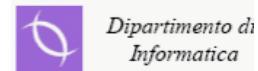

Biometric Systems

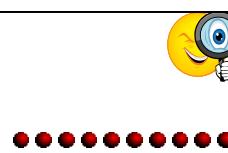
Lesson 13: Example solutions



Maria De Marsico
demarsico@di.uniroma1.it



Face Recognition



PIFS

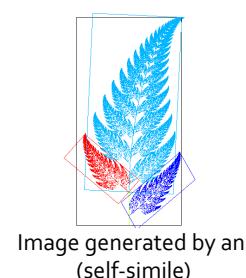
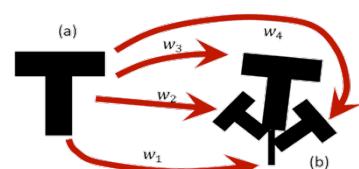
- PIFS = Partitioned Iterated Function System
- A powerful fractal-based approach to image compression and indexing
- Exploits and codes the image self-similarities
- The image is partitioned in square non-overlapping regions called ranges
- Further square overlapping regions, called domains, are also identified (side length = 2 side length of ranges)
- Each range is coded through the best approximating domain after a suitable affine transformation

Partitioned Iterated Function Systems

Self-similarities

FARO

- Evolution of **IFS** o **Iterated Function System**
- Arbitrary image-> affine transformations-> final image



- **It is possible to only store transformations**
- Real/natural images are not self-simile

Partitioned Iterated Function Systems



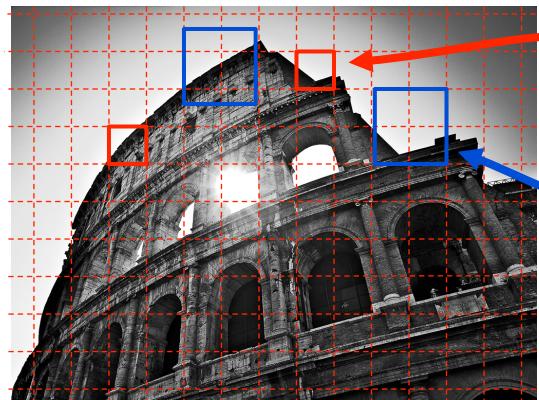
Image Partitioning



- The image is formed by copies of parts of itself

For each range we identify the best approximating domain ...

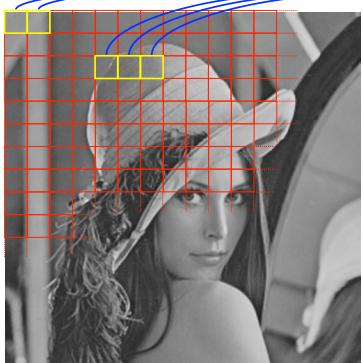
... after an affine transformation



PIFS: self-similarities coding



PIFS: self-similarities coding (range location)



- They represent a coverage of the image.

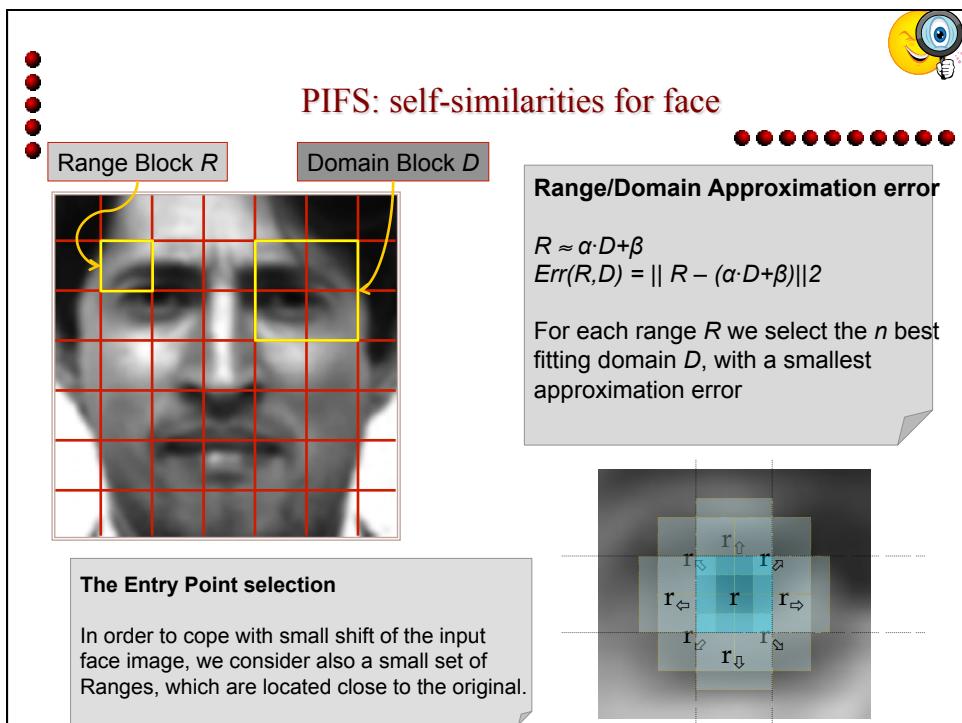
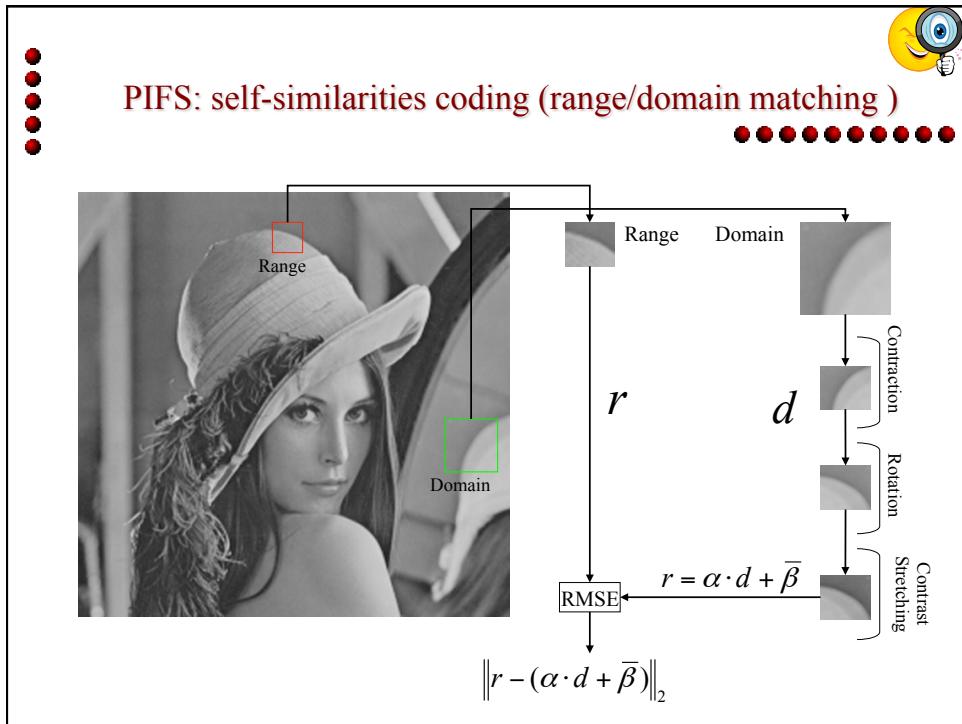
$$I = \bigcup_i r_i$$
$$r_i \cap r_j = \emptyset, \quad \forall i \neq j$$

This means 2^{12} 8×8 ranges,
on a 512×512 pixel image.

PIFS: self-similarities coding (domain location)



This means 2^{18} 16×16 domain,
for a 512×512 pixel image.



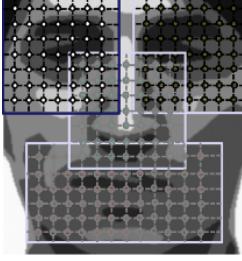


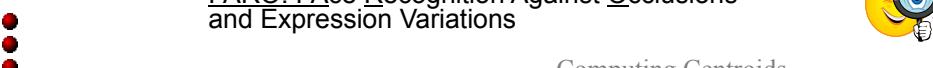
PIFS: self-similarities for segmented face

Face Segmentation

The face image is segmented in four different regions (eyes, nose, mouth) and each one is segmented independently.

In this way, the feature extraction process is made local and the effect of partial occlusions on the face image is mitigated.





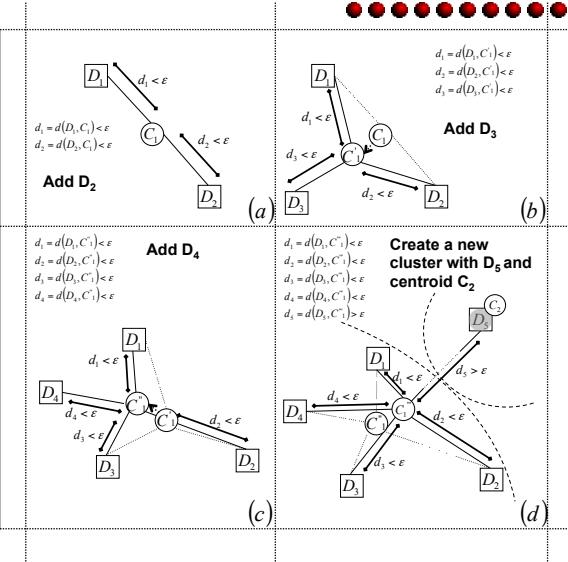
FARO: FAcce Recognition Against Occlusions and Expression Variations

The List of Centroids

The domains extracted during the range indexing process are organized in a set of clusters $C = \{C_1, C_2, \dots, C_m\}$, each of them represented by its centroid c_j , while centroids are stored in memory as a list.

Each centroid in the list stores its spatial coordinates computed as the mean of the coordinates of the domains belonging to the cluster and the average approximation error between domains and the prefixed block.

Computing Centroids



FARO: FAc Recognition Against Occlusions and Expression Variations

Extracted Features

Three real values are associated to each centroid in the list:

- X_i is the X coordinate of the centroid C_i ;
- Y_i is the Y coordinate of the centroid C_i ;
- Var_Err_i is the mean value of the range/domain approximation errors, computed on the domains belonging to the cluster of C_i .

Comparing Features

The centroid C_i in the Image 1 is compared only with the centroid C_j in the Image 2, whose position (X_j, Y_j) is nearest to (X_i, Y_i) , in an Euclidean sense.

$$d_j = \min_k \|C_{X_i Y_i} - C_{X_k Y_k}\|_2$$

FARO: FAc Recognition Against Occlusions and Expression Variations

Peano visit of the list of Centroids

The spatial coordinates of the centroids are visited according to a Peano scanning of the space, so that the centroids can be organized in a linear list.

The Peano scanning of a bi-dimensional space allows its linearization holding the spatial information.

The Peano scanning of two integers x and y consists in interleaving their bits, from the most significant bit to the less significant one.

FARO: FAce Recognition Against Occlusions and Expression Variations



Measuring distance between feature vectors

Given two feature vectors V_1 and V_2 obtained with the Peano linearization, the nearest element in V_2 is searched for the first item in V_1 , saving its position j . Then for the second item in V_1 the search start from the position j in V_2 and so on, summing the corresponding approximation error difference. The process is repeated swapping V_1 and V_2 .

$$\tau(i) = \arg \min_j |a_i - b_{\tau(i)}|$$

$$d = \sum_{i=1}^{|V_1|} \gamma_i \cdot |a_i - b_{\tau(i)}| + \sum_{k=1}^{|V_2|} \gamma_k \cdot |b_k - a_{\tau(k)}|$$

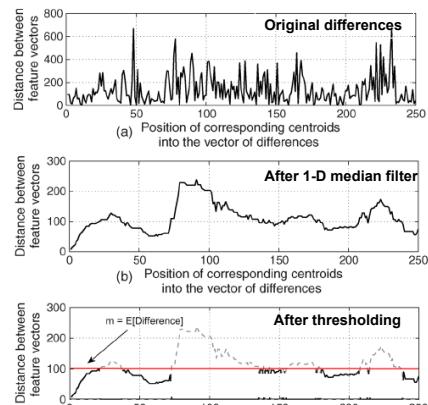
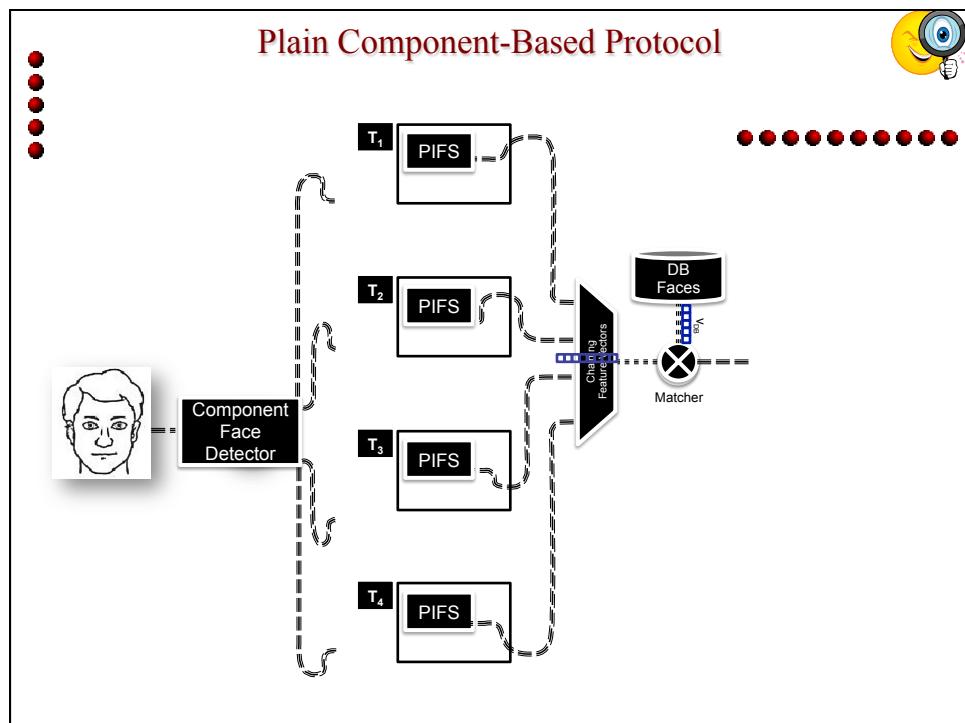
Feature Extraction

$V_1 = \{a_1, a_2, \dots, a_r\}$
 \dots
 $V_2 = \{b_1, b_2, \dots, b_s\}$
 \dots

Original differences
 Distance between feature vectors vs Position of corresponding centroids into the vector of differences

After 1-D median filter
 Distance between feature vectors vs Position of corresponding centroids into the vector of differences

After thresholding
 Distance between feature vectors vs Position of corresponding centroids into the vector of differences
 $m = E[\text{Difference}]$

FARO: FAce Recognition Against Occlusions and Expression Variations

Some results

AR-Faces database contains images of 126 persons (70males and 56 females). 13 image sets per session, which differ in expression
- 1 neutral, 2 smile, 3 anger, and 4 scream
illumination
- 5 left light, 6 right light, and 7 all side light
presence/absence of occlusions
- 8 sunglasses and 11 scarf combinations
- 9 sunglasses and left light, 10 sunglasses and right light, 12 scarf and left light, and 13 scarf and right light
Sets from 14 to 26 present the same conditions as 1–13, yet at a different time.
Neutral images from set 1 of the first session have been considered as the system gallery

Method	RANK				
	1	2	3	4	5
Set 5	0.96	0.98	0.99	1.00	1.00
Set 6	0.87	0.91	0.91	0.93	0.94
Set 8	0.90	0.93	0.95	0.96	0.97
Set 11	0.85	0.87	0.92	0.93	0.96
WFND	0.80	0.87	0.88	0.90	0.91
Set 6	0.77	0.85	0.93	0.94	0.96
Set 8	0.50	0.60	0.62	0.68	0.73
Weighted Fractal Neighbor	0.25	0.44	0.48	0.50	0.55
Distance	Set 5	0.77	0.87	0.89	0.88
LPP	Set 6	0.34	0.49	0.65	0.73
Set 8	0.32	0.44	0.52	0.58	0.62
Set 11	0.16	0.34	0.43	0.48	0.54
NPE	Set 5	0.89	0.91	0.93	0.96
Set 6	0.66	0.73	0.80	0.84	0.85
Set 8	0.44	0.51	0.58	0.61	0.64
Set 11	0.18	0.27	0.30	0.31	0.38



FACE: Face Analysis for Commercial Entities

Face Location

Face is detected by the Viola-Jones




[c] S.Milborrow, F. Nicolls., Locating facial features with an extended active shape model, European Conf. Computer Vision, pp. 504–513, 2008.

