Biometric Systems
Lesson 9: Ear recognition

Maria De Marsico
demarsico@di.uniroma1.it

SAPIENZA
UNIVERSITÀ DI ROMA
Dipartimento di Informatica

Credits

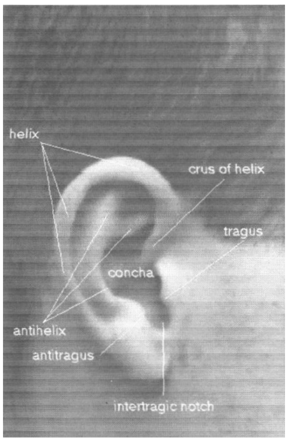
- This lesson is mostly the translation of a lesson by Prof. Daniel Riccio of University of Naples – BIPLAB at University of Salerno … with some additions

Ear as a biometrics

- There are several adjectives to describe a face and its components, but very few to describe an ear:
  - Shape: oval, round, ...
  - Appearance: delicate, plump, ...
  - Facial features: eye color, shape of the cheekbones, the shape of the nose and chin, lips thickness , ..
  - .
- People are used to recognize a person by the face, but not by the ear.
- These factors have fed skepticism towards this biometrics.

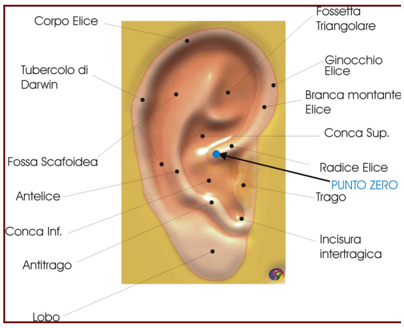
Ear anatomy

- **The ear has a structure that is not random, but well-defined just like the face** :
  - **the external border (helix);**
  - **the protrusion (anti-helix) running inside and parallel to the external border;**
  - **the lobe;**
  - **the u-shaped socket known as intertragic notch (tacca intratragica) between the entry of the inner ear (meatus) and the lobe**



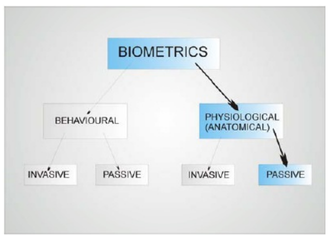
Ear anatomy

- **The ear has a structure that is not random, but well-defined just like the face :**
  - **the external border (helix);**
  - **the protrusion (anti-helix) running inside and parallel to the external border;**
  - **the lobe;**
  - **the u-shaped socket known as intertragic notch between the entry of the inner ear (meatus) and the lobe**



Positive aspects

- Ear is a passive and static biometrics
- Since it is associated with one of human senses, it is usually uncovered for better hearing
- If poorly visible it is possible to interact with the user



Positive aspects … But …

- The ear has been compared with other biometrics, especially the face. Some advantages have been underlined:
  - less details, therefore requiring lower resolution;
  - more uniform color distribution;
  - lower if no sensitiveness to expression variations

- … but the inherent 3D structure makes it sensible to illumination and pose variations and …

- … the small size is a double-sided medal …

- The use of such biometrics is still relatively limited to forensics (crime scene, e.g., latent ear print analysis).

The ear as a new biometrics

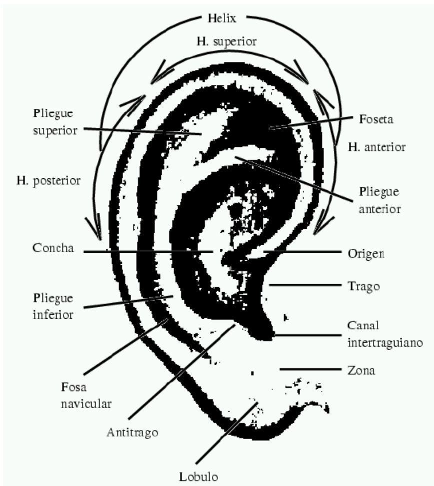
- In 1989, Alfred Iannarelli performs two investigations.
- The first one involves 10000 random samples selected in California, and its results demonstrate that:

