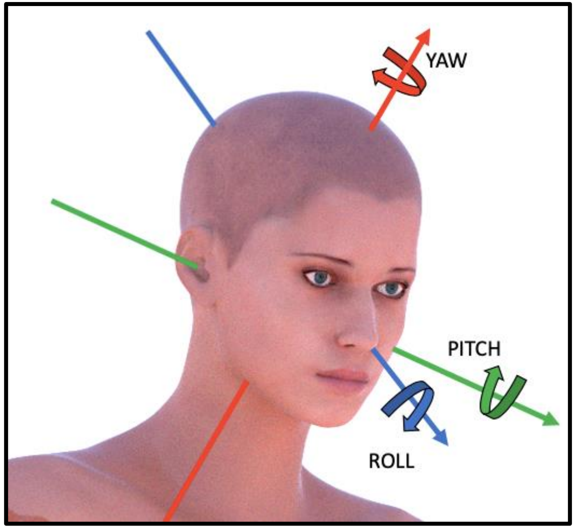
HP^2IFS: HEAD POSE ESTIMATION EXPLOITING PARTITIONED ITERATED FUNCTION SYSTEMS

Carmen Bisogni, Michele Nappi, Chiara Pero and Stefano Ricciardi

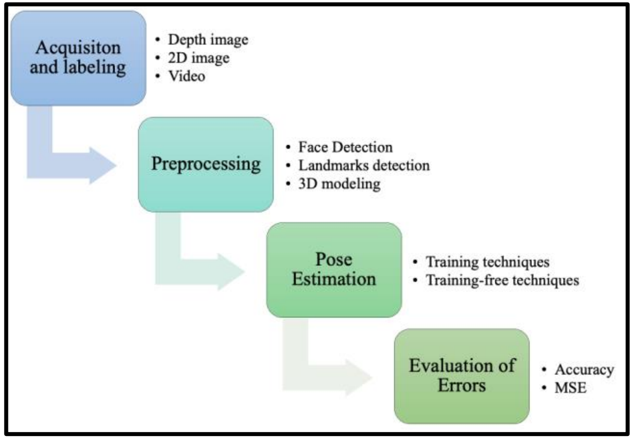
UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI INFORMATICA

UNIVERSITÀ DEGLI STUDI DI SALERNO

HPE Applications

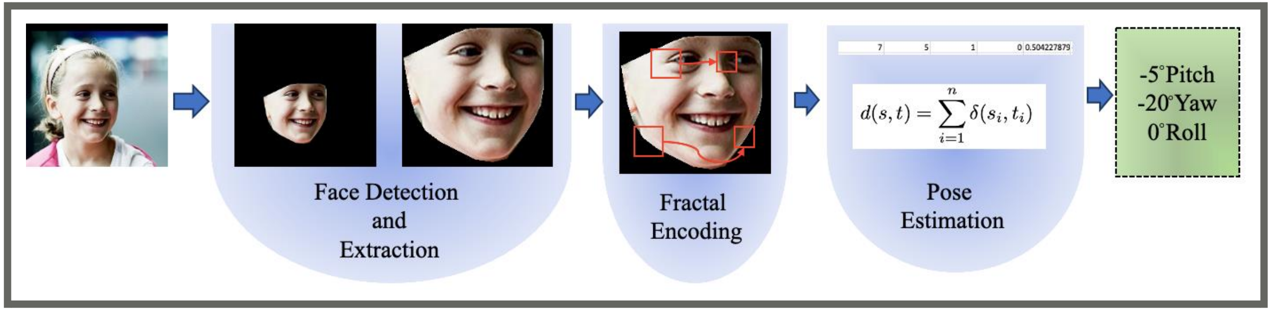


- To frontalize a face image in recognition.
- To detect subject's behaviour.
- To select the best frame in a video sequence.