FACE: Face Analysis for Commercial Entities



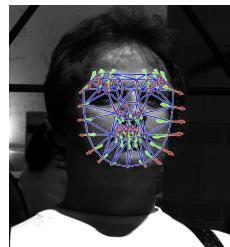
Face Location



Viola-Jones is also applied to locate eyes



The eyes are detected and used as guideline for the active shape model




68 facial points located on the face

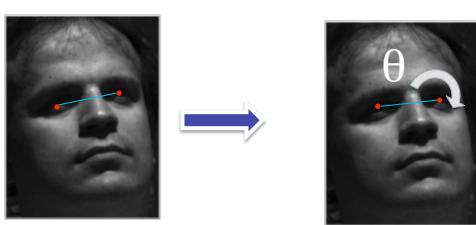
FACE: Face Analysis for Commercial Entities



Pose Correction



The pose correction process implemented in FACE is divided in different phases:

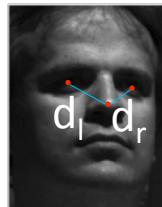


The centre of the eyes is used to correct head rolling

FACE: Face Analysis for Commercial Entities



Pose Correction



$$d_l > d_r$$



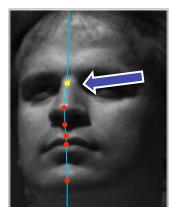
The distances between the external corners of the left and right eye and the tip of the nose, represented by d_l and d_r , allow to locate the better exposed half of the face

FACE: Face Analysis for Commercial Entities

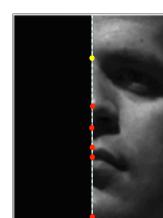


Pose Correction

The vertical points provide the exact face profile.



The highlighted is not provided by STAMS, but is computed as the median point of the link between the first two points upwards on nose sides



The face profile and the image borders delimit the right and left face regions.

A stretching operation is applied to all rows of the right region, so to obtain the same constant length.

FACE: Face Analysis for Commercial Entities

Pose Correction

The so processed right half face is divided into horizontal and vertical bands, which are delimited by the lines passing through the interest points

Lines are resized, to make these interest points fall in pre-determined positions

Finally, the left half face is reconstructed by reflecting the right half

FACE: Face Analysis for Commercial Entities

Illumination Correction

isotropic smoothing: useful information is lost on contour regions

anisotropic smoothing: It performs better than isotropic, but it is more computationally expensive.

It represents a good compromise between quality of the corrected images and efforts spent to obtain it.

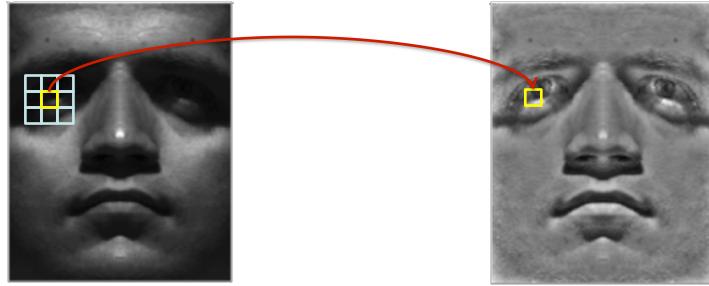
self quotient image

FACE: Face Analysis for Commercial Entities



Illumination Correction

FACE implements illumination correction using the SQI (Self Quotient Image) approach.



The value of each image pixel is divided by the mean of the values in its neighbourhood, represented by a square mask of size $k \times k$ in our case $k=8$)

FACE: Face Analysis for Commercial Entities



Matching

In its global form, given two images A and B and the respective mean values of their pixels \bar{A} and \bar{B} , it is defined as:

$$s(A, B) = \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (A(i, j) - \bar{A})(B(i, j) - \bar{B})}{\sqrt{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (A(i, j) - \bar{A})^2 \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (B(i, j) - \bar{B})^2}}$$

All correlation indices (local maxima) are then summed, to obtain a global correlation index.



FACE performs the matching by a localized version of the spatial correlation index.

FACE: Face Analysis for Commercial Entities



Matching



Correlation is computationally expensive

However some quantities can be pre-computed, and represent the biometric key

$$s(A, B) = \frac{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (A(i, j) - \bar{A})(B(i, j) - \bar{B})}{\sqrt{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (A(i, j) - \bar{A})^2} \sqrt{\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (B(i, j) - \bar{B})^2}}$$

Only the cross product has to be computed on-line

$$\sum_{i=0}^{n-1} \sum_{j=0}^{m-1} (A(i, j) - \bar{A})(B(i, j) - \bar{B})$$

FACE: Face Analysis for Commercial Entities



Some results



TABLE I
RR ACCURACY OF METHODS ACROSS FIVE DATA SETS

DB	Method						
	FACE	SVM	ISVM	PCA	ILDA	ICA	HMLBP
FERET	93%						
<i>fa</i>	(only LCM)	75%	78%	74%	79%	73%	84%
CDB	82%						
<i>(only LCM)</i>	44%	43%	27%	38%	31%	66%	
CDB (3 img)	60%						
<i>(only LCM)</i>	38%	42%	28%	45%	30%	51%	
LFW	54%						
<i>(only LCM)</i>	44%	42%	35%	46%	35%	47%	
SCface	81%						
<i>(only LCM)</i>	70%	69%	50%	62%	55%	73%	

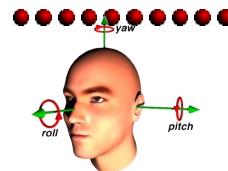
TABLE II
RR ACCURACY OF METHODS ACROSS FIVE DATA SETS

DB	Method						
	FACE	SVM	ISVM	PCA	ILDA	ICA	HMLBP
FERET	95%						
<i>fa</i>	79%	80%	77%	81%	78%	89%	
CDB	87%	55%	53%	29%	42%	28%	72%
CDB (3 img)	75%	45%	48%	33%	50%	33%	57%
LFW	61%	45%	42%	37%	48%	41%	49%
SCface	89%	80%	79%	65%	76%	73%	81%

Face Quality Measures



Pose



The index related to pose distortion is given by the linear combination of three components, which are respectively inversely proportional to roll, yaw, and pitch.

We are interested in an estimation of the influence of pose distortion on the recognition process.

All three factors range from 0 (almost no distortion) to 1 (worst situation, corresponding to higher distortion). The Sample Pose (SP) index is defined as a weighted linear combination:

$$SP = \alpha \cdot (1 - roll) + \beta \cdot (1 - yaw) + \gamma \cdot (1 - pitch)$$

with $\alpha + \beta + \gamma = 1$. We used training data to estimate how much each kind of deviation may contribute to the need for normalization. We found empirically that the set of values $\alpha = 0.1$, $\beta = 0.6$, $\gamma = 0.3$ work pretty well.

Face Quality Measures



Pose

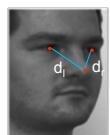


$$roll = \min\left(\left|\frac{2 \cdot \theta}{\pi}\right|, 1\right)$$

Roll is approximated with the angle θ between the line passing through the centres of the eyes, and the x-axis.

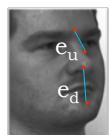


We then measure the left distance d_l and right distance d_r , between the external corner of each eye and the nose tip, which tend to be equal in a frontal pose.



$$yaw = \frac{\max(d_l, d_r) - \min(d_l, d_r)}{\max(d_l, d_r)}$$

A similar argument holds for the pitch, where the distances of concern are e_u and e_d , which are respectively the distances to the root of the nose and to the chin from the nose tip.



$$pitch = \frac{\max(e_u, e_d) - \min(e_u, e_d)}{\max(e_u, e_d)}$$

Face Quality Measures



Symmetry

Face is an object with a vertical symmetry axis that divides it into two parts which are almost perfectly symmetrical.

Let us consider any facial point $P_i(x_i, y_i)$ returned by the face detection module on the left half of the face at hand.

It is possible to consider the homologous $P_j(x_j, y_j)$ of this point on the right half face

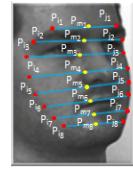


Each point along the line between these two extremes can be denoted as $P_k(x_k, y_k=r(x_k))$, where r is the line itself $r(x_k)=\varphi x_k+\omega$



Each point $L_i(x_i+t, r(x_i+t))$ on the $\overline{P_i P_m}$ half of the segment $P_i P_j$, has its corresponding point $R_i(x_i-t, r(x_i-t))$ on the other half $P_m P_j$ of the same segment.

The symmetry score of the pair of points $< P_i(x_i, y_i), P_j(x_j, y_j) >$ is defined by: $sym(P_i, P_j) = \sum_t \left(\frac{\min(I(L_i), I(R_i))}{\max(I(L_i), I(R_i))} \right)$



The global symmetry score SY for the whole face is computed as the sum of the symmetry scores of a limited set $X=\{(i,j)\}$ of predetermined pairs of points:

$$SY = \sum_{(i,j) \in X} sym(P_i, P_j).$$

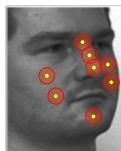
Face Quality Measures



Illumination

When a face is lit up in an optimal way, some face regions present a quite similar grey level distribution (grey level histogram) since they are smoothed surfaces.

We identified eight reference points, which are related to those face regions.

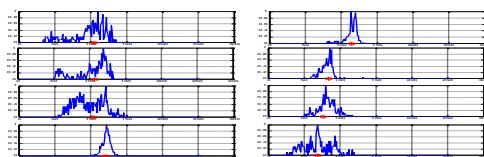


For each point, we select a portion of the image which is proportional to the square containing the face. For each portion w , we compute the histogram h and its centre of mass, using the formula:

$$mc(w) = \left(\sum_{i=0}^{255} i \cdot h_w(i) \right) / \sum_{i=0}^{255} h_w(i)$$

SI is then defined as a scalar computed from the variance of the elements of vector mc .

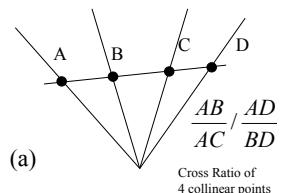
$$SI = 1 - F(std(mc))$$



Anti-spoofing for Moving Faces



They are shape descriptors, that are not affected by object pose and scale, by perspective projection and intrinsic parameters of the camera. They are expressed as Ratios of distances/measures or as a combination of 3D/2D coordinates of the points of the object



Cross Ratio of 4 collinear points

Geometric Invariants

$$c = \frac{M(1,2,4)M(1,3,5)}{M(1,2,5)M(1,3,4)}$$

$$M(j,k,l) = \begin{pmatrix} x_j & x_k & x_l \\ y_j & y_k & y_l \\ 1 & 1 & 1 \end{pmatrix}$$

$$c = \frac{P_{A^ABD} \cdot P_{F^E}}{P_{A^ABC} \cdot P_{F^ED}}$$

All these invariants can be calculated from a single view of the object, then we can refer to them as 2D/3D invariants.

They can be used as a preliminary coarse grain description of the object shape (face).

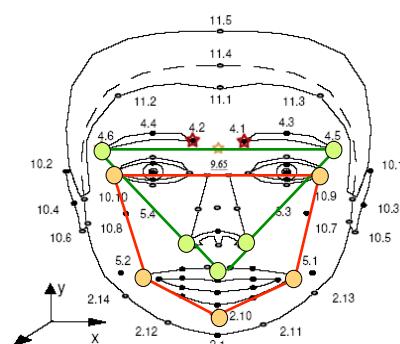
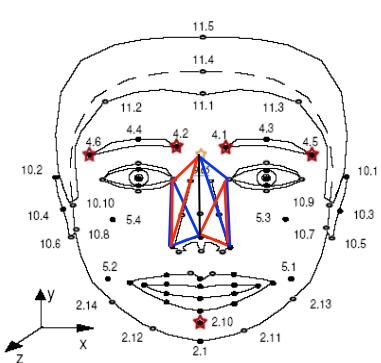
Anti-spoofing for Moving Faces



Face is not a rigid surface and to find control points which both respect the required hypotheses and that are easy to locate, turns in a difficult task.

Cross Ratio of the area of triangles on adjacent planes

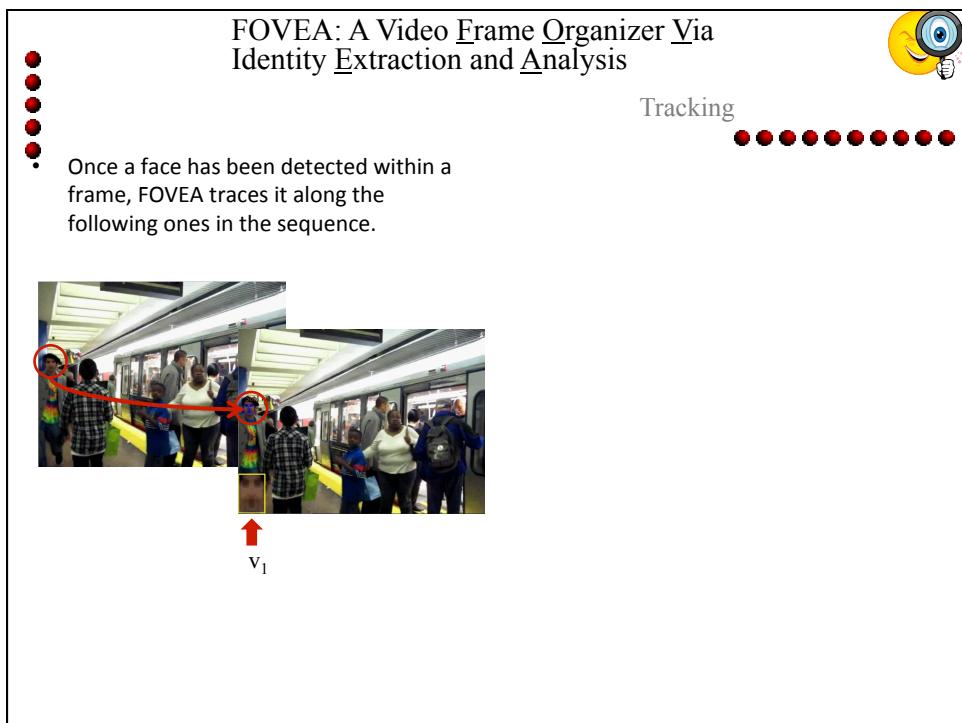
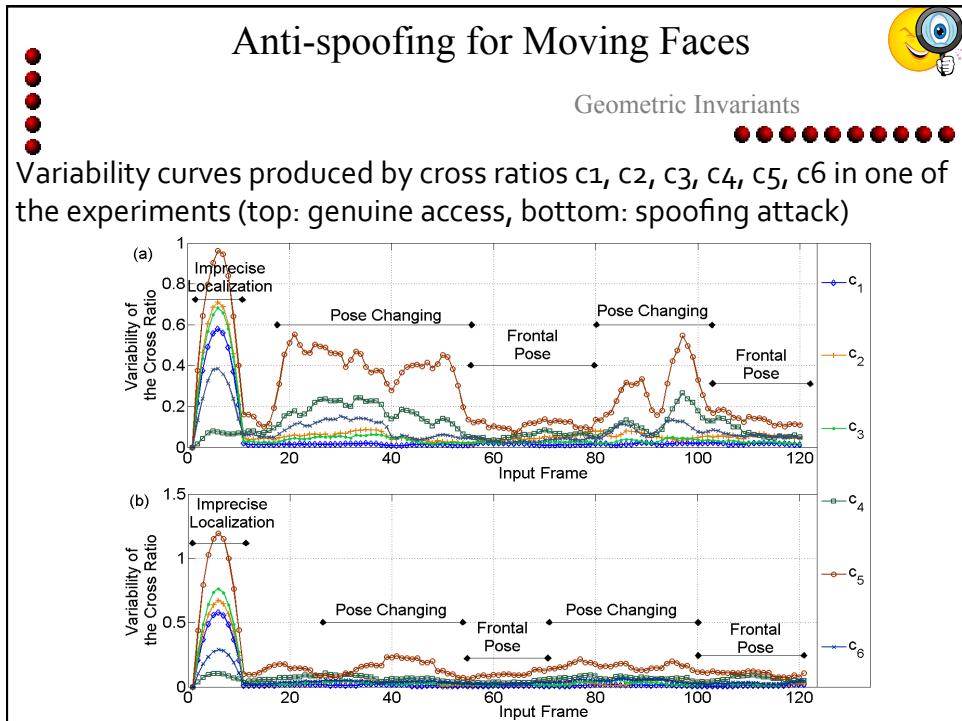
— 9.65, 3.8, 9.2, 9.3, 9.1 and 3.11



Cross Ratio of 5 coplanar points

— 4.6, 9.2, 8.1, 9.1 and 4.5

— 3.12, 8.4, 2.10, 8.3 and 3.7

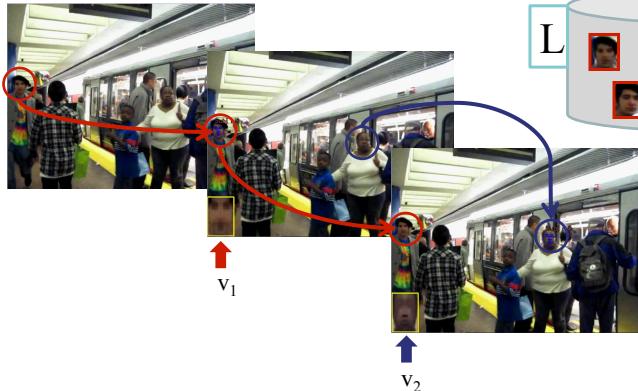


FOVEA: A Video Frame Organizer Via Identity Extraction and Analysis



- Once a face has been detected within a frame, FOVEA traces it along the following ones in the sequence.