  - Human ear has a sufficient variability to distinguish two different subjects

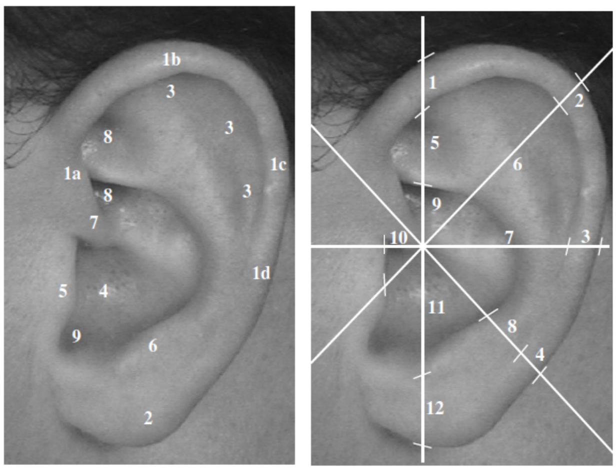
  - The human ear complies with the fundamental properties of a biometrics (universal, unique, permanent, collectable)

The ear as a new biometrics

- The second study involves brothers and identical twins, to verify if they have different features.
- A further aspect verified is the permanence over time:
  - The ear growth is proportional from birth to the first 4 months
  - Between 4 months and 8 years the lobe undergoes a greater vertical elongation due to gravity force
  - The size of the lobe is constant between 8 and 70 years; after this it further elongates due to tissue relaxation



Iannarelli's measurements

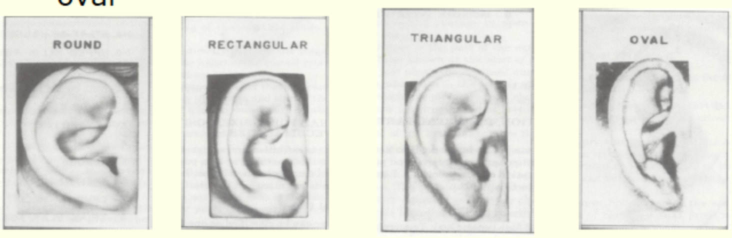


(a) Anatomy. (b) Measurements.

From Burge & Burger, 1996

A rough classification

- Iannarelli's classification:
  - round
  - rectangle
  - triangle
  - oval



Ear Identification (Forensic Identification Series) by Alfred V. Iannarelli

The ear as a new biometrics

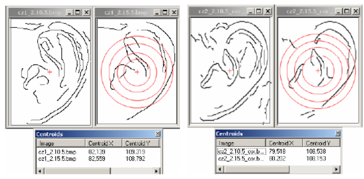
- Further minor works testify the suitability of the ear as a biometric trait :
  - Imhofer, in 1906, observes that given a set of 500 eras, 4 features are sufficient to distinguish them
  - La Bromba Gmbh, a company working with biometric systems, has compared a number of biometrics with respect to permanence over time

Biometric Trait	*Permanence over time*
Fingerprint (Minutia)	000000
Signature (dynamic)	0000
Facial structure	00000
Iris pattern	000000000
Retina	00000000
Hand geometry	0000000
Finger geometry	0000000
Vein structure of the back of the hand	000000
Ear form	000000
Voice (Tone)	000
DNA	000000000
Odor	000000?
Keyboard strokes	0000
Comparison: Password	00000

Ear localization and extraction

- The ear has a lower size as well as a lower complexity with respect to the face, but at the same time it has the same color of surrounding skin. This makes extraction more difficult.

