Using Sparse Coding For Landmark Localization In Facial Expressions

Vittorio Cuculo, Raffaella Lanzarotti, Giuseppe Boccignone
vittorio.cuculo@unimi.it, {lanzarotti,boccignone}@di.unimi.it

Problem

Landmark detection is an interesting and yet unsolved problem, especially when performed “in the wild”, dealing with variability in pose, race, gender, age, lighting, make-up, etc.

Solution: an unified model based on mixture of trees with a shared pool of parts, that realizes:

- face detection
- pose estimation
- landmark localization

adopting a local sparse coding representation, denoted as Histogram of Sparse Codes [1].

Mixture of trees

Local and global information is merged via a mixture of m trees of connected patches covering the landmarks of the face:

$$Q(x) = \sum_{k=1}^{m} \lambda_k T^k(x) \qquad (1)$$

with

$$\lambda_k \geq 0, \;\; k = 1, \ldots, m; \qquad \sum_{k=1}^{m} \lambda_k = 1$$

and its learning is solved via the search of the maximum likelihood tree adopting Chow-Liu.

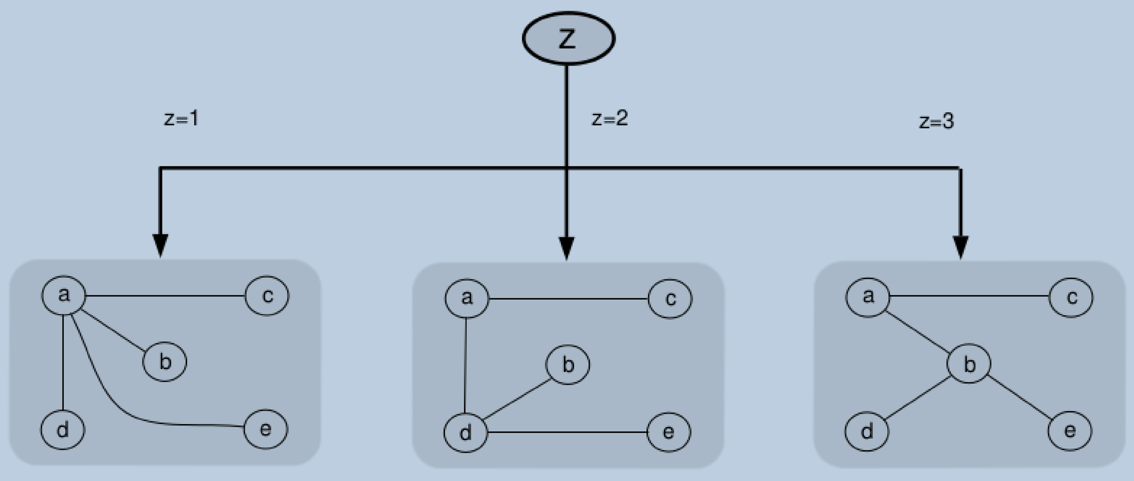

Fig.1: *Mixture of trees consisting of random variables, where z is a hidden choice variable.*

Sparse coding

General goal: given a matrix of image patches $\mathbf{X}$ and an overcomplete dictionary $\mathbf{W}$, find the associated sparse code matrix $\mathbf{Z}$ of a pre-defined K sparsity level.

$\Downarrow$

Rephrased as the minimization of the negative log-likelihood (NLL):

$$\min_{\mathbf{W},\mathbf{Z}} \{\|\mathbf{X} - \mathbf{W}\mathbf{Z}\|^2\} \quad \text{s.t.} \quad \forall i, \|\mathbf{z}_i\|_0 \leqslant K \qquad (2)$$

and solved applying the K-SVD algorithm [2].

