



DEEPER FLOW: DETECTION, POSE ESTIMATION & RECOGNITION BASED ON FACE DYNAMICS AND LIPS MOVEMENT

SILVIO BARRA

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCES, UNIVERSITY OF CAGLIARI

SILVIO.BARRA@UNICA.IT

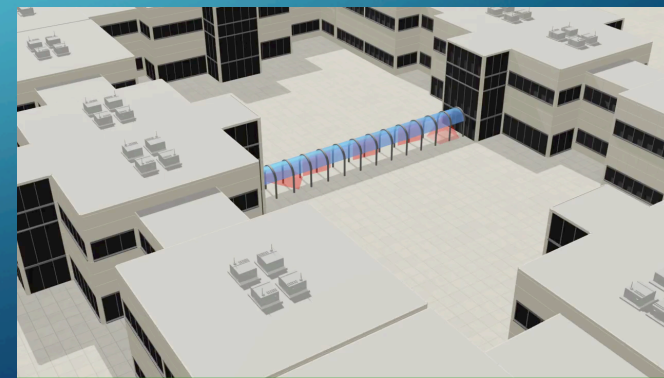
ANDREA ABATE, MICHELE NAPPI

DEPARTMENT OF COMPUTER SCIENCES, UNIVERSITY OF SALERNO

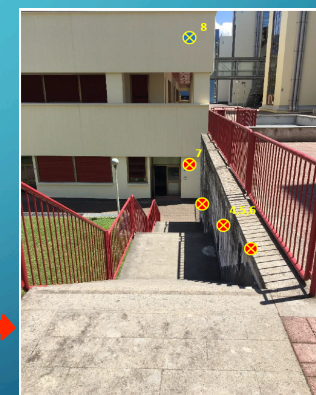
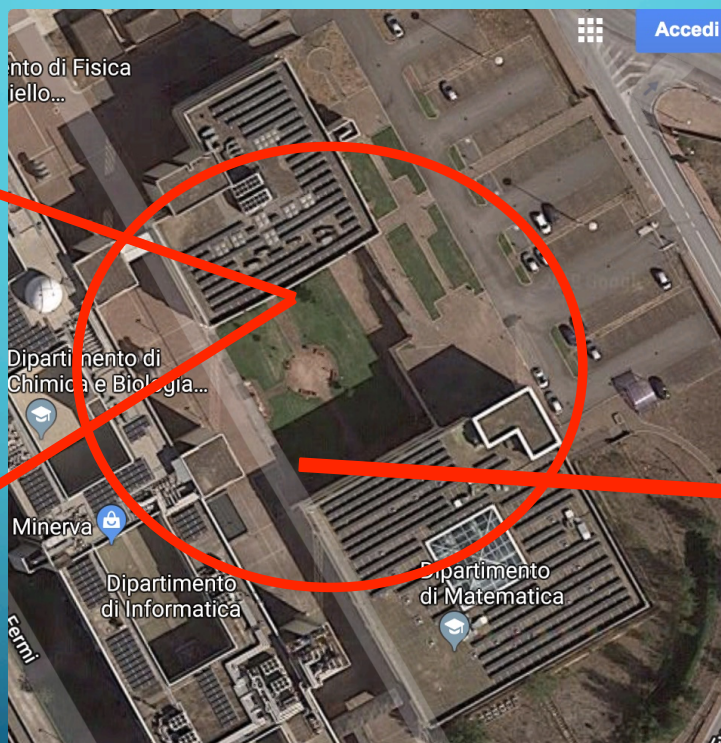
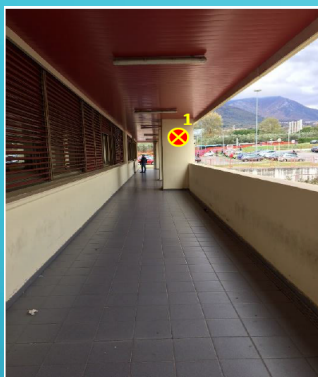
{ABATE,MNAPPI}@UNISA.IT