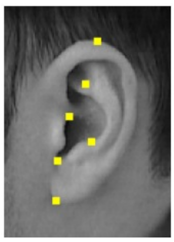
- There are different approaches in literature to ear extraction, among which:
  - Localization of points of interest
  - General object detection
  - Geometric 3D methods



Ear localization
(localization of points of interest)

- A possible approach uses Neural Networks, that must be trained
- A large amount of images is acquired
- On each image interest points (in yellow) are selected, which are a subset of those identified by Iannarelli and mentioned by Burge & Burger, 1996 (white numbers)

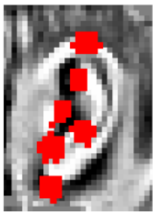
1. Helix
2. Lobule
3. Antihelix
4. Concha
5. Tragus
6. Antitragus
7. Crus of Helix
8. Triangular Fossa
9. Insicure Intertragica



Ear localization
(localization of points of interest)

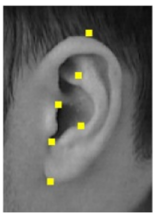
- Since the training for neutral networks takes a time which is proportional to the input image size, and since in this phase we are not much interested in fine details, the images are resized to a lower resolution

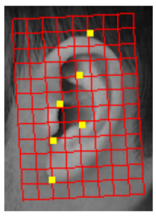
- Contrast in each image is normalized for better results
- Red points are an example of localization



Ear localization
(localization of points of interest)

- The selected points determine a reference system on the image
- All other points can be considered with respect to the reference ones as on a map



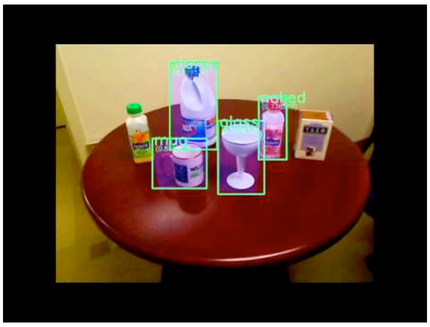


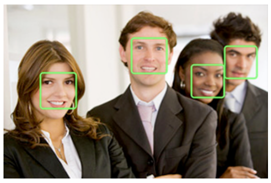
- In this way it is possible to extract the rectangle including the ear and possible normalize its size for matching purposes

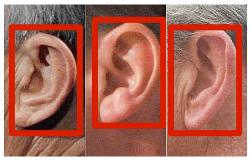
Ear localization
(general object detection)

- Object Detection refers to the localization of an object of a certain class within an image, in other words to the identification of its position

- A statistical approach performs a training step for complex objects





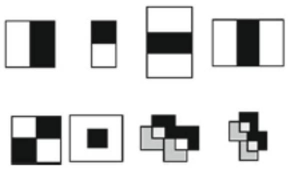


Ear localization
(general object detection)

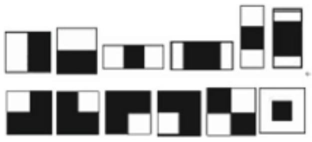
- An old friend: AdaBoost
- Training is based on:
  - multiple istances of the class of the relevant objects or positive samples
  - multiple negative samples, i.e. images that do not contain the object of interest

- The union of the two sets makes up the training set. During training the relevat features are extracted and selected according to their ability to classify the interesting objects

- Is it possible to add new positive or negative samples and repeat the training if respectively either the rate of missed objects or the rate of false alarms grow too much.



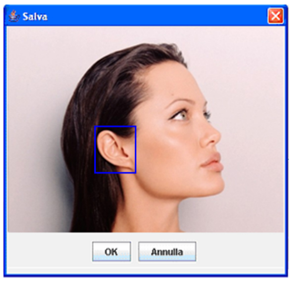
Haar features used in
Islam SMS, Bennamoun M, Davies R (2008)



Haar features used in
Yuan L, Zhang F (2009)