- The tracking maintains a set V of the faces being currently examined, which is updated on a frame-by-frame basis.

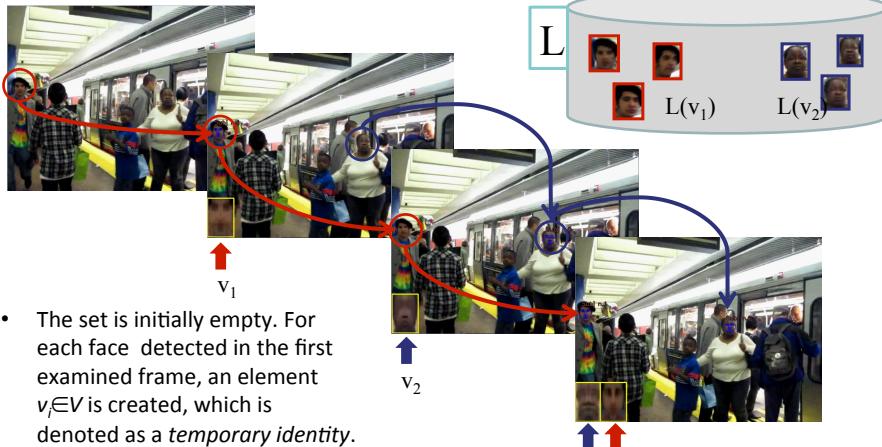


FOVEA: A Video Frame Organizer Via Identity Extraction and Analysis



- Once a face has been detected within a frame, FOVEA traces it along the following ones in the sequence.

- The tracking maintains a set V of the faces being currently examined, which is updated on a frame-by-frame basis.



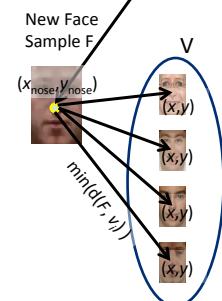
- The set is initially empty. For each face detected in the first examined frame, an element $v_i \in V$ is created, which is denoted as a *temporary identity*.

FOVEA: A Video Frame Organizer Via Identity Extraction and Analysis



Temporary Identities

- For each *temporary identity*, the coordinates $(x_{\text{nose}}, y_{\text{nose}})$ of the nose tip are recorded together with a list $L(v_i)$ of its possible further samples that may appear later in consecutive frames.
- The face tracking module computes the distance $d(F, v_i)$ between coordinates $(newx_{\text{nose}}, newy_{\text{nose}})$ of new sample and those of each *temporary identity* v_i already in the set.
- If there exist an identity $v_i \in V$ for which $d(F, v_i)$ is lower than a threshold th_{nose} , the new sample is assigned to $L(v_i)$. Otherwise, a new element is added to V .
- The coordinates $(x_{\text{nose}}, y_{\text{nose}})$ of the matching identity are updated with $(newx_{\text{nose}}, newy_{\text{nose}})$.

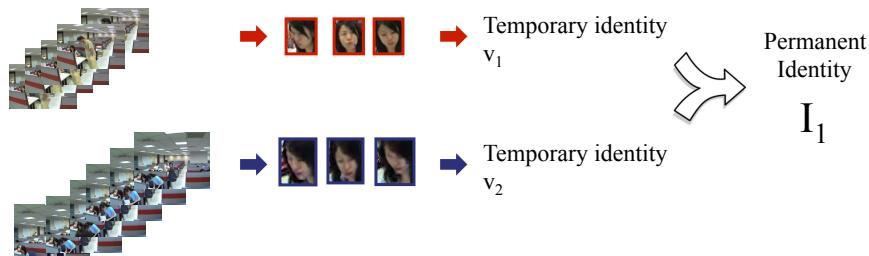


FOVEA: A Video Frame Organizer Via Identity Extraction and Analysis



Temporary Vs. Permanent

- FOVEA distinguishes between two kind of identities:
 - temporary identity* – it is an identity created to temporarily represent a set of samples with coherent positions within temporally close frames;
 - permanent identity* – it represents the final outcome of the classification process.



FOVEA: A Video Frame Organizer Via
Identity Extraction and Analysis



Identity Mapping



- When no instance (face sample) can be associated to the *temporary identity* v_i for F_{not} consecutive frames, such identity must be mapped onto a *permanent identity* I_k , through an *identity mapping* operation:
 - permanent identity search;
 - reduction of representative face samples contained in the gallery;



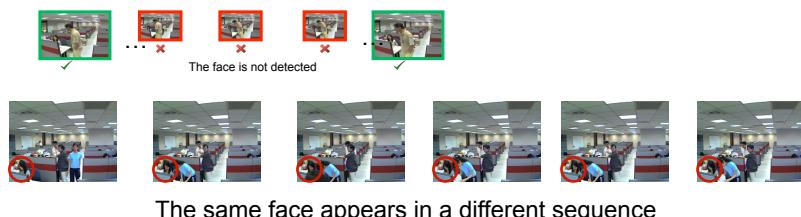
FOVEA: A Video Frame Organizer Via
Identity Extraction and Analysis



Identity Mapping



- When no instance (face sample) can be associated to the *temporary identity* v_i for F_{not} consecutive frames, such identity must be mapped onto a *permanent identity* I_k , through an *identity mapping* operation:
 - permanent identity search;
 - reduction of representative face samples contained in the gallery;

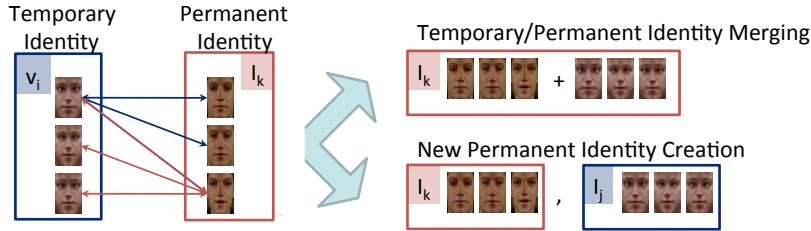


FOVEA: A Video Frame Organizer Via Identity Extraction and Analysis



Identity Mapping

- An *identity mapping* operation consists in:
 - comparing the samples of *temporary identity* v_i with those of existing *permanent identities* I_j , and possibly creating a new permanent identity I_k ;
 - selecting the most representative samples for identity I_k , if this has been newly created from v_i , or merging samples associated to v_i with those of the I_k with which v_i has been identified, by selecting the most representative ones from both sets.

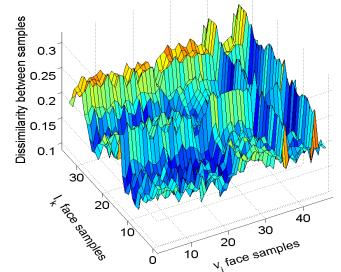
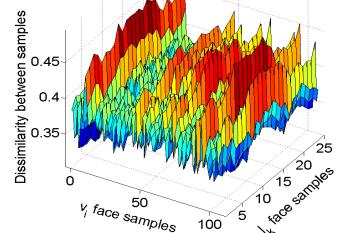


FOVEA: A Video Frame Organizer Via Identity Extraction and Analysis



Identity Mapping

- Compatibility between identities is measured by computing a distance matrix $M(n, m)$, where $n = |L(v_i)|$ and $m = |L(I_k)|$.
 M is defined such that $M(h, j) = 1 - S(v_{i,h}, I_{k,j})$.



- FOVEA counts the number C_M of face samples (of v_i and I_k) that match. C_M is normalized with respect to the number of considered pairs, so to obtain a value in the interval $[0, 1]$.
- A threshold th_M , which is set by the user, allows to decide if *temporary identity* v_i can be associated to the *permanent identity* I_k ; the association will be performed if $C_M > th_M$.

FOVEA: A Video Frame Organizer Via
Identity Extraction and Analysis



Identity Mapping

- FOVEA takes into account the information given by the concomitant appearance of more subjects within the same scene.
- This information allows to decide with certainty that two identities are not compatible, reducing the algorithm computational cost.
- Two different strategies exist:
 - off-line – all extracted temporary identities are mapped onto permanent identities at the end of the overall extraction process from the video sequence;
 - on-line – temporary identities are mapped onto permanent identities as they are identified.

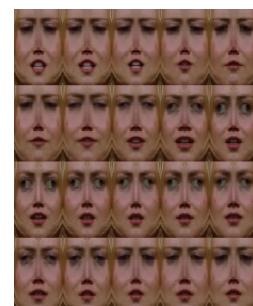
FOVEA: A Video Frame Organizer Via
Identity Extraction and Analysis



Sample Selection

- FOVEA identifies the row h_{min} with the lowest mean value and the corresponding face sample enters the new set $L'(I_k)$ for the permanent identity I_k .

- In order to increase intra-class variability, it scans all entries in h_{min} row and, select the face sample $I_{k,j} \in L(I_k)$ to be included into $L'(I_k)$, only if the distance between face samples j and k is lower than a given threshold th_s .



Before Merging

After Merging



- In a symmetric way, FOVEA operates on columns.
- At the end of this process, the set $L(I_k)$ for the permanent identity I_k is substituted with $L'(I_k)$.

Face Recognition at BIPLAB



References

M. De Marsico, M. Nappi, D. Riccio. FARO: FAcE Recognition Against Occlusions and Expression Variations. IEEE Transactions on Systems, Man, and Cybernetics — Part A: Systems and Humans, Vol. 40, No. 1, Jan. 2010, pp. 121-132

M. De Marsico, M. Nappi, and D. Riccio. Face Authentication with Undercontrolled Pose and Illumination. *Signal Image and Video Processing*. - Vol. 5 N. 4 - Special Issue on Unconstrained Biometrics: Advances and Trends. pp. 401-413, November 2011

M. De Marsico, M. Nappi, D. Riccio, H. Wechsler. Robust Face Recognition for Uncontrolled Pose and Illumination Changes. Accepted for publication on IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 43, No. 1, January 2013, pp. 149-163

M. De Marsico, G. Doretto, D. Riccio. M-VIVIE: a Multi-thread Video Indexer Via Identity Extraction. Pattern Recognition Letters – Special Issue on Novel Pattern Recognition-Based Methods for Reidentification in Biometric Context- BIOCON – Volume 33, Issue 14, October 2012, pp. 1882–1890

M. De Marsico, M. Nappi, D. Riccio. FACE: Face Analysis for Commercial Entities. Proceedings of ICIP, 2010 International Conference on Image Processing, September 26-29 2010, Honk Kong, pp. 1597-1600

M. De Marsico, M. Nappi, D. Riccio. Measuring Sample Distortions in Face Recognition. Proceedings of Multimedia in Forensics, Security and Intelligence - MiFor2010 - October 2010 Firenze, Italy - held in conjunction with ACM MultiMedia 2010, pp. 83-88

M. De Marsico, M. Nappi, D. Riccio, H. Wechsler. Robust Face Recognition after Plastic Surgery Using Local Region Analysis. Proceedings of International Conference on Image Analysis and Recognition - ICIAR 2011, June 22-24, 2011, Burnaby, BC, Canada, Lecture Notes in Computer Science, 2011, Volume 6754/2011, pp. 191-200

Face Recognition at BIPLAB



References

A. Abate, M. De Marsico, M. Nappi, D. Riccio. VIVIE: A Video-surveillance Indexer Via Identity Extraction. In Proceedings of IEEE International Workshop on Advances in Automated Multimedia Surveillance for Public Safety (AAMS-PS 2011) In conjunction with IEEE International Conference on Multimedia and Expo (ICME 2011) July 15, 2011, Barcelona, Spain.

M. De Marsico, M. Nappi, D. Riccio. FOVEA: A video frame organizer via identity extraction and analysis . Proceedings of BIOMS 2011- 2011 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications. Milan, Italy, 28 September 2011. pp. 1 – 4.

M. De Marsico, M. Nappi, D. Riccio. Measuring Measures for Face Sample Quality. Proceedings of International ACM Workshop on Multimedia in Forensics and Intelligence (MiFor'11) - November 29th, 2011 Scottsdale, Arizona - USA - held in conjunction with ACM MultiMedia 2011, pp. 7-12.

M. De Marsico, M. Nappi, D. Riccio, J.-L. Dugelay. Moving Face Spoofing Detection via 3d Projective Invariants. Proceedings of the 5th IAPR International Conference on Biometrics - ICB 2012, March 29 - April 1, 2012 New Delhi, India, pp.1-6.

S. Barra, M. De Marsico, C. Galdi. D. Riccio, H. Wechsler. FAME: Face Authentication for Mobile Encounter. 2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS 2013), Napoli (Italy), September 9 2013.

M. De Marsico, M. Nappi, D. Riccio, S. Vitulano. Face Searching in Large Databases. In: Yu-Jin Zhang (Editor): Advances in Face Image Analysis: Techniques and Technologies, pp. 16-40. Medical Information science reference. IGI Global, Hershey • New York, 2010.

M. De Marsico, M. Nappi, D. Riccio. Biometrics Using 3D Vision Techniques. In: Song Zhang (Editor). Handbook of 3D Machine Vision: Optical Metrology and Imaging, pp. 361-385. CRC Press, Taylor & Francis Group, Abingdon, UK, 2013

Ear Recognition



• • • • • • • •

•
•
•
•

HERO: Human Ear Recognition against Occlusions



PIFS

• • • • • • • •

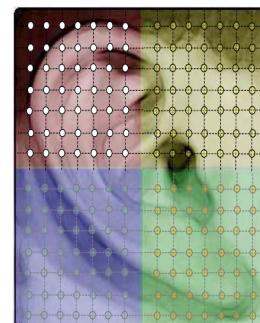
•
•
•
•

The image is split in four subregions (upper-left, lowerleft, upper-right and lower-right squares) we call tiles

In order to get the system robust with respect to local changes, the four tiles are independently indexed by means of the PIFS systems as separate regions of interest.

For each of the segmented regions a set of entry points is selected.

An entry point locates the upper-left corner of a range to be matched by domains extrated from the same region.



HERO: Human Ear Recognition against Occlusions

The diagram illustrates the HERO feature extraction process. It starts with an **Input Image** (a grayscale ear image) which is divided into **Tiles** (small square regions). These tiles are processed by **PIFS** (Pixel Intensity Feature Selection) to identify domains and ranges. An **Affine Transform** is applied to each tile. The resulting **Centroids** (represented as circles with centers C_i) are then compared using a **Peano key** to form a **Feature Vector**.