COSMOS: OBIETTIVI

- **Progettazione, Creazione e Testing di un sistema multibiometrico/multimodale contactless per verifica/riconoscimento**
- **Massivo utilizzo di dispositivi di acquisizione *mobile* (GoPro, Cellulari, Droni etc.)**
- **Creazione di un database multibiometrico in contesti non collaborativi (in the wild)**

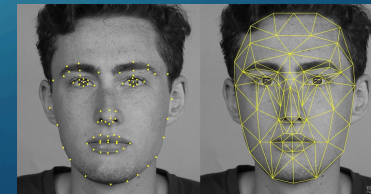
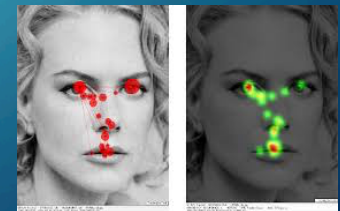
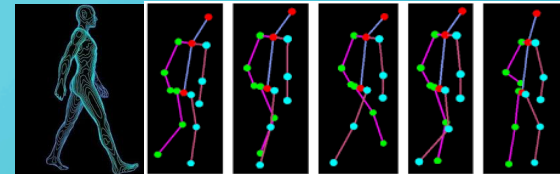


COSMOS: IL CORRIDOIO BIOMETRICO



COSMOS

- Andatura (Gait)
- Contactless FingerPrint
- Iris & Periocular
- Hand
- Ear
- Gaze
- Face 2D/3D
-

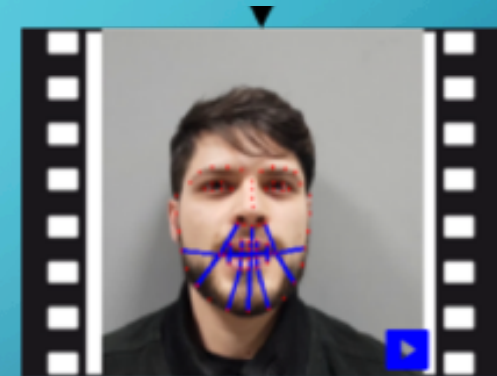
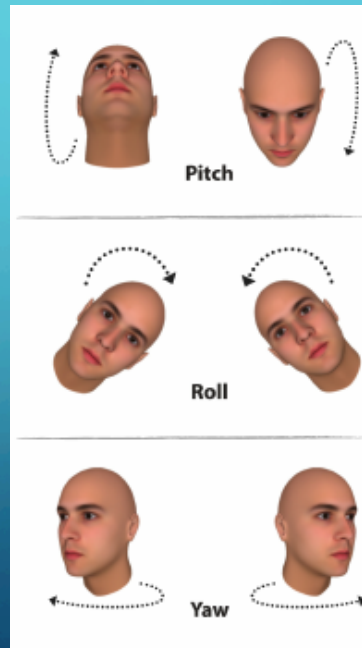


DEEPER FLOW: THE PROJECT



- Automatic guidance
- Activation by movement detection
- Biometric detection

- Frontal image selection
- Head Pose Estimation
 - 2D images



- Face Dynamics from the frontal pose
- Face Landmark distances computation and tracking

DRONE BIOMETRIC DETECTION

- Flussi di persone in percorsi predefiniti (contesti semicontrollati)
- Aeroporti, Stazioni,....
- Hotspot (flussi di migranti)