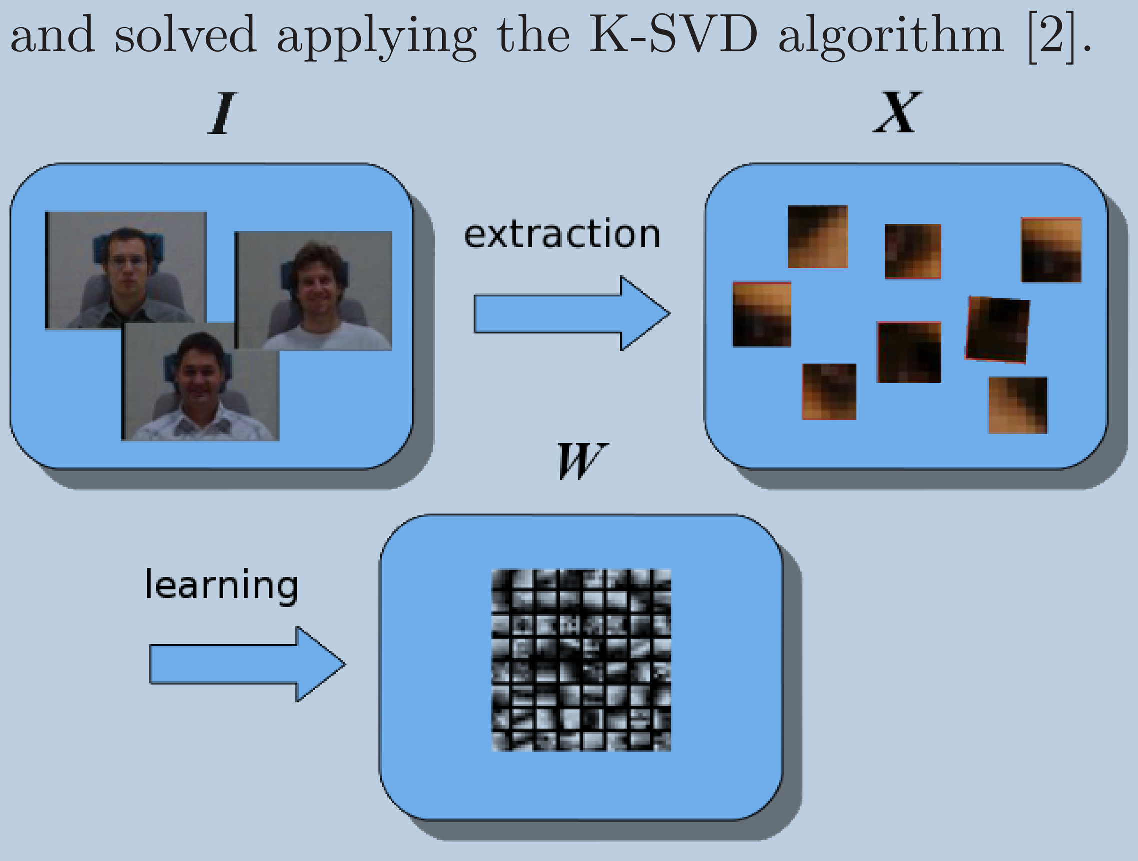

Fig.2: *Dictionary learning offline phase.*

Sparse coding framework

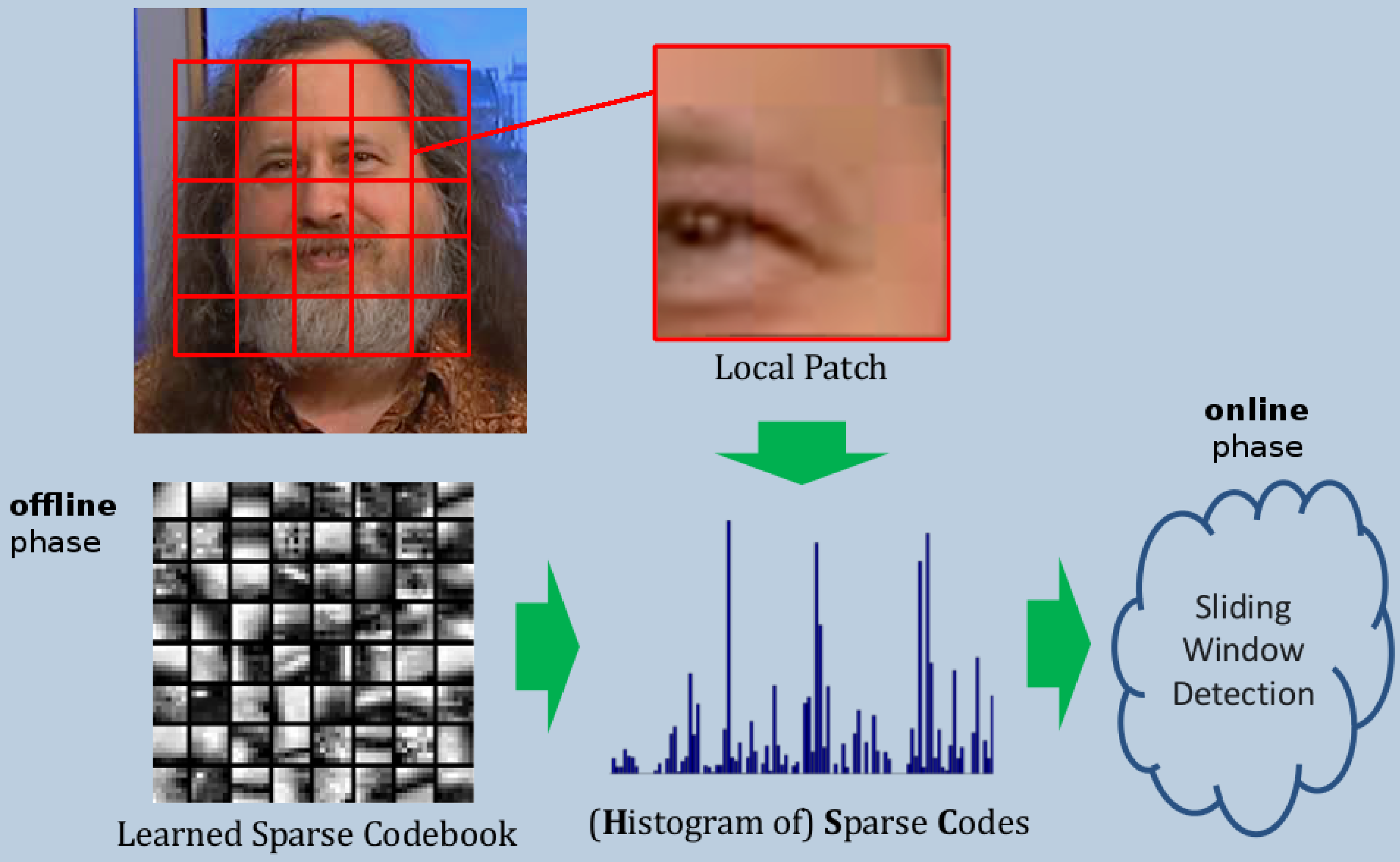

Fig.3: *Landmark localization framework.*

Once the dictionary $\mathbf{W}$ is learned, it is possible to use OMP to compute sparse codes at every location in an image. We adopt the Histograms of Sparse Codes (HSC) representation, a (semi-)dense feature vector F obtained averaging the sparse codes of patches in a 16×16 neighborhood, to sample the local response $\mathbf{f}^i$.

Experimental results

Experiments on a subset of CMU Multi-PIE database and comparison with Histogram of Gradients (HoG) feature descriptor, using 5×5 and 7×7 image patches and variable sparsity level $1 \leq T \leq 3$.