Extracted Features

Three real values are associated to each centroid in the list:

- X_i is the X coordinate of the centroid C_i ;
- Y_i is the Y coordinate of the centroid C_i ;
- Var_i is the mean value of the domain variance, computed on the domains belonging to the cluster of C_i .

Comparing Features

The centroid C_i in the Image 1 is compared only with the centroid C_j in the Image 2, whose position (X_j, Y_j) is the closest to (X_i, Y_i) , in an Euclidean sense.

Ear Recognition by means of a Rotation Invariant Descriptor

The ear region is extracted from the profile face by means of a Haar based object detector. The edge map of the region is obtained by means of the Canny's operator and its mass center is computed. A circular region centered in (X_c, Y_c) is cropped and resized to a 64 pixel radius.

$$X_C = \frac{1}{M} \cdot \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} i \cdot I(i, j)$$
 where $M = \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} I(i, j)$

$$Y_C = \frac{1}{M} \cdot \sum_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} j \cdot I(i, j)$$

Ear Recognition by means of a Rotation
Invariant Descriptor



Feature Extraction



Polar Fourier Transform: However, problems arise also applying the 2-D FT directly

$$PF_2(\rho, \phi) = \sum_r \sum_{\theta_i} f(r, \theta_i) \times e^{[j2\pi\left(\frac{r}{R}\rho + \frac{2\pi i}{T}\phi\right)]}$$

where $0 \leq r < R$ and $\theta_i = i(2\pi/T)$, $0 \leq i < T$; $0 \leq \rho < R$ and $0 \leq \phi < T$. R and T are the radial and angular resolutions. $f(x, y)$ is a binary function in shape application. The ρ and ϕ are simply the number of radial frequencies selected and the number of angular frequencies selected.

Ear Recognition by means of a Rotation
Invariant Descriptor

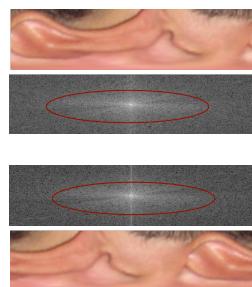
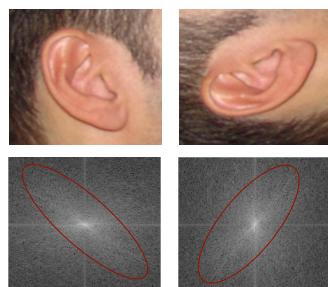


Feature Extraction



The PF operator is then applied on the normalized image, selecting n radii and m angles, and resulting real values are organized in a linear feature vector as follows:

$$GFD = \left[\frac{|PF_2(0,0)|}{area}, \frac{|PF_2(0,1)|}{|PF_2(0,0)|}, \dots, \frac{|PF_2(0,n)|}{|PF_2(0,0)|}, \dots, \frac{|PF_2(m,0)|}{|PF_2(0,0)|}, \dots, \frac{|PF_2(m,n)|}{|PF_2(0,0)|} \right]$$



Ear Recognition at BIPLAB



BIPLAB

References



M. De Marsico, M. Nappi, D. Riccio. HERO: Human Ear Recognition Against Occlusions. Proceedings of IEEE Computer Society Workshop on Biometrics - In association with IEEE Conference on Computer Vision and Pattern Recognition - CVPR 2010 –18 June San Francisco – USA, pp. 178-183.

A. F. Abate, M. Nappi, S. Ricciardi, D. Riccio, Ear Recognition by means of a Rotation Invariant Descriptor, Proceedings of the Int. Conf. On Pattern Recognition ICPR 2006, Hong Kong, 20-24 August 2006.

M. De Marsico, M. Nappi, D. Riccio. Unconstrained ear processing: what is possible and what must be done. In: Signal and Image Processing for Biometrics : State of the Art and Recent Advances. To be published by Springer.

Iris Recognition



ISIS: Iris Segmentation for Identification Systems



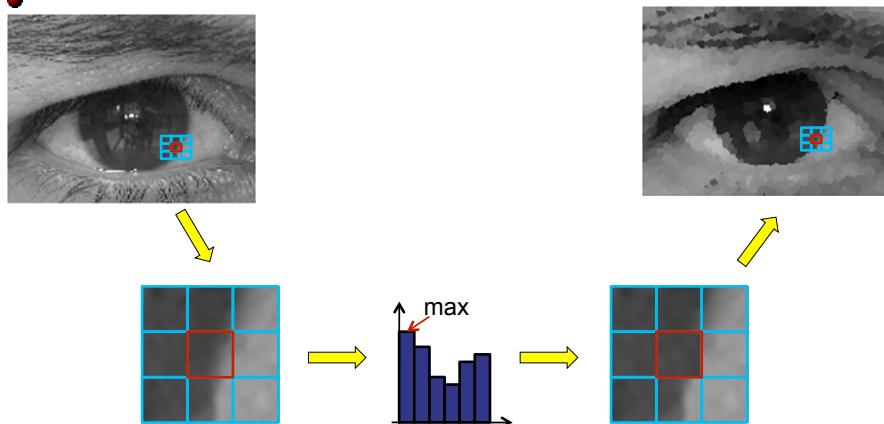
Preprocessing

- Details of sclera vessels, skin pores, or eyelashes shape = complex patterns that can negatively interfere with edge detection
- A posterization filter FE (*Enhance*) is applied
 - a square window W is moved over the whole image, pixel by pixel
 - a histogram h_W is computed for the region in W
 - the value with the maximum occurrence is substituted for the central position.

ISIS: Iris Segmentation for Identification Systems



Preprocessing



Posterization: example

ISIS: Iris Segmentation for Identification Systems

Preprocessing

- Canny filtering is applied with ten different thresholds $th=0.05, 0.10, 0.15, \dots, 0.55$.

(a) $th=0.05$
(b) $th=0.15$
(c) $th=0.35$

ISIS: Iris Segmentation for Identification Systems

Pupil Location

Many approaches search circular shapes through Hough transform or its adaptations = high computational cost.

Using ellipse fitting, the presence of noise (e.g. spurious branches by Canny filter) may cause the erroneous detection of an elliptical shape even where the expected result would be a (quasi)circular one.

Our algorithm detects circular objects using a precise and fast *circle detection* procedure presented by Taubin.

The pupil is not a perfect circle, but searching for a circle causes a lower error than obtaining a noise-conditioned ellipse.

Taubin's method

Ellipse Fitting

Taubin G., Estimation Of Planar Curves, Surfaces And Nonplanar Space Curves Defined By Implicit Equations, With Applications To Edge And Range Image Segmentation, IEEE Trans. on PAMI, vol. 13, pp 1115-1138, 1991.

ISIS: Iris Segmentation for Identification Systems



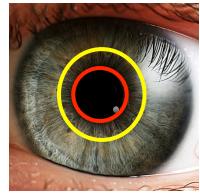
Candidate Pruning

Many circles are found while searching for the pupil.

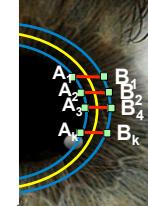
IS_{IS} adopts two ranking criteria to only select the best candidate pupil circle:

Homogeneity – Each circular region gets a score according to the homogeneity of its pixels

$$s_H = \max_i[H(i)] / \sum_{i=0}^{255} H(i)$$



Separability – Given a candidate circle C we measure the difference between the grey levels of corresponding pixels on circle that includes/is included in C.



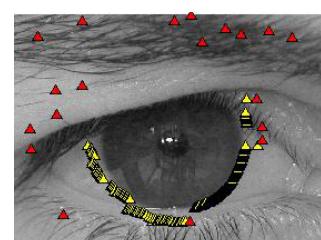
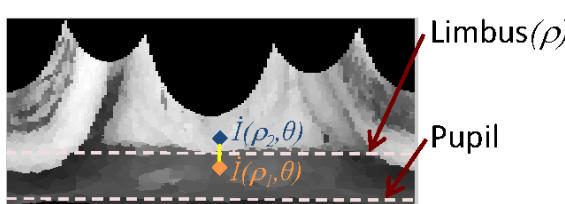
ISIS: Iris Segmentation for Identification Systems



Linearization

The eye image is transformed from Cartesian coordinates space to polar coordinates space.

Along the vertical direction, it is possible to identify in an extremely precise way the limbo boundary region which separates iris from sclera.



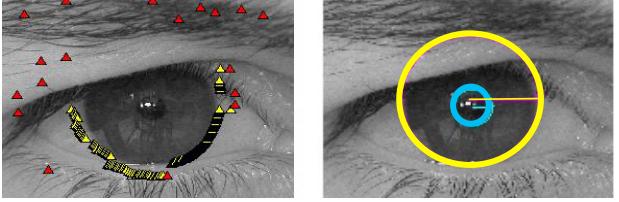
ISIS: Iris Segmentation for Identification Systems



Limbus Detection



- Points in F belong to a polar space = their ρ component should remain about constant while θ describes an approximate circle
- Smoothness criterion = avoid outliers
- We compute the median value value ρ_{med} over F , and compute a relative error

$$err = \frac{|\rho_i - \rho_{med}|}{\max_i |\rho_i - \rho_{med}|}$$


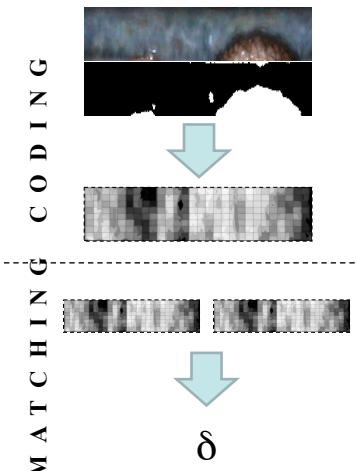
Iris Matching for Identification Systems



Pipeline



- **Coding**
 - **Input:**
 - Normalized iris and mask
 - **Output:** coding
 - **Function:** create a biometric template
- **Matching**
 - **Input:**
 - Probe coding, gallery coding
 - **Output:** distance value
 - **Function:** compute the lowest distance



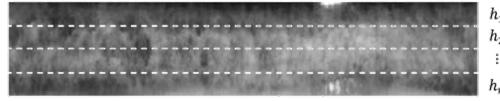
Iris Matching for Identification Systems

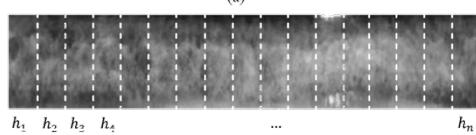


Partitioned LBP



- Sun et al.* divide the normalized iris in blocks.
- We tried the division in horizontal (anatomically meaningful) or vertical bands.


(a)


Code= Set of histograms + Mask

*Sun Z., Tan T., Qiu X.^(b) 2006. Graph Matching Iris Image Blocks with Local Binary Pattern. In Proceedings of the International Conference on Biometrics, pp.366–372

Iris Matching for Identification Systems



LBP Matching

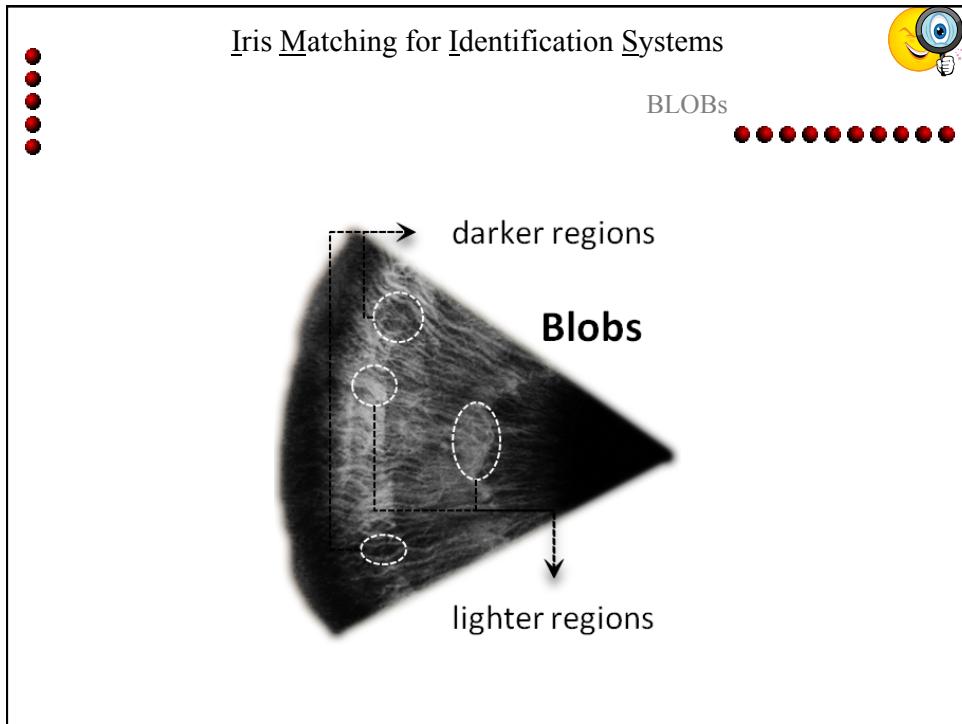


Matching:

- given $C = (H_1, H_2, \dots, H_n, M) \quad P = (K_1, K_2, \dots, K_n, N)$
- given a measure of similarity between histograms δ
- similarity between codes will be:

$$\frac{1}{n} \sum_{i=1}^n \delta(H_i, K_i) \left(1 - \frac{\overline{\text{noise}_i}}{\text{totpixel}} \right)$$

- As we will see from results:
 - Horizontal division better than vertical one
 - No too narrow bands



Iris Matching for Identification Systems

BLOB Detection

- ❖ Furrows and crypts make up the blobs
- ❖ Use of Laplacian of Gaussian (LoG)*
- Gaussian:

$$f(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2+y^2}{2t}}$$

- Laplacian:
 - 2° order differential operator
 - 2D is defined by:

$$\Delta = \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

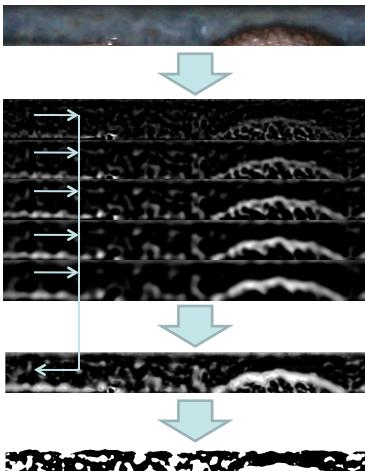
*Chenhong L., Zhaoyang L., 2008. Local feature extraction for iris recognition with automatic scale selection. Image and Vision Computing, vol. 26, no. 7, pp. 935–940.

Iris Matching for Identification Systems



- Different scales
- Normalization
- Fusion:
 - Point with higher LOG
- Binarization
- Matching = Hamming distance weighted by segmentation mask
- N-IRIS also considers shifts of 10 pixels -> the final distance = computed on the alignment returning the maximum match

Multi-scale BLOBs



$$\nabla_{\text{norm}}^2 f = t \cdot (\nabla^2 f)$$

Iris recognition at BIPLAB



References

- M. De Marsico, M. Nappi, D. Riccio, H. Wechsler. Iris segmentation using pupil location, linearization, and limbus boundary reconstruction in ambient intelligent environments. Journal of Ambient Intelligence and Humanized Computing. Vol. 2, 2011, pp. 153-162
- M. De Marsico, M. Nappi, and D. Riccio. Noisy Iris Recognition Integrated Scheme. Pattern Recognition Letters- Special issue on the Recognition of Visible Wavelength Iris Images Captured At-a-distance and On-the-move Volume 33, Issue 8, 1 June 2012, Pages 1006–1011
- M. De Marsico, M. Nappi, D. Riccio. ISIS: Iris Segmentation for Identification Systems. Proceedings of 20-th International Conference on Pattern Recognition - ICPR 2010 – 23-26 August 2010 – Istanbul – Turkey, pp 2857-2860
- D. Riccio, M. De Marsico. Iris Recognition in Visible Light Domain. Proceedings of ICPRAM 2012 - 1-st International Conference on Pattern Recognition Applications and Methods. Vilamoura, Algarve, Portugal, 6-8 February 2012, pp. 55-62.

Multi Biometric Systems Normalization



What about data normalization?



- A number of different solutions have been proposed in literature to solve this problem.