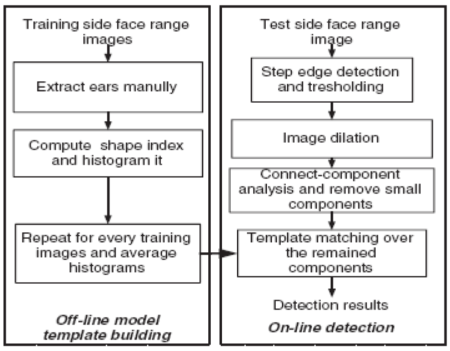
Ear localization
(general object detection)

- This method is implemented in OpenCV (C/C++) library
- OpenCV is an library for image analysis developed by Intel and continuously updated
- Many machine vision algorithms are already implemented there



Ear localization
(geometric 3D methods)

- 3D models allow to exploit the additional information provided by depth for localization
- Localization methods from 3D models use depth and surface
- Most attempts found in literature to capture a 3D model of the ear are related to its detection from a range image of the side face.



Ear localization
(geometric 3D methods)

- Training is performed offline:
  - Starting from a 3D model of the face profile, the points of maximum curvature are identified
  - A binary image is created (pixel 0-black, 1-white).
  - The region corresponding to the ear is manually extracted


(a) (b) (c) (d)

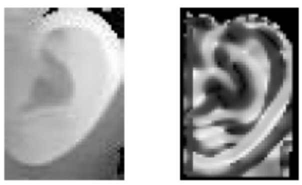
In Chen and Bhanu, 2004, ears are extracted exploiting a template matching-based detection method. The model template is built by manually extracting ear regions from training images and is represented by an average histogram (obtained from the histograms computed for the single training images) of the shape index of this ear region. A shape index is a quantitative measure of the shape of a surface at a point p, and is defined by

$$S(p) = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)}$$

with k_1 and k_2 maximum and minimum principal curvatures

Ear localization
(geometric 3D methods)

- Extracted regions are fused together to make up a reference model (template)


(a) (b)

Ear localization
(geometric 3D methods)

- Testing is performed on-line:
  - The binary image is computed for the new model
  - The points of maximum/minimum curvature are identified
  - Regions corresponding to the template are searched for (blu ones)

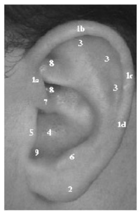
(a) (b)

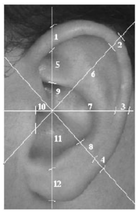
Ear recognition

- Recognition approaches can be classified in different ways.
- As an example, we can identify 3 classes of techniques:
  - 2D geometric (global) approaches based on curves/ landmarks
  - 3D Models
  - thermograms

Ear recognition
(global approaches: Iannarelli system)

- In this kind of approaches the whole ear is considered, starting from a 2D image and extracting relevant features from the Region of Interest (ROI) identified during localization
- Iannarelli's system is an example of this class





- The ear ROI is normalized with respect to its size
- The crus of helix is identified and set as the origin of the measurement system
- Starting from this point, 12 different measurements are performed (integer values)
- The feature vector includes gender, ethnicity and the 12 measures
- The main problem with this method is the accuracy required in identifying the central point
- If the central point identification is incorrect, all measures are wrong

Ear recognition
(global approaches: Voronoi diagrams)

- Burge e Burger solve the problem of exact point localization by using Voronoi diagrams and graph distance
- Within the ROI, edges are computed by Canny operator, and then edge relaxation is used to form larger curve segments
- Differences in illumination and positioning would undermine this method, therefore the authors describe the relations between the curves in a way invariant to affine transformations and to small shape changes caused by different illumination: a Voronoi neighborhood graph of the curves
- A Voronoi diagram is a partitioning of a plane into regions based on 'closeness' to 'seed' points in a specific subset of the plane. For each seed there is a corresponding region consisting of all points closer to that seed than to any other. These regions are called Voronoi cells
- The matching process searches for subgraph isomorphisms also considering possibly broken curves
- The method has not been experimented. However, ear segmentation is very sensitive to pose and illumination

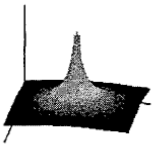
(a) Ear print.