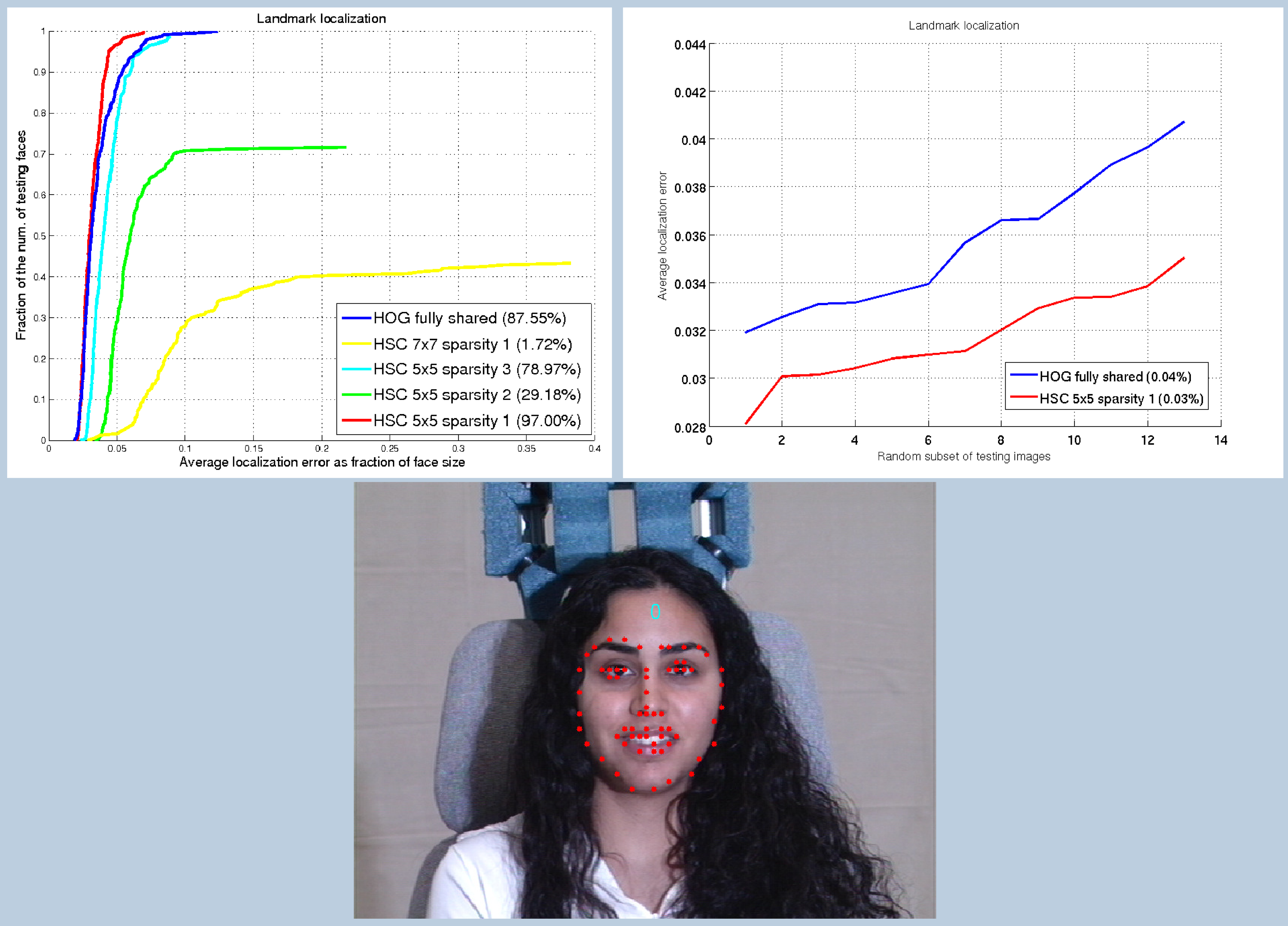

Fig.4: a) *Mean landmark localization error, in percentage of faces whose localization error is less than 5% of the face size.* b) *Sorted mean landmark localization error on random subset of images.* c) *Example of landmark localizations on non neutral expression.*

References

[1] X. Ren and D. Ramanan, "Histograms of sparse codes for object detection", in *IEEE Conf. Comput. Vision Pattern Anal. (CVPR)*. IEEE, 2013.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K -SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation", *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

Acknowledgements

The research was carried out as part of the project "Interpreting emotions: a computational tool integrating facial expressions and biosignals based shape analysis and bayesian networks", which is supported by the Italian Government, managed by MIUR, financed by the Future in Research Fund.