Normalization Functions

$$\text{Min/Max} \quad s'_k = \frac{s_k - \min}{\max - \min}$$

$$\text{Z-score} \quad s'_k = \frac{s_k - \mu}{\sigma}$$

$$\text{Median/Mad} \quad s'_k = \frac{s_k - \text{median}}{\text{MAD}}$$

$$\text{Sigmoid} \quad s'_k = \frac{1}{1 + ce^{-ks_k}}$$

$$\text{Tanh} \quad s'_k = \frac{1}{2} \left[\tanh \left(0.01 \frac{(s_k - E[s_k])}{\sigma(s_k)} \right) + 1 \right]$$

- When minimum and maximum values are known, the normalization process is trivial.
- For this reason, we assumed to miss an exact estimate of the maximum value
- We chose the average value in its place, in order to stress normalization functions even more.

Testing the existing normalization functions

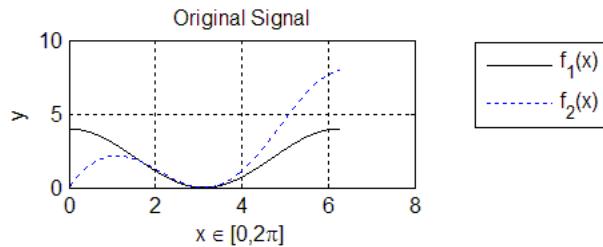
- we chose the two following test functions:

$$f_1(x) = 2 \cdot (\cos(x) + 1)$$

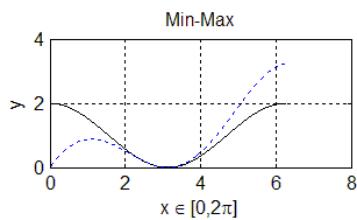
and

$$f_2(x) = 2 \cdot \log(x+1) \cdot (\cos(x) + 1)$$

in $[0, 2\pi]$ interval.



The Min/Max Function



The **Min-max** normalization technique performs a “mapping” (shifting + compression/dilation) of the interval between the minimum and maximum values in the interval between 0 and 1

Normalization Functions

Min/Max $s'_k = \frac{s_k - \min}{\max - \min}$

Z-score $s'_k = \frac{s_k - \mu}{\sigma}$

Median/Mad $s'_k = \frac{s_k - \text{median}}{\text{MAD}}$

Sigmoid $s'_k = \frac{1}{1 + ce^{-ks_k}}$

Tanh $s'_k = \frac{1}{2} \left[\tanh \left(0.01 \frac{(s_k - E[s_k])}{\sigma(s_k)} \right) + 1 \right]$

Such technique assumes that the minimum and maximum ever generated by a matching module are known.

The Z-Score function

Normalization Functions

Min/Max	$s'_k = \frac{s_k - \min}{\max - \min}$
Z-score	$s'_k = \frac{s_k - \mu}{\sigma}$
Median/Mad	$s'_k = \frac{s_k - \text{median}}{\text{MAD}}$
Sigmoid	$s'_k = \frac{1}{1 + ce^{-ks_k}}$
Tanh	$s'_k = \frac{1}{2} \left[\tanh \left(0.01 \frac{(s_k - E[s_k])}{\sigma(s_k)} \right) + 1 \right]$

The **Z-score** technique is the most widespread and uses arithmetic average and standard deviation of scores returned by the single subsystem.

μ represents the arithmetic average of scores and σ is the standard deviation.

Z-score does not guarantee a common interval for normalized values coming from different subsystems.

The Median/MAD function

Normalization Functions

Min/Max	$s'_k = \frac{s_k - \min}{\max - \min}$
Z-score	$s'_k = \frac{s_k - \mu}{\sigma}$
Median/Mad	$s'_k = \frac{s_k - \text{median}}{\text{MAD}}$
Sigmoid	$s'_k = \frac{1}{1 + ce^{-ks_k}}$
Tanh	$s'_k = \frac{1}{2} \left[\tanh \left(0.01 \frac{(s_k - E[s_k])}{\sigma(s_k)} \right) + 1 \right]$

The **Median/MAD** technique uses the median and the MAD (median of absolute values).

Median/MAD is less effective, most of all when values have a non-Gaussian distribution; in such cases it neither preserves the original value distribution nor transforms the values in a common numeric interval.

The Sigmoid function

A Sigmoid function has the open interval $(0,1)$ as codomain.

It has two drawbacks:

- the distortion introduced by the function when x tends to the extremes of the interval is excessive;
- the shape of the function depends on the two parameters c and k that in turn strongly depend on the domain of x parameter.

Normalization Functions

Min/Max	$s'_k = \frac{s_k - \min}{\max - \min}$
Z-score	$s'_k = \frac{s_k - \mu}{\sigma}$
Median/Mad	$s'_k = \frac{s_k - \text{median}}{\text{MAD}}$
Sigmoid	$s'_k = \frac{1}{1 + ce^{-ks_k}}$
Tanh	$s'_k = \frac{1}{2} \left[\tanh \left(0.01 \frac{(s_k - E[s_k])}{\sigma(s_k)} \right) + 1 \right]$

The Tanh function

The Tanh function guarantees data to be projected in the open interval $(0,1)$.

It excessively concentrates values around the centre of the interval (0.5) .

Normalization Functions

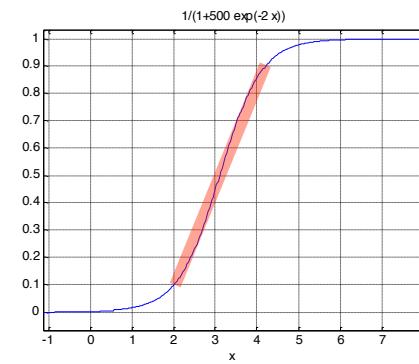
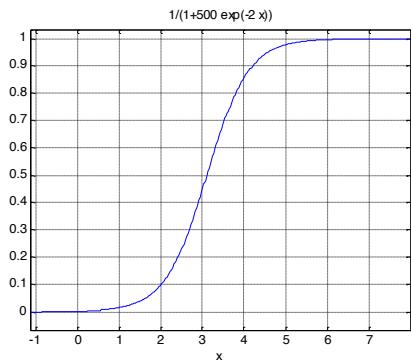
Min/Max	$s'_k = \frac{s_k - \min}{\max - \min}$
Z-score	$s'_k = \frac{s_k - \mu}{\sigma}$
Median/Mad	$s'_k = \frac{s_k - \text{median}}{\text{MAD}}$
Sigmoid	$s'_k = \frac{1}{1 + ce^{-ks_k}}$
Tanh	$s'_k = \frac{1}{2} \left[\tanh \left(0.01 \frac{(s_k - E[s_k])}{\sigma(s_k)} \right) + 1 \right]$

A new normalization function *Quasi-Linear Sigmoid (QLS)*

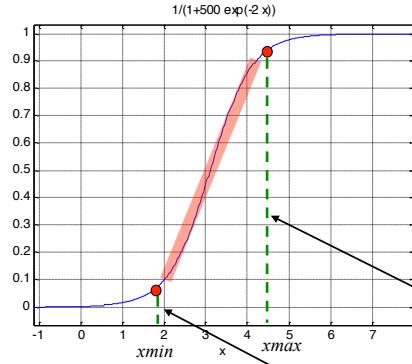
- The desired properties of a new normalization function are:
 - The (0,1) codomain;
 - Minimal distortion of the input data distribution.
 - High robustness to imprecise maximum estimations.
 - A limited number of parameters.

A new normalization function

- It is possible to reduce the distortion of the Sigmoid function $s'_k = \frac{1}{1+e^{-kx}}$ by deriving a new function $F(x)$ from $f(x)$, with a pseudo-linear behaviour in the whole codomain though preserving the property such that $F(x) \in [0,1]$



Quasi-Linear Sigmoid (QLS)



[x_{min}, x_{max}] is the range in which the sigmoidal function assumes a pseudo-linear trend.

- We find the null points of the third derivative:

$$f'''(x) = 6 \frac{c^3 k^3 e^{(-kx)^3}}{(1 + ce^{-kx})^4} - 6 \frac{c^2 k^3 e^{(-kx)^2}}{(1 + ce^{-kx})^3} + \frac{ck^3 e^{(-kx)}}{(1 + ce^{-kx})^2}$$

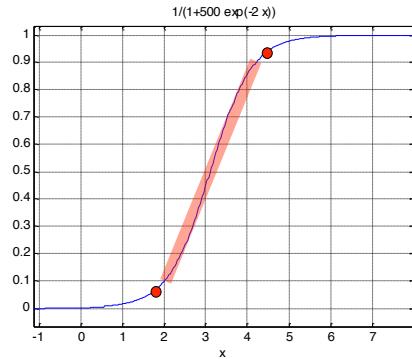
- Which are

$$x_{max} = -\frac{1}{k} \log\left(\frac{2 - \sqrt{3}}{c}\right)$$

- And

$$x_{min} = -\frac{1}{k} \log\left(\frac{2 + \sqrt{3}}{c}\right)$$

Quasi-Linear Sigmoid (QLS)



- Knowing that $x_{min}=0$ and combining the two equations we can write:

$$c = 2 + \sqrt{3}$$

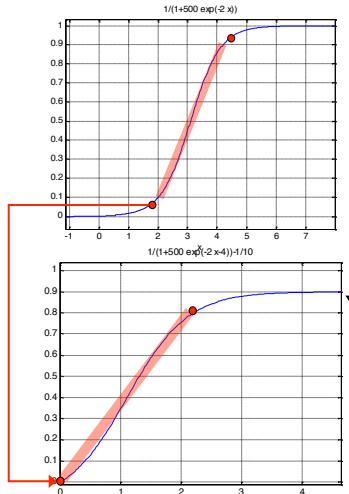
- And

$$k = -\frac{1}{x_{max}} \log\left(\frac{2 - \sqrt{3}}{2 + \sqrt{3}}\right)$$

x_{max} is the only parameter we have to know.



Mapping $f(x_{min})$ to 0



- To map $f(x_{min})$ to 0 we define a new function:

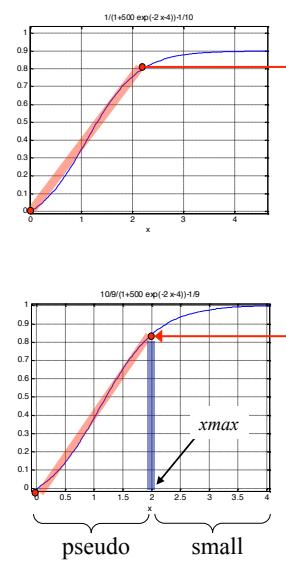
$$g(x) = f(x) - f(x_{min})$$

$$= f(x) - f(0)$$

The upper limit of the function $g(x)$ has to be mapped on 1.



Mapping $f(\infty)$ to 1



- To map $f(\infty)$ to 1 we compute:

$$L = \lim_{x \rightarrow \infty} g(x) = \frac{2 + \sqrt{3}}{3 + \sqrt{3}}$$

- and, finally, we define:

$$F(x) = \frac{1}{L} g(x) = \frac{1 - b^{\frac{x}{x_{max}}}}{ab^{\frac{x}{x_{max}}} + 1}$$

- with

$$a = (2 + \sqrt{3}) \quad \text{and} \quad b = (7 - 4\sqrt{3})$$

A.F.Abate, M.Nappi, D.Riccio, M.DeMarsico, Data Normalization and Fusion in Multibiometric Systems, in: International Conference on Distributed Multimedia Systems, DMS2007, 2007, pp.87-92

Summary of results with monodimensional functions

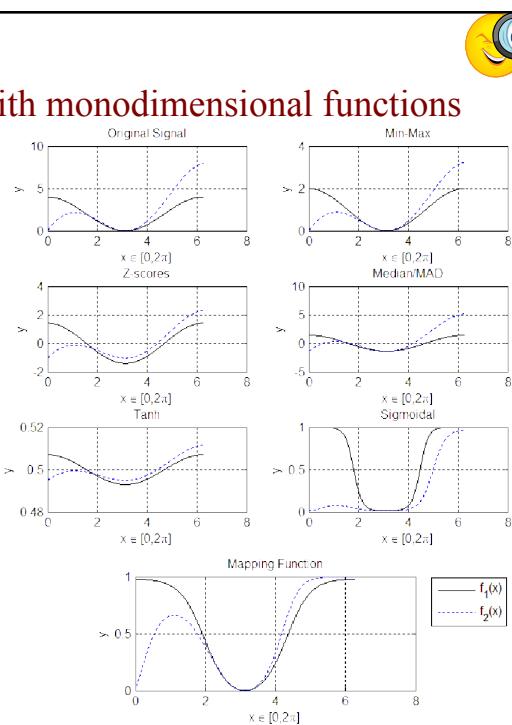
- Normalization techniques:

- Min-Max
- Z-score
- Median/MAD
- Tanh Estimator
- Sigmoidale
- QL-Sigmoidale

- Test functions

$$f_1(x) = 2 \cdot \cos(x) + 1 \quad f_2(x) = 2 \cdot \log(x) \cdot (\cos(x) + 1)$$

- The first three do not assure a mapping of original value onto the common interval $[0,1]$
- Tanh and Sigmoid in $(0,1)$ with too central values for Tanh and distortion near 0 for Sigmoid.
- QL-Sigmoidal assures a common interval $[0,1]$ and preserves the original data distribution.



Experiments with biometric data

The used databases were:

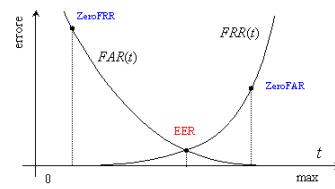
- Face*: FERET e AR-Faces (first 100 subjects).



- Ear*: Notre-Dame (first 100 subjects).



Performances were measured in terms of Recognition Rate and Equal Error Rate (EER).





Performance of biometric systems for different normalization functions with correct x_{max} estimation



System		Performances				
		min max	z scores	Median mad	sigmoid	QLS
Face	RR	93%	93%	93%	93%	93%
	EER	0.03	0.23	0.12	0.04	0.03
Ear	RR	72%	72%	72%	72%	72%
	EER	0.14	0.25	0.17	0.16	0.14
<i>Face</i> ⊕ <i>Ear</i>	RR	95%	93%	93%	94%	98%
	EER	0.018	0.23	0.11	0.02	0.015

Min-Max vs QLS with a wrong estimation of the maximum face score



System		Overestimated Maximum Score		Underestimated Maximum Score	
		Min/max	QLS	Min/max	QLS
Face	RR	93%	93%	38%	93%
	EER	0.04	0.04	0.81	0.034
Ear	RR	72%	72%	72%	72%
	EER	0.14	0.14	0.14	0.14
Face ⊕ Ear	RR	78%	78%	81%	97%
	EER	0.08	0.08	0.10	0.058



Min-Max vs QLS

with a wrong estimation of the maximum face score

System		Overestimated Maximum Score		Underestimated Maximum Score	
		Min/ max	QLS	Min/ max	QLS
Face	RR	93%	93%	38%	93%
	EER	0.04	0.04	0.81	0.034
Ear	RR	72%	72%	72%	72%
	EER	0.14	0.14	0.14	0.14
Face ⊕ Ear	RR	78%	78%	81%	97%
	EER	0.08	0.08	0.10	0.058



Min-Max vs QLS

with a wrong estimation of the maximum face score

System		Overestimated Maximum Score		Underestimated Maximum Score	
		Min/ max	QLS	Min/ max	QLS
Face	RR	93%	93%	38%	93%
	EER	0.04	0.04	0.81	0.034
Ear	RR	72%	72%	72%	72%
	EER	0.14	0.14	0.14	0.14
Face ⊕ Ear	RR	78%	78%	81%	97%
	EER	0.08	0.08	0.10	0.058

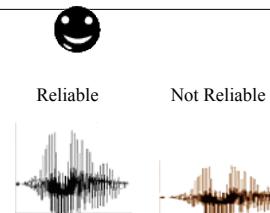
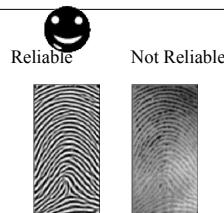
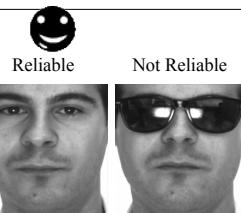
Multibiometric systems System Response Reliability



The reliability of identification systems

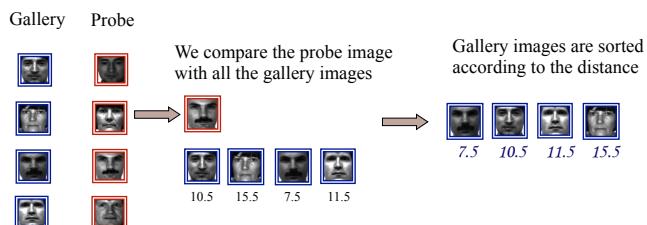


- Due to the possibly different quality of data inputted to each subsystem, and to the possibly different accuracy of exploited recognition procedures, it could happen that not all responses are equally reliable.
- The definition of a measure for the response reliability of the single subsystems would be significant for fusing the single results in an overall final response.