(b) Voronoi diagram.

(c) Neighbor graph.

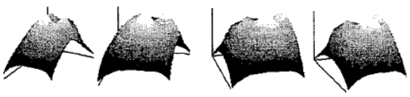
Ear recognition
(global approaches: Force Fields)

- Proposed by Hurley, Nixon, and Carter (2002)
- Each pixel is considered as a loaded particle (0 neutral, 255 maximum load)



- Each pixel in the image is treated as a Gaussian attractor, and is the source of a spherically symmetric force field, acting upon all the other pixels in a way which is directly proportional to pixel intensity and inversely proportional to the square of distance.

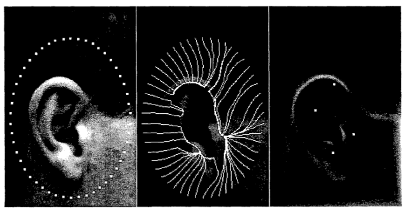
$$\mathbf{F}_i(\mathbf{r}_j) = P(\mathbf{r}_i)\frac{\mathbf{r}_i - \mathbf{r}_j}{\left|\mathbf{r}_i - \mathbf{r}_j\right|^3}$$

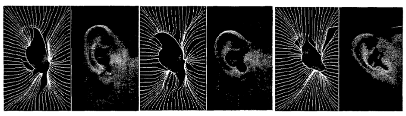


- For each pixel the method computes the force that all the others apply on it (module, direction and orientation) (modulo, direzione e verso)

Ear recognition
(global approaches: Force Fields)

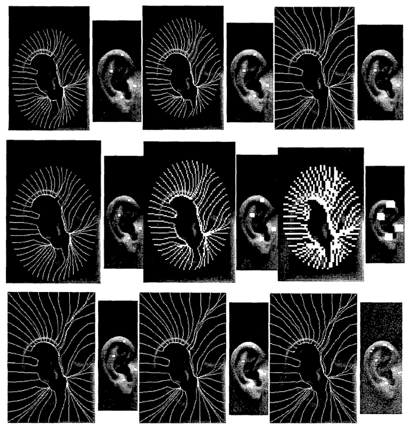
- A series of pints is fixed along an ellipse around the ear
- Starting from each point, the attraction of the force field is followed
- If two paths join each other, they cannot divide anymore
- Field lines converge in points defined as sinks





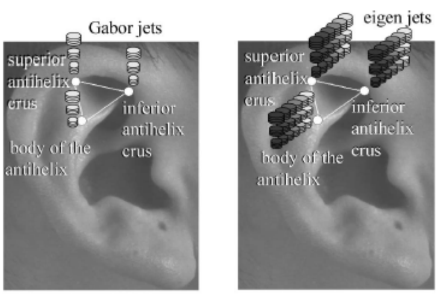
Ear recognition
(global approaches: Force Fields)

- This method is quite robust with respect to:
  - starting points
  - image resolution
  - noise



(a) No Noise (b) Medium Noise (c) Severe Noise

Ear recognition
(more with interest points)



an ear graph (a) the eigen ear graph (b)

- Jets (remember Gabor wavelets) (Watabe at al. 2008)
  - A number of kernel functions determining Gabor filters with orientation selectivity are convoluted with the image. This produces features vectors defined as Gabor Jets.
  - Gabor Jet function is used as a visual feature of the gray scale image I(v) at each image point v. The presented approach introduces an "ear graph" whose vertices are labeled by the Gabor Jets at the body of the antihelix, superior antihelix crus, and inferior antihelix crus. These Jets are stored as ear graphs in the gallery, and PCA is used to obtain the eigenear graph; an ear is detected using the similarity between sampled Jets and those reconstructed by the probe

Ear recognition
(more with interest points)

- Angle vectors
  - In Shailaja and Gupta P (2006), after edge detection, the extracted features are all angles. The angles are divided into two vectors. The first vector contains angles corresponding to edges in the outer shape of the ear, and the second vector is composed examining all other edges. The approach is based on the definition of max-line and normal lines.
  - Max-line is defined as the longest line with both endpoints on the edges of the ear. If edges are correctly identified and connected, the max-line has both its end points on the outer edge of the image. Features corresponding to each possibly detected max-line are to be extracted.
  - Normal lines are those which are perpendicular to the max-line and whose intersections divide the max-line into (n+1) equal parts, with n a positive integer.