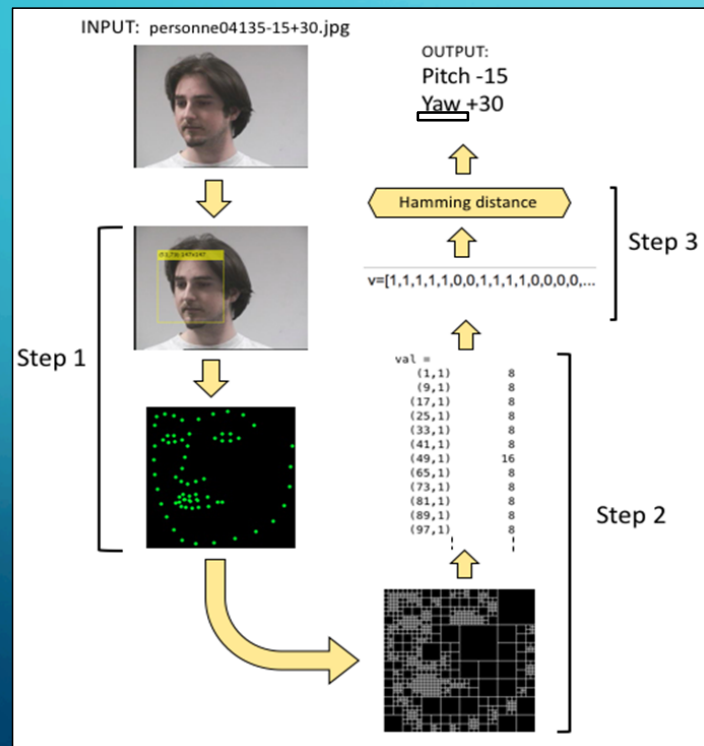
DRONE FUTURE WORKS

NEW

- Trajectory drawing
 - Minimum covering path
 - 3 axis gimbal
 - Google Maps
- Stealth mode
 - Gaze analysis at-a-distance
 - Head pose estimation enhancement



HEAD POSE ESTIMATION



- **Step 1, Landmarks detection:** by means of Viola-Jones algorithm to detect face and then applying the method described in CVPR 2014 [V. Kaemi, J. Sullivan: One Millisecond Face Alignment with an Ensemble of Regression Trees.].
- **Step 2, Quad-tree decomposition:** to obtain the sparse matrix which represents the tree, quad-tree decomposition is applied with minimum block size 4, and the median as the function.
- **Step 3, Pose estimation:** sparse matrix of the previous step is transformed in a tree-array and compared to 35 model trees.

HEAD POSE ESTIMATION

For each of the faces in the database pitch and yaw rotations were correctly estimated with a maximum angular error within 15° , that represents the discretization step adopted for the synthetic reference dataset used for comparison.

With regard to proposed method's efficiency, the overall time required for pose estimate since input image is provided amounts to approximately 0.22 seconds on average.

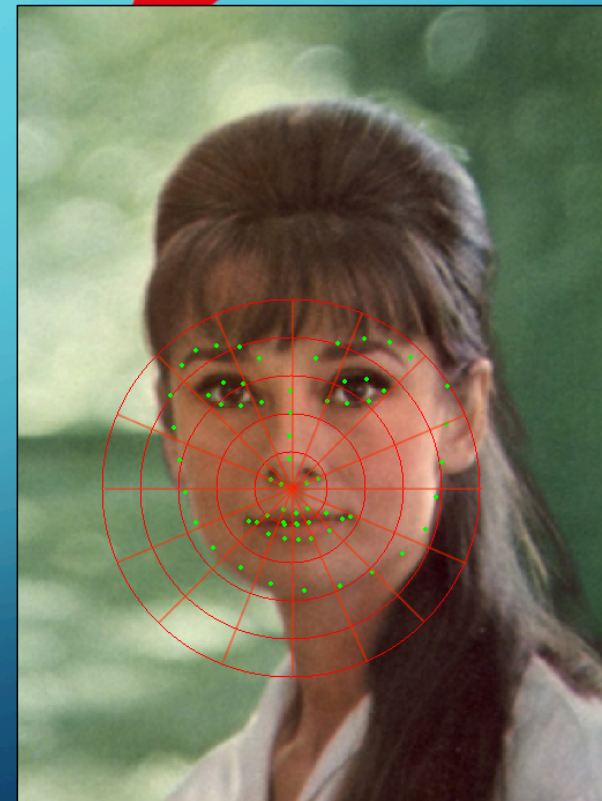


(30, -45)	(30, -30)	(30, -15)	(30, 0)	(30, 15)	(30, 30)	(30, 45)
(15, -45)	(15, -30)	(15, -15)	(15, 0)	(15, 15)	(15, 30)	(15, 45)
(0, -45)	(0, -30)	(0, -15)	(0, 0)	(0, 15)	(0, 30)	(0, 45)
(-15, -45)	(-15, -30)	(-15, -15)	(-15, 0)	(-15, 15)	(-15, 30)	(-15, 45)
(-30, -45)	(-30, -30)	(-30, -15)	(-30, 0)	(-30, 15)	(-30, 30)	(-30, 45)