The Identification Process

- Let A be an identification system and G its gallery of genuine subjects who were correctly enrolled.
 - Assume there are at least $n > 0$ acquisitions for each.
 - Let p be a person to be identified.



SRR: System Response Reliability

Confusion Measures

There is a major difference between a quality measure for an input sample and a reliability measure for the response of a biometric system.

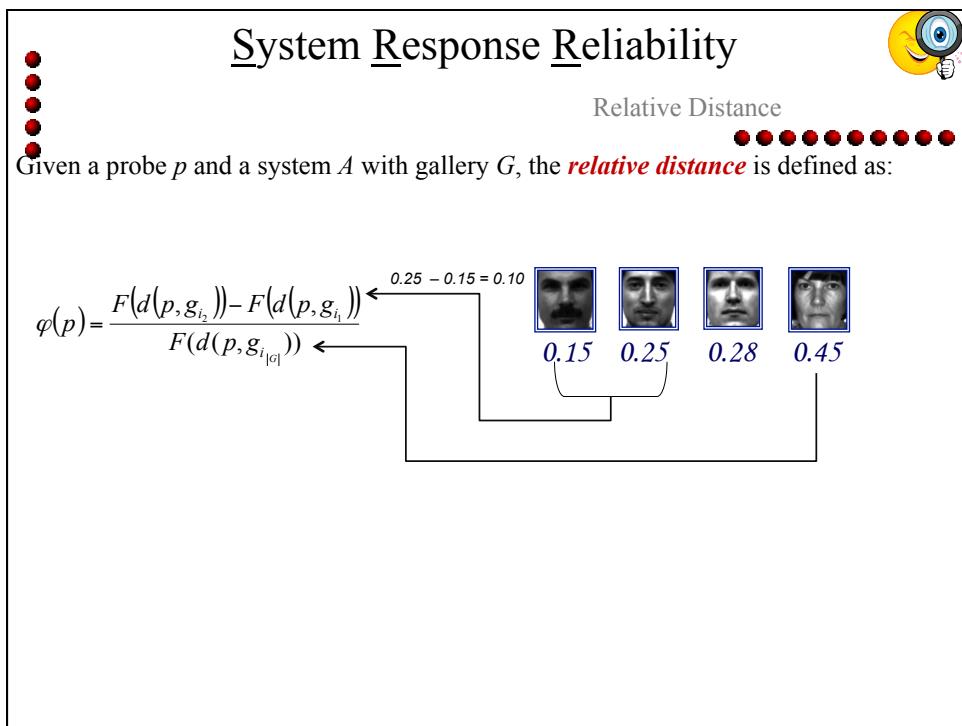
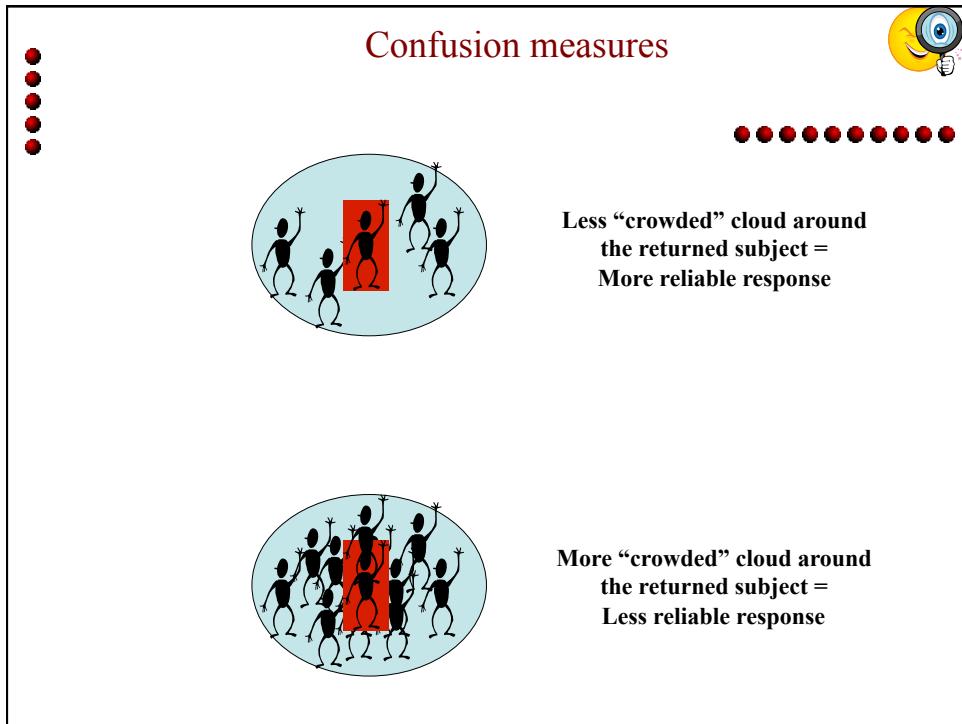
System Response Reliability ($srr \in [0, 1]$) index measures the ability of an identification system to separate genuine subjects from impostors on a single probe basis.

The SRR relies on different versions of function φ . We defined and tested three different φ functions:

- Relative distance;
 - Density ratio;
 - A combination of previous ones.

All three functions measure the amount of “confusion” among possible candidates.

We assume that the result of an identification operation is the whole gallery ordered by distance from the probe.



System Response Reliability



Density Ratio

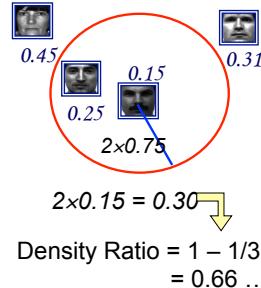


Given a probe p and a system A with gallery G , the **density ratio** is defined as:

$$\varphi(p) = 1 - |N_b| / |G|$$

With

$$N_b = \{g_{i_k} \in G \mid F(d(p, g_{i_k})) < 2 \cdot F(d(p, g_1))\}$$



This function is less sensible to outliers, and in fact usually performs better than φ_1 .

However, its definition takes to consider narrower clouds when the first retrieved identity is closer to the probe. On the contrary, a large distance takes to a larger cloud, which can be expected to be more crowded in any case.

System Response Reliability



Adaptive Density Ratio

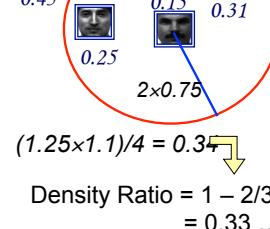


To further improve the behavior of φ , we define here a new version N_c of the term used to compute φ such that the cloud radius depends on the difference between the first two distances:

$$\varphi(p) = 1 - |N_c| / |G|$$

With

$$N_c = \left\{ g_{i_k} \in G \mid d(p, g_{i_k}) < \frac{(1 + d(p, g_{i_2}))(1 + d(p, g_{i_2}) - d(p, g_{i_1}))}{4} \right\}$$



The farther the second returned subject from the probe, also with respect to the first one, the wider the cloud we inspect.

We also use the appropriate normalization factor since the value of d is in $[0,1]$, and the maximum value for the numerator is 4.

System Response Reliability

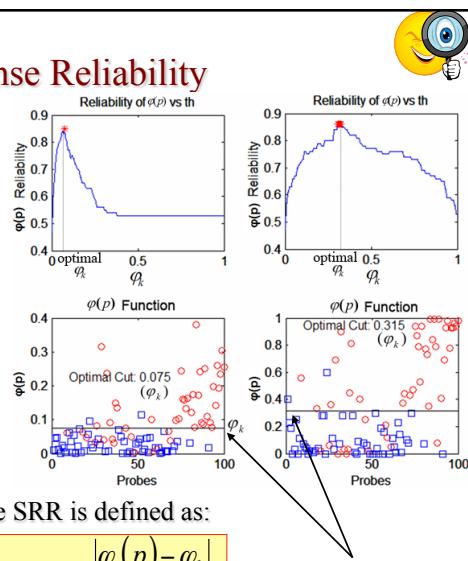
We need to establish a value φ_k for the reliability index separating genuine subjects from impostor ones

The optimal φ_k is given by that value able to minimize the wrong estimates of function $\varphi(p)$, i.e. impostors with $\varphi(p)$ higher than φ_k or genuine subjects with $\varphi(p)$ lower than φ_k

SRR gets high values both for $\varphi(p)$ much higher than φ_k (genuine subjects) and $\varphi(p)$ much lower than φ_k (impostors).

The distance between $\varphi(p)$ and φ_k is significant for reliability, but also the maximum achievable

$$S(\varphi(p), \varphi_k) = \begin{cases} 1 - \varphi_k & \text{if } \varphi(p) > \varphi_k \\ \varphi_k & \text{otherwise} \end{cases}$$



The SRR is defined as:

$$SRR = \frac{|\varphi(p) - \varphi_k|}{S(\varphi(p), \varphi_k)}$$

How to integrate SRR index into the fusion protocol

- Let us assume to have a system S composed by N subsystems T_1, \dots, T_N , each able to produce a sorted list $T_i(1, \dots, |G|)$ of $|G|$ subjects and a SRR value srr_i
- In order to guarantee a consistent fusion we define
 $w_i = \frac{srr_i}{\sum_{j=1}^N srr_j}, \forall i$ to assure $\sum_i w_i = 1$
- A consistent threshold th is estimated for each subsystem T_i above which we can consider its reliability satisfactory enough.

Threshold setup

- Thresholds th_i for each subsystem are automatically estimated according to a certain number M of subsequent observations.

$$\bar{S}_i = \{srr_i^1, \dots, srr_i^M\}$$

- The desirable characteristic for a certain T_i subsystem is that its vector has a high mean value (the system is generally reliable) and a low value for the variance (basically stable system).

- We can summarize this in the formula :

$$th_i = \left| \frac{E[\bar{S}_i]^2 - \sigma[\bar{S}_i]}{E[\bar{S}_i]} \right|$$

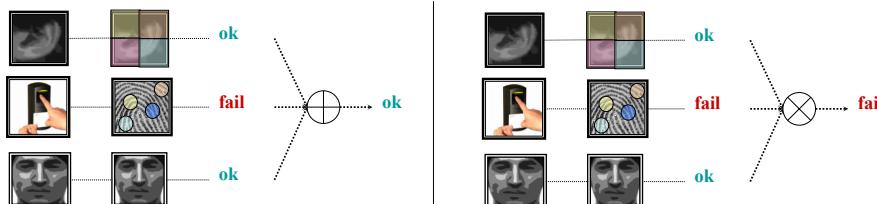
A. F. Abate, M. Nappi, D. Riccio, M. De Marsico, "Data Normalization and Fusion in Multibiometric Systems", Proceedings of The Thirteenth International Conference on Distributed Multimedia Systems DMS 2007, September 6-8 2007, San Francisco, USA, pp. 87-92

How to integrate SRR index into the fusion protocol

- The main integration policies are:

OR

AND



We apply a decision fusion techniques to the set of reliability indexes associated to returned responses, before applying a further fusion technique to the actual responses.

Rule	DESCRIPTION
Or	the combined response is valid only if at least one subsystem response reliability is above the corresponding threshold; the system returns the first identity from the list of the subsystem with the higher reliability above the corresponding threshold
And	the combined response is valid only if all subsystem response reliabilities are above the corresponding thresholds; the system returns the identity with the minimum weighted sum of distances from the probe, where weights are the reliability degrees of the different subsystems



Performances of different fusion rules

••••••••••

Database		Statistics				
		None	SRR I		SRR II	
			SIMPLE	OR	AND	OR
Feret Fafb	RR	98%	99%	100%	96%	100%
	EER	0.028	0.016	0.003	0.015	0.000
	NR	100	75	63	94	38
Feret Fafc	RR	55%	76%	100%	84%	-
	EER	0.167	0.153	0.002	0.117	-
	NR	100	85	2	74	0
Feret Dup I	RR	75%	81%	100%	87%	100%
	EER	0.238	0.228	0.001	0.177	0.000
	NR	100	91	18	84	22



Performances of different fusion rules

••••••••••

Database		Statistics				
		None	SRR I		SRR II	
			SIMPLE	OR	AND	OR
Feret Fafb	RR	98%	99%	100%	96%	100%
	EER	0.028	0.016	0.003	0.015	0.000
	NR	100	75	63	94	38
Feret Fafc	RR	55%	76%	100%	84%	-
	EER	0.167	0.153	0.002	0.117	-
	NR	100	85	2	74	0
Feret Dup I	RR	75%	81%	100%	87%	100%
	EER	0.238	0.228	0.001	0.177	0.000
	NR	100	91	18	84	22

Performances of SRR I and SRR II



Face distortion		Performance					
		Face	Ear	Face ⊕ Ear			
					SRR I	SRR II	
Left light	RR	93%	72%	RR	100%	100%	
	EER	0.09	0.12	EER	0.001	0.008	
				NRR	37	70	
Sad	RR	100%	72%	RR	100%	100%	
	EER	0.07	0.12	EER	0.005	0.002	
				NRR	86	43	
Scarf	RR	80%	72%	RR	100%	100%	
	EER	0.17	0.12	EER	0.015	0.020	
				NRR	70	70	
Scream	RR	47%	72%	RR	100%	100%	
	EER	0.18	0.12	EER	0.001	0.020	
				NRR	23	46	
Glasses	RR	90%	72%	RR	100%	100%	
	EER	0.14	0.12	EER	0.016	0.010	
				NRR	87	70	

Performances of SRR I and SRR II



Face distortion		Performance					
		Face	Ear	Face ⊕ Ear			
					SRR I	SRR II	
Left light	RR	93%	72%	RR	100%	100%	
	EER	0.09	0.12	EER	0.001	0.008	
				NRR	37	70	
Sad	RR	100%	72%	RR	100%	100%	
	EER	0.07	0.12	EER	0.005	0.002	
				NRR	86	43	
Scarf	RR	80%	72%	RR	100%	100%	
	EER	0.17	0.12	EER	0.015	0.020	
				NRR	70	70	
Scream	RR	47%	72%	RR	100%	100%	
	EER	0.18	0.12	EER	0.001	0.020	
				NRR	23	46	
Glasses	RR	90%	72%	RR	100%	100%	
	EER	0.14	0.12	EER	0.016	0.010	
				NRR	87	70	