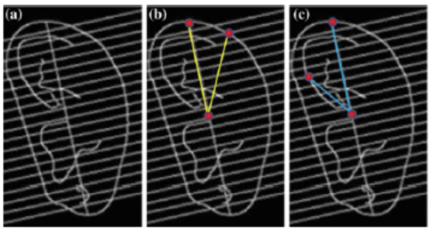


Fig. 6.9 **a** Example of max-line and normal lines; **b** example angle in the first vector; **c** example angle in the second vector

Ear recognition
(more with interest points)

- Angle vectors
  - Given the points where the outer edge intersects the normal lines, each angle stored in the first vector is formed by the segment connecting one such point with the center of the max-line, and the segment connecting the upper end point of the max-line and its center. The second feature vector is calculated similarly, and all the points where all the edges of the ear intersect the normal lines are considered, except for those already used for the first feature vector, i.e. those belonging to the outer curve.
  - The distance between two samples is computed in a hierarchicalway, in two steps: the first vector is used first, and the second one later. In the first step, two distance measures are computed. The first one is given by the sum of absolute difference between corresponding vector elements, while the second one counts the number of similar elements, i.e. those whose difference is under a threshold.
  - In order to pass the first check, both measures must comply with the respective thresholds. In the second step, two points are considered to match if their difference is under a threshold and they also belong to corresponding normal lines.

Ear recognition
(alignment)

- Active Shape Model (ASM)
  - Active shape models (ASMs) are statistical models of the shape of objects which iteratively deform to fit to an example of the object in a new image, developed by Tim Cootes and Chris Taylor in 1995.

- Scale Invariant Feature Transform (SIFT)
  - Lowe's method (1999) transforms an image into a large collection of feature vectors, each of which is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and robust to local geometric distortion. Key locations are defined as maxima and minima of the result of difference of Gaussians function applied in scale space to a series of smoothed and resampled images.

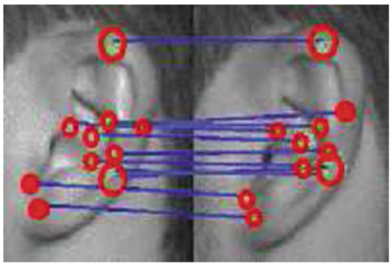
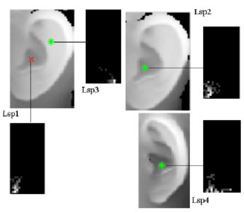


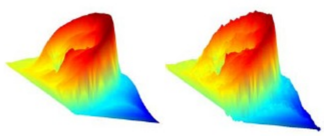
Fig. 6.8 Examples of point correspondence computed by SIFT

Ear recognition
(3D)

- These methods evaluate depth and curvature of relevant ear regions

- In some cases matching is performed between corresponding regions of two 3D models called patches

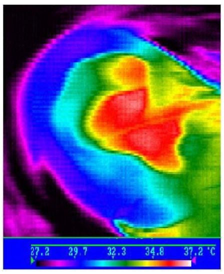




(a) No added noise (b) $\sigma = 0.6mm$

Ear recognition
(thermograms)