HP^2IFS

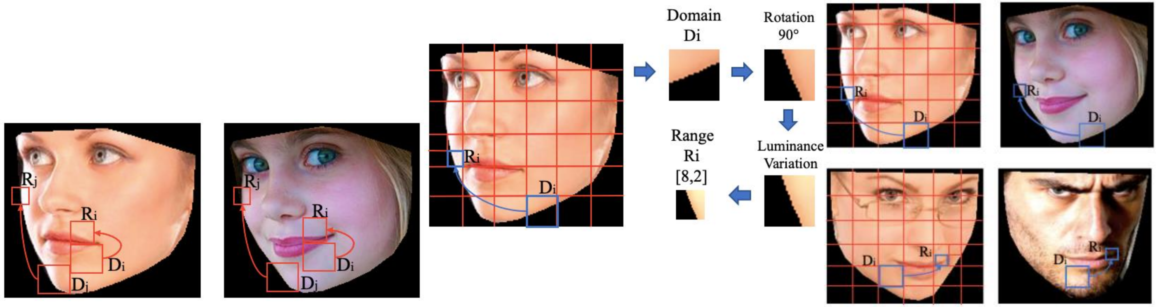
The Fractal Encoding approach proposed in HP2IFS is based on the face appearance.
The face is detected by a HoGs-based method and the landmarks by a pre-trained regression method.
The landmarks are then used to build a mask of the face.
This mask will be encoded using a fractal encoding algorithm. The obtained encoding is then compared with a reference model using the Hamming Distance. The reference model is built using part of the database involved in the tests.



Details

The fractal encoding is based on the concept of self-similarity and its step can be summarized as follow:

- Let M the metric space, $D_i \in$ M a collection of sub-domains and f_i a collection of contractive maps.
- The image to be encoded is partitioned in R_i , non overlapping range blocks.
- The image is then partitioned in larger non overlapping blocks D_j called domain blocks.
- For every range block R_i , a domain block D_{R_i} is found such that a contractive affine transformation f, transform this Domain block in a good approximation of the range block.

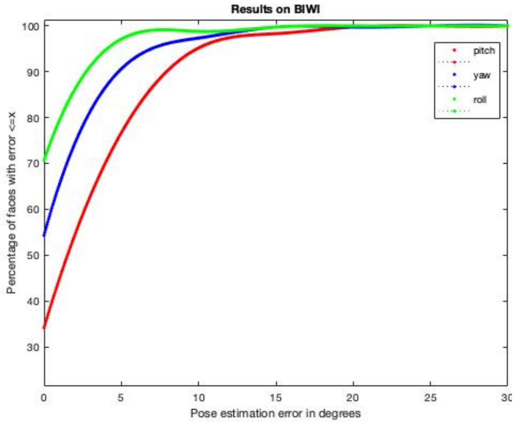


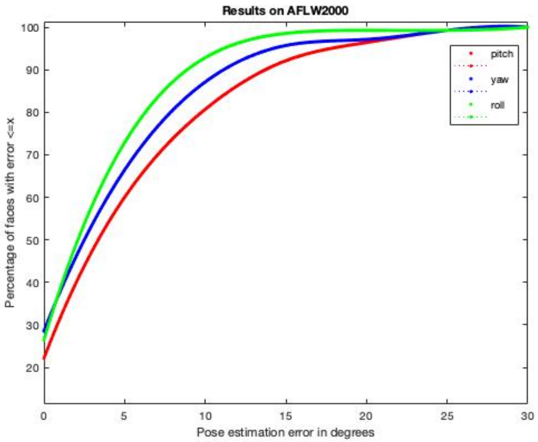
Results

Biwi AFLW2000

Dataset	E_yaw	E_pitch	E_roll	MAE
BIWI	4.05	6.23	3.30	4.52
AFLW2000	6.28	7.46	5.53	6.42

