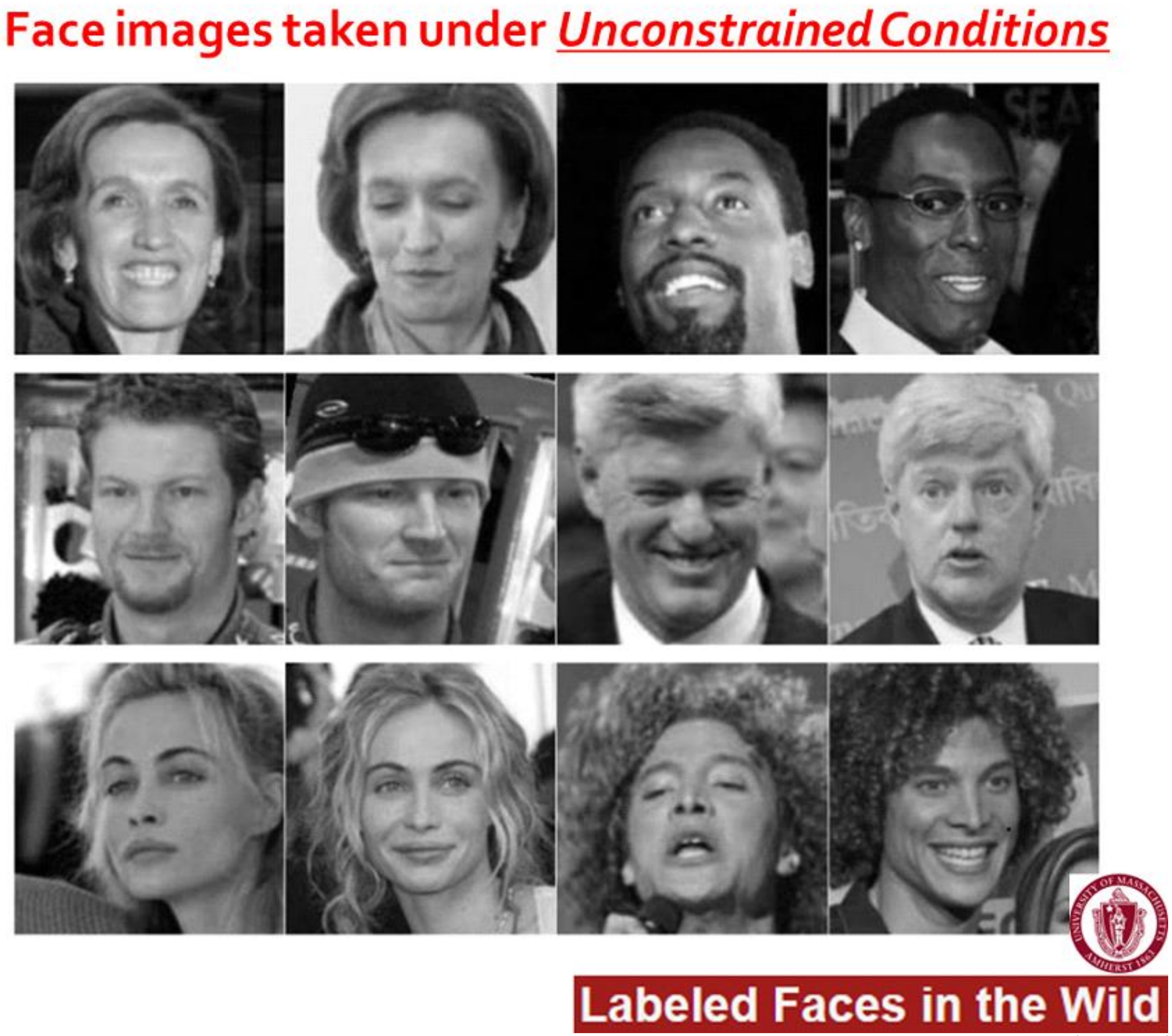


4624 scans of 123 persons using the non-contact 3D digitizer, Minolta Vivid 910 Created in 2004



http://biometrics.idealtest.org/dbDetailForUser.do?id=8

Popular databases: LFW (Labeled Faces in the Wild)(2007)



The data set contains more than 13,000 images of faces collected from the web. Each face has been labeled with the name of the person pictured. 1680 of the people pictured have two or more distinct photos in the data set. The only constraint on these faces is that they were detected by the Viola-Jones face detector.

http://vis-www.cs.umass.edu/lfw/

Popular databases: YouTube faces (2011)

The data set contains **3,425** videos of **1,595** different people. All the videos were downloaded from YouTube. An average of 2.15 videos are available for each subject. The shortest clip duration is 48 frames, the longest clip is 6,070 frames, and the average length of a video clip is 181.3 frames. The creation started by using the 5,749 names of subjects included in the LFW data set to search YouTube for videos of these same individuals. The top six results for each query were downloaded.

Number of videos per person:

#videos	1	2	3	4	5	6
#people	591	471	307	167	51	8

http://www.cs.tau.ac.il/~wolf/ytfaces/