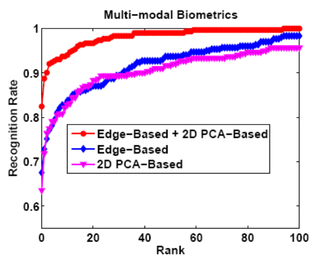
- Ear image is captured by a thermal camera
- Advantages:
  - the ear is easily locatable
  - robustness to hair occlusion;
  - different color facilitate segmentation
- Disadvantages:
  - sensitive to movement
  - low resolution
  - high costs



Ear recognition
(in multimodal systems)

- A number of systems combine 2D and 3D information to improve recognition performance
- The core issue is to select the best fusion strategy

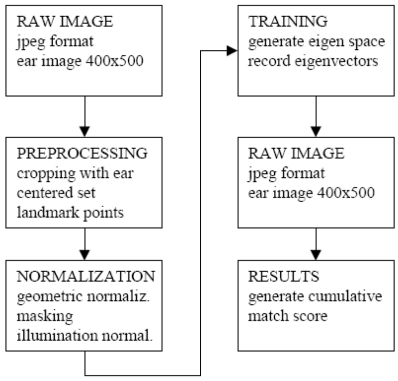
Multi-modals	MIN	Simple SUM	Advanced Sum
2D PCA + 3D ICP	76.4%	81.1%	82.5%
2D PCA + 3D PCA	72.2%	78.8%	79.1%
2D PCA + 3D Edge	73.5%	80.5%	82.5%



Face and ear

- Victor, Bowyer e Sarkar in 2002 compared face and ear recognition

- Their approach uses PCA and entails three steps:
  - Preprocessing
  - Normalization
  - Identification

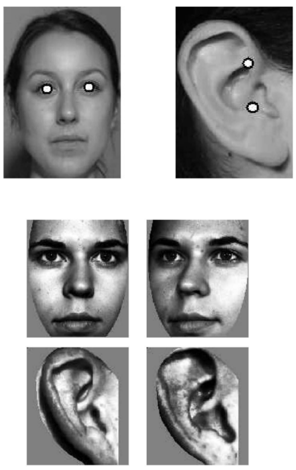


Face and ear

- Preprocessing entails in resizing each image at a resolution of 400×500 pixels

- As for normalization, two reference points are provided, both for face and for ear, and photometric normalization is performed too

- Regions not belonging to the ear are masked out



Face and ear

- Tests were performed on a set of 294 subjects, with a total of 808 images
- For each subject, one image for the face and one for the ear were provided at least
- Three types of experiments :
  - Gallery and probe same day
  - Gallery e probe in different days with same expression
  - Gallery e probe in different days with different expression

Face and ear

- According to this study, face always provides better performance than ear

Experiment #	Face/Ear compared		Expected Result	Result
1	Same day, different expression	Same day, opposite ear	Greater variation in expressions than ears; ears perform better	Face performs better
2	Different day, similar expresion	Different day, same ear	Greater variation in expression across days; ears perform better	Face performs better
3	Differet day, different expression	Different day, opposite ear	Greater variation in face expression than ear; ears perform better	Face performs better

Volto e Orecchio

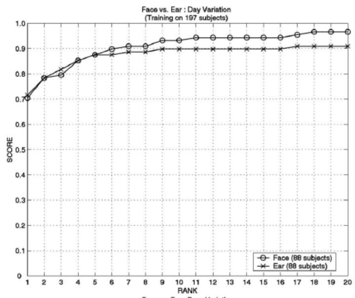
- Chang, Bowyer e Sarkar repeat a similar study in 2003

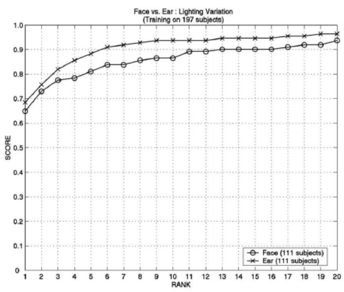
- Three variations are considered:
  - Images captured in different days
  - Illumination variations
  - Pose variations