HEAD POSE ESTIMATION

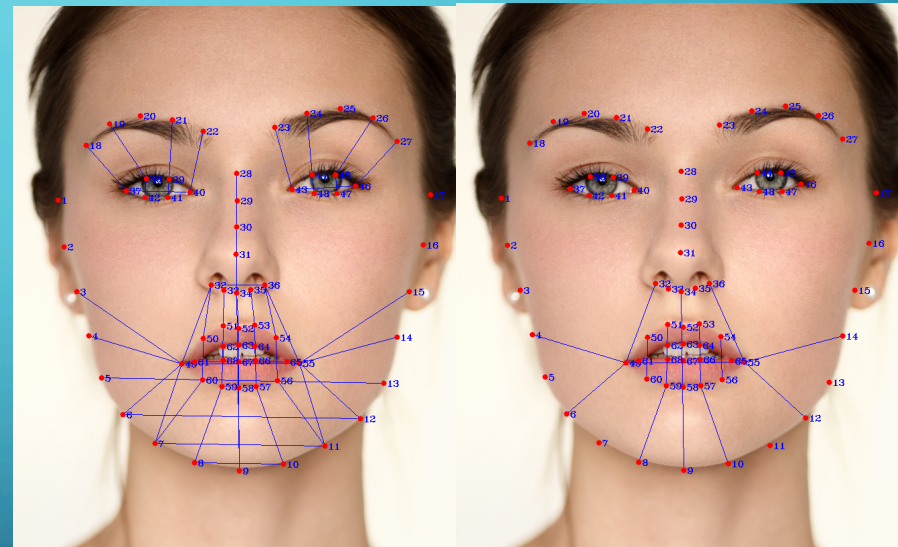
NEW

- No SLIC decomposition
- All the operations involve basilar computations
- Angular Error from 15° to 5°
- Computing time from 0.22 secs to 0.09 secs



FACE DYNAMICS

- Hello
- Excuse me
- I am sorry
- Thank you
- Goodbye
- See you
- Nice to meet you
- You are welcome
- How are you
- Have a good time

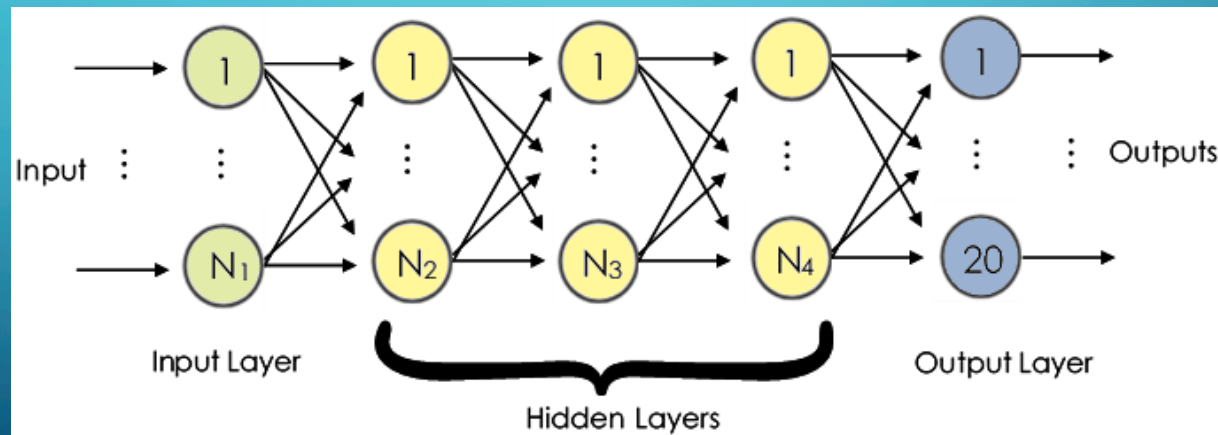


RESULTS

- A former approach exploited the Dynamic time Warping (DTW) suitable for computing distances between vectors built over time.
- The dataset has built ad-hoc and consists of 20 subject; each of them pronounces the sentence 9 times in 2 different sessions
- The results have shown an accuracy of over 85%