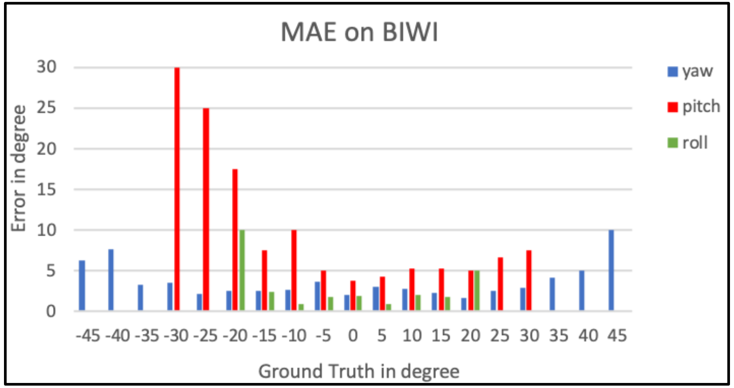
More than 95% of images have an error lower than 10° on BIWI.
More than 85% of images have an error lower than 10° on AFLW2000.

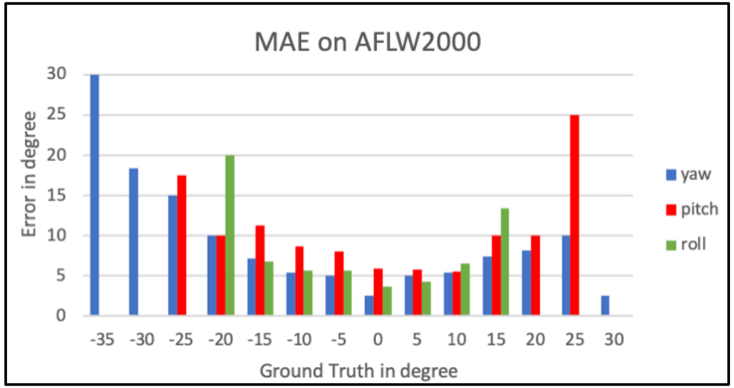




Results

Dependence of the errors compared to the angles ground truth:





Comparisons with the state-of-the-art

BIWI

Method	Yaw	Pitch	Roll	MAE
Multi-Loss ResNet50 [16]	5.17	6.97	3.39	5.177
GPR [15]	7.72	9.64	6.01	7.79
PLS [2]	7.35	7.87	6.11	7.11
SVR [17]	6.98	7.77	5.14	6.63
hGLLiM [7]	6.06	7.65	5.62	6.44
QT PYR [1]	5.41	12.80	6.33	8.18
FSA-Net [21]	4.27	4.96	2.76	3.996
Coarse-to-Fine [20]	4.76	5.48	4.29	4.84
QuatNet [10]	4.01	5.49	2.93	4.14
HP^2IFS	4.05	6.23	3.30	4.52

AFLW2000

Method	Yaw	Pitch	Roll	MAE
Multi-Loss ResNet50 [16]	6.470	6.559	5.436	6.155
Hyperface [14]	7.61	6.13	3.92	5.89
KEPLER [12]	6.45	5.85	8.75	7.01
3DDFA [24]	5.400	8.530	8.250	7.393
FAN [4]	6.358	12.277	8.714	9.116
QT PYR [1]	7.6	7.6	7.17	7.45
QuatNet [10]	3.973	5.615	3.92	4.503
HP^2IFS	6.28	7.46	5.53	6.42

Conclusions

The proposed approach is explicitly is aimed at achieving accurate angular values for yaw, pitch and roll axis without requiring any neural network architecture and the related training overhead.

Future improvements could involve the use of non-linear transformation in the Fractal Encodings or the study of different metrics to perform the distance between fractal encodings.

Carmen Bisogni, Michele Nappi, Chiara Pero and Stefano Ricciardi

Thank you

UNIVERSITÀ DEGLI STUDI DI SALERNO

UNIVERSITÀ DEGLI STUDI DI SALERNO
DIPARTIMENTO DI INFORMATICA