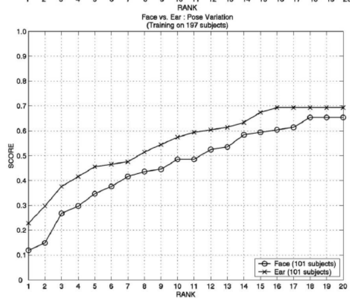
Gallery image — Day variation — Lighting variation — Pose variation

Face and ear

- Results are quite different:

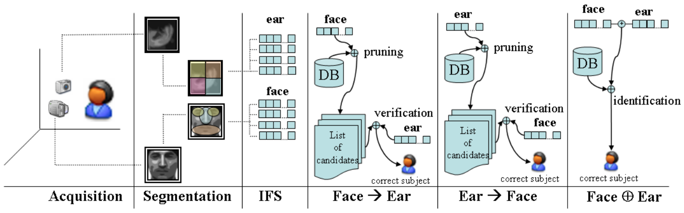






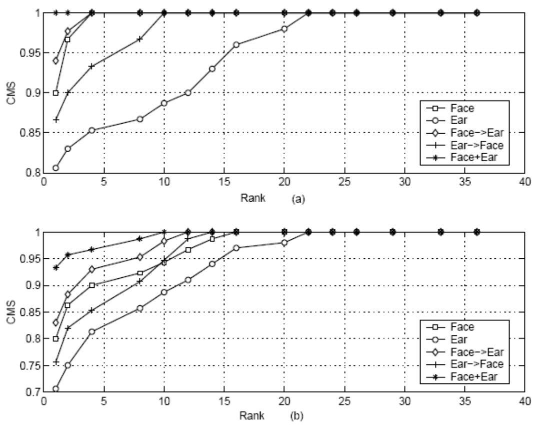
Face and ear

- It is possible to combine face and ear:
  - Face ⊕ Ear
  - Face → Ear
  - Ear → Face



Face and ear

- Testing on a significant number of subjects demonstrated that the best option is the parallel combination of face and ear



Some readings

- Iannarelli, A. V. (1964). *The Iannarelli system of ear identification*. Foundation Press.
- Iannarelli, A. (1989). Ear Identification, Forensic Identification series, Fremont. *Paramont Publishing, Calif, ISBN*, *10*, 0962317802.
- Burge, M., & Burger, W. (1996). Ear biometrics. In *Biometrics* (pp. 273-285). Springer US.
- Islam SMS, Bennamoun M, Davies R (2008) Fast and fully automatic ear detection using cascaded adaBoost. In: Proceedings of the IEEEworkshop on applications of computer vision– WACV 2008, pp 1–6
- Yuan L, Zhang F (2009) Ear detection based on improvedAdaBoost algorithm. In: Proceedings of the International Conference on Machine Learning and Cybernetics, vol 4, pp 2414–2417
- Chen H, Bhanu B (2004) Human ear detection from side face range images. In: Proceedings of the International Conference on Pattern Recognition (ICPR 2004), vol 3, pp 574–577
- Hurley DJ, Nixon MS, Carter JN (2002) Force field energy functionals for image feature extraction. Image Vision Comput 20(5–6):311–317

Some readings

- Watabe D, Sai H, Sakai K, Nakamura O (2008) Ear biometrics using jet space similarity. In: Proceedings of the Canadian conference on electrical and computer engineering–CCECE 2008, pp 1259–1264
- Shailaja D, Gupta P (2006) A simple geometric approach for ear recognition. In: Proceedings of the 9th international conference on information technology–ICIT 2006, pp 164–167
- Victor B, Bowyer K, Sarkar S (2002) An evaluation of face and ear biometrics. In: Proceedings of 16th international conference on pattern recognition, vol 1, pp 429–432
- ChangK,Victor B, BowyerKW, Sarkar S (2003) Comparison and combination of ear and face images in appearance-based biometrics. IEEE Trans Pattern Anal Mach Intell 25(8):1160–1165