Performances of SRR I and SRR II

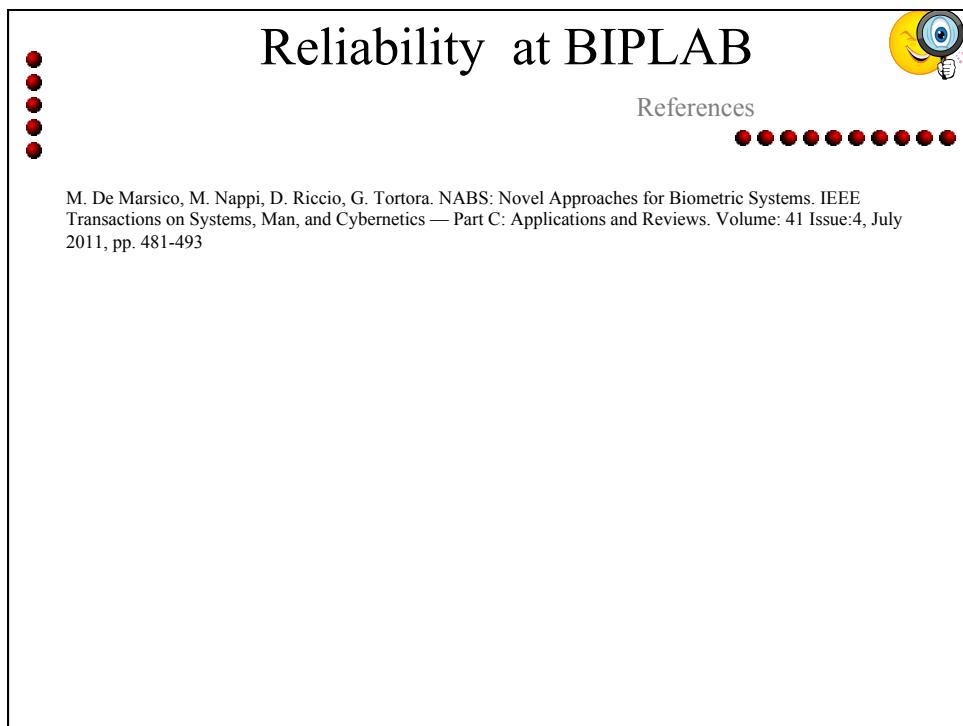
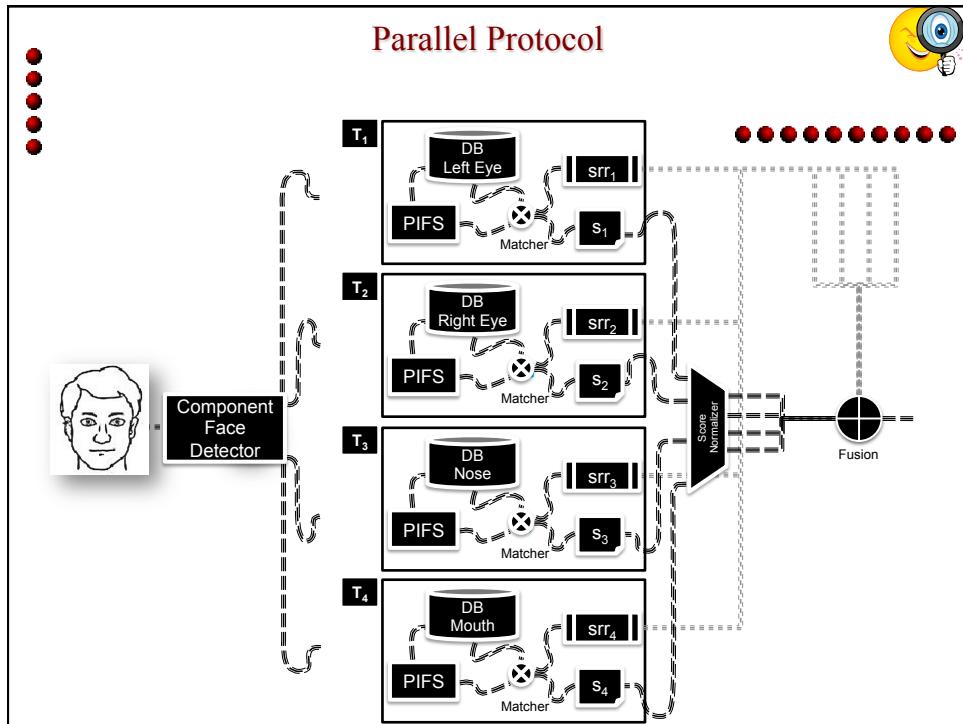


Face distortion		Performance					
		Face		Ear		Face ⊕ Ear	
							SRR I
Left light	RR	93%	72%	RR		100%	100%
	EER	0.09	0.12	EER		0.001	0.008
				NRR		37	70
Sad	RR	100%	72%	RR		100%	100%
	EER	0.07	0.12	EER		0.005	0.002
				NRR		86	43
Scarf	RR	80%	72%	RR		100%	100%
	EER	0.17	0.12	EER		0.015	0.020
				NRR		70	70
Scream	RR	47%	72%	RR		100%	100%
	EER	0.18	0.12	EER		0.001	0.020
				NRR		23	46
Glasses	RR	90%	72%	RR		100%	100%
	EER	0.14	0.12	EER		0.016	0.010
				NRR		87	70

A novel approach



- We pushed the multibiometric approach to divide the face into distinct components
- Each component is processed by a separate classifier module
- Modules are embedded in a multicomponent architecture
- Reliability measures and self-tuning policies enhance the simple result fusion



Presentation Outline

- Biometric Systems
 - Short introduction
 - Face recognition
 - Ear Recognition
 - Iris recognition
- Multibiometric Systems
 - Short introduction
 - Data Normalization
 - System Response Reliability
 - **Supervised Fusion**
 - Cross Testing Protocol
- Introduction to Ambient Intelligence
 - Definitions and trends
 - Interacting with an intelligent ambient
- Conclusions

The Supervisor

Case I: an identity got more votes

T ₁	T ₂	T ₃	T ₄
----------------	----------------	----------------	----------------

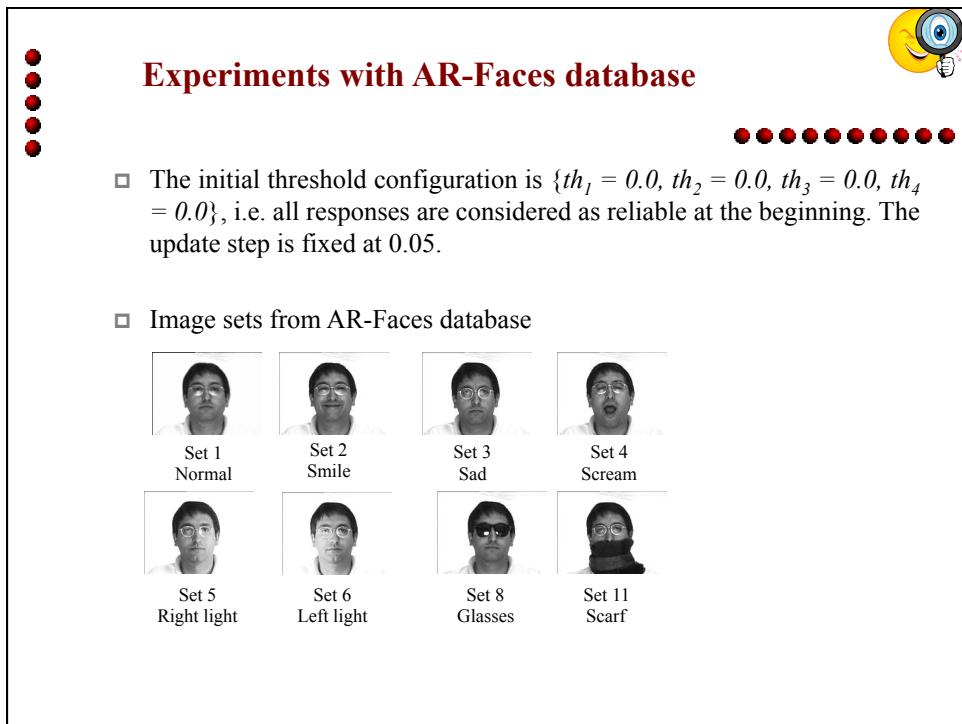
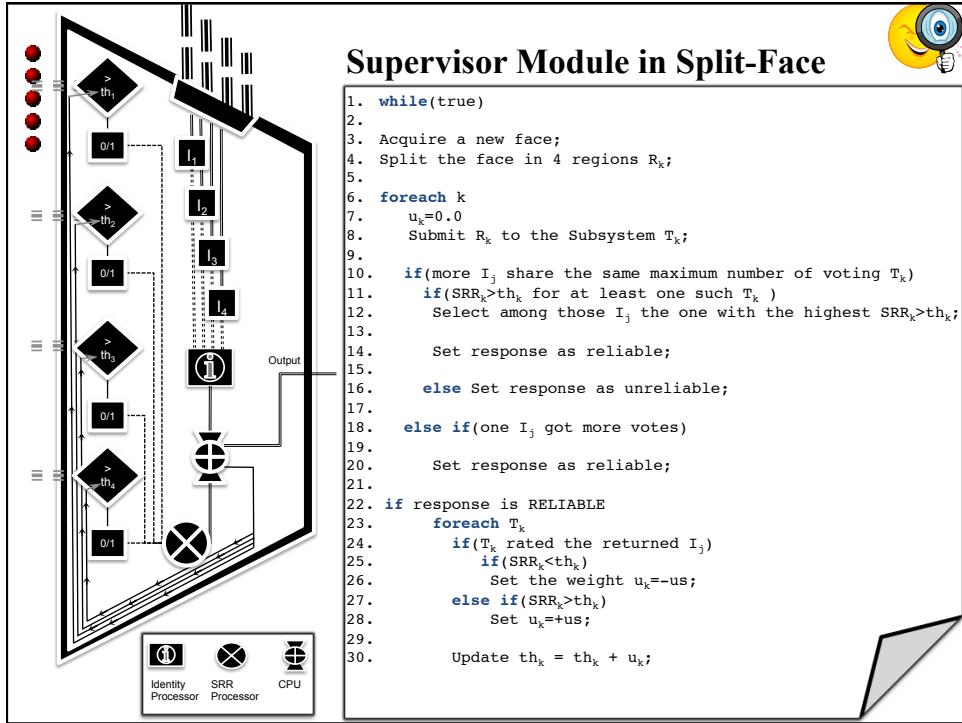
If $srr_k < th_k \Rightarrow$ decrease th_k , $k=\{1,2,3\}$
 If $srr_k > th_k \Rightarrow$ increase th_k , $k=\{4\}$

Case II: more identities share the maximum number of votes

T ₁	T ₂	T ₃	T ₄
T ₁	T ₂	T ₃	T ₄

$\exists k \ni srr_k > th_k$ with $k=\{1,2,\dots\}$
 $k_{max} = argmax \{ srr_k \mid srr_k > th_k \}$
 Suppose $k_{max} = 2$
 For $k=\{2,4\}$ If $srr_k < th_k \Rightarrow$ decrease th_k
 For $k=\{1,3\}$ If $srr_k > th_k \Rightarrow$ increase th_k ,
 else
 the response is unreliable

M. De Marsico, M. Nappi, D. Riccio, G. Tortora. A multiexpert collaborative biometric system for people identification. Journal of Visual Languages & Computing, Volume 20, Issue 2, April 2009, Pages 91-100

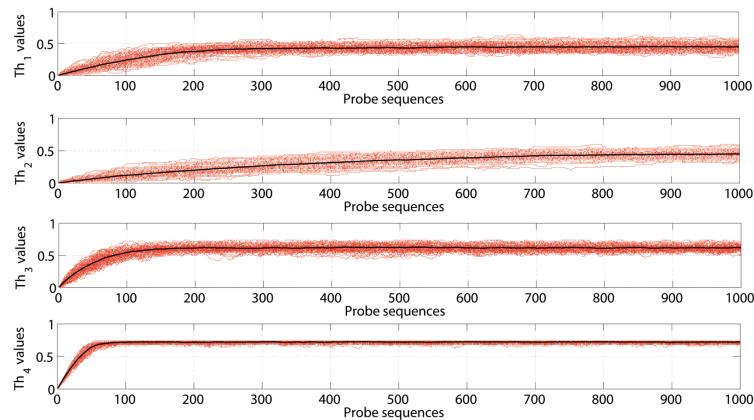


Question - 1

- Do the thresholds converge ?

- For this experiment, set 1 is used as gallery, while 100 probe sequences are extracted from set 2, 6 and 11 (smile, left light, scarf).
- Each probe sequence is built by randomly extracting 1000 times one of the 126 images from the probe set.

Answer - 1



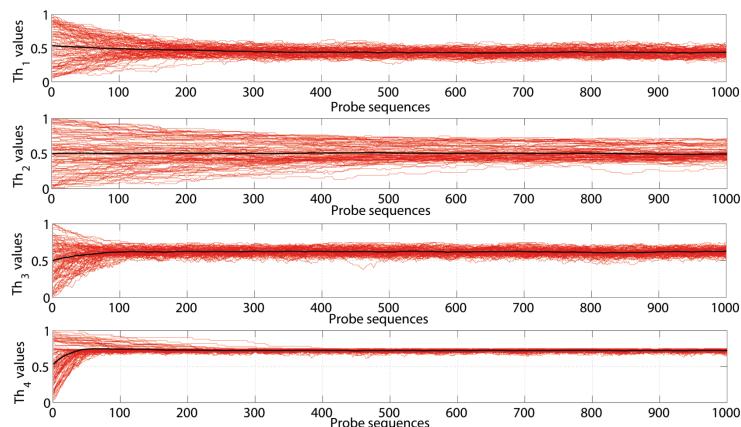
- Thresholds th_1 and th_2 (right and left eye) tend to assume lower values than th_3 and th_4 (nose and mouth). The latter values show an initial variation, and then stay constant for all the remaining part of the probe sequence. Notice the higher values for the right eye, which in set 6 is poorly lit.
- This can be explained by observing that, since images in set 2 belong to smiling subjects, nose and mouth show a higher variability than eyes, making the corresponding systems T_3 e T_4 less reliable, and therefore demanding higher values for the respective thresholds.
- The darker line (in black) is the mean value of the 100 computed curves and represents the mean trend for thresholds variation.

Question - 2

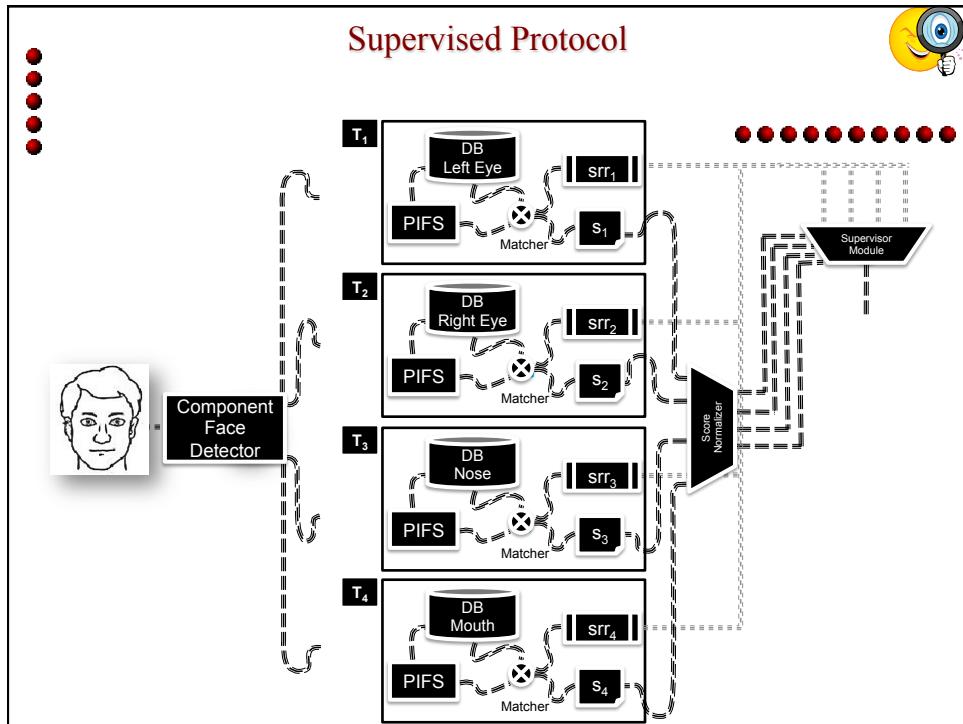
- Does the initial setting of thresholds influence the system behaviour?

- Even in this case, we considered 100 probe sequences of 1000 images randomly extracted among the 126 of set 2.
- For each system run, the initial values for thresholds are randomly chosen (all values are equally probable) in the interval $[0, 1]$

Answer - 2



- Results on set 2 for different initial thresholds show that curves generated by the different probe sequences tend to always concentrate in a relatively small final interval. This confirms the convergence of the updating procedure.



Experimental Results on AR-Faces (Face Database)

- In most cases, PP offers worse performances than PCBPs, which is in general robust to occlusions and local distortions. Such result can be ascribed to the fact that single subsystems do not have any information about all the others.