EXHANCED METHODOLOGY

Before the computation of the accuracy, each of the distance has been weighted according



RESULTS

Useful-Speech-Downsampled		Useful-Weighted-Downsampled		Useful-Challenge-Downsampled	
Distanze per frame	14	Distanze per frame	14	Distanze per frame	14
Dimensione campione	140	Dimensione campione	140	Dimensione campione	140
Dimensione train/test set	800/200	Dimensione train/test set	800/200	Dimensione train/test set	960/240
Nodi dei livelli interni	140 – 100 – 60	Nodi dei livelli interni	140 – 100 – 60	Nodi dei livelli interni	140 – 100 – 60
Tempo di train	27s	Tempo di train	29s	Tempo di train	38s
Gallery totale	1000	Gallery totale	1000	Gallery totale	1200
Percentuale di accuratezza	98.2	Percentuale di accuratezza	96.9	Percentuale di accuratezza	94.4



ESPERIMENTO - USEFUL-SPEECH-DOWNSAMPLED

Distanze per frame	14
Dimensione campione	140
Dimensione train/test set	800/200
Nodi dei livelli interni	140 – 100 – 60
Tempo di train	27s
Gallery totale	1000
Percentuale di accuratezza	98.2

ESPERIMENTO - USEFUL-WEIGHTED-DOWNSAMPLED

Distanze per frame	14
Dimensione campione	140
Dimensione train/test set	800/200
Nodi dei livelli interni	140 – 100 – 60
Tempo di train	29s
Gallery totale	1000
Percentuale di accuratezza	96.9

ESPERIMENTO - USEFUL-CHALLENGE-DOWNSAMPLED

Distanze per frame	14
Dimensione campione	140
Dimensione train/test set	960/240
Nodi dei livelli interni	140 – 100 – 60
Tempo di train	38s
Gallery totale	1200
Percentuale di accuratezza	94.4

ESPERIMENTO - USEFUL-SPEECH-UPSAMPLED

Distanze per frame	14
Dimensione campione	854
Dimensione train/test set	800/200
Nodi dei livelli interni	854 – 575 – 297
Tempo di train	3m 26s
Gallery totale	1000
Percentuale di accuratezza	96.4

ESPERIMENTO - USEFUL-CHALLENGE-UPSAMPLED

Distanze per frame	14
Dimensione campione	854
Dimensione train/test set	960/240
Nodi dei livelli interni	854 – 575 – 297
Tempo di train	3m 51s
Gallery totale	1200
Percentuale di accuratezza	97.3

ESPERIMENTO - ALL-SPEECH-DOWNSAMPLED

Distanze per frame	59
Dimensione campione	590
Dimensione train/test set	800/200
Nodi dei livelli interni	590 – 399 – 209
Tempo di train	2m 10s
Gallery totale	1000
Percentuale di accuratezza	95.1

ESPERIMENTO - ALL-CHALLENGE-DOWNSAMPLED

Distanze per frame	59
Dimensione campione	590
Dimensione train/test set	960/240
Nodi dei livelli interni	590 – 399 – 209
Tempo di train	2m 41s
Gallery totale	1200
Percentuale di accuratezza	94.1

ESPERIMENTO - ALL-SPEECH-UPSAMPLED

Distanze per frame	59
Dimensione campione	3599
Dimensione train/test set	800/200
Nodi dei livelli interni	3599 – 2404 – 1212
Tempo di train	41m 19s
Gallery totale	1000
Percentuale di accuratezza	0

ESPERIMENTO – ALL-CHALLENGE-UPSAMPLED

Distanze per frame	59
Dimensione campione	3599
Dimensione train/test set	960/240
Nodi dei livelli interni	599 – 2404 – 1212
Tempo di train	43m 01s
Gallery totale	1200
Percentuale di accuratezza	0.1



CONCLUSIONS

- ◆ Many tasks have already been accomplished...
- ◆ ...but still many need to be done
- ◆ Drone independency
 - ◆ Stronger relationship with the server
 - ◆ Automatic profiling
- ◆ Improvement of the Frontal face image extraction (up to real time!)
- ◆ Going DEEPER with Face Dynamics



GRAZIE MILLE PER L'ATTENZIONE