Subset	Expression Variations							
	PCBP	PP	SP					
			PERF.	th ₁	th ₂	th ₃	th ₄	
SET 2 SMILE	RR EER NRR	0.92 0.07 126	0.89 0.05 38	0.94 0.03 120	0.15	0.30	0.40	0.70
SET 3 ANGRY	RR EER NRR	0.95 0.05 126	0.98 0.03 56	0.94 0.03 125	0.20	0.25	0.40	0.50
SET 4 SCREAM	RR EER NRR	0.48 0.15 126	0.36 0.29 33	0.76 0.12 50	0.05 0.00 0.65	0.00	0.65	0.70

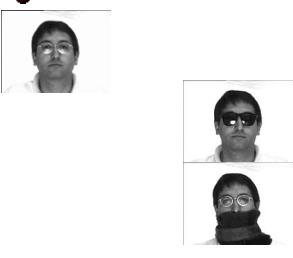
Experimental Results on AR-Faces (Face Database)




Subset	Illumination Variations					
	PCBP	PP	SP			
			PERF.	th ₁	th ₂	th ₃
SET 5	RR	0.92	1.00	0.96		
LEFT	EER	0.03	0.02	0.02	0.45	0.50
LIGHT	NRR	126	30	112		0.65
						0.60
SET 6	RR	0.94	0.97	0.96		
RIGHT	EER	0.05	0.07	0.03	0.00	0.75
LIGHT	NRR	126	37	107		0.75

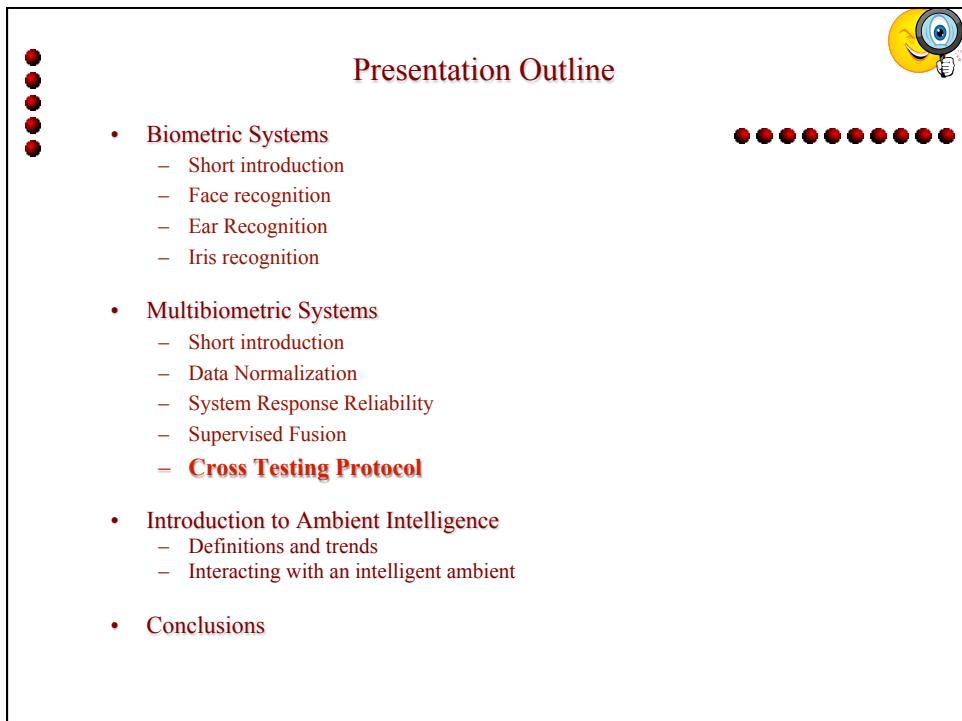
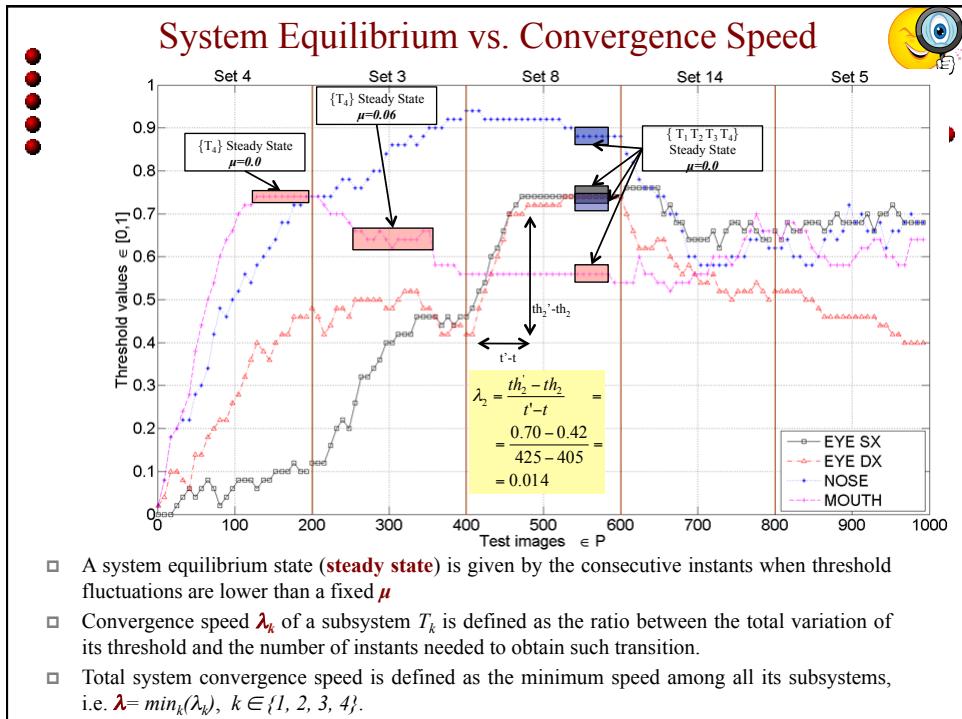
- As expected, PCBP performances are quite constantly worse than those obtained with SP. We can observe that, even when the accuracy of SP drops slightly below that of PP (sets 5 and 6), this is counterbalanced by a much higher number of reliable responses.

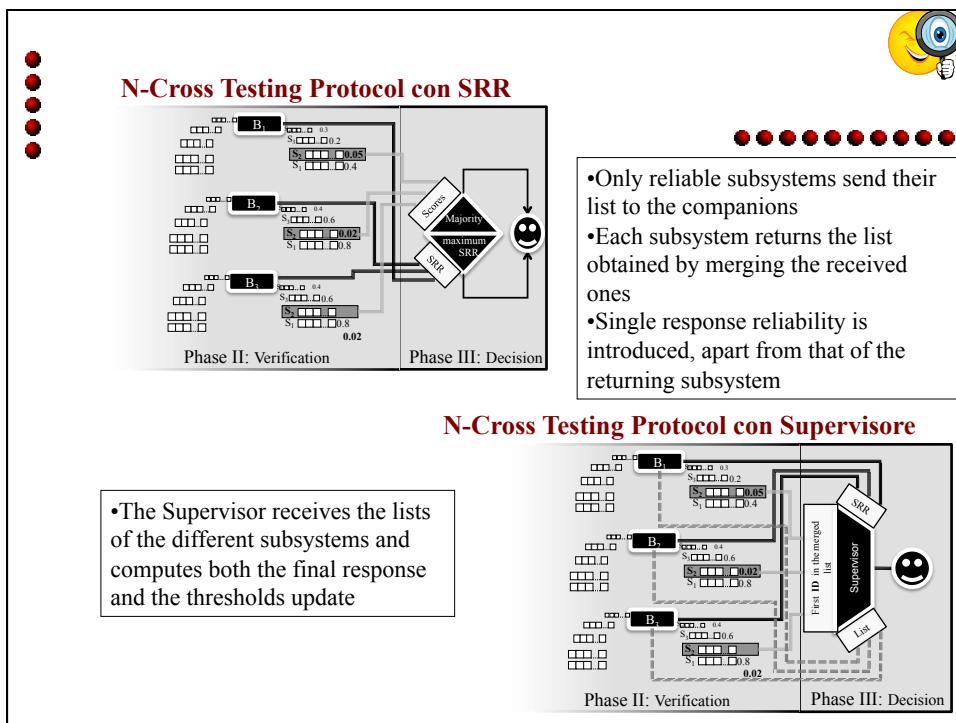
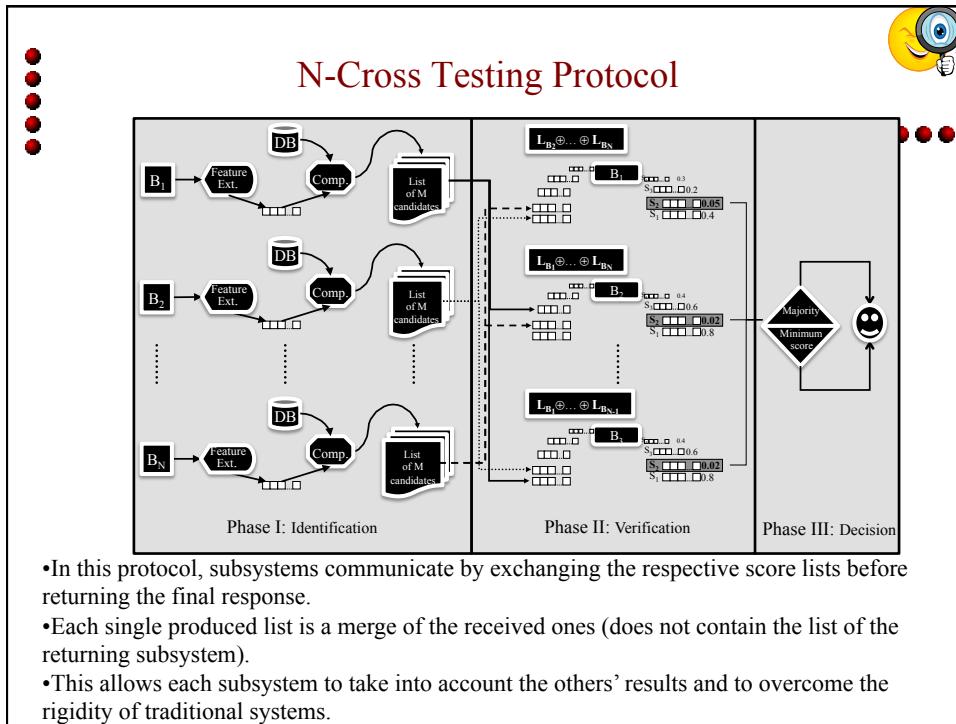
Experimental Results on AR-Faces (Face Database)

Subsets	OCCLUSIONS					
	PCBP	PP	SP			
			PERF.	th ₁	th ₂	th ₃
SET 8	RR	0.71	0.25	0.98		
SUN	EER	0.09	0.23	0.04	0.65	0.60
GLASSE S	NRR	126	20	50		0.60
						0.00
SET 11	RR	0.85	0.61	0.92		
SCARF	EER	0.09	0.19	0.02	0.35	0.45
	NRR	126	23	115		0.75
						0.75

- The sets of equilibrium thresholds reached by the system perfectly agree with the variations introduced by the different sets of face images.
- The number of reliable responses for SP drops to 50 for sun glasses (set 8) and to 115 for scarf (set 11). This agrees with our expectations, as the distortions introduced involve a larger face area.
- However, out of a lower number of reliable responses, the system is able in both cases to guarantee a significantly higher accuracy than PCBP (RR of 0.98 versus 0.71 and of 0.92 versus 0.85) and lower EER







N-Cross Testing Protocol - Results




ARCHITECTURE									
DATA SETS	SIMPLE N-CROSS-TESTING			RELIABLE N-CROSS-TESTING			SUPERVISED N-CROSS-TESTING		
	RR	EER	NRR	RR	EER	NRR	RR	EER	NRR
SET 2	0.962	0.018	126	0.989	0.005	115	0.990	0.004	121
SET 3	0.971	0.014	126	0.987	0.006	96	0.989	0.005	116
SET 4	0.652	0.17	126	0.933	0.033	35	0.962	0.018	94
SET 5	0.744	0.127	126	0.925	0.037	95	0.940	0.029	118
SET 6	0.584	0.207	126	0.825	0.087	94	0.905	0.047	112
SET 8	0.522	0.238	126	0.839	0.080	65	0.849	0.075	102
SET 11	0.359	0.320	126	0.975	0.023	61	0.975	0.012	94

M. De Marsico, M. Nappi, D. Riccio, G. Tortora. A multiexpert collaborative biometric system for people identification. Journal of Visual Languages & Computing, Volume 20, Issue 2, April 2009, Pages 91-100



Multimodal protocols at BIPLAB




References

M. De Marsico, M. Nappi, D. Riccio, G. Tortora. "A Multiexpert Collaborative Biometric System for People Identification". Journal of Visual Languages & Computing, Volume 20, Issue 2, April 2009, Pages 91-100

A. F. Abate, M. De Marsico, D. Riccio and G. Tortora. MUBAI: multiagent biometrics for ambient intelligence. Journal of Ambient Intelligence and Humanized Computing Vol. 2, 2011, pp. 81-89

M. De Marsico, M. Nappi, D. Riccio, CABALA—Collaborative architectures based on biometric adaptable layers and activities. Pattern Recognition, Volume 45, Issue 6, June 2012, Pages 2348–2362.

C11. A. F. Abate, M. Nappi, D. Riccio, M. De Marsico, "Data Normalization and Fusion in Multibiometric Systems"; Proceedings of The Thirteenth International Conference on Distributed Multimedia Systems DMS 2007, September 6-8 2007, San Francisco, USA, pp. 87-92

A. F. Abate, M. Nappi, D. Riccio, M. De Marsico, "Face, Ear and Fingerprint: Designing Multibiometric Architectures", Proceedings of the 14th International Conference on Image Analysis and Processing ICIAP 2007, 10-14 September 2007, Modena, Italy, pp. 437-442

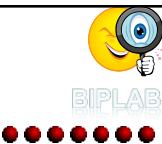
M. De Marsico, D. Riccio, "A New Data Normalization Function for Multibiometric Contexts: a Case Study", Proceedings of International Conference on Image Analysis and Recognition - ICIAR 08, June 25-27, 2008 Povoa De Varzim, Portugal, LNCS 5112, pp. 1033-1040

C16. M. De Marsico, M. Nappi, D. Riccio. A Self-Tuning People Identification System from Split Face Components. Proceedings of The 3rd Pacific-Rim Symposium on Image and Video Technology, PSIVT2009, January 13th–16th, 2009, Tokyo, Japan, LNCS 5414 pp. 1-12

M. De Marsico, M. Nappi, D. Riccio. Multibiometric People Identification: A Self-Tuning Architecture. Proceedings of the Third International Conference on Biometrics, ICB 2009, Alghero, Italy, June 2-5, 2009, LNCS 5558, pp. 980-989

A. F. Abate, M. De Marsico, M. Nappi, and D. Riccio. A Self-updating Multiexpert System for Face Identification. In: P. Foggia, C. Sansone, and M. Vento (Eds.); Proceedings of the 15-th International Conference on Image Analysis and Processing, ICIAP 2009, Vietri sul Mare (SA), Italy, September 8-11 2009, LNCS 5716, pp. 346–354, 2009.

Multimodal protocols at BIPLAB



References

M. De Marsico, M. Nappi, D. Riccio, and G. Tortora. Automatic Template Labeling in Extensible Multiagent Biometric Systems. In: G. Maino and G.L. Foresti (Eds.): Proceedings of International Conference on Image Analysis and Processing - ICIAP 2011, Part II, LNCS 6979, pp. 313–322, 2011, Ravenna, Italy, 14-16 Settembre 2011.

M. De Marsico, M. Nappi, D. Riccio. Fusion of multi-biometric recognition results by representing score and reliability as a complex number. Accepted for 18th Iberoamerican Congress on Pattern Recognition (CIARP 2013) to be held in Havana CUBA, November 20-23 Nov 2013.

M. De Marsico, R. Distasi, M. Nappi, and D. Riccio. Fractal Indexing in Multimodal Biometric Contexts. In: L. Kocarev, Z. Galias, and S. Lian (Eds.): Intelligent Computing Based on Chaos, Studies in Computational Intelligence - SCI 184, pp. 93–120. Springer-Verlag Berlin Heidelberg 2009.

M. De Marsico, R. Distasi, M. Nappi, D. Riccio. Multiple Traits for People Identification: Face, Ear and Fingerprints. In: H. T. Sencar, and L. Kocarev, Z. Galias and Shiguo Lian (Eds.): Intelligent Multimedia Analysis for Security Applications, Studies in Computational Intelligence - SCI 282, pp.79-95. Springer-Verlag Berlin Heidelberg